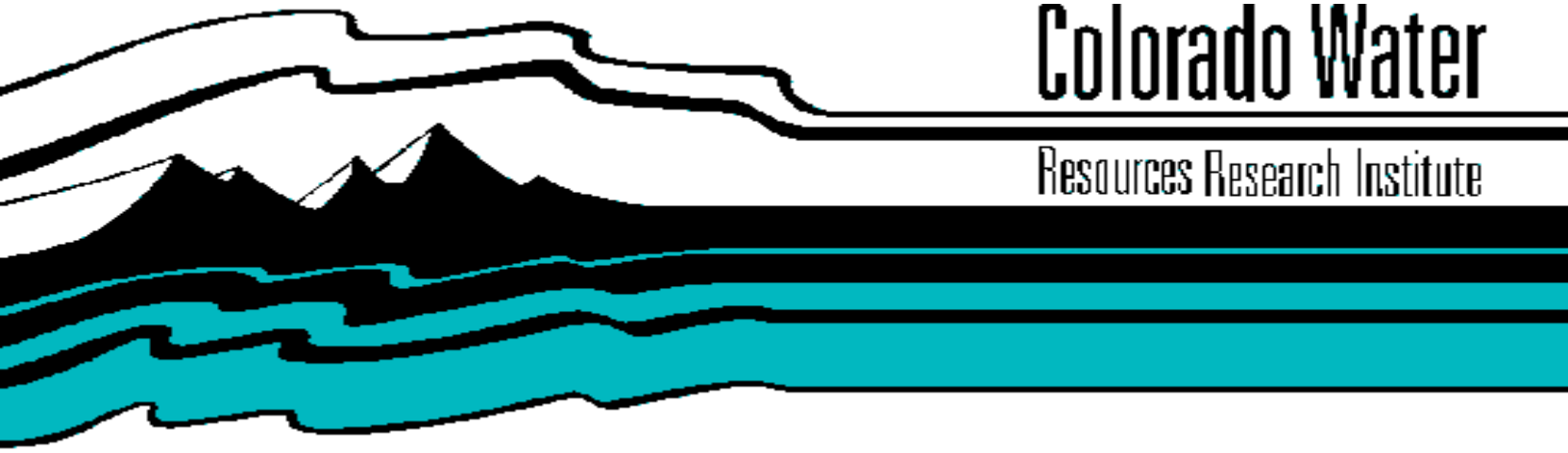
FORECASTING STREAMFLOW FOR COLORADO RIVER SYSTEMS

by

D.C. Wang and J.D. Salas



Completion Report No. 164

Colorado State University

FORECASTING STREAMFLOW FOR COLORADO RIVER SYSTEMS

by

D. C. Wang and J. D. Salas
Department of Civil Engineering
Colorado State University

December 1991

Grant N. 14-08-0001-G1551-01
Project N. 10
and
Grant N. 14-08-001-G1551-04
Project N. 07

The research on which this report is based was financed in part by the U.S. Department of the Interior, Geological Survey, through the Colorado Water Resources Research Institute; and the contents of this publication do not necessarily reflect the views and policies of the U.S. Department of the Interior, nor does mention of trade names or commercial products constitute their endorsement by the United States Government.

COLORADO WATER RESOURCES RESEARCH INSTITUTE
Colorado State University
Fort Collins, Colorado 80523

Robert C. Ward, Director

ABSTRACT

FORECASTING MONTHLY STREAMFLOW FOR COLORADO RIVER SYSTEMS

Forecasting water supply is critical for meeting the growing demands for in-state use of Colorado's water resources. The intensifying out-of-state demands for waters originating in the State of Colorado underscores the necessity for maximizing beneficial use of water within the state. The State of Colorado has 16 compacts related to water use with other states thus, Colorado's problem in this regard is on how to make the maximum use of our water resources while complying with compacts with other states. Likewise, operation of complex systems such as the Colorado-Big Thompson Project, for example, requires forecasts of inflows into the reservoirs during the spring and summer in order to plan supplemental releases to meet irrigation, domestic and hydropower demands. The research reported herein was aimed at improving current forecasting procedures used in the State of Colorado.

In this study, single input-single output and multiple input-single output periodic transfer function models have been developed for forecasting monthly streamflow. First, the monthly streamflow is deseasonalized and filtered by a periodic autoregressive (PAR) model. Then, a transfer function model, in which the deseasonalized snow water equivalent is the input, and the streamflow residual from the PAR model is the output, has been formulated. The building of the transfer function model has been carried out by using spectral analysis and the non-linear least squares method was used for parameter estimation. The models have been applied to forecast monthly flows of the Rio Grande watershed system in Southern Colorado. Tests and comparisons of the proposed forecast method were made versus multiple regression approaches currently used by agencies responsible for managing the Colorado water resources. It was shown that the single input-single output periodic transfer function model gives better monthly streamflow forecasts than the multiple regression approach. Likewise, the addition of more inputs in the transfer function model improves the forecasts. It is concluded that the forecasting approaches developed in this research may be useful for forecasting monthly flows in the Rio Grande system. It is expected that similar results would be obtained for other river systems in the state.

ACKNOWLEDGEMENT

Research leading to this paper has been supported by Projects G1551-01/10 and G1551-04/07 of the Colorado Water Resources Research Institute. Discussions held with Mr. H. Simpson and S. Rao of the Water Resources Division of the State Engineer Office of Colorado and Dr. J. Scott and co-workers of the Northern Colorado Water Conservancy District, Loveland, Colorado were very useful in this research.

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION	1-1
	1.1 General	1-1
	1.2 Objectives of the Research	1-4
II.	LITERATURE REVIEW	2-1
	2.1 General Remarks	2-1
	2.2 Univariate Time Series Modeling of Seasonal Streamflow	2-3
	2.3 Input-Output Modeling of Seasonal Streamflow	2-4
	2.4 Spectral Analysis	2-6
III.	BACKGROUND ANALYSIS AND MODELS	3-1
	3.1 General Remarks	3-1
	3.2 Deseasonalized Autoregressive Moving-Average Model	3-3
	3.2.1 Model Description	3-3
	3.2.2 Parameters Estimation, Diagnostic Checking and Model Selection	3-4
	3.2.3 Forecast Function	3-5
	3.3 Periodic Autoregressive Moving-Average Model	3-8
	3.3.1 Model Description	3-8
	3.3.2 Parameter Estimation and Diagnostic Checking	3-10
	3.3.3 Forecast Function	3-10
	3.4 Transfer Function Model	3-12
	3.4.1 Model Description	3-12
	3.4.2 The Difficulties of Modeling the Multiple Input-Multiple Output Transfer Function Model With the Time Domain Approach	3-15
	3.5 Nonlinear Least Squares Method	3-16
	3.6 The Forecasting of Log-Transformed Series	3-20
	3.7 Spectral Analysis	3-21
	3.7.1 Introduction	3-21
	3.7.2 Discrete Fourier Transform	3-22
	3.7.3 Power Spectrum and Cross Spectrum	3-23
	3.7.4 Periodogram Based Spectrum Estimation	3-25
	3.7.5 Covariance Based Spectrum Estimation	3-30
IV.	SINGLE INPUT-OUTPUT PERIODIC TRANSFER FUNCTION MODEL	4-1
	4.1 Introduction	4-1
	4.2 Model Description	4-1
	4.3 Model Identification	4-3
	4.4 Parameter Estimation	4-8

4.5 Diagnostic Checking 4-9
4.6 Forecast Function 4-10

V. MULTIPLE INPUT-SINGLE OUTPUT PERIODIC TRANSFER FUNCTION MODEL 5-1
5.1 Introduction 5-1
5.2 Multiple Input-Single Output Periodic Transfer Function Model 5-1
5.2.1 Model Identification 5-1
5.2.2 Parameter Estimation and Diagnostic Checking 5-9
5.2.3 Forecast Function 5-9

VI. DATA DESCRIPTION 6-1
6.1 Watershed Description 6-1
6.2 Monthly Streamflow Data 6-2
6.3 Snow Water Equivalent Data 6-3
6.3.1 The SWE of the Los Pinos River Basin 6-4
6.3.2 The SWE of the Rio Grande River Basin Above Del Norte . . 6-5
6.3.3 The SWE of the Conejos River Basin 6-6
6.4 Temperature and Precipitation Data 6-6
6.4.1 The Climatic Data of the Upper Rio Grande River Basin 6-7
6.4.2 The Climatic Data of the Conejos River Basin 6-7

VII. FORECASTING MONTHLY STREAMFLOW BASED ON THE SINGLE INPUT-SINGLE OUTPUT MODEL 7-1
7.1 Univariate Time Series Model 7-1
7.1.1 Deseasonalized ARMA Model 7-1
7.1.2 The PAR(1) Model 7-4
7.1.3 The PARMA(1,1) Model 7-7
7.1.4 Comment 7-9
7.2 Forecasting Snow Water Equivalent 7-10
7.3 Rational Single Input-Single Output Transfer Function Model 7-11
7.3.1 Log-Transformed Monthly Flow Modeling 7-11
7.3.2 No Transformed Monthly Flow Modeling 7-17
7.4 Single Input-Single Output Periodic Transfer Function Model 7-20
7.4.1 Transfer Function Model with PAR(1) Noise 7-20
7.4.2 Periodic-Transfer Function Model 7-22
7.5 Existing Forecasting Model 7-26
7.6 Comparison of the Forecast 7-33
7.6.1 The Los Pinos River Basin 7-34
7.6.2 The Conejos River Basin 7-39
7.6.3 The Rio Grande River Basin 7-46
7.6.4 Summary of the Forecasting Results 7-47

7.7 Forecasting with the Current Data Available 7-48
7.7.1 The Los Pinos River Basin 7-48
7.7.2 The Conejos River Basin 7-49
7.7.3 The Rio Grande River Basin 7-50
7.7.4 Summary of the Forecasting Results 7-51

VIII. FORECASTING BASED ON THE MULTIPLE INPUT-SINGLE OUTPUT MODEL 8-1
8.1 Introduction 8-1
8.2 Multiple Input-Single Output Forecasting 8-2
8.2.1 Introduction 8-2
8.2.2 Using SWE and Temperature to Forecast Streamflow 8-2
8.2.3 Using SWE and Precipitation to Forecast Streamflow 8-6
8.2.4 Using SWE, Temperature and Precipitation to Forecast Streamflow 8-8

IX. CONCLUSIONS 9-1

REFERENCES R-1

CHAPTER I

INTRODUCTION

1.1 General

Accurate water supply forecasting is critical for meeting the growing demands for in-state use of Colorado's water resources. The intensifying out-of-state demands for waters originating in our state underscore the necessity for maximizing beneficial use within Colorado. The State of Colorado has 16 compacts related to water use with other states. Thus, Colorado's problem in this regard is how to make the maximum use of our water resources while complying with compacts with other states. For instance, streamflow forecasts at the state line on the Rio Grande River are routinely made in order to administrate more rationally the water resources available in the Rio Grande basin. Reliable estimates of water available over the next day, week, month or season increases the likelihood of effectively matching these supplies with estimated demands.

Water commissioners under the Office of the State Engineer and various other water officials need accurate streamflow forecasts for filling reservoirs and delivering supplies with a minimum of waste, in accordance with water right seniority. Likewise, operation of complex systems such as the Colorado-Big Thompson Project, for example, requires forecasts of inflows into the reservoirs during the spring and summer in order to plan supplemental releases to meet irrigation, domestic and hydropower demands.

Most of the monthly streamflow in the Western United States originates as snowfall that has accumulated in the mountains through the late fall, winter, and early

spring. Forecasts of April through December aggregate (cumulative) runoff are generally made from April through December by the Water Resources Division of the State Engineer Office on the first of the month after the results of snow course surveys for the previous month becomes available. These total seasonal runoff forecasts are further divided into monthly values (the difference of two successive forecast values). These forecasts are needed for maximizing use of the water resources of the state. Likewise, shorter time interval forecasts of the order of 10 days or a week are usually needed for the same purposes. Local agencies such as the Northern Colorado Water Conservancy District (NCWCD) use such forecast information to plan their annual operations and to decide on quota allocations for supplemental water for irrigation and domestic uses of their clients. Likewise, certain state water right administrative agencies also find this information useful for obtaining more efficient use of available water resources within legal priorities.

Existing methods used for forecasting streamflow in the State of Colorado are based on the usual multiple regression analysis. This method has been traditionally used for a number of prediction and forecasting problems in water resources. However, it has a number of shortcomings as has been documented in the literature (Tabios and Salas, 1982; Haltiner and Salas, 1988). Research developed in the past 20 years or so at a number of universities, including Colorado State University, indicates that approaches based on more structured models such as the so-called transfer function and ARMAX models and threshold ARMAX models are better suited for forecasting purposes than the conventional regression models.

Advances in time series analysis (Box and Jenkins, 1976; Salas et al., 1980; Brockwell and Davis, 1987) offer significant advantages over the current procedure, such as the ability to identify the best model structure, more efficient updating of model parameters, and the possibility of determining forecast errors (confidence limits). The records of historical flows at a given site are analyzed and the underlying mechanism generating this system output is identified through time series analysis. Likewise, the system inputs such as snow-water equivalent and temperature can be analyzed and modeled. Then, a transfer function-noise model relating the inputs to output can be developed with consideration of the delay response time between the input occurrence and the significant output response.

The Rio Grande River system has been selected as the pilot project to implement and test the procedures to be developed in this study. The main reason for considering this river as the pilot system is because of its importance in agricultural development in the southern part of the state and because of the obligation that the state has to deliver water in the Rio Grande at the Colorado-New Mexico state line (Rio Grande Basin Compact). These characteristics make it extremely important for the state to forecast streamflow at key points along the Rio Grande River. More specifically, to comply with Rio Grande Compact the state needs to forecast streamflow at the Del Norte gaging station along the main stem of the river, at the Mogote gaging station along the Conejos River, Los Pinos near the Ortiz gaging station along the Los Pinos River, and at the San Antonio at Ortiz gaging station along the San Antonio River. Thus, the main thrust of the project has been to forecast seasonal streamflows at these four gaging stations.

1.2 Objectives of the Research

The specific objectives of the research are:

(a) To develop new methods for streamflow forecasting in selected systems of the State of Colorado. The Rio Grande River system was selected as the pilot case study to apply the new methods.

(b) To test and compare the forecasting methods to be developed here with approaches currently used by agencies responsible for administering Colorado's water resources.

(c) To document and make available to the public the models and procedures to be developed and the results obtained in the study.

CHAPTER II

LITERATURE REVIEW

2.1 General Remarks

In analyzing and modeling streamflow processes in places as in the State of Colorado, snowmelt runoff plays an important role. There are a number of driving meteorological variables for melting snow such as solar radiation, temperature, humidity and wind. Rainfall represents an immediate input to the runoff process. Snowfall has to turn into snowmelt before it becomes runoff. Liquid water in the snow pack, either from snowmelt or rainfall, must percolate through the pack before it can appear as runoff. The overall effect of snowpack is to produce a lag between the time which melt water is produced and the time it reaches the bottom of the snowpack. Then at the bottom of the snow pack, the liquid water can either enter the soil system or flow through the soil surface in a saturated zone. This produces another time lag before the water ultimately reaches the stream.

A number of researchers have developed precipitation-runoff models that apply to various types of watersheds. The models attempt to describe physically or conceptually the actual physical processes of the hydrologic cycle so as to simulate hydrologic events such as the evolution of streamflow hydrographs. While these models can be useful for determining runoff forecasts, they usually require extensive data for calibration and operation which often is unnecessary for runoff forecasts at the monthly and seasonal time scales. A simple conceptual model is the "Snowmelt Runoff Model" (SRM) which

has been used by Rango and Martinec (1979), Shafer et al. (1982) and Rango (1983). In this study we will not consider the physics of the snowmelt process. The system that traces precipitation from snowfall and from rainfall to streamflow will be considered as a black box.

A great number of methods have been studied in the area of seasonal streamflow forecasting. The multiple regression analysis method has been traditionally used in water resources. U.S. Soil Conservation Service (1970) forecasts the seasonal flow quantity using multiple regression techniques. These relate the runoff to base flow, fall precipitation, spring precipitation and the snow water equivalent. Tabios and Salas (1982) reported that improved forecasts could be obtained by applying Kalman filtering and spatial interpolation techniques. Lang (1986) forecasts the meltwater runoff from glacier basins by using multiple regression.

Stochastic methods have been also applied for streamflow forecasting. Although the seasonal autoregressive integrated moving average (SARIMA) model forecasts seasonal economic data well, this model may not be appropriate for modeling and forecasting monthly streamflow (McLeod et al., 1987). Other stochastic approaches have been used by hydrologists in forecasting streamflow series. One of the approaches is to analyze streamflow based on univariate time series models. Others use an input-output relationship for modeling the streamflow series. The literature review in this report includes three categories: (a) the modeling of seasonal streamflow by using univariate time series models, deseasonalized ARMA models and periodic ARMA processes; (b) the input-output models, such as the transfer function model, the ARMAX model and the

state space model; and (c) a brief overview of spectral analysis, a tool which is used in this research.

2.2 Univariate Time Series Modeling of Seasonal Streamflow

Time series are periodic when their statistical properties change periodically during the year. Before using the stationary ARMA model, the seasonal streamflow should remove the seasonality by subtracting the seasonal mean and dividing by the seasonal standard deviation (Salas et al., 1980). In this way, the streamflow is transformed into a zero-mean and unit-variance series. Then, the deseasonalized series may be analyzed by using the stationary ARMA process if this is applicable. The ARMA process can be used to forecast streamflow from current and past values. Most deseasonalized ARMA models used for data generation preserve the underlying historical statistical properties, but when it is used for forecasting monthly flows, the ARMA model underestimates the high flow in the summer time. The deseasonalized ARMA process was used for forecasting monthly streamflow by Noakes et al. (1985). The result of this study concluded that the deseasonalized ARMA processes does not give an accurate forecast to streamflow.

It is known that streamflow series exhibit an autocorrelation structure that depends not only on the time lag between observations but also on the season of the year. Thomas and Fiering (1962) suggested an AR(1) model with periodic coefficients that may be used to fit a time series presenting lag-one correlations. Yevjevich (1971) suggested the AR(p) models with periodic parameters for modeling the seasonal hydrologic time series. Salas and Yevjevich (1972) derived moment estimates of periodic parameters. Salas (1972)

derived Yule-Walker equations for these models. Models similar to the family of PAR models have been employed by Clarke (1973). Delleur et al. (1976) gave approximate parameter estimates of the periodic MA(1), MA(2) and ARMA(1,1) models. These were patterned after the Yule-Walker equations for the ARMA(p,q) process which assumes constant parameters. A Yule-Walker equation of the periodic ARMA(p,q) process was derived by Salas et al. (1982). They showed that the periodic autoregressive parameters of ARMA(p,1) models can be computed by solving a system of linear equations, and that the periodic moving average parameters can be solved iteratively.

An exact likelihood function of the PARMA(1,1) model was developed by Vecchia (1983), and an algorithm for computing the approximate maximum likelihood estimates is presented. Thompstone (1983) formulated a parsimonious periodic autoregressive model, which groups seasons that present similar autoregressive characteristics in order to consider a single AR model for those seasons. Noakes et al. (1985) used the PAR model in forecasting the monthly streamflow. They concluded that the PAR model provided the most accurate forecast.

2.3 Input-Output Modeling of Seasonal Streamflow

A number of papers dealing with input-output modeling, state-space approaches and Kalman filtering estimation are included in Chiu (1978). Mizumura and Chiu (1985) using a combination of the tanks and autoregressive models, identified parameters using the Kalman filtering technique for forecasting combined snowmelt and rainfall during the snowmelt period. Burn and McBean (1985) used a different technique that incorporates the Kalman filtering technique to reflect uncertainty in the measured data and model

errors for forecasting the daily streamflow resulting from the combined snowmelt and rainfall.

Some papers deal with the distribution of the forecasted runoff volume. Hoshi and Burges (1980) suggested a method that can forecast the seasonal streamflow volumes conditional on the forecasted total runoff volume. Given an imperfect forecast of the total runoff volume for a season, the distribution of the forecasted runoff volume in each subperiod was studied by Pei et al. (1987). They developed a method for finding the conditional distribution of runoff volume in each subperiod, given the total forecast. Krzysztofowicz and Watada (1986) described the uncertainties in the forecast of seasonal snowmelt runoff volume by using a discrete-time, finite, continuous-space, nonstationary Markov process. They pointed out that Bayesian forecasts offer a more explicit and complete assessment of forecast uncertainties. This assessment can be used for optimal decision making.

In recent years, some input-output models for streamflow forecasting have been developed. Although these models have relatively simple formulation and data requirements, they can be combined with modern filtering techniques to provide accurate forecasts. These models are based on the need to formulate operational relationships between input and output with acceptable accuracy without regard to the physics of nature. Several black box approaches were studied during the 1960's. Prasad (1967) presented a non-linear model to handle the nonlinearity inherent in the rainfall-runoff process. The publication of the book on time series by Box and Jenkins (1970) summarized and formalized the time series method, representing the beginning of a

period of increasing interest and research in the application of time series. The transfer function model presented by Box and Jenkins has become popular. Cooper and Wood (1980; 1982a; 1982b) suggested a variation of Akaikie's canonical correlation procedure to determine model order of the autoregressive moving average with exogenous input (ARMAX) model. The ARMAX model needs a good initial parameter estimates for the recursive parameter estimation. Haltiner and Salas (1988) use the ARMAX model for forecasting the daily riverflow resulting from combined snowmelt and rainfall.

The application of the single input-single output transfer function model to the Arctic data was studied by Baracos et al. (1981). The use of a transfer function model in relating monthly river flows with precipitation and temperature was presented by Hipel et al. (1982). Nicklin (1985) use the transfer function model for modeling irrigation return flow problem. The forecasting of quarter-monthly riverflows by using a transfer function model was also presented by Thompstone et al. (1985).

2.4 Spectral Analysis

Spectrum describes how the variation in a time series may be accounted for by cyclic components at different frequencies. The procedure for estimating the spectrum is called spectral analysis. Spectral analysis is mainly concerned with estimating the spectrum over a whole range of frequencies. Bartlett (1966) and Tukey (1967) have been prominent in the development of modern spectral analysis, and their techniques are now widely used. There are a number of references which are useful for studying the frequency domain approach. Jenkins and Watts (1968), Koopmans (1974), Bloomfield (1976), Otnes and Enochson (1978) and Chatfield (1980) provide elementary descriptions

of spectral analysis. Brillinger (1981) provides a classic study on the advanced level. Priestley (1981) and Shumway (1988) gave a comprehensive treatment of spectral methods for an intermediate level. Yevjevich (1972) used the spectrum to analyze hydrologic time series. Canfield (1982), Canfield and Bowles (1985) suggest the generation of the autocorrelation function by generating the Fourier coefficients. Since the spectral density function and the autocorrelation function are equivalent representations of the serial association characteristics of a time series, this can be achieved. Padmanabhan and Rao (1988) applied a maximum entropy spectral analysis to find the periodicities of hydrologic data.

CHAPTER III

BACKGROUND ANALYSIS AND MODELS

3.1 General Remarks

Hydrological time series has been successfully used for forecasting and data generation studies. Some of the stochastic models that will be used for modeling the seasonal hydrological time series are: (a) deseasonalized autoregressive moving-average (ARMA) model, (b) periodic autoregressive moving-average (PARMA) model, and (c) transfer function model.

Among these three models, the transfer function model will become the main subject of this research. In addition to the time domain time series analysis, the frequency domain time series analysis is also considered here. Spectral analysis is useful, since it can simplify the analysis of time invariant linear systems. Spectral analysis is a non-parametric approach, which is useful as an exploratory tool for suggesting models which can be fitted parametrically.

First, the basic statistical characteristics of time series which will be needed in this research are described. Let x_t, $t=1,\ldots,N$ be a stationary time series, in which N is the sample size. The population and sample statistical characteristics are denoted as:

(1) The mean

$$\mu_x = E[x_t]. \tag{3.1.1}$$

The sample mean

$$\bar{x} = \frac{1}{N}\sum_{t=1}^{N} x_t. \tag{3.1.2}$$

(2) The variance

$$\sigma^2 = E[(x_t - \mu)^2]. \tag{3.1.3}$$

The sample variance

$$s_x^2 = \frac{1}{N-1}\sum_{t=1}^{N} (x_t - \bar{x})^2. \tag{3.1.4}$$

(3) The autocovariance function

$$\gamma_k = Cov(x_{t+k}, x_t) = E[(x_{t+k} - \mu_x)(x_t - \mu_x)]. \tag{3.1.5}$$

The sample autocovariance

$$\hat{\gamma} = \frac{1}{N}\sum_{t=1}^{N-k} (x_{t+k} - \bar{x})(x_t - \bar{x}). \tag{3.1.6}$$

(4) The autocorrelation function

$$\rho_k = \frac{\gamma_k}{\gamma_0}. \tag{3.1.7}$$

The sample autocorrelation function

$$\hat{\rho}_k = \frac{\sum_{t=1}^{N-k} (x_{t+k} - \bar{x})(x_t - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2}. \tag{3.1.8}$$

(5) The cross-covariance function

$$\gamma_{xy}(k) = Cov(x_{t+k}, y_t) = E[(x_{t+k} - \mu_x)(y_t - \mu_y)]. \tag{3.1.9}$$

The sample cross-covariance function

$$\hat{\gamma}_{xy}(k) = \begin{cases} \dfrac{\sum_{t=1}^{N-k} (x_{t+k} - \bar{x})(y_t - \bar{y})}{N} & \text{for } k=0,\ 1,\ 2,... \\ \dfrac{\sum_{t=1-k}^{N} (x_{t+k} - \bar{x})(y_t - \bar{y})}{N} & \text{for } k=0,-1,-2,... \end{cases} \tag{3.1.10}$$

(6) The cross-correlation function

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{\sqrt{\gamma_x(0)\,\gamma_y(0)}}. \tag{3.1.11}$$

The sample cross-correlation function

$$\hat{\rho}_{xy}(k) = \frac{\hat{\gamma}_{xy}(k)}{\sqrt{s_x^2\ s_y^2}}. \tag{3.1.12}$$

3.2 Deseasonalized Autoregressive Moving-Average Model

3.2.1 Model Description

One approach for modeling seasonal hydrological data is to deseasonalize the series and fit an ARMA process to the deseasonalized data (see Salas et al., 1980). Let us assume that $y_{v,\tau}$ represents a periodic hydrologic time series, where v is the year and τ is the month within the year. Then, the series $y_{v,\tau}$ can be deseasonalized by

$$z_{v,\tau} = \frac{y_{v,\tau} - \mu_\tau}{\sigma_\tau}, \tag{3.2.1}$$

where μ_τ and σ_τ are the periodic mean and periodic standard deviation, which can be

estimated by

$$\hat{\mu}_\tau = \frac{\sum_{v=1}^{N} x_{v,\tau}}{N}, \tag{3.2.2}$$

$$\hat{\sigma}_\tau = \sqrt{\frac{\sum_{v=1}^{N} (x_{v,\tau} - \hat{\mu}_\tau)^2}{N-1}}, \tag{3.2.3}$$

where N is the total number of years of record.

Now, the $z_{v,\tau}$ series may be represented by a non-seasonal ARMA model. Consider the series $z_{v,\tau}$ of Eq. (3.2.1) to be represented by z_t with $t=(v-1)\cdot 12+\tau$. Then, an ARMA(p,q) process can be written as

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \tag{3.2.4}$$

where ε_t is a white noise with mean zero and variance σ_ε^2; and ϕ_1, ϕ_2,..., ϕ_p and θ_1, θ_2,..., θ_q are the other parameters of the model.

For hydrologic series, the orders p and q are usually small, i.e., of the order of one or two (Salas et al., 1980; p. 125), thus low order models such as ARMA(1,0) ARMA(2,0), ARMA(0,1), ARMA(0,2) and ARMA(1,1) model are usually popular for most practical applications.

3.2.2 Parameters Estimation, Diagnostic Checking and Model Selection

Estimation of stationary ARMA models by the method of moments, method of least squares and method of maximum likelihood are well understood (Box and Jenkins, 1976; Salas et al., 1980; Brockwell and Davis, 1987). The method of least squares is an approximation of the method of maximum likelihood. It can be implemented by

minimizing the sum of squares of the residuals

$$S(\underline{\phi},\underline{\theta}) = \sum (\varepsilon_t)^2 \tag{3.2.5}$$

The sum of the squares of the residuals depends on the parameters $\underline{\phi}$ and $\underline{\theta}$ and the starting values of the ε's. Therefore, the set of parameters $\hat{\underline{\phi}}$ and $\hat{\underline{\theta}}$ which minimizes the sum of the squares function is sought. At last, the independence and normality of the residuals should be checked.

Among competing ARMA(p,q) models, the correct model can be determined by using the Akaike Information Criterion (Akaike, 1974; Salas et al, 1980)

$$AIC(p,q) = N \ln(\hat{\sigma}_\varepsilon^2) + 2(p+q) \tag{3.2.6}$$

where N is the sample size and $\hat{\sigma}_\varepsilon^2$ is the maximum likelihood estimate of the residual variance.

3.2.3 Forecast Function

The ARMA(p,q) model of Eq. (3.2.4)

$$z_t = \phi_1 z_{t-1} + \ldots + \phi_p z_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

can be written as

$$(1 - \phi_1 B - \ldots - \phi_p B^p) z_t = (1 - \theta_1 B - \ldots - \theta_q B^q) \varepsilon_t$$

where

$$\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p \tag{3.2.7a}$$

$$\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q \tag{3.2.7b}$$

and B is the backward shift operator, defined by $B^j z_t = z_{t-j}$. Then, Eq. (3.2.4) can be written as

$$z_t = \frac{\theta(B)}{\phi(B)}\varepsilon_t = \psi(B)\varepsilon_t \tag{3.2.8a}$$

or

$$z_t = \psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \cdots\cdots \tag{3.2.8b}$$

where $\Psi(B) = \Psi_0 + \Psi_1 B + \Psi_2 B^2 + \cdots$.

Consider now that z_{t+L} is the process defined at time t+L which means a lead time of L units after the current time t. Then, from Eq. (3.2.8b) the forecasting function can be expressed by

$$\begin{aligned} z_{t+L} &= \sum_{j=0}^{\infty} \psi_j \varepsilon_{t+L-j} \\ &= (\psi_0\varepsilon_{t+L} + \psi_1\varepsilon_{t+L-1} + \cdots + \psi_{L-1}\varepsilon_{t+1}) + (\psi_L\varepsilon_t + \psi_{L+1}\varepsilon_{t-1} + \cdots) \\ &= \sum_{j=0}^{L-1} \psi_j \varepsilon_{t+L-j} + \sum_{j=0}^{\infty} \psi_{L+j}\varepsilon_{t-j} \end{aligned} \tag{3.2.9}$$

Now suppose, standing at present time t, we want to make a forecast of z_{t+L}, that is we want to forecast z at L time steps ahead given that we know the information of z up to present time t. This mathematically can be represented by the conditional expectation of z_{t+L} (for $L \geq 1$) given that z_t, z_{t-1}, z_{t-2},... are known. Let us denote by $\hat{z}(L)$ such conditional expectation, i.e.

$$\hat{z}_t(L) = E[z_{t+L} | z_t, z_{t-1}, z_{t-2}, \ldots], \quad L \geq 1 \tag{3.2.10}$$

Then by using Eq. (3.2.9), we can assume that

$$\hat{z}_t(L) = \sum_{j=0}^{\infty} \psi^*_{L+j}\varepsilon_{t-j} \tag{3.2.11}$$

where the weights Ψ_L^*, Ψ_{L+1}^*, ... are to be determined.

From Eqs. (3.2.9) and (3.2.11), we have

$$z_{t+L}-\hat{z}_t(L)=\sum_{j=0}^{L-1}\psi_j\varepsilon_{t+L-j}+\sum_{j=0}^{\infty}(\psi_{L+j}-\psi_{L+j}^*)\varepsilon_{t-j} \tag{3.2.12}$$

which is the error of the forecast. The mean square error of the forecast becomes

$$E[z_{t+L}-\hat{z}_t(L)]^2 = \sigma_\varepsilon^2\sum_{j=0}^{L-1}\psi_j^2 + \sigma_\varepsilon^2\sum_{j=0}^{\infty}(\psi_{L+j}-\psi_{L+j}^*)^2 \tag{3.2.13}$$

which is minimized only if $\Psi_{L+j}^* = \Psi_{L+j}$. Thus the minimum mean square error forecast $\hat{z}_t(L)$ of z_{t+L} at origin t is given by the conditional expectation of z_t at t. This expectation is conditional on knowledge of the series from infinite past up to the present origin t.

In addition, let us denote $E[z_{t+L}|z_t, z_{t-1}, \ldots]$, the conditional expectation of z_{t+L} given knowledge of all the z's up to time t by $[z_{t+L}]$. From Eq. (3.2.4), the lead time L forecast equation can be written as

$$\begin{aligned}\hat{z}_t(L) = {} & \phi_1[z_{t+L-1}] + \ldots + \phi_p[z_{t+L-p}] - [\varepsilon_{t+L}] \\ & - \theta_1[\varepsilon_{t+L-1}] - \ldots - \theta_q[\varepsilon_{t+L-q}]\end{aligned} \tag{3.2.14}$$

where

$$[z_{t+L}] = E[z_{t+L}|z_t,z_{t-1},\ldots.]$$

$$[\varepsilon_{t+L}] = E[\varepsilon_{t+L}|\varepsilon_t,\varepsilon_{t-1},\ldots.]$$

$$[z_{t+j}] = \begin{cases} z_{t+j} & j\le 0 \\ \hat{z}_t(j) & j>0 \end{cases}$$

$$[\varepsilon_{t+j}] = \begin{cases} \varepsilon_{t+j} & j \le 0 \\ 0 & j > 0 \end{cases}$$

From Eq. (3.2.12), the forecast error for lead time L may be written as

$$e_t(L) = \sum_{j=0}^{L-1} \psi_j \varepsilon_{t+L-j} \tag{3.2.15}$$

Since $E[e_t(L)]=0$, the forecast error is unbiased. Also, the variance of the forecast error of Eq. (3.2.15) can be written as

$$\hat{\sigma}_e^2(L) = Var\left[\sum_{j=0}^{L-1} \psi_j \varepsilon_{t+L-j}\right]$$

$$= \sigma_\varepsilon^2 \sum_{j=0}^{L-1} \psi_j^2 \tag{3.2.16}$$

If the forecast error is normal distributed, the 95% probability limits of forecast with lead time L are given by

$$z_{t+L}(\pm) = \hat{z}_t(L) \pm 1.96\,\hat{\sigma}_e(L) \tag{3.2.17}$$

3.3 Periodic Autoregressive Moving-Average Model

3.3.1 Model Description

Periodic hydrologic series such as monthly flow series, generally have a correlation structure which depends not only on the time lag between observations but also depends on the season. The periodic autoregressive moving average model can represent this seasonally-varying correlation structure. In here, the PAR(1) and PARMA(1,1) are included as examples. The parameter estimation will follow the procedures suggested by Salas et al. (1980; 1982).

The PAR(1) Model

Let us assume that $z_{v,\tau}$ represents a periodic time series, where v is the year and τ is the season within the year. Then a PAR(1) model is defined as

$$z_{v,\tau} = \varphi_{1,\tau} z_{v,\tau-1} + \varepsilon_{v,\tau} \tag{3.3.1}$$

where $\varepsilon_{v,\tau}$ is uncorrelated, has zero mean and has variance $\sigma_\tau^2(\varepsilon)$. The moment estimation of the parameters can be written as

$$\hat{\varphi}_{1,\tau} = \hat{\rho}_{1,\tau} \tag{3.3.2}$$

and

$$\sigma_\tau^2(\varepsilon) = 1 - \hat{\varphi}_{1,\tau}^2 \tag{3.3.3}$$

where $\rho_{1,\tau}$ is the seasonal correlation, which can be estimated by

$$\hat{\rho}_{k,\tau} = \frac{\frac{1}{N}\sum_{v=1}^{N} (z_{v,\tau} - \bar{z}_\tau)(z_{v,\tau-k} - \bar{z}_{\tau-k})}{s_\tau \; s_{\tau-k}}. \tag{3.3.4}$$

in which $\bar{z}_t$ and s_τ represent the mean and standard deviation of the z's for season τ.

The PARMA(1,1) Model

The PARMA(1,1) model is defined as

$$z_{v,\tau} = \varphi_{1,\tau} z_{v,\tau-1} + \varepsilon_{v,\tau} - \theta_{1,\tau} \varepsilon_{v,\tau-1}. \tag{3.3.5}$$

The moment estimates of $\varphi_{1,\tau}$ and $\theta_{1,\tau}$ can be obtained by solving the equations (see Salas et al., 1982)

$$\varphi_{1,\tau} = \frac{\rho_{2,\tau}}{\rho_{1,\tau-1}} \quad for \; \tau = 1,..,\omega \tag{3.3.6}$$

and

$$\theta_{1,\tau} = \varphi_{1,\tau} + \frac{1-\varphi_{1,\tau}\rho_{1,\tau}}{\varphi_{1,\tau}-\rho_{1,\tau}} - \frac{\varphi_{1,\tau+1}-\rho_{1,\tau+1}}{(\varphi_{1,\tau}-\rho_{1,\tau})\ \theta_{1,\tau+1}}, \tag{3.3.7}$$

where $\tau=1,\ldots,\omega$, and ω is the number of seasons in the year. Once $\varphi_{1,\tau}$ and $\theta_{1,\tau}$ are determined, the noise variances $\sigma_\tau^2(\varepsilon)$ can be estimated by solving the equation

$$\varphi_{1,\tau} - \theta_{1,\tau}\ \sigma^2_{\tau-1}(\varepsilon) = \rho_{1,\tau}, \tag{3.3.8}$$

for $\tau = 1,\ldots, \omega$.

3.3.2 Parameter Estimation and Diagnostic Checking

While the previous equations provide the moment estimates of model parameters, least squares estimation procedure will be used here. For this purpose, the Marquardt algorithm will be used for minimizing the sum of the squares function by using the moment estimates as initial values. The adequacy of the PARMA model can be checked by examining the independence of the residuals.

3.3.3 Forecast Function

Since the forecast function of the PAR(1) model can be derived from the forecast function of the PARMA(1,1) model, only the later function is presented here.

A PARMA(1,1) model can be written as

$$z_{\nu,\tau} = \frac{1-\theta_{1,\tau}B}{1-\varphi_{1,\tau}B}\varepsilon_{\nu,\tau} = \psi_\tau(B)\varepsilon_{\nu,\tau} \tag{3.3.9}$$

where

$$\psi_\tau(B) = \psi_{0,\tau} + \psi_{1,\tau}B + \psi_{2,\tau}B^2 + \ldots$$

Then, the forecasting model can be written as

$$z_{v,\tau+L}=\sum_{j=0}^{\infty}\psi_{j,\tau+L}\varepsilon_{v,\tau+L-j}$$

$$=\sum_{j=0}^{L-1}\psi_{j,\tau+L}\varepsilon_{v,\tau+L-j}+\sum_{j=0}^{\infty}\psi_{L+j,\tau+L}\varepsilon_{v,\tau-j}. \qquad (3.3.10)$$

By using the same method of Section 3.2.3, we know that the minimum mean square error forecast $\hat{z}_{v,\tau}(L)$ of $z_{v,\tau+L}$ made at current time τ can be written as

$$\hat{z}_{v,\tau}(L)=\varphi_{1,\tau+L}[z_{v,\tau+L-1}]-[\varepsilon_{v,\tau+L}]-\theta_{1,\tau+L}[\varepsilon_{v,\tau+L-1}] \qquad (3.3.11)$$

where

$$[z_{v,\tau+L}] = E[z_{v,\tau+L}|z_{v,\tau},z_{v,\tau-1},....]$$

$$[\varepsilon_{v,\tau+L}] = E[\varepsilon_{v,\tau+L}|\varepsilon_{v,\tau},\varepsilon_{v,\tau-1},....]$$

and

$$[z_{v,\tau+j}] = \begin{cases} z_{v,\tau+j} & j\le 0 \\ \hat{z}_{v,\tau}(j) & j>0 \end{cases}$$

$$[\varepsilon_{v,\tau+j}] = \begin{cases} \varepsilon_{v,\tau+j} & j\le 0 \\ 0 & j>0 \end{cases}$$

As in Eq. (3.2.15), the forecast error for lead time L is a linear function of future noises $\varepsilon_{v,\tau+1}$, $\varepsilon_{v,\tau+2}$,..... Thus,

$$e_{v,\tau}(L)=\sum_{j=0}^{L-1}\psi_{j,\tau+L}\varepsilon_{v,\tau+L-j}. \qquad (3.3.12)$$

Likewise, $E[e_{v,\tau}(L)]=0$ and the variance of the forecast error can be written as

$$\hat{\sigma}_{e_\tau}(L) = \sum_{j=0}^{L-1} \psi_{j,\tau+L}^2 \; \sigma_{\tau+L-j}^2(\varepsilon) \tag{3.3.13}$$

3.4 Transfer Function Model

3.4.1 Model Description

The transfer function model is a model that forecasts future values of a given time series on the basis of current and past values of the given series and current and past values of other time series related to it (Bowerman and O'Connell, 1987).

Suppose that x_t and y_t series are, respectively, the input and output of a dynamic system. Then, a transfer function model can be written as

$$y_t = \sum_{j=0}^{\infty} a_j x_{t-j} + N_t = A(B)x_t + N_t \tag{3.4.1}$$

where $A(B) = a_0 + a_1B + a_2B^2 + \ldots$ is the transfer function, a_j, $j=0,1,..$ is the impulse response function of the system, and N_t is a zero mean stationary process, uncorrelated with the input series x_t.

Assuming that x_t and y_t are stationary time series with zero mean, then a systematic identification procedure can be implemented (see Box and Jenkins, 1970; Brockwell and Davis, 1987). First, we need to prewhiten the input series x_t. For instance, an ARMA model can be fitted as

$$\phi'(B)x_t = \theta'(B)z_t \tag{3.4.2a}$$

or

$$\pi'(B)x_t = z_t \tag{3.4.2b}$$

where

$$\phi'(B) = 1 - \phi_1' B - .... - \phi_p' B^p$$
$$\theta'(B) = 1 - \theta_1' B - .... - \theta_q' B^q$$
$$\pi'(B) = \pi_0' + \pi_1' B + + \pi_2' B^2 + ...$$

and z_t has zero mean and variance σ_z^2. Applying the operator $\pi'(B)$ to each side of Eq. (3.4.1), we have

$$\pi'(B) y_t = A(B)\pi'(B) x_t + \pi'(B) N_t$$

or

$$q_t = A(B) z_t + N_t' = \sum_{j=0}^{\infty} a_j z_{t-j} + N_t' \qquad (3.4.3)$$

where $\{N_t'\}$ is a zero-mean stationary process, uncorrelated with $\{z_t\}$.

By multiplying each side of Eq. (3.4.3) by z_{t-k} and taking expectations, we have

$$E[q_t z_{t-k}] = \sum_{j=0}^{\infty} a_j E[z_{t-j} z_{t-k}] + E[N_t' z_{t-k}]$$
$$Cov[q_t z_{t-k}] = a_k \sigma_z^2$$

or

$$a_k = \frac{\rho_{qz}(k)\,\sigma_q \sigma_z}{\sigma_z^2} = \rho_{qz}(k)\frac{\sigma_q}{\sigma_z} \qquad (3.4.4)$$

where $\rho_{qz}(k)$ is the cross-correlation coefficient of q_t and z_t.

Equation (3.4.4) gives a preliminary estimate of the impulse response function a_k for lag k. Box and Jenkins (1976) recommend choosing A(B) as the ratio of two polynomials, i.e.

$$A(B) = \frac{\omega(B)B^{b}}{\delta(B)}, \tag{3.4.5}$$

where $\omega(B) = \omega_0 - \omega_1 B - \ldots - \omega_s B^s$, $\delta(B) = 1 - \delta_1 B - \ldots - \delta_r B^r$ and b is a delay parameter. By observing the shape of the cross correlation function $\rho_{qz}(k)$, and using the properties of $\rho_{qz}(k)$, the degrees of the polynomials $\omega(B)$ and $\delta(B)$ and the delay parameter b can be obtained. Then, the transfer function A(B) can be computed by the Eq. (3.4.5).

The noise N_t is estimated from Eq. (3.4.1), as

$$\hat{N}_t = y_t - A(B)x_t. \tag{3.4.6}$$

Preliminary identification of a suitable model for the noise sequence is carried out by fitting an ARMA processes

$$\phi(B)\hat{N}_t = \theta(B)\varepsilon_t$$

or

$$\hat{N}_t = \frac{\theta(B)}{\phi(B)}\varepsilon_t \tag{3.4.7}$$

in which ε_t has zero mean and variance σ_ε^2. Substituting Eqs. (3.4.7) and (3.4.5) into (3.4.1) gives the preliminary transfer function model

$$y_t = \frac{\omega(B)B^{b}}{\delta(B)}x_t + \frac{\theta(B)}{\phi(B)}\varepsilon_t \tag{3.4.8}$$

Then, the non-linear least squares method can be used to minimize the sum of square errors $\Sigma\, \hat{\varepsilon}_t^2$, so the final parameters of the model be estimated.

The diagnostic checking of the models includes the checking of the properties of the residual ε_t and checking the independence of the two series x_t and N_t. The forecast function of the model shown above will not be presented here. A periodic transfer

function model identified by spectral analysis will be presented in the next chapter, and the forecast function of the transfer function model will be discussed there.

3.4.2 The Difficulties of Modeling the Multiple Input-Multiple Output Transfer Function Model with the Time Domain Approach.

When a dynamic system has two or more inputs and only one output, the common procedure for reducing the model complexity is to combine similar types of series to form a single input system (Hipel et al., 1982). Consider the case where two input covariate series, $x_t^{(1)}$ and $x_t^{(2)}$ are to be combined to form a single input covariate series x_t. If $x_t^{(i)}$ causes y_t instantaneously, then the transfer function noise models for the two series would be

$$y_t = \omega_0^{(1)} x_t^{(1)} + N_t^{(1)} \tag{3.4.9a}$$

$$y_t = \omega_0^{(2)} x_t^{(2)} + N_t^{(2)} \tag{3.4.9b}$$

In this case, the two series x_{t1} and x_{t2} would be combined using the relative ratio of the transfer function coefficients such that

$$\frac{\omega_0^{(1)}}{\omega_0^{(1)} + \omega_0^{(2)}} x_t^{(1)} + \frac{\omega_0^{(2)}}{\omega_0^{(1)} + \omega_0^{(2)}} x_t^{(2)} = x_t. \tag{3.4.10}$$

The application of two inputs, one output transfer function model was presented by Thompstone et al. (1985). In this case, the model which gives the relationship between the deseasonalized quarter monthly rainfall $x_{t,1}$ and the deseasonalized log transformed quarter monthly inflow z_t was selected to be

$$z_t = \frac{\omega_0^{(1)} - \omega_1^{(1)}B}{1-\delta_1^{(1)}B} x_t^{(1)} + N_t \qquad (3.4.11)$$

The second covariate series was a quarter-monthly snowmelt series, and it was extracted from a conceptual model. The model relating deseasonalized quarter-monthly snowmelt $x_{t,2}$ and deseasonalized log-transformed quarter-monthly inflow z_t was given by

$$z_t = \frac{\omega_0^{(2)} - \omega_1^{(2)}B}{1 - \delta_1^{(2)}B} x_t^{(2)} + N_t. \qquad (3.4.12)$$

Then, they assumed that the form of the transfer functions (3.4.11) and (3.4.12) was conserved, and a two inputs, one output transfer function noise model was given as

$$z_t = \frac{\omega_0^{(1)} - \omega_1^{(1)}B}{1 - \delta_1^{(1)}B} x_t^{(1)} + \frac{\omega_0^{(2)} - \omega_1^{(2)}B}{1 - \delta_1^{(2)}B} x_t^{(2)} + N_t \qquad (3.4.13)$$

The method of combining two equations of a "single input-single output model" into a "two input-single output model" shown above, is not generally applicable. The two inputs may be correlated between them, but the foregoing formulation does not take that relationship into account. The time domain approach has problem in formulating multiple input-multiple output transfer function model. Hence, the formulation of multiple input-multiple output model by using the frequency domain technique will be presented later.

3.5 Nonlinear Least Squares Method

Let us assume that y_t is defined by a model with parameter set θ. The least squares estimate of θ is found by minimizing the sum of squares errors

$$S = \sum_{t=1}^{N} (y_t - \hat{y}_t)^2 \tag{3.5.1}$$

where $\hat{y}_t$ is the predicted or calculated value, y_t is the historical value and N is the number of observations.

The sum of the squares of the residuals at all points gives an indication of goodness of fit. Clearly, if the objective function S is equal to zero, then the parameters has led to a perfect fit, the data points fall exactly on the predicted curve. The best values of the model parameters are obtained when the objective function is minimized. Methods for minimizing the sum of Eq. (3.5.1) include linear and nonlinear methods. The nonlinear least squares methods are iterative in nature, in which beginning with some initial values, the sum function is determined and sequentially decreased until a convergence criterion is satisfied. In here, the Marquardt algorithm (Marquardt, 1963; Kuester and Mize, 1973) will be presented.

The Marquardt algorithm solves the parameters in a multivariable, nonlinear model $\hat{y} = F(x_1, x_2, \ldots, x_K; \hat{A}_1, \hat{A}_2, \ldots, \hat{A}_3)$. The method allows for convergence with relatively poor starting values for the unknown parameters. Let the model be linearized by expanding $\hat{y}_i$ in a Taylor series about current trial values for the coefficients and retaining the linear terms only,

$$\hat{y}_1 = \hat{y}_1^* + \left(\frac{\partial \hat{y}_1}{\partial \hat{A}_1}\right)^* (\hat{A}_1 - \hat{A}_1^*) + \left(\frac{\partial \hat{y}_1}{\partial \hat{A}_2}\right)^* (\hat{A}_2 - \hat{A}_2^*)$$
$$+ \quad \ldots\ldots\ldots \quad + \left(\frac{\partial \hat{y}_1}{\partial \hat{A}_M}\right)^* (\hat{A}_M - \hat{A}_M^*)$$

$$\hat{y}_2 = \hat{y}_2^* + \left(\frac{\partial \hat{y}_2}{\partial \hat{A}_1}\right)^* (\hat{A}_1 - \hat{A}_1^*) + \left(\frac{\partial \hat{y}_2}{\partial \hat{A}_2}\right)^* (\hat{A}_2 - \hat{A}_2^*)$$
$$+ \quad \cdots\cdots\cdots \quad + \left(\frac{\partial \hat{y}_2}{\partial \hat{A}_M}\right)^* (\hat{A}_M - \hat{A}_M^*) \tag{3.5.2}$$

.
.
.

$$\hat{y}_M = \hat{y}_M^* + \left(\frac{\partial \hat{y}_M}{\partial \hat{A}_1}\right)^* (\hat{A}_1 - \hat{A}_1^*) + \left(\frac{\partial \hat{y}_M}{\partial \hat{A}_2}\right)^* (\hat{A}_2 - \hat{A}_2^*)$$
$$+ \quad \cdots\cdots\cdots \quad + \left(\frac{\partial \hat{y}_M}{\partial \hat{A}_M}\right)^* (\hat{A}_M - \hat{A}_M^*)$$

in which the asterisk designates quantities evaluated at the initial trial values. Then Eq. (3.5.1) can be written as

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_M \end{bmatrix} = \begin{bmatrix} \hat{y}_1^* \\ \hat{y}_2^* \\ \vdots \\ \hat{y}_M^* \end{bmatrix} + \begin{bmatrix} \left(\frac{\partial \hat{y}_1}{\partial \hat{A}_1}\right)^* & \left(\frac{\partial \hat{y}_1}{\partial \hat{A}_2}\right)^* & \cdots & \left(\frac{\partial \hat{y}_1}{\partial \hat{A}_M}\right)^* \\ \left(\frac{\partial \hat{y}_2}{\partial \hat{A}_1}\right)^* & \left(\frac{\partial \hat{y}_2}{\partial \hat{A}_2}\right)^* & \cdots & \left(\frac{\partial \hat{y}_2}{\partial \hat{A}_M}\right)^* \\ \vdots & \ddots & \ddots & \vdots \\ \left(\frac{\partial \hat{y}_M}{\partial \hat{A}_1}\right)^* & \left(\frac{\partial \hat{y}_M}{\partial \hat{A}_2}\right)^* & \cdots & \left(\frac{\partial \hat{y}_M}{\partial \hat{A}_M}\right)^* \end{bmatrix} \begin{bmatrix} \hat{A}_1 - \hat{A}_1^* \\ \hat{A}_2 - \hat{A}_2^* \\ \vdots \\ \hat{A}_M - \hat{A}_M^* \end{bmatrix} \tag{3.5.3}$$

or

$$\hat{\underline{y}} - \hat{\underline{y}}^* = \underline{D} \cdot \Delta\hat{\underline{A}}. \tag{3.5.4}$$

in which D is the square matrix of derivatives and $\Delta\hat{A}$ is the vector of differences of the A's. The linearized model is substituted into the objective function (3.5.1) and the

normal equations are formed by setting the partial derivatives of the objective function with respect to each coefficient equal to zero,

$$\frac{\partial S}{\partial \hat{A}_j} = 0, \qquad j=1,2,...,M$$

The resulting normal equations will be of the form

$$\underline{D}^T \underline{D}\, \underline{\Delta \hat{A}} = \underline{D}^T (\underline{\hat{y}} - \underline{\hat{y}}^*) \tag{3.5.5}$$

The normal equations are a system of linear algebraic equations and are solved by an appropriate technique for $\underline{\Delta \hat{A}}$. The $\underline{\Delta \hat{A}}$ vector and the objective function S will approach zero as convergence is achieved.

Equation (3.5.5) needs a good set of starting estimates of the unknown parameters to obtain the convergence. Hence, a modified equation was suggested by Marquardt (1963), i.e.

$$(\underline{D}^T \underline{D} + \lambda \underline{I}) \cdot \underline{\Delta \hat{A}} = \underline{D}^T (\underline{\hat{y}} - \underline{\hat{y}}^*) \tag{3.5.6}$$

where λ is a factor and $\underline{I}$ is the identity matrix. The λ is added to each term of the main diagonal of the $\underline{A}^T\underline{A}$ matrix. Thus, when λ is large in Eq. (3.5.6), it would be expected to converge for poor starting values but requires a lengthy solution time. When the λ equals zero, it will converge rapidly for good starting estimates. If the convergence is achieved, the final coefficients are calculated from

$$\hat{A}_j = \hat{A}_j^* + \underline{\Delta \hat{A}_j}, \qquad j=1,2,...,M$$

If convergence is not achieved, $\underline{\hat{A}}^*$ is updated by replacing the old values by the new values and the process repeated.

3.6 The Forecasting of Log-Transformed Series

Monthly streamflow is a skewed hydrologic time series. A transformation can be used to ensure that the model residuals are approximately normally distributed. For instance, a transformation function of the form (Salas et al., 1980)

$$y_{v,\tau} = a_\tau \, (x_{v,\tau} - c_\tau)^{b_\tau} \tag{3.6.1}$$

may be used.

However, the problem of using the log or square root transformation is that the forecast of actual flows is always underestimated, because the log function is a concave function. A function f(x) is said to be convex in an interval if $f''(x) \geq 0$, and concave if $f''(x) < 0$. Consider a random variable y with mean E[y] and g(.) be a concave function. Then, by the Jensen inequality (Mood et al., 1974)

$$E[g(y)] < g(E[y]) \tag{3.6.2}$$

Thus, if the flow x_t is transformed as

$$y_t = \ln(x_t) \tag{3.6.3}$$

then, the following will hold

$$E[\ln(x_{t+L}|x_1,\ldots,x_t)] < \ln(E[x_{t+L}|x_1,\ldots,x_t])$$

or

$$\hat{y}_{t+L} < \ln(\hat{x}_{t+L})$$

$$e^{\hat{y}_{t+L}} < \hat{x}_{t+L} \tag{3.6.4}$$

From Eq. (3.6.4), we can see that after we find a model to forecast the log-flow, we can not use the inverse log transformation to forecast the actual flow, i.e.

$$\hat{x}_{t+L} \neq e^{\hat{y}_{t+L}} \tag{3.6.5}$$

Since the interest is in forecasting the actual streamflow, one should use instead the method of minimum mean squared error forecasts of the actual flow, which can be written as (Granger and Newbold, 1976)

$$\hat{x}_{t+L} = \exp\left[\hat{y}_{t+L} + \frac{1}{2}\hat{\sigma}^{*2}_{t+L}(e_t)\right] \tag{3.6.6}$$

where $\hat{y}_{t+L}$ is the L step ahead forecast produced by the model for the logged data and $(\hat{\sigma}^{*}_{t+L}(e_t))^2$ is the variance of the L step ahead forecast error of y_t, i.e.

$$\hat{y}_t(L) = y_{t+L} + e_t(L) \tag{3.6.7}$$

and

$$\hat{\sigma}^{*2}_{t+L}(e_t) = Var[e_t(L)] \tag{3.6.8}$$

where $e_t(L)$ is the L steps ahead forecast error of y_t series.

3.7 Spectral Analysis

3.7.1 Introduction

In this research, the spectral analysis is used for identifying the transfer function model. It is specially useful for formulating the transfer function model with multiple-inputs and multiple outputs. The periodogram based spectrum estimation and the covariance based spectrum estimation are discussed. The periodogram based spectrum is needed for some derivations. The covariance based spectrum is difficult of programming but it is fast to compute. Both, the periodogram based spectrum and covariance based spectrum estimation are presented here.

3.7.2 Discrete Fourier Transform

A physical process can be described either in the time domain or in the frequency domain. The Fourier transform is the vehicle which transforms from one space to the other (Jenkins and Watts, 1968). The discrete Fourier transform equations has been discussed by Press et al. (1986). It is at least heuristically reasonable that a matching or correlating of a sine or cosine function with strong periodic components should yield large values when the time series looks like sine or cosine with which it is matched and should give small values otherwise (Shumway, 1988). The two random variables that can be formed in this manner are the cosine transform

$$H_c(f_j) = \sum_{k=0}^{N-1} h_k \cos(2\pi f_j k) \tag{3.7.1}$$

and the sine transform

$$H_s(f_j) = \sum_{k=0}^{N-1} h_k \sin(2\pi f_j k), \tag{3.7.2}$$

for j=0,1,..., N-1, where the frequencies measured in cycles per point are

$$f_j = \frac{j}{N} \tag{3.7.3}$$

In common situations, the discrete transform pair can be written as

$$H(f_j) = \sum_{k=0}^{N-1} h_k \, e^{-i\,2\pi f_j k} = \sum_{k=0}^{N-1} h_k \, e^{-i\,(\frac{2\pi j}{N} k)} \tag{3.7.4}$$

and

$$h_k = \frac{1}{N} \sum_{j=0}^{N-1} H(f_j) \, e^{i\,2\pi f_j k} = \frac{1}{N} \sum_{j=0}^{N-1} H(f_j) \, e^{i\,(\frac{2\pi j}{N} k)} \tag{3.7.5}$$

The relations between the h_k and $H(f_j)$ are shown below:

h_k	$H(f_j)$
real and even	real and even
real and odd	imaginary and odd
imaginary and even	imaginary and even
imaginary and odd	real and odd

3.7.3 Power Spectrum and Cross Spectrum

Spectral analysis is a non-parametric approach, its usefulness in the area of model building is limited. However, it is sometimes useful as an exploratory tool for suggesting models which can then be fitted parametrically (Jenkins and Watts, 1968). Many approaches can be used to estimate the power spectrum, among those, the theory described in Jenkins and Watts (1968), Brillinger (1981), Chatfield (1980), Priestley (1981), Robinson (1983) and Shumway (1988) will be followed.

The autocovariance depends on the lag k can be expressed as

$$\gamma_x(k) = E[(x_{t+k} - \mu)(x_t - \mu)] \tag{3.7.6}$$

In order to develop the frequency-dependent characteristics of a stationary time series x_t, $t = 0, \pm 1, \pm 2, \ldots$, it is essential to begin with the Fourier representation of the autocovariance function $\gamma_x(k)$. If the autocorvariance $\gamma_x(k)$ satisfies

$$\sum_{k=-\infty}^{\infty} |\gamma_x(k)| < \infty.$$

Then, there exists a non-negative function $S_x(f)$ satisfying

$$\gamma_x(k) = \int_{-1/2}^{1/2} S_x(f)\, e^{i\,2\pi fk}\, df \tag{3.7.7}$$

for k = 0, ± 1, ± 2,..... The function $S_x(f)$ is called the power spectrum of the process x_t. The inverse relation

$$S_x(f) = \sum_{k=-\infty}^{\infty} \gamma_x(k)\, e^{-i\,2\pi fk} \tag{3.7.8}$$

is also useful in certain circumstances. Since $\gamma_x(k) = \gamma_x(-k)$, it follows that $S_x(-f) = S_x(f)$.

The definition can be extended by noting that for two jointly stationary series x_t and y_t, the cross covariance

$$\gamma_{xy}(k) = E[(x_{t+k} - \mu_x)(y_t - \mu_y)] \tag{3.7.9}$$

and the cross spectrum $S_{xy}(f)$ are related by

$$\gamma_{xy}(k) = \int_{-1/2}^{1/2} S_{xy}(f)\, e^{i\,2\pi fk}\, df \tag{3.7.10}$$

$$S_{xy}(f) = \sum_{k=-\infty}^{\infty} \gamma_{xy}(k)\, e^{-i\,2\pi fk} \tag{3.7.11}$$

The cross spectrum is generally a complex function, and writing out the exponential part of (3.7.11) gives

$$S_{xy}(f_j) = \sum_{k=-\infty}^{\infty} \gamma_{xy}(k)\ [\cos(2\pi f_j k)\ -\ i\ \sin(2\pi f_j k)]$$

$$= \sum_{k=-\infty}^{\infty} \gamma_{xy}(k)\cos(\frac{2\pi jk}{N}) - i\sum_{k=-\infty}^{\infty} \gamma_{xy}(k)\sin(\frac{2\pi jk}{N})$$

$$= c_{xy}(f)\ -\ i\ q_{xy}(f) \qquad (3.7.12)$$

where $c_{xy}(f)$ is called the co-spectrum and q_{xy} is the quad-spectrum.

Now, since $\gamma_{yx}(k) = \gamma_{xy}(-k)$, it follows that

$$S_{yx}(f) = \sum_{k=-\infty}^{\infty} \gamma_{yx}(k)\ e^{-i2\pi fk}$$

$$= \sum_{k=-\infty}^{\infty} \gamma_{xy}(-k)\ e^{-i2\pi fk}$$

$$= \sum_{k=-\infty}^{\infty} \gamma_{xy}(k)\ e^{i2\pi fk}$$

$$= \bar{S}_{xy}(f) \qquad (3.7.13)$$

where the overbar denotes the complex conjugate.

3.7.4 Periodogram Based Spectrum Estimation

Since most of the spectrum theory is related to the periodogram, some basic knowledge about the periodogram is presented. An observed zero-mean stationary time series $\{x_t,\ t=0,1,\ldots,N-1\}$ has a discrete Fourier transform

$$X(f_j) = \sum_{t=0}^{N-1} x_t\ e^{-i2\pi f_j t}, \qquad (3.7.14)$$

where f_j is defined in Eq. (3.7.3).

Then, the periodogram is defined as

$$P_x(f_j) = \frac{1}{N} X(f_j)\ \bar{X}(f_j)$$

$$= \frac{1}{N} |X(f_j)|^2 \qquad (3.7.15)$$

The periodogram appears to be a natural way of estimating the power spectrum, but we shall see that for a process with a continuous spectrum, it provides a poor estimate and needs to be modified.

The periodogram $P_x(f_j)$ and the autocovariance $\gamma_x(k)$ are both quadratic forms of the data $\{x_t\}$. If x_t is a zero mean stationary series, Eq. (3.7.15) can be written as

$$P_x(f_j) = \frac{1}{N} \left| \sum_{t=0}^{N-1} x_t\ e^{-i2\pi f_j t} \right|^2$$

$$= \frac{1}{N} \{ [\sum_{t=0}^{N-1} x_t \cos(2\pi f_j t)]^2 + [\sum_{t=0}^{N-1} x_t \sin(2\pi f_j t)]^2 \}$$

$$= \frac{1}{N} \{ \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} x_t \cos(2\pi f_j t)\ x_s \cos(2\pi f_j s)$$

$$+ \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} x_t \sin(2\pi f_j t)\ x_s \cos(2\pi f_j s) \}$$

$$= \frac{1}{N} \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} x_s x_t [\cos(2\pi f_j t) \cos(2\pi f_j s)$$

$$+ \sin(2\pi f_j t) \sin(2\pi f_j s)]$$

Defining s=t+k, we have

$$
\begin{aligned}
P_x(f_j) &= \frac{1}{N}\sum_{k=0}^{N-1}\sum_{t=0}^{N-1} x_{t+k}x_t \cos(2\pi f_j k) \\
&= \frac{\sum_{t=0}^{N-1} x_t^2}{N} + 2\sum_{k=1}^{N-1}\frac{\sum_{t=0}^{N-k} x_{t+k}x_t}{N}\cos(2\pi f_j k) \\
&= \hat{\gamma}_x(0) + 2\sum_{k=1}^{N-1}\hat{\gamma}_x(k)\cos(2\pi f_j k)
\end{aligned}
\tag{3.7.16}
$$

Since $\hat{\gamma}_x(-k) = \hat{\gamma}_x(k)$, we can rewrite the above equation to another form, i.e.

$$
\begin{aligned}
P_x(f_j) &= \hat{\gamma}_x(0)e^{-i2\pi f_j 0} \\
&\quad + \sum_{k=1}^{N-1}\hat{\gamma}_x(k)\,[\cos(2\pi f_j k) + i\sin(2\pi f_j k)] \\
&\quad + \sum_{k=1}^{N-1}\hat{\gamma}_x(k)[\cos(2\pi f_j k) - i\sin(2\pi f_j k)] \\
&= \hat{\gamma}_x(0)\,e^{-i2\pi f_j 0} + \sum_{s=-(N-1)}^{-1}\hat{\gamma}_x(-s)\,e^{-i2\pi f_j s} \\
&\quad + \sum_{k=1}^{N-1}\hat{\gamma}_x(k)\,e^{-i2\pi f_j k} \\
&= \sum_{k=-(N-1)}^{N-1}\hat{\gamma}_x(k)\,e^{-i2\pi f_j k}
\end{aligned}
\tag{3.7.17}
$$

Equation (3.7.17) says that the finite Fourier transform of $\hat{\gamma}_x(k)$, which appears to be the obvious estimate of the power spectrum

$$
S_x(f) = \sum_{k=-\infty}^{\infty}\gamma_x(k)\,e^{-i2\pi f k}
$$

Although periodogram $P_x(f_j)$ is asymptotically unbiased, i.e.

$$E[P_x(f_j)] \rightarrow S_x(f) \qquad if\ N \rightarrow \infty$$

the variance of $P_x(f_j)$ does not decrease as N increases. Thus, $P_x(f_j)$ is not a consistent estimator of $S_x(f)$ (see Chatfield, 1980).

The alternative way of estimating a power spectrum can be accomplished by smoothing the periodogram. In order to develop an average spectral estimator in the neighborhood of some center frequency f_k, we define the notion of a frequency band centered on f_k as (Shumway, 1988)

$$B_k = \{f; f_k - \frac{L}{2N} \le f \le f_k + \frac{L}{2N}\}$$

with a sample series x_t observed at the point t=0,1,...,N-1, and the L considerably less then N, where L is some integer. The length of the interval defined by the set B_k is called the bandwidth, which is equal to B=L/N cycles per point. We may now consider the smoothed spectral estimator, defined over the frequency interval B_k, as the average of the periodogram (L is an odd number)

$$\begin{aligned} \hat{S}_x(f_k) &= \frac{1}{L} \sum_{j=-\frac{L-1}{2}}^{\frac{L-1}{2}} P_x(f_k + \frac{j}{N}) \\ &= \frac{1}{L} \sum_{j=-\frac{L-1}{2}}^{\frac{L-1}{2}} \frac{1}{N} \left| X(f_k + \frac{j}{N}) \right|^2 \end{aligned} \tag{3.7.18}$$

over the interval. This approach smooths the periodogram ordinates in sets of size L and

finds their average value. It is clear that the estimator of Eq.(3.7.18) may be biased, since

$$E[\hat{S}_x(f_k)] \simeq \frac{1}{L} \sum S_x(f_k)$$

which is equal to $S_x(f_k)$ only if the spectrum is linear over the interval. The bias will be unimportant provided that $S_x(f_k)$ is a reasonably smooth function at f_k and L is not too large compared with N (Chatfield, 1980).

The L value has to be chosen so as to balance resolution against variance. The larger the value of L the larger will be the bias, and if L is too large then interesting features of $S_x(f)$, such as peaks, may be smoothed out. As N increases, we can allow L to increase too. It seems advisable to try several values, in the region of N/40 (Chatfield, 1980). A high value should give some idea where the large peaks in $S_x(f)$ are, but the curve is likely to be too smooth. A low value is likely to produce a curve showing a large number of peaks, some of which may be spurious. A compromise can then be made.

The cross-spectral estimator is defined similarly in terms of discrete Fourier transforms $X(f_k)$ and $Y(f_k)$ of the series x_t and y_t. That is, using the cross periodogram

$$P_{xy}(f_k) = \frac{1}{N} X(f_k) \bar{Y}(f_k) \qquad (3.7.19)$$

then the smoothed cross-spectral estimator can be obtained by

$$\hat{S}_{xy}(f_k) = \frac{1}{L} \sum_{j=-(L-1)/2}^{(L-1)/2} P_{xy}(f_k + \frac{j}{N})$$

$$= \frac{1}{L} \sum_{j=-(L-1)/2}^{(L-1)/2} \frac{1}{N} X(f_k + \frac{j}{N}) \bar{Y}(f_k + \frac{j}{N}) \qquad (3.7.20)$$

Historically, the periodogram based spectrum estimation was not used until recent years because it requires much computation. Calculating the periodogram from

$$P_x(f_j) = \frac{1}{N} |X(f_j)|^2$$

$$= \frac{1}{N} \left| \sum_{t=0}^{N-1} x_t \, e^{-i2\pi f_j t} \right|^2$$

$$= \frac{1}{N} \{ [\sum_{t=0}^{N-1} x_t \cos(2\pi f_j t)]^2$$

$$+ [\sum_{t=0}^{N-1} x_t \sin(2\pi f_j t)]^2 \}$$

at f_j for j=1,2,...,N/2, would require about N^2 arithmetic operations (each one a multiplication and an addition). This problem can be solved by using the technique of Fast Fourier Transform (Cooley and Tukey, 1965; Cooley et al., 1977; Thrall, 1983).

3.7.5 Covariance Based Spectrum Estimation

For a zero mean stationary time series x_t, the autocovariance $\gamma_x(k)$ can be estimated by

$$\hat{\gamma}_x(k) = \frac{1}{N} \sum_{t=0}^{N-|k|} x_{t+|k|} x_t \qquad (3.7.21)$$

for k = 0, ± 1, $\pm 2,\ldots,\pm$(N-1). When the lag value k is large, the sample autocovariance is a poor estimate, since in this tail region $\hat{\gamma}_x(k)$ is based on just a small number of pairs of observations. The periodogram of Eq. (3.7.17)

$$P_x(f_j) = \sum_{k=-(N-1)}^{N-1} \hat{\gamma}_x(k)\ e^{-i2\pi f_j k}$$

shows that the periodogram contains too many sample autocovariances, which is the reason why the variance of periodogram does not tend to zero as N increases (Priestley, 1981).

One way of obtaining an expression with a reduced variance is simply to omit some of the terms in Eq. (3.7.17). If we omit only those terms which correspond to the tail of the sample autocovariance function then hopefully the bias will not be affected too seriously. These ideas suggest that we might consider as an estimator of $S_x(f_j)$ an expression of the form

$$S_x^*(f_j) = \sum_{k=-M}^{M} \hat{\gamma}_x(k)\ e^{-i2\pi f_j k} \qquad (3.7.22)$$

where M is a truncation point.

The expectation of the sample autocovariance $\hat{\gamma}_x(k)$ can be derived from Eq. (3.7.21) as

$$E[\hat{\gamma}_x(k)] = \frac{1}{N} \{\sum_{t=0}^{N-|k|} E[x_{t+|k|}\, x_t]\}$$
$$= \frac{N-|k|}{N}\, \gamma(k)$$
$$= (1 - \frac{|k|}{N})\, \gamma(k) \qquad (3.7.23)$$

From Eqs. (3.7.22) and (3.7.23), we have

$$E[S_x^*(f_j)] = \sum_{k=-M}^{M} E[\hat{\gamma}_x(k)]\; e^{-i2\pi f_j k}$$
$$= \sum_{k=-M}^{M} (1-\frac{|k|}{N})\; \gamma(k)\; e^{-i2\pi f_j k} \qquad (3.7.24)$$

and

$$\lim_{N\to\infty} E[S_x^*(f_j)] \to S_x(f)$$

Hence, if we make the value of M to depend on N in such a way that $M\to\infty$ as $N\to\infty$, then $S_x^*(f_j)$ will be an asymptotically unbiased estimator of $S_x(f_j)$. Thus we may let $M\to\infty$ as $N\to\infty$, but sufficiently slowly relative to N so that $(M/N)\to 0$ as $N\to\infty$, then both the bias and variance of $S_x^*(f_j)$ will be a consistent estimator of $S_x(f_j)$.

The general form of the power spectrum estimator can be written as

$$\hat{S}_x(f_j) = \sum_{k=-M}^{M} w(k)\; \hat{\gamma}_x(k)\; e^{-i2\pi f_j k} \qquad (3.7.25)$$

where w(k) is called the lag window. From Eq. (3.7.25), we can see that the values of $\hat{\gamma}_x(k)$ for $M<k<N$ are no longer used, while values of $\hat{\gamma}_x(k)$ for $k\le M$ are weighted by lag window w(k). The lag window used in this research is the Parzen window

(Parzen,1961; Robinson, 1983), which for M even is

$$w(k)=\begin{cases} 1-6\left(\frac{k}{M}\right)^2+6\left|\frac{k}{M}\right|^3 & for\ |k|=0,1,\ldots,\frac{M}{2} \\ 2\left(1-\left|\frac{k}{M}\right|\right)^3 & for\ |k|=\frac{M}{2}+1,\ldots,M \\ 0 & for\ |k|>M \end{cases} \tag{3.7.26}$$

The choice of the truncation point M is rather difficult. The smaller the value of M, the smaller will be the variance of $\hat{S}_x(f)$, but the larger will be the bias. Jenkins and Watts (1968) suggest trying three different values of M. A low value will give a smooth curve. A high value will produce a curve with a large number of peaks. Then, a compromise can be achieved with the third value of M. Chatfield (1980) suggested to choose M to be about 2 times the square root of N, so that as $N\rightarrow\infty$, so does that $M\rightarrow\infty$, but in such a way that $M/N \rightarrow 0$. Robinson (1983) suggested that the truncation point M can be 10 or 20% of the number N of observations in the finite-length time series.

The cross spectrum can be estimated by using

$$\hat{S}_{xy}(f_j) = \sum_{k=-M}^{M} w(k)\,\hat{\gamma}_{xy}(k)\; e^{-i2\pi f_j k} \tag{3.7.27}$$

Before estimating the cross spectrum, the cross covariance should be estimated. However, there is an important difference between the autocovariance and the cross-covariance. The autocovariance $\gamma_x(k)$ is symmetric about the origin, but the cross-covariance is not. If a process is an exact copy of the other process but delayed by D, then the cross-correlation function will be identical to the autocorrelation function but centered at lag D instead of zero. The trouble is that the estimation of cross spectrum

will be appreciable bias if the base width M of lag window is less than the delay D (Jenkins and Watts, 1968). Further, if D is large, the numerical value of the lag window w(k) will be small when D equal to M, and hence very large value of M may be needed in order to reduce the bias to a reasonable size. This effect can be remedied by aligning the two series to ensure that the cross correlation function has a maximum in the neighborhood of zero lag (Brillinger, 1981; Shumway, 1988).

CHAPTER IV

SINGLE INPUT-OUTPUT PERIODIC TRANSFER FUNCTION MODEL

4.1 Introduction

In this chapter the forecast procedure is developed assuming that the transfer function model consists of one input and one output. Furthermore, it is assumed that the input and output are periodic processes.

4.2 Model Description

We are interested in the monthly flow forecast by using the transfer function model with intermittent snow water equivalent data as the input. The PAR(1) model has been considered as the model describing the input and output process. Before using the transfer function model, a periodic autoregressive PAR(1) model was fitted to the monthly streamflow $Q_{v,\tau}$. This is deseasonalized by

$$z_{v,\tau} = \frac{Q_{v,\tau} - \hat{\mu}_\tau(Q)}{\hat{\sigma}_\tau(Q)} \tag{4.2.1}$$

where $\hat{\mu}_\tau(Q)$ is the sample mean and $\hat{\sigma}_\tau(Q)$ is the sample standard deviation of monthly streamflow for season τ. Then, a PAR(1) model was fitted to the deseasonalized streamflow $z_{v,\tau}$ as

$$z_{v,\tau} = \varphi_{1,\tau}\, z_{v,\tau-1} + y_{v,\tau} \tag{4.2.2}$$

where the series $y_{v,\tau}$ is not necessarily a white noise. The parameter $\varphi_{1,\tau}$ can be estimated by

$$\hat{\varphi}_{1,\tau} = \hat{\rho}_{1,\tau} \tag{4.2.3}$$

The input to the transfer function model is the monthly snow water equivalent which is an intermittent series, since in the summer time there is insignificant snow in the mountains. The non-zero part of the snow water equivalent $W_{v,\tau}$ was deseasonalized by

$$x_{v,\tau} = \frac{W_{v,\tau} - \hat{\mu}_\tau(W)}{\hat{\sigma}_\tau(W)} \quad for\ W_{v,\tau} \neq 0$$

$$x_{v,\tau} = 0 \quad for\ W_{v,\tau}=0 \tag{4.2.4}$$

where $\hat{\mu}_\tau(W)$ is the sample mean and $\hat{\sigma}_\tau(W)$ is the sample standard deviation of monthly snow water equivalent for month τ.

Then, a transfer function model is formulated with deseasonalized snow water equivalent $x_{v,\tau}$ as the input, and with the residual series $y_{v,\tau}$ of Eq. (4.2.2) as the output. The model is written as

$$y_t = \sum_{j=0}^{\infty} a_j x_{t-j} + N_t \tag{4.2.5}$$

where $t=(v-1)\cdot 12+\tau$, and the noise N_t has zero mean and is assumed to be uncorrelated with the input x_t . After identifying the model order, the noise N_t can be derived and an ARMA process fitted to N_t

$$\phi(B)N_t = \theta(B)\varepsilon_t \tag{4.2.6}$$

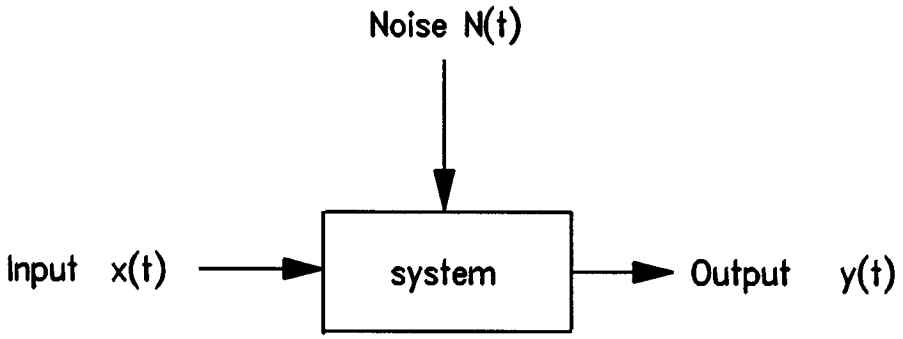
where $\varepsilon_t \sim WN(0,\sigma_\varepsilon^2)$. Finally, the periodic transfer function model can be obtained from Eqs. (4.2.2), (4.2.5) and (4.2.6).

$$z_{v,\tau} = \varphi_{1,\tau} z_{v,\tau-1} + \sum_{j=0}^{D} a_j x_{v,\tau-j} + \frac{\theta(B)}{\phi(B)} \varepsilon_{v,\tau} \qquad (4.2.7)$$

where D is the order of the transfer function model.

4.3 Model Identification

When the structure of the system is unknown, we need to solve the problem by examining the relationship between input and output so as to infer the properties of the system. This procedure is called the identification of the system. In our research here, we are interested in the forecast of monthly streamflow in which SWE is an input. Here we assume the physical system is linear over the range of interest. But many systems may be contaminated by the noise N_t as shown below:



The noise N_t may not be white noise, but is usually assumed to be uncorrelated with the input process x_t and have zero mean. If the input x_t and output y_t are stationary with zero mean, the Eq. (4.2.5) can be rewritten as

$$y_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j} + N_t \qquad (4.3.1)$$

with $a_j = 0$, for $j<0$. Note the integration from $-\infty$ to ∞ is needed if one uses an estimation or identification technique based on Fourier transformation.

The objective at this stage is to obtain some idea of the order D of the transfer function

$$A(B)=...+a_0+a_1B^1+...+a_DB^D+...$$

and to derive the initial guesses of the parameters $a_0, a_1,...,a_D$. The impulse response function on the negative side, i.e. $a_{-1}, a_{-2},...$, was assumed to be zero. In this research, the identification of the transfer function model was made by using spectral analysis.

First, we need to multiply Eq. (4.3.1) by x_{t-k} and take expectations as

$$E[y_t x_{t-k}]=\sum_{j=-\infty}^{\infty} a_j E[x_{t-j}x_{t-k}] + E[N_t x_{t-k}]$$

Since N_t is uncorrelated with input x_t, we have

$$\gamma_{yx}(k) = \sum_{j=-\infty}^{\infty} a_j \gamma_x(k-j) \qquad (4.3.2)$$

Take Fourier transforms on both sides of Eq. (4.3.2) by multiplying them by $e^{-i2\pi fk}$ and summing from $k=-\infty$ to $+\infty$. Then, we have

$$\sum_{k=-\infty}^{\infty} \gamma_{yx}(k) e^{-i2\pi fk} = \sum_{k=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} a_j \gamma_x(k-j) e^{-i2\pi fk}$$

or

$$S_{yx}(f) = \sum_{j=-\infty}^{\infty} a_j e^{-i2\pi fj} \sum_{k=-\infty}^{\infty} \gamma_x(k-j)\, e^{-i2\pi f(k-j)}$$

$$= A(f)\ S_x(f)$$

Then,

$$A(f) = \frac{S_{yx}(f)}{S_x(f)} \tag{4.3.3}$$

where the A(f) is called the frequency response function.

The spectrum $S_x(f)$ can be computed by Eqs. (3.1.6), (3.7.25) and (3.7.26), and the cross spectrum $S_{yx}(f)$ can be computed by Eqs. (3.1.10), (3.7.26) and (3.7.27). Then, the frequency response function A(f) can be computed by Eq. (4.3.3). The impulse response function a_j is the Fourier transform of the frequency response function A(f), which can be computed by

$$\hat{a}_j = \frac{1}{M} \sum_{k=0}^{M-1} \hat{A}(f_k)\, e^{i2\pi f_k j} \tag{4.3.4}$$

where M is the base width of lag window.

Let y be in a Hilbert space H and consider a subspace G (y is not in G). If $\hat{y}$ is the projection of y on M, then y can be uniquely represented in the form

$$y = \hat{y} + v$$

where $\hat{y}$ is in M and v is orthogonal to M, that is

$$E[v\hat{y}] = E[(y-\hat{y})\ \hat{y}] = 0 \tag{4.3.5}$$

Now, in our case here, let the projection of the output y_t on input x_t be the form

$$\hat{y}_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j}$$

then, the mean square error can be written as

$$\begin{aligned} MSE &= E[(y_t - \hat{y}_t)^2] \\ &= E[y_t^2 - 2y_t\hat{y}_t + y_t^2] \\ &= E[y_t(y_t - \hat{y}_t) - \hat{y}_t(y_t - \hat{y}_t)] \\ &= E[(y_t - \hat{y}_t)y_t] - E[(y_t - \hat{y}_t)\hat{y}_t] \\ &= E[(y_t - \hat{y}_t)y_t] \end{aligned}$$

since from Eq. (4.3.5) we know that $E[(y_t - \hat{y}_t)\hat{y}_t] = 0$. Finally, we have

$$\begin{aligned} MSE &= E[(y_t - \sum_{j=-\infty}^{\infty} a_j x_{t-j})y_t] \\ &= E[y_t^2] - \sum_{j=-\infty}^{\infty} a_j E[x_{t-j}y_t] \\ &= \gamma_y(0) - \sum_{j=-\infty}^{\infty} a_j \gamma_{xy}(-j) \\ &= \int_{-1/2}^{1/2} S_y(f)df - \sum_{j=-\infty}^{\infty} a_j \int_{-1/2}^{1/2} S_{xy}(f)e^{i2\pi f(-j)}df \\ &= \int_{-1/2}^{1/2} s_y(f)df - \int_{-1/2}^{1/2} [\sum_{j=-\infty}^{\infty} a_j e^{-i2\pi fj}] S_{xy}(f)df \end{aligned}$$

or

$$MSE = \int_{-1/2}^{1/2} [S_y(f) - A(f) S_{xy}(f)] df$$

$$= \int_{-1/2}^{1/2} [S_y(f) - \frac{S_{yx}(f) S_{xy}(f)}{S_x(f)}] df$$

$$= \int_{-1/2}^{1/2} S_{y \cdot x}(f) df \qquad (4.3.6)$$

where

$$S_{y \cdot x}(f) = S_y(f) - \frac{S_{yx}(f) S_{xy}(f)}{S_x(f)} \qquad (4.3.7)$$

is called the conditional power spectrum (Shumway, 1988).

The properties of $\hat{a}_j$ of Eq. (4.3.4) follow as in Wahba (1969) and Brillinger (1981) by considering the distributional properties, conditional on the independent process $\{x_t\}$. This leads to the conclusion that $\hat{a}_j$ is approximately normal with mean given by (4.3.4) and variance estimated by

$$Var(\hat{a}_j) = \frac{1}{M^3} \sum_{k=0}^{M-1} \frac{\hat{S}_{y \cdot x}(f_k)}{\hat{S}_x(f_k)} \qquad (4.3.8)$$

By using this equation, we can give an approximate 100 (1-α) percent confidence interval of the form

$$\hat{a}_j \pm u_{\alpha/2} \sqrt{Var(\hat{a}_j)} \qquad (4.3.9)$$

for the jth component a_j.

In order for a system to be physically realizable, it is necessary that the system

respond only to the past inputs (Bendat and Piersol, 1971). This implies that

$$a_j = 0 \qquad \text{for } j < 0 \tag{4.3.10}$$

By using Eqs. (4.3.4), (4.3.9) and (4.3.10), the order and the initial guess of impulse response function a_j can be determined.

Then, the noise sequence in Eq. (4.3.1) can be estimated by

$$\hat{N}_t = y_t - \sum_{j=0}^{D} \hat{a}_j x_{t-j} \tag{4.3.11}$$

where D is the order of the transfer function. Preliminary identification of a suitable model for the noise sequence is carried out by fitting an ARMA model,

$$\phi(B)\hat{N}_t = \theta(B)\,\varepsilon_t \tag{4.3.12}$$

to $\hat{N}_t$ in which $\varepsilon_t \sim WN(0,\sigma_\varepsilon^2)$.

Considering $t=(v-1)\cdot 12+\tau$, the preliminary transfer function model can be obtained by combining Eqs. (4.2.2), (4.2.5) and (4.3.12), i.e.

$$z_{v,\tau} = \varphi_{1,\tau} z_{v,\tau-1} + \sum_{j=0}^{D} a_j x_{v,\tau-j} + \frac{\theta(B)}{\phi(B)}\varepsilon_{v,\tau} \tag{4.3.13}$$

which include parameters $\varphi_{1,\tau}$, a_j, $\theta(B)$ and $\phi(B)$. The estimation of those parameters is presented in the next section.

4.4 Parameter Estimation

Now we need to find jointly the parameters $\varphi_{1,\tau}$, a_j, $\theta(B)$, and $\phi(B)$ in the tentatively identified model (4.3.13). More explicitly,

$$z_{v,\tau} = (\phi_1 B + \ldots + \phi_p B^p) z_{v,\tau}$$
$$+ (1 - \phi_1 B - \ldots - \phi_p B^p)(\varphi_{1,\tau} z_{v,\tau-1} + \sum_{j=0}^{D} a_j x_{v,\tau-j})$$
$$+ (1 - \theta_1 B - \ldots - \theta_q B^q)\varepsilon_{v,\tau} \qquad (4.4.1)$$

The goal of estimating these parameters is to obtain the best model fitting utilizing the experimental data. The criterion for determining the best model parameters requires to minimize the least squares objective function

$$S = \sum (z_{v,\tau} - \hat{z}_{v,\tau})^2 = \sum \varepsilon_{v,\tau}^2 \qquad (4.4.2)$$

The non-linear least squares method is used in this research for estimating the model parameters. Let the conditional sum of squares function be

$$S(\varphi_{1,\tau}, a_j, \underline{\theta}, \underline{\phi}) = \sum \varepsilon_{v,\tau}{}^2(\varphi_{1,\tau}, a_j, \underline{\theta}, \underline{\phi}, | z_{v,\tau}, x_{v,\tau}) \qquad (4.4.3)$$

Then, a nonlinear least squares algorithm suggested by Marquardt (Kuester and Mize, 1973) can be used for estimating the parameters. The Marquardt algorithm has estimation routines which require to give the initial guesses of parameters and constrains. The preliminary estimates of parameters obtained from Section 4.3 can be used as initial values in the minimization.

4.5 Diagnostic Checking

The diagnostic checks are applied with the object of uncovering possible lack of fit and diagnosing the cause. If no lack of fit is indicated, the model is ready to use. If any inadequacy is found, the model should be modified.

All the assumptions in the model should be checked. The assumption of the residual $\varepsilon_{v,\tau}$ is white noise process should be checked first. Violation of this assumption

can lead to serious model error. For the test of independence, the Anderson test of the correlogram and the Porte Manteau lack of fit test are usually applied for testing the independence of a hydrologic time series (Salas, et al., 1980). For an independent series with sample size N, Anderson (1941) gave probability limits for the correlogram of an independent series as

$$r_k(95\%) = \frac{-1 \pm 1.96\sqrt{N-k-1}}{N-k} \qquad (4.5.1)$$

where N is the sample size.

The other assumption which should be checked is that the noise N_t and the input series x_t should be uncorrelated. The procedure of testing the independence between two stationary time series was discussed by Brockwell and Davis (1987; p. 402). Before computing the cross-correlation of the two series, we need to prewhiten one or both of the two series. The reason of transforming the series to white noise is because we need to know the bounds for the independence testing. The theory says that if one of the two processes to be checked independence is white noise, then the cross-correlation $\hat{\rho}_{12}(h)$ is approximately $N(0,N^{-1})$. Thus, it is simple to test the hypothesis that $\rho_{12}(h)=0$.

4.6 Forecast Function

Forecasting future monthly streamflows is the objective of this research. Suppose we have an observed streamflow series $x_{v,\tau}$, with $v=1,\ldots,N$ and $\tau=1,\ldots,\omega$. Then the problem is to forecast $x_{v,\tau+L}$. The forecast of $x_{v,\tau+L}$ made at time τ will be denoted by $\hat{x}_{v,\tau}(L)$. The integer L is called the lead time. In this research, the lead time is less than nine months. The larger the lead time, the more unreliable the forecast. The major application of the transfer function model is to provide more accurate forecast of the

output $z_{v,\tau}$ based on past observation of $z_{v,\tau}$ and $x_{v,\tau}$ than could be obtained by forecasting $z_{v,\tau}$ based solely on past observation $z_{v,\tau}$.

The model suggested in here is the model of Eq. (4.3.13), i.e.

$$(1-\varphi_{1,\tau}B)z_{v,\tau} = A(B)x_{v,\tau} + \frac{\theta(B)}{\phi(B)}\varepsilon_{v,\tau} \tag{4.6.1}$$

where

$$A(B) = \sum_{j=0}^{D} a_j B^j \tag{4.6.2}$$

Now, assume that an ARMA model was fitted to the non-zero deseasonalized SWE data, i.e.

$$\phi_x(B)x_t = \theta_x(B)\zeta_t \tag{4.6.3}$$

with $\zeta_t \sim \mathrm{WN}(0,\sigma_\zeta^2)$. Then, Eq. (4.6.1) can be written as

$$\begin{aligned} z_{v,\tau} &= \frac{A(B)}{(1-\varphi_{1,\tau}B)}\frac{\theta_x(B)}{\phi_x(B)}\zeta_{v,\tau} + \frac{1}{(1-\varphi_{1,\tau}B)}\frac{\theta(B)}{\phi(B)}\varepsilon_{v,\tau} \\ &= \nu_\tau(B)\zeta_{v,\tau} + \psi_\tau(B)\varepsilon_{v,\tau} \end{aligned} \tag{4.6.4}$$

where

$$A(B) = a_0 + a_1 B + a_2 B^2 + \ldots. \tag{4.6.5a}$$

$$\nu_\tau(B) = \nu_{0,\tau} + \nu_{1,\tau}B + \nu_{2,\tau}B^2 + \ldots. \tag{4.6.5b}$$

$$\psi_\tau(B) = 1 + \psi_{1,\tau}B + \psi_{2,\tau}B^2 + \ldots. \tag{4.6.5c}$$

and

$$\nu_\tau(B) = \frac{A(B)}{(1-\varphi_{1,\tau}B)}\frac{\theta_x(B)}{\phi_x(B)}$$

$$\psi_\tau(B) = \frac{1}{(1-\varphi_{1,\tau}B)}\frac{\theta(B)}{\phi(B)}$$

From Eqs. (4.6.4) and (4.6.5), the forecasting model can be expressed by

$$z_{v,\tau+L} = \sum_{j=0}^{\infty} \nu_{j,\tau+L}\zeta_{v,\tau+L-j} + \sum_{j=0}^{\infty} \psi_{j,\tau+L}\varepsilon_{v,\tau+L-j} \tag{4.6.6}$$

or

$$\begin{aligned} z_{v,\tau+L} &= (\nu_{0,\tau+L}\zeta_{v,\tau+L} + \nu_{1,\tau+L}\zeta_{v,\tau+L-1} + \ldots + \nu_{L-1,\tau+L}\zeta_{v,\tau+1}) \\ &\quad + (\nu_{L,\tau+L}\zeta_{v,\tau} + \nu_{L+1,\tau+L}\zeta_{v,\tau-1} + \ldots.) \\ &\quad + (\psi_{0,\tau+L}\varepsilon_{v,\tau+L} + \psi_{1,\tau+L}\varepsilon_{v,\tau+L-1} + \ldots + \psi_{L-1,\tau+L}\varepsilon_{v,\tau+1}) \\ &\quad + (\psi_{L,\tau+L}\varepsilon_{v,\tau} + \psi_{L+1,\tau+L}\varepsilon_{v,\tau-1} + \ldots.) \\ &= \sum_{j=0}^{L-1} \nu_{j,\tau+L}\zeta_{v,\tau+L-j} + \sum_{j=0}^{\infty} \nu_{L+j,\tau+L}\zeta_{v,\tau-j} \\ &\quad + \sum_{j=0}^{L-1} \psi_{j,\tau+L}\varepsilon_{v,\tau+L-j} + \sum_{j=0}^{\infty} \psi_{L+j,\tau+L}\varepsilon_{v,\tau-j} \end{aligned} \tag{4.6.7}$$

with the $\zeta_{v,\tau}$ and the $\varepsilon_{v,\tau}$ statistically independent. Now, suppose standing at time τ, the L steps ahead forecast $\hat{z}_{v,\tau}(L)$ of $z_{v,\tau+L}$ is a linear function of current and previous $\zeta_{v,\tau}$, $\zeta_{v,\tau-1},\ldots$ and $\varepsilon_{v,\tau}$, $\varepsilon_{v,\tau-1},\ldots$ and has the form of

$$\hat{z}_{v,\tau}(L) = \sum_{j=0}^{\infty} \nu^*_{L+j,\tau+L}\zeta_{v,\tau-j} + \sum_{j=0}^{\infty} \psi^*_{L+j,\tau+L}\varepsilon_{v,\tau-j} \tag{4.6.8}$$

where $\nu^*_{L,\tau+L}$, $\nu^*_{L+1,\tau+L},\ldots$ and $\Psi^*_{L,\tau+L}$, $\Psi^*_{L+1,\tau+L},\ldots$ are to be determined.

From Eqs. (4.6.7) and (4.6.8), we have

$$z_{v,\tau+L}-\hat{z}_{v,\tau}(L)=\sum_{j=0}^{L-1}(\nu_{j,\tau+L}\zeta_{v,\tau+L-j}+\psi_{j,\tau+L}\varepsilon_{v,\tau+L-j})$$

$$+\sum_{j=0}^{\infty}[(\nu_{L+j,\tau+L}-\nu^{*}_{L+j,\tau+L})\zeta_{v,\tau-j}$$

$$+(\psi_{L+j,\tau+L}-\psi^{*}_{L+j,\tau+L})\varepsilon_{v,\tau-j}\ ] \qquad (4.6.9)$$

and

$$E[z_{v,\tau+L}-\hat{z}_{v,\tau}(L)]^2=(\nu^2_{0,\tau}+\nu^2_{1,\tau}+\ldots+\nu^2_{L-1,\tau})\sigma^2_{\zeta}$$
$$+(\ \psi^2_{0,\tau}\ +\psi^2_{1,\tau}+\ldots+\psi^2_{L-1,\tau})\sigma^2_{\varepsilon}$$
$$+\sum_{j=0}^{\infty}[(\nu_{L+j,\tau}-\nu^{*}_{L+j,\tau})^2\sigma^2_{\zeta}$$
$$+(\psi_{L+j,\tau}-\psi^{*}_{L+j,\tau})^2\sigma_{\varepsilon}{}^2] \qquad (4.6.10)$$

which is minimized only if $\nu^{*}_{L+j,\tau+L}=\nu_{L+j,\tau+L}$ and $\Psi^{*}_{L+j,\tau+L}=\Psi_{L+j,\tau+L}$. Thus the minimum mean square error forecast $\hat{z}_{v,\tau}(L)$ of $z_{v,\tau+L}$ at time τ is the conditional expectation of $z_{v,\tau+L}$ at time τ given that the series from infinite past up to the present origin τ is known.

The process equation defined at time $v,\tau+L$ can be obtained from Eq. (4.3.13) as

$$z_{v,\tau+L}=\varphi_{1,\tau+L}z_{v,\tau+L-1}+\sum_{j=0}^{\infty}a_j x_{v,\tau+L-j}$$

$$+\frac{1-\theta_1 B-\ldots-\theta_q B^q}{1-\phi_1 B-\ldots-\phi_p B^p}\varepsilon_{v,\tau+L} \qquad (4.6.11)$$

or

$$z_{v,\tau+L}=\phi_1 z_{v,\tau+L-1}+\ldots+\phi_p z_{v,\tau+L-p}$$

$$+\ (\varphi_{1,\tau+L} z_{v,\tau+L-1}-\phi_1\varphi_{1,\tau+L-1} z_{v,\tau+L-2}-\cdots$$

$$-\phi_p\varphi_{1,\tau+L-p} z_{v,\tau+L-p-1})+\ (\sum_{j=0}^{\infty} a_j x_{v,\tau+L-j}$$

$$-\phi_1\sum_{j=0}^{\infty} a_j x_{v,\tau+L-j-1}-\ldots-\phi_p\sum_{j=0}^{\infty} a_j x_{v,\tau+L-j-p})$$

$$+\ \ (\varepsilon_{v,\tau+L}-\theta_1\varepsilon_{\tau+L-1}-\ldots-\theta_q\varepsilon_{v,\tau+L-q})$$

Then, by letting square brackets denote conditional expectation, we have

$$\hat{z}_{v,\tau}(L)=\phi_1[z_{v,\tau+L-1}]+\phi_2[z_{v,\tau+L-2}]+\ldots+\phi_p[z_{v,\tau+L-p}]$$

$$+\varphi_{1,\tau+L}[z_{v,\tau+L-1}]-\phi_1\varphi_{1,\tau+L-1}[z_{v,\tau+L-2}]-\cdots$$

$$-\phi_p\varphi_{1,\tau+L-p}[z_{v,\tau+L-p-1}]+\sum_{j=0}^{\infty} a_j[x_{v,\tau+L-j}]$$

$$-\phi_1\sum_{j=0}^{\infty} a_j[x_{v,\tau+L-j-1}]-\ldots-\phi_p\sum_{j=0}^{\infty} a_j[x_{v,\tau+L-j-p}]$$

$$+[\varepsilon_{v,\tau+L}]-\theta_1[\varepsilon_{v,\tau+L-1}]-\ldots-\theta_q[\varepsilon_{v,\tau+L-q}] \qquad (4.6.12)$$

where $[x_{v,\tau+L}] = E[x_{v,\tau+L}|x_{v,\tau},\ x_{v,\tau-1},\ldots]$, $[z_{v,\tau+L}] = E[z_{v,\tau+L}|z_{v,\tau},\ z_{v,\tau-1},\ldots]$, $[\varepsilon_{v,\tau+L}]=E[\varepsilon_{v,\tau+L}|\varepsilon_{v,\tau},\ \varepsilon_{v,\tau-1},\ldots]$. Furthermore,

$$[z_{v,\tau+j}] = \begin{cases} z_{v,\tau+j} & j\le 0 \\ \hat{z}_{v,\tau}(j) & j>0 \end{cases} \qquad (4.6.13a)$$

$$[x_{v,\tau+j}] = \begin{cases} x_{v,\tau+j} & j\le 0 \\ \hat{x}_{v,\tau}(j) & j>0 \end{cases} \qquad (4.6.13b)$$

and

$$[\varepsilon_{v,\tau+j}] = \begin{cases} \varepsilon_{v,\tau+j} & j \le 0 \\ 0 & j > 0 \end{cases} \tag{4.6.13c}$$

Note that, in Eq. (4.6.12) the forecast of the input series $[x_{v,\tau+j}]$ for $j > 0$ is based on the AR(p) process.

From Eq. (4.6.7), the forecast error for lead time L is a linear function of $\varepsilon_{v,\tau+1}$, $\varepsilon_{v,\tau+2}$,..., i.e.

$$e_{v,\tau}(L) = \sum_{j=0}^{L-1} \nu_{j,\tau+L} \zeta_{v,\tau+L-j} + \sum_{j=0}^{L-1} \psi_{j,\tau+L} \varepsilon_{v,\tau+L-j} \tag{4.6.14}$$

Since $E[e_{v,\tau}(L)] = 0$, the forecast is unbiased. Also, the variance of the forecast error can be written as

$$\begin{aligned} Var[e_{v,\tau}(L)] &= (\nu_{0,\tau+L}^2 + \nu_{1,\tau+L}^2 + \ldots + \nu_{L-1,\tau+L}^2)\sigma_\zeta^2 \\ &\quad + (\psi_{0,\tau+L}^2 + \psi_{1,\tau+L}^2 + \ldots + \psi_{L-1,\tau+L}^2)\sigma_\varepsilon^2 \\ &= \sigma_\zeta^2 \sum_{j=0}^{L-1} \nu_{j,\tau+L}^2 + \sigma_\varepsilon^2 \sum_{j=0}^{L-1} \psi_{j,\tau+L}^2 \end{aligned} \tag{4.6.15}$$

If the forecast error is a normally distributed, the 95% probability limits of the forecast with lead time L can be written as

$$z_{v,\tau+L}(\pm) = \hat{z}_{v,\tau}(L) \pm 1.96 \sqrt{\sigma_\zeta^2 \sum_{j=0}^{L-1} \nu_{j,\tau+L}^2 + \sigma_\varepsilon^2 \sum_{j=0}^{L-1} \psi_{j,\tau+L}^2} \tag{4.6.16}$$

CHAPTER V

MULTIPLE INPUT-SINGLE OUTPUT PERIODIC TRANSFER FUNCTION MODEL

5.1 Introduction

In this chapter, the "multiple input-single output" periodic transfer function forecast model will be presented. A two input-single output transfer function model will be presented as the basis for deriving the general form of the multiple input-single output transfer function model. In addition, some computing procedure will be presented.

5.2 Multiple Input-Single Output Periodic Transfer Function Model

5.2.1 Model Identification

A system with two inputs $x_t^{(1)}$ and $x_t^{(2)}$ and a single output y_t will be presented first. Then, it well be generalized. This system has the impulse response function $a_j^{(1)}$ and $a_j^{(2)}$ and the corresponding frequency response functions $A_1(f)$ and $A_2(f)$. Assume that the physical system is linear over the range of interest, and the system can be contaminated by the noise N_t. Let the inputs $x_t^{(1)}$ and $x_t^{(2)}$ be the deseasonalized SWE and temperature series respectively, and $z_{v,\tau}$ be the deseasonalized monthly streamflow. First, $z_{v,\tau}$ is filtered by a PAR(1) model

$$z_{v,\tau} = \varphi_{1,\tau} z_{v,\tau-1} + y_{v,\tau} \qquad (5.2.1)$$

Then, the residual $y_{v,\tau}$ is considered the output of the system. The response functions $a_j^{(1)}$ and $A_1(f)$ correspond to the input $x_t^{(1)}$ and the response functions $a_j^{(2)}$ and $A_2(f)$ correspond to the input $x_t^{(2)}$.

The output y_t ($t=(v-1)12+\tau$) as a function of the inputs $x_t^{(1)}$ and $x_t^{(2)}$ may be

expressed as

$$y_t = \sum_{j=-\infty}^{\infty} a_j^{(1)} x_{t-j}^{(1)} + \sum_{j=-\infty}^{\infty} a_j^{(2)} x_{t-j}^{(2)} + N_t$$
$$= \sum_{j=-\infty}^{\infty} \underline{a}_j' \underline{x}_{t-j} + N_t \qquad (5.2.2)$$

in which the symbol ' denotes transpose of a matrix, N_t is assumed to be uncorrelated with inputs $x_t^{(1)}$ and $x_t^{(2)}$, and

$$\underline{a}_j = \begin{bmatrix} a_j^{(1)} \\ a_j^{(2)} \end{bmatrix} , \qquad \underline{x}_{t-j} = \begin{bmatrix} x_{t-j}^{(1)} \\ x_{t-j}^{(2)} \end{bmatrix}$$

Multiplying both sides of Eq. (5.2.2) by $[x_{t-k}^{(1)}, x_{t-k}^{(2)}]$, we have

$$[y_t][x_{t-k}^{(1)} \; x_{t-k}^{(2)}] = \sum_{j=-\infty}^{\infty} \begin{bmatrix} a_j^{(1)} & a_j^{(2)} \end{bmatrix} \begin{bmatrix} x_{t-j}^{(1)} \\ x_{t-j}^{(2)} \end{bmatrix} \begin{bmatrix} x_{t-k}^{(1)} & x_{t-k}^{(2)} \end{bmatrix}$$
$$+ [N_t][x_{t-k}^{(1)} \; x_{t-k}^{(2)}]$$

or

$$\begin{bmatrix} y_t x_{t-k}^{(1)} & y_t x_{t-k}^{(2)} \end{bmatrix} = \sum_{j=-\infty}^{\infty} \begin{bmatrix} a_j^{(1)} & a_j^{(2)} \end{bmatrix} \begin{bmatrix} x_{t-j}^{(1)} x_{t-k}^{(1)} & x_{t-j}^{(1)} x_{t-k}^{(2)} \\ x_{t-j}^{(2)} x_{t-k}^{(1)} & x_{t-j}^{(2)} x_{t-k}^{(2)} \end{bmatrix}$$
$$+ \begin{bmatrix} N_t x_{t-k}^{(1)} & N_t x_{t-k}^{(2)} \end{bmatrix} \qquad (5.2.3)$$

Then, by taking expectation in Eq. (5.2.3) gives

$$[R_{yx_1}(k) \; R_{yx_2}(k)] = \sum_{j=-\infty}^{\infty} [a_j^{(1)} \; a_j^{(2)}] \begin{bmatrix} R_{x_1x_1}(k-j) & R_{x_1x_2}(k-j) \\ R_{x_2x_1}(k-j) & R_{x_2x_2}(k-j) \end{bmatrix}$$

in which $R_{yx}(k)$ denotes the lag-k cross-covariance between y and x.

However, from Eq. (3.7.10), we have

$$[\int_{-1/2}^{1/2} S_{yx_1}(f)e^{i2\pi fk}df \qquad \int_{-1/2}^{1/2} S_{yx_2}(f)e^{i2\pi fk}df]$$

$$= \sum_{j=-\infty}^{\infty} [a_j^{(1)} \; a_j^{(2)}]$$

$$\begin{bmatrix} \int_{-1/2}^{1/2} S_{x_1x_1}(f)e^{i2\pi f(k-j)}df & \int_{-1/2}^{1/2} S_{x_1x_2}e^{i2\pi f(k-j)}df \\ \int_{-1/2}^{1/2} S_{x_2x_1}e^{i2\pi f(k-j)}df & \int_{-1/2}^{1/2} S_{x_2x_2}(f)e^{i2\pi f(k-j)}df \end{bmatrix}$$

$$=[\ \{\ \int_{-1/2}^{1/2} S_{x_1x_1}(f)e^{i2\pi fk}(\sum_{j=-\infty}^{\infty} a_j^{(1)}e^{-i2\pi fj})df$$

$$+ \int_{-1/2}^{1/2} S_{x_2x_1}(f)e^{i2\pi fk}(\sum_{j=-\infty}^{\infty} a_j^{(2)}e^{-i2\pi fj})df\ \}$$

$$\{\int_{-1/2}^{1/2} S_{x_1x_2}(f)e^{i2\pi fk}(\sum_{j=-\infty}^{\infty} a_j^{(1)}e^{-i2\pi fj})df$$

$$+ \int_{-1/2}^{1/2} S_{x_2x_2}(f)e^{i2\pi fk}(\sum_{j=-\infty}^{\infty} a_j^{(2)}e^{-i2\pi fj})df\ \}\ ]$$

Hence

$$[\ S_{yx_1}(f) \qquad S_{yx_2}(f)\ ]$$

$$=[A_1(f)S_{x_1x_1}(f)+A_2(f)S_{x_2x_1}(f) \qquad A_1(f)S_{x_1x_2}(f)+A_2(f)S_{x_2x_2}(f)]$$

$$=[A_1(f) \quad A_2(f)]\begin{bmatrix} S_{x_1x_1}(f) & S_{x_1x_2}(f) \\ S_{x_2x_1}(f) & S_{x_2x_2}(f) \end{bmatrix} \qquad (5.2.4)$$

where

$$A_1(f) = \sum_{j=-\infty}^{\infty} a_j^{(1)} e^{-i2\pi fj}$$

$$A_2(f) = \sum_{j=-\infty}^{\infty} a_j^{(2)} e^{-i2\pi fj}$$

Furthermore, Eq. (5.2.4) can be expressed as

$$\underline{S}'_{yx} = \underline{A}'(f)\underline{S}_{xx}(f) \tag{5.2.5}$$

Then, the frequency response function matrix can be estimated by

$$\underline{A}'(f) = \underline{S}'_{yx}(f)\ \underline{S}^{-1}_{xx}(f) \tag{5.2.6}$$

or

$$[A_1(f)\ A_2(f)] = [S_{yx_1}(f)\ S_{yx_2}(f)] \begin{bmatrix} S_{x_1x_1}(f) & S_{x_1x_2}(f) \\ S_{x_2x_1}(f) & S_{x_2x_2}(f) \end{bmatrix}^{-1} \tag{5.2.7}$$

Likewise, the impulse response function can be obtained by using the inverse Fourier transform of Eq. (5.2.7), i.e.

$$a_j^{(1)} = \int_{-1/2}^{1/2} A_1(f) e^{i2\pi fj} df \tag{5.2.8}$$

$$a_j^{(2)} = \int_{-1/2}^{1/2} A_2(f) e^{i2\pi fj} df \tag{5.2.9}$$

The matrix $\underline{S}_{xx}(f)$ can be obtained by computing the co-spectrum and quad-spectrum, i.e.

$$\underline{S}_{xx}(f) = \begin{bmatrix} c_{x_1x_1}(f) + i\ q_{x_1x_1}(f) & c_{x_1x_2}(f) + i\ q_{x_1x_2}(f) \\ c_{x_2x_1}(f) + i\ q_{x_2x_1}(f) & c_{x_2x_2}(f) + i\ q_{x_2x_2}(f) \end{bmatrix}$$

$$= \begin{bmatrix} c_{x_1x_1}(f) & c_{x_1x_2}(f) \\ c_{x_2x_1}(f) & c_{x_2x_2}(f) \end{bmatrix} + i \begin{bmatrix} q_{x_1x_1}(f) & q_{x_1x_2}(f) \\ q_{x_2x_1}(f) & q_{x_2x_2}(f) \end{bmatrix}$$

$$= \underline{c}_{xx}(f) + i\ \underline{q}_{xx}(f) \tag{5.2.10}$$

Let the inverse matrix of $\underline{S}_{xx}(f)$ be defined as

$$\underline{S}^{-1}_{xx}(f) = \underline{c}^{*}_{xx}(f) + i\underline{q}^{*}_{xx}(f) \tag{5.2.11}$$

Multiplying Eq. (5.2.10) by Eq. (5.2.11) gives

$$(\underline{c}_{xx}(f) + i\underline{q}_{xx}(f))(\underline{c}^{*}_{xx}(f) + i\underline{q}^{*}_{xx}(f)) = \underline{I} + i\underline{0} \tag{5.2.12}$$

where $\underline{I}$ is the identity matrix and $\underline{0}$ is a zero matrix. Then,

$$\underline{c}_{xx}(f)\underline{c}^{*}_{xx}(f) - \underline{q}_{xx}(f)\underline{q}^{*}_{xx}(f) = \underline{I} \tag{5.2.13}$$

$$\underline{c}_{xx}(f)\underline{q}^{*}_{xx}(f) + \underline{q}_{xx}(f)\underline{c}^{*}_{xx}(f) = \underline{0} \tag{5.2.14}$$

Thus, $\underline{q}^{*}_{xx}(f)$ can be obtained from Eq. (5.2.14) as

$$\underline{q}^{*}_{xx}(f) = -\underline{c}_{xx}(f)^{-1}\underline{q}_{xx}(f)\underline{c}^{*}_{xx}(f) \tag{5.2.15}$$

Substituting Eq. (5.2.15) into Eq. (5.2.13) gives

$$\underline{c}_{xx}(f)\underline{c}^{*}_{xx}(f) + \underline{q}_{xx}(f)\underline{c}_{xx}(f)^{-1}\underline{q}_{xx}(f)\underline{c}^{*}_{xx}(f) = \underline{I}$$

or

$$\underline{c}^{*}_{xx}(f) = (\ \underline{c}_{xx}(f) + \underline{q}_{xx}(f)\underline{c}_{xx}(f)^{-1}\underline{q}_{xx}(f)\ )^{-1} \tag{5.2.16}$$

Hence, the matrix $\underline{S}^{1}{}_{xx}(f)$ of Eq. (5.2.11) can be obtained.

Next, write

$$
\begin{aligned}
&[S_{yx_1}(f) \quad S_{yx_2}(f)] \\
&= [c_{yx_1}(f)+iq_{yx_1}(f) \quad c_{yx_2}(f)+iq_{yx_2}(f)] \\
&= [c_{yx_1}(f) \quad c_{yx_2}(f)]+i[q_{yx_1}(f) \quad q_{yx_2}(f)] \\
&= \underline{c}_{yx}(f) + i\,\underline{q}_{yx}
\end{aligned} \tag{5.2.17}
$$

Then, the frequency response function of Eq. (5.2.6) can be computed based on Eqs. (5.2.11), (5.2.15), (5.2.16) and (5.2.17), i.e.

$$
\begin{aligned}
\underline{A}'(f) &= (\underline{c}_{yx}+i\underline{q}_{yx})(\underline{c}^*_{xx}(f)+i\underline{q}^*_{xx}(f)) \\
&= (\,\underline{c}_{yx}(f)\underline{c}^*_{xx}(f) - \underline{q}_{yx}(f)\underline{q}^*_{xx}(f)\,) \\
&\quad + i\,(\,\underline{c}_{yx}(f)\underline{q}^*_{xx}(f) + \underline{q}_{yx}(f)\underline{c}^*_{xx}(f)\,)
\end{aligned} \tag{5.2.18}
$$

where

$$
\underline{c}_{yx}(f) = [c_{yx_1}(f) \quad c_{yx_2}(f)]
$$

$$
\underline{q}_{yx}(f) = [q_{yx_1}(f) \quad q_{yx_2}(f)]
$$

$$
\underline{c}^*_{xx}(f) = \begin{bmatrix} c^*_{x_1x_1}(f) & c^*_{x_1x_2}(f) \\ c^*_{x_2x_1}(f) & c^*_{x_2x_2}(f) \end{bmatrix}
$$

$$
\underline{q}^*_{xx}(f) = \begin{bmatrix} q^*_{x_1x_1}(f) & q^*_{x_1x_2}(f) \\ q^*_{x_2x_1}(f) & q^*_{x_2x_2}(f) \end{bmatrix}
$$

From the above development of two input-single output system, a general form of the N input-single output system can be written as

$$y_t = \sum_{j=-\infty}^{\infty} a_j^{(1)} x_{t-j}^{(1)} + \sum_{j=-\infty}^{\infty} a_j^{(2)} x_{t-j}^{(2)} + \ldots + \sum_{j=-\infty}^{\infty} a_j^{(N)} x_{t-j}^{(N)} + N_t$$

$$= \sum_{j=-\infty}^{\infty} \underline{a}_j' \, \underline{x}_{t-j} + N_t \qquad (5.2.19)$$

where N_t is uncorrelated with inputs $x_t^{(1)}$, $x_t^{(2)}, \ldots, x_t^{(3)}$, and

$$\underline{a}_j = \begin{bmatrix} a_j^{(1)} \\ a_j^{(2)} \\ \cdot \\ \cdot \\ \cdot \\ a_j^{(N)} \end{bmatrix}, \qquad \underline{x}_{t-j} = \begin{bmatrix} x_{t-j}^{(1)} \\ x_{t-j}^{(2)} \\ \cdot \\ \cdot \\ \cdot \\ x_{t-j}^{(N)} \end{bmatrix}$$

From Eq. (5.2.6) the frequency response function matrix becomes

$$\underline{A}'(f) = \underline{S}_{yx}'(f) \, \underline{S}_{xx}(f)^{-1} \qquad (5.2.20)$$

where

$$\underline{A}(f) = \begin{bmatrix} A_1(f) \\ A_2(f) \\ \vdots \\ A_N(f) \end{bmatrix}, \qquad \underline{S}_{yx}(f) = \begin{bmatrix} S_{yx_1}(f) \\ S_{yx_2}(f) \\ \vdots \\ S_{yx_N}(f) \end{bmatrix}$$

and

$$\underline{S}_{xx}(f) = \begin{bmatrix} S_{x_1x_1}(f) & \cdots & S_{x_1x_N}(f) \\ \vdots & \ddots & \vdots \\ S_{x_Nx_1}(f) & \cdots & S_{x_Nx_N}(f) \end{bmatrix}$$

Finally, the impulse response function can be estimated by

$$\underline{a}_j' = \int_{-1/2}^{1/2} \underline{A}'(f) e^{i2\pi fj} df \tag{5.2.21}$$

The mean square error of the transfer function model can now be obtained by using the same method as in Eq. (4.3.6). Thus, we have

$$\begin{aligned} MSE &= \int_{-1/2}^{1/2} [S_y(f) - \underline{A}'(f)\underline{S}_{xy}(f)] df \\ &= \int_{-1/2}^{1/2} [S_y(f) - \underline{S}_{yx}'(f)\underline{S}_{xx}^{-1}(f)\underline{S}_{xy}(f)] df \\ &= \int_{-1/2}^{1/2} \underline{S}_{y\cdot x}(f) df \end{aligned} \tag{5.2.22}$$

where

$$\underline{S}_{y\cdot x}(f) = S_y(f) - \underline{S}_{yx}'(f)\underline{S}_{xx}^{-1}(f)\underline{S}_{xy}(f) \tag{5.2.23}$$

is called the conditional power spectrum.

Now, let M be the bandwidth parameter of the windows used for estimation of spectrum. The properties of the impulse response function estimator of Eq. (5.2.21) follow as in Wahba (1968; 1969), Brillinger (1981) and Shumway (1988). This leads to the conclusion that the estimator of the impulse response function has variance-covariance matrix estimated by

$$Cov(\underline{\hat{a}}_t) = \frac{1}{M^3} \sum_{k=0}^{M-1} \underline{\hat{S}}_{y\cdot x}(f_k) [\underline{\hat{S}}_{xx}'(f_k)]^{-1} \tag{5.2.24}$$

The noise term N_t in Eq. (5.2.19) can now be estimated. An ARMA process can be fitted to N_t, i.e.

$$\phi(B)N_t = \theta(B)\,\varepsilon_t \tag{5.2.25}$$

in which ε_t has zero mean and variance σ_ε^2. Therefore, substituting Eqs. (5.2.25) and (5.2.19) into (5.2.1) gives a preliminary model

$$z_{v,\tau} = \varphi_{1,\tau} z_{v,\tau-1} + \sum_{j=0}^{D} \underline{a}_j' \underline{x}_{v,\tau-j} + \frac{\theta(B)}{\phi(B)}\varepsilon_{v,\tau} \tag{5.2.26}$$

Hence the structure of the multiple input-single output transfer function model has been identified.

5.2.2 Parameter Estimation and Diagnostic Checking

The parameters in Eq. (5.2.26), $\varphi_{1,\tau}$, ϕ(B), θ(B) and $\underline{a}_j$, can be estimated by using the nonlinear least squares method. The Marquardt algorithm of nonlinear least squares method is used in this research. The diagnostic checking will include the checking of the residual ε_t to be white noise, and checking that the noise N_t and the inputs $x_t^{(1)}$, $x_t^{(2)},\ldots,x_t^{(N)}$ are uncorrelated.

5.2.3 Forecast Function

The multiple input-single output transfer function model of Eq. (5.2.26) can be written as

$$z_{v,\tau+L} = \varphi_{1,\tau+L} z_{v,\tau+L-1} + \sum_{j=0}^{D_1} a_j^{(1)} x_{v,\tau+L-j}^{(1)} + \sum_{j=0}^{D_2} a_j^{(2)} x_{v,\tau+L-j}^{(2)} + \ldots$$
$$+ \sum_{j=0}^{D_N} a_j^{(N)} x_{v,\tau+L-j}^{(N)} + \frac{1-\theta_1 B-\ldots-\theta_q B^q}{1-\phi_1 B-\ldots-\phi_p B^p}\varepsilon_{v,\tau+L} \tag{5.2.27}$$

or

$$
\begin{aligned}
z_{v,\tau+L} &= \phi_1 z_{v,\tau+L-1} + \ldots + \phi_p z_{v,\tau+L-p} \\
&+ (\varphi_{1,\tau+L} z_{v,\tau+L-1} - \phi_1 \varphi_{1,\tau+L-1} z_{v,\tau+L-2} - \cdots \\
&- \phi_p \varphi_{1,\tau+L-p} z_{v,\tau+L-1-p}) + (\sum_{j=0}^{D_1} a_j^{(1)} x_{v,\tau+L-j}^{(1)} \\
&- \phi_1 \sum_{j=0}^{D_1} a_j^{(1)} x_{v,\tau+L-j-1}^{(1)} - \ldots - \phi_p \sum_{j=0}^{D_1} a_j^{(1)} x_{v,\tau+L-j-p}^{(1)}) \\
&+ \ldots\ldots \\
&+ (\sum_{j=0}^{D_N} a_j^{(N)} x_{v,\tau+L-j}^{(N)} - \phi_1 \sum_{j=0}^{D_N} a_j^{(N)} x_{v,\tau+L-j-1}^{(N)} - \ldots\ldots \\
&- \phi_p \sum_{j=0}^{D_N} a_j^{(N)} x_{v,\tau+L-j-p}^{(N)}) \\
&+ (\varepsilon_{v,\tau+L} - \theta_1 \varepsilon_{v,\tau+L-1} - \ldots - \theta_q \varepsilon_{v,\tau+L-q})
\end{aligned}
$$

Then, the forecast equation with lead time L can be written as

$$
\begin{aligned}
\hat{z}_{v,\tau}(L) &= (\phi_1 [z_{v,\tau+L-1}] + \ldots + \phi_p [z_{v,\tau+L-p}]) \\
&+ (\varphi_{1,\tau+L} [z_{v,\tau+L-1}] - \phi_1 \varphi_{1,\tau+L-1} [z_{v,\tau+L-2}] - \cdots \\
&- \phi_p \varphi_{1,\tau+L-p} [z_{v,\tau+L-1-p}]) + (\sum_{j=0}^{D_1} a_{j,1} [x_{v,\tau+L-j}^{(1)}] \\
&- \phi_1 \sum_{j=0}^{D_1} a_{j,1} [x_{v,\tau+L-j-1}^{(1)}] - \ldots - \phi_p \sum_{j=0}^{D_1} a_{j,1} [x_{v,\tau+L-j-p}^{(1)}] \quad \cdots \\
&+ (\sum_{j=0}^{D_N} a_{j,N} [x_{v,\tau+L-j}^{(N)}] - \phi_1 \sum_{j=0}^{D_N} a_{j,N} [x_{v,\tau+L-j-1}^{(N)}] \\
&- \phi_p \sum_{j=0}^{D_N} a_{j,N} [x_{v,\tau+L-j-p}^{(N)}]) \\
&+ ([\varepsilon_{v,\tau+L}] - \theta_1 [\varepsilon_{v,\tau+L-1}] - \ldots - \theta_q [\varepsilon_{v,\tau+L-q}])
\end{aligned}
\qquad (5.2.28)
$$

where

$$[z_{\nu,\tau+j}] = \begin{cases} z_{\nu,\tau+j} & j \le 0 \\ \hat{z}_{\nu,\tau}(j) & j > 0 \end{cases} \tag{5.2.29a}$$

$$[x^{(k)}_{\nu,\tau+j}] = \begin{cases} x^{(k)}_{\nu,\tau+j} & j \le 0 \\ \hat{x}^{(k)}_{\nu,\tau}(j) & j > 0 \end{cases} \tag{5.2.29b}$$

$$[\varepsilon_{\nu,\tau+j}] = \begin{cases} \varepsilon_{\nu,\tau+j} & j \le 0 \\ 0 & j > 0 \end{cases} \tag{5.2.29c}$$

In the foregoing formulation it is also necessary to forecast the inputs. Let us assume that the deseasonalized inputs were fitted by the ARMA processes

$$\left.\begin{aligned} \Phi_x^{(1)}(B)x_t^{(1)} &= \Theta_x^{(1)}(B)\zeta_t^{(1)} \\ \Phi_x^{(2)}(B)x_t^{(2)} &= \Theta_x^{(2)}(B)\zeta_t^{(2)} \\ &\vdots \\ \Phi_x^{(N)}(B)x_t^{(N)} &= \Theta_x^{(N)}(B)\zeta_t^{(N)} \end{aligned}\right\} \tag{5.2.30}$$

where $\zeta_t^{(1)}$, $\zeta_t^{(2)}$, ..., $\zeta_t^{(N)}$ are white noise with mean zero and variance $\sigma_\zeta^2(j)$. Then, Eq. (5.6.26) can be written as

$$\begin{aligned} z_{\nu,\tau} &= \frac{A_1(B)}{(1-\varphi_{1,\tau}B)}\frac{\Theta_{x,1}(B)}{\Phi_{x,1}(B)}\zeta_{\nu,\tau}^{(1)} + \frac{A_2(B)}{(1-\varphi_{1,\tau}B)}\frac{\Theta_{x,2}(B)}{\Phi_{x,2}(B)}\zeta_{\nu,\tau}^{(2)} + \cdots \\ &\quad + \frac{A_N(B)}{(1-\varphi_{1,\tau}B)}\frac{\Theta_{x,N}(B)}{\Phi_{x,N}(B)}\zeta_{\nu,\tau}^{(N)} + \frac{1}{(1-\varphi_{1,\tau}B)}\frac{\theta(B)}{\phi(B)}\varepsilon_{\nu,\tau} \\ &= \nu_\tau^{(1)}(B)\zeta_{\nu,\tau}^{(1)} + \nu_\tau^{(2)}(B)\zeta_{\nu,\tau}^{(2)} + \cdots + \nu_\tau^{(N)}(B)\zeta_{\nu,\tau}^{(N)} \\ &\quad + \psi_\tau(B)\varepsilon_{\nu,\tau} \end{aligned} \tag{5.2.31}$$

where

$$A_j(B) = a_0^{(j)} + a_1^{(j)}B + a_2^{(j)}B^2 + \ldots.$$

$$\nu_\tau^{(j)}(B) = \nu_{0,\tau}^{(j)} + \nu_{1,\tau}^{(j)}B + \nu_{2,\tau}^{(j)}B^2 + \ldots.$$

$$\psi_\tau(B) = \psi_{0,\tau} + \psi_{1,\tau}B + \psi_{2,\tau}B^2 + \ldots.$$

and

$$\nu_\tau^{(j)}(B) = \frac{A_j(B)}{(1-\varphi_{1,\tau}B)}\frac{\theta_x^{(j)}(B)}{\Phi_x^{(j)}(B)}$$

$$\psi_\tau(B) = \frac{1}{(1-\varphi_{1,\tau}B)}\frac{\theta(B)}{\phi(B)}$$

Finally Eq. (5.2.31) can be rewritten as

$$z_{v,\tau+L} = \sum_{j=0}^{\infty} \nu_{j,\tau+L}^{(1)}\zeta_{v,\tau+L-j}^{(1)} + \sum_{j=0}^{\infty} \nu_{j,\tau+L}^{(2)}\zeta_{v,\tau+L-j}^{(2)} + \ldots$$

$$+ \sum_{j=0}^{\infty} \nu_{j,\tau+L}^{(N)}\zeta_{v,\tau+L-j}^{(N)} + \sum_{j=0}^{\infty} \psi_{j,\tau+L}\varepsilon_{v,\tau+L-j}$$

with $\zeta_{v,\tau}^{(1)}$, $\zeta_{v,\tau}^{(2)}, \ldots, \zeta_{v,\tau}^{(N)}$ and $\varepsilon_{v,\tau}$ uncorrelated.

From the discussion of Section 4.6, we know that the minimum mean square error forecast $\hat{z}_{v,\tau}(L)$ of $z_{v,\tau+L}$ at time τ is given by the conditional expectation of $z_{v,\tau+L}$ at time τ. Hence, the forecast error for lead time L can be expressed as

$$e_{v,\tau}(L) = \sum_{j=0}^{L-1} \nu_{j,\tau+L}^{(1)}\zeta_{v,\tau+L-j}^{(1)} + \sum_{j=0}^{L-1} \nu_{j,\tau+L}^{(2)}\zeta_{v,\tau+L-j}^{(2)} + \ldots$$

$$+ \sum_{j=0}^{L-1} \nu_{j,\tau+L}^{(N)}\zeta_{v,\tau+L-j}^{(N)} + \sum_{j=0}^{L-1} \psi_{j,\tau+L}\varepsilon_{v,\tau+L-j} \qquad (5.2.32)$$

CHAPTER VI

DATA DESCRIPTION

6.1 Watershed Description

The Rio Grande River system in Colorado was selected as the pilot project to implement and test the procedures developed in this research. The Rio Grande River begins in Southern Colorado and passes through the states of Colorado, New Mexico and Texas finally ending in the Republic of Mexico. It is an important water supply for these states. The streamflow forecasting of the Rio Grande Basin in Colorado is important to both agricultural development and fulfilling Colorado State's obligation to deliver water from the Rio Grande to the Colorado-New Mexico stateline.

The release of any acre-foot of water from the Rio Grande to New Mexico is considered a loss to Colorado. The amount of water sent to New Mexico each year depends on the amount of water available. In normal years, Colorado is required under "the Rio Grande Basin Compact" to provide about 300,000 acre-feet of water to New Mexico and Texas. One acre-foot is about 325,000 gallons, or the amount of water used by a typical family of four in a year. According to the Rio Grande Basin Compact, a stream gaging station should be operated and maintained at each of the following points in Colorado (see Fig. 6.1): (a) the Rio Grande River near Del Norte, (b) the Conejos River near Mogote, (c) the Los Pinos River near Ortiz, and (d) the San Antonio River at Ortiz.

The streamflow of the Rio Grande River Basin is dominated by snowpack accumulation and melt. This results in a high flow beginning in early April and

persisting through the summer. Snowpack, precipitation in general and temperature are important factors for streamflow forecast. This research uses snow water equivalent data, temperature data, and precipitation data as inputs to the system, and streamflow data as the output.

6.2 Monthly Streamflow Data

The monthly streamflow (1948-1982) of the Los Pinos River near Ortiz is shown in Fig. 6.2. The gaging station is located at latitude 36:58:56 and longitude 106:04:23, and elevation 8040 feet above sea level. It has a drainage area of 167 sq-mi and the streamflow record is from 1915 to the present. Since the record of snow water equivalent is available only from 1948, the streamflow record used in this report will be also from 1948. The monthly streamflow from 1948 to 1977 has been used by the Colorado State Engineer's Office to develop a multiple regression forecasting model. The monthly means and monthly standard deviations based on the record 1948-1977 are shown in Table 6.1 and Fig. 6.3. They show that flow after July has nearly returned to a baseflow condition. In order to use the time series technique for modeling streamflow, the monthly flow record was deseasonalized by using Eq. (3.2.1). This is shown in Fig. 6.4.

The monthly streamflow (1949-1982) of the Rio Grande River near Del Norte is shown in Fig. 6.5. The gaging stations located at latitude 37:41:22, longitude 106:27:38 and elevation 7980.25 feet, and it has a drainage area of 1320 sq-mi. The statistical properties of monthly streamflow 1949-1977 are shown in Table 6.2 and Fig. 6.6. The deseasonalized flows are shown in Fig. 6.7.

The monthly streamflow (1949-1982) of the Conejos River near Mogote is shown in Fig. 6.8. The gaging station is located at latitude 37:03:14, longitude 106:11:13, and elevation 8271.54 feet. It has a drainage area of 282 sq-mi and the streamflow record is from 1903 to the present. The statistical properties of monthly streamflow (1949-1977) are shown in Table 6.3 and Fig. 6.9. The deseasonalized flows are shown in Fig. 6.10.

6.3 Snow Water Equivalent Data

Snow pack implies storage of water and makes a great contribution to streamflow. In most cases, direct observations of snowcover are necessary to achieve reliable results in snowmelt run-off forecasting. Manual point measurements are the classical way to obtain high precision information on snowdepth and the snow water equivalent. Manual point measurements need a trained staff for data collection. It is a very efficient and reliable method especially in small basins or in simple terrain, where a few points are representative of the whole area. The disadvantages of point field measurements are that data collection is time consuming, and often delayed because of bad weather conditions. This is particularly true in mountain areas where safety considerations must be taken into account.

Traditionally, measurements of snow water equivalent are obtained once a month at snow courses throughout the mountainous region. The Soil Conservation Service each year publishes a series of snow survey data. This information summarizes the monthly snow water equivalent data collection during each water year. Recording instrumentation has been installed in recent years. Since these records of snow water equivalent data are short, of the order of 5 to 10 years, we will not consider this data in our analysis. In this research, only the snow water equivalent obtained by the "manual" point measurements

will be used.

There is no general rule that can be given for determining the areal values from point measurements. For the computation of basin values of average snow water equivalent, the methods are similar to those for precipitation. In this research, the Thiessen Polygon Approach is used to calculate the mean snow water equivalent.

The snow water equivalent data collection station of the Soil Conservation Service in the Rio Grande Basin of Southern Colorado is shown in Fig. 6.1. The manual point measurements of the snow water equivalent is always measured at the end of each month or on the first day of the next month. For example, the March 30 or the April 1st snow water equivalent will represent the existing snow water equivalent at the end of March. Another problem with manual point measurements is that the data collection of SWE (Snow Water Equivalent) are only available for the months of January through April, for the other months the data were missing.

6.3.1. The SWE of the Los Pinos River Basin

There is only one location (Cumbres Pass) for collecting the SWE in the Los Pinos River basin. Cumbres Pass is located at latitude 37°01', longitude 106°27', and elevation 10020 feet. The SWE at Cumbres Pass (1948-1986) is shown in Fig. 6.11. A more clear structure of the SWE is shown in Fig. 6.12 for the year 1948-1952. The SWE has missing observations from May to December. In order to find a more complete record of SWE, telemetered or recording instrumented SWE data at Cumbres Trestle, close to the Cumbres Pass, was examined. The snow pill at Cumbres Trestle (Oct. 1981 - Sep. 1989) and the SWE at Cumbres Pass are plotted in Fig. 6.13. This figure shows that the SWE in September can be assumed to be zero and that the SWE from September

to January can be assumed to increase linearly. Moreover, from Fig. 6.13, we can see that the SWE of June, July and Aug. can be assumed to be zero. Now, the only unknown SWE is in May. This can be forecasted by using a stochastic model. Thus, a complete data of SWE at Cumbres Pass can be obtained. Based on such complete data, the statistical properties of SWE at Cumbres Pass were computed and are shown in Table 6.4 and Fig. 6.14.

6.3.2 The SWE of the Rio Grande River Basin above Del Norte

In the Rio Grande River basin above Del Norte, there are four sites for collecting SWE data. The name and location of those four sites are listed below :

SITE NAME	ELEVATION (ft)	LATITUDE	LONGITUDE
Upper Rio Grande	9400	37°43'	107°15'
Santa Maria	9600	37°49'	107°07'
Lake Humphrey	9200	37°40'	106°52'
Wolf Creek Pass	11000	37°28'	106°48'

The missing SWE record for Oct., Nov. and Dec. was filled in by using linear interpolation. The SWE for May has some non-zero value at Wolf Creek Pass, but is equal to zero at the other three sites. In order to find the average SWE of this basin, the Thiessen polygon approach was used. The weights for each gaging station are shown below:

SITE NAME	SWE	WEIGHT
Upper Rio Grande	S_1	0.186
Santa Maria	S_2	0.221
Lake Humphrey	S_3	0.450
Wolf Creek Pass	S_4	0.143

Then, the average SWE can be obtained by using

$$average\ SWE = 0.186S_1 + 0.221S_2 + 0.45S_3 + 0.143S_4$$

The average SWE is shown in Fig. 6.15. The statistical properties of the average SWE are shown in Table 6.5 and are plotted in Fig. 6.16.

6.3.3 The SWE of the Conejos River

There are two gaging stations for collecting the SWE data in the Conejos River Basin. The name and location of these two stations are listed below :

SITE NAME	ELEVATION (ft)	LATITUDE	LONGITUDE
Platoro	9950	37°21'	106°33'
River Springs	9300	37°04'	106°16'

The missing SWE data for the Oct., Nov., and Dec. were filled in by using linear interpolation. The SWE of June, July, Aug. and Sep. is equal to zero in this basin. The average SWE was calculated using the Thiessen polygon approach. The weights for each gaging station are :

SITE NAME	SWE	WEIGHT
Platoro	S_1	0.5
River Springs	S_2	0.5

Thus,the average SWE is simply the average of the 2 sites. The average SWE is shown in Fig. 6.17. The statistical properties of the average SWE are shown in Table 6.6 and Fig. 6.18.

6.4 Temperature and Precipitation Data

The National Climate Data Center in Asheville, North Carolina, publishes each

year a series of climatological data which include: daily precipitation, daily maximum temperature and daily minimum temperature. In the Rio Grande River basin of southern Colorado, some of the climatological data records are too short, and some are missing. We found that only the gaging stations of the Hermit 7 ESE, Del Norte City and Manassa City have a complete record of the precipitation and temperature data. The location of these three gaging stations are shown in Fig. 6.1.

6.4.1 The Climatic Data of the Upper Rio Grande River Basin

The gaging stations of Hermit 7 ESE and Del Norte are located in the upper Rio Grande River basin. The name and the location of these two sites are listed below :

SITE NAME	ELEVATION (ft)	LATITUDE	LONGITUDE
Hermit 7 ESE	9000	37°46'	107°08'
Del Norte	7880	37°40'	106°21'

The average climate data of these two gaging stations will be used as inputs to the upper Rio Grande River basin. The monthly precipitation (1949-1983) is plotted in Fig. 6.19. The monthly statistical properties are shown in Table 6.7 and Fig. 6.20. The temperature data available from the National Climate Data Center is the daily maximum and daily minimum temperature. After computing the average temperature, the monthly average temperature can be obtained. The monthly temperature (1949-1983) is plotted in Fig. 6.21. The monthly statistical properties are shown in Table 6.8 and Fig. 6.22.

6.4.2 The Climatic Data of the Conejos River Basin

From Fig. 6.1, we can see that no climatic data gaging stations are located in the Conejos River Basin. The only climatic data available and close to the Conejos River

Basin is the climatic data of Manassa City. Hence the climatic data of Manassa City will be used as inputs to the Conejos River system. Since the historical precipitation record of Manassa City has a lot of missing data before 1966, the climate data after 1966 will be used for finding a model. The monthly precipitation (1966-1983) is shown in Fig. 6.23. The monthly statistical properties are shown in Table 6.9 and Fig. 6.24. The temperature data (1966-1983) is plotted in Fig. 6.25. The monthly statistical properties are shown in Table 6.10 and Fig. 6.26.

Table 6.1 The Monthly Mean and Standard Deviation of Streamflow (Los Pinos River near Ortiz, 1948-1977)

MONTH	MEAN	STANDARD DEVIATION
January	837.100	234.723
February	909.233	277.908
March	1903.867	1013.500
April	12062.870	6888.070
May	32025.830	16963.010
June	16308.870	14120.930
July	3839.067	3687.079
August	2031.200	1496.741
September	1223.000	855.834
October	1303.400	725.049
November	1140.633	467.002
December	951.300	302.836

Table 6.2 The Monthly Mean and Standard Deviation of Streamflow (Rio Grande River near Del Norte, 1949-1977)

MONTH	MEAN	STANDARD DEVIATION
January	9560.725	2139.658
February	9592.793	1825.739
March	14189.000	4070.236
April	39190.960	16929.990
May	135088.700	47149.870
June	162736.600	79068.310
July	78091.100	56614.470
August	41093.310	24659.170
September	23404.100	14423.590
October	20252.170	8474.996
November	13264.660	4449.579
December	10345.600	2422.469

Table 6.3 The Monthly Mean and Standard Deviation of Streamflow (Conejos River near Mogote, 1949-1977)

MONTH	MEAN	STANDARD DEVIATION
January	2814.759	550.494
February	2800.759	484.685
March	4783.069	1572.386
April	16372.900	7433.529
May	59158.790	18454.680
June	62670.000	29654.320
July	26202.960	23213.990
August	11595.590	8722.024
September	6070.896	4961.366
October	5611.828	3592.012
November	8413.793	7294.174
December	2968.345	798.695

Table 6.4 The Monthly Mean and Standard Deviation of SWE (Cumbres Pass; Los Pinos River Basin, 1948-1977)

MONTH	MEAN	STANDARD DEVIATION
January	13.056	5.504
February	16.540	6.459
March	18.757	6.716
April	13.600	9.586
May	-----	-----
June	0.000	0.000
July	0.000	0.000
August	0.000	0.000
September	0.000	0.000
October	3.293	1.361
November	6.585	2.723
December	9.878	4.084

Table 6.5 The Monthly Mean and Standard Deviation of SWE (Rio Grande River Basin above the Del Norte, 1949-1977)

MONTH	MEAN	STANDARD DEVIATION
January	7.418	3.287
February	9.461	3.531
March	0.287	4.866
April	5.903	4.052
May	1.051	1.712
June	0.000	0.000
July	0.000	0.000
August	0.000	0.000
September	0.000	0.000
October	1.802	0.799
November	3.605	1.598
December	5.407	2.397

Table 6.6 The Monthly Mean and Standard Deviation of SWE (Conejos River Basin above the Mogote, 1949-1977)

MONTH	MEAN	STANDARD DEVIATION
January	8.052	3.795
February	10.195	4.716
March	11.145	5.970
April	6.060	4.902
May	-----	-----
June	0.000	0.000
July	0.000	0.000
August	0.000	0.000
September	0.000	0.000
October	1.970	0.951
November	3.941	1.902
December	5.911	2.853

Table 6.7 The Monthly Mean and Standard Deviation of Precipitation (Rio Grande River Basin above the Del Norte, 1949-1983)

MONTH	MEAN	STANDARD DEVIATION
January	0.608	0.517
February	0.460	0.350
March	0.877	0.586
April	0.914	0.663
May	0.986	0.667
June	0.711	0.562
July	1.909	0.964
August	1.847	0.810
September	1.118	0.742
October	1.236	1.140
November	0.797	0.638
December	0.860	0.836

Table 6.8 The Monthly Mean and Standard Deviation of Temperature (Rio Grande River Basin above Del Norte, 1949-1983)

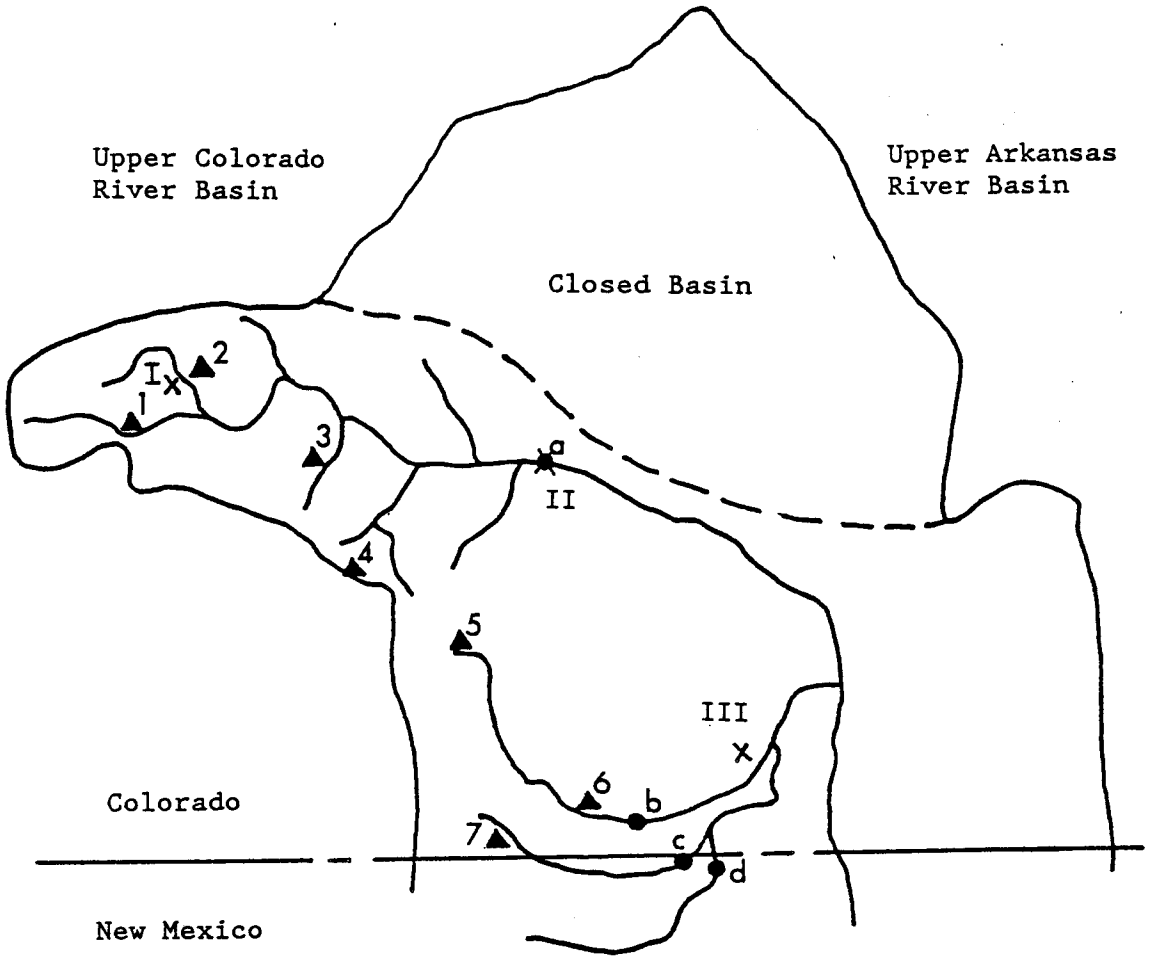
MONTH	MEAN	STANDARD DEVIATION
January	6.671	4.267
February	20.929	4.668
March	27.229	3.618
April	36.957	3.368
May	46.264	1.938
June	54.436	1.943
July	59.707	1.407
August	57.950	1.640
September	51.771	1.726
October	42.579	2.342
November	29.307	4.119
December	18.836	4.391

Table 6.9 The Monthly Mean and Standard Deviation of Precipitation (Conejos River Basin, Manassa City, 1966-1983)

MONTH	MEAN	STANDARD DEVIATION
January	0.249	0.264
February	0.230	0.198
March	0.328	0.252
April	0.287	0.183
May	0.751	0.504
June	0.604	0.687
July	1.379	0.897
August	1.424	0.696
September	0.860	0.729
October	0.610	0.421
November	0.306	0.256
December	0.348	0.262

Table 6.10 The Monthly Mean and Standard Deviation of Temperature (Conejos River Basin, Manassa City, 1966-1983)

MONTH	MEAN	STANDARD DEVIATION
January	18.417	5.131
February	24.611	4.831
March	33.944	3.105
April	41.250	3.223
May	50.056	1.939
June	58.972	2.226
July	64.139	1.281
August	61.861	1.830
September	55.528	1.490
October	44.333	2.646
November	31.417	3.679
December	20.750	4.638



STREAM GAGING STATIONS:

a. Rio Grande River near Del Norte
b. Conejos River near Mogote
c. Los Pinos River near Ortiz
d. San Antonio River near Ortiz

SNOW WATER EQUIVALENT DATA SITES:

1. Upper Rio Grande
2. Santa Maria
3. Lake Humphrey
4. Wolf Creek Pass
5. Platoro
6. River Springs
7. Cumbres Pass and Cumbres Trestle

TEMPERATURE AND PRECIPITATION DATA SITES

I. Hermit 7 ESE
II. Del Norte
III. Manassa

Fig. 6.1 The gaging stations of flow and climatological data

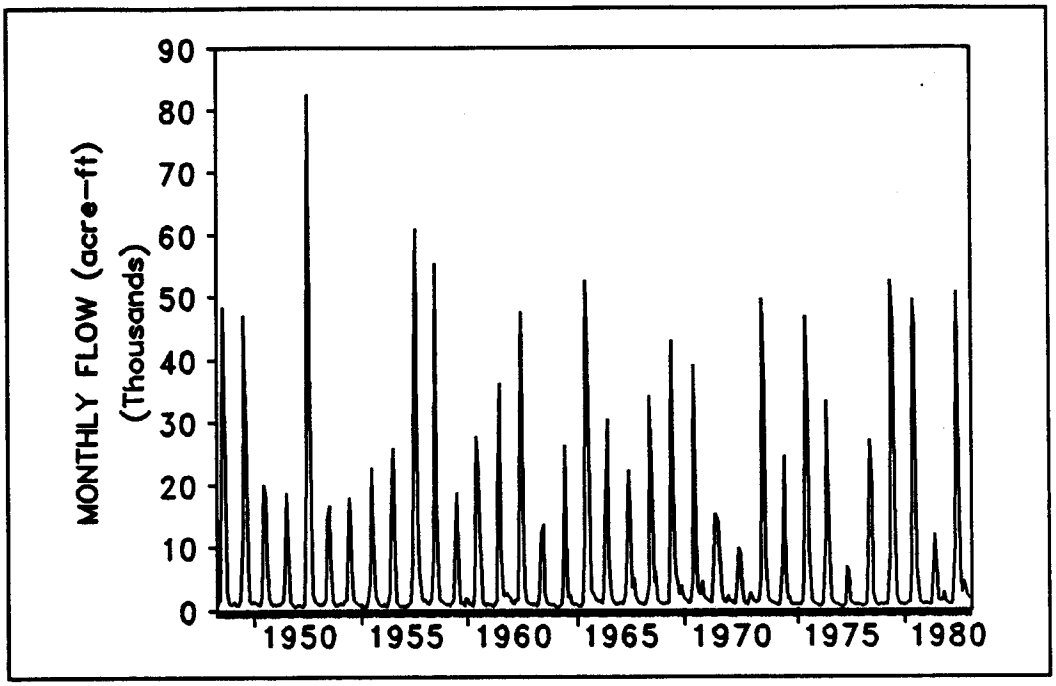


Fig. 6.2 Monthly streamflow of the Los Pinos River near Ortiz (1948-1982)

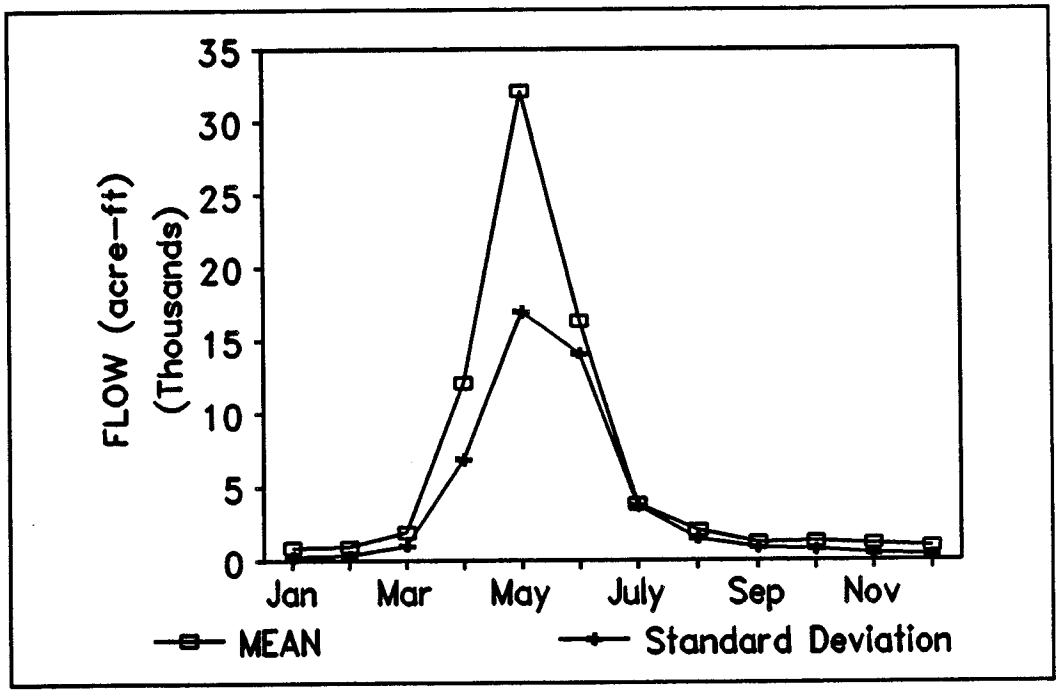


Fig. 6.3 The statistical properties of monthly streamflow (Los Pinos River near Ortiz, 1948-1977)

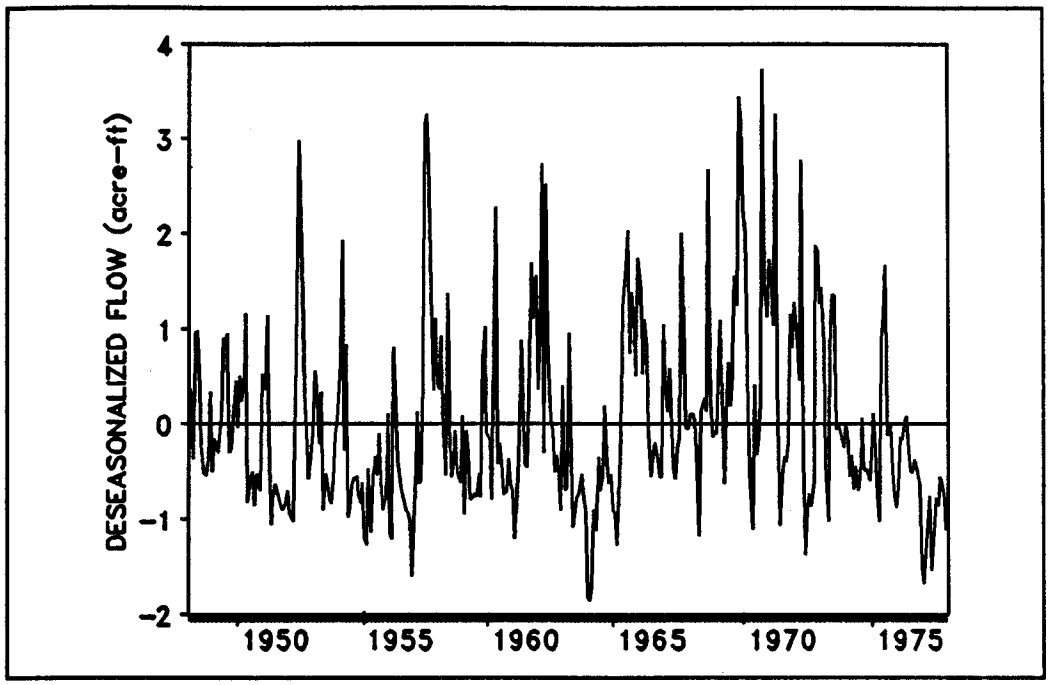


Fig. 6.4 The deseasonalized flow of the Los Pinos River near Ortiz (1948-1977)

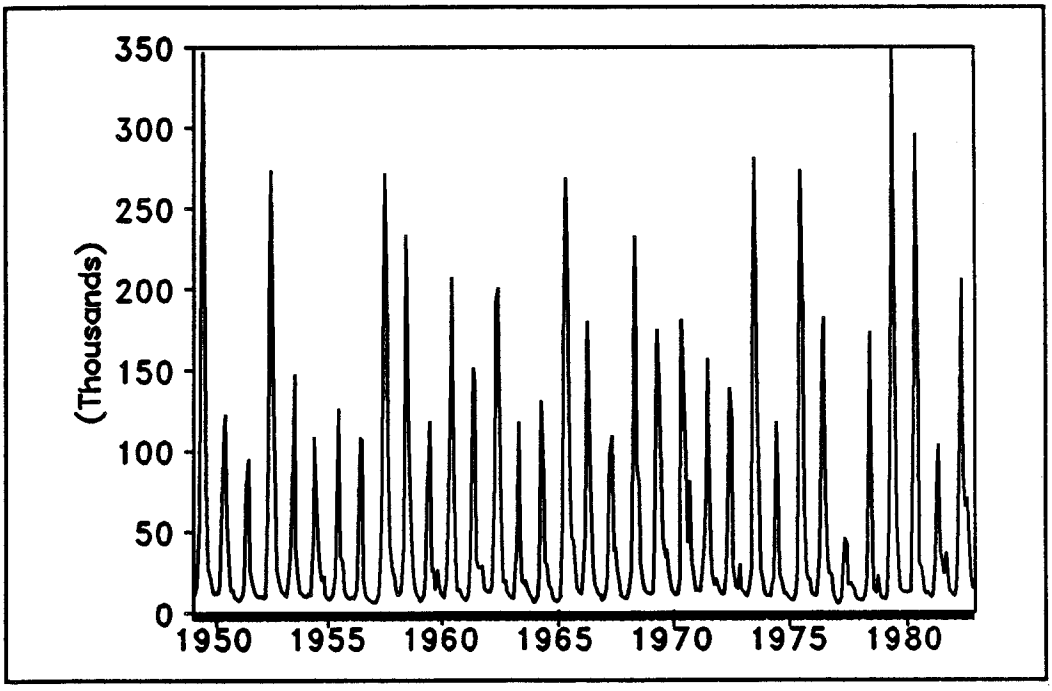


Fig. 6.5 Monthly streamflow of the Rio Grande River near Del Norte (1949-1982)

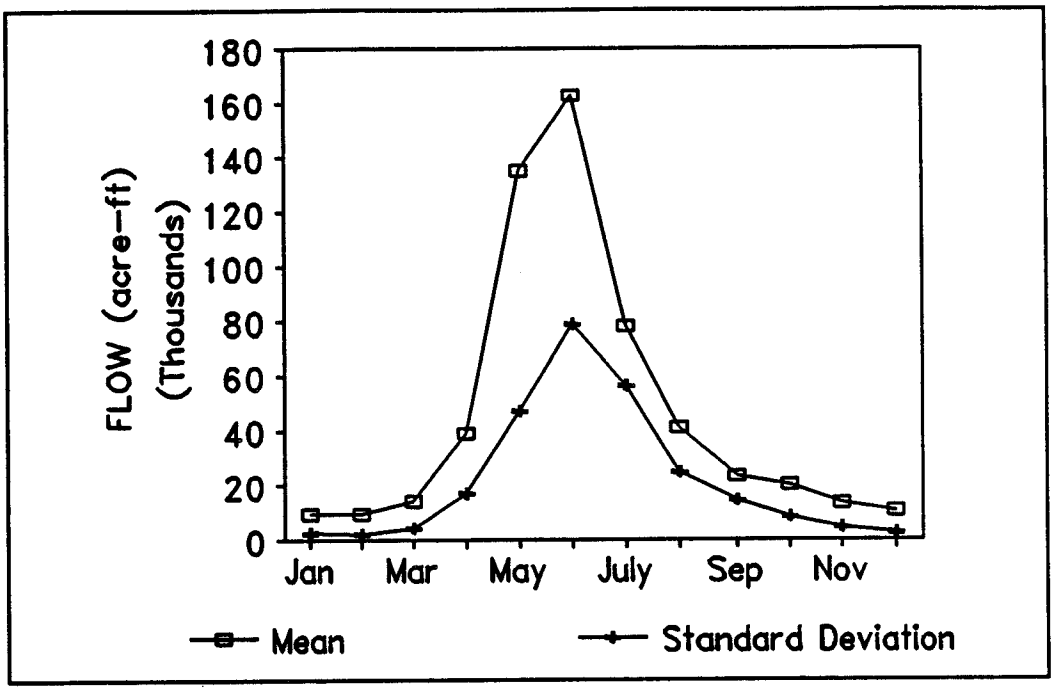


Fig. 6.6 The statistical properties of monthly streamflow (Rio Grande River near Del Norte, 1949-1977)

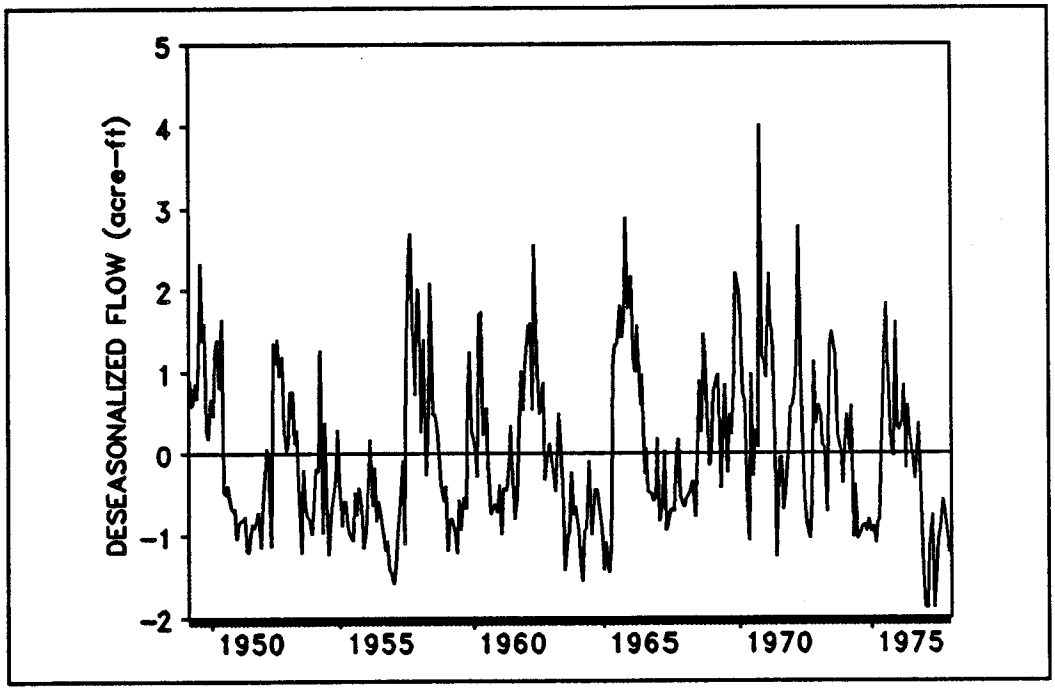


Fig. 6.7 The deseasonalized streamflow (Rio Grande River near Del Norte, 1949-1977)

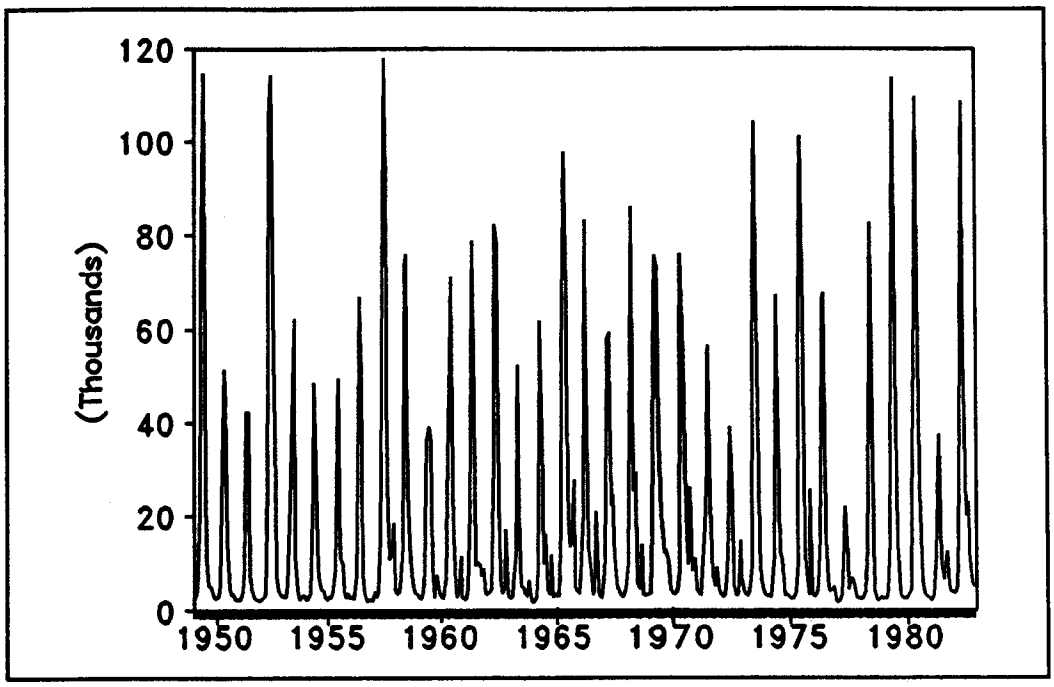


Fig. 6.8 Monthly streamflow of the Conejos River near Mogote (1949-1982)

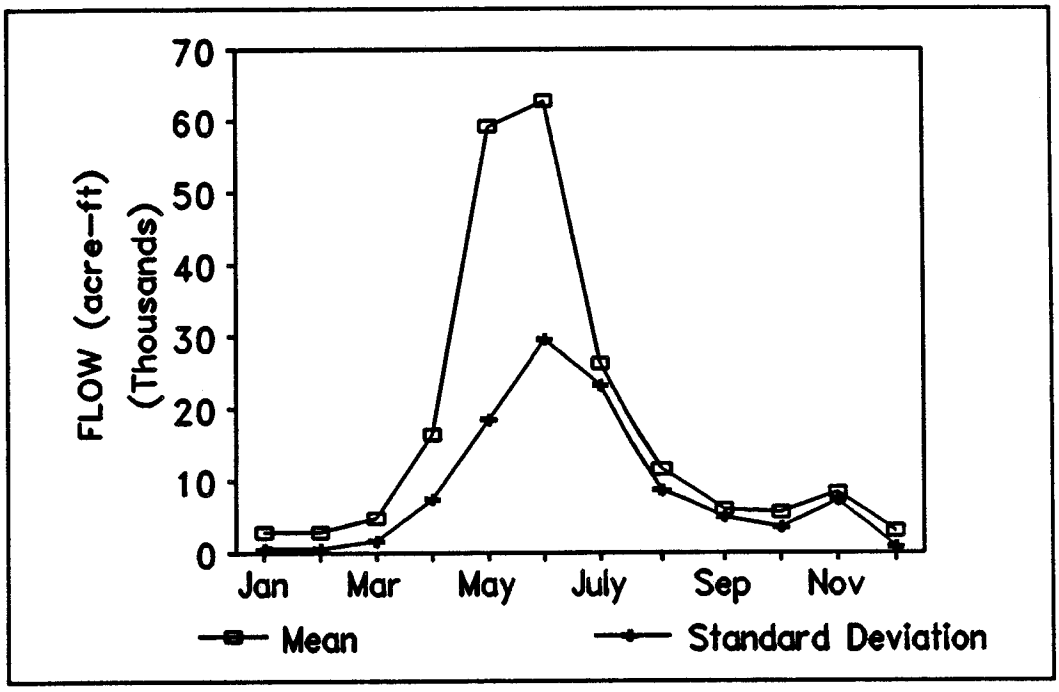


Fig. 6.9 The statistical properties of monthly streamflow (Conejos River near Mogote, 1949-1977)

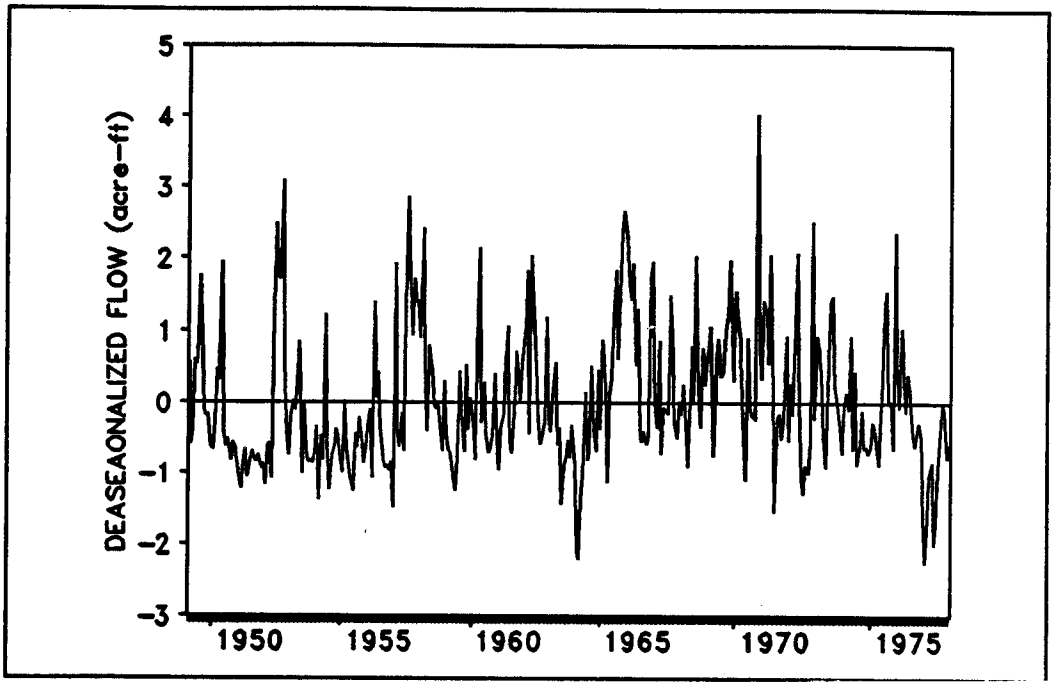


Fig. 6.10 The deseasonalized flow of the Conejos River near Mogote (1949-1977)

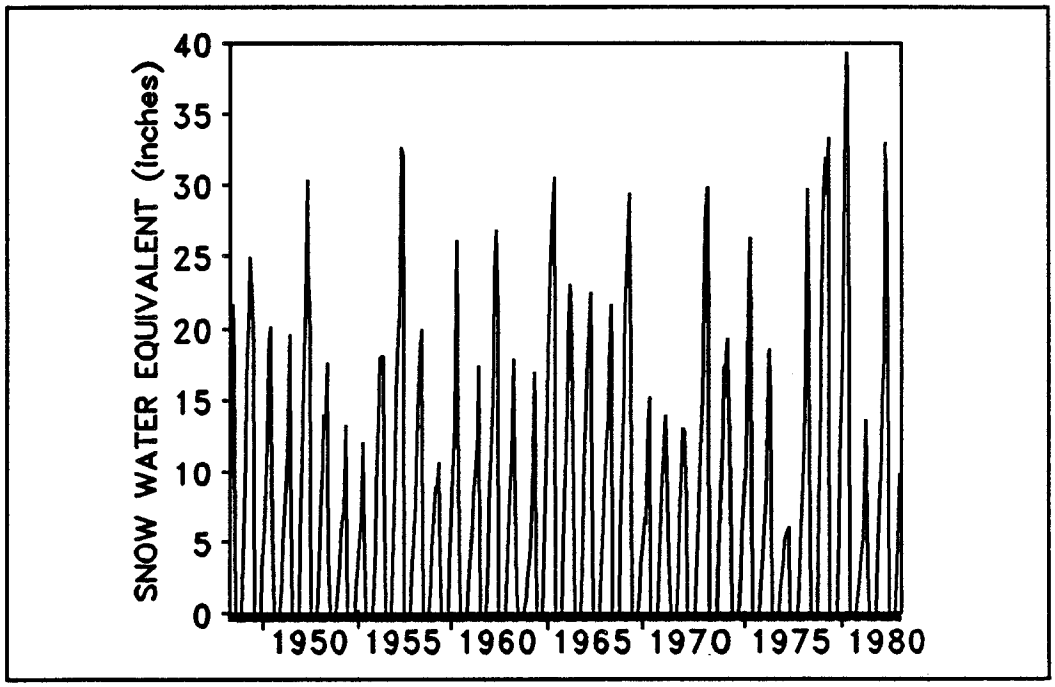


Fig. 6.11 The snow water equivalent at Cumbres Pass (1948-1982)

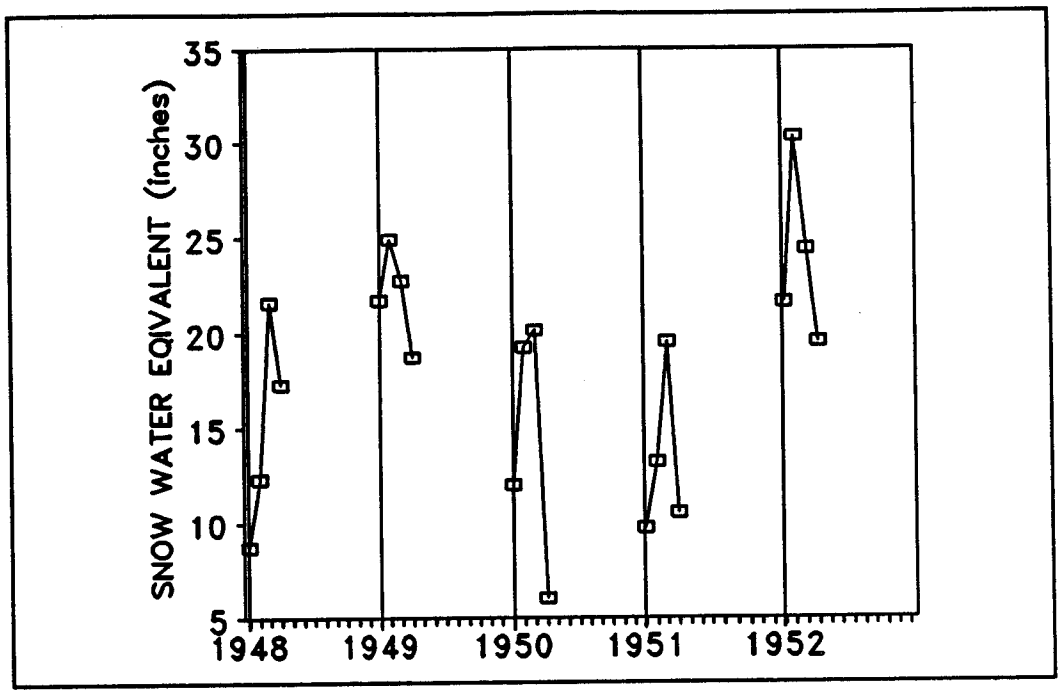


Fig. 6.12 The snow water equivalent at Cumbres Pass (1948-1952)

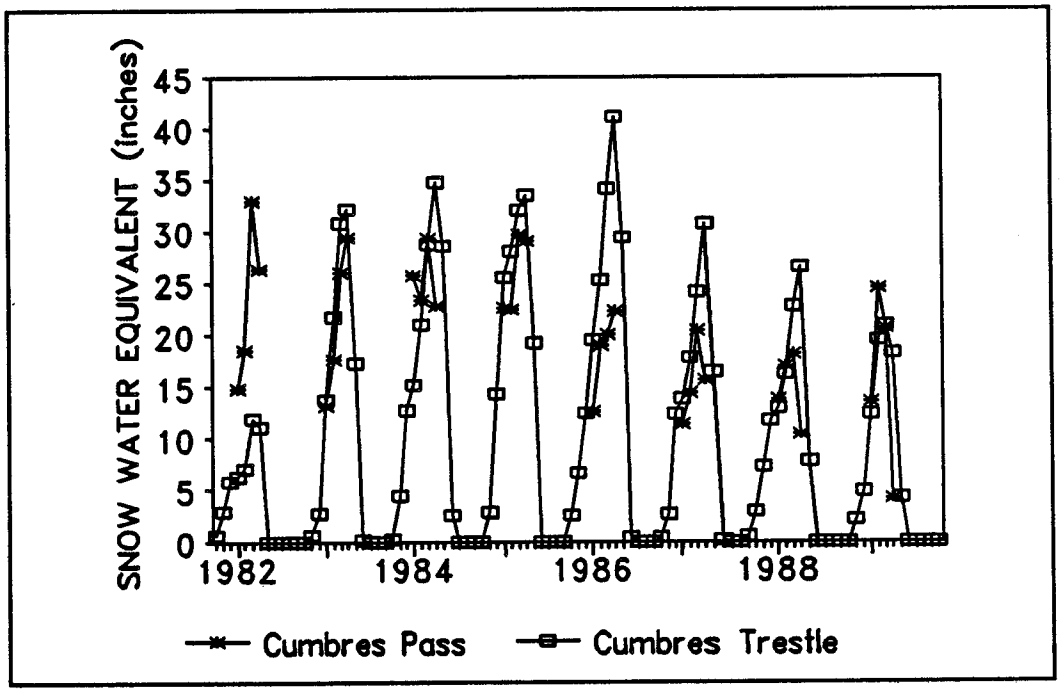


Fig. 6.13 The snow water equivalent at Cumbres Trestle (1981-1989)

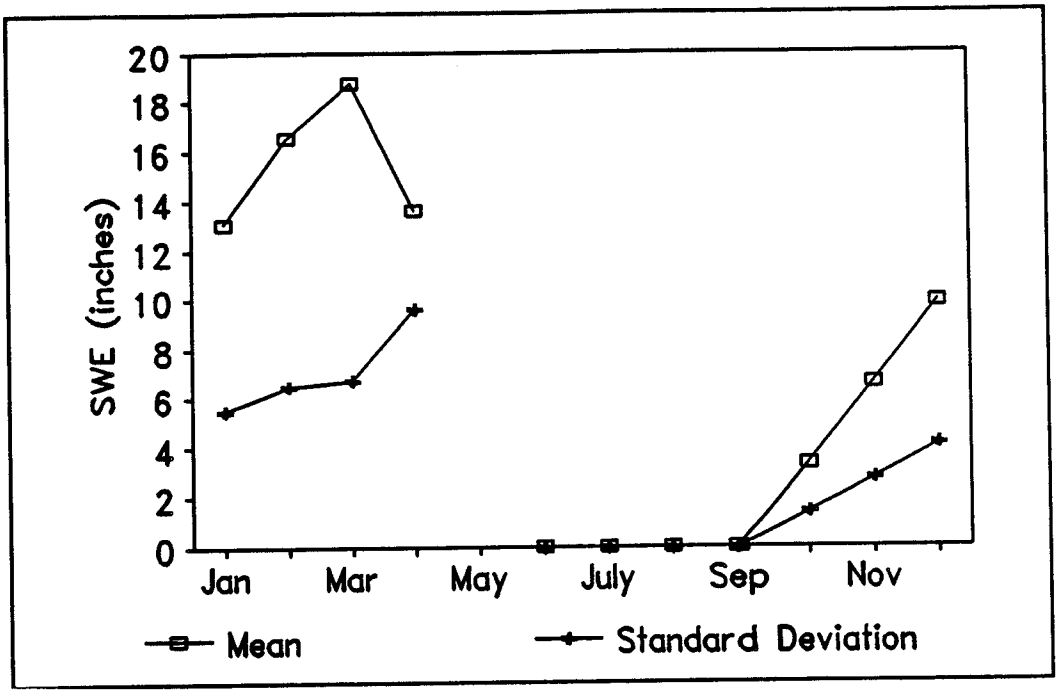


Fig. 6.14 The statistical properties of SWE at Cumbres Pass (1948-1977)

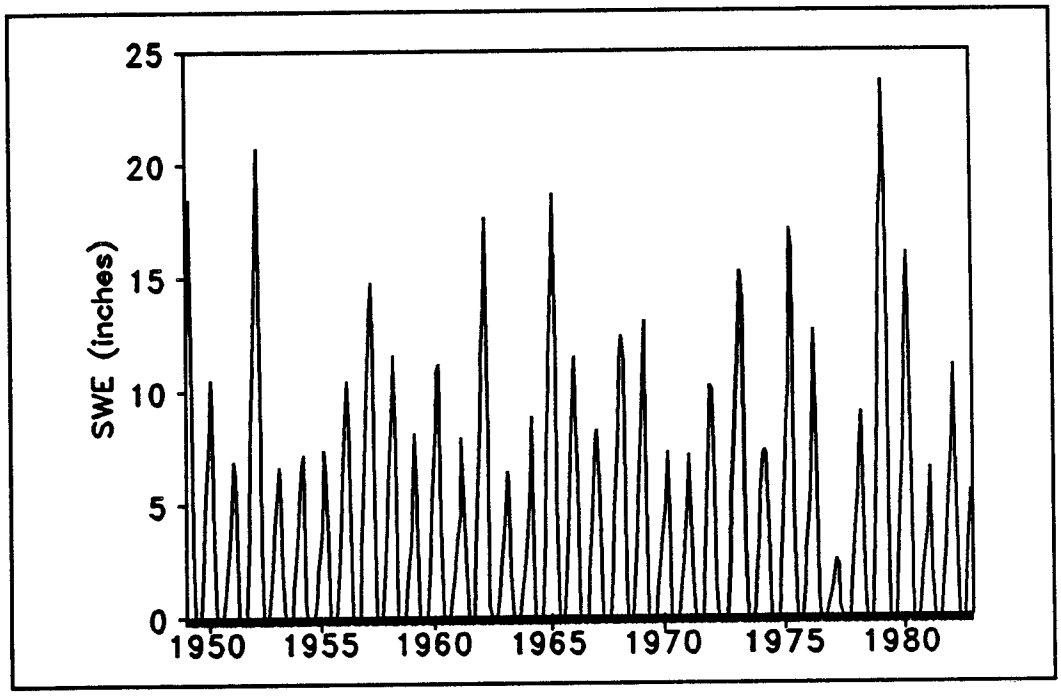


Fig. 6.15 The average snow water equivalent (Rio Grande River Basin above Del Norte, 1949-1982)

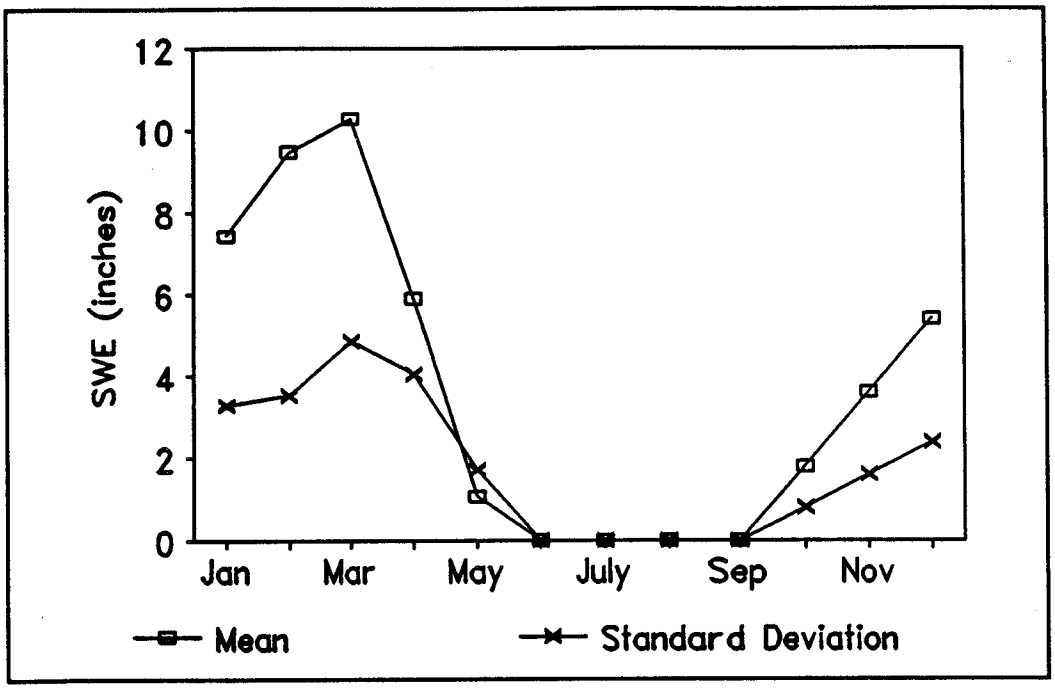


Fig. 6.16 The statistical properties of the average SWE (Rio Grande River Basin above Del Norte, 1949-1977)

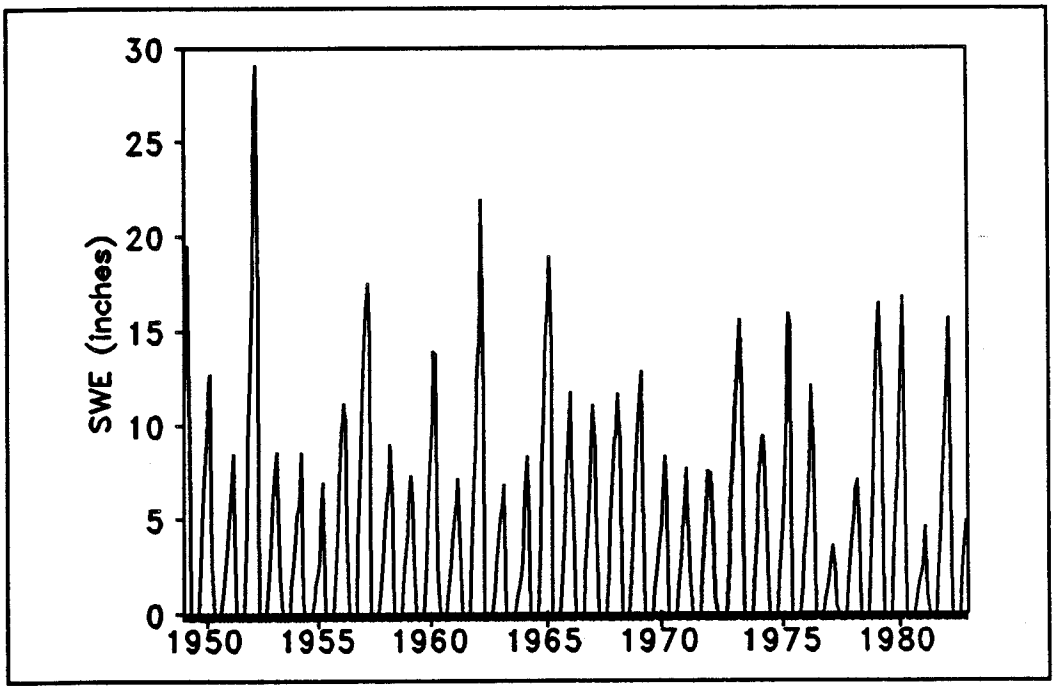


Fig. 6.17 The average snow water equivalent 1949-1982 (Conejos River near Mogote)

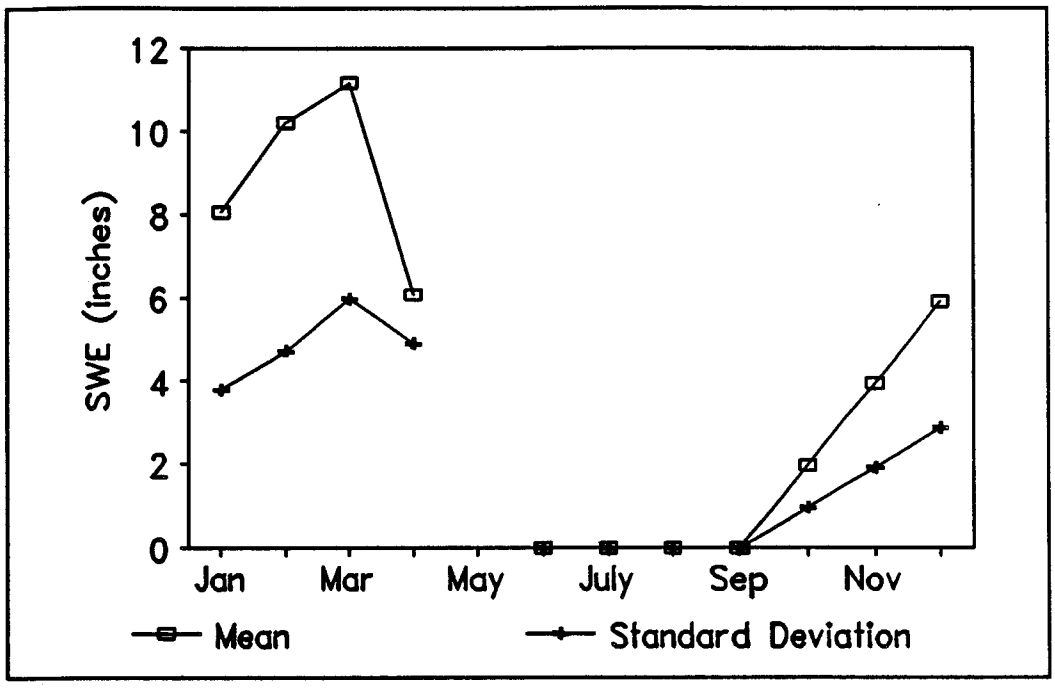


Fig. 6.18 The statistical properties of the average SWE (Conejos River near Mogote, 1949-1977)

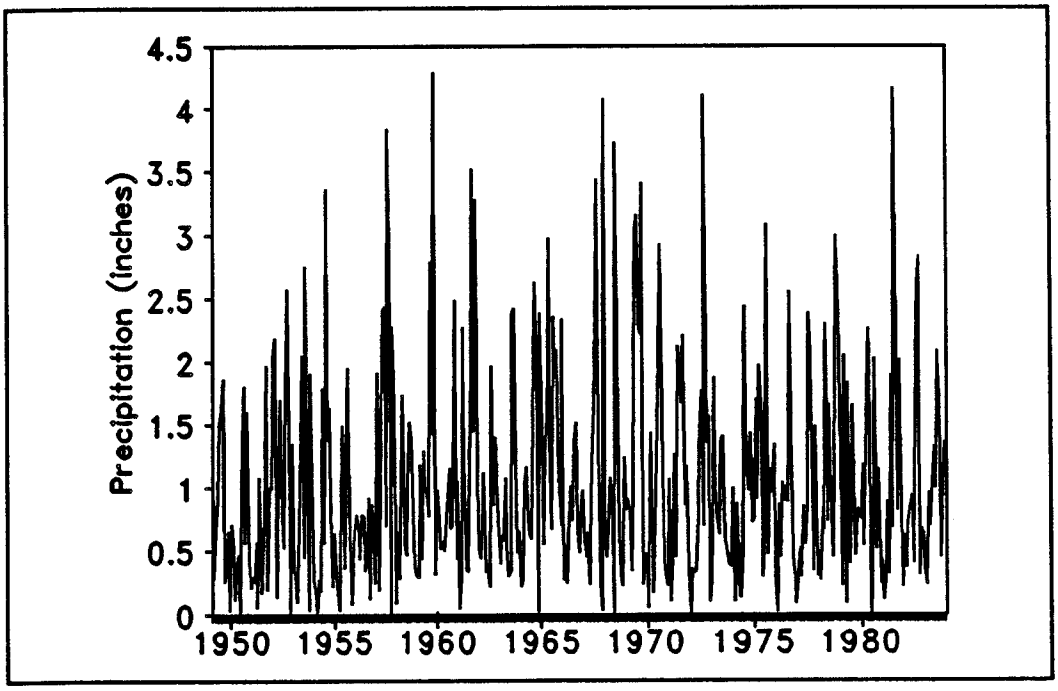


Fig. 6.19 The monthly precipitation of the Rio Grande River Basin (1949-1983)

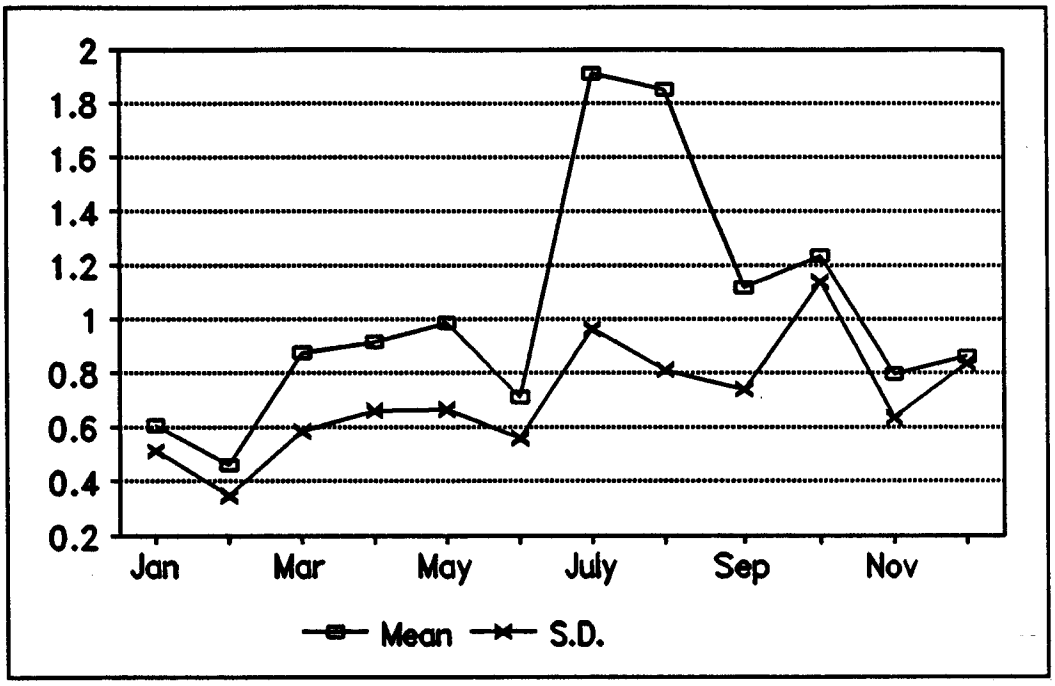


Fig. 6.20 The statistical properties of precipitation (Rio Grande River Basin, 1949-1983)

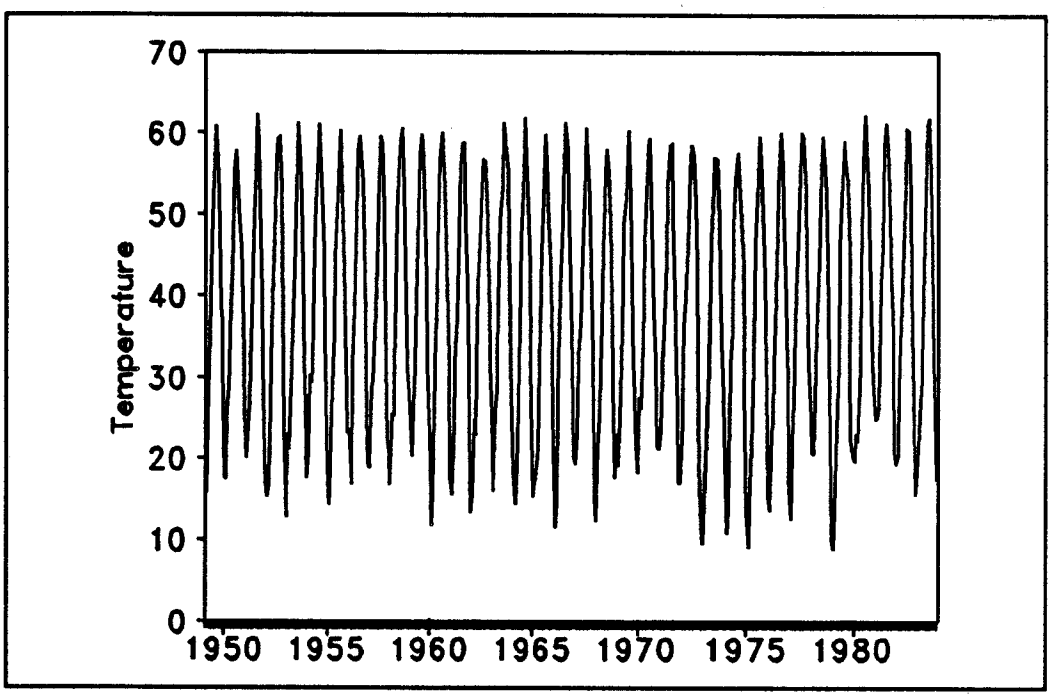


Fig. 6.21 The monthly temperature of the Rio Grande River Basin (1949-1983)

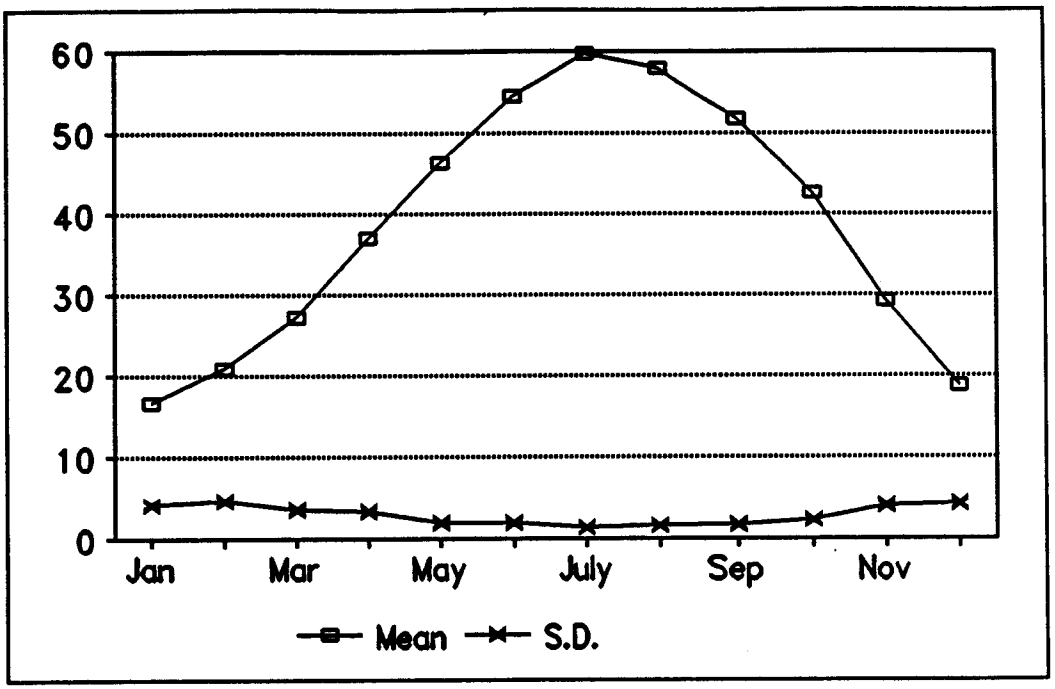


Fig. 6.22 The statistical properties of temperature (Rio Grande River Basin, 1949-1983)

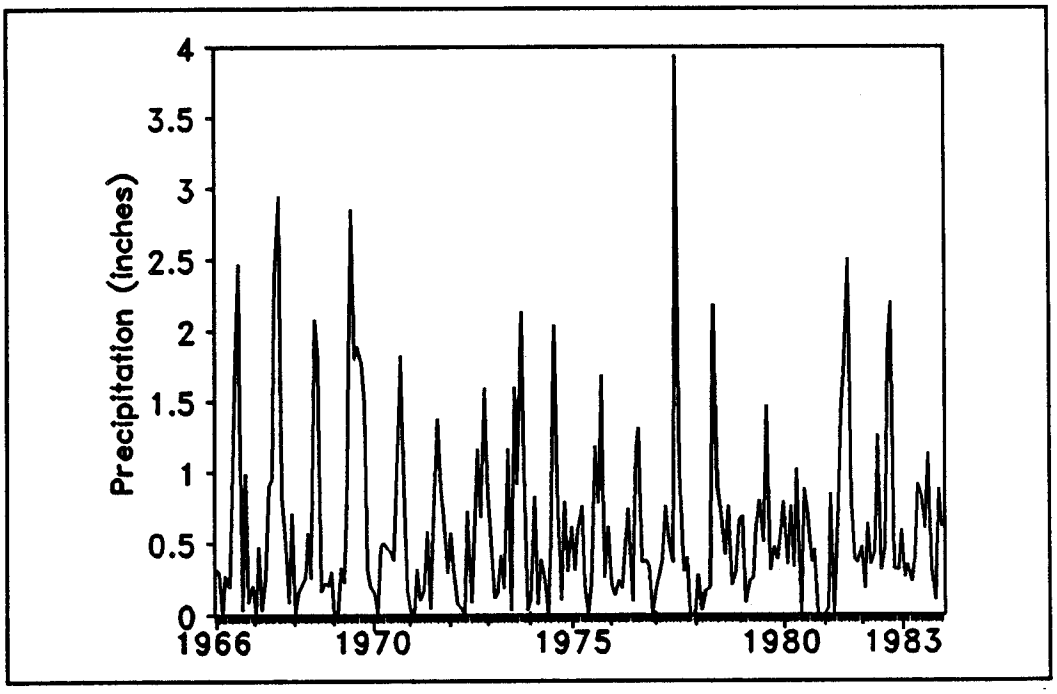


Fig. 6.23 The monthly precipitation of the Conejos River Basin (1966-1983)

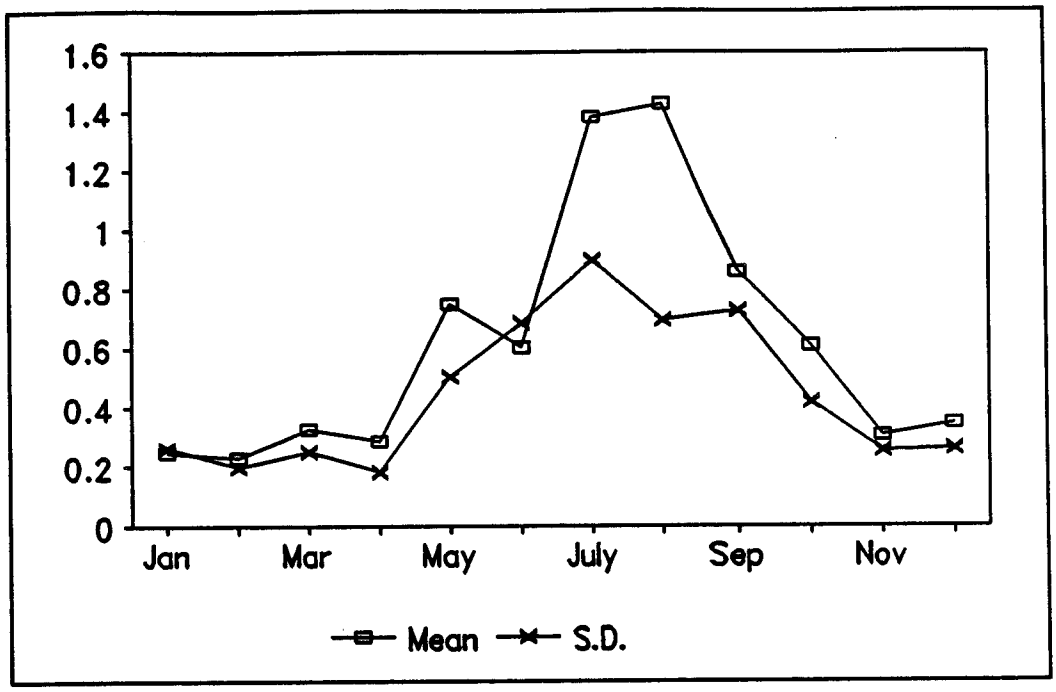


Fig. 6.24 The statistical properties of the precipitation (Conejos River Basin, 1966-1983)

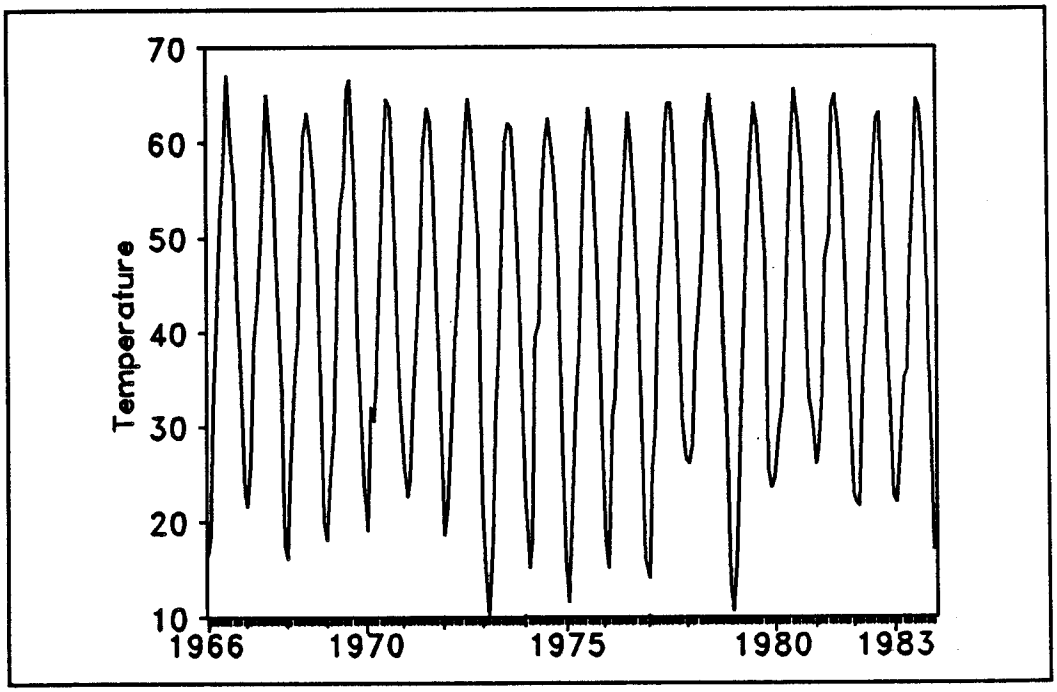


Fig. 6.25 The monthly temperature of the Conejos River Basin (1966-1983

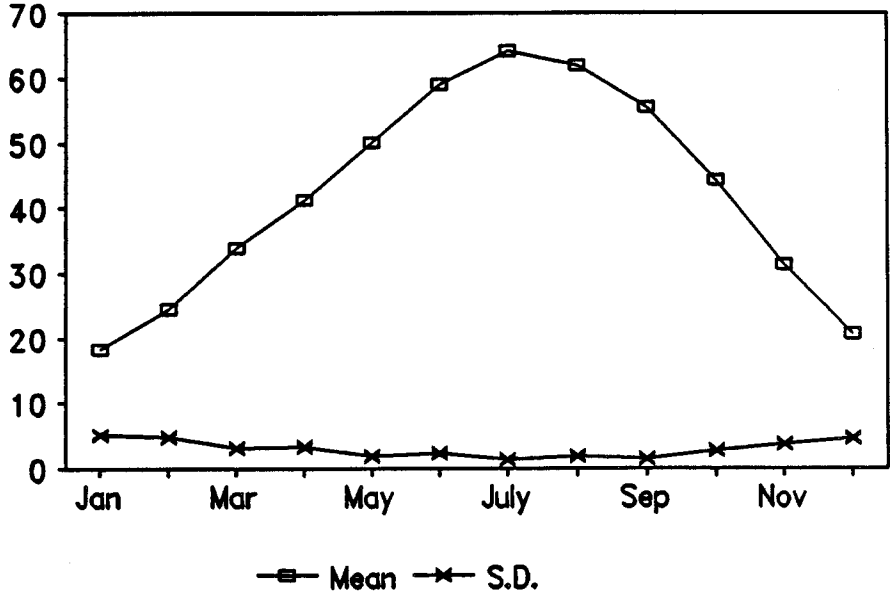


Fig. 6.26 The statistical properties of the temperature (Conejos River Basin, 1966-1983)

CHAPTER VII

FORECASTING MONTHLY STREAMFLOW BASED ON THE SINGLE INPUT-SINGLE OUTPUT MODEL

7.1 Univariate Time Series Model

Many types of stochastic models have been developed for modelling seasonal hydrological time series. Seasonal autoregressive integrated moving-average (SARIMA) models, deseasonalized autoregressive moving average (ARMA) models, and periodic autoregressive (PAR) models are most commonly used for hydrological series modelling. These three models have been employed to forecast monthly river flows by Noakes et al., 1985. In that paper, they concluded that although the SARIMA model forecasts seasonal economic data well, this model may not be appropriate for modeling and forecasting monthly river flows. The PAR model appears to be a more suitable model for monthly river flow modeling.

In this study, the deseasonalized ARMA model, the periodic AR(1) and periodic ARMA(1,1) will be used for modelling monthly streamflow. Only the monthly streamflow of the Los Pinos River near Ortiz will be analyzed by using these three models. The forecast results show that the univariate time series models underestimate the high flow in the summer time.

7.1.1 Deseasonalized ARMA Model

The monthly flows $Q_{v,\tau}$ of the Los Pinos River (1948-1977) was transformed by using the log transformation (base-10), i.e.

$$Q'_{\nu,\tau} = \log(Q_{\nu,\tau}) \tag{7.1.1}$$

Then, the log transformed flows were deseasonalized by

$$z_{\nu,\tau} = \frac{Q'_{\nu,\tau} - \hat{\mu}_\tau(Q')}{\hat{\sigma}_\tau(Q')} \tag{7.1.2}$$

The time series $z_{\nu,\tau}$ is now denoted as z_t, t=1,...,360 (t=(v-1)12+τ). The $z_{\nu,\tau}$ was checked to be normally distributed with mean zero and variance equal to 0.97. The autocorrelation function and the partial autocorrelation function of the deseasonalized flow were computed and are shown in Figs. 7.1 and 7.2, respectively. Then an ARMA model was fitted to the deseasonalized log-transformed flow. The AIC Information Criterion of Eq. (3.2.6) was used for selecting the correct model among competing ARMA models (see Salas et al., 1980).

The analysis of the log-transformed deseasonalized monthly streamflow of the Los Pinos River, gave the AR(4) model as the model with the smallest AIC value. Hence, the AR(4) model will be used for streamflow forecasting. The AR(4) model of the deseasonalized Los Pinos monthly flows is

$$z_t = 0.636\, z_{t-1} + 0.130\, z_{t-3} - 0.100\, z_{t-4} + \varepsilon_t \tag{7.1.3}$$

The residual ε_t has mean of zero and a variance of 0.54. It is shown in Fig. 7.3, and the autocorrelation function of ε_t is shown in Fig. 7.4. In addition, since

$$Q = N \sum_{k=1}^{20} \hat{\rho}_k^2(\varepsilon_t) = 7.326 < \chi^2_{0.95}(16) = 26.30$$

the normality hypothesis of the residuals was not rejected by using the chi-square test.

The forecast function is the conditional expectation of Eq. (7.1.3), i.e.

$$\hat{z}_t(L) = 0.636E[z_{t+L-1}] + 0.13E[z_{t+L-3}] - 0.1E[z_{t+L-4}] \tag{7.1.4}$$

The AR(4) model of Eq. (7.1.3) can be written as

$$z_t = (1 + 0.636B + 0.404B^2 + 0.387B^3 + \ldots)\varepsilon_t$$

The forecast of the deseasonalized flows z_t with lead time L=1,...,9 and 95% confidence limits are shown in Fig. 7.5. The one step ahead forecasts of "log-flows" Q_t' (1978-1982) are shown in Fig. 7.6. From Fig. 7.5-7.6, we can observe that the forecast of log-flows is fairly good.

After obtaining the log-flows Q_t', the actual flows Q_t are determined by Eq. (7.1.1), i.e.

$$\hat{Q}_t = 10^{\hat{Q}_t'} \tag{7.1.5}$$

From the discussion of Section 3.6, we know that Eq. (7.1.5) always underestimates the forecast, hence the minimum mean squared error forecasts of the actual flows given by Eq. (3.6.6) will be used. It is known that the forecast error of the deseasonalized streamflow z_t can be written as

$$e_t(L) = \sum_{j=0}^{L-1} \psi_j \varepsilon_{t+L-j} \tag{7.1.6}$$

Hence, the forecast error of the log flows $Q_{v,\tau}'$ can be written as

$$e^*_{v,\tau}(L) = \left(\sum_{j=0}^{L-1} \psi_j \varepsilon_{v,\tau+L-j}\right) \sigma_{\tau+L}(Q') + \mu_{\tau+L}(Q') \tag{7.1.7}$$

and the variance of the forecast error is

$$\sigma^{*2}_{\nu,\tau+L}(e_t)=\sigma^2_{\tau+L}(Q') \ \sigma^2_\varepsilon \ \sum_{j=0}^{L-1}\psi_j^2 \tag{7.1.8}$$

Finally, the minimum mean squared error forecasts of the actual flows can be obtained by

$$\hat{Q}_{\nu,\tau}(L) \ = \ 10^{\hat{Q}'_{\nu,\tau}(L)+\frac{1}{2}\hat{\sigma}^{*2}_{\nu,\tau+L}(e_t)} \tag{7.1.9}$$

Based on Eq. (7.1.9) the one step ahead forecasts of the actual flows are shown in Fig. 7.7 and Table 7.1. In addition, the forecasts using Eq. (7.1.5) are also shown in Fig. 7.7. The figure says that the forecast of the actual flow is still underestimated in the summer time.

7.1.2 The PAR(1) Model

The AR model with periodic coefficients was applied. After taking the logarithm of the monthly flows $Q_{\nu,\tau}$, it was deseasonalized as in Eq. (7.1.2). The monthly lag-1 autocorrelation of $z_{\nu,\tau}$ was calculated by Eq. (3.3.4). It is shown in Fig. 7.8. The PAR(1) model can be written as

$$z_{\nu,\tau} = \hat{\Phi}_{1,\tau} \ z_{\nu,\tau-1} \ + \ \varepsilon_{\nu,\tau} \tag{7.1.10}$$

in which $\hat{\Phi}_{1,\tau} = \hat{\rho}_{1,\tau}$. The estimated residual ε_t is shown in Fig. 7.9, the autocorrelation function of the estimated residuals is shown in Fig. 7.10, and the seasonal correlation of the estimated residuals is shown in Fig. 7.11. The residual $\varepsilon_{\nu,\tau}$ has $\sigma_\tau^2(\varepsilon)$ as shown below:

τ	1	2	3	4	5	6
$\sigma_\tau^2(\varepsilon)$	0.331	0.276	0.889	0.882	0.993	0.230
τ	7	8	9	10	11	12
$\sigma_\tau^2(\varepsilon)$	0.170	0.533	0.426	0.593	0.116	0.528

Overall the ε_t's are verified to be normal with mean zero and variance 0.48.

The forecast function based on the PAR(1) model has the form of

$$\hat{z}_{\nu,\tau}(L) = \hat{\rho}_{1,\tau+L}\, E[z_{\nu,\tau+L-1}] \tag{7.1.11}$$

Equation (7.1.10) can also be rewritten as

$$z_{\nu,\tau} = \frac{1}{1-\hat{\rho}_{1,\tau}B}\varepsilon_{\nu,\tau}$$

$$= \psi_\tau(B)\varepsilon_{\nu,\tau} \tag{7.1.12}$$

with

τ	1	2	3	4	5	6
$\hat{\rho}_{1,\tau}$	0.815	0.851	0.334	0.344	0.086	0.878
τ	7	8	9	10	11	12
$\hat{\rho}_{1,\tau}$	0.911	0.683	0.758	0.638	0.940	0.687

Then the forecast error of the deseasonalized flow $z_{\nu,\tau}$ can be written as

$$e_{\nu,\tau}(L) = \sum_{j=0}^{L-1} \psi_{j,\tau+L}\varepsilon_{\nu,\tau+L-j} \tag{7.1.13}$$

The hypothesis of normality of the whole 360 historical residuals ε_t is rejected by the chi-square test. Several distributions including lognormal, gamma, Weibull,...etc. were used for fitting the residuals, but no distribution was accepted.

Since we are only interested in the 95% quantiles of the forecast error, the cumulative relative frequency of the forecast error will be simulated. For example, by standing in March, 1978, the forecast error with lead time L=1,2,... can be written as

$$\left.\begin{aligned} e_{1978,3}(1) &= \sum_{j=0}^{0} \psi_{j,4}\varepsilon_{1978,4-j} \\ e_{1978,3}(2) &= \sum_{j=0}^{1} \psi_{j,5}\varepsilon_{1978,5-j} \\ &\vdots \end{aligned}\right\} \qquad (7.1.14)$$

The 95% probability limits of the forecast error with lead L=1, can be obtained using the empirical cumulative relative frequency. For the lead time L greater than 1, we can put the historical residuals ε_τ in a black box and sample from it randomly with replacement. Then, the forecast error can be simulated 1000 times by using Eq. (7.1.14). The simulated cumulative relative frequency of the forecast errors are shown in Fig. 7.12-7.18, and the 2.5% and 97.5% quantiles are listed below:

	Lead Time L						
	1	2	3	4	5	6	7
$q_{97.5}$	1.79	1.68	2.35	2.22	1.87	2.27	2.27
$q_{2.5}$	-1.40	-1.97	-2.24	-2.19	-1.50	-1.47	-1.37

Then, the forecast of the deseasonalized streamflow $z_{v,\tau}$ with lead time L and the 95% probability limit bound can be constructed. The forecast of the deseasonalized flow (1978-1982) with lead time L=1,...,9 and the 95% confidence limits are shown in Fig. 7.19. This shows that most of the historical data fall inside the bounds.

The forecast error of the streamflow $Q_{v,\tau}'$ can be written as

$$e^{*}_{\nu,\tau}(L) = \left(\sum_{j=0}^{L-1} \psi_{j,\tau+L}\varepsilon_{\nu,\tau+L-j}\right) \sigma_{\tau+L}(Q') + \mu_{\tau+L}(Q') \tag{7.1.15}$$

Hence, the variance of the forecast error of $Q_{\nu,\tau}'$ can be written as

$$\sigma^{*2}_{\nu,\tau+L}(e_t) = \sigma^{2}_{\tau+L}(Q') \sum_{j=0}^{L-1} \psi^{2}_{j,\tau+L}\ \sigma^{2}_{\tau+L-j}(\varepsilon_t) \tag{7.1.16}$$

and the minimum mean squared error forecast of the actual flow can be obtained by

$$\hat{Q}_{\nu,\tau}(L) = 10^{Q'_{\nu,\tau}(L)+\frac{1}{2}\sigma^{*2}_{\nu,\tau+L}(e_t)} \tag{7.1.17}$$

The one step ahead forecast of the monthly flows using the PAR(1) model is shown in Fig. 7.20 and Table 7.1. In addition, the forecast results using the deseasonalized AR model are also shown in Fig. 7.20. The results show that the forecast using the PAR(1) model has improved a little.

7.1.3 The PARMA(1,1) Model

The deseasonalized log-transformed flow series $z_{\nu,\tau}$ can be fitted to a periodic ARMA(1,1) model, i.e.

$$z_{\nu,\tau} = \varphi_{1,\tau} z_{\nu,\tau-1} + \varepsilon_{\nu,\tau} - \theta_{1,\tau}\varepsilon_{\nu,\tau-1} \tag{7.1.18}$$

The moment estimation of the parameters $\phi_{1,\tau}$ and $\theta_{1,\tau}$ can be determined by solving Eqs. (3.3.6) and (3.3.7) in which the population coefficients $\rho_{1,\tau}$ and $\rho_{2,\tau}$ are substituted by the corresponding sample estimates. The estimated parameters are:

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	1.1	0.8	0.5	0.6	-1.1	0.3
$\hat{\theta}_{1,\tau}$	0.6	-0.3	0.7	0.3	-1.5	-0.7
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.8	0.7	0.9	0.5	0.9	0.8
$\hat{\theta}_{1,\tau}$	-0.6	-0.1	0.3	-0.2	-0.1	0.6

By using the moment estimates as the initial values the final parameters $\varphi_{v,\tau}$ and $\theta_{v,\tau}$ are determined by the Marquardt algorithm of least squares method. The results are:

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	1.0	0.8	0.5	0.4	-0.7	0.6
$\hat{\theta}_{1,\tau}$	0.6	-0.3	0.7	0.3	-1.1	-0.4
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.8	0.6	1.0	0.5	0.9	0.8
$\hat{\theta}_{1,\tau}$	-0.6	-0.6	0.5	-0.3	0.02	0.6

Then, the residuals ε_t are estimated and plotted in Fig. 7.21. The autocorrelation function of the estimated residuals is shown in Fig. 7.22 and the seasonal correlation of the estimated residual is shown in Fig. 7.23. The ε_t is checked to be WN(0,0.47), and the $\varepsilon_{v,\tau}$ has a monthly variance $\sigma_\varepsilon^2(\tau)$ shown below:

τ	1	2	3	4	5	6
$\sigma_\tau^2(\varepsilon)$	0.291	0.257	0.860	0.948	1.036	0.222
τ	7	8	9	10	11	12
$\sigma_\tau^2(\varepsilon)$	0.176	0.530	0.365	0.575	0.117	0.492

Then, the forecast function can be written as

$$\hat{z}_{\nu,\tau}(L) = \hat{\varphi}_{1,\tau+L}\ E[z_{\nu,\tau+L-1}] \ - \ \hat{\theta}_{1,\tau+L} E[\varepsilon_{\nu,\tau+L-1}] \tag{7.1.19}$$

Eq. (7.1.18) can also be rewritten as

$$z_{\nu,\tau} = \frac{(1-\theta_{1,\tau}B)}{(1-\varphi_{1,\tau}B)}\varepsilon_{\nu,\tau} = \psi_\tau(B)\varepsilon_{\nu,\tau} \tag{7.1.20}$$

The normality hypothesis of the residual ε_t is rejected by the chi-squares test. Hence the simulated cumulative distribution frequency of the simulated forecast error is constructed using (7.1.14). Then, the forecast of the deseasonalized flow with lead time L=1,... and the 95% confidence limits are shown in Fig. 7.24.

The variance of the forecast error of log flow $Q_{\nu,\tau}$ can be obtained using Eq. (7.1.16). Then, the minimum mean squared error forecast of the actual flow can be obtained from Eq. (7.1.17).

The one step ahead forecast using the PARMA(1,1) model is shown in Fig. 7.25 and Table 7.1. In addition, the forecast results using the PAR(1) model is also shown in Fig. 7.25 and Table 7.1. This figure shows that the PARMA(1,1) is better in certain cases than the PAR(1) model. The PARMA(1,1) and PAR(1) model are both better than the AR(4) model for forecasting monthly flows. However, by observing Fig. 7.25, the forecast is still not satisfactor because most of the forecast values have a big difference from the actual flows. In later sections, an input-output model will be used in order to improve the forecast of monthly flows.

7.1.4 Comment

The univariate time series model can be used to generate streamflow. This may preserve the statistical properties of the flow series, but when used for forecasting

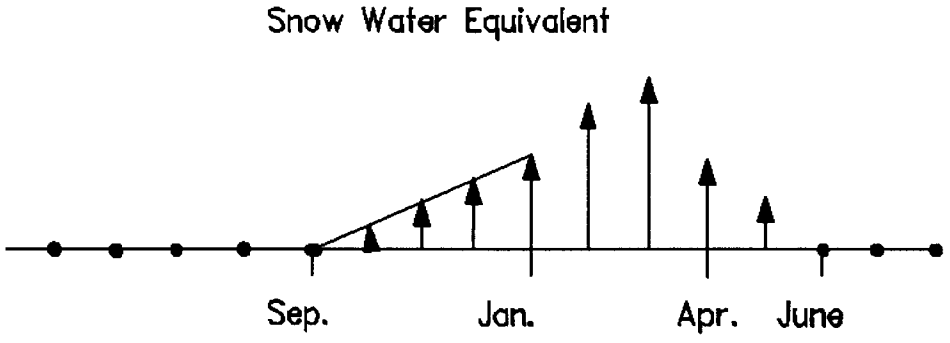
monthly flows, the high flows in the summer time is always underestimated. A good forecast of the "log-flow" does not mean a good forecast of the actual flow. The forecasted result should not be compared in the log domain.

7.2 Forecasting Snow Water Equivalent

In Section 6.3, we have discussed the properties of the snow water equivalent (SWE). Now, we need to find a model for forecasting SWE.

The SWE data first occurs in the month of October. Figures 7.26-7.29 show the plot of the SWE (1948-1977) of October vs. the SWE of April, March, February, and January, respectively. They show that the SWE of October does not dependent on the SWE of the previous months. Hence, the forecast of the SWE in October will be taken as the historical mean value for October.

Since the SWE of June, July, Aug. and Sep. is equal to zero, and the SWE of Oct., Nov., Dec. can be obtained by using linear interpolation between zero in September and the value of January. The SWE of Jan., Feb., Mar., April and May can be fitted by using ARMA model. The whole structure of SWE is shown below:



Since the streamflow in January, February, and March are small, the agency administering the Rio Grande Basin does not need to forecast the streamflow of these months. The forecast of monthly streamflow from April to December is what is needed. Before forecasting the streamflow using the input-output model, we need to forecast SWE. First, we can forecast the SWE using the SWE series of each month. For example, from the historical SWE record of 1948-1977, we have 30 SWE readings of January data. We then attempt to find some correlation between these 30 data readings. The SWE of January, February, March, April and their autocorrelations are shown in Figs. 7.30-7.37. These show that the SWE is independent of each month, i.e. there is no correlation between the SWE in one month to that same month in a previous year.

The next step is to fit the deseasonalized SWE of January, February, March and April using a time series model. The autoregressive model will be used.

Because the SWE of June, July, Aug., Sep. are equal to zero, we do not need to do any forecast for these months. The SWE of Oct., Nov. and Dec. can be estimated by using the historical mean value of each month.

7.3 Rational Single Input-Single Output Transfer Function Model

The univariate time series model does not give a satisfactory result for forecasting monthly flows as shown in Section 7.1. An input-output transfer function model will be tried now. Both the transformed and the no-transformed modeling will be compared.

7.3.1 Log-Transformed Monthly Flow Modeling

The monthly flows $Q_{v,\tau}$ of the Los Pinos River near Ortiz is transformed by using the equation

$$Q'_{v,\tau} = \log(Q_{v,\tau}) \tag{7.3.1}$$

Then, the monthly mean $\hat{\mu}_\tau(Q')$ and monthly standard deviation $\hat{\sigma}_\tau(Q')$ of log-flow $Q_{v,\tau}'$ is calculated. The log-flow $Q_{v,\tau}'$ is deseasonalized using

$$z_{v,\tau} = \frac{Q'_{v,\tau} - \hat{\mu}_\tau(Q')}{\hat{\sigma}_\tau(Q')} \tag{7.3.2}$$

The snow water equivalent $W_{v,\tau}$ is also transformed using

$$W'_{v,\tau} = \begin{cases} log(W_{v,\tau}) & for\ W_{v,\tau} \neq 0 \\ 0 & for\ W_{v,\tau} = 0 \end{cases} \tag{7.3.3}$$

Then the log-SWE is deseasonalized using

$$x_{v,\tau} = \begin{cases} \dfrac{W'_{v,\tau} - \hat{\mu}_\tau(W')}{\hat{\sigma}_\tau(W')} & for\ W'_{v,\tau} \neq 0 \\ 0 & for\ W'_{v,\tau} = 0 \end{cases} \tag{7.3.4}$$

The SWE of Jan. through April has an AR(1) structure, and the SWE of May is filled in using this structure, i.e.

$$x_t = 0.728\ x_{t-1} + \zeta_t \tag{7.3.5}$$

in which ζ_t is normal with mean zero and variance 0.209. Then, by denoting $z_{v,\tau}$ as z_t and $x_{v,\tau}$ as x_t, for t=1,2,..., and estimating the deseasonalized SWE of May using Eq. (7.3.5), a single input-output transfer function model can be written as

$$z_t = \sum_{j=0}^{\infty} a_j x_{t-j} + N_t \tag{7.3.6}$$

The impulse response function a_j is estimated by calculating the frequency

response function

$$A(f) = \frac{\hat{S}_{zx}(f)}{\hat{S}_x(f)} = \frac{\hat{c}_{zx}(f) + i\,\hat{q}_{zx}(f)}{\hat{c}_x(f) + i\,\hat{q}_x(f)} \tag{7.3.7}$$

Since the input x_t is a real number, the imaginary part of spectrum $S_x(f)$ is equal to zero. Then, the frequency response function can be written as

$$A(f) = \frac{\hat{c}_{zx}(f)}{\hat{c}_x(f)} + i\,\frac{\hat{q}_{zx}(f)}{\hat{c}_x(f)} \tag{7.3.8}$$

Equation (7.3.8) needs to estimate the auto-spectrum $S_x(f)$. First, the autocovariance γ_x of input series x_t is estimated by using Eq. (3.1.6). Then the Parzen window of Eq. (3.7.26) and the estimated autocovariance are substituted into Eq. (3.7.25) for estimating the auto-spectrum $S_x(f)$. Since we are not interested in the peak of the spectrum, the truncation point M in Eq. (3.7.25) is used following the suggestion of Chatfield (1980). He suggests using "2 times the square root of the total number of data N". The estimation of auto-spectrum $S_x(f)$ is shown in Fig. 7.38.

The cross covariance γ_{yx} of the input x_t and y_t is estimated by using Eq. (3.1.10). Then, the Parzen window and the estimated autocovariance are substitute into Eq. (3.7.27) to estimate the cross spectrum $S_{yx}(f)$. The co-spectrum c_{yx} and the quad-spectrum q_{yx} can then be obtained. These are shown in Fig. 7.39 and Fig. 7.40.

By substituting the estimated values of $c_x(f)$, $c_{yx}(f)$ and $q_{yx}(f)$ into Eq. (7.3.8), the frequency response function A(f) can be estimated. Then, the impulse response function a_j can be estimated by using the Fourier transform of Eq. (3.7.5). The variance of the impulse response function is computed using Eq. (4.3.8). It has a value of 0.194.

The impulse response function is shown in Fig. 7.41. This figure shows that the values of a_0, a_1, and a_2 are significantly different from zero. Hence, Eq. (7.3.6) can be written as

$$z_t = -0.228x_t + 0.181x_{t-1} + 0.264x_{t-2} + N_t \tag{7.3.9}$$

Thus N_t can be estimated by

$$\hat{N}_t = z_t - (-0.228x_t + 0.181x_{t-1} + 0.264x_{t-2}) \tag{7.3.10}$$

Preliminary identification of a suitable model for the noise $\hat{N}_t$ is carried out by fitting an ARMA model, i.e.

$$\hat{N}_t = 0.64\hat{N}_{t-1} + \varepsilon_t = \frac{\varepsilon_t}{1-0.64B} \tag{7.3.11}$$

in which ε_t is normal with mean zero and variance 0.509.

Substituting (7.3.11) into (7.3.9) gives

$$\begin{aligned} z_t = {} & 0.64z_{t-1} - 0.228x_t + 0.327x_{t-1} \\ & + 0.148x_{t-2} - 0.169x_{t-3} + \varepsilon_t \end{aligned} \tag{7.3.12}$$

Since the agency administering the Rio Grande river in southern Colorado, does not need to forecast the flows of January, February and March, the forecast of monthly streamflow is only for the months of April through December. In this research, minimizing the sum of square errors will be only for the monthly flow from April to December. This considers the estimated monthly flow equal the historical monthly flow $z_{v,\tau}$ for $\tau=1,2,3$. Then, the series z_t represented by $z_{v,\tau}$ ($t=(v-1)\cdot 12+\tau$) is

$$\hat{z}_{v,\tau} = \begin{cases} 0.64z_{v,\tau-1} - 0.228x_{v,\tau} + 0.327x_{v,\tau-1} + 0.148x_{v,\tau-2} \\ -0.169x_{v,\tau-2}, & \text{for } \tau = 4,5,\ldots,12 \\ z_{v,\tau} & \text{for } \tau = 1,2,3. \end{cases}$$

where v = 1,2,...,N and N = the number of years of historical monthly flow. After minimizing the sum of square errors $\Sigma\varepsilon_t^2$, the final model can be obtained. It can be written as

$$z_t = 0.627 z_{t-1} - 0.266 x_t + 0.316 x_{t-1} + 0.16 x_{t-2} - 0.01 x_{t-3} + \varepsilon_t^* \quad (7.3.13)$$

The ε_t^* is checked to be white noise with mean zero and variance 0.498, which has smaller variance then the variance of ε_t in Eq. (7.3.11). Then the new estimated noise N_t was obtained from Eq. (7.3.13),i.e.

$$(1-0.627B) z_t = (-0.266 + 0.316B + 0.160B^2 - 0.01B^3) x_t + \varepsilon_t^*$$

$$z_t \simeq (-0.266 + 0.149B + 0.16B^3 + 0.1B^3) x_t + \hat{N}_t$$

$$\textit{where } \hat{N}_t = \frac{\varepsilon_t^*}{(1-0.627B)}$$

Then, the cross correlation between the input x_t and the residual ε_t^* of the new noise $\hat{N}_t$ is checked. The result indicates that the x_t and $\hat{N}_t$ are uncorrelated. Hence, Eq. (7.3.13) is accepted.

From Eq. (7.3.13), the forecast function of the deseasonalized log-transformed flow z_t can be written as

$$\hat{z}_t(L) = 0.627 E[z_{t+L-1}] - 0.266 E[x_{t+L}] + 0.316 E[x_{t+L-1}] + 0.160 E[x_{t+L-2}] - 0.01 E[x_{t+L-3}] \quad (7.3.14)$$

In Eq. (7.3.14), we need to forecast the deseasonalized log-transformed SWE x_t. The deseasonalized log-transformed SWE of January using through April has an autocorrelation coefficient of $\hat{\rho}_1$=0.52. After using the least squares method, we find that

the deseasonalized SWE can be fitted using an AR(1) model as shown in Eq. (7.3.5), i.e.

$$\hat{x}_t(L) = 0.728\, E[x_{t+L-1}] \tag{7.3.15}$$

Equation (7.3.13) can be rewritten as

$$\begin{aligned} z_{\nu,\tau} &= (-0.266 - 0.044B + 0.221B^2 + \ldots)\zeta_{\nu,\tau} \\ &\quad + (1 + 0.627B + 0.393B^2 + \ldots)\varepsilon_{\nu,\tau} \\ &= \sum_{j=0}^{\infty} \nu_j \zeta_{\nu,\tau-j} + \sum_{j=0}^{\infty} \psi_j \varepsilon_{\nu,\tau-j} \end{aligned} \tag{7.3.16}$$

Then, the forecast error of the deseasonalized flow $z_{\nu,\tau}$ can be written as

$$e_{\nu,\tau}(L) = \sum_{j=0}^{L-1} \nu_j \zeta_{\nu,\tau+L-j} + \sum_{j=0}^{L-1} \psi_j \varepsilon_{\nu,\tau+L-j} \tag{7.3.17}$$

Since the normality hypothesis of the residual ε_t is rejected by the chi-square test, we do not need to check the normality of ζ_t. Only the linear combination of the normal variables will give a normal variable. Then, 1000 forecast errors with lead time L=2,...,7 were simulated. After obtaining the 95% quantiles, the forecast of deseasonalized flow with lead time L=1,... is shown in Fig. 7.42.

The log-transformed flow $Q_{\nu\tau}'$ can be obtained by using Eq. (7.3.4), i.e.

$$\hat{Q}'_{\nu,\tau}(L) = \hat{z}_{\nu,\tau}(L) * \hat{\sigma}_\tau(Q') + \hat{\mu}_\tau(Q') \tag{7.3.18}$$

The one step ahead forecast of the log-flow of the Los Pinos River (1978-1982) is shown in Fig. 7.43. Notice that this figure closely resembles the given forecast. The forecast error of the un-deseasonalized flow $Q_{\nu,\tau}'$ can be written as

$$e^*_{\nu,\tau}(L) = \left(\sum_{j=0}^{L-1} \nu_j \zeta_{\nu,\tau+L-j} + \sum_{j=0}^{L-1} \psi_j \varepsilon_{\nu,\tau+L-j}\right)\sigma_{\tau+L}(Q') + \mu_{\tau+L}(Q') \tag{7.3.19}$$

Hence, the variance of the forecast error of $Q_{v,\tau}$ can be written as

$$\sigma^{*2}_{v,\tau}(L)=\sigma^{2}_{\tau+L}(Q')[\sigma^{2}_{\zeta}\sum_{j=0}^{L-1}\nu_j^2 + \sigma^{2}_{\varepsilon}\sum_{j=0}^{L-1}\psi_j^2] \tag{7.3.20}$$

Then, the minimum mean square error forecast of the actual flow can be obtained by

$$\hat{Q}_{v,\tau}(L) = 10^{\hat{Q}'_{v,\tau}(L)+\frac{1}{2}\sigma^{*2}_{v,\tau}(L)} \tag{7.3.21}$$

The one step ahead forecast of the actual monthly flow (1978-1982) by using the Transfer Function Model (TFM) is shown in Fig. 7.44 and Table 7.1. In addition, the forecast results using the PAR(1) model are also shown in Fig. 7.44. From Fig. 7.44, we can see that the forecast using the TFM still underestimated the high flow in the summer time, but it is an improvement on the univariate time series model.

7.3.2 No Transformed Monthly Flow Modeling

From the forecast result shown in the last section, we can see that using log-transformation to forecast actual flow always underestimates the high flow in the summer time. Because this model minimizes the sum of square errors in the log domain, the forecast of the log-flow is better than the forecasted value of the actual flow. Now, we will not take any transformation of the original monthly flow, and attempt to fit a transfer function noise model to it.

The monthly flow $Q_{v,\tau}$ of the Los Pinos river near Ortiz was deseasonalized by using

$$z_{v,\tau} = \frac{Q_{v,\tau} - \hat{\mu}_{\tau}(Q)}{\hat{\sigma}_{\tau}(Q)} \tag{7.3.22}$$

and the monthly SWE $W_{v,\tau}$ of Cumbres Pass is deseasonalized by

$$x_{v,\tau} = \frac{W_{v,\tau} - \hat{\mu}_\tau(W)}{\hat{\sigma}_\tau(W)} \tag{7.3.23}$$

It has an AR(1) structure, and the deseasonalized SWE of May filled in by

$$x_t = 0.774\ x_{t-1} + \zeta_t \tag{7.3.24}$$

in which ζ_t zero mean and variance 0.311. Then the $z_{v,\tau}$ and $x_{v,\tau}$ was denoted as z_t and x_t, respectively. The impulse response function a_j of the single input-output transfer function model is estimated and is shown in Fig. 7.45. The transfer function model can then be written as

$$z_t = -0.286x_t + 0.175x_{t-1} + 0.270x_{t-2} + N_t \tag{7.3.25}$$

And the noise N_t can be estimated by

$$\hat{N}_t = z_t - (-0.286x_t + 0.175x_{t-1} + 0.27x_{t-2}) \tag{7.3.26}$$

This is plotted in Fig. 7.46. The ARMA model is fitted to the noise $\hat{N}_t$, which is an AR(1) model of

$$N_t = 0.593\ N_{t-1} + \varepsilon_t = \frac{\varepsilon_t}{1-0.593B} \tag{7.3.27}$$

Substituting (7.3.27) into (7.3.25) gives

$$\begin{aligned} z_t = {} & 0.593z_{t-1} - 0.286x_t + 0.345x_{t-1} + 0.166x_{t-2} \\ & - 0.160x_{t-3} + \varepsilon_t \end{aligned}$$

After using the least squares algorithm, we have

$$\begin{aligned} z_t = {} & 0.581z_{t-1} - 0.340x_t + 0.323x_{t-1} + 0.199x_{t-2} \\ & - 0.016x_{t-3} + \varepsilon_t^* \end{aligned} \tag{7.3.28}$$

The residual ε_t^* shown in Fig. 7.47 has mean zero, variance 0.555 and skewness

coefficient 1.18. The autocorrelation function of the residual ε_t^* is shown in Fig. 7.48. This figure shows that the residual ε_t^* is white noise. The hypothesis of normality of the residual ε_t^* is rejected by the skewness test of normality.

Then the cross correlation of input x_t and the residual of the new noise $\hat{N}_t$ is checked. It is concluded that the input x_t is uncorrelated with the noise $\hat{N}_t$, hence the Eq. (7.3.29) is accepted.

The forecast function of Eq. (7.3.29) can be written as

$$\hat{z}_t(L) = 0.581E[z_{t+L-1}] - 0.34E[x_{t+L}] + 0.323E[x_{t+L-1}] + 0.199E[x_{t+L-2}] - 0.016E[x_{t+L-3}] \tag{7.3.29}$$

where the SWE can be forecasted using Eq. (7.3.25). Then Eqs. (7.3.25), (7.3.30) and (7.3.23) can be used for forecasting the monthly streamflow.

Equation (7.3.29) can be rewritten as

$$\begin{aligned} z_{\nu,\tau} &= (-0.34-0.1377B+0.165B^2+0.2695B^3+...)\zeta_{\nu,\tau} \\ &\quad +(1+0.581B+(0.581B)^2+(0.581B)^3+....)\varepsilon_{\nu,\tau} \\ &= \sum_{j=0}^{\infty} \nu_j \zeta_{\nu,\tau-j} + \sum_{j=0}^{\infty} \psi_j \varepsilon_{\nu,\tau-j} \end{aligned} \tag{7.3.30}$$

Then the forecast error can be written as

$$e_{\nu,\tau}(L) = \sum_{j=0}^{L-1} \nu_j \zeta_{\nu,\tau+L-j} + \sum_{j=0}^{L-1} \psi_j \varepsilon_{\nu,\tau+L-j} \tag{7.3.31}$$

From (7.3.31), we can see that $e_{\nu,\tau}(L)$ has a mean zero, but with an unknown probability distribution. Since the distribution of ζ_t and ε_t are not normally distributed, a simulation of the forecast error will be used. After obtaining the 95% quantiles of the simulated

forecast error, the result of the forecasted deseasonalized flow with lead time L=1,...,7 and the 95% confidence is shown in Fig. 7.49. It shows that most of the historical and forecasted flow fall inside the 95% confidence limit bound.

The one step ahead forecast of monthly flow (1978-1982) of the Los Pinos river near Ortiz is shown in Fig. 7.50 and Table 7.1. In addition, the forecast result of the log-transformed transfer function model is shown in Fig.7.50. We can see that the no-log transformation forecasting is better then the log-transformation forecasting.

7.4 Single Input-Single Output Periodic Transfer Function Model

The rational transfer function model gave some improvement in forecasting the streamflow when the original flow data were not transformed. Two other forecasting schemes in using transfer function models are tried here. First is by fitting a PAR(1) model to the noise N_t and second is by filtering the deseasonalized flows with a PAR(1) model before applying the transfer function model. The two schemes are further discussed below.

7.4.1 Transfer Function Model With PAR(1) Noise

The noise N_t of Eq. (7.3.26) of Section 7.3.3 is plotted in Fig. 7.46, and its seasonal correlation is plotted in Fig. 7.51. Figure 7.51 shows that the noise N_t has periodic properties. Hence, a PAR(1) model may be fitted to N_t instead of fitting an ARMA model.

If the noise N_t is fitted by a PAR(1) model, it can be written as

$$N_{\nu,\tau} = \varphi_{1,\tau} N_{\nu,\tau-1} + \varepsilon_{\nu,\tau} \tag{7.4.1}$$

Now the transfer function model can be rewritten as

$$z_{v,\tau} = \sum_{j=0}^{\infty} a_j x_{v,\tau-j} + N_{v,\tau} \tag{7.4.2}$$

Substituting (7.4.1) into (7.4.2) gives

$$z_{v,\tau} = \sum_{j=0}^{\infty} a_j x_{v,\tau-j} + \frac{\varepsilon_{v,\tau}}{1-\varphi_{1,\tau}B}$$

or

$$z_{v,\tau} = \varphi_{1,\tau} z_{v,\tau-1} + \sum_{j=0}^{\infty} a_j x_{v,\tau-j} - \varphi_{1,\tau} \sum_{j=0}^{\infty} a_j x_{v,\tau-j-1} + \varepsilon_{v,\tau} \tag{7.4.3}$$

From Section 7.3.3, we know that the transfer function noise model of the Los Pinos River monthly streamflow can be written as (see Eq. (7.3.25))

$$z_{v,\tau} = -0.286\, x_{v,\tau} + 0.175\, x_{v,\tau-1} + 0.270\, x_{v,\tau-2} + N_{v,\tau} \tag{7.4.4}$$

The parameters of the PAR(1) model of Eq.(7.4.1) fitted to the noise $N_{v,\tau}$ gave

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.829	0.744	0.329	0.148	0.134	0.603
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.926	0.567	0.548	0.555	0.946	0.705

In addition, substituting (7.4.1) into (7.4.4) gives

$$\begin{aligned} z_{v,\tau} = \hat{\varphi}_{1,\tau} z_{v,\tau-1} + (1-\hat{\varphi}_{1,\tau}B)(-0.286 x_{v,\tau} + 0.175 x_{v,\tau-1} \\ + 0.270 x_{v,\tau-2}) + \varepsilon_{v,\tau} \end{aligned} \tag{7.4.5}$$

After minimizing the sum of square errors, we have

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.829	0.752	0.326	0.237	0.017	0.483
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	1.0	0.655	0.548	0.581	0.955	0.748

and

$$z_{v,\tau} = \hat{\varphi}_{1,\tau} z_{v,\tau-1} + (1-\hat{\varphi}_{1,\tau}B)(-0.105 x_{v,\tau} + 0.170 x_{v,\tau-1} + 0.334 x_{v,\tau-2}) + \varepsilon^{*}_{v,\tau} \tag{7.4.6}$$

The cross correlation between the input x_t and the residual of the new noise N_t is examined and found to be uncorrelated. Hence, Eq. (7.4.6) is accepted.

The forecast function of Eq. (7.4.6) can be written as

$$\hat{z}_{v,\tau}(L) = \hat{\varphi}_{1,\tau+L} E[z_{v,\tau+L-1}] + (1-\hat{\varphi}_{1,\tau+L}B)(-0.105 E[x_{v,\tau+L}] + 0.170 E[x_{v,\tau+L-1}] + 0.334 E[x_{v,\tau+L-2}]) \tag{7.4.7}$$

The one step ahead forecast of the Los Pinos River monthly flow using the foregoing transfer function model (TFM-PAR(1) Noise model) is shown in Fig. 7.52 and Table 7.1. In addition, the forecast using the rational transfer function model (without log-transformation) of Eq. (7.3.30) is also shown in Fig. 7.52. From Fig. 7.52, we can see that the transfer function model incorporating PAR(1) noise gives a similar forecast result of the high flow in the summer time. In fact, this can be observed from Table 7.1.

7.4.2 Periodic-Transfer Function Model

In the previous section, the noise N_t was fitted by a PAR(1) model. Now, the deseasonalized streamflow $z_{v,\tau}$ is first filtered by a PAR(1) model before the transfer function model is applied. The deseasonalized flows of the Los Pinos River has a monthly correlation $\rho_{k,\tau}$ as shown in Fig. 7.53. Let the original monthly flow be denoted

by $Q_{v,\tau}$. Then the deseasonalized flow $z_{v,\tau}$ can be computed by

$$z_{v,\tau} = \frac{Q_{v,\tau} - \hat{\mu}_\tau(Q)}{\hat{\sigma}_\tau(Q)} \tag{7.4.8}$$

The deseasonalized flow is then filtered by PAR(1) model using

$$z_{v,\tau} = \hat{\varphi}_{1,\tau} z_{v,\tau-1} + y_{v,\tau} \tag{7.4.9}$$

where

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.829	0.771	0.226	0.227	0.170	0.751
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.960	0.584	0.548	0.587	0.950	0.714

Then, a transfer function model between the deseasonalized SWE $x_{v,\tau}$ as the input and the variable $y_{v,\tau}$ as the output can be written as

$$y_{v,\tau} = \sum_{j=0}^{\infty} a_j x_{v,\tau-j} + N_{v,\tau} \tag{7.4.10}$$

The impulse response function is estimated using spectral analysis and was shown in Fig. 7.54. Then Eq. (7.4.10) can be written as

$$y_{v,\tau} = -0.173 x_{v,\tau} + 0.269 x_{v,\tau-1} + 0.161 x_{v,\tau-2} + N_{v,\tau} \tag{7.4.11}$$

The noise $N_{v,\tau}$ is plotted in Fig. 7.55. An ARMA model is fitted to $\hat{N}_t$ ($t=(v-1)12+\tau$), but we find that the $\hat{N}_t$ is white noise with mean zero and variance 0.504. Hence, Eq.(7.4.11) can be written as

$$y_{v,\tau} = -0.173 x_{v,\tau} + 0.269 x_{v,\tau-1} + 0.161 x_{v,\tau-2} + \varepsilon_{v,\tau} \tag{7.4.12}$$

By substituting (7.4.12) into (7.4.9), we have

$$z_{v,\tau} = \hat{\varphi}_{1,\tau} z_{v,\tau-1} - 0.173 x_{v,\tau} + 0.269 x_{v,\tau-1} + 0.161 x_{v,\tau-2} + \varepsilon_{v,\tau} \tag{7.4.13}$$

Minimizing the sum of square errors, a final model is given as

$$z_{v,\tau} = \hat{\varphi}_{1,\tau} z_{v,\tau-1} - 0.207 x_{v,\tau} + 0.241 x_{v,\tau-1} + 0.324 x_{v,\tau-2} + \varepsilon^{*}_{v,\tau} \tag{7.4.14}$$

where

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.829	0.771	0.226	0.216	0.126	0.462
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.797	0.584	0.548	0.575	0.954	0.731

and v=1,2,...,N and τ=4,5,...,12 and $\varepsilon^{*}_{v,\tau}$ has mean zero and variance 0.352.

The hypothesis of independence of residuals $\varepsilon^{*}_{v,\tau}$ has been accepted. The cross correlation of $x_{v,\tau}$ and the residual $\hat{N}_{v,\tau}$ is shown in Fig. 7.56. It says that the two variables are uncorrelated.

Then, the forecast function $\hat{z}_{v,\tau}(L)$ of Eq.(7.4.14) can be written as

$$\hat{z}_{v,\tau}(L) = \hat{\varphi}_{1,\tau+L} E[z_{v,\tau+L-1}] - 0.207 E[x_{v,\tau+L}] + 0.241 E[x_{v,\tau+L-1}] + 0.324 E[x_{v,\tau+L-2}] \tag{7.4.15}$$

and

$$\hat{x}_{v,\tau}(L) = 0.774 \; E[x_{x,\tau+L-1}] \tag{7.4.16}$$

By using Eqs. (7.4.15) and (7.4.16), the deseasonalized flow $z_{v,\tau}$ can be forecasted. Then, actual flow is forecasted and shown in Eq. (7.4.8), i.e

$$\hat{Q}_{v,\tau} = \hat{z}_{v,\tau} \; \hat{\sigma}_{\tau}(Q) + \hat{\mu}_{\tau}(Q) \tag{7.4.17}$$

Equation (7.4.14) can be rewritten as

$$z_{\nu,\tau}=\frac{(-0.207+0.241B+0.324B^2)}{(1-\hat{\varphi}_{1,\tau}B)}\frac{1}{(1-0.774B)}\zeta_{\nu,\tau}$$

$$+\frac{1}{(1-\hat{\varphi}_{1,\tau}B)}\varepsilon_{\nu,\tau}$$

$$= \nu_\tau(B)\zeta_{\nu,\tau} + \psi_\tau(B)\varepsilon_{\nu,\tau}$$

where ν_τ(B) and Ψ_τ(B) can be estimated by substituting $\hat{\varphi}_{1,\tau}$ into above equation, for τ=1,2,...,12.

The forecast error is obtained using Eq.(4.6.14). The simulated cumulative relative frequency of the forecast error with τ=3, and lead time L=2,..,7 , is computed, and the 2.5% and 97.5% quantiles are:

	Lead Time L						
	1	2	3	4	5	6	7
$q_{97.5}$	2.19	2.00	2.27	2.01	2.39	2.17	3.07
$q_{2.5}$	-0.87	-0.71	-0.88	-1.12	-1.18	-1.04	-1.22

The forecast of deseasonalized flow $z_{\nu,\tau}$ with lead time L=1,2,..,7 and the bound of 95% confidence limits are shown in Fig. 7.57. This shows that most of the historical and forecasted flow fall inside the bound.

The one step ahead forecast using the transfer function model (periodic-TFM model) is shown in Fig. 7.58 and Table 7.1. For comparison, the results of the previous section (TFM-PAR(1) Noise model) are also shown in Fig. 7.58. From Fig. 7.58 and Table 7.1, we can see that using the residual of the PAR(1) model as the output of the transfer function model (periodic-TFM model) gives a much better forecast result.

7.5 Existing Forecasting Model

The forecasting method currently used by the agency responsible for administering the Rio Grande River basin in Colorado is the multiple regression (MR) model. Multiple regression analysis offers a way to fit linear models in which the variables are independent of time. Standing on April 1st (or standing on March 31), the SWE and flow of January, February and March are known. Hence, a forecast of April's streamflow can be made depending on the flow and SWE of January, February and March. The Colorado State Engineer's Office uses multiple regression (MR) equations for forecasting the "cumulative" monthly streamflow of the Rio Grande River basin. The forecast results will be compared to the results of the periodic-TFM model. This will be shown in the next section.

The monthly streamflows (1948-1977) are analyzed by using multiple regression. The equations are shown below :

Q1 = Cumulative flow through March

Q2 = Cumulative flow through April

Q3 = Cumulative flow through May

Q4 = Cumulative flow through June

Q5 = Cumulative flow through July

Q6 = Cumulative flow through August

Q7 = Cumulative flow through September

Q8 = Cumulative flow through October (except for Los Pinos River, it is April through October)

Q9 = Cumulative flow through November

Q10= Cumulative flow through December

WC1= April 1st (March 31) snow water content

WC2= May 1st (April 30) snow water content

I. Los Pinos River near Ortiz

(1) April first forecast

$$Q2 = 2.725\ Q1 + 13.68\ WC1 + 5199.7$$

$$Q3 = -0.419\ Q1 + 1794.26\ WC1 + 15565.1$$

$$Q4 = -2.580\ Q1 + 3277.27\ WC1 + 11866.8$$

$$Q5 = -3.019\ Q1 + 3671.35\ WC1 + 9972.6$$

$$Q6 = -3.183\ Q1 + 3772.40\ WC1 + 10704.5$$

$$Q7 = -3.247\ Q1 + 3797.30\ WC1 + 11691.1$$

$$Q8 = -4.146\ Q1 + 3797.60\ WC1 + 12637.9$$

(2) May first forecast

$$Q3 = 1.410\ Q2 - 3.38\ Q1 + 1811.9\ WC1 + 0.0\ WC2 + 4352.0$$

$$Q4 = 2.081\ Q2 - 4.99\ Q1 + 0.0\ WC1 + 2442.9\ WC2 + 16923.8$$

$$Q5 = 2.084\ Q2 - 5.00\ Q1 + 0.0\ WC1 + 2753.4\ WC2 + 16581.8$$

$$Q6 = 2.025\ Q2 - 4.86\ Q1 + 0.0\ WC1 + 2846.3\ WC2 + 17741.5$$

$$Q7 = 2.004\ Q2 - 4.81\ Q1 + 0.0\ WC1 + 2888.7\ WC2 + 18526.8$$

$$Q8 = 1.985\ Q2 - 4.76\ Q1 + 0.0\ WC1 + 2945.8\ WC2 + 15560.0$$

(3) June first forecast

$$Q4 = -0.807\ Q2 + 1.394\ Q3 + 980.238\ WC1 + 0.0\ WC2 - 8528.412$$

$$Q5 = -0.997\,Q2 + 1.455\,Q3 + 1311.911\,WC1 + 0.0\,WC2 - 10799.074$$

$$Q6 = -1.079\,Q2 + 1.452\,Q3 + 1436.511\,WC1 + 0.0\,WC2 - 9730.125$$

$$Q7 = -0.705\,Q2 + 1.313\,Q3 + 882.054\,WC1 + 656.969\,WC2 - 6161.412$$

$$Q8 = -0.813\,Q2 + 1.338\,Q3 + 852.876\,WC1 + 690.022\,WC2 - 7880.633$$

(4) July first forecast

$$Q5 = 200.738\,WC1 + 1.063\,Q4 - 3864.884$$

$$Q6 = 314.149\,WC1 + 1.061\,Q4 - 3857.851$$

$$Q7 = 314.398\,WC1 + 1.069\,Q4 - 3158.031$$

$$Q8 = 316.462\,WC1 + 1.080\,Q4 - 6193.543$$

(5) August first forecast

$$Q6 = 1.316\,Q5 - 0.33\,Q4 + 1528.8$$

$$Q7 = 1.407\,Q5 - 0.42\,Q4 + 2407.9$$

$$Q8 = 1.566\,Q5 - 0.58\,Q4 - 302.0$$

(6) September first forecast

$$Q7 = 1.290\,Q6 - 0.287\,Q5 + 411.15$$

$$Q8 = 1.584\,Q6 - 0.575\,Q5 - 2916.90$$

(7) October first forecast

$$Q8 = 1.600\,Q7 - 0.594\,Q6 - 3501.30$$

II. <u>The Conejos River at Mogote</u>

(1) April first forecast

$$Q2 = 1.858\,Q1 + 16.302\,WC1 + 7478.868$$

$$Q3 = 3.352\,Q1 + 2672.829\,WC1 + 22101.048$$

Q4 = 5.800 Q1 + 6699.421 WC1 + 14374.223

Q5 = 6.302 Q1 + 9263.342 WC1 + 6021.483

Q6 = 5.891 Q1 + 9980.287 WC1 + 13637.581

Q7 = 5.777 Q1 + 10037.763 WC1 + 20134.652

Q8 = 6.123 Q1 + 10056.591 WC1 + 21830.556

Q9 = 6.510 Q1 + 10450.114 WC1 + 21507.972

Q10= 6.613 Q1 + 10465.899 WC1 + 23212.880

(2) May first forecast

Q3 =-1.757 Q1 + 1.097 Q2 + 2459.202 WC1 + 0. WC2 + 48409.308

Q4 = 3.322 Q1 + 1.292 Q2 + 2518.432 WC1 + 5134.838 WC2 + 22064.553

Q5 = 5.231 Q1 + 0.515 Q2 + 3767.340 WC1 + 7636.339 WC2 + 20029.543

Q6 = 5.770 Q1 + 0. Q2 + 4444.404 WC1 + 8080.084 WC2 + 29454.527

Q7 = 5.648 Q1 + 0. Q2 + 4098.860 WC1 + 8668.324 WC2 + 37103.312

Q8 = 5.989 Q1 + 0. Q2 + 3926.704 WC1 + 8947.083 WC2 + 39344.727

Q9 = 6.365 Q1 + 0. Q2 + 3778.325 WC1 + 9738.032 WC2 + 40570.700

Q10= 6.465 Q1 + 0. Q2 + 3696.151 WC1 + 9881.006 WC2 + 42556.016

(3) June first forecast

Q4 = 2.991 Q1 +0. Q2 +0.822 Q3 + 1977.3 WC1 + 3687.1 WC2 + 3434.121

Q5 = 5.231 Q1 +0.515 Q2 +0. Q3 + 3767.3 WC1 + 7636.3 WC2 + 20029.543

Q6 = 5.770 Q1 +0. Q2 +0. Q3 + 4444.4 WC1 + 8080.1 WC2 + 29454.527

Q7 = 4.764 Q1 +0. Q2 +0.264 Q3 + 3448.7 WC1 + 8588.0 WC2 + 31115.731

Q8 = 5.260 Q1 +0. Q2 +0.218 Q3 + 3390.1 WC1 + 8880.8 WC2 + 34403.601

Q9 = 6.365 Q1 +0. Q2 +0. Q3 + 3778.3 WC1 + 9738.0 WC2 + 40570.700

Q10= 6.465 Q1 +0. Q2 +0. Q3 + 3696.2 WC1 + 9881.0 WC2 + 42556.016

(4) July first forecast

Q5 = -0.826 Q3 + 1.479 Q4 + 2035.159 WC2 + 13737.839

Q6 = -0.941 Q3 + 1.524 Q4 + 3020.592 WC2 + 22485.564

Q7 = -0.971 Q3 + 1.541 Q4 + 3218.481 WC2 + 27368.193

Q8 = -1.019 Q3 + 1.556 Q4 + 3316.443 WC2 + 34343.324

Q9 = -1.049 Q3 + 1.550 Q4 + 4057.991 WC2 + 41688.414

Q10= -1.051 Q3 + 1.549 Q4 + 4123.882 WC2 + 44545.056

(5) August first forecast

Q6 = -0.421 Q2 +0. Q3 +0. Q4 +1.064 Q5 +340.1 WC1 + 7613.067

Q7 = -0.624 Q2 +0.233 Q3 -0.333 Q4 +1.308 Q5 -141.3 WC1 + 11253.088

Q8 = 0. Q2 +0. Q3 -0.388 Q4 +1.414 Q5 -402.2 WC1 + 12903.124

Q9 = -0.526 Q2 +0.444 Q3 -0.871 Q4 +1.719 Q5 -566.6 WC1 + 17315.350

Q10= -0.532 Q2 +0.473 Q3 -0.920 Q4 +1.759 Q5 -678.1 WC1 + 19293.695

(6) September first forecast

Q7 = 1.179 Q6 - 0.170 Q5 + 2219.267

Q8 = 1.243 Q6 - 0.228 Q5 + 6127.322

Q9 = 1.261 Q6 - 0.203 Q5 + 6449.658

Q10= 1.281 Q6 - 0.221 Q5 + 8658.509

(7) October first forecast

Q8 = 1.350 Q7 - 0.347 Q6 + 2840.691

Q9 = 1.701 Q7 - 0.663 Q6 + 2393.926

Q10= 1.772 Q7 - 0.731 Q6 + 4266.150

(8) November first forecast

Q9 = 1.640 Q8 - 0.603 Q7 - 2469.715

Q10= 1.787 Q8 - 0.747 Q7 - 955.047

(9) December first forecast

Q10 = 1.067 Q9 - 0.064 Q8 + 1937.609

III. The Rio Grande River at Del Norte

(1) April first forecast

Q2 = 1.293 WC1 + 2.365 Q1 - 5856.478

Q3 = 8606.811 WC1 + 4.602 Q1 - 42285.713

Q4 = 22370.001 WC1 + 5.635 Q1 - 69655.461

Q5 = 31514.500 WC1 + 5.322 Q1 - 85845.171

Q6 = 35314.086 WC1 + 4.665 Q1 - 66098.511

Q7 = 36096.973 WC1 + 4.259 Q1 - 38175.529

Q8 = 36663.294 WC1 + 4.198 Q1 - 22496.189

Q9 = 37012.085 WC1 + 4.163 Q1 - 13696.109

Q10= 37218.594 WC1 + 4.076 Q1 - 1246.523

(2) May first forecast

Q3 = 2.55836 Q1+0.86449 Q2 + 7131.365 WC1 -20097.870

Q4 = 3.56534 Q1+1.75166 Q2 + 8916.738 WC1 +14057.034 WC2-73928.143

Q5 = 3.03514 Q1+2.39070 Q2+10436.153 WC1+22835.454 WC2-104461.262

Q6 = 0. Q1+4.25523 Q2 + 0. WC1 +36508.462 WC2 - 75066.235

Q7 = 0. Q1+4.18416 Q2 + 0. WC1 +37656.716 WC2 - 54726.133

Q8 = 0. Q1+4.25090 Q2 + 0. WC1 +38137.450 WC2 - 42891.121

Q9 = 0. Q1+4.25746 Q2 + 0. WC1 +38540.780 WC2 - 33047.004

Q10= 0. Q1+4.27500 Q2 + 0. WC1 +38776.700 WC2 - 25721.400

(3) June first forecast

Q4 = 22167.564 WC1 - 0.859 Q2 + 0.406 Q3 + 99145.52

Q5 = 31323.307 WC1 - 0.832 Q2 + 0.393 Q3 + 73038.82

Q6 = 35146.496 WC1 - 0.832 Q2 + 0.393 Q3 + 70424.24

Q7 = 35943.968 WC1 - 1.041 Q2 + 0.492 Q3 + 78923.35

Q8 = 36512.481 WC1 - 1.088 Q2 + 0.514 Q3 + 91280.87

Q9 = 36862.529 WC1 - 1.134 Q2 + 0.536 Q3 + 97644.04

Q10= 37072.164 WC1 - 1.083 Q2 + 0.512 Q3 + 108491.5

(4) July first forecast

Q5 = 0. Q1 -0.375 Q3 +1.392 Q4 + 0. WC1 +4374.646 WC2 -1144.079

Q6 = 0. Q1 -0.604 Q3 +1.479 Q4 +4334.632 WC1 +4158.530 WC2 -12675.132

Q7 =-1.943 Q1 -0.520 Q3 +1.648 Q4 + 0. WC1 +4713.998 WC2 +40520.499

Q8 =-2.426 Q1 -0.497 Q3 +1.686 Q4 + 0. WC1 +4054.769 WC2 +62172.846

Q9 =-2.569 Q1 -0.528 Q3 +1.717 Q4 + 0. WC1 +3944.711 WC2 +75966.596

Q10=-2.661 Q1 -0.548 Q3 +1.741 Q4 + 0. WC1 +3765.227 WC2 +85979.350

(5) August first forecast

Q6 = 1812.232 WC1 + 1.064 Q5 - 8732.255

Q7 = 506.971 WC1 + 1.131 Q5 - 1024.515

Q8 = 231.634 WC1 + 1.158 Q5 + 10011.53

Q9 = 640.883 WC1 + 1.156 Q5 + 17800.47

Q10= - 29.040 WC1 + 1.184 Q5 + 24613.21

(6) September first forecast

Q7 = 1.337 Q6 - 0.342 Q5 + 11276.498

Q8 = 1.406 Q6 - 0.398 Q5 + 23075.514

Q9 = 1.511 Q6 - 0.503 Q5 + 31805.717

Q10= 1.561 Q6 - 0.551 Q5 + 39086.83

(7) October first forecast

Q8 = 1.323 Q7 - 0.314 Q6 + 8021.127

Q9 = 1.472 Q7 - 0.455 Q6 + 14381.271

Q10= 1.548 Q7 - 0.526 Q6 + 20408.532

(8) November first forecast

Q9 = 1.394 Q8 - 0.390 Q7 + 3012.132

Q10=1.572 Q8 - 0.563 Q7 + 7694.574

(9) December first forecast

Q10 = 1.402 Q9 - 0.399 Q8 + 3574.11

7.6 Comparison of the Forecast

From the forecast results shown in sections 7.1-7.4 , we can see that the periodic-TFM model gives the best forecast results. In this section, we will compare the forecasting ability of the periodic-TFM model to the Multiple Regression model. A

comparison of the forecast results will be assessed using a graphic comparison, the root mean square deviation (RMSD) and the maximum absolute deviation (MAD), i.e.

$$RMSD = \sqrt{\frac{\sum_{k=1}^{m} e_{t+k}^2}{m}}$$

and

$$MAD = \max|e_{t+k}|, \quad for\ k=1,2,...,m$$

where e_t is the deviation between the forecasted value and the historical value.

Since the monthly streamflows of 1948-1977 have been analyzed by the Colorado State Engineer's Office using the multiple regression (MR) method, testing of its forecasting ability will be proceeded by an analysis of the flow data from 1948 to 1977 and validation of the monthly flow in the years of 1978-1982. Some numerical computation procedure will be presented.

7.6.1 The Los Pinos River Basin

The periodic-TFM method presented in Section 7.4 for forecasting the monthly streamflow of the Los Pinos River, gives a better forecast result than other models studied. A demonstration of the computation procedure will be presented. The forecast equation of the periodic-TFM method for the Los Pinos River near Ortiz is given by Eq. (7.4.15) and (7.4.16), i.e

$$\hat{z}_{v,\tau}(L) = \hat{\varphi}_{1,\tau+L} E[z_{v,\tau+L-1}] - 0.207\, E[x_{v,\tau+L}] + 0.241\, E[x_{v,\tau+L-1}] + 0.324\, E[x_{v,\tau+L-2}]$$

and

$$\hat{x}_{\nu,\tau}(L) = 0.774\ E[x_{\nu,\tau+L-1}]$$

where $\hat{\varphi}_{1,\tau}$ has the values shown below:

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.829	0.771	0.226	0.216	0.126	0.462
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.797	0.584	0.548	0.575	0.954	0.731

Assume we want to forecast the streamflow of April by standing in March (or April 1) 1978. We already know the actual streamflow and the SWE of January, February and March. These are shown below:

	Streamflow	SWE
Jan 1978	602	12.1
Feb. 1978	730	17.1
Mar. 1978	1073	29.7

Then, by subtracting the historical monthly mean and dividing the historical monthly standard deviation, the deseasonalized streamflow and the deseasonalized SWE are shown below:

	Deseasonalized Flow	Deseasonalized SWE
Jan 1978	-1.0016	-0.1738
Feb. 1978	-0.6449	0.0867
Mar. 1978	-0.8198	1.6293

The one step ahead forecast of the April deseasonalized streamflow can be computed by

using Eq. (7.4.15), i.e

$$\begin{aligned}\hat{z}_{1978,3}(1) &= \hat{\varphi}_{1,4}E[z_{1978,3}] - 0.207\,E[x_{1978,4}] \\ &\quad + 0.24\,E[x_{1978,3}] + 0.324\,E[x_{1978,2}] \\ &= \hat{\varphi}_{1,4}z_{1978,3} - 0.207\,\hat{x}_{1978,3}(1) \\ &\quad + 0.24x_{1978,3} + 0.324x_{1978,2}\end{aligned} \tag{7.6.1}$$

From Eq. (7.4.16), we know that

$$\hat{x}_{1978,3}(1) = 0.774x_{1978,3} = 1.2610 \tag{7.6.2}$$

Substituting (7.6.2) into (7.6.1) gives a forecast value for the April deseasonalized streamflow; i.e.

$$\begin{aligned}\hat{z}_{1978,3}(1) &= (0.216)(-0.8198) - (0.207)(1.2610) \\ &\quad + (0.24)(1.6293) + (0.324)(0.0867) = -0.01898\end{aligned}$$

The forecasted actual April streamflow can be obtained by multiplying the historical April standard deviation and adding the April mean value, i.e.

$$\begin{aligned}\hat{y}_{1978,4} &= (-0.01898)(6888.070) + 12062.87 \\ &= 11932.13\end{aligned}$$

Using the same method, when we stand at the end of April (or May 1), the one step ahead forecast for May streamflow can be obtained by using

$$\begin{aligned}\hat{z}_{1978,4}(1) &= \hat{\varphi}_{1,5}E[z_{1978,4}] - 0.207\,E[x_{1978,5}] \\ &\quad + 0.24\,E[x_{1978,4}] + 0.324\,E[x_{1978,3}] \\ &= \hat{\varphi}_{1,5}z_{1978,4} - 0.207\,\hat{x}_{1978,4}(1) \\ &\quad + 0.24x_{1978,4} + 0.324x_{1978,4}\end{aligned} \tag{7.6.3}$$

where only the deseasonalized SWE of May is unknown. It can be forecasted using

(7.6.4)

$$\hat{x}_{1978,4}(1) = 0.774 x_{1978,4}$$

Then May streamflow can be forecasted by Eqs. (7.6.3) and (7.6.4).

The one step ahead forecast results of the streamflow (1978-1982) using the periodic-TFM method are shown in Fig. 7.59. In addition, the forecast results of using multiple regression (MR) equations of section 7.5 are also plotted in Fig. 7.59. One can observe that the periodic-TFM method gives better forecast results than the multiple regression model.

Next, we forecast the streamflow with lead time L=1,2,...,7. For example, if we need to forecast the April, May, June,... streamflow by standing at the end of March 1978, then the only information available is the streamflow and SWE before April 1978. The streamflow forecast can be computed by using Eqs. (7.4.15), (7.4.16) and (7.4.17), i.e

For lead time L=1

Based on Eq. (7.6.1), the forecasted values are

$$\text{April deseasonalized flow} \quad \hat{z}_{1978,3}(1) = -0.01898 \tag{7.6.5}$$

$$\text{April actual flow} \quad \hat{y}_{1978,4} = 11932.13$$

For lead time L=2

$$\begin{aligned} \hat{z}_{1978,3}(2) &= \hat{\varphi}_{1,5} E[z_{1978,4}] - 0.207 E[x_{1978,5}] \\ &\quad + 0.24 E[x_{1978,4}] + 0.324 E[x_{1978,3}] \\ &= \hat{\varphi}_{1,5} \hat{z}_{1978,3}(1) - 0.207 \hat{x}_{1978,3}(2) \\ &\quad + 0.24 \hat{x}_{1978,3}(1) + 0.324 x_{1978,3} \end{aligned} \tag{7.6.6}$$

where $\hat{x}_{1978,3}(1) = 0.774\ x_{1978,3} = 1.261$ (7.6.7)

and $\hat{x}_{1978,3}(2) = 0.774\ \hat{x}_{1978,3}(1) = 0.976$ (7.6.8)

Substituting Eqs. (7.6.5), (7.6.7) and (7.6.8) into (7.6.6) gives the forecasted May deseasonalized streamflow

$$\hat{z}_{1978,3}(2) = (0.126)(-0.01898)-(0.207)(0.976)+(0.24)(1.261)$$
$$+(0.324)(1.6293) = 0.6261$$

Then, May streamflow forecast is computed by

$$\hat{y}_{1978,5} = (0.6261)(16963.010) + 32025.83 = 42646.37$$

By using the same method shown above, the forecast with lead time L=3,4,.. can be obtained.

When we stand at the end of June, the SWE does not contribute to the streamflow of July, August,...,December. This can be seen from the impulse response function of Fig. 7.54. Hence, the streamflow forecast after June will be different from those of earlier months. By looking at Eq. (7.4.14), we can delete all SWE terms which gives

$$z_{v,\tau} = \varphi_{1,\tau}\ z_{v,\tau-1} \quad (7.6.9)$$

This is a PAR(1) model. If we re-estimate the parameters of $\varphi_{1,\tau}$ by using the least squares method, we get

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.829	0.771	0.226	0.227	0.171	0.752
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.960	0.584	0.548	0.587	0.951	0.714

By standing at the end of March, the streamflow forecasts for Los Pinos River

near Ortiz with lead times L=1,2,...,7 are shown in Fig. 7.60. Then the cumulative streamflow was computed and was shown in Fig. 7.61. By standing at the end of April and May, the forecast of the May, June actual streamflow and the cumulative streamflow with lead time L=1,2,...,6 are shown in Figs. 7.62-7.65. By standing in the June, July, August and September, the forecast of the July, August, September and October, are shown in Figs. 7.66-7.73.

Then, the forecasted root mean square deviation (RMSD) and the maximum absolute deviation (MAD) are used for comparing the forecast ability of the periodic-TFM method and the multiple regression method. The computed values of the RMSD and MAD are shown in Table 7.2. From the Figs. 7.60-7.73 and Table 7.2, we can see that the periodic-TFM method gives better results for forecasting the monthly streamflow of the Los Pinos River near Ortiz than the multiple regression method.

7.6.2 The Conejos River Basin

There are two SWE collecting sites in the Conejos River Basin, the Platoro and River Springs stations (see Fig. 6.1). The average SWE of the Platoro and the River Springs sites will be used as the input to the Conejos River system. After the SWE's were deseasonalized, an AR(1) model can be fitted to the deseasonalized SWE of January, February, March and April. This is written as

$$x_t = 0.903\ X_{t-1} + \varepsilon_t \qquad (7.6.10)$$

and it can be used for forecasting the deseasonalized SWE of May.

Then the streamflow $Q_{v,\tau}$ is also deseasonalized by using

$$z_{\nu,\tau} = \frac{Q_{\nu,\tau} - \hat{\mu}_\tau(Q)}{\hat{\sigma}_\tau(Q)} \tag{7.6.11}$$

and the deseasonalized streamflow was filtered by a PAR(1) model

$$z_{\nu,\tau} = \hat{\varphi}_{1,\tau} z_{\nu,\tau-1} + y_{\nu,\tau} \tag{7.6.12}$$

After estimating the residual $y_{v,\tau}$, an input-output model with the residual $y_{v,\tau}$ as the output and with the deseasonalized SWE as the input, can be formulated. The impulse response function is computed and is shown in Fig. 7.74. This figure suggests the model of

$$\begin{aligned} y_{v,\tau} = {} & -0.178\, x_{v,\tau} + 0.400\, x_{v,\tau-1} + 0.054\, x_{v,\tau-2} + 0.196\, x_{v,\tau-3} \\ & -0.230\, x_{v,\tau-4} - 0.032\, x_{v,\tau-5} + 0.221\, x_{v,\tau-6} + N_{v,\tau} \end{aligned} \tag{7.6.13}$$

The noise $N_{v,\tau}$ is estimated from equation (7.6.19) and has an AR(1) structure given by

$$N_t = -0.114\, N_{t-1} + \varepsilon_t \tag{7.6.14}$$

with $\varepsilon_t \sim$ WN(0,0.6). Substituting (7.6.13) and (7.6.14) into (7.6.12) gives

$$\begin{aligned} z_{v,\tau} = {} & (\hat{\varphi}_{1,\tau} - 0.114) z_{v,\tau-1} + 0.114\, \hat{\varphi}_{1,\tau-1}\, z_{v,\tau-2} - 0.178\, x_{v,\tau} \\ & + 0.38\, x_{v,\tau-1} + 0.100\, x_{v,\tau-2} + 0.202\, x_{v,\tau-3} - 0.208\, x_{v,\tau-4} \\ & - 0.06\, x_{v,\tau-5} + 0.217\, x_{v,\tau-6} + 0.025\, x_{v,\tau-7} + \varepsilon_{v,\tau} \end{aligned}$$

where $\varphi_{1,\tau}$ is estimated by $\rho_{1,\tau}$. Then the non-linear least squares method is used to estimate the parameters. The final model is

$$z_{v,\tau} = (\hat{\varphi}_{1,\tau}\text{-}0.263)z_{v,\tau\text{-}1} + 0.263\ \hat{\varphi}_{1,\tau\text{-}1}\ z_{v,\tau\text{-}2} - 0.329\ x_{v,\tau}$$
$$+ 0.363\ x_{v,\tau\text{-}1} + 0.167\ x_{v,\tau\text{-}2} + 0.321\ x_{v,\tau\text{-}3} + 0.01\ x_{v,\tau\text{-}4}$$
$$- 0.351\ x_{v,\tau\text{-}5} + 0.369\ x_{v,\tau\text{-}6} - 0.039\ x_{v,\tau\text{-}7} + \varepsilon^{*}_{v,\tau} \qquad (7.6.15)$$

with $\varepsilon^{*}_{v,\tau} \sim WN(0,0.42)$ and $\hat{\varphi}_{1,\tau}$ has values of

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.557	0.689	0.273	0.421	0.288	0.196
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.744	0.569	0.532	0.622	0.613	0.910

The cross correlation between the residual $\varepsilon^{*}_{v,\tau}$ of the noise $N_{v,\tau}$ and the input series $x_{v,\tau}$ was checked to be uncorrelated, hence Eqs. (7.6.15) and (7.6.10) can be used to forecast the streamflow of the Conejos River.

The one step ahead forecast of the streamflow (1978-1982) is shown in Fig. 7.75. The forecasts with lead time L=1,2,.. by standing in the March is shown in Fig. 7.76. In addition, the multiple regression equations of Section 7.5 are also used, and the forecast results are shown in Figs. 7.75-7.76. From Figs. 7.75-7.76, we can see that both forecasting methods underestimate the high flows in the summer time.

In order to overcome underestimation of the high flow, the SWE data was re-examined. The SWE at Cumbres Pass in the Los Pinos River basin (see Fig. 6.1) was found to have some contribution to streamflow in the Conejos River basin. By taking the SWE at Cumbres Pass to be one of the inputs to the Conejos River basin, the average SWE in the Conejos River basin is computed by

Average SWE = 0.44 Cumbres Pass + 0.12 Platoro + 0.44 River Springs

The average SWE is deseasonalized and fitted to an AR(1) model, i.e.

$$x_t = 0.88\, x_{t-1} + \varepsilon_t \tag{7.6.16}$$

By using the residual of Eq. (7.6.12) as the output, and using the new deseasonalized SWE as the input to the Conejos River system, the impulse response function can be computed and shown in Fig. 7.77. This suggests the input-output model can be written as

$$y_{v,\tau}=0.36x_{v,\tau-1} +0.062x_{v,\tau-2} +0.15x_{v,\tau-3} -0.183x_{v,\tau-4} +N_{v,\tau} \tag{7.6.17}$$

The noise $N_{v,\tau}$, estimated from Eq. (7.6.23), is model as

$$N_t = -0.039\, N_{t-1} + \varepsilon_t \tag{7.6.18}$$

with $\varepsilon_t \sim$ WN(0,0.68). Substituting Eqs. (7.6.17) and (7.6.18) into (7.6.12) gives

$$\begin{aligned} z_{v,\tau} = {} & (\hat{\varphi}_{1,\tau} -0.039)z_{v,\tau-1} +0.039\, \hat{\varphi}_{1,\tau-1}\, z_{v,\tau-2} +0.36\, x_{v,\tau-1} +0.076\, x_{v,\tau-2} \\ & + 0.152\, x_{v,\tau-3} -0.177\, x_{v,\tau-4} -0.007\, X_{V,\tau-5} + \varepsilon_{V,\tau} \end{aligned} \tag{7.6.19}$$

After re-estimating the parameters of Eq. (7.6.19) using the non-linear least squares algorithm, the input-output model is obtained as

$$\begin{aligned} z_{v,\tau}=(\hat{\varphi}_{1,\tau}-0.263)z_{v,\tau-1} & +0.263\hat{\varphi}_{1,\tau-1}z_{v,\tau-2} +0.040x_{v,\tau-1} +0.233x_{v,\tau-2} \\ & + 0.267\, x_{v,\tau-3} +0.0197\, x_{v,\tau-4} +0.031\, X_{V,\tau-5} + \varepsilon^*_{V,\tau} \end{aligned} \tag{7.6.20}$$

with $\varepsilon^*_{v,\tau} \sim$ WN(0,0.45) and $\hat{\varphi}_{1,\tau}$ has the values of

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.557	0.689	0.258	0.425	0.283	0.487
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.620	0.533	0.480	0.681	0.645	0.883

Then, the residual $\varepsilon^*_{v,\tau}$ of the new noise is checked to be uncorrelated to the input $x_{v,\tau}$ as shown in the Fig.7.78. Hence, Eq. (7.6.26) and (7.6.22) can be used to forecast the streamflow of the Conejos River.

In order to forecast the streamflow after June, it is assumed that the SWE does not contribute to the streamflow after June. Hence the SWE of Eq. (7.6.26) is deleted and the parameters re-estimated. The equation for forecasting the streamflow of July, August, September and October can be written as

$$z_{v,\tau} = (\hat{\varphi}_{1,\tau} - 0.287)\, z_{v,\tau-1} + 0.287\, \hat{\varphi}_{1,\tau-1}\, z_{v,\tau-2} \qquad (7.6.21)$$

where $\hat{\varphi}_{1,\tau}$ has the values of

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.557	0.689	0.323	0.205	0.275	0.000
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	1.000	0.675	0.497	0.703	0.653	0.900

In order to compare the forecasting ability of the periodic-TFM method with the multiple regression method, the multiple regression equations in Section 7.5 of the Conejos River basin are re-estimated by using new SWE input data. The multiple regression equations are shown below:

(1) April first forecast

Q2 = 1.967 Q1 + 472.63 WC1 + 364.948

Q3 = 3.336 Q1 + 2674.360 WC1 + 17548.369

Q4 = 4.461 Q1 + 7402.101 WC1 + 8969.909

Q5 = 3.649 Q1 + 10740.964 WC1 + 1553.323

Q6 = 2.831 Q1 + 11649.765 WC1 + 10204.634

Q7 = 2.637 Q1 + 11805.732 WC1 + 16328.286

Q8 = 2.958 Q1 + 11884.324 WC1 + 17616.392

(2) May first forecast

Q3 = 1.127 Q1 + 1.123 Q2 + 2143.700 WC1 + 0. WC2 + 17138.616

Q4 = 1.530 Q1 + 1.666 Q2 + 3087.415 WC1 + 3981.357 WC2 + 18805.765

Q5 = 1.551 Q1 + 1.327 Q2 + 4882.531 WC1 + 5904.384 WC2 + 16557.157

Q6 = 3.238 Q1 + 0. Q2 + 7491.555 WC1 + 4693.325 WC2 + 22515.994

Q7 = 3.107 Q1 + 0. Q2 + 7011.242 WC1 + 5411.486 WC2 + 30523.499

Q8 = 3.460 Q1 + 0. Q2 + 6751.924 WC1 + 5792.881 WC2 + 32812.069

(3) June first forecast

Q4 = 1.853 Q1 +0. Q2 +0.842 Q3 + 3078.4 WC1 + 2337.7 WC2 + 321.436

Q5 = 1.551 Q1 +1.327 Q2 +0. Q3 + 4882.5 WC1 + 5904.4 WC2 + 16557.157

Q6 = 3.238 Q1 +0. Q2 +0. Q3 + 7491.6 WC1 + 4693.3 WC2 + 22515.994

Q7 = 1.497 Q1 +0. Q2 +0.492 Q3 + 5366.9 WC1 + 5782.2 WC2 + 22861.493

Q8 = 1.982 Q1 +0. Q2 +0.452 Q3 + 5241.8 WC1 + 6133.3 WC2 + 25775.437

(4) July first forecast

$$Q5 = -0.727\ Q3 + 1.475\ Q4 + 1715.977\ WC2 + 5032.148$$

$$Q6 = -0.818\ Q3 + 1.554\ Q4 + 2316.507\ WC2 + 7977.331$$

$$Q7 = -0.778\ Q3 + 1.516\ Q4 + 2837.810\ WC2 + 12476.207$$

$$Q8 = -0.787\ Q3 + 1.494\ Q4 + 3124.460\ WC2 + 19625.725$$

(5) August first forecast

$$Q6 = -0.375\ Q2 + 0.\ \ Q3 + 0.\ \ Q4 + 1.070\ Q5 + 347.1\ WC1 + 4978.874$$

$$Q7 = -0.639\ Q2 + 0.238\ Q3 - 0.267\ Q4 + 1.283\ Q5 - 319.4\ WC1 + 8466.812$$

$$Q8 = 0.\ \ Q2 + 0.\ \ Q3 - 0.339\ Q4 + 1.407\ Q5 - 686.7\ WC1 + 11164.030$$

(6) September first prediction

$$Q7 = 1.152\ Q6 - 0.139\ Q5 + 2145.523$$

$$Q8 = 1.195\ Q6 - 0.175\ Q5 + 5879.422$$

(7) October first forecast

$$Q8 = 1.340\ Q7 - 0.337\ Q6 + 3061.268$$

The one step ahead forecast of the streamflow is shown in Fig. 7.79. This shows that both periodic-TFM method and the multiple regression method have improved the forecast of the high flow in the summer time compared with Fig. 7.75.

Then, by standing at the end of March, April,..., September , the forecasts of the streamflow and the cumulative streamflow with lead time L=1,2,... are shown in Figs. 7.80-7.93. The root mean square deviation and the maximum absolute deviation of both models are shown in Table 7.3. From Figs. 7.79-7.93 and Table 7.3, we can see that both the periodic-TFM method and the multiple regression method give a good forecast of the Conejos River streamflow, but the periodic-TFM method gives better results.

7.6.3 The Rio Grande River Basin

Using the same method shown in the previous sections, the deseasonalized SWE of January, February, March and April can be fitted by

$$x_t = 0.906\, x_{t-1} + \varepsilon_t \tag{7.6.22}$$

After the deseasonalized streamflow was filtered by the PAR(1) model, the impulse response function of the Rio Grande River basin is shown in Fig. 7.94. This figure suggests an input-output model given as

$$y_{v,\tau} = -0.185\, x_{v,\tau} + 0.389\, x_{v,\tau-1} + N_{v,\tau} \tag{7.6.23}$$

The noise $N_{v,\tau}$, estimated from Eq. (7.6.23), id fitted by

$$N_t = -0.189\, N_{t-1} + \varepsilon_t \tag{7.6.24}$$

with $\varepsilon_t \sim WN(0,0.402)$. Substituting (7.6.23) and (7.6.24) into (7.6.12) gives

$$z_{v,\tau} = (\hat{\varphi}_{1,\tau} - 0.189)z_{v,\tau-1} + 0.189\, \hat{\varphi}_{1,\tau-1}\, z_{v,\tau-2} - 0.185\, x_{v,\tau} + 0.354\, x_{v,\tau-1} + 0.07\, x_{v,\tau-2} + \varepsilon_{v,\tau} \tag{7.6.25}$$

After estimating the parameters of Eq. (7.6.25) using the non-linear least squares method, the model can be written as

$$z_{v,\tau} = (\hat{\varphi}_{1,\tau} - 0.246)z_{v,\tau-1} + 0.246\, \hat{\varphi}_{1,\tau-1}\, z_{v,\tau-2} - 0.276\, x_{v,\tau} + 0.384\, x_{v,\tau-1} + 0.323\, x_{v,\tau-2} + \varepsilon^*_{v,\tau} \tag{7.6.26}$$

with $\varepsilon^*_{v,\tau} \sim WN(0,0.278)$ and $\hat{\phi}_{1,\tau}$ has values of

τ	1	2	3	4	5	6
$\hat{\phi}_{1,\tau}$	0.800	0.883	1.000	0.468	0.494	0.344
τ	7	8	9	10	11	12
$\hat{\phi}_{1,\tau}$	0.843	0.973	0.553	0.756	0.970	0.933

In order to forecast the streamflow after June, the SWE in Eq. (7.6.26) is deleted to obtain a forecast equation for the months of July, August, September and October. After re-estimating the parameters, the model can be written as

$$z_{v,\tau} = \left(\hat{\varphi}_{1,\tau} - 0.173\right) z_{v,\tau-1} + 0.173\, \hat{\varphi}_{1,\tau-1}\, z_{v,\tau-2} + \varepsilon_{v,\tau} \qquad (7.6.27)$$

where $\hat{\varphi}_{1,\tau}$ has the values of

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.800	0.883	0.566	0.433	0.687	0.667
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	1.000	0.931	0.559	0.756	0.922	0.908

The one step ahead forecast of the streamflow of the Rio Grande River near Del Norte is shown in Fig. 7.95. The forecast of streamflow and cumulative streamflow with lead time L=1,2,... are shown in Figs. 7.96-7.109. The root mean square deviation and the maximum absolute deviation are listed in Table 7.4. From Figs. 7.95-7.109 and Table 7.4, we can see that both the periodic-TFM method and the multiple regression method give a good forecast of the streamflow, but the periodic-TFM method gives a better forecast results for the one step ahead forecast, and a better forecast for lead time L=1,2,....

7.6.4 Summary of the Forecasting Results

(1) The forecasted monthly streamflow of the Los Pinos River near Ortiz shows that the periodic-TFM method gives a much better forecast than the multiple regression method.

(2) The forecast of the monthly streamflow of the Conejos River near Mogote using

the average SWE of the Platoro and River Springs under-estimates the high flow in the summer time.

(3) The forecasts of the Conejos River streamflow where the information of the SWE at Cumbres Pass was added to the Conejos River basin gives a much better result. The periodic-TFM method still shows a better forecasting ability than the multiple regression approach.

(4) The Rio Grande River near Del Norte has more SWE input information than the other basins. The forecasted results show that both the periodic-TFM method and the multiple regression method give a good forecast, but the periodic-TFM method gives slightly better results.

7.7 Forecasting with the Current Data Available

A comparison of the model presented in section 6-6 and the multiple regression model was made using data from 1948-1977 to forecast streamflows for the years 1978-1982. The years 1948-1977 were used because the Colorado State Engineer's Office has already analyzed the data from this time period.

Since the SWE is available from 1949 to present, the data from 1949-1983 will be analyzed, and a validation done for the years of 1984-1988. No detail of the analysis will be presented, only the final equation and some graphs will be given.

7.7.1 The Los Pinos River Basin

After analyzing the data of the Los Pinos River basin, the periodic -TFM method gives

$$z_{v,\tau}=\varphi_{1,\tau}\, z_{v,\tau-1}-0.208\, x_{v,\tau}+0.195\, x_{v,\tau-1}+0.332\, x_{v,\tau-2}+\varepsilon_{v,\tau} \qquad (7.7.1)$$

with $\varepsilon^*_{v,\tau} \sim WN(0,0.333)$ and $\hat{\varphi}_{1,\tau}$ values of

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.838	0.745	0.229	0.252	0.123	0.483
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.765	0.561	0.491	0.606	0.910	0.777

The equation for the streamflow after June can be written as

$$z_{v,\tau} = \hat{\varphi}_{1,\tau}\, z_{v,\tau-1} \tag{7.7.2}$$

where $\hat{\varphi}_{1,\tau}$ has values of

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.838	0.745	0.229	0.257	0.126	0.765
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.946	0.561	0.491	0.626	0.911	0.750

The one step ahead forecast of the streamflow is shown in the Fig. 7.110. The forecasts by standing in the March with lead time L=1,2,...,7 is shown in Fig. 7.111. The result shows that the high flow of summer time in the 1985 and 1986 were under-estimated.

7.7.2 The Conejos River Basin

Using the new SWE of Section 7.6.2 as the input, the data of the year 1949-1983 were analyzed. They give the model

$$\begin{aligned} z_{v,\tau} = {} & (\hat{\varphi}_{1,\tau} - 0.238)\, z_{v,\tau-1} + 0.238\, \hat{\varphi}_{1,\tau-1}\, z_{v,\tau-2} + 0.003\, x_{v,\tau-1} \\ & + 0.258\, x_{v,\tau-2} + 0.324\, x_{v,\tau-3} + 0.075\, x_{v,\tau-4} + 0.202\, x_{v,\tau-5} \\ & + 0.131\, x_{v,\tau-6} + 0.050\, x_{v,\tau-7} + \varepsilon^*_{v,\tau} \end{aligned} \tag{7.7.3}$$

with $\varepsilon^*_{v,\tau} \sim WN(0,0.453)$ and $\hat{\varphi}_{1,\tau}$ has the values of

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.636	0.519	0.000	0.420	0.248	0.264
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.541	0.545	0.469	0.716	0.559	0.746

The equation for forecasting the streamflow of July, August, September and October can be written as

$$z_{v,\tau} = (\hat{\varphi}_{1,\tau} - 0.281)\, z_{v,\tau-1} + 0.281\, \hat{\varphi}_{1,\tau-1}\, z_{v,\tau-2} \qquad (7.7.4)$$

where $\hat{\varphi}_{1,\tau}$ has the values of

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.636	0.519	0.230	0.202	0.304	0.304
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	1.000	0.684	0.487	0.744	0.610	0.782

The one step ahead forecast of the streamflow is shown in Fig.7.112. The forecast by standing in the March with lead time L=1,2,...,7 is shown in Fig. 7.113. The result shows that the high flow of summer time in the 1985 and 1986 are under-estimated.

7.7.3 The Rio Grande River Basin

The data of the years 1949-1983 was analyzed, which give an input-output model as

$$z_{v,\tau} = (\hat{\varphi}_{1,\tau} - 0.100) z_{v,\tau-1} + 0.100\, \hat{\varphi}_{1,\tau-1}\, z_{v,\tau-2} - 0.228\, x_{v,\tau}$$
$$+ 0.270\, x_{v,\tau-1} + 0.229\, x_{v,\tau-2} + 0.268\, x_{v,\tau-3} + \varepsilon^*_{v,\tau} \qquad (7.7.5)$$

with $\varepsilon^*_{v,\tau} \sim WN(0,0.289)$ and $\hat{\varphi}_{1,\tau}$ has values of

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.832	0.887	1.000	0.388	0.305	0.167
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.554	0.700	0.566	0.711	0.868	0.900

The equation for the streamflow after June can be written as

$$z_{v,\tau} = (\hat{\varphi}_{1,\tau} - 0.173)z_{v,\tau-1} + 0.173\,\hat{\varphi}_{1,\tau-1}\,z_{v,\tau-2} + \varepsilon^{*}_{v,\tau} \qquad (7.7.5)$$

where $\hat{\varphi}_{1,\tau}$ has the values of

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.832	0.887	0.545	0.423	0.481	1.000
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.921	0.874	0.548	0.739	0.854	0.865

The one step ahead forecast of the streamflow is shown in the Fig. 7.114, and the forecasting by standing at the end of March with lead time L=1,2,...,7 is shown in Fig. 7.115. The results show that the high flow in summer time of 1985 and 1986 were also underestimated.

7.7.4 Summary of the Forecasting Results

The forecast results of the monthly streamflow (1984-1988) show that the streamflow were under-estimated in the years of 1985 and 1986. Since all three basins have the same problem of underestimating the flow in the year of 1985 and 1986, we will assume that there must be some other input factor besides the SWE existing in the river basin without being considered. In order to solve this problem, the additional input factors will be put into the multiple input-multiple output model of the next chapter.

Table 7.1 Comparison of one step ahead forecast (Los Pinos River near Ortiz, 1978-1982)

Model		RMSD	MAD
Deseasonalized ARMA	[see Fig. 7.7]	8870	27988
PAR(1)	[see Fig. 7.20]	7504	23316
PARMA(1,1)	[see Fig. 7.25]	8577	29394
Log-TFM	[see Fig. 7.44]	6714	22323
No Log-TFM	[see Fig. 7.50]	6097	21681
TFM-PAR(1) Noise	[see Fig. 7.52]	6573	24286
Periodic TFM	[see Fig. 7.58]	4918	13801
Mul-Regression	[see Fig. 7.59]	7021	78732

Table 7.2 The Comparison of the forecast ability (Los Pinos River near Ortiz, 1978-1982)

	Root Mean Square Deviation		Maximum Absolute Deviation	
	Periodic-TFM	Multi-Regres	Periodic-TFM	Multi-Regres
Apr	6036	7838	15636	28652
May	4918	7506	13801	27345
June	2336	3760	4493	7507
July	1510	1790	4105	3709
Aug	1121	1219	1438	1143
Sep	1179	1408	444	499
Oct	670	1187	263	353

Table 7.3 The Comparison of the forecast ability (Conejos River near Mogote, 1978-1982)

	Root Mean Square Deviation		Maximum Absolute Deviation	
	Periodic-TFM	Multi-Regres	Periodic-TFM	Multi-Regres
Apr	55906	61844	21055	21635
May	56048	64419	23135	23867
June	50064	81071	21178	43920
July	35226	49492	19162	21566
Aug	27062	32800	15093	20598
Sep	23141	26627	14922	15280
Oct	16956	23345	11731	12610

Table 7.4 The Comparison of the forecast ability (Rio Grande River near Del Norte, 1978-1982)

	Root Mean Square Deviation		Maximum Absolute Deviation	
	Periodic-TFM	Multi-Regres	Periodic-TFM	Multi-Regres
Apr	46782	48112	61786	63098
May	47578	53881	48572	74751
June	53110	39761	67243	49684
July	44282	44908	44396	48988
Aug	36902	37722	45643	45557
Sep	29311	29279	41452	39801
Oct	15798	17838	17262	35629

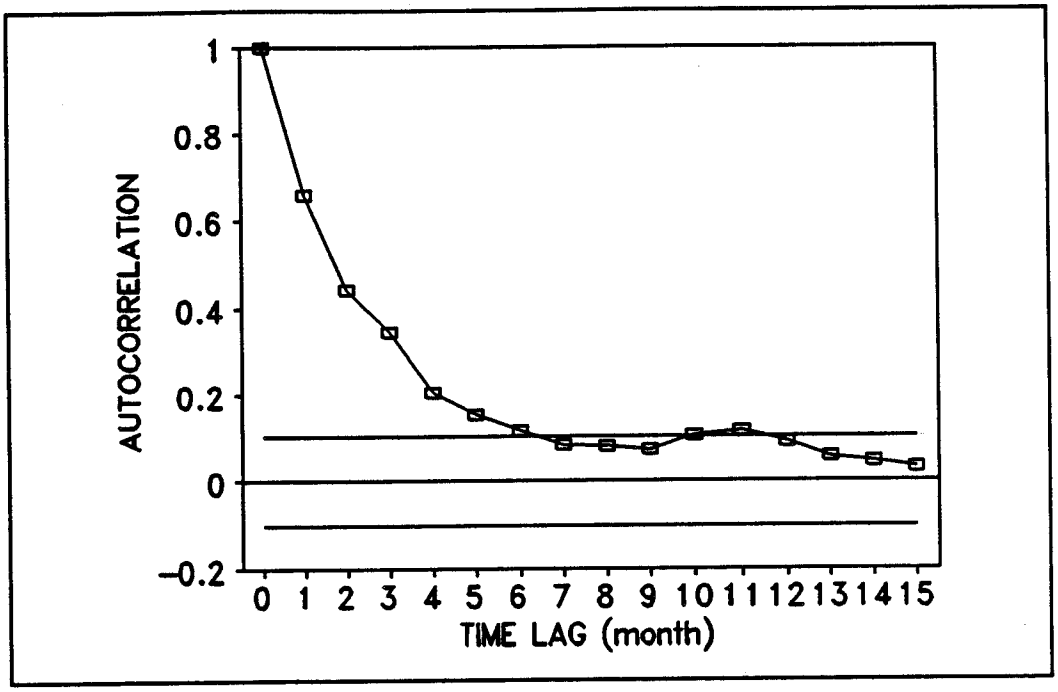


Fig. 7.1 The autocorrelation function (Los Pinos River, 1948-1977)

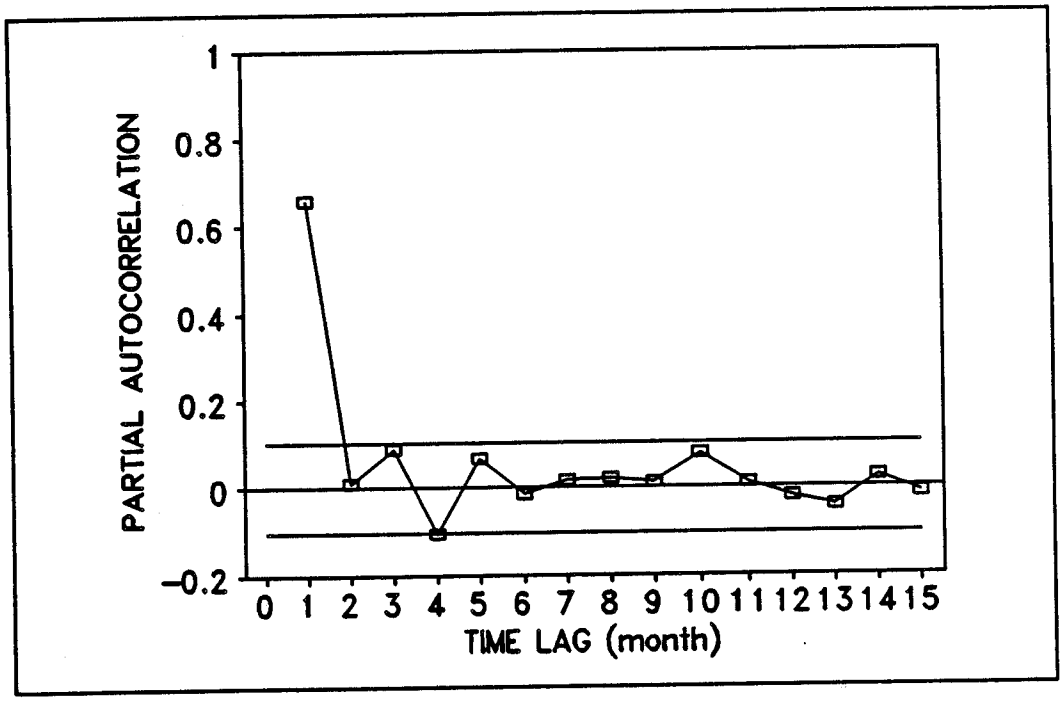


Fig. 7.2 The partial autocorrelation function (Los Pinos River, 1948-1977)

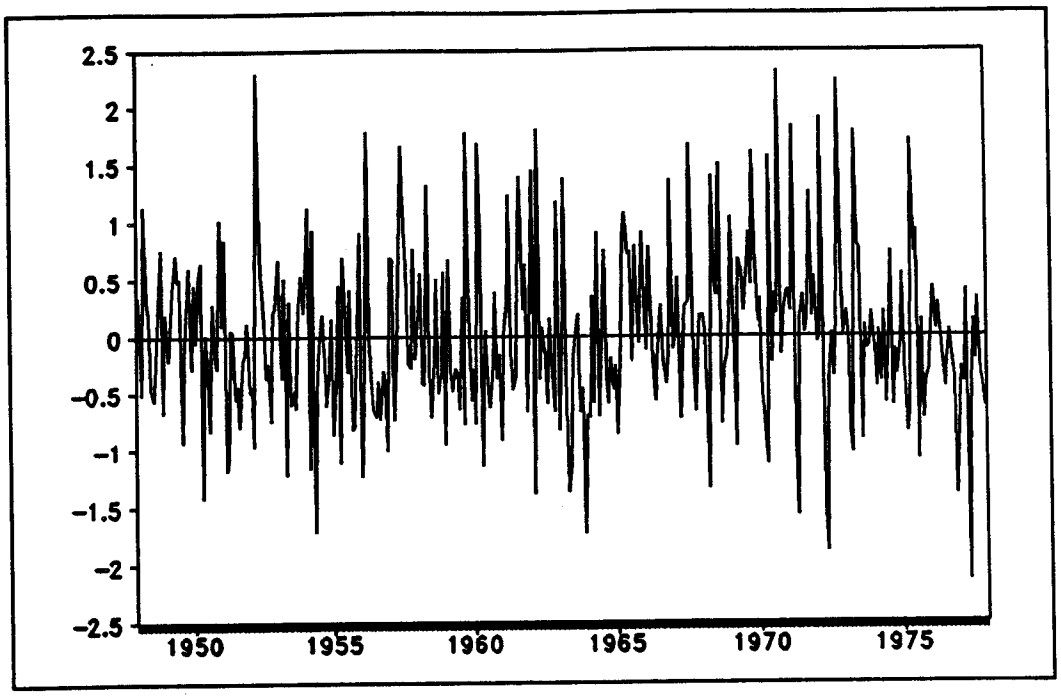


Fig. 7.3 The residual of the AR(4) model (Los Pinos River near Ortiz)

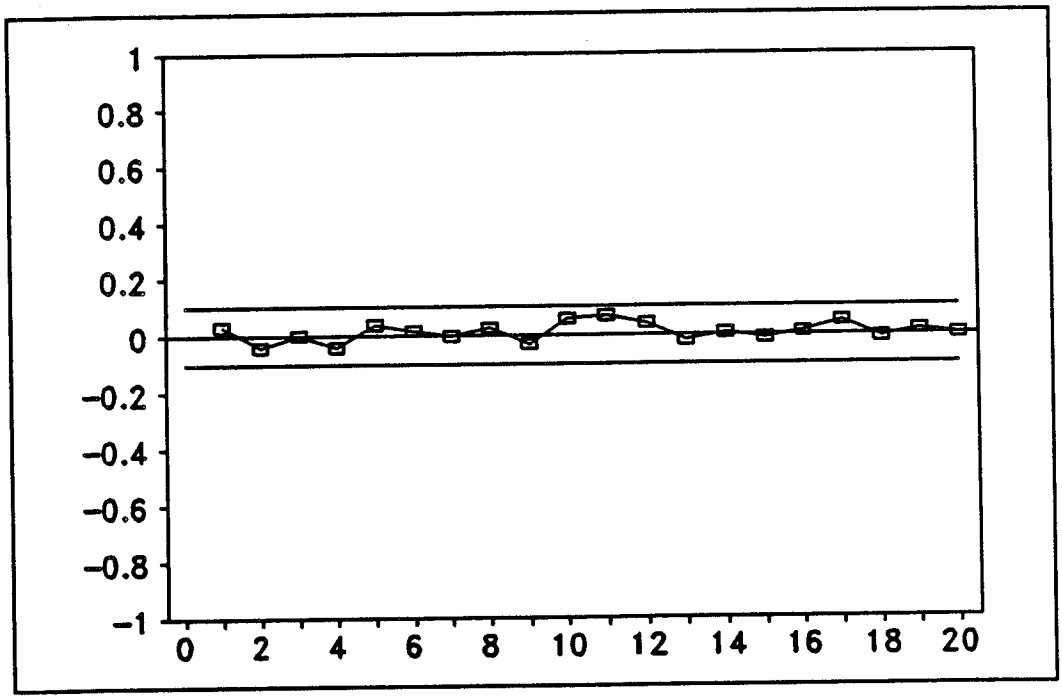


Fig. 7.4 The autocorrelation function of residual (Los Pinos River, deseasonalized AR(4) model)

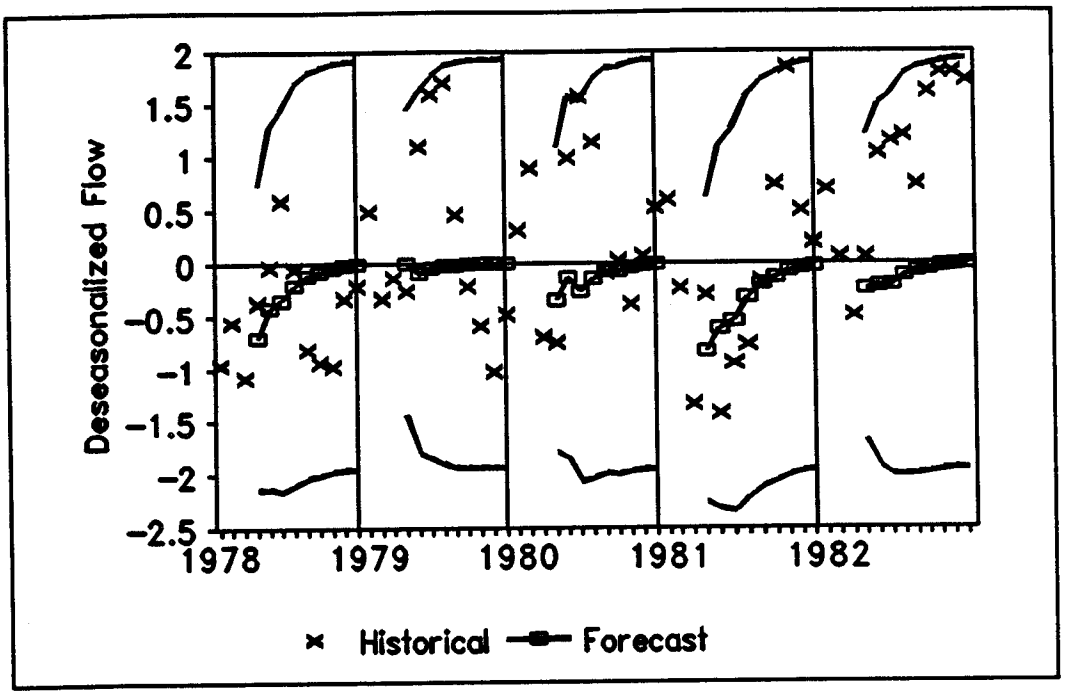


Fig. 7.5 The forecast of the deseasonalized flow (Los Pinos River, 1948-1977, AR(4) model)

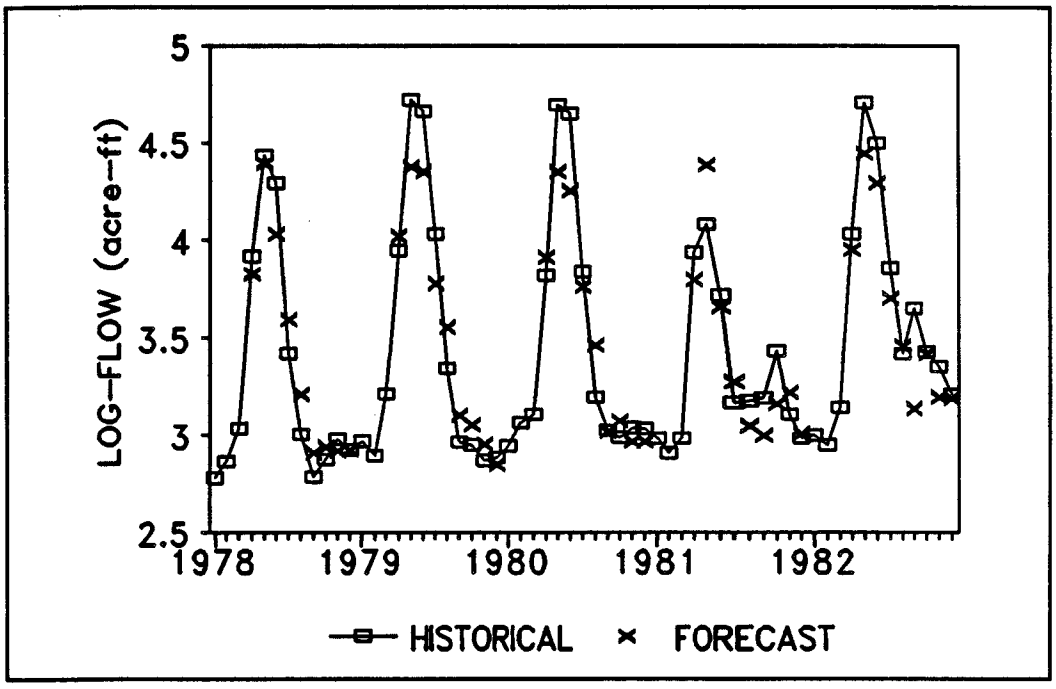


Fig. 7.6 The one step ahead forecast of log-flow (Los Pinos River, deseasonalized ARMA model)

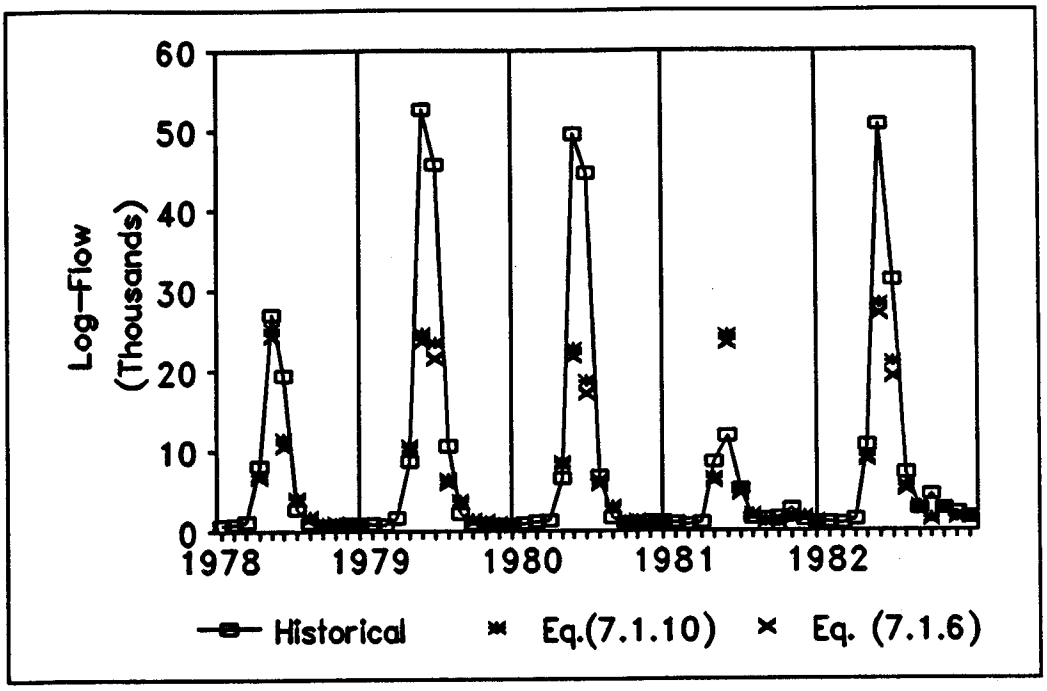


Fig. 7.7 The one step ahead forecast of monthly flow (Los Pinos River, 1978-1982, AR(4) model)

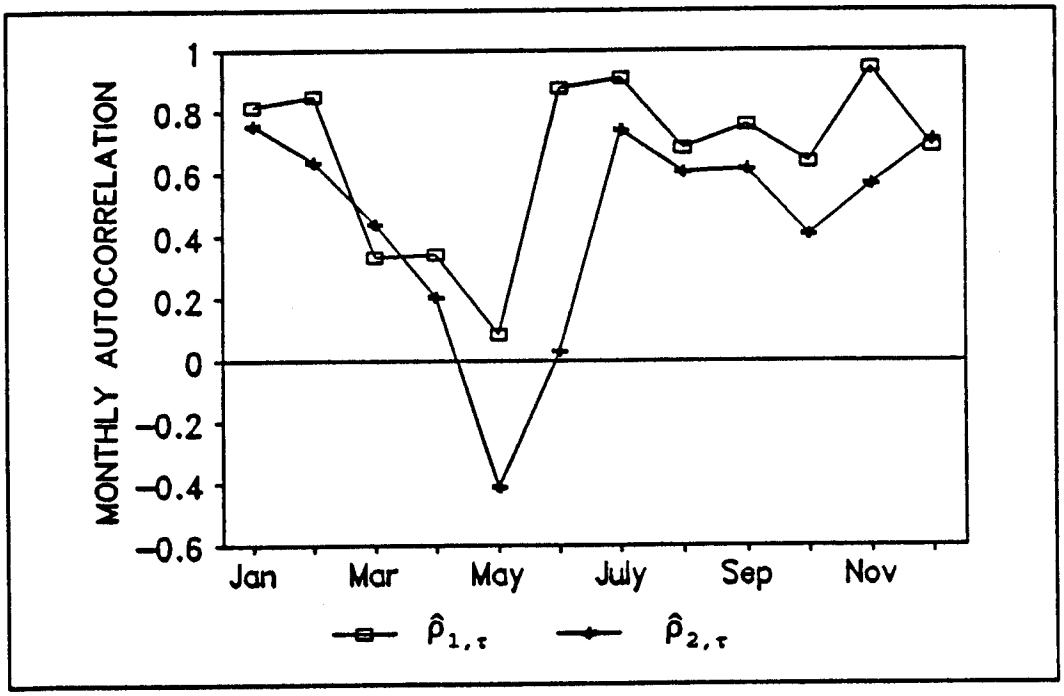


Fig. 7.8 The monthly autocorrelation of the deseasonalized flow (Los Pinos River Basin, 1948-1977)

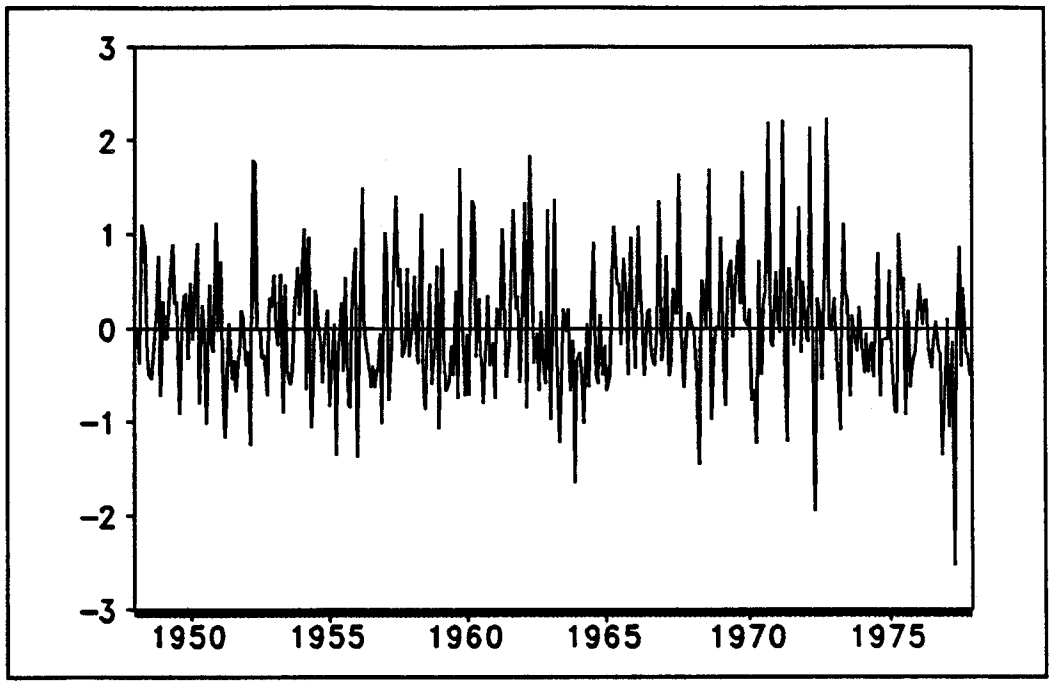


Fig. 7.9 The residual of the par(1) model (Los Pinos River Basin, 1948-1977)

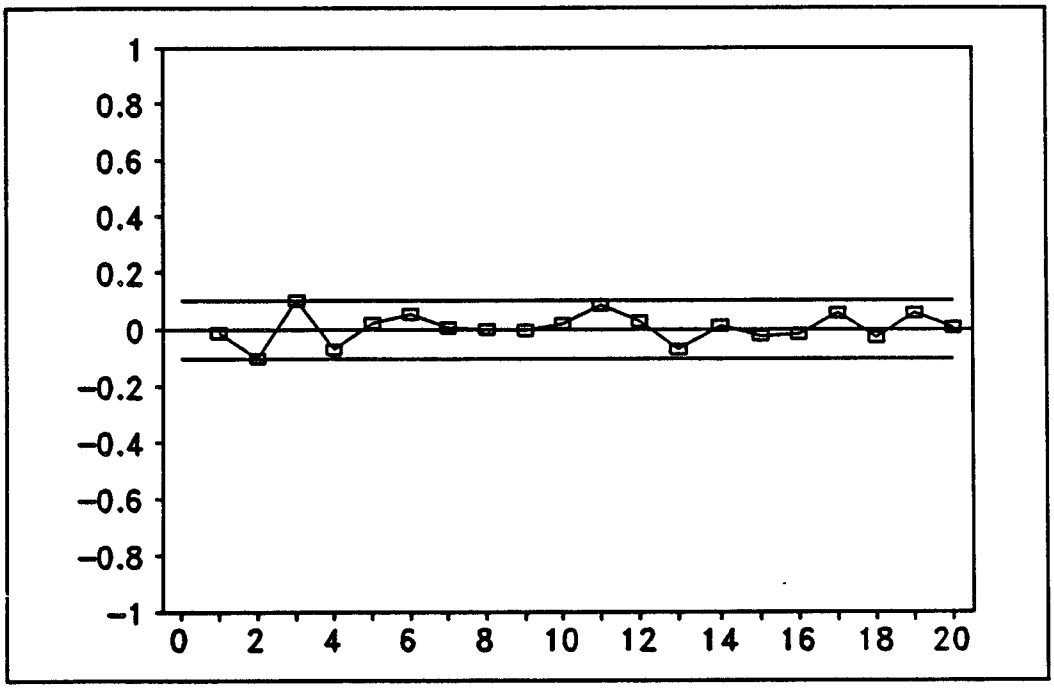


Fig. 7.10 The autocorrelation of the residual (Los Pinos River Basin, 1948-1977; PAR(1) model)

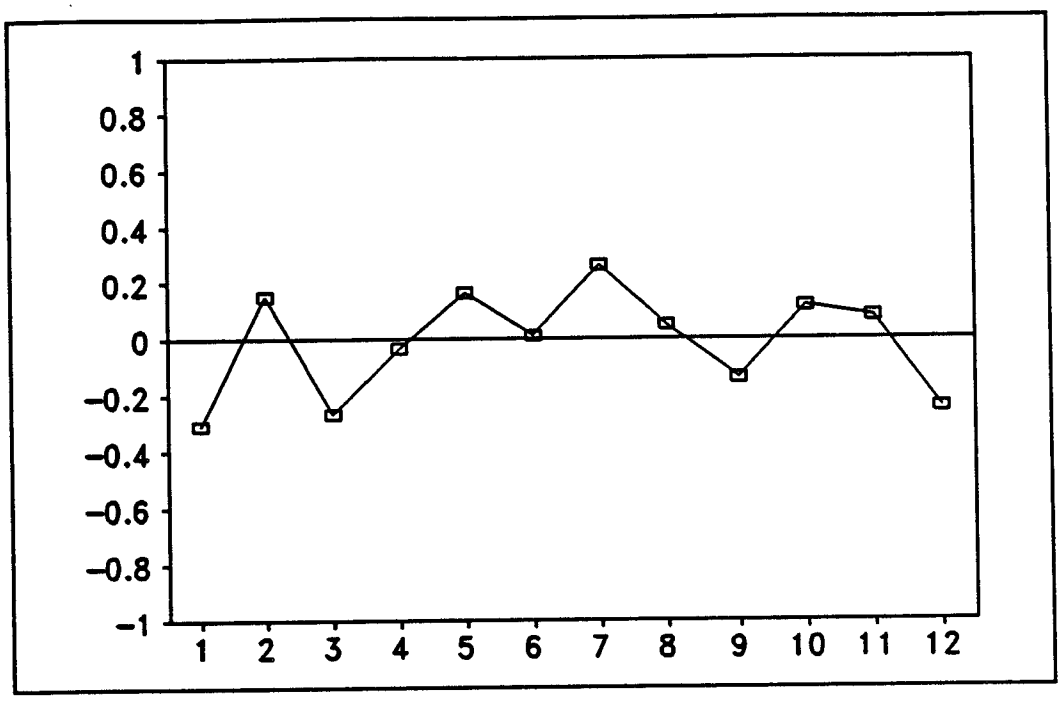


Fig. 7.11 The monthly correlation of the residual (Los Pinos River Basin, 1948-1977; PAR(1) model)

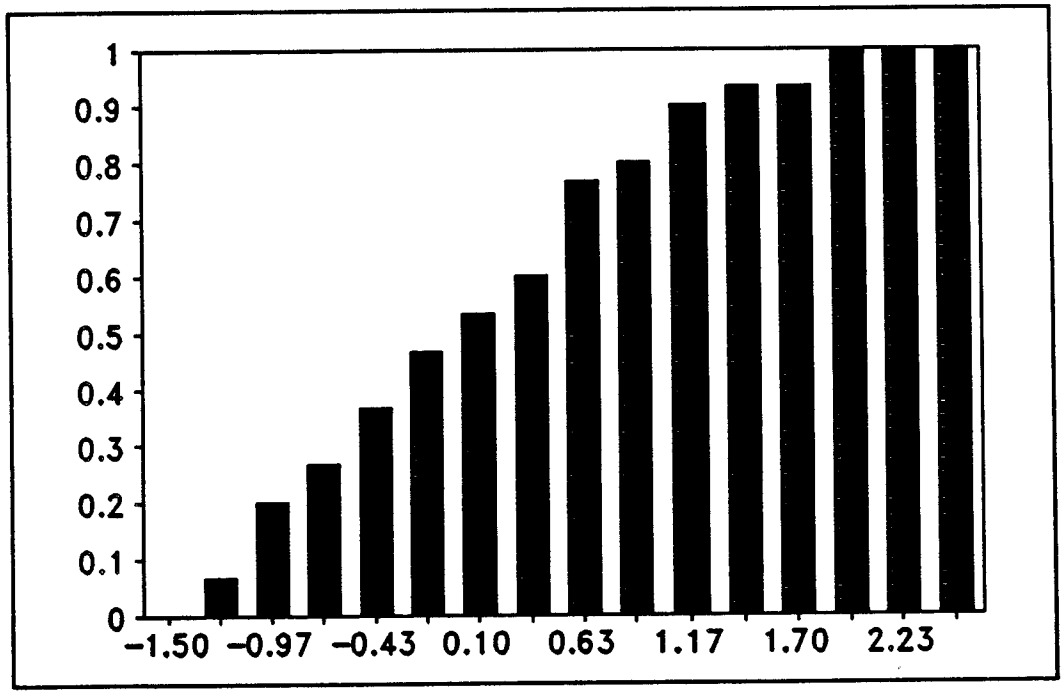


Fig. 7.12 The empirical CDF of the forecast error, L=1 (Los Pinos River Basin, 1948-1977; PAR(1) model)

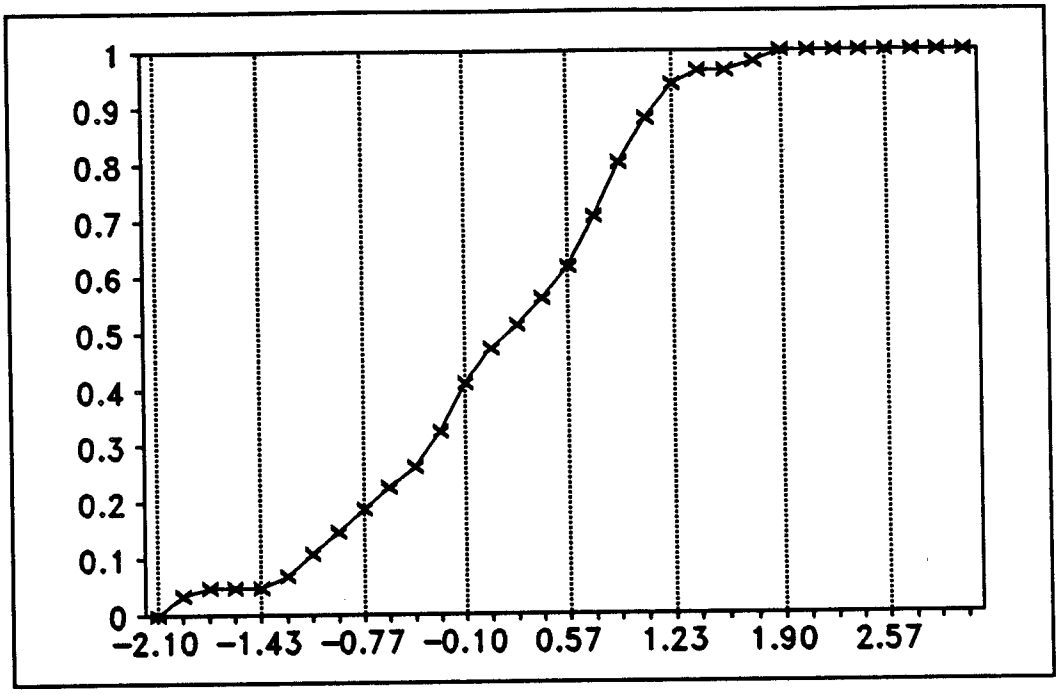


Fig. 7.13 The simulated CDF of the forecast error, L=2 (Los Pinos River; par(1) model)

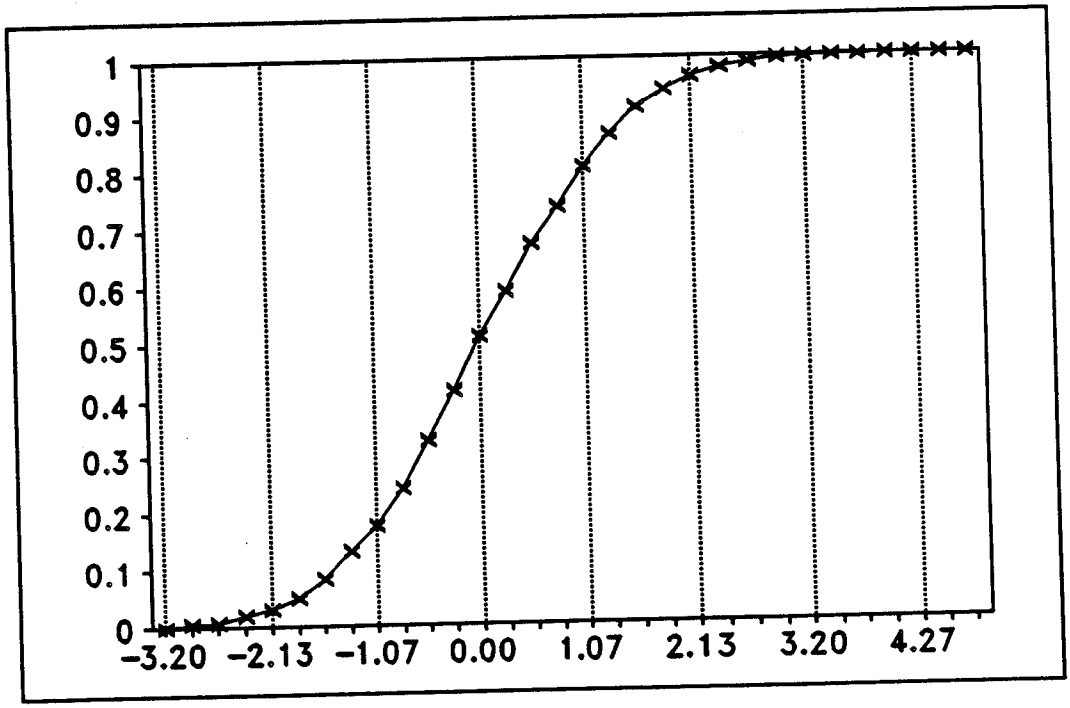


Fig. 7.14 The simulated CDF of the forecast error, L=3 (Los Pinos River, PAR(1) model)

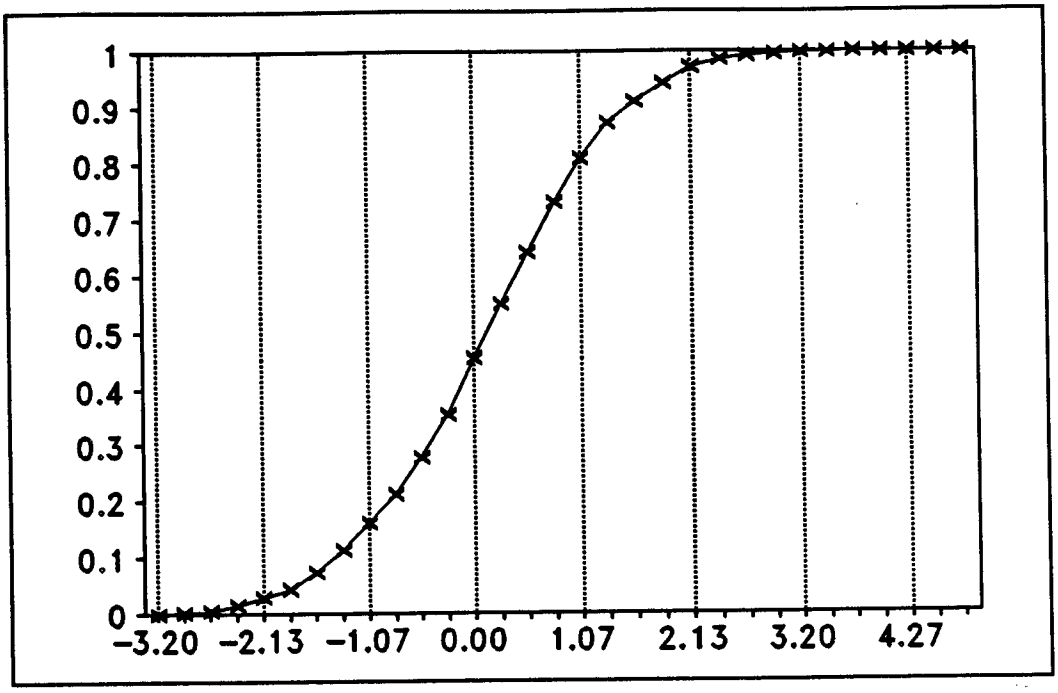


Fig. 7.15 The simulated CDF of forecast error, L=4 (Los Pinos River; PAR(1) model)

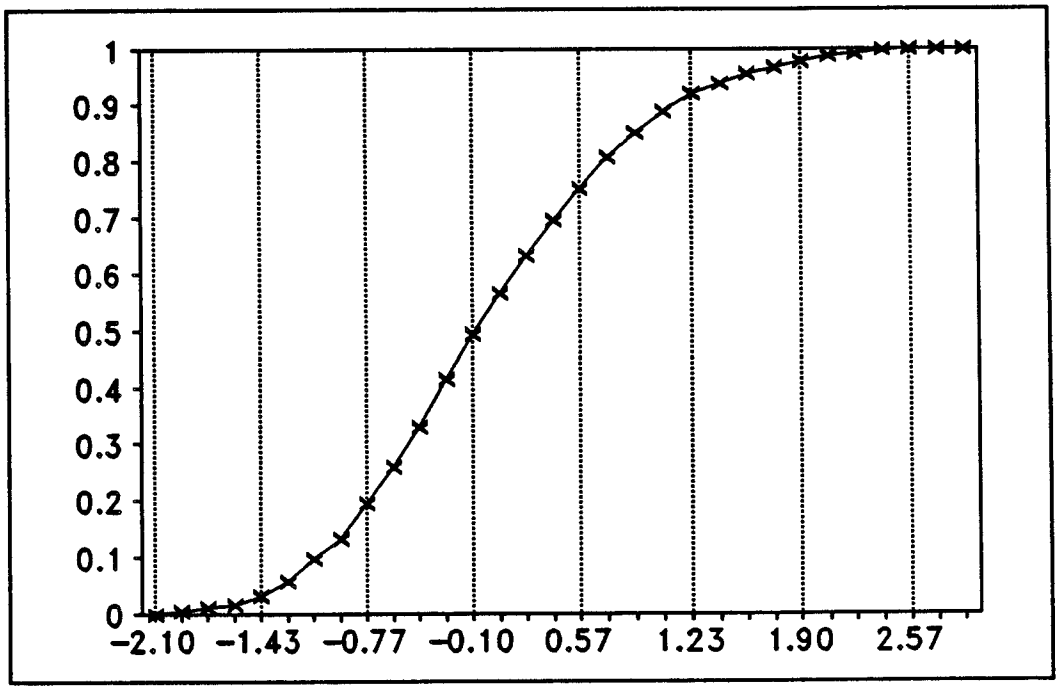


Fig. 7.16 The simulated CDF of forecast error, L=5 (Los Pinos River; PAR(1) model)

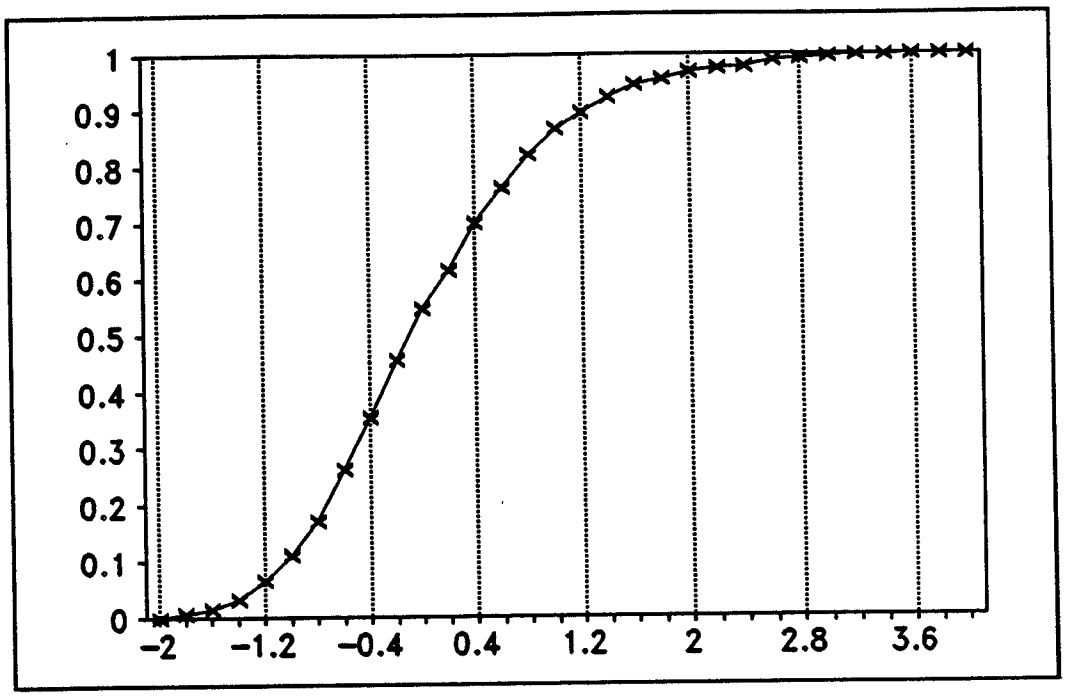


Fig. 7.17 The simulated CDF of forecast error, L=6 (Los Pinos River; PAR(1) model)

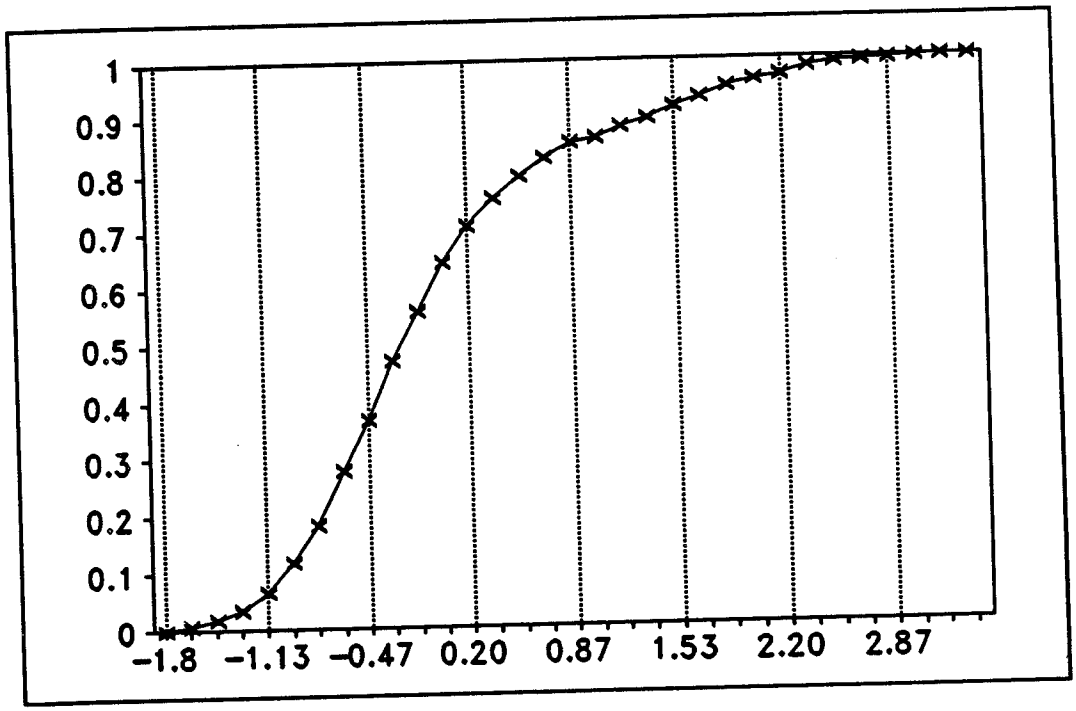


Fig. 7.18 The simulated CDF of forecast error, L=7 (Los Pinos River, PAR(1) model)

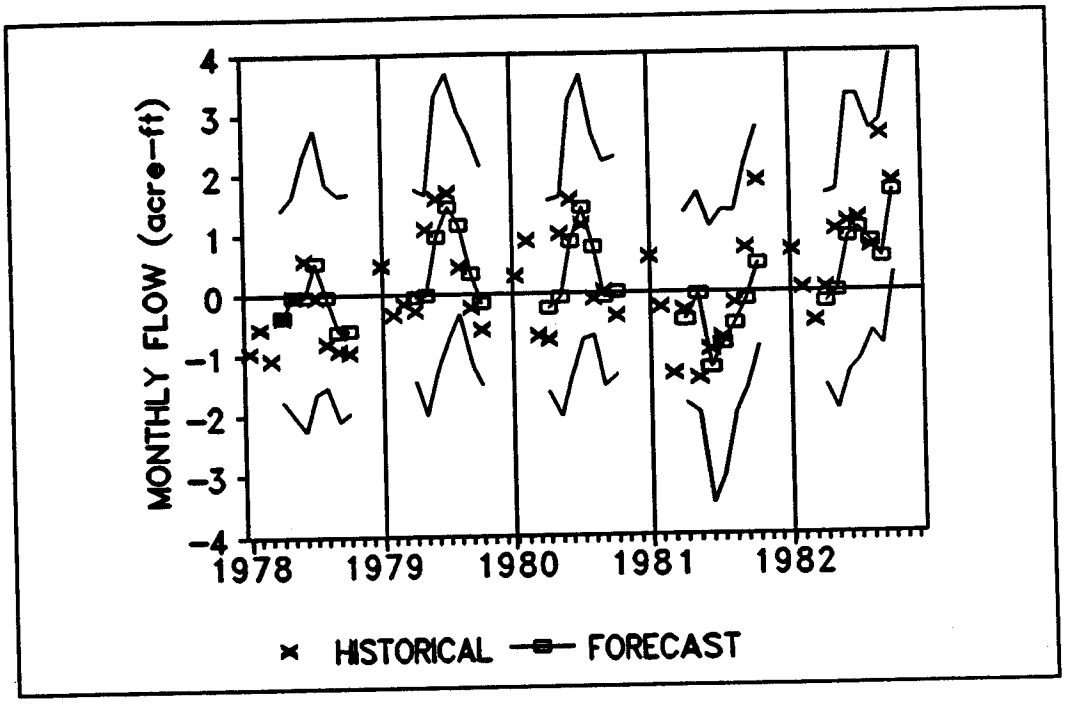


Fig. 7.19 The forecast of deseasonalized flow, L=1,...,9 (Los Pinos River, PAR(1) model)

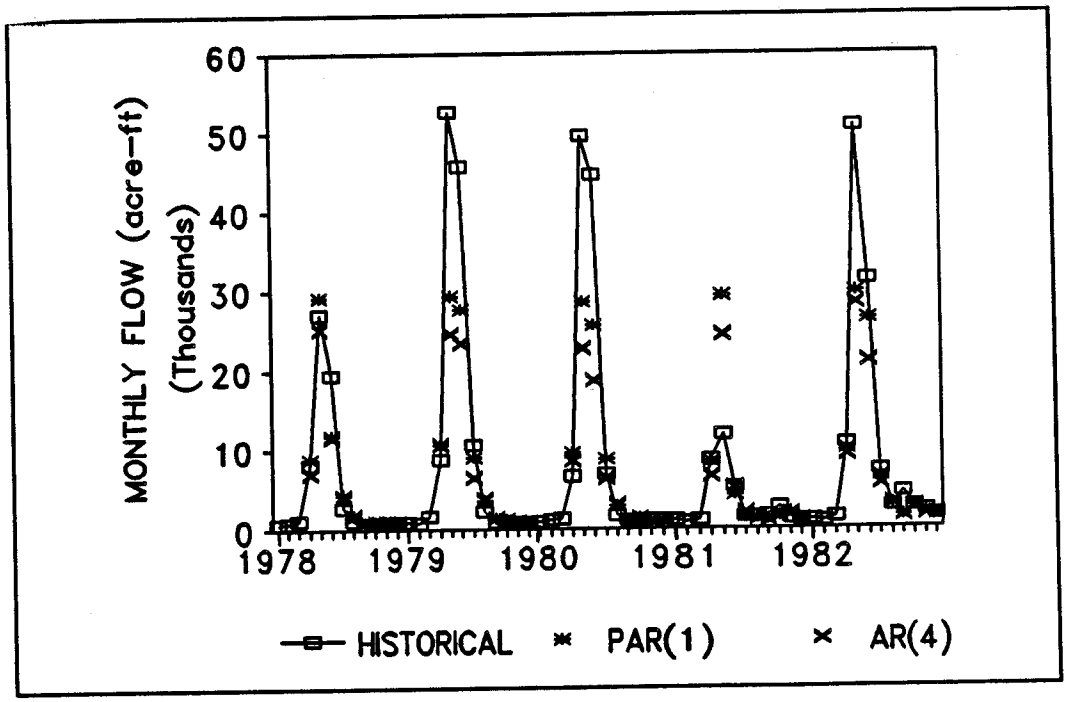


Fig. 7.20 The one step ahead forecast of the monthly flow (Los Pinos River Basin; PAR(1) model)

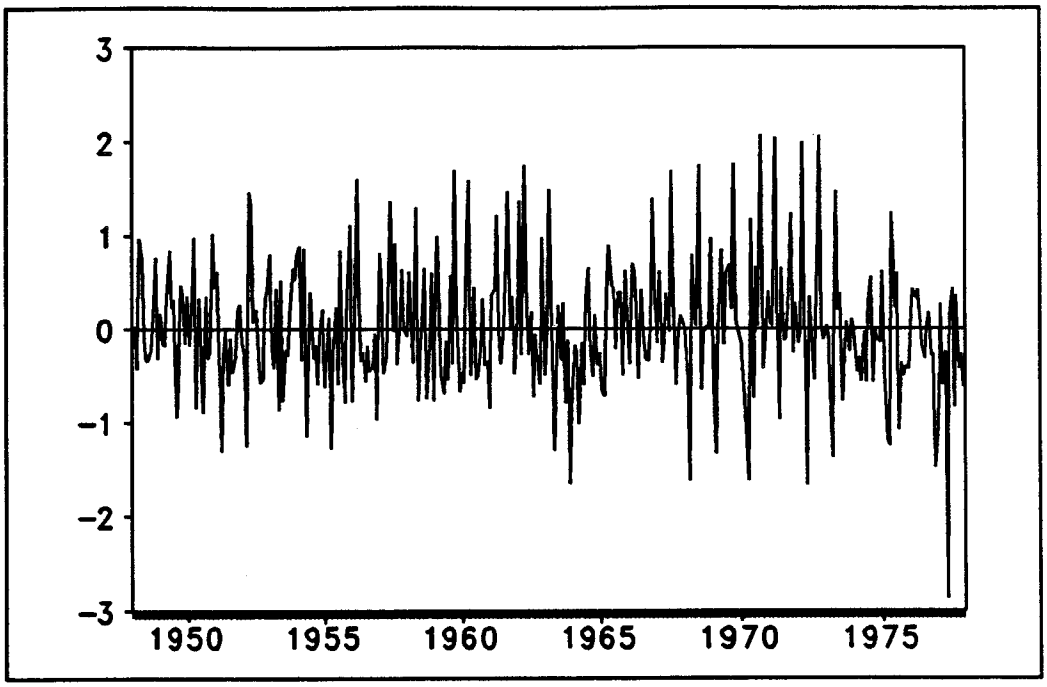


Fig. 7.21 The residual of the PARMA(1,1) model (Los Pinos River Basin, 1948-1977)

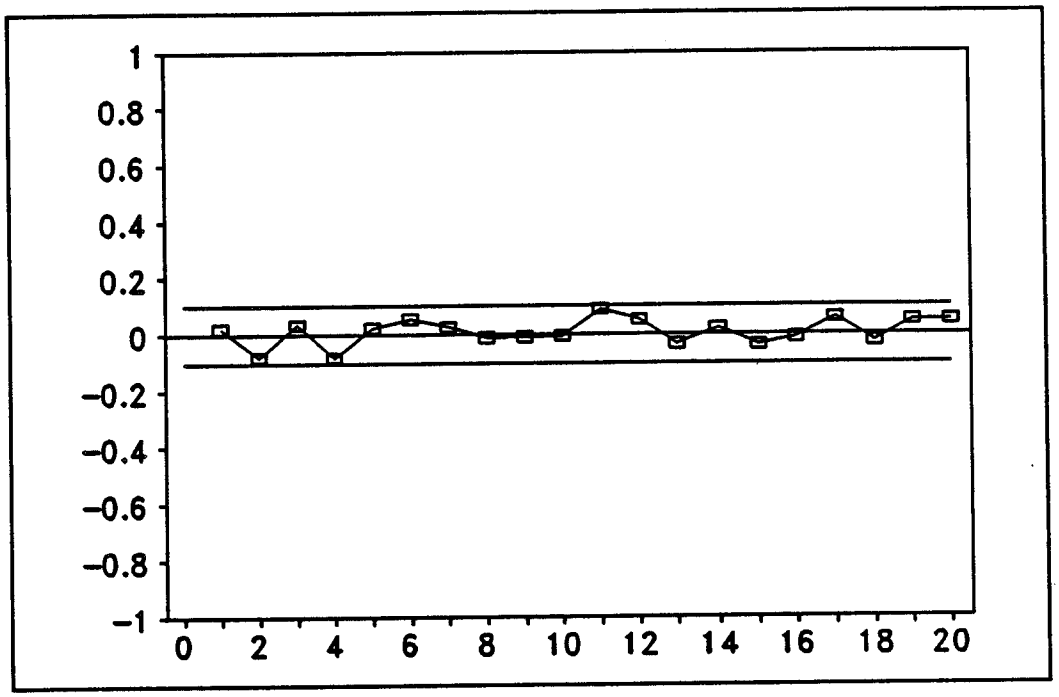


Fig. 7.22 The autocorrelation of the residual (Los Pinos River Basin; PARMA(1,1) model)

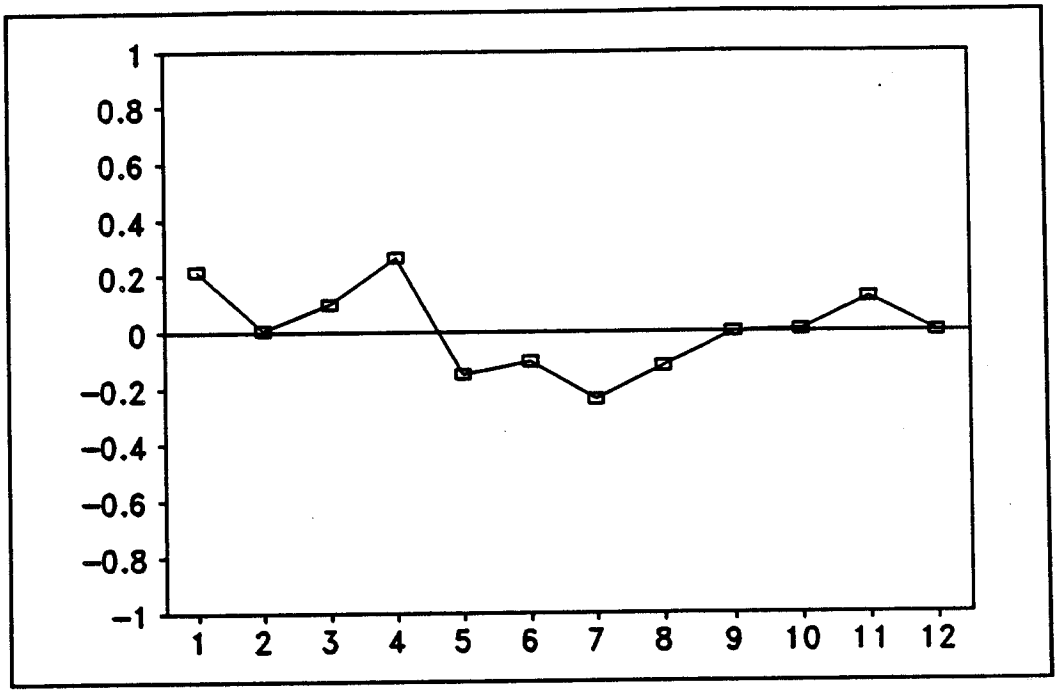


Fig. 7.23 The monthly correlation of the residual (Los Pinos River Basin; PARMA(1,1) model)

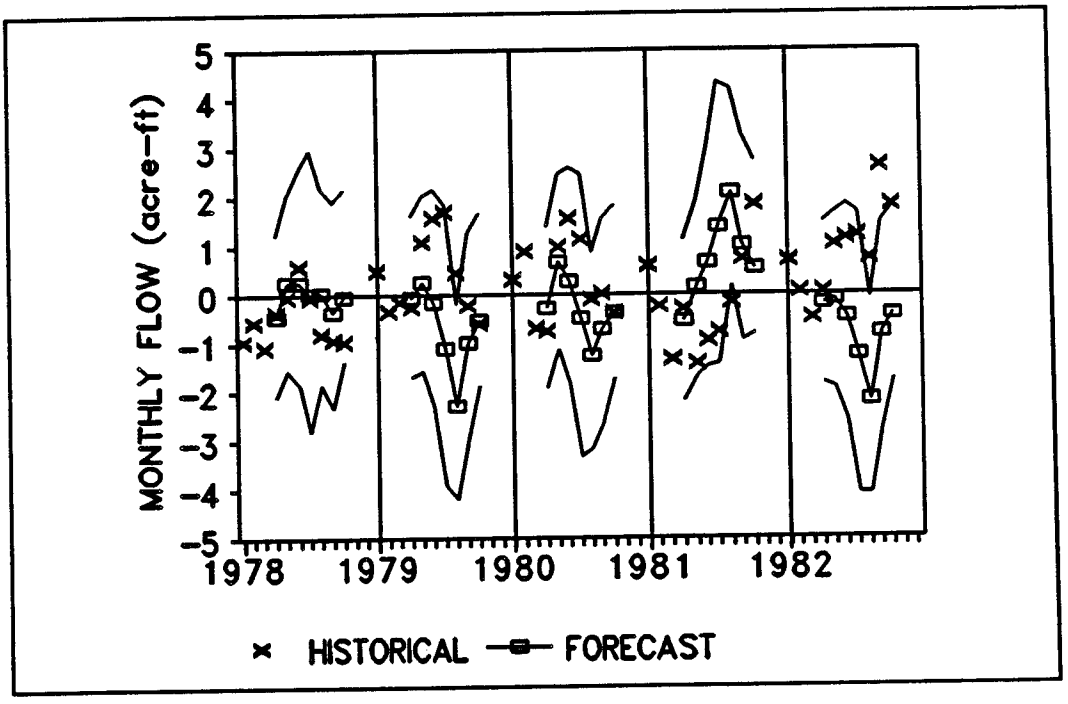


Fig. 7.24 The forecast of deseasonalized flow, L=1,...,9 (Los Pinos River Basin; PARMA(1,1) model)

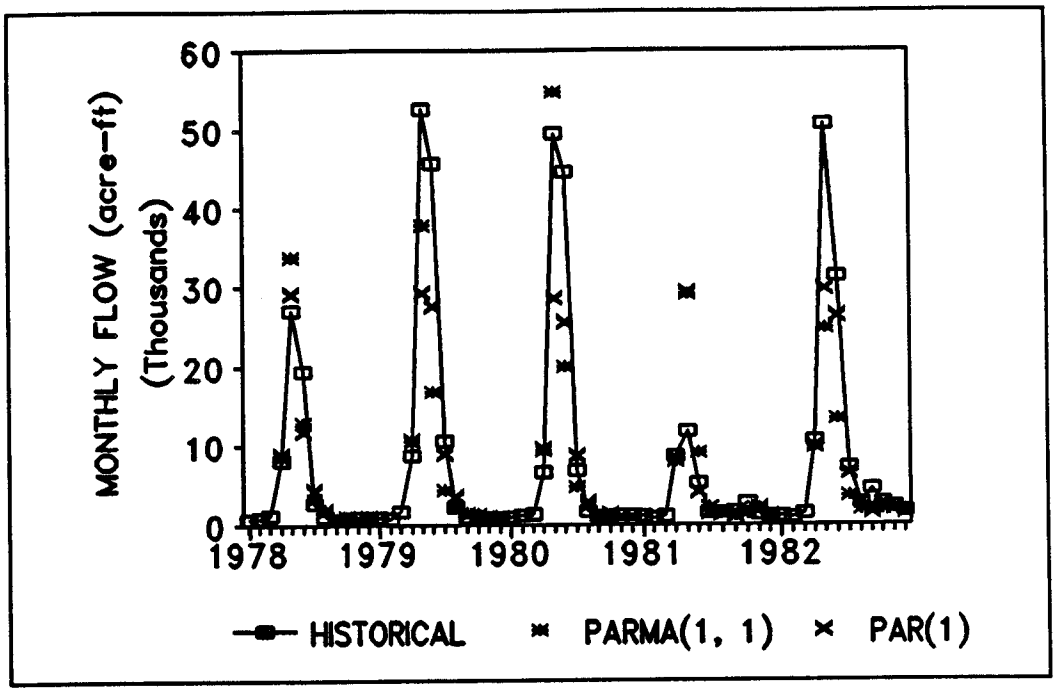


Fig. 7.25 One step ahead forecast of the monthly flow (Los Pinos River; PARMA(1,1) model)

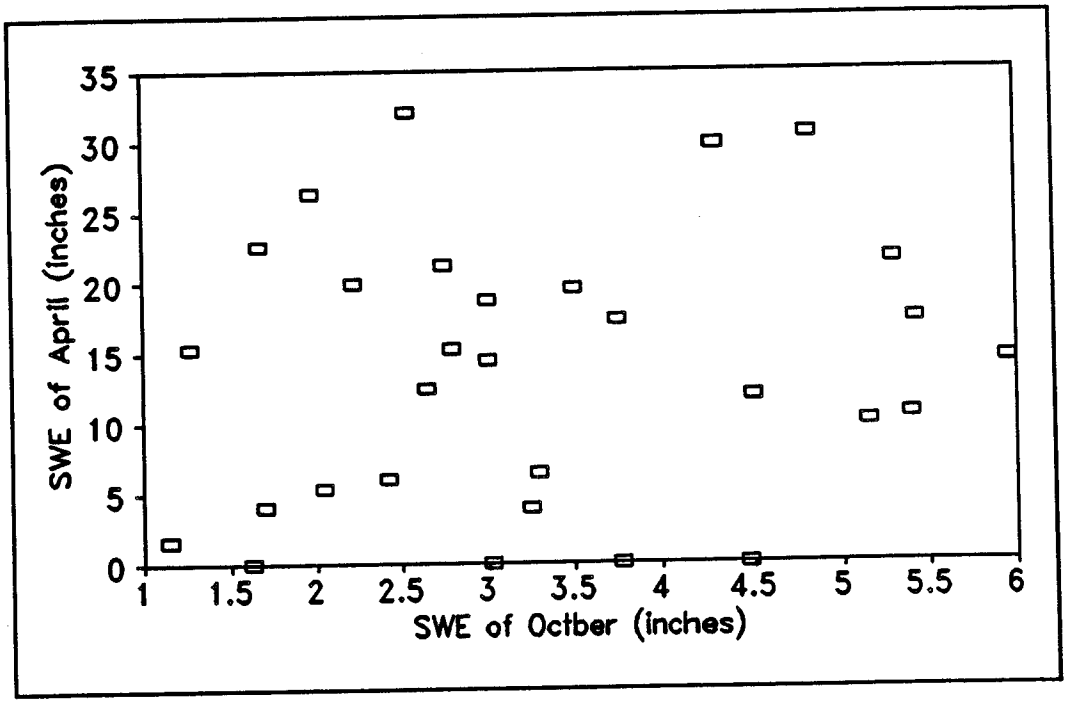


Fig. 7.26 The SWE of October vs. the SWE of April (Los Pinos River Basin, 1948-1977)

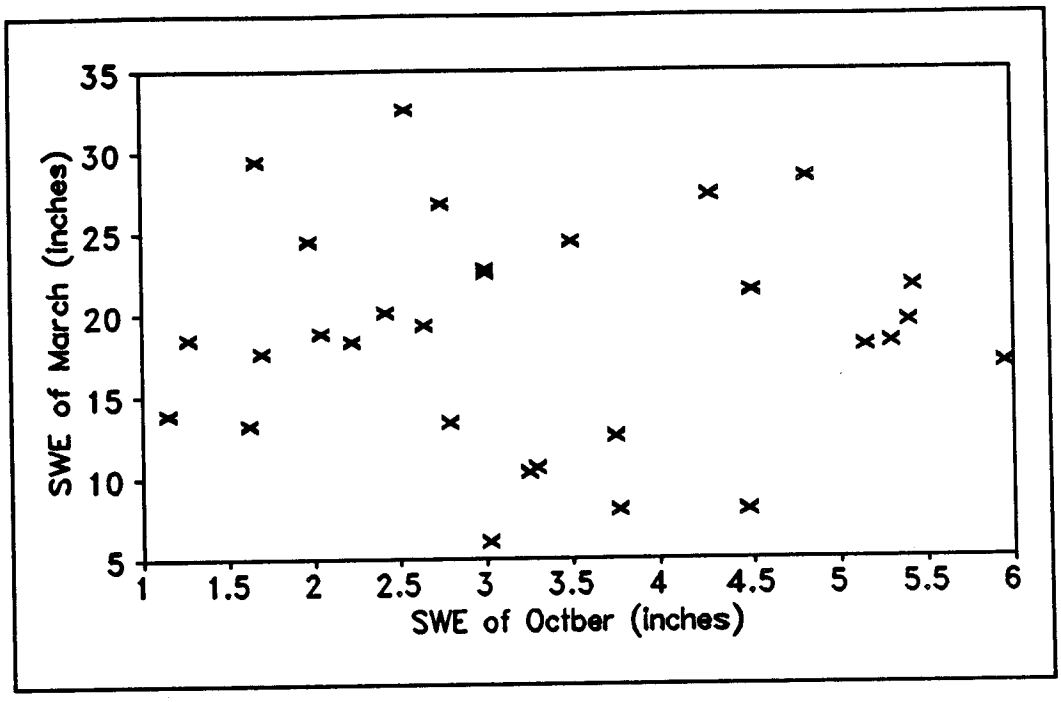


Fig. 7.27 The SWE of October vs. the SWE of March (Los Pinos River Basin, 1948-1977)

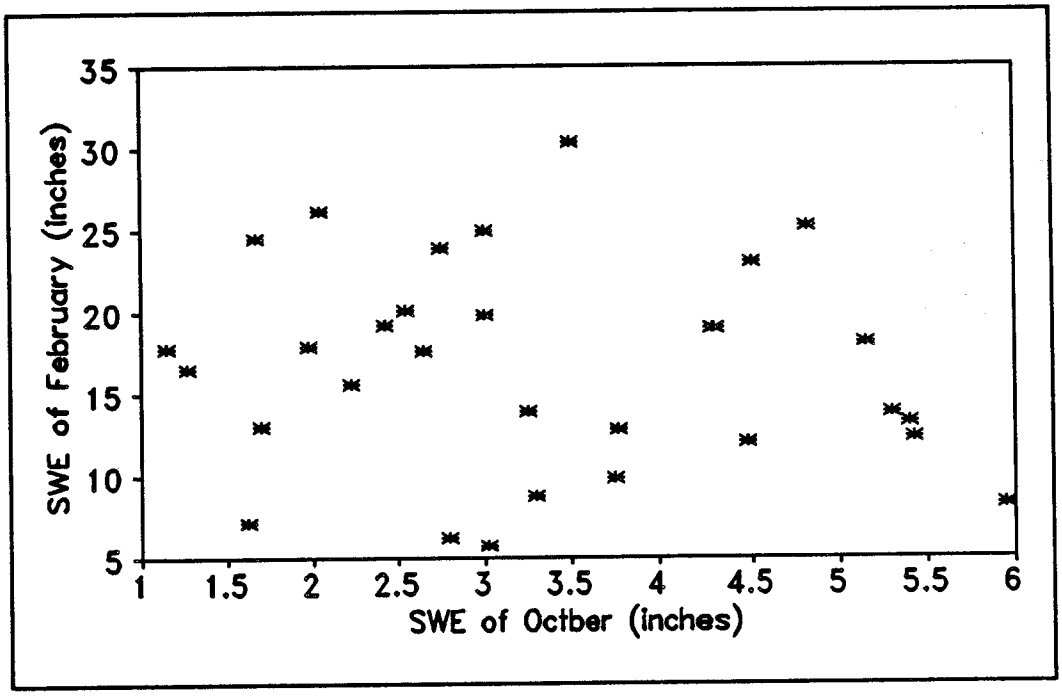


Fig. 7.28 The SWE of October vs. the SWE of February (Los Pinos River Basin, 1948-1977)

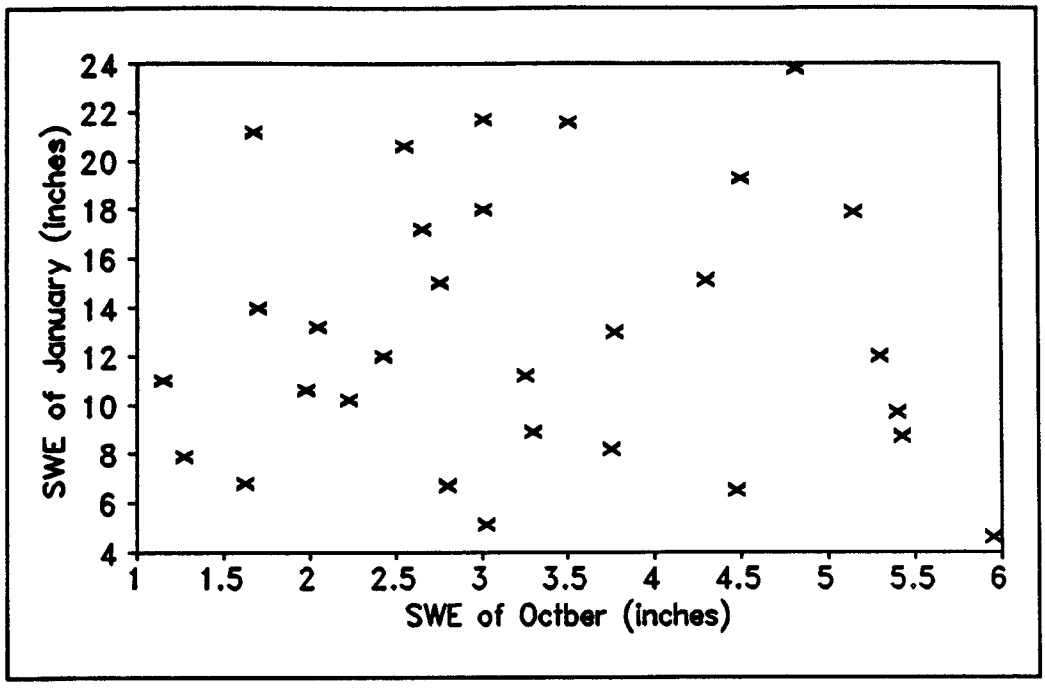


Fig. 7.29 The SWE of October vs. the SWE of January (Los Pinos River Basin, 1948-1977)

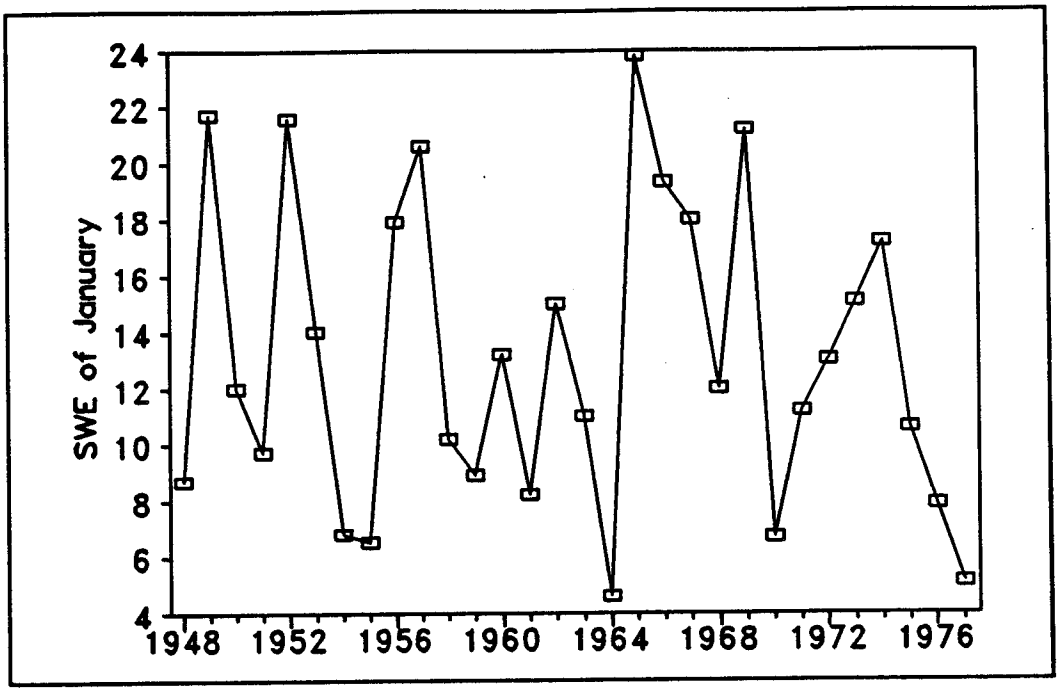


Fig. 7.30 The SWE of January (Los Pinos River Basin, 1948-1977)

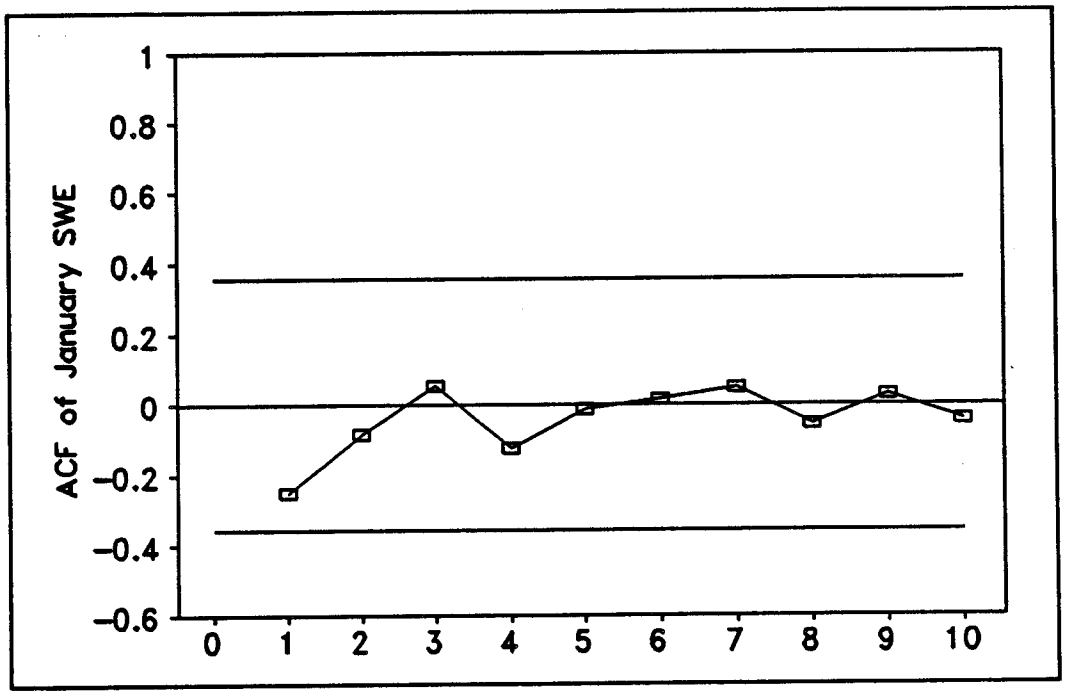


Fig. 7.31 The autocorrelation of January SWE (Los Pinos River Basin, 1948-1977)

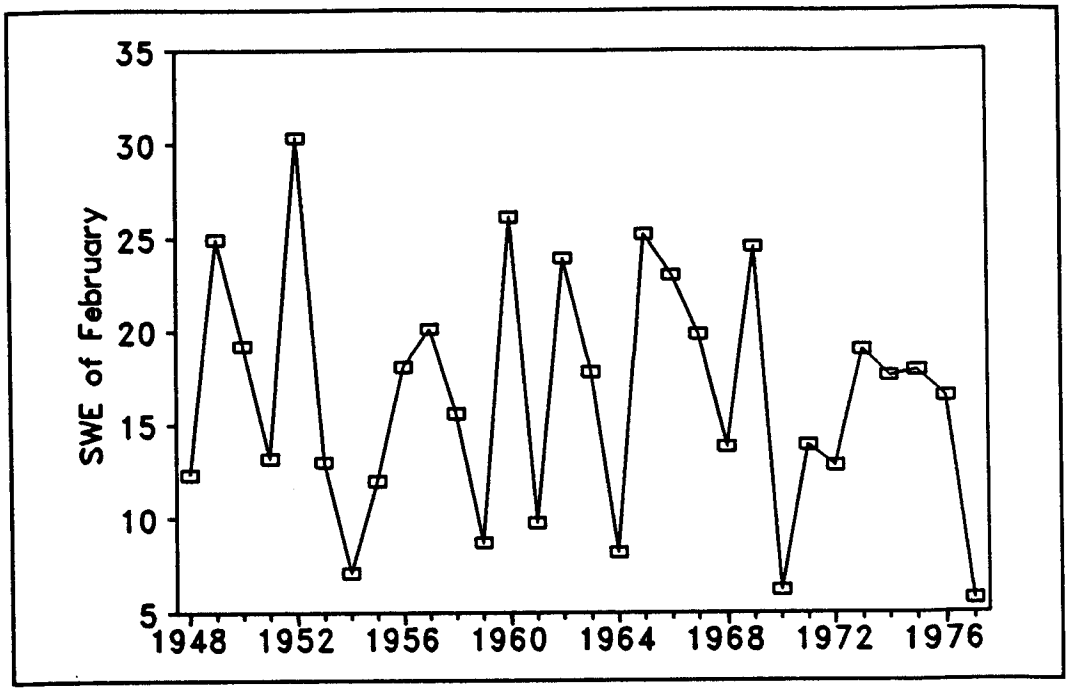


Fig. 7.32 The SWE of February (Los Pinos River Basin, 1948-1977)

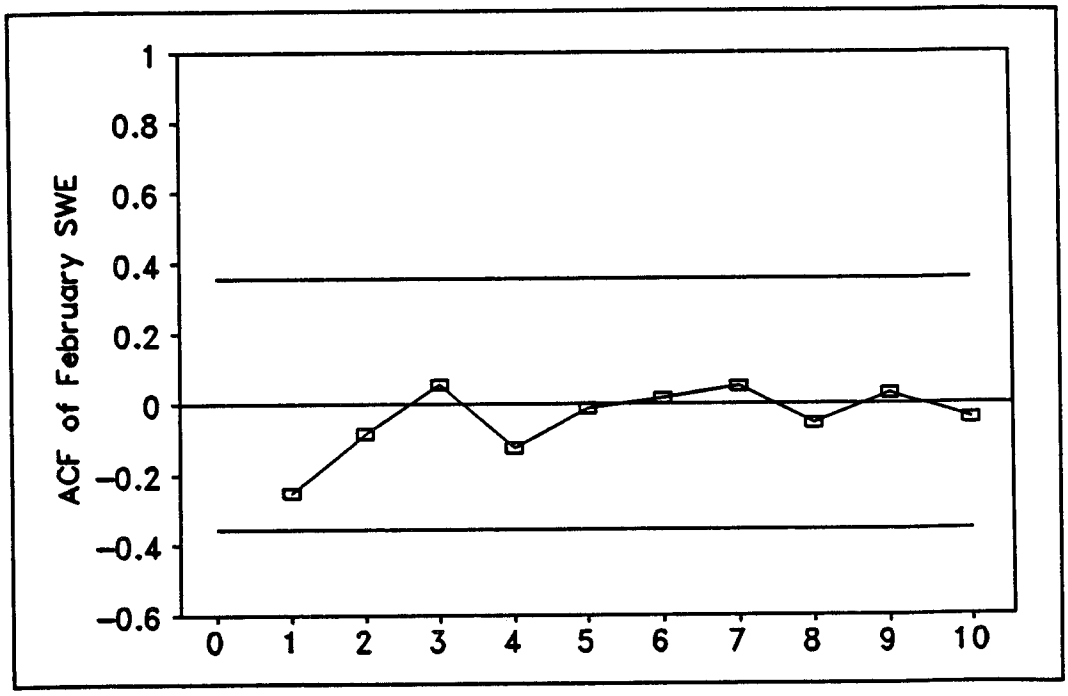


Fig. 7.33 The autocorrelation of February SWE (Los Pinos River Basin, 1948-1977)

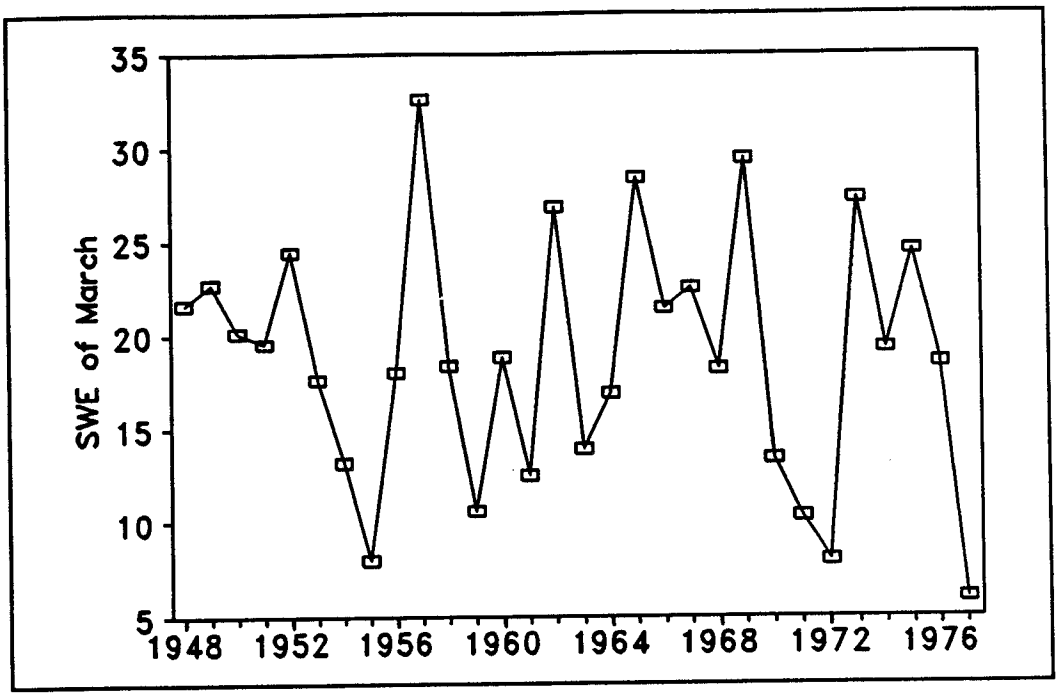


Fig. 7.34 The SWE of March (Los Pinos River Basin, 1948-1977)

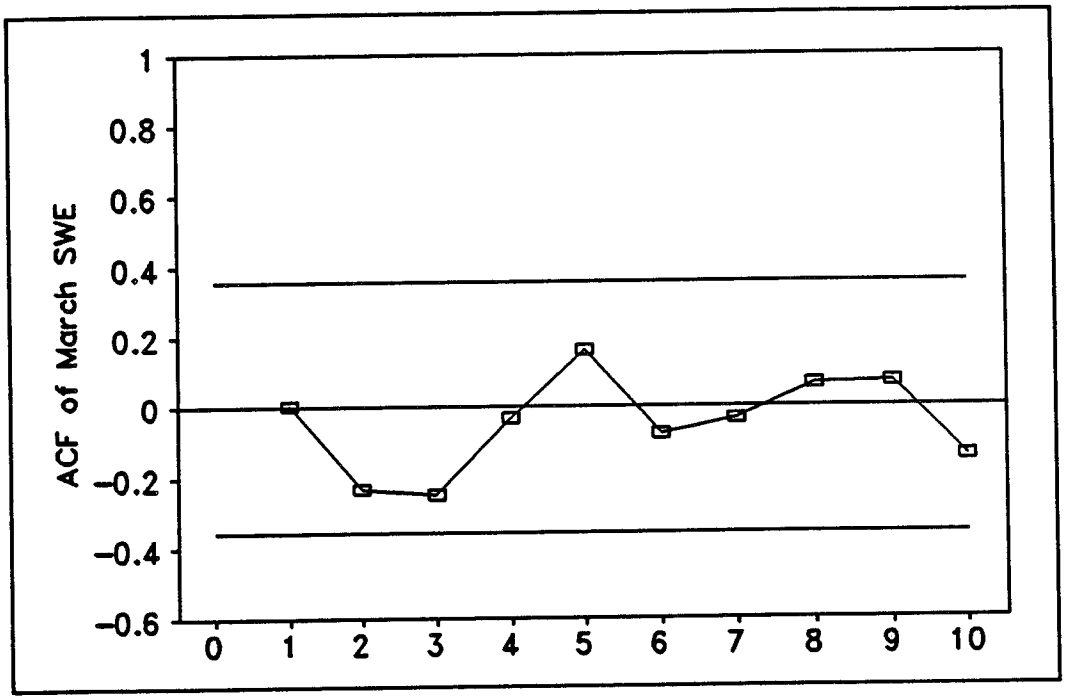


Fig. 7.35 The autocorrelation of March SWE (Los Pinos River Basin, 1948-1977)

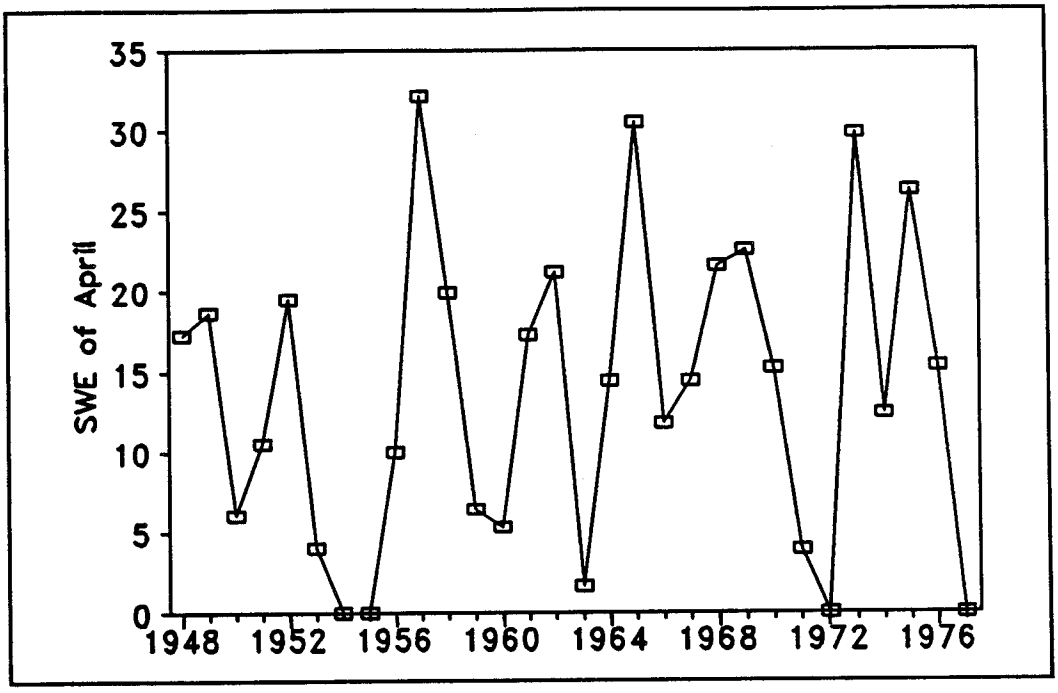


Fig. 7.36 The SWE of April (Los Pinos River Basin, 1948-1977)

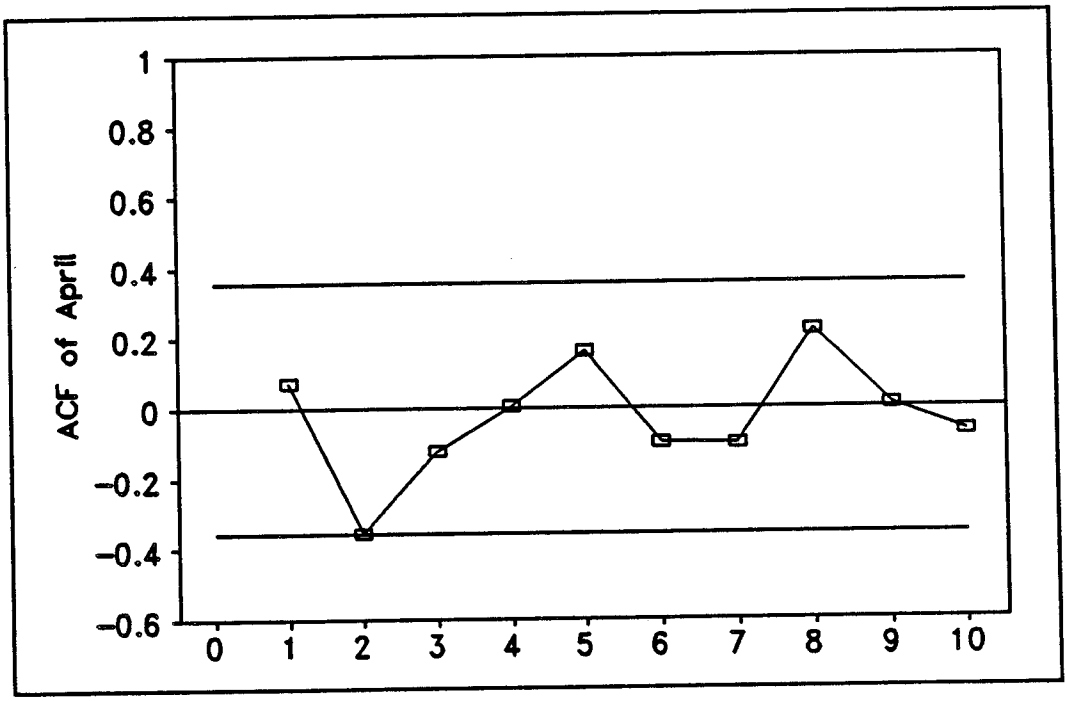


Fig. 7.37 The autocorrelation of April SWE (Los Pinos River Basin, 1948-1977)

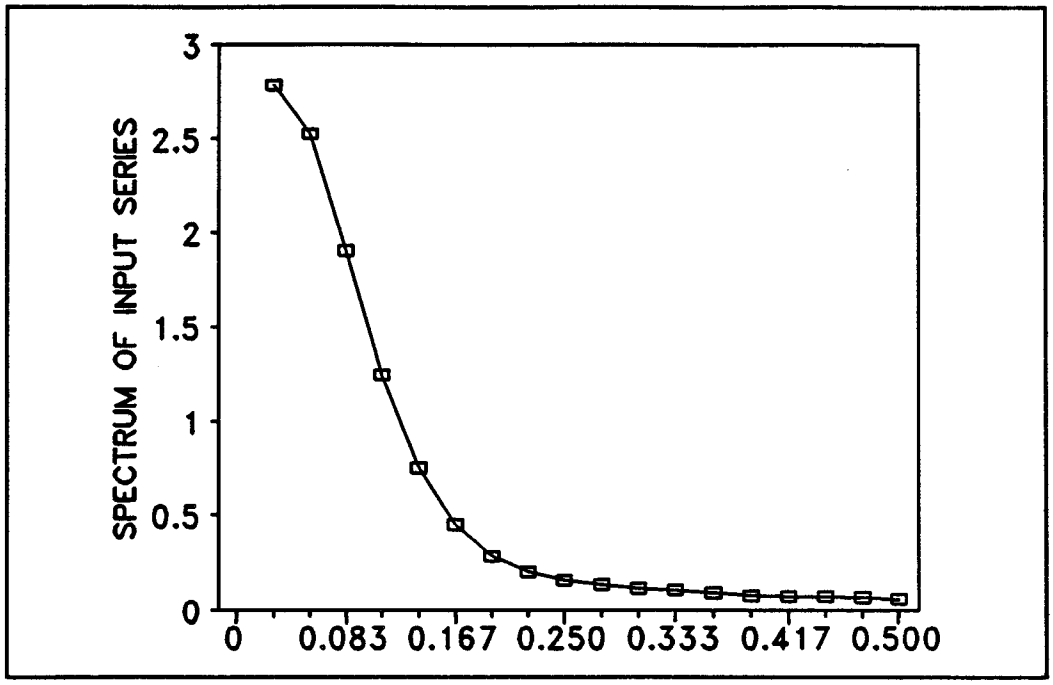


Fig. 7.38 The auto-spectrum of the deseasonalized SWE (Los Pinos River Basin, 1948-1977)

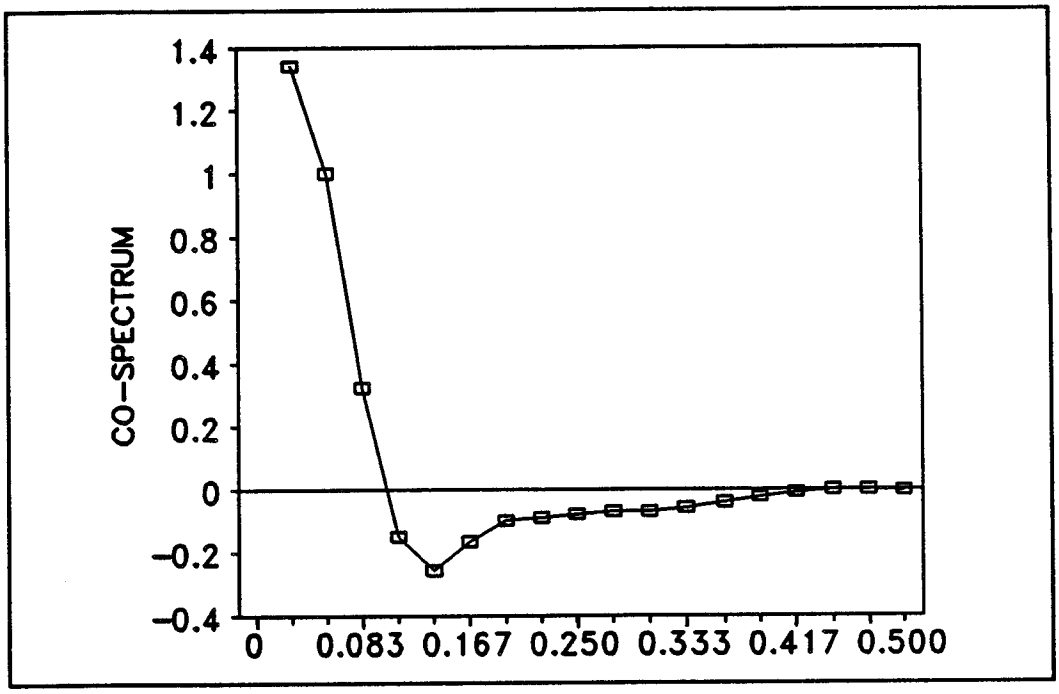


Fig. 7.39 The co-spectrum of the deseasonalized SWE and flow (Los Pinos River Basin, 1948-1977)

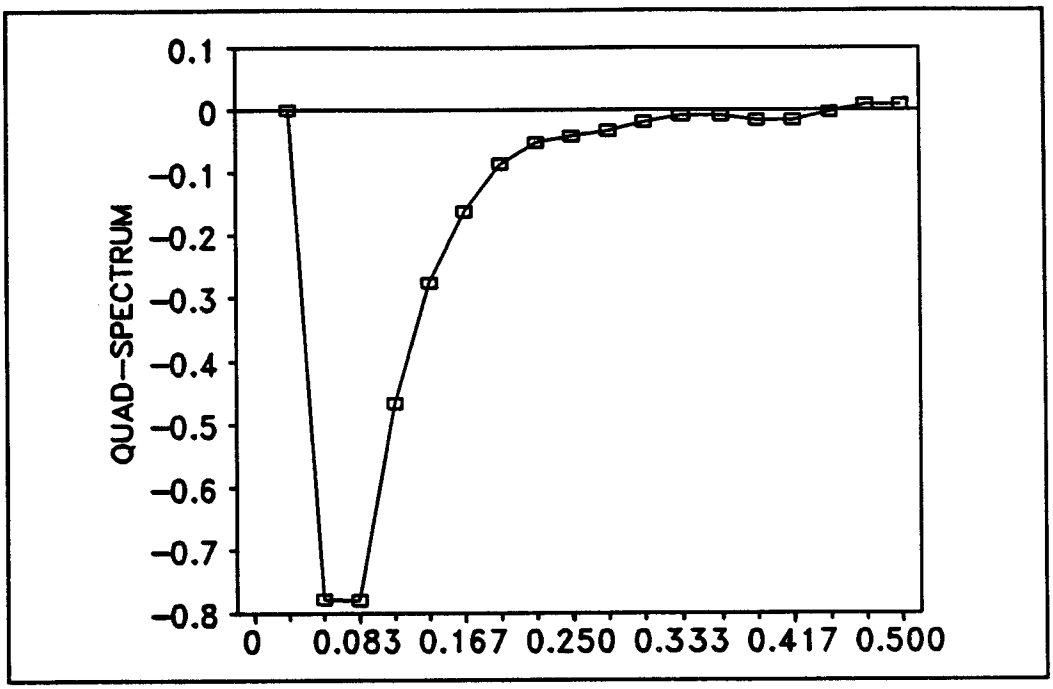


Fig. 7.40 The quad-spectrum of the deseasonalized SWE and flow (Los Pinos River Basin, 1948-1977)

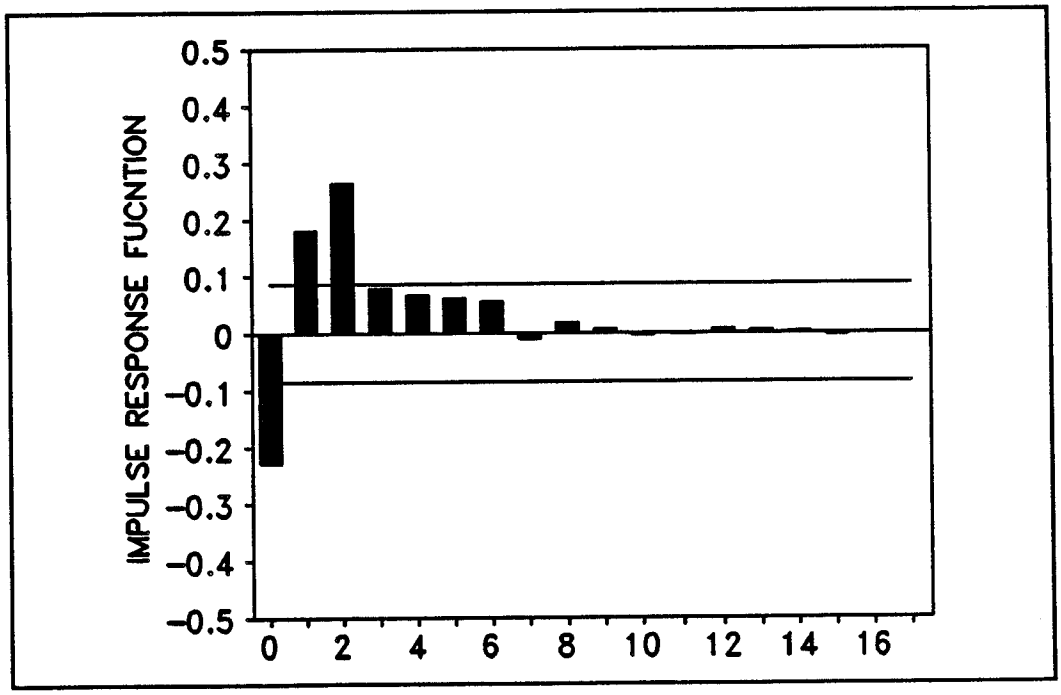


Fig. 7.41 The impulse response function (rational TFM model) (Los Pinos River Basin, 1948-1977; log-transformation)

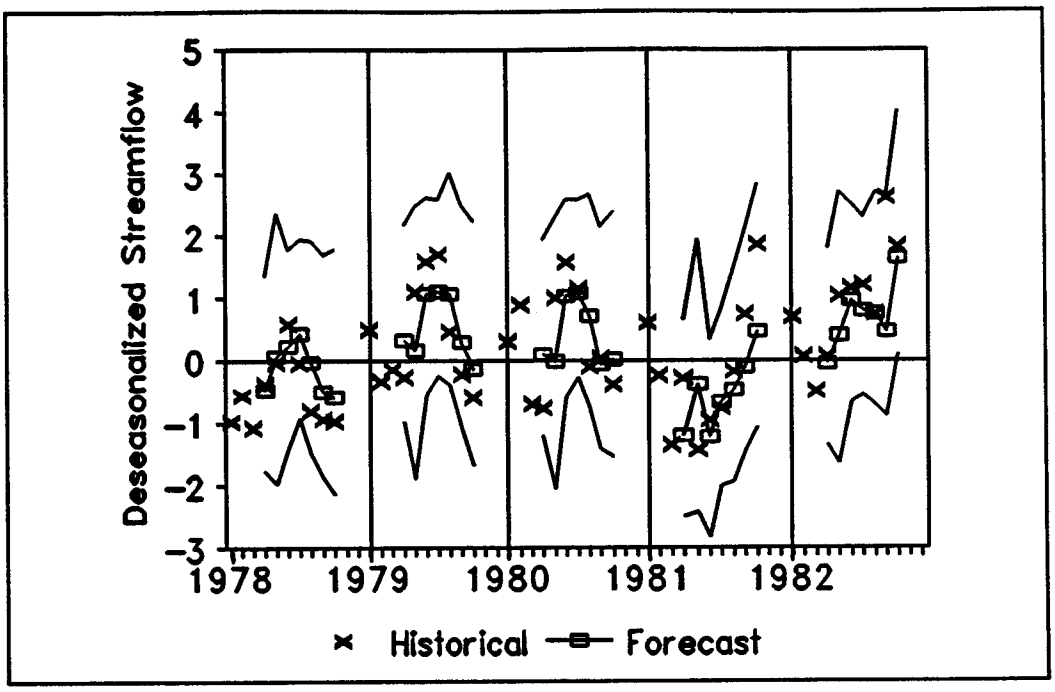


Fig. 7.42 Forecasting of the deseasonalized flow L=1,...,7 (Los Pinos River Basin, 1948-1977; rational TFM)

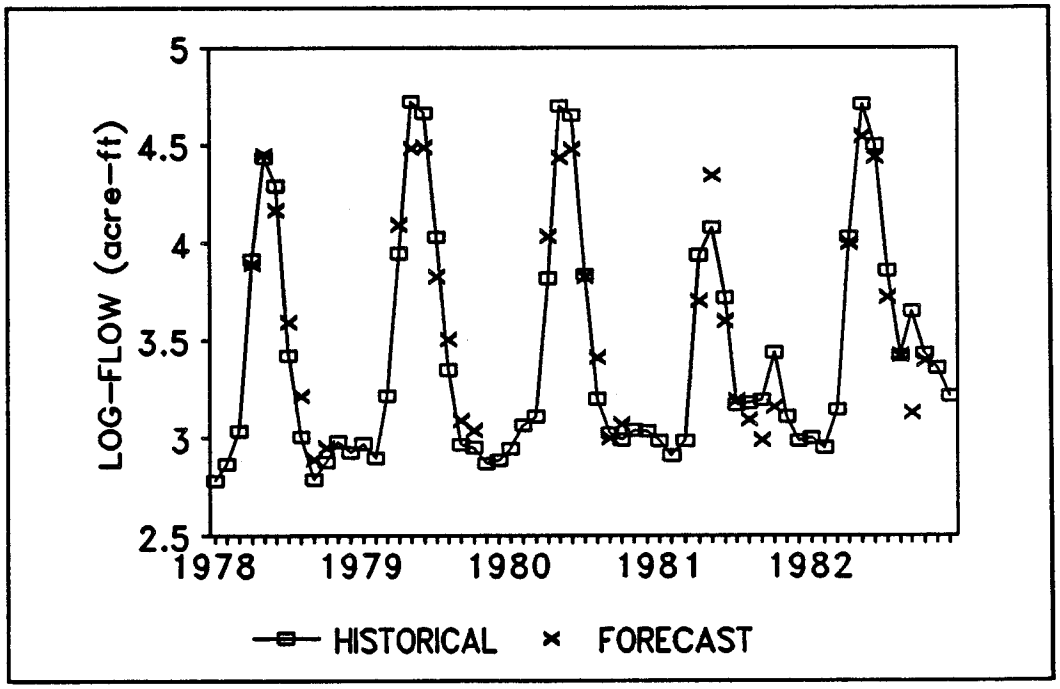


Fig. 7.43 One step ahead forecast of the log-flow (Los Pinos River Basin near Ortiz; rational TFM model)

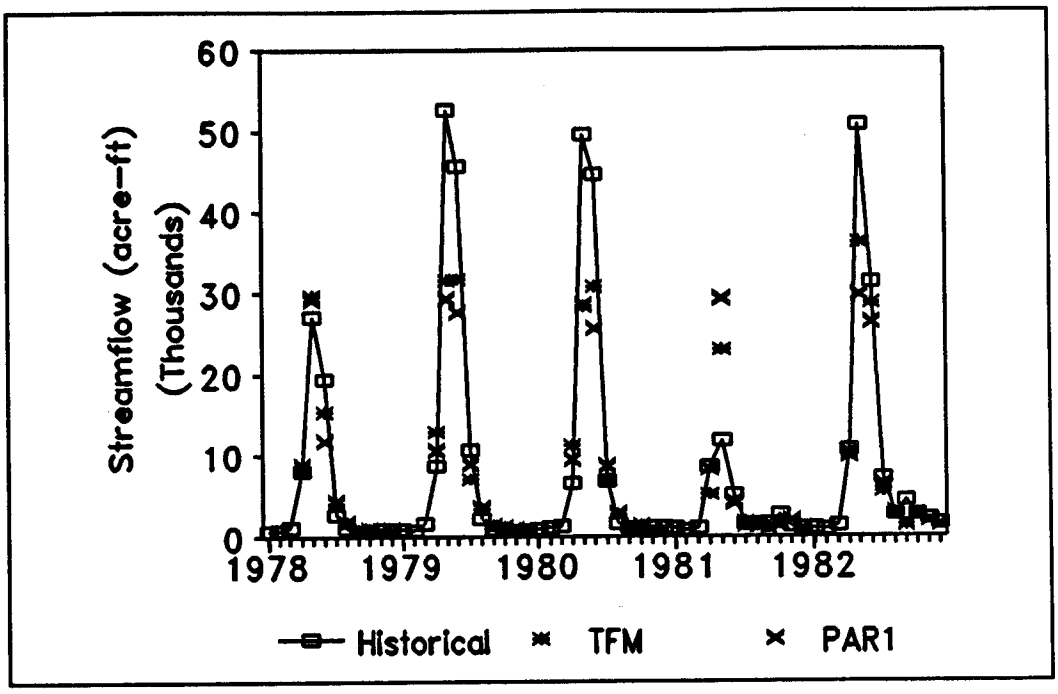


Fig. 7.44 One step ahead forecast of the actual monthly flow (Los Pinos River Basin near Ortiz; rational TFM model)

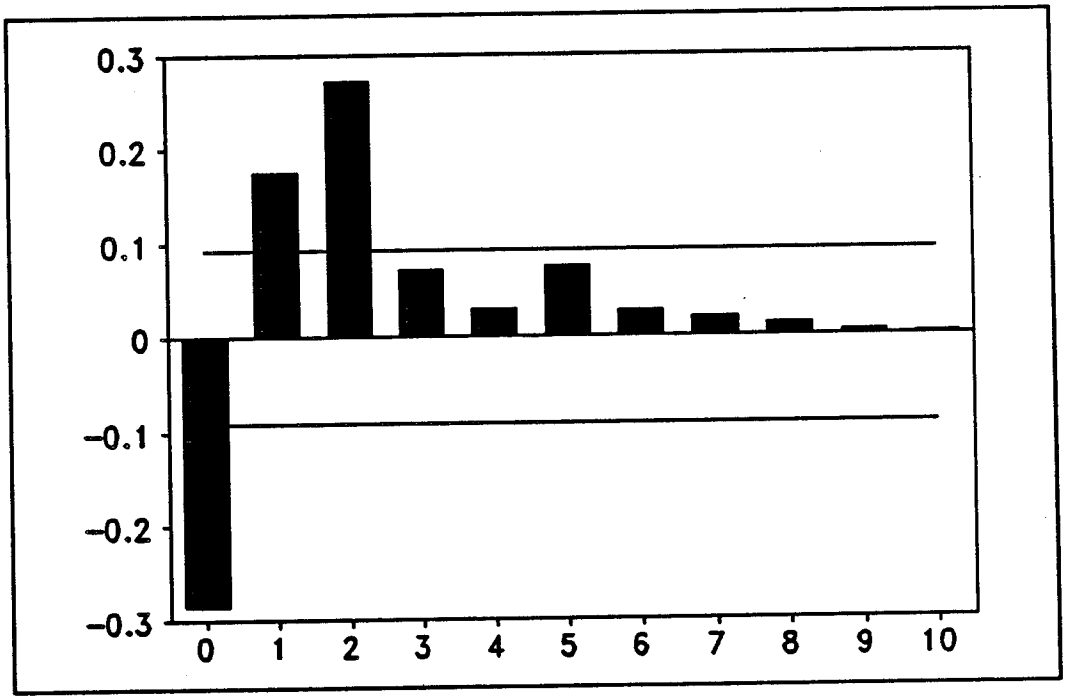


Fig. 7.45 The impulse response function (rational TFM model) (Los Pinos River Basin; no-log transformation)

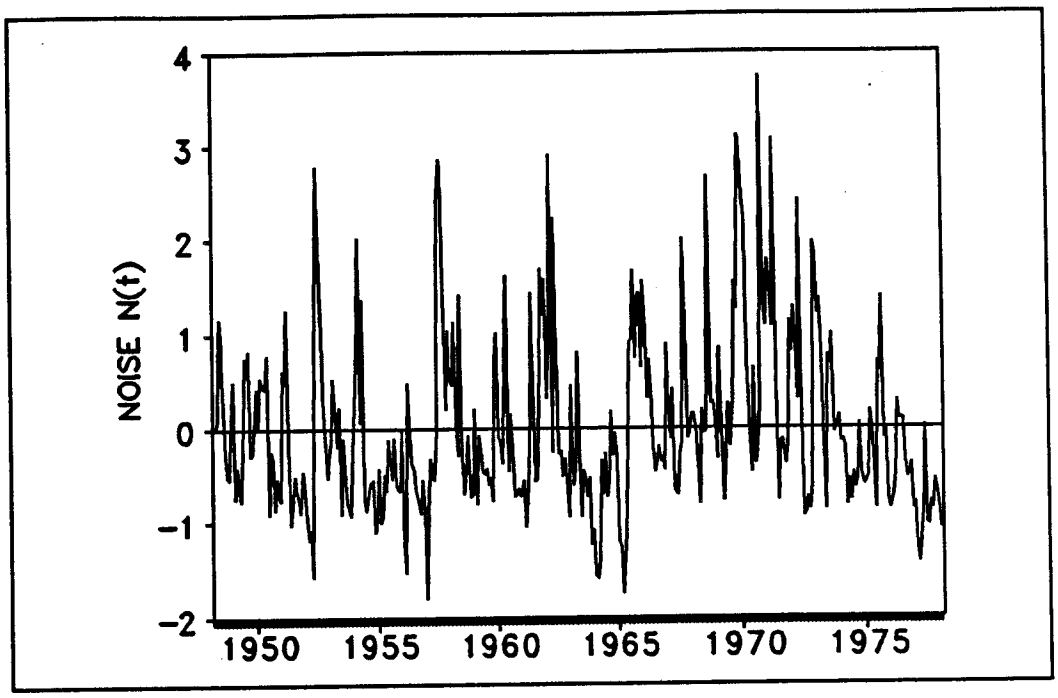


Fig. 7.46 The estimated noise (rational TFM model) (Los Pinos River Basin; no-log transformation)

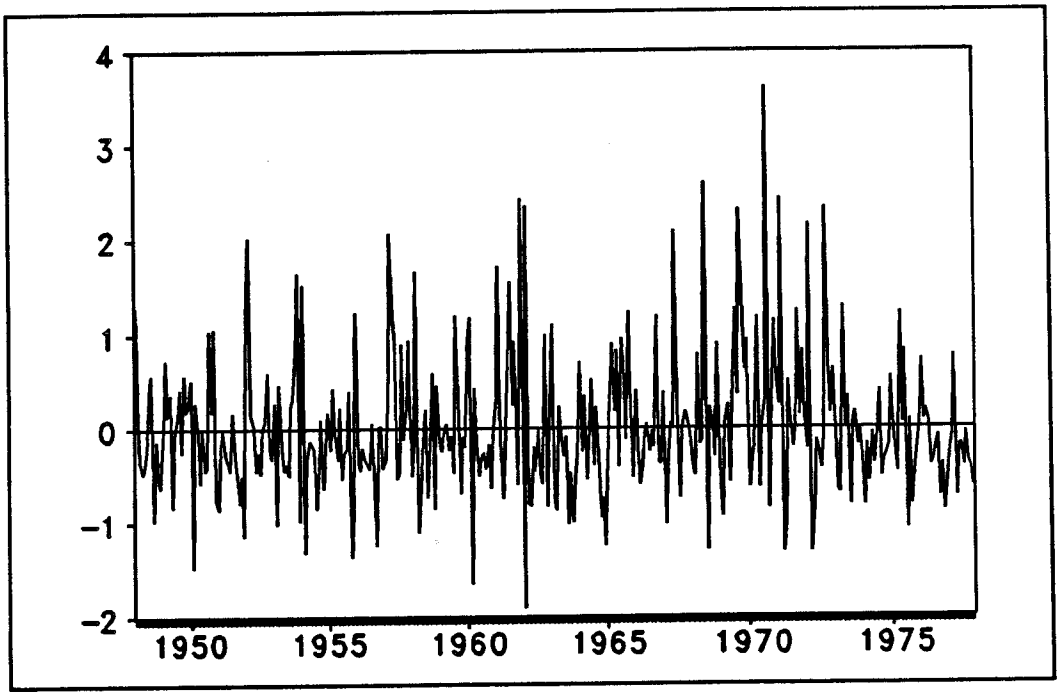


Fig. 7.47 The estimated residual (traditional TFM model (Los Pinos River Basin; no-log transformation)

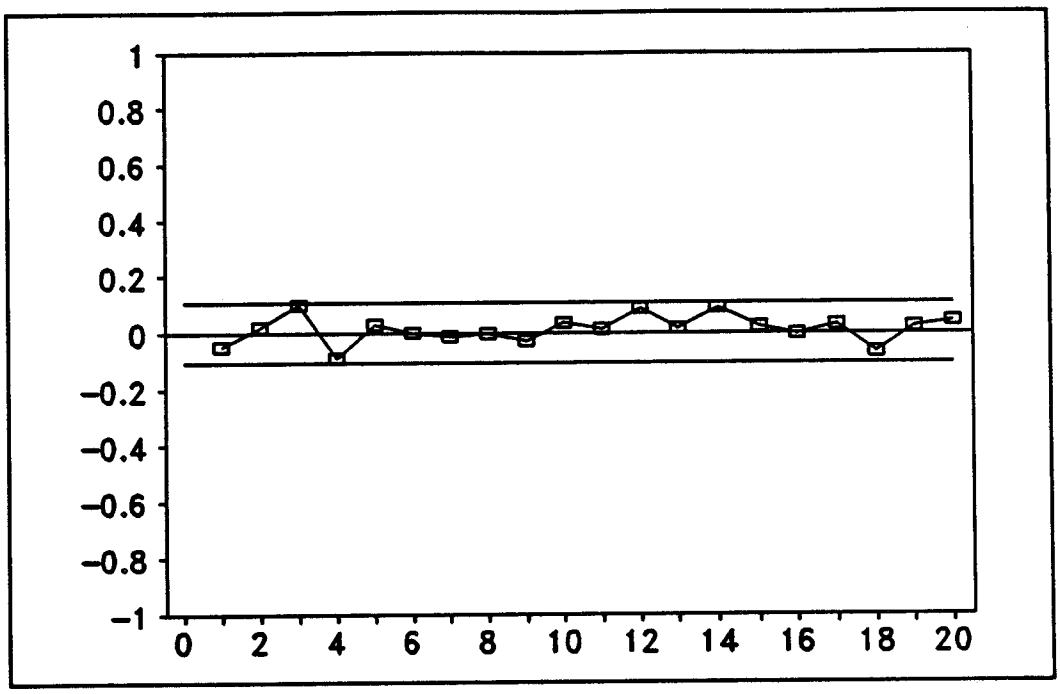


Fig. 7.48 Autocorrelation function of residual (Los Pinos River Basin near Ortiz; rational TFM model)

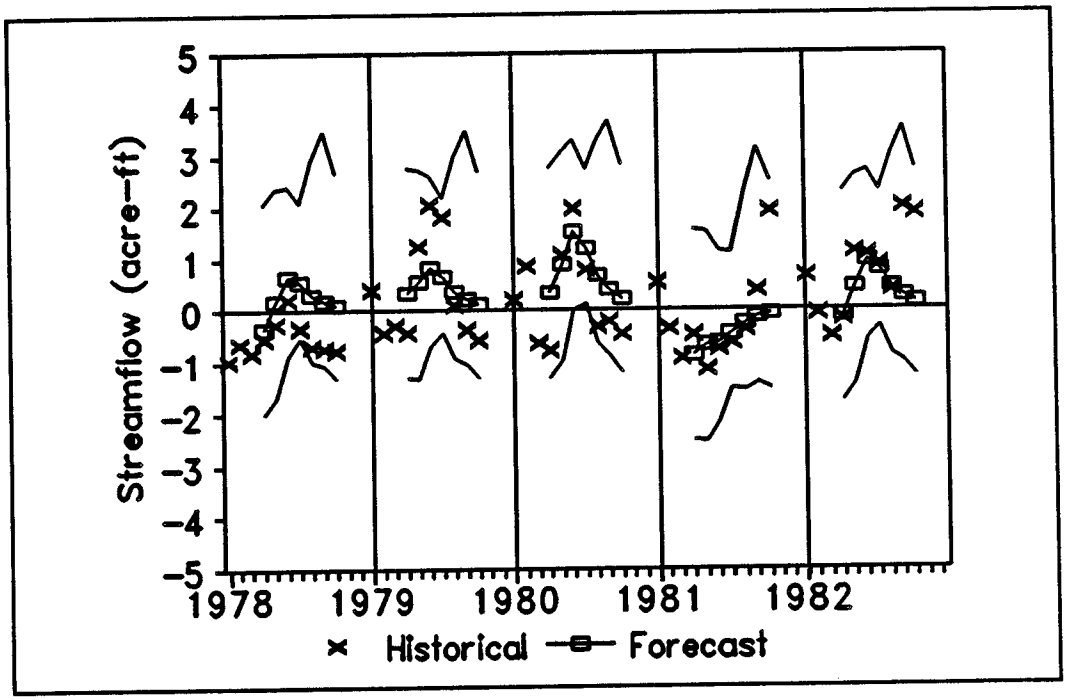


Fig. 7.49 Forecasting of deseasonalized flow, L=1,...,7 (Los Pinos River Basin near Ortiz; rational TFM model)

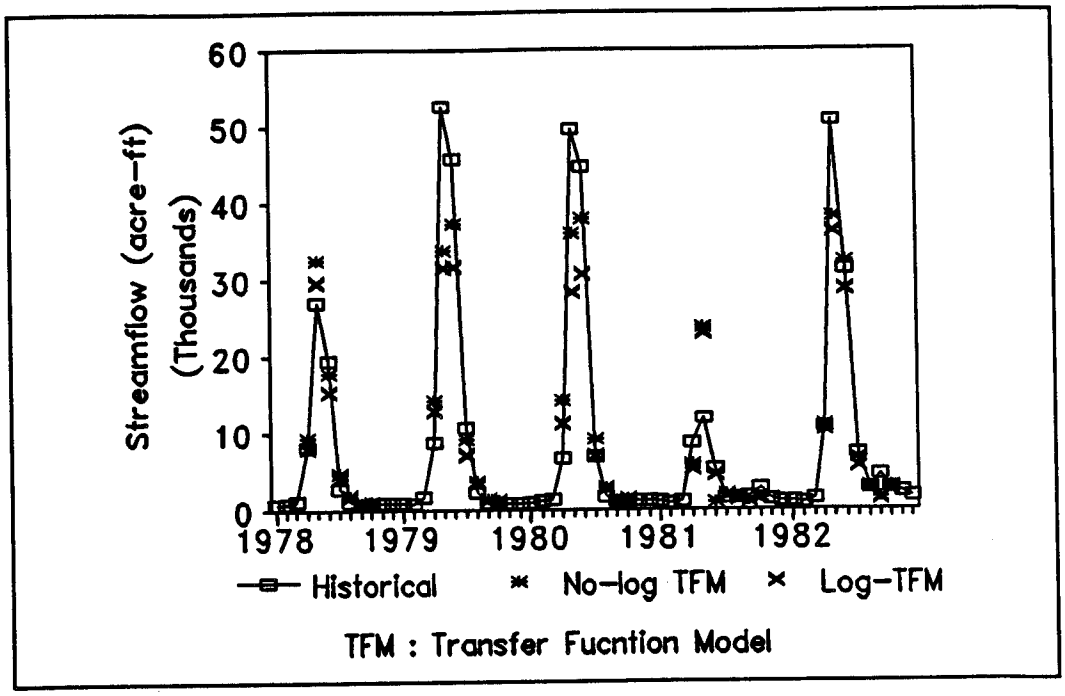


Fig. 7.50 One step ahead forecast of the monthly flow (Los Pinos River Basin near Ortiz; rational TFM model)

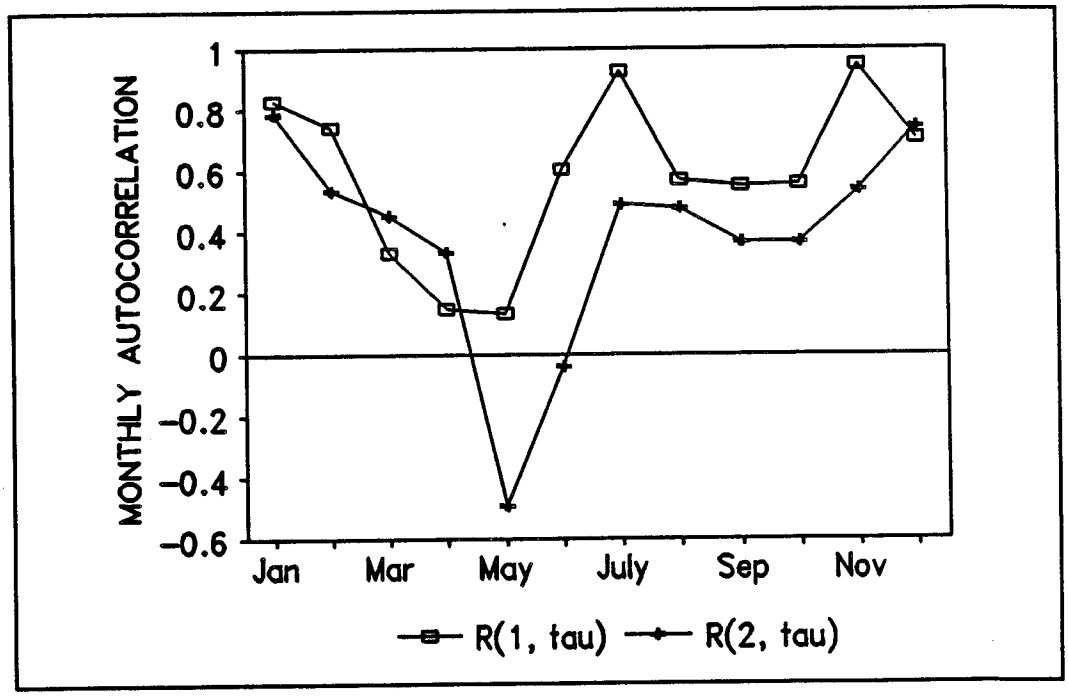


Fig. 7.51 Seasonal correlation of the noise (Los Pinos River Basin near Ortiz; 1948-1977)

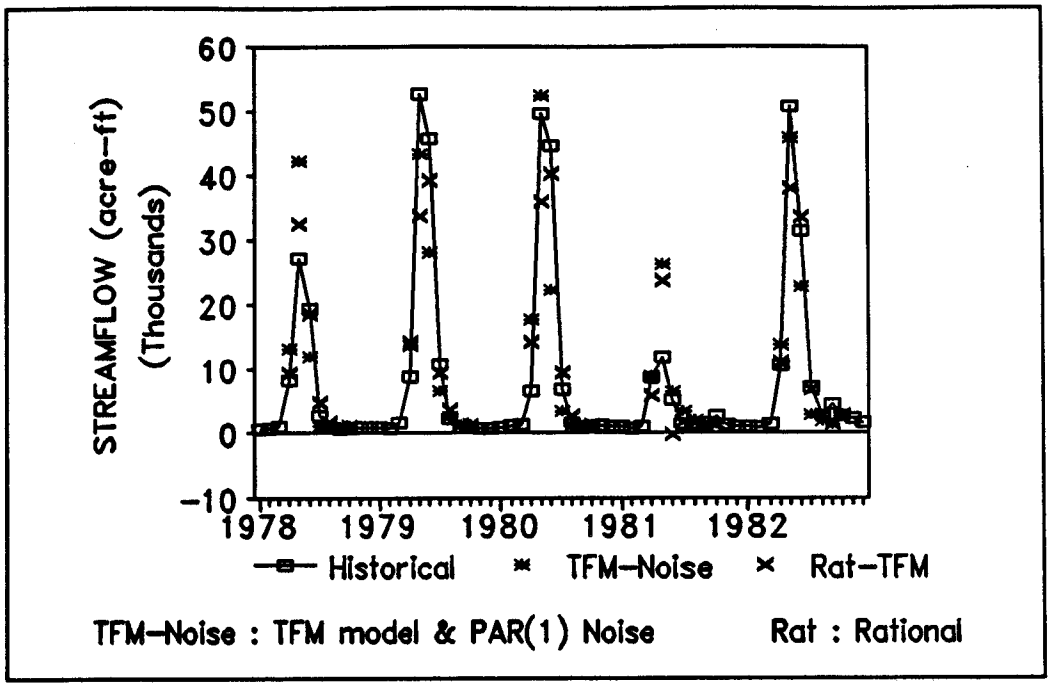


Fig. 7.52 One step ahead forecast of the monthly flow (Los Pinos River Basin near ortiz; TFM-PAR(1) noise model)

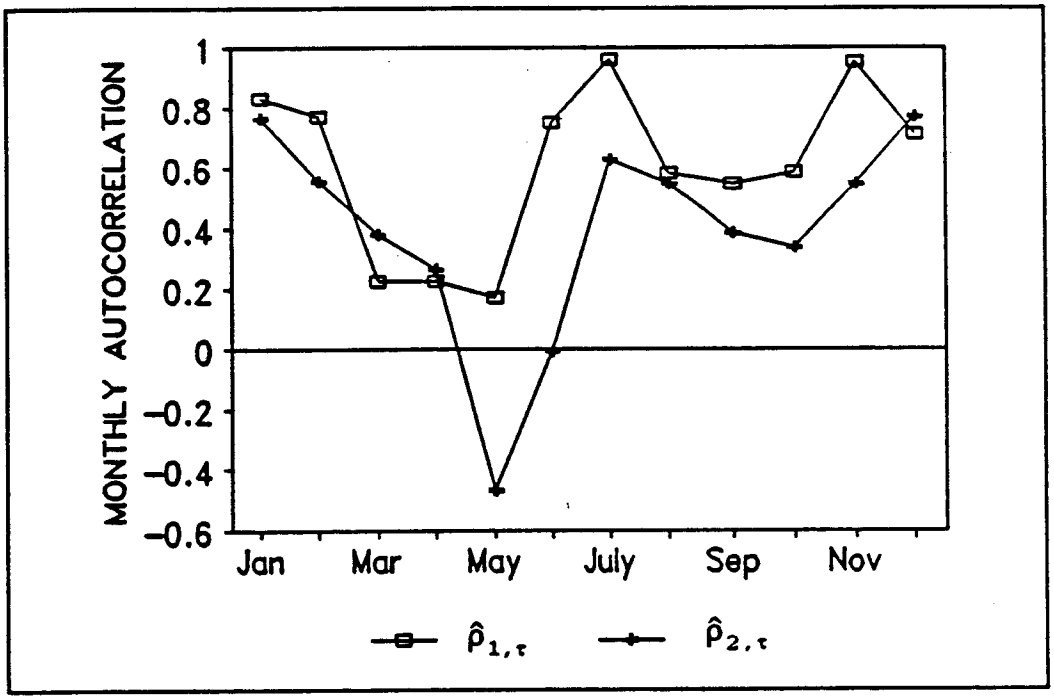


Fig. 7.53 The monthly correlation of the deseasonalized flow (Los Pinos River Basin, near Ortiz; 1948-1977)

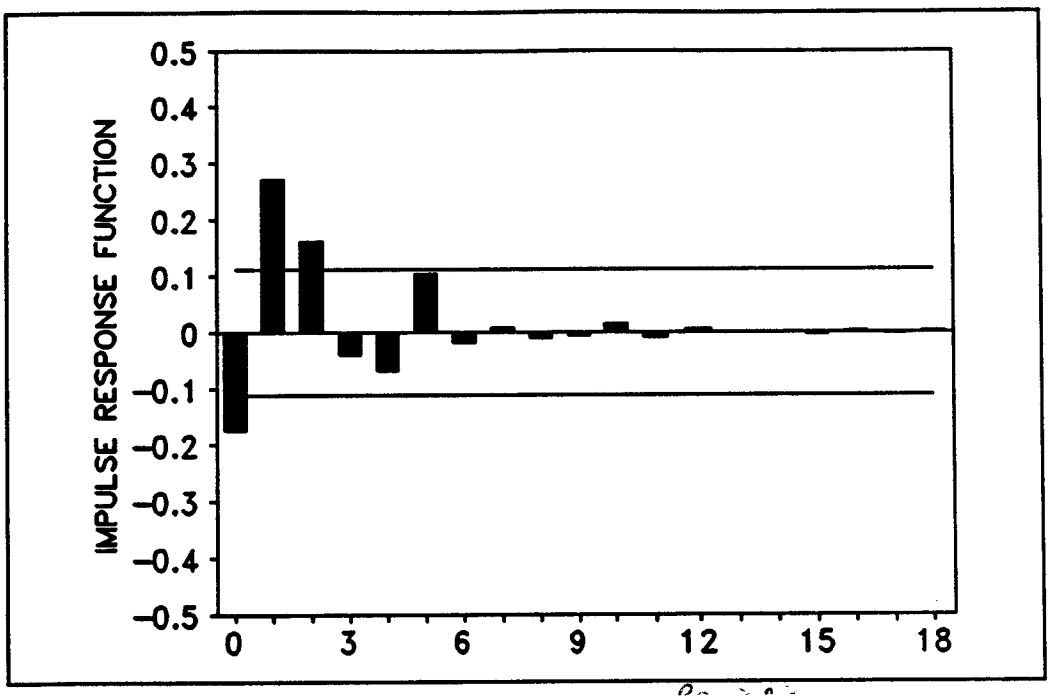


Fig. 7.54 Impulse response function (Periodic-TFM model) (Los Pinos River Basin near Ortiz; 1948-1977)

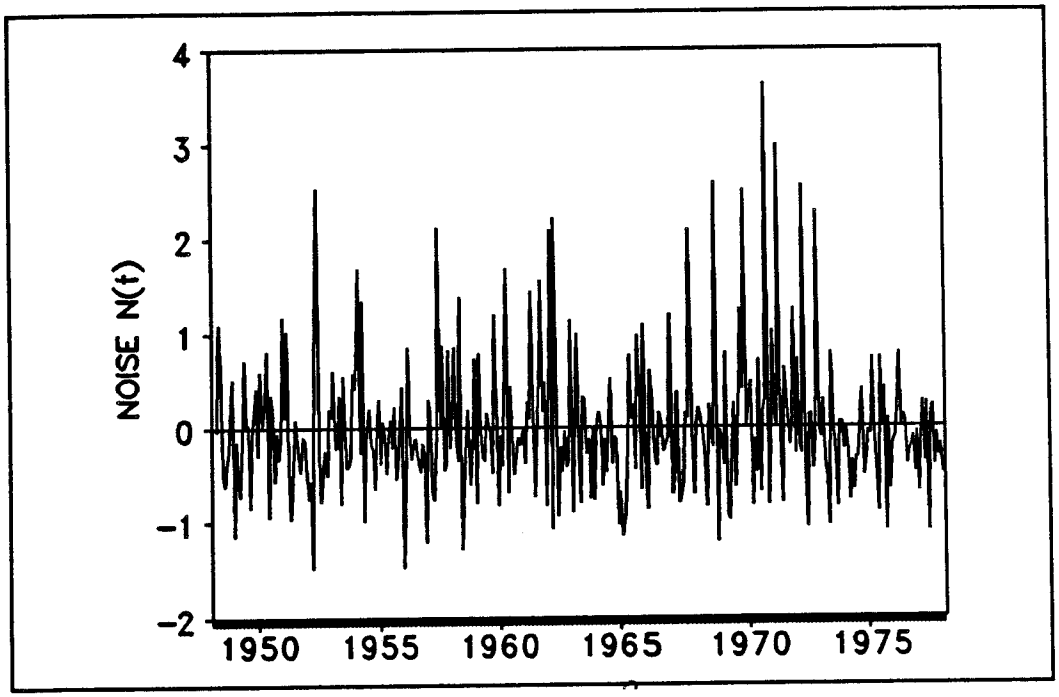


Fig. 7.55 The estimated noise (periodic-TFM model) (Los Pinos River Basin near Ortiz; 1948-1977)

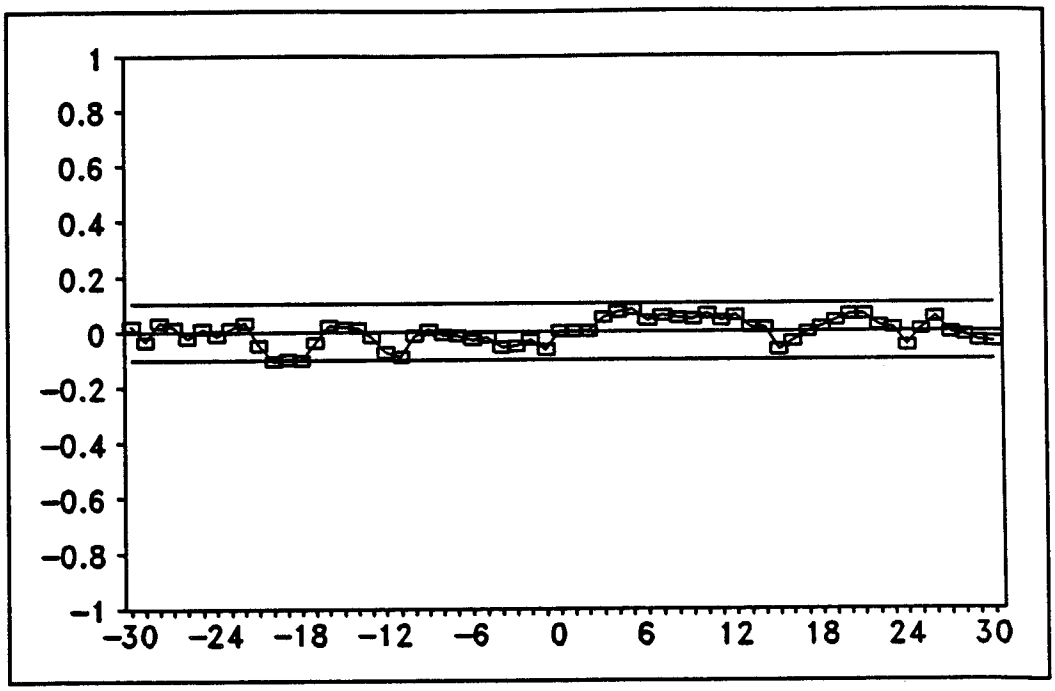


Fig. 7.56 The cross correlation function of the noise and input (Los Pinos River Basin near Ortiz; Periodic-TFM)

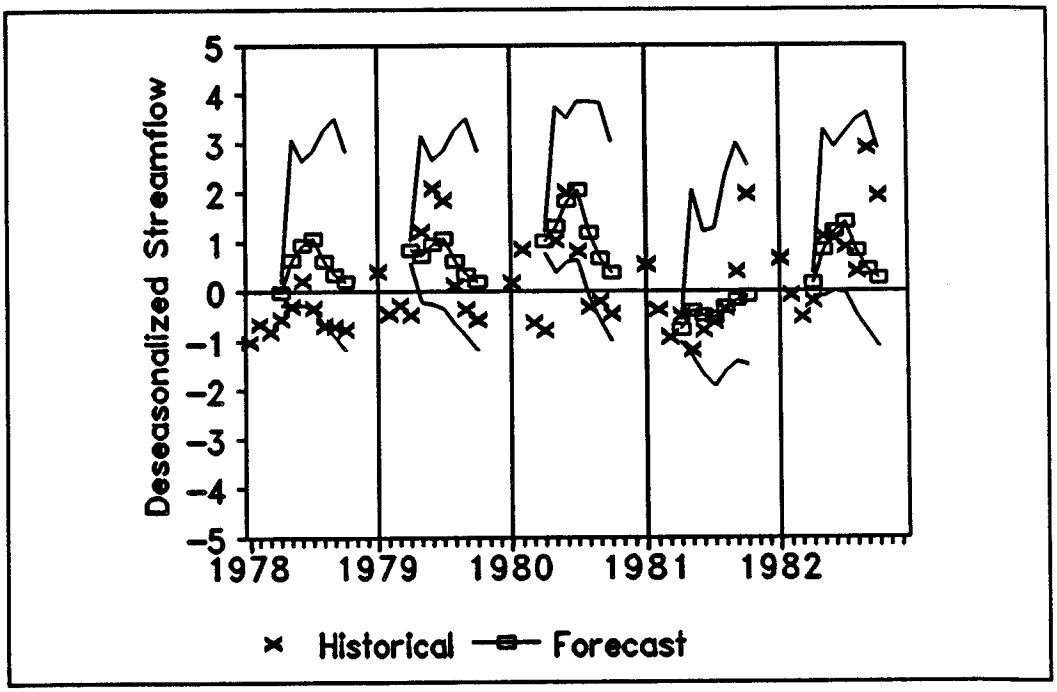


Fig. 7.57 Forecasting of deseasonalized flow with L=1,...,7 (Los Pinos River Basin near Ortiz; periodic-TFM)

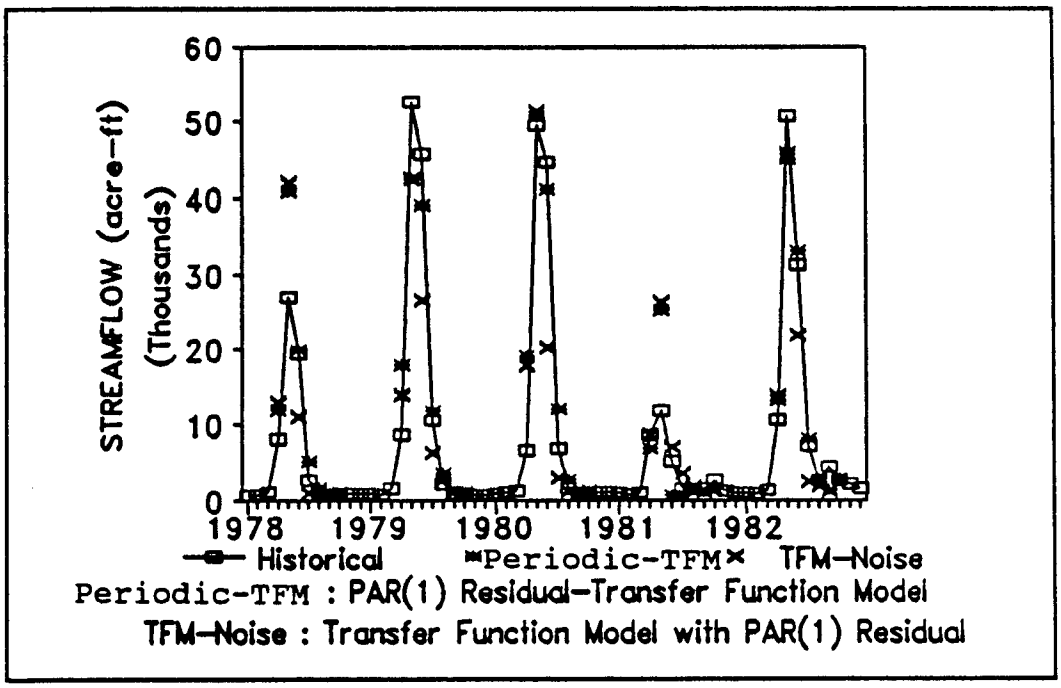


Fig. 7.58 One step ahead forecast of the monthly flow (Los Pinos River Basin near Ortiz; periodic-TFM and TFM-noise model)

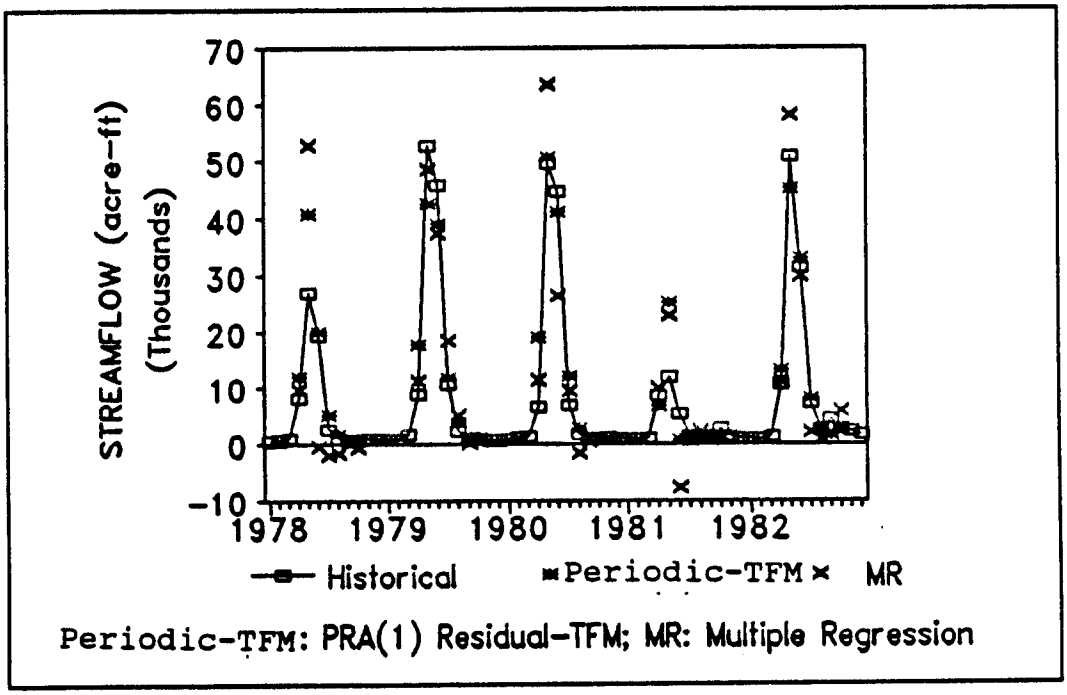


Fig. 7.59 One step ahead forecast of the monthly flow (Los Pinos River Basin near Ortiz; periodic-TFM and MR model)

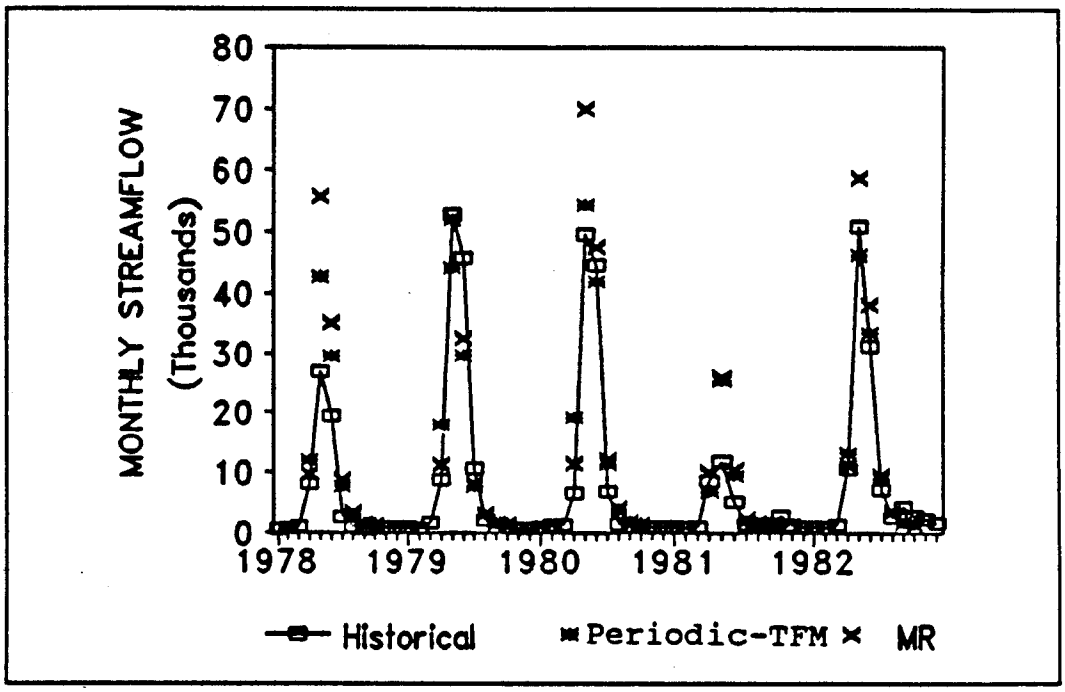


Fig. 7.60 April 1st forecasting of the monthly flow (Los Pinos River Basin)

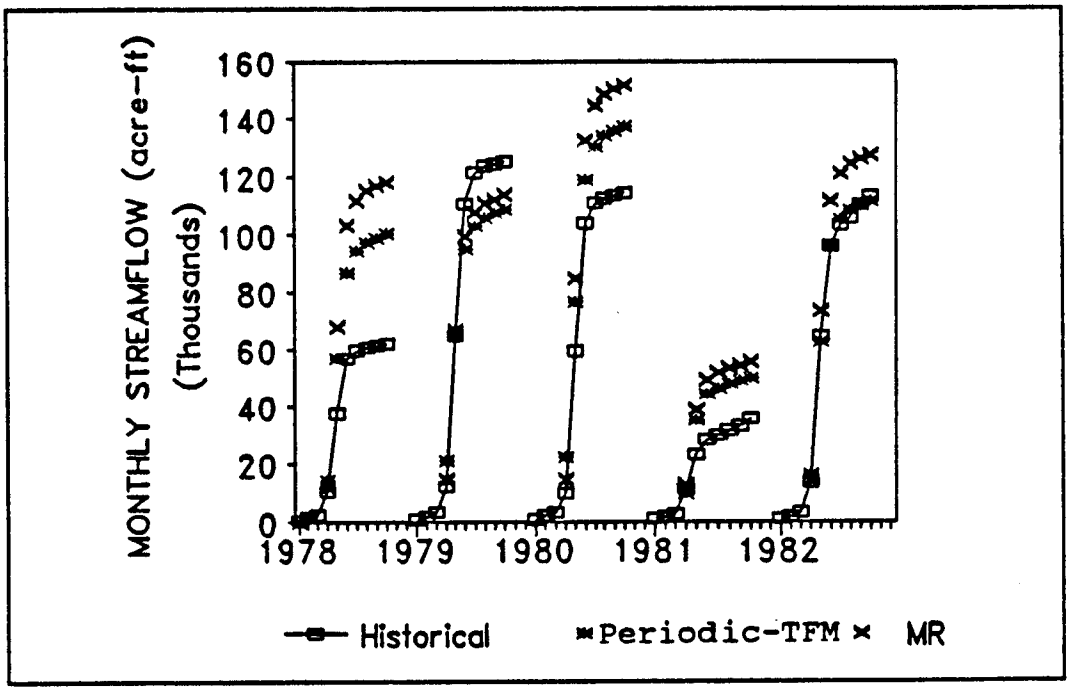


Fig. 7.61 April 1st forecasting of the cumulative flow (Los Pinos River Basin)

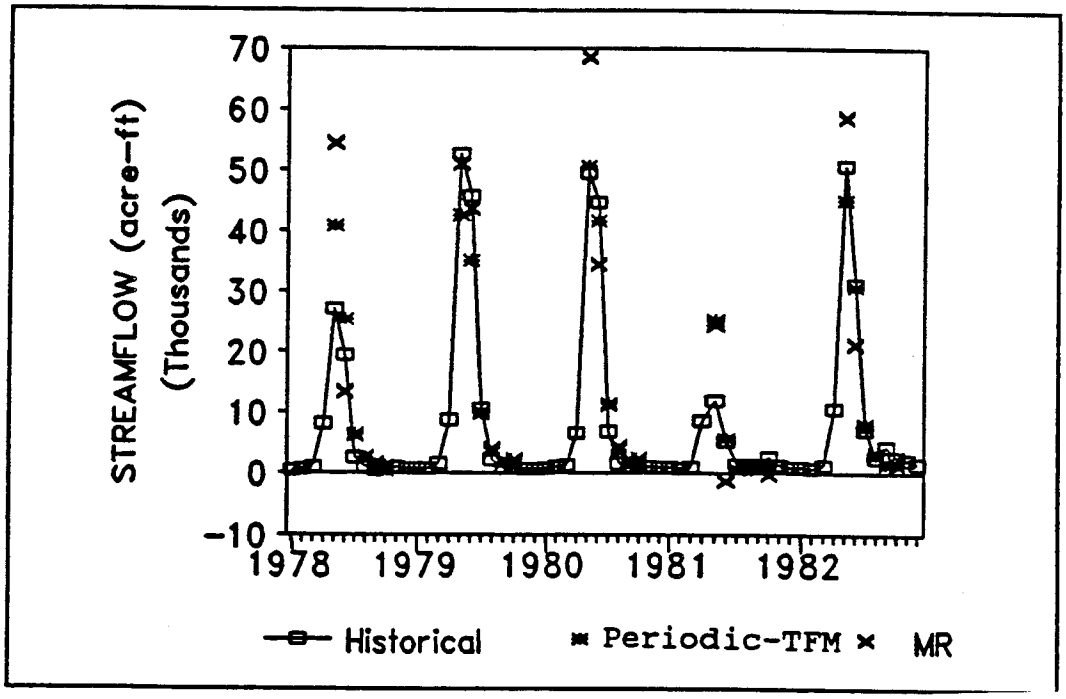


Fig. 7.62 May 1st forecasting of the monthly flow (Los Pinos River Basin)

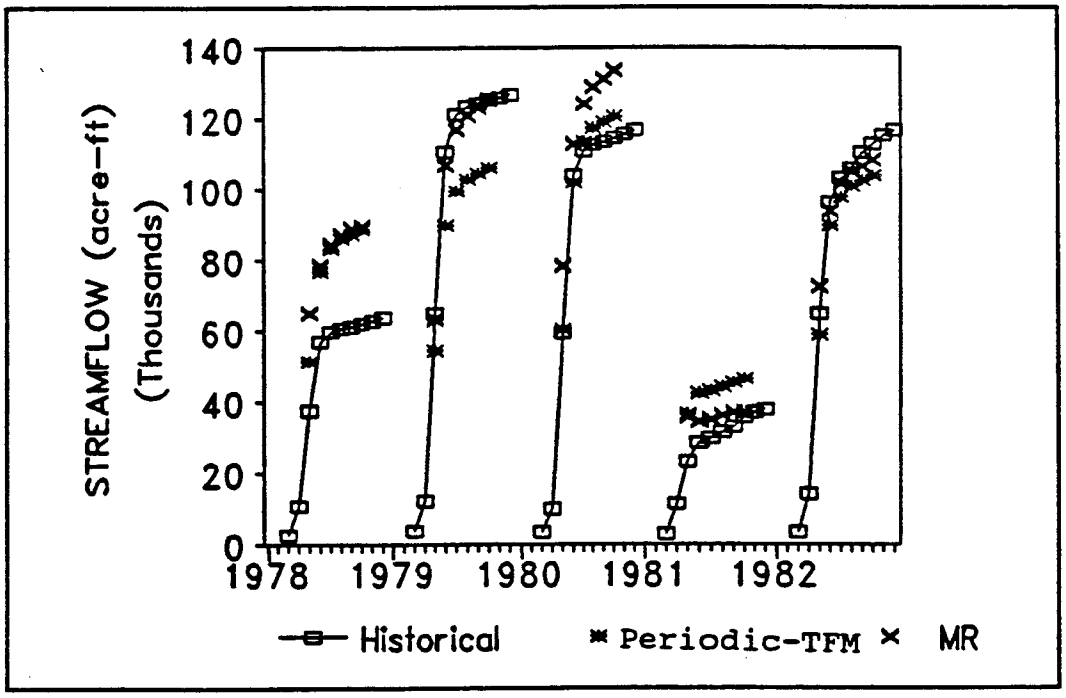


Fig. 7.63 May 1st forecasting of the cumulative flow (Los Pinos River Basin)

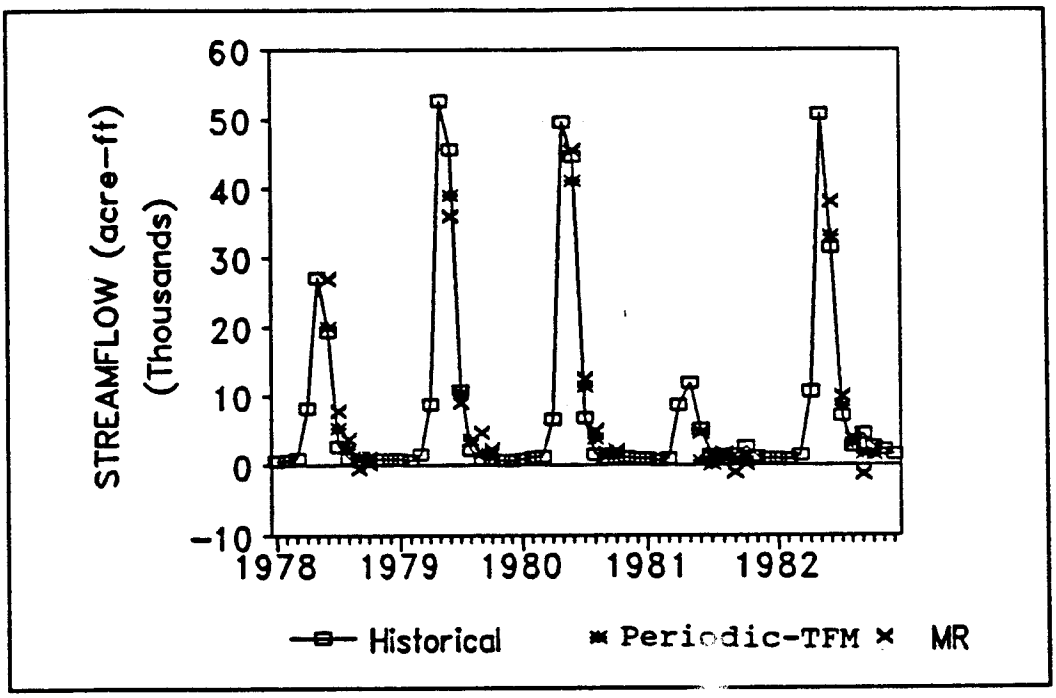


Fig. 7.64 June 1st forecasting of the monthly flow (Los Pinos River Basin)

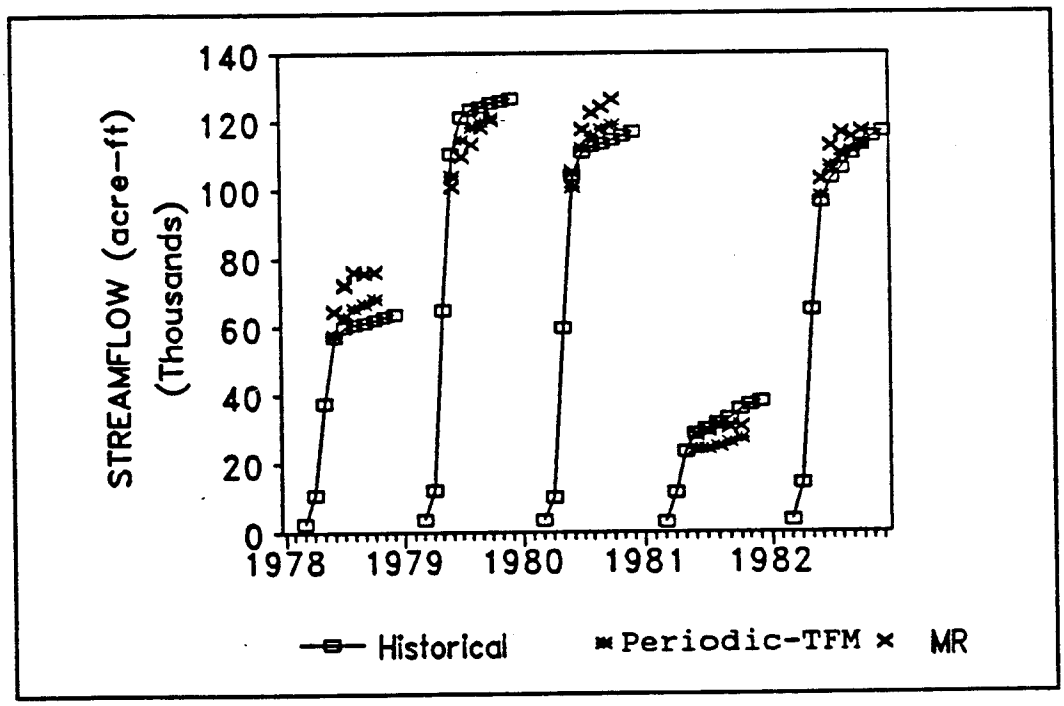


Fig. 7.65 June 1st forecasting of the cumulative flow (Los Pinos River Basin)

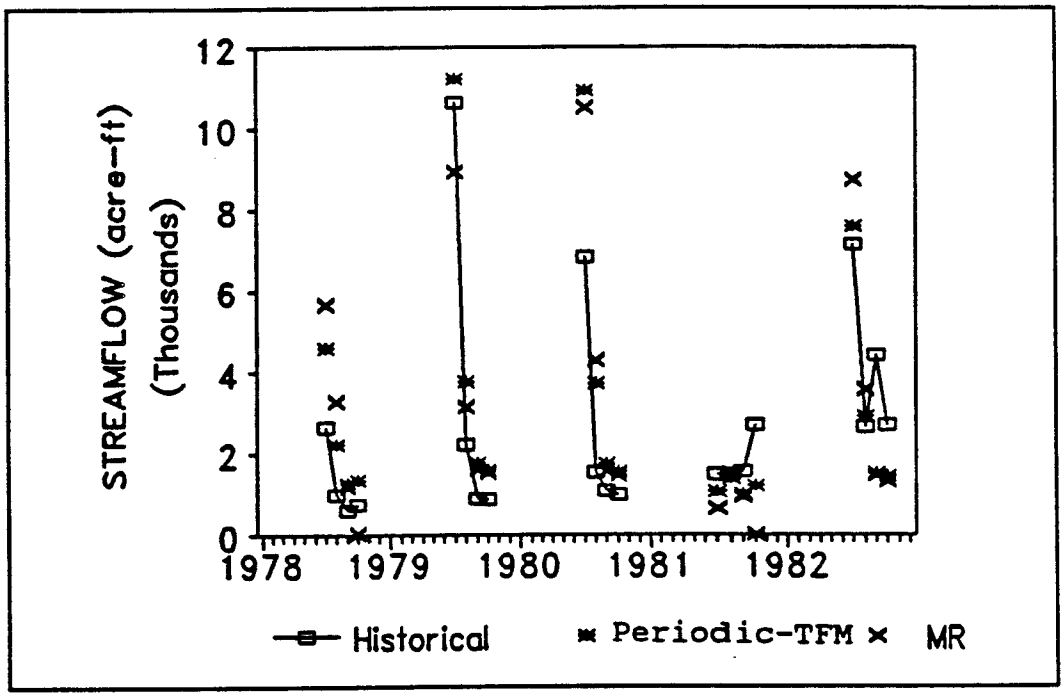


Fig. 7.66 July 1st forecasting of the monthly flow (Los Pinos River Basin)

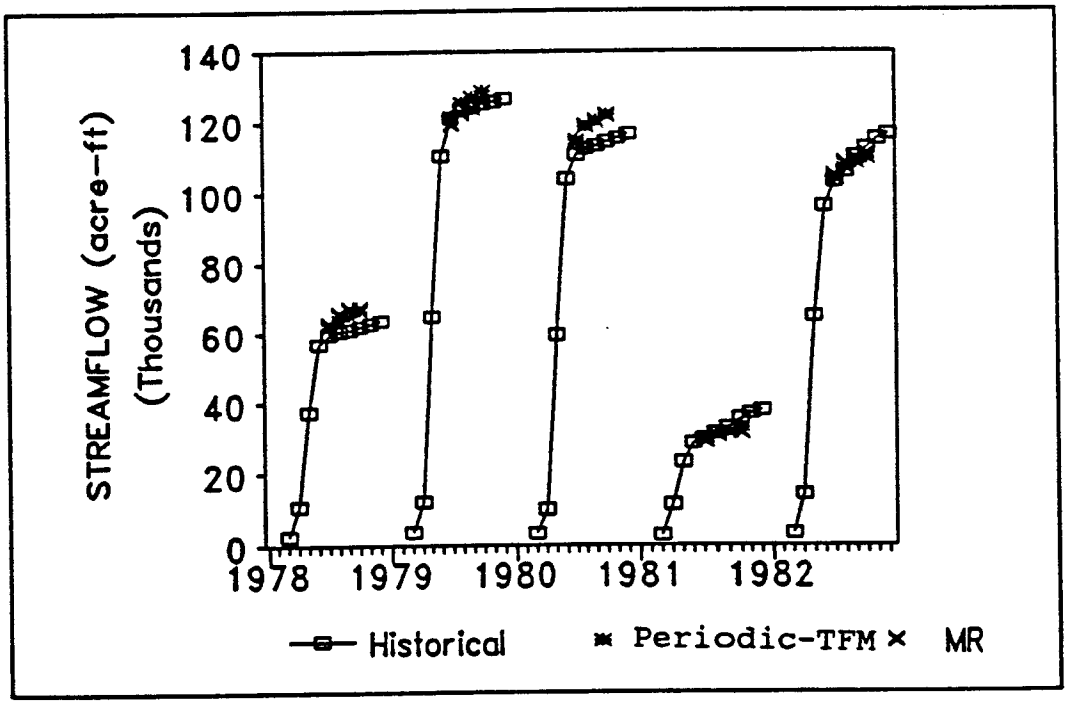


Fig. 7.67 July 1st forecasting of the cumulative flow (Los Pinos River Basin)

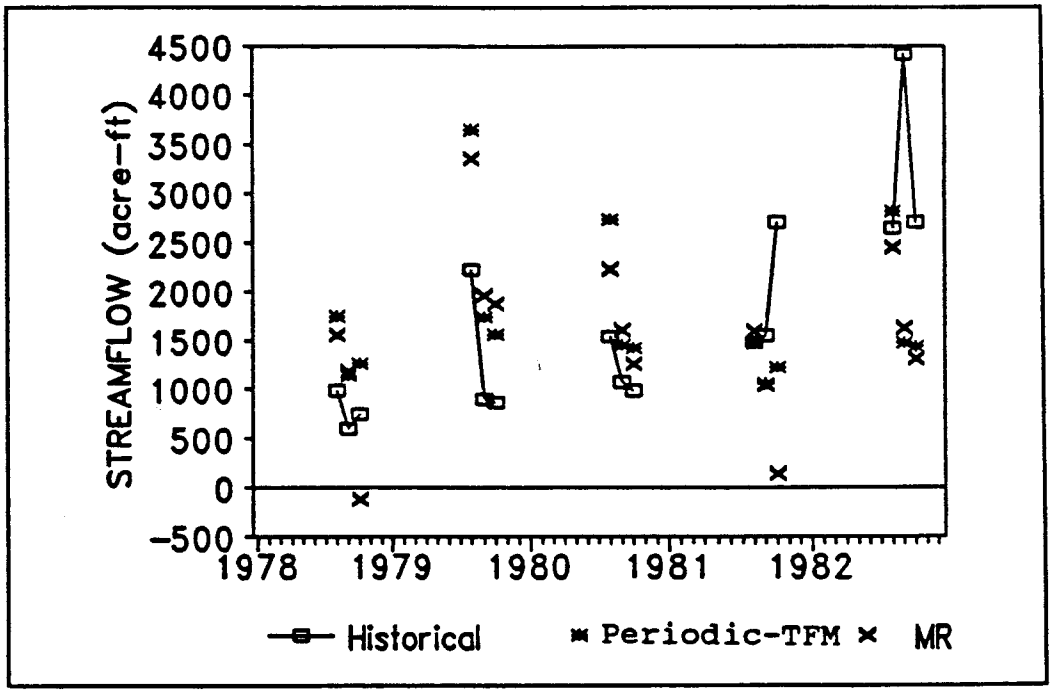


Fig. 7.68 August 1st forecasting of the monthly flow (Los Pinos River Basin)

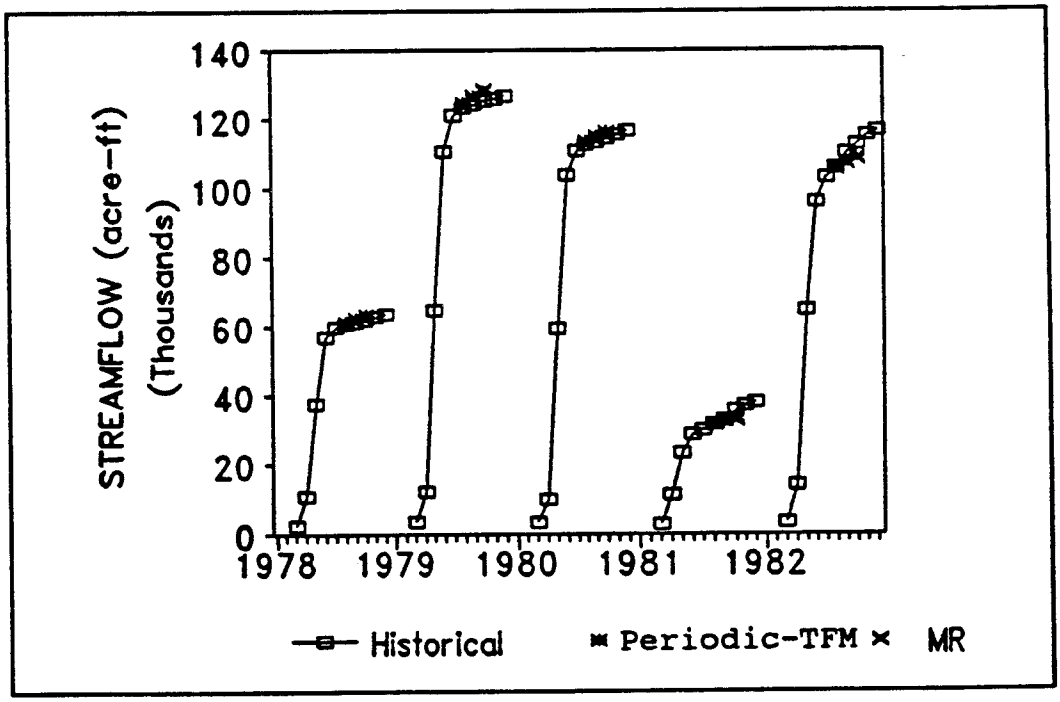


Fig. 7.69 August 1st forecasting of the cumulative flow (Los Pinos River Basin)

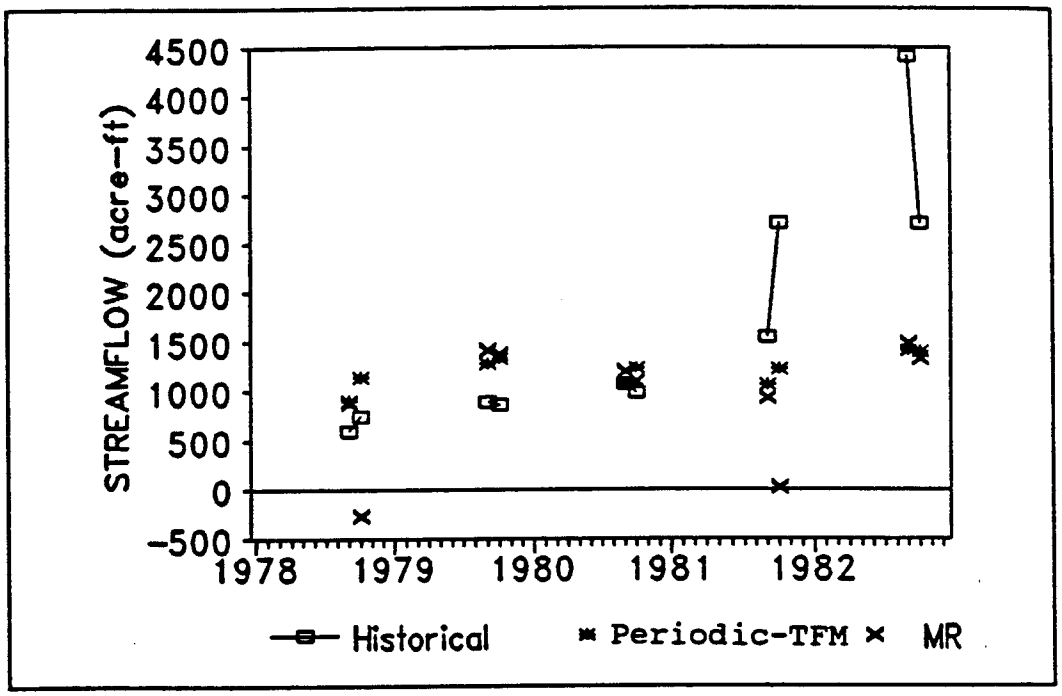


Fig. 7.70 September 1st forecasting of the monthly flow (Los Pinos River Basin)

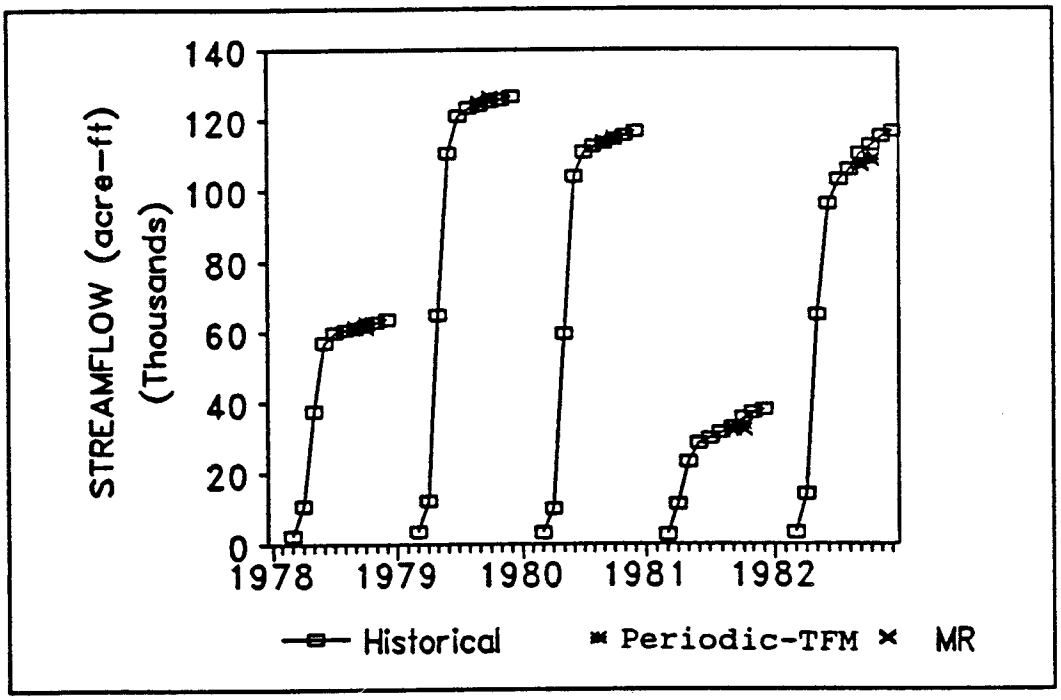


Fig. 7.71 September 1st forecasting of the cumulative flow (Los Pinos River Basin)

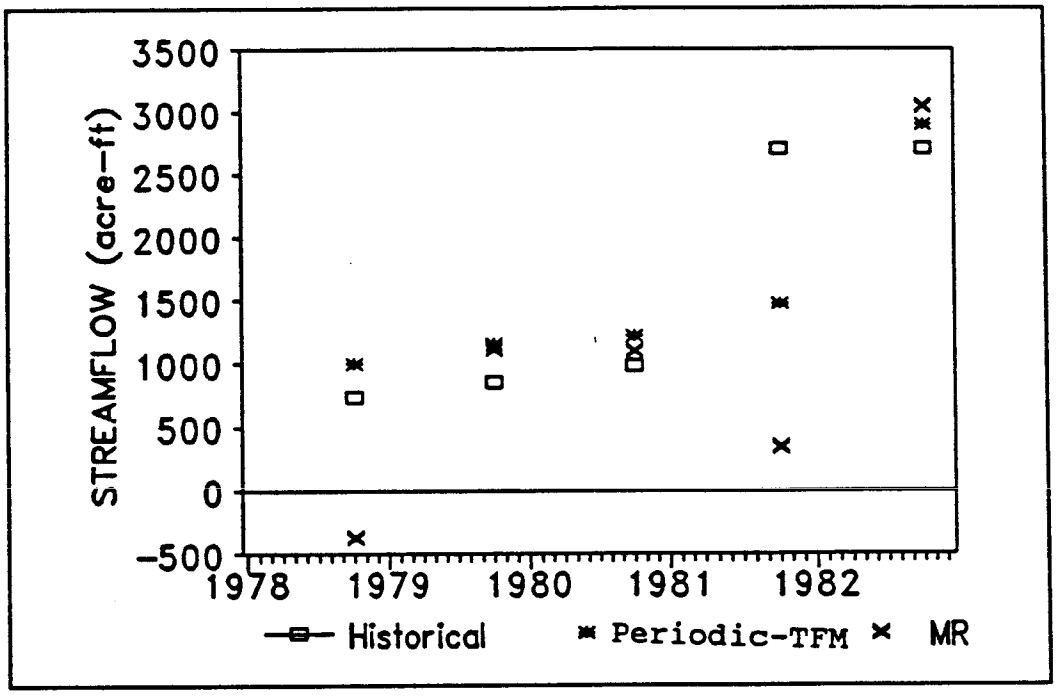


Fig. 7.72 October 1st forecasting of the monthly flow (Los Pinos River Basin)

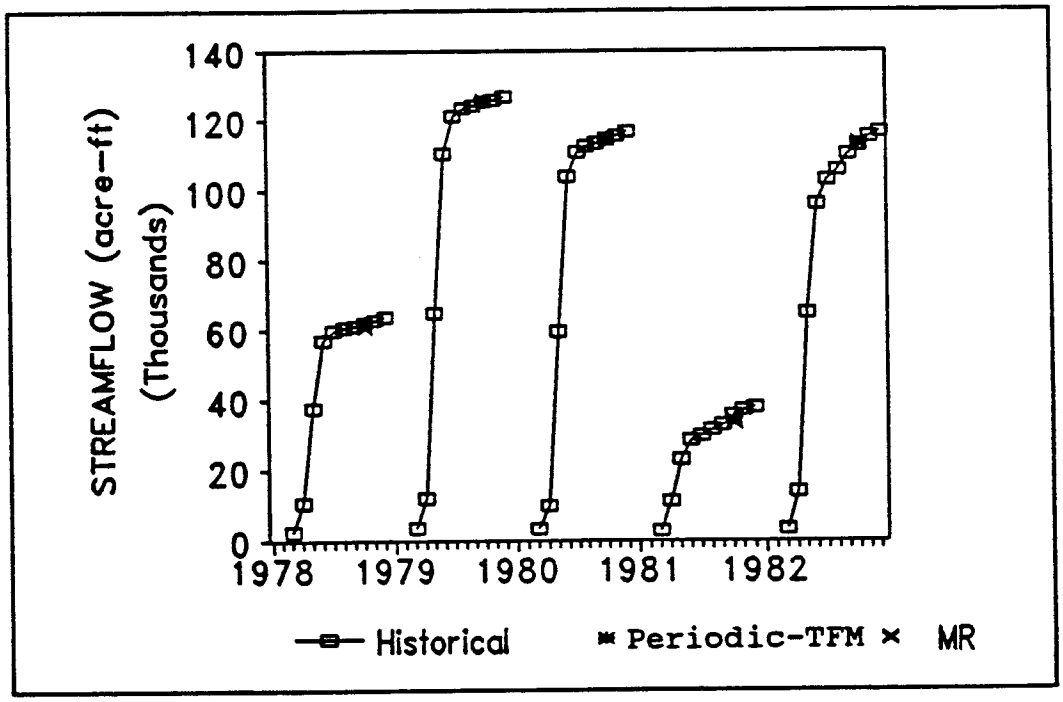


Fig. 7.73 October 1st forecasting of the cumulative flow (Los Pinos River Basin)

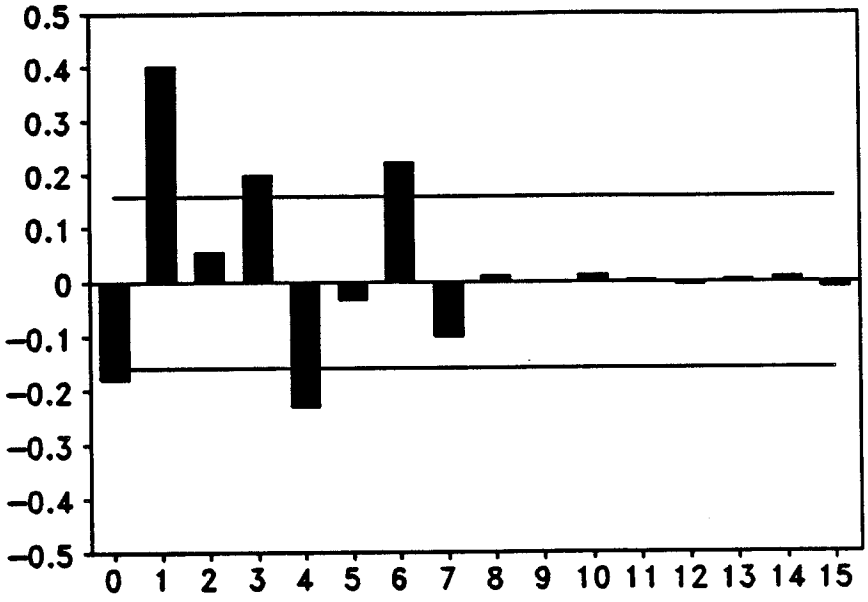


Fig. 7.74 The impulse response function (Conejos River near Mogote)

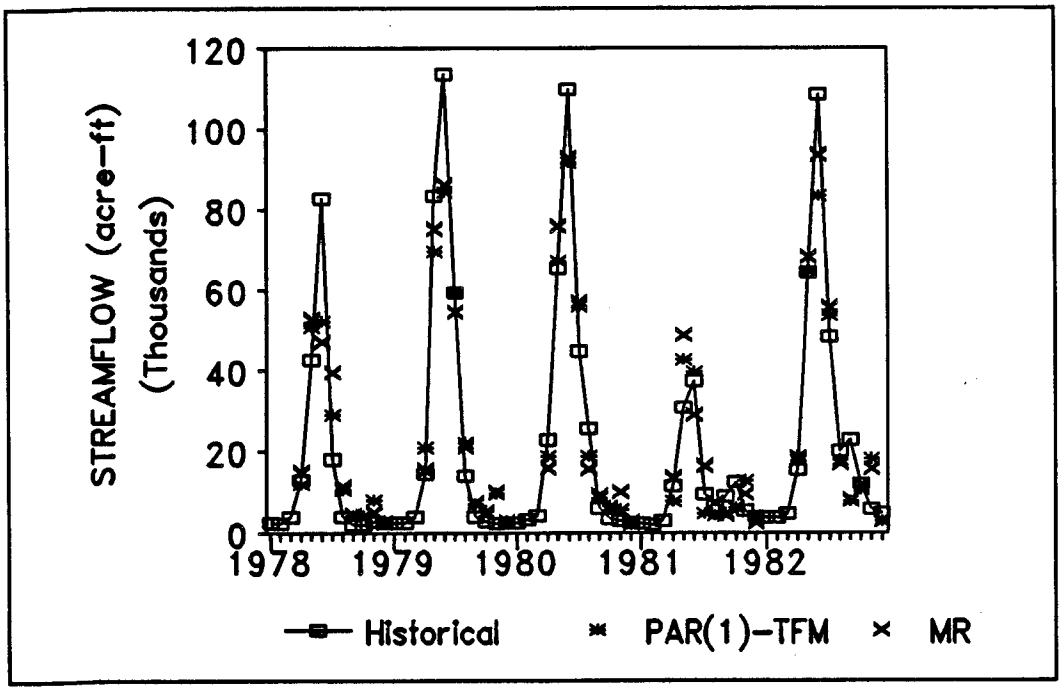


Fig. 7.75 The one step ahead forecast of the monthly flow (Conejos River near Mogote)

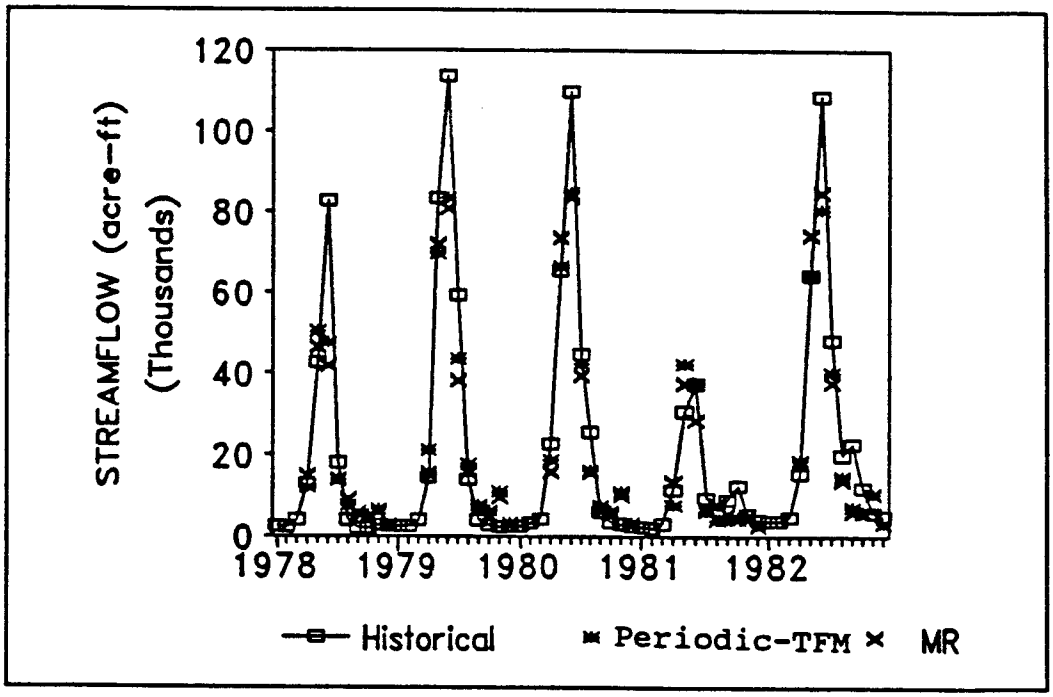


Fig. 7.76 April 1st forecasting of the monthly flow (Conejos River near Mogote)

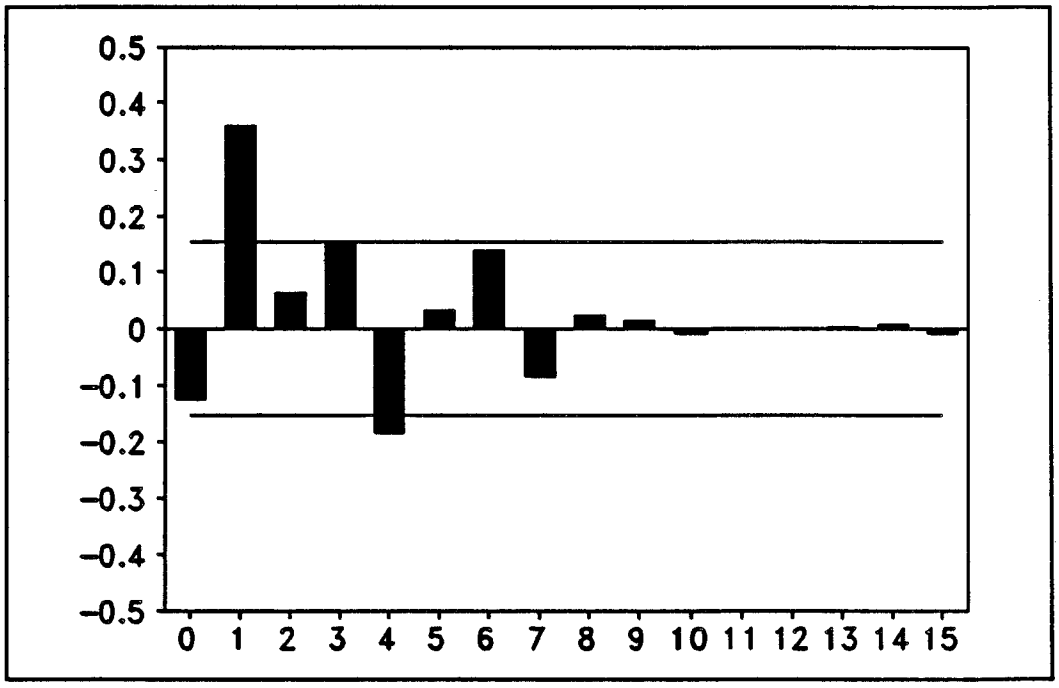


Fig. 7.77 The impulse response function (Conejos River near Mogote; new SWE)

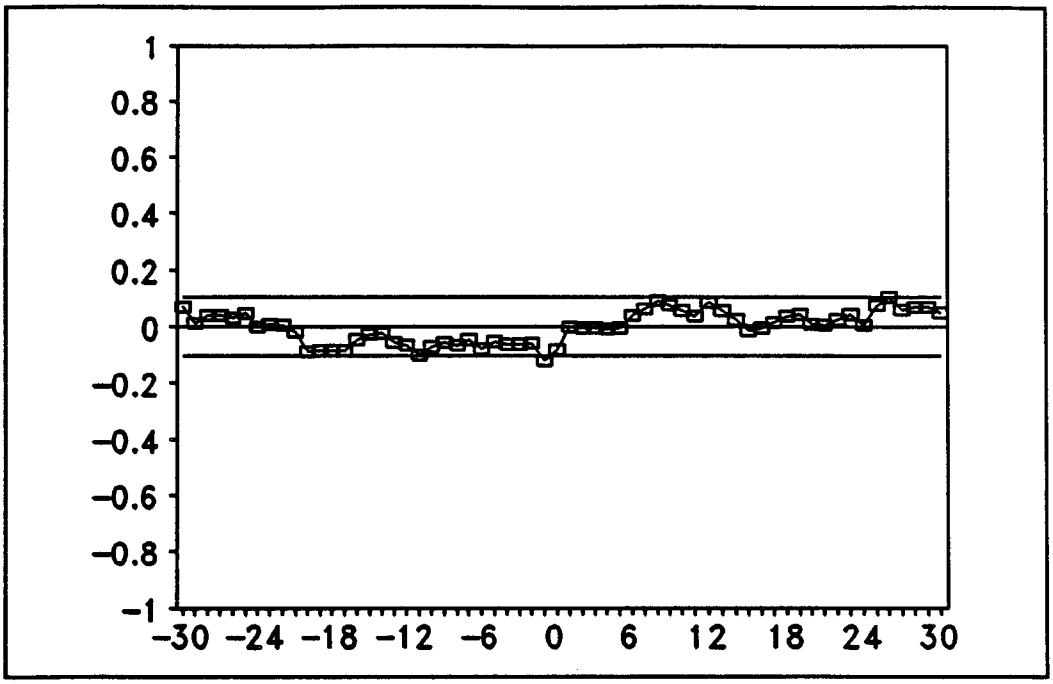


Fig. 7.78 The cross correlation of noise and input (Conejos River near Mogote; new SWE)

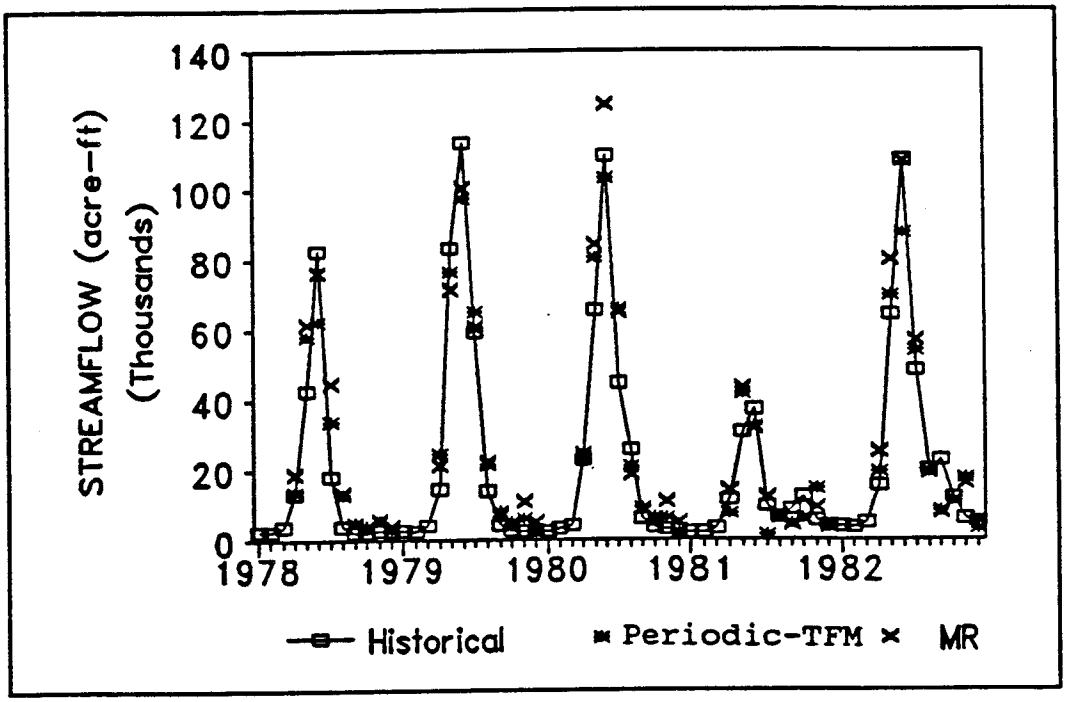


Fig. 7.79 The one step ahead forecast of the monthly flow (Conejos River near Mogote; new SWE)

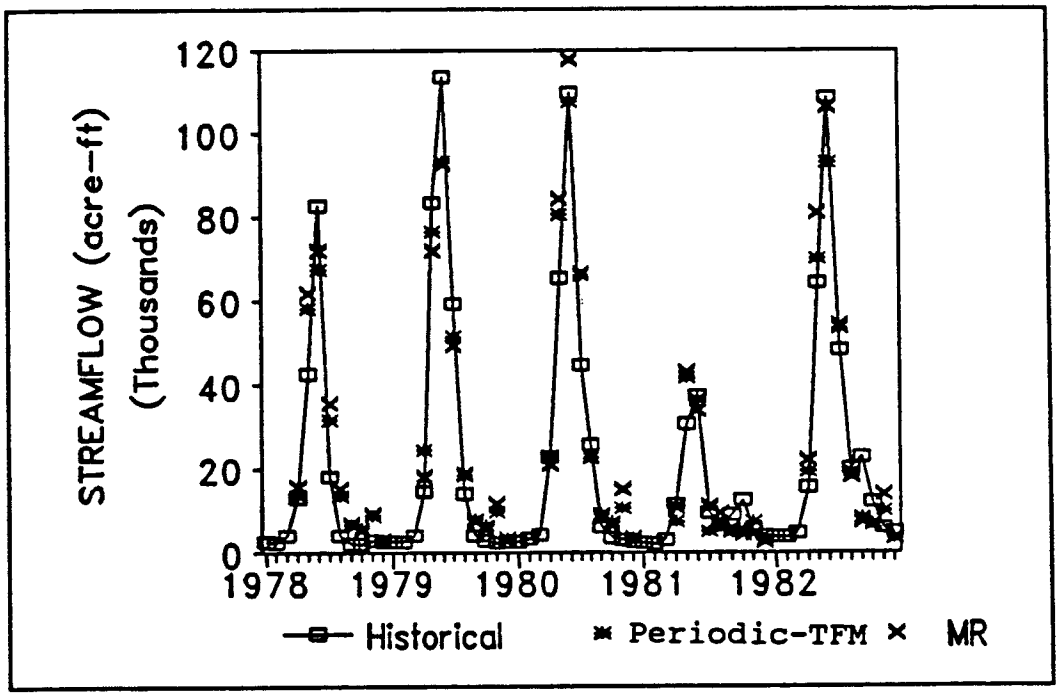


Fig. 7.80 April 1st forecasting of the monthly flow (Conejos River near Mogote; new SWE)

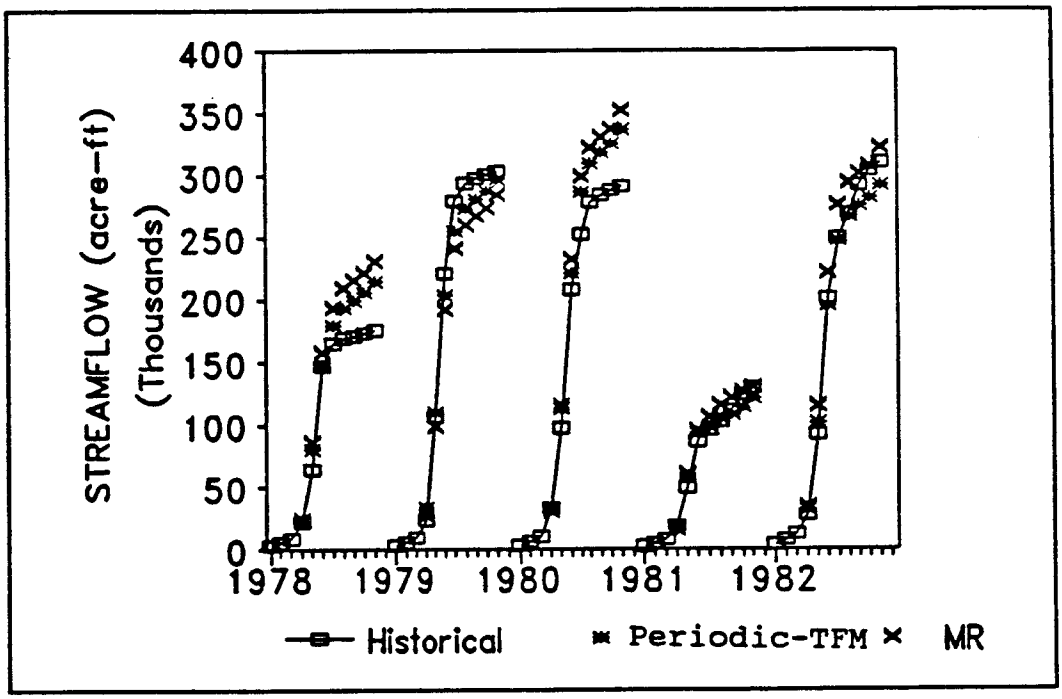


Fig. 7.81 April 1st forecasting of the cumulative flow (Conejos River near Mogote; new SWE)

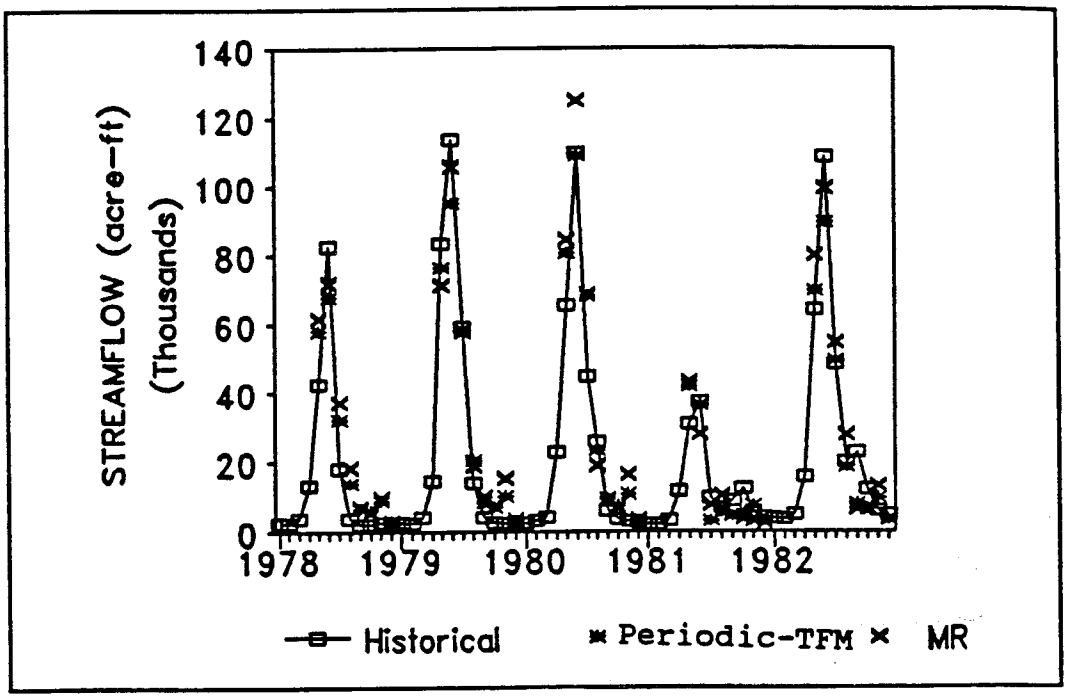


Fig. 7.82 May 1st forecasting of the monthly flow (Conejos River near Mogote; new SWE)

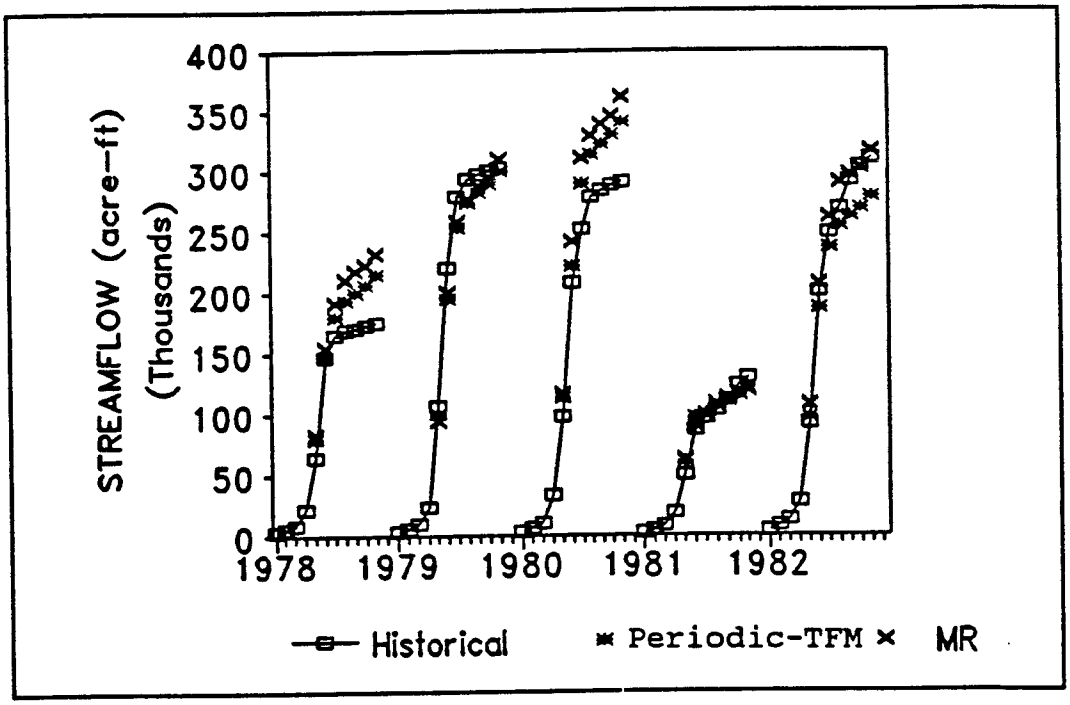


Fig. 7.83 May 1st forecasting of the cumulative flow (Conejos River near Mogote; new SWE)

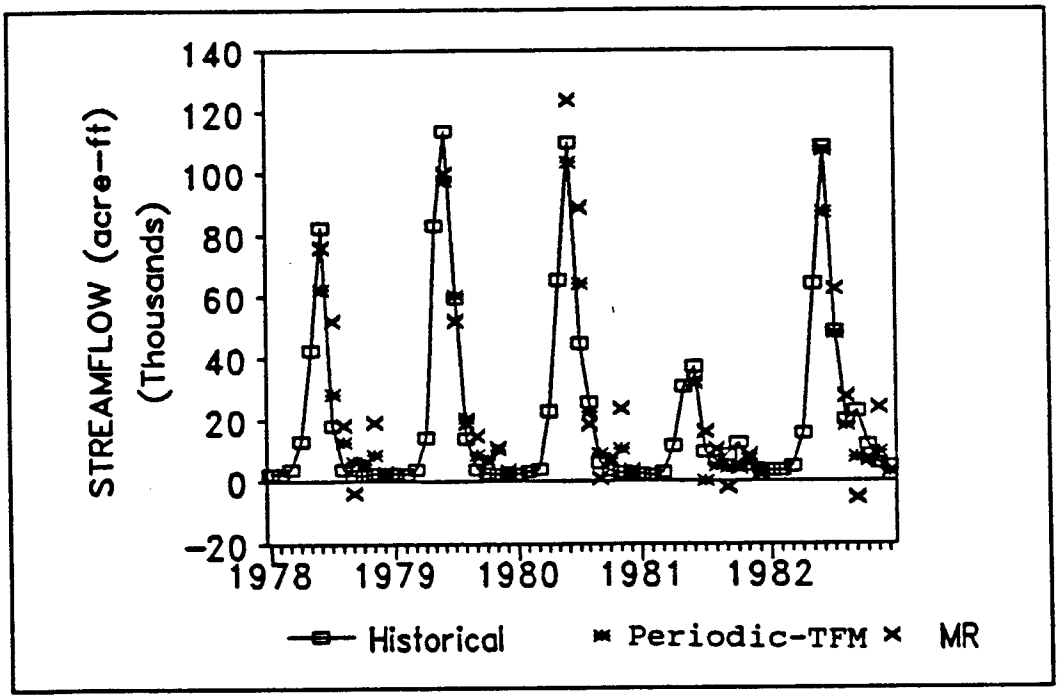


Fig. 7.84 June 1st forecasting of the monthly flow (Conejos River near Mogote; new SWE)

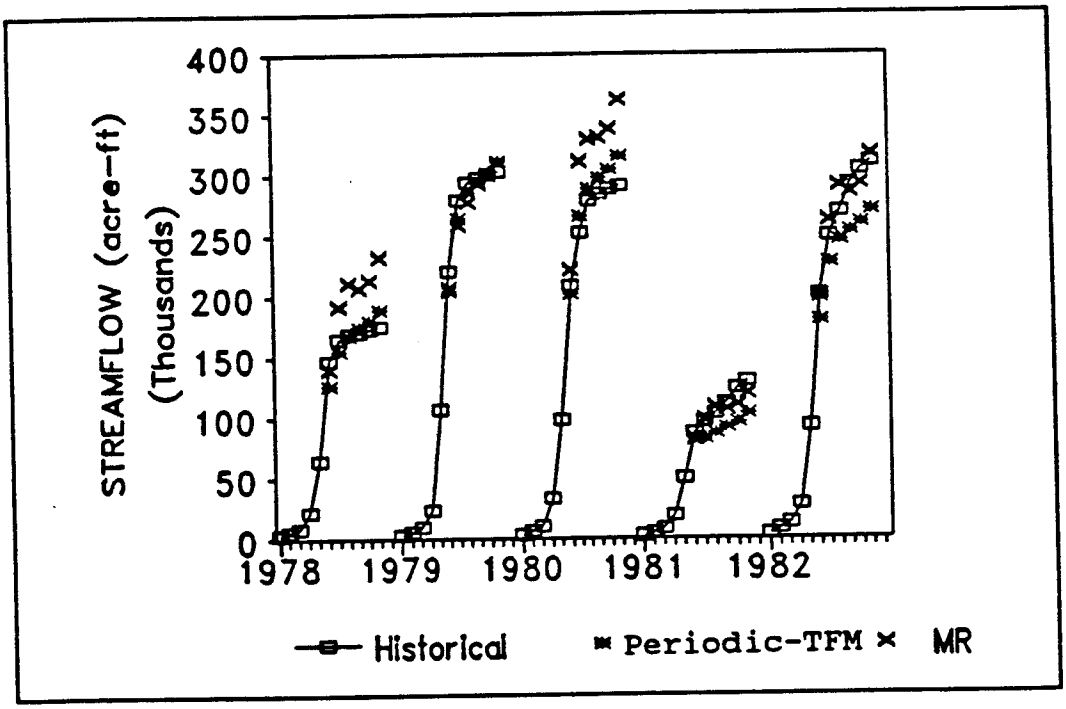


Fig. 7.85 June 1st forecasting of the cumulative flow (Conejos River near Mogote; new SWE)

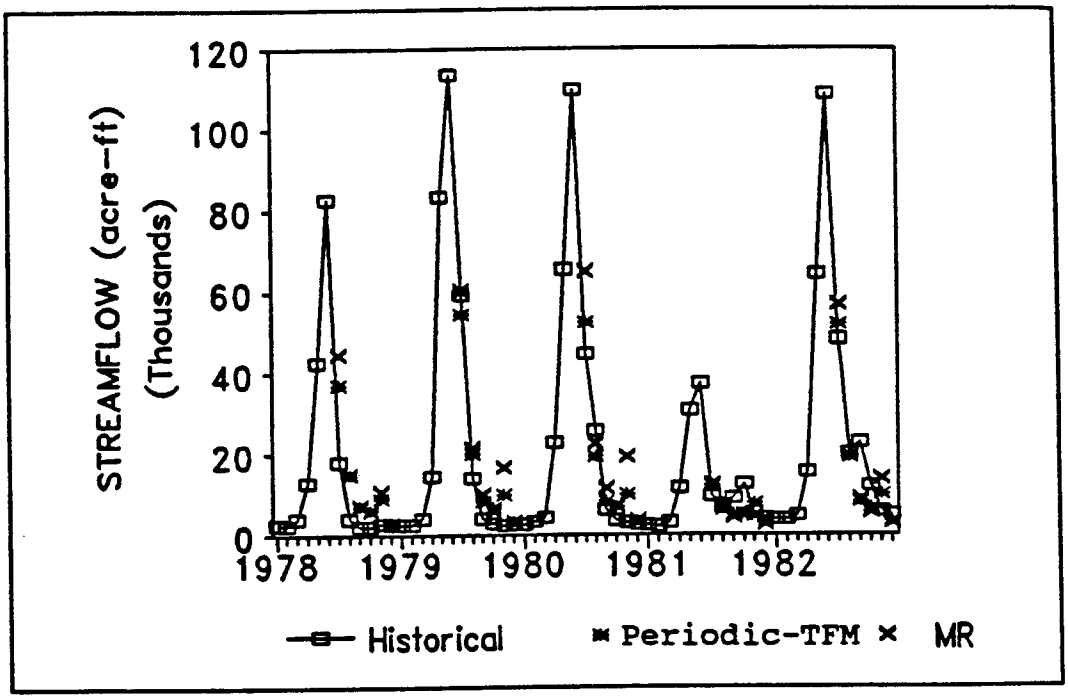


Fig. 7.86 July 1st forecasting of the monthly flow (Conejos River near Mogote; new SWE)

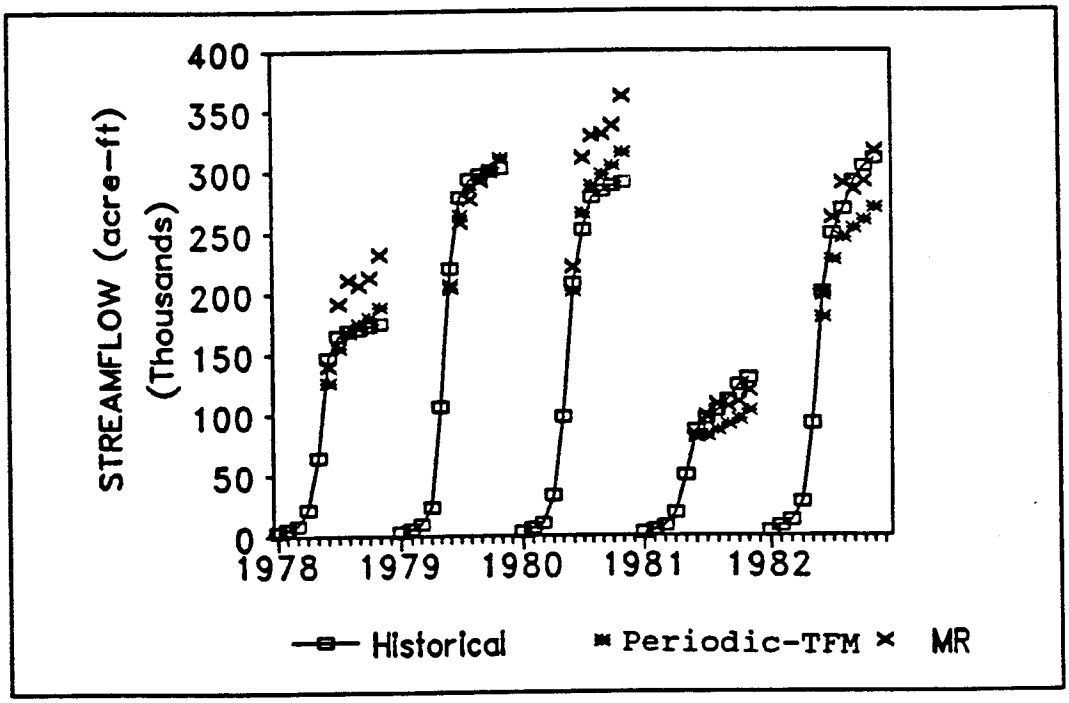


Fig. 7.87 July 1st forecasting of the cumulative flow (Conejos River near Mogote; new SWE)

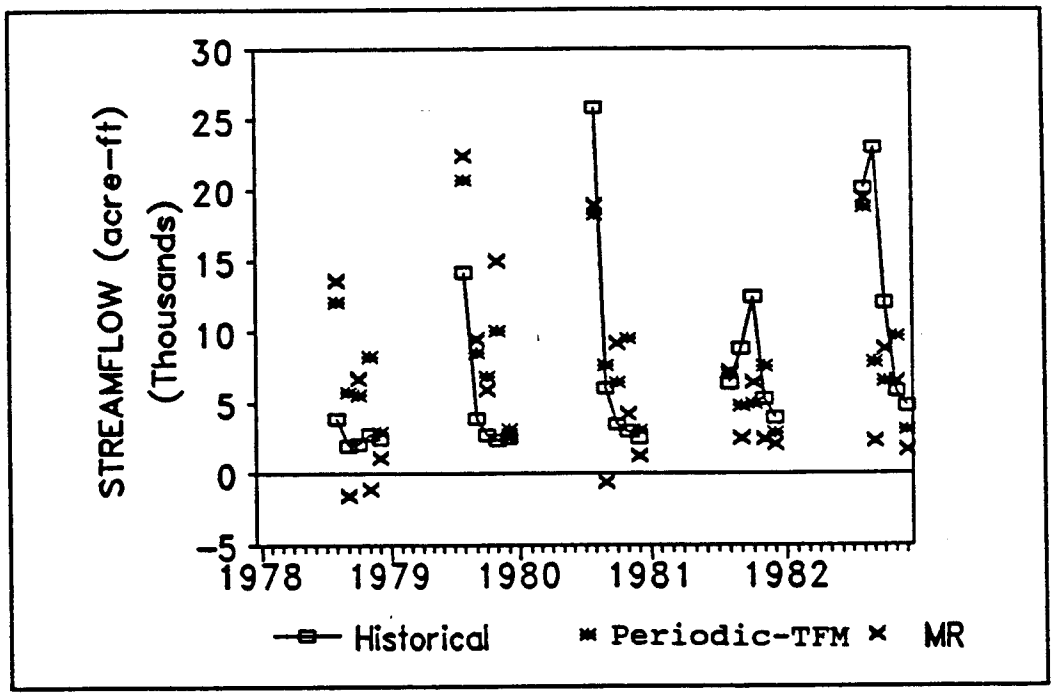


Fig. 7.88 August 1st forecasting of the monthly flow (Los Pinos River near Ortiz; new SWE)

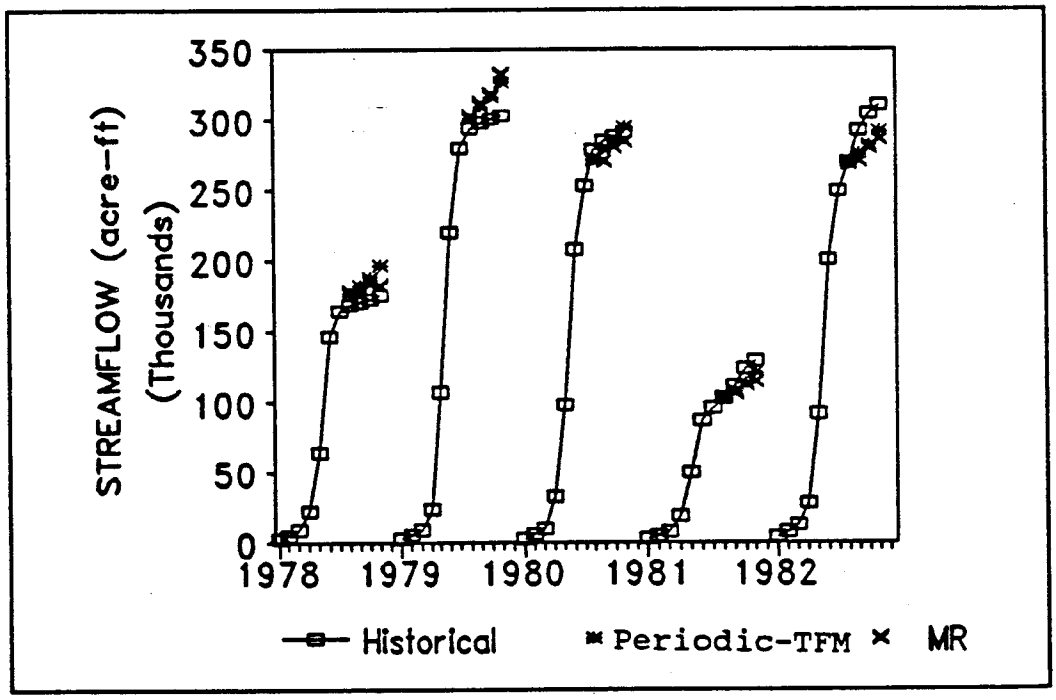


Fig. 7.89 August 1st forecasting of the cumulative flow (Los Pinos River near Ortiz; new SWE)

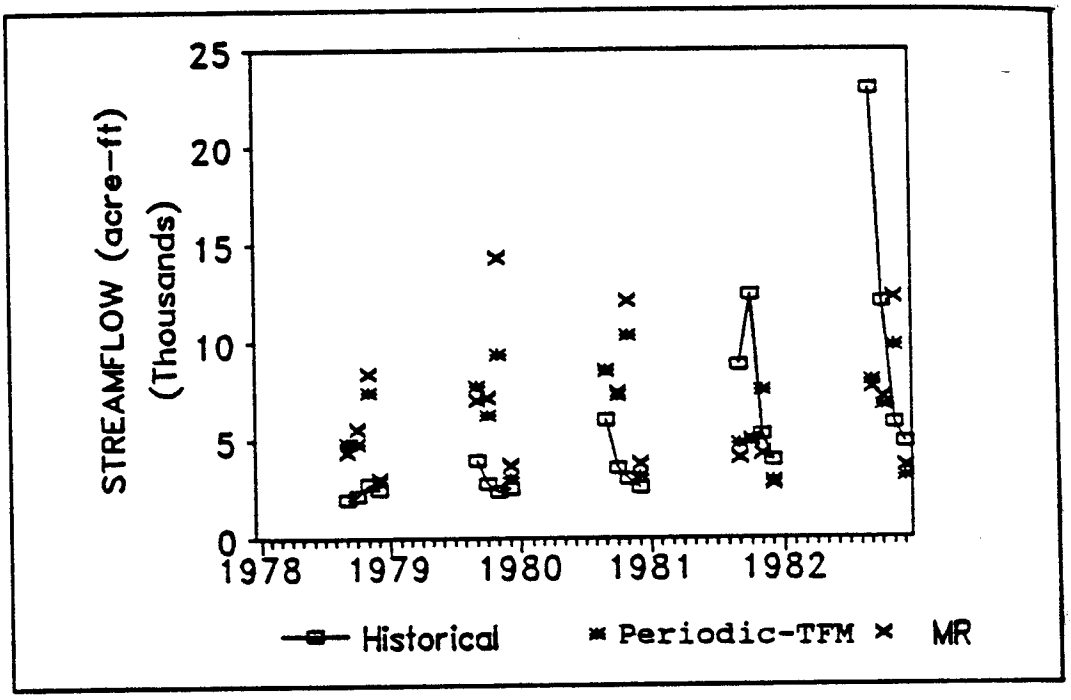


Fig. 7.90 September 1st forecasting of the monthly flow (Los Pinos River near Ortiz; new SWE)

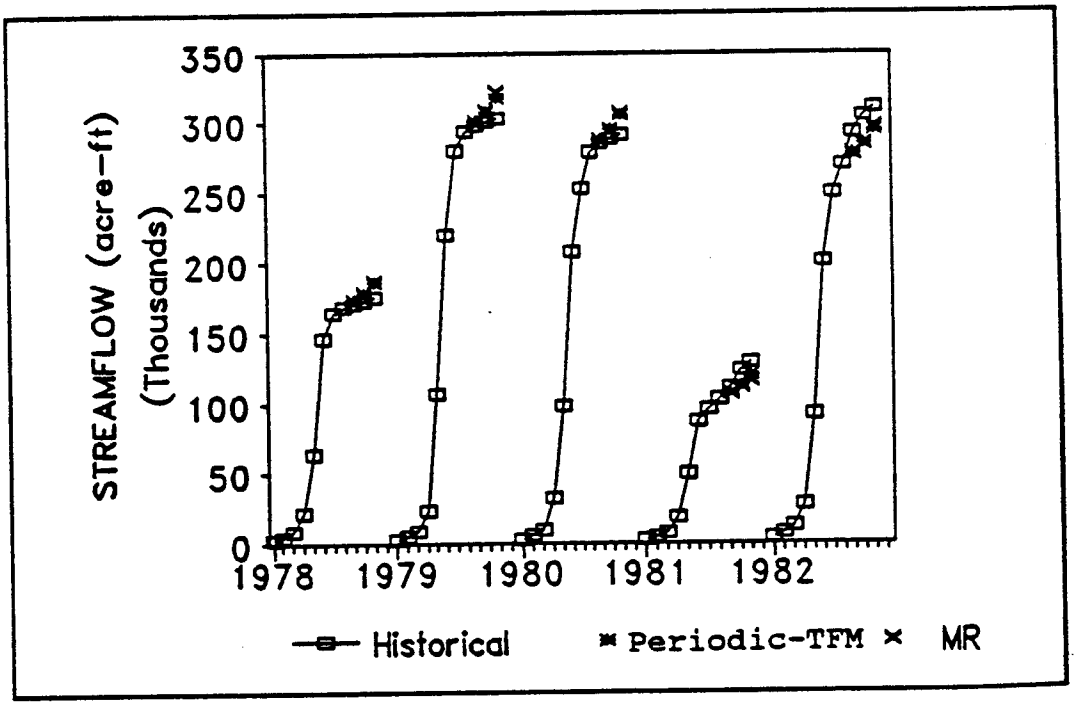


Fig. 7.91 September 1st forecasting of the cumulative flow (Los Pinos River near Ortiz; new SWE)

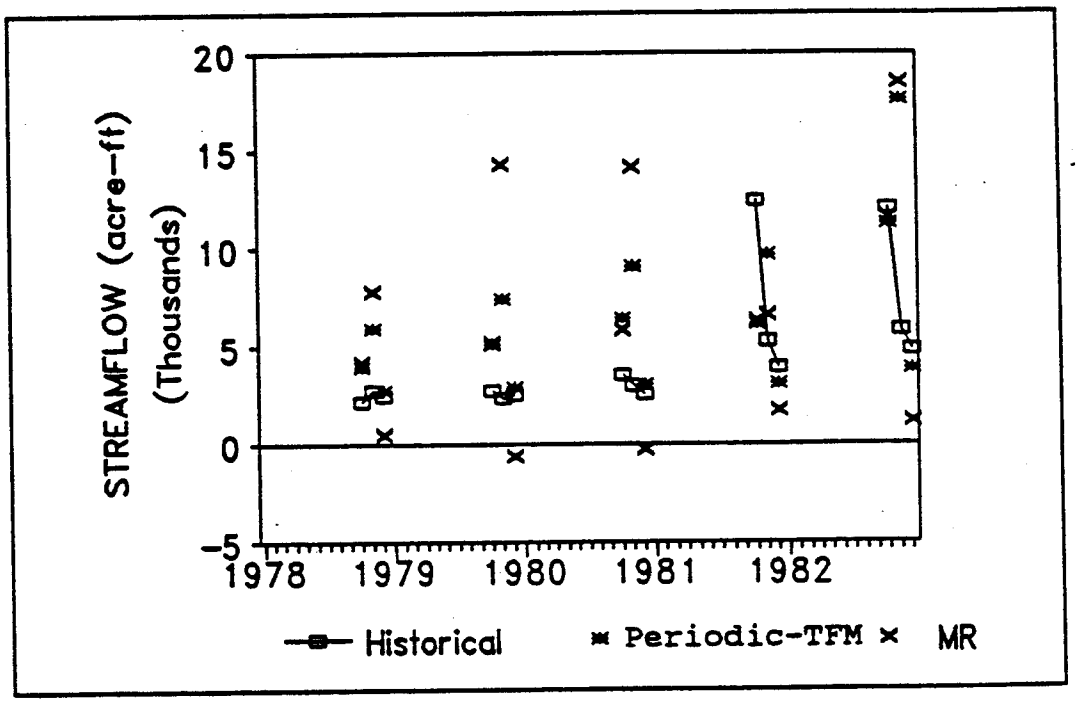


Fig. 7.92 October 1st forecasting of the monthly flow (Los Pinos River near Ortiz; new SWE)

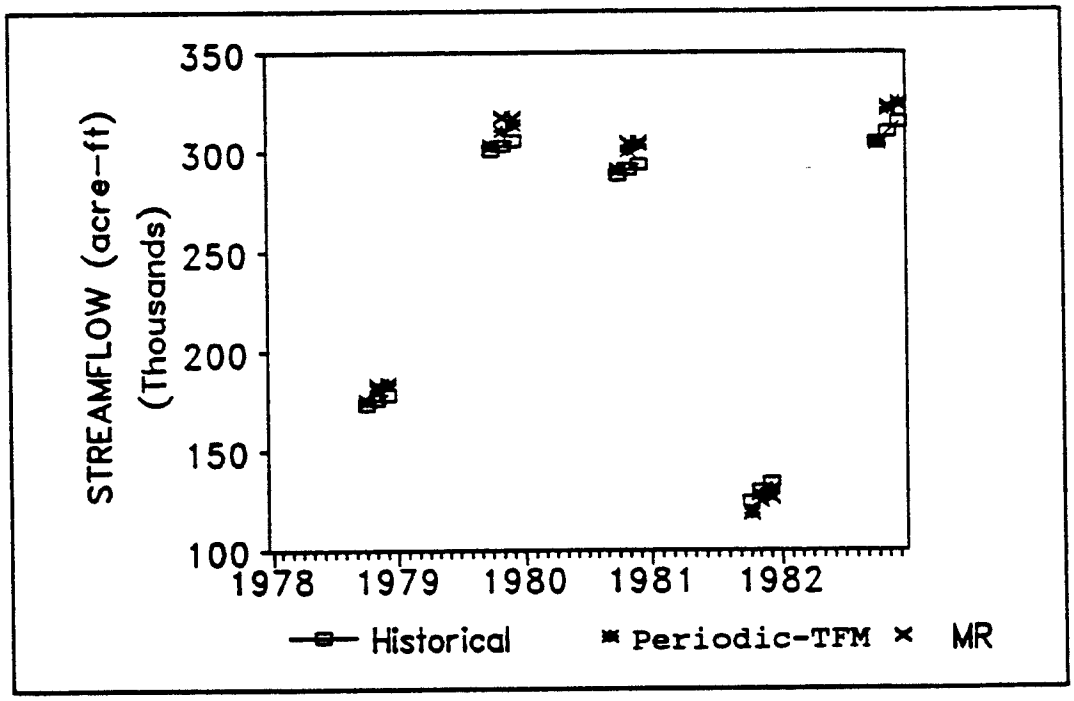


Fig. 7.93 October 1st forecasting of the cumulative flow (Los Pinos River near Ortiz; new SWE)

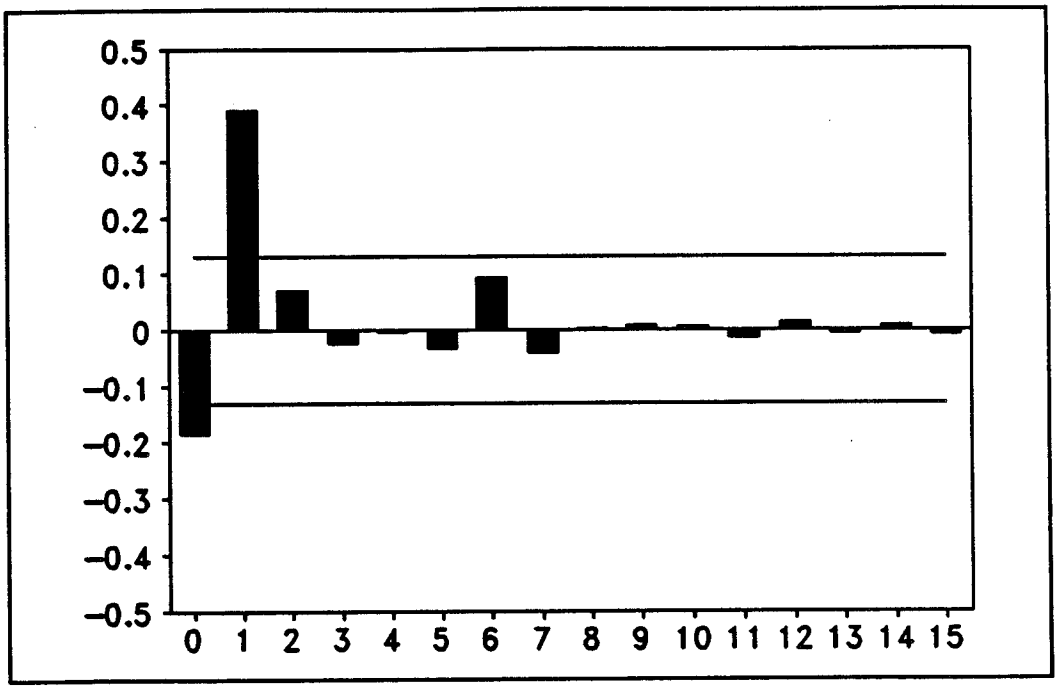


Fig. 7.94 The impulse response function (Rio Grande River near Del Norte)

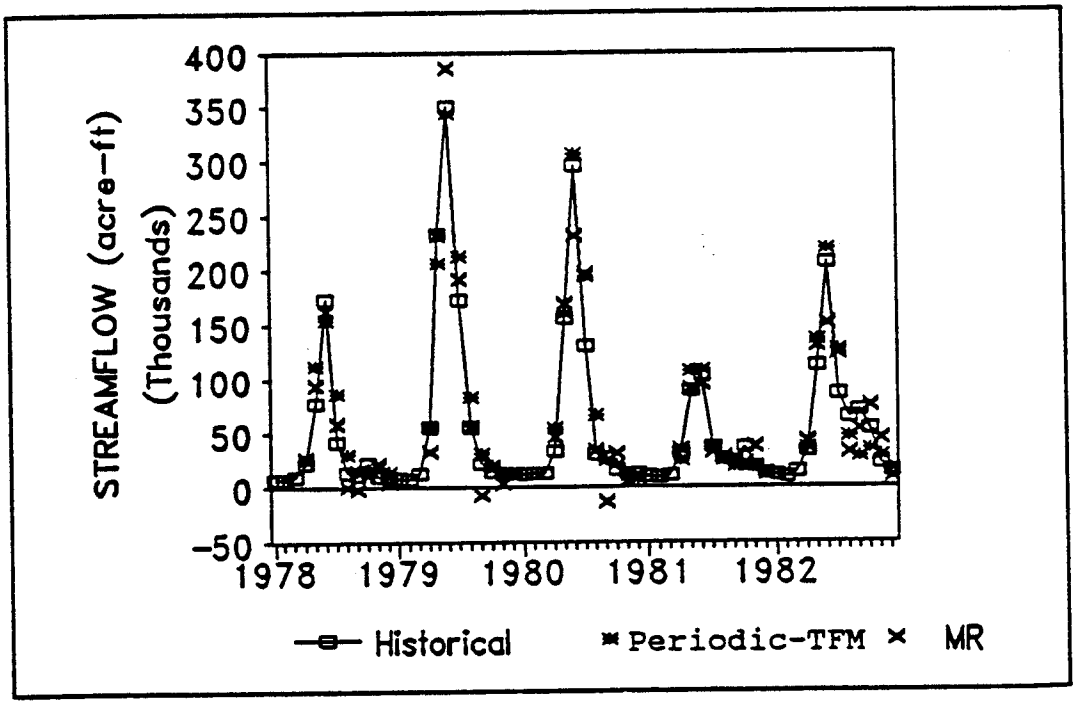


Fig. 7.95 One step ahead forecasting of the monthly flow (Rio Grande River near Del Norte)

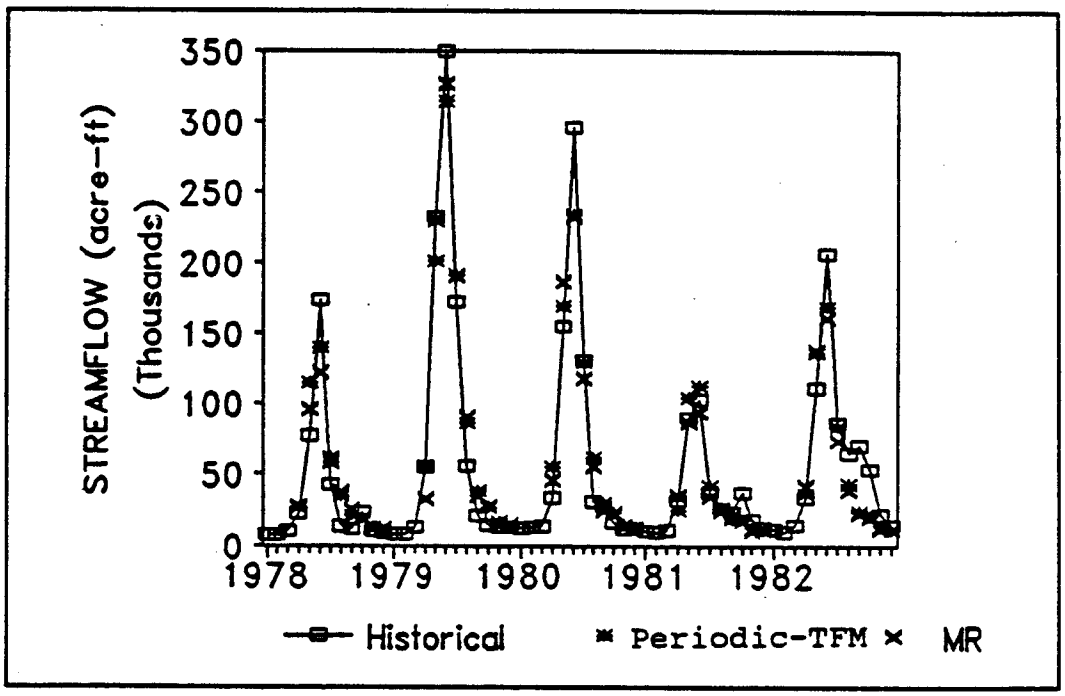


Fig. 7.96 April 1st forecasting of the monthly flow (Rio Grande River near Del Norte)

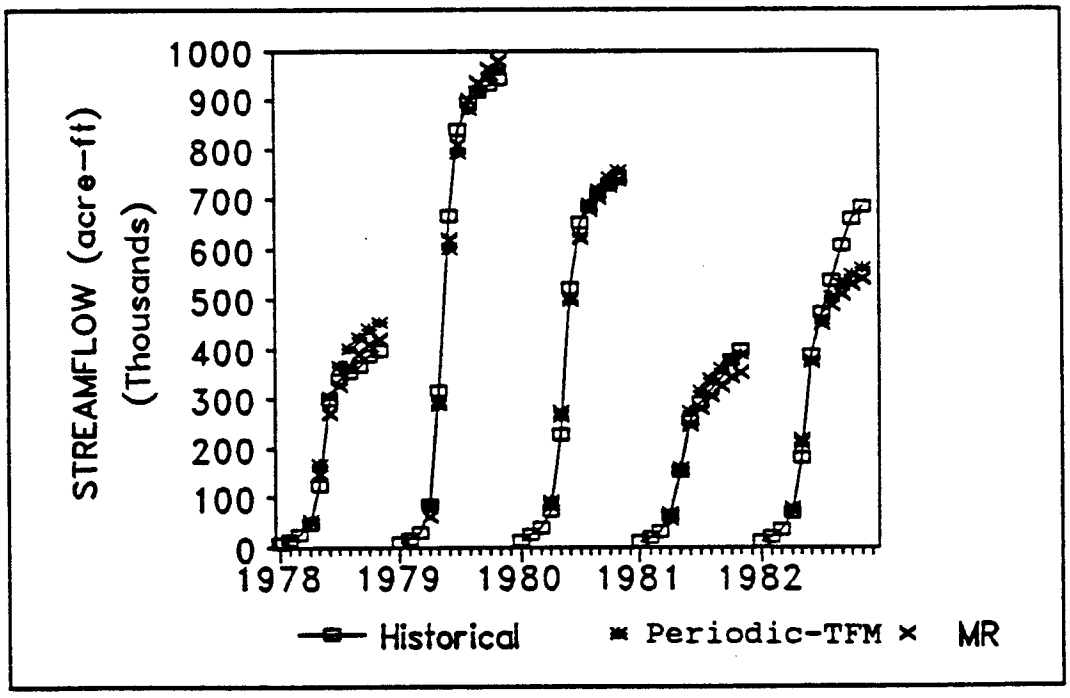


Fig. 7.97 April 1st forecasting of the cumulative flow (Rio Grande River near Del Norte)

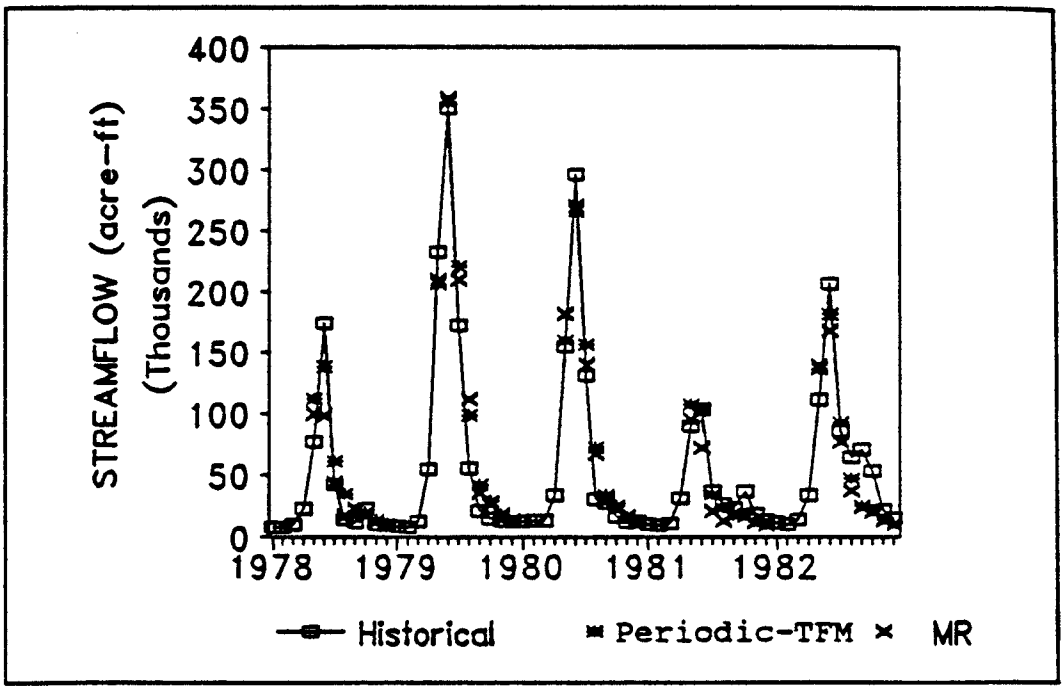


Fig. 7.98 May 1st forecasting of the monthly flow (Rio Grande River near Del Norte)

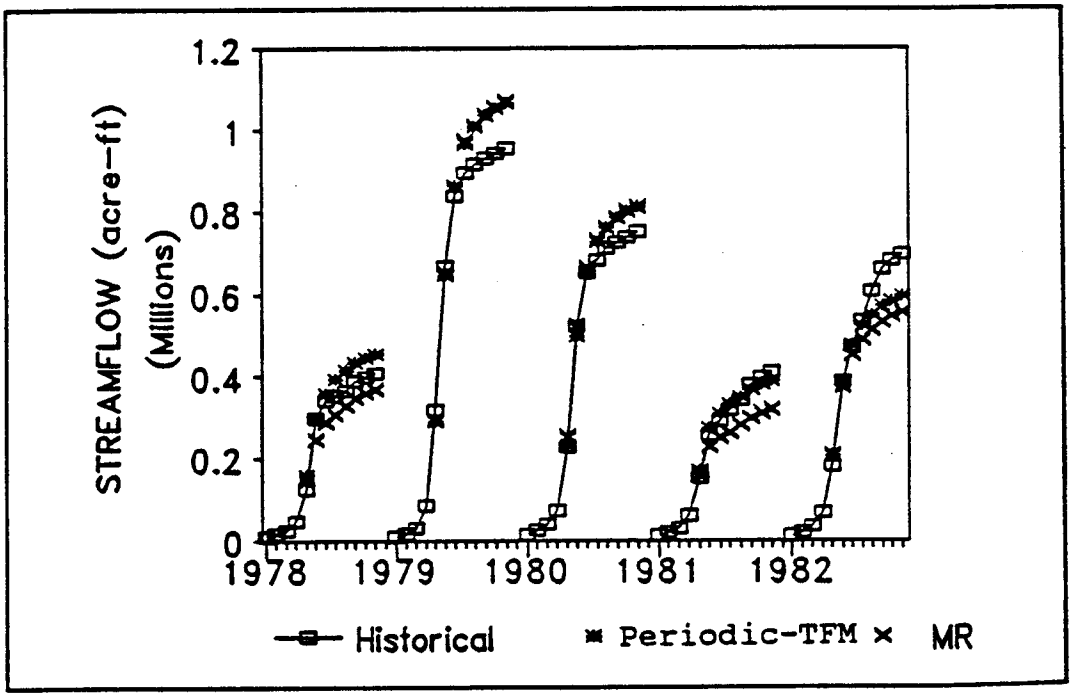


Fig. 7.99 May 1st forecasting of the cumulative flow (Rio Grande River near Del Norte)

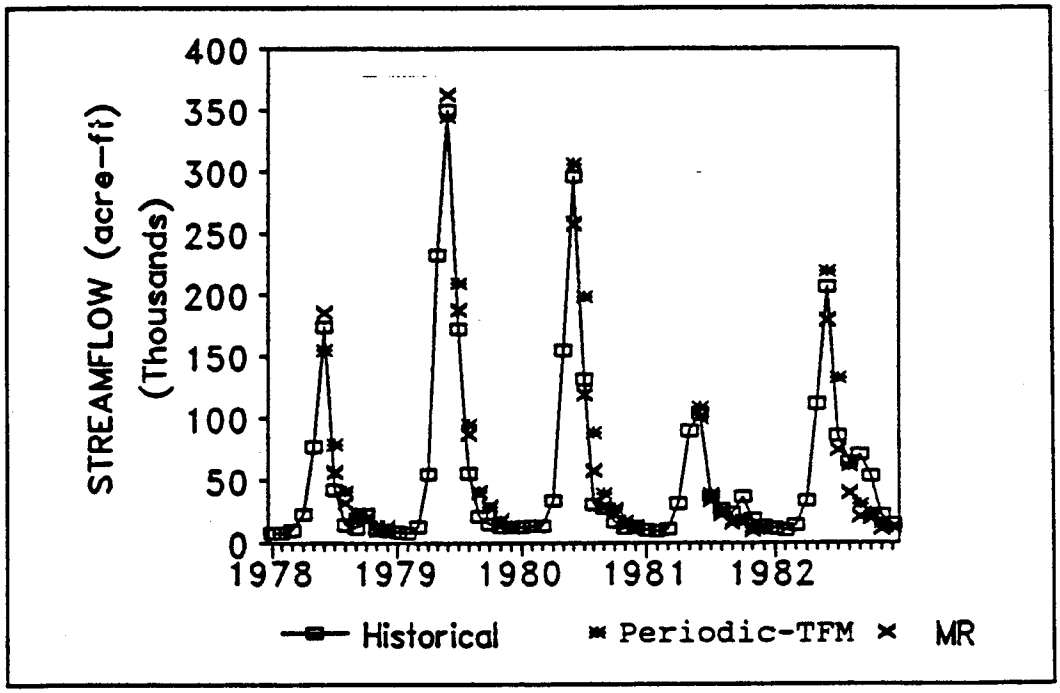


Fig. 7.100 June 1st forecasting of the monthly flow (Rio Grande River near Del Norte)

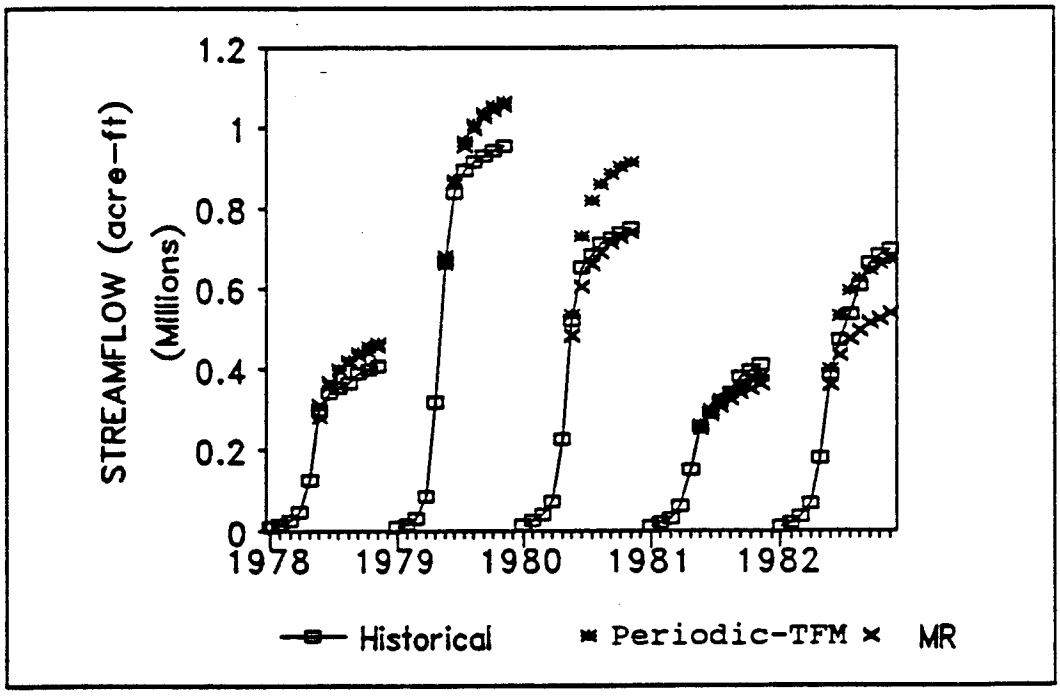


Fig. 7.101 June 1st forecasting of the cumulative flow (Rio Grande River near Del Norte)

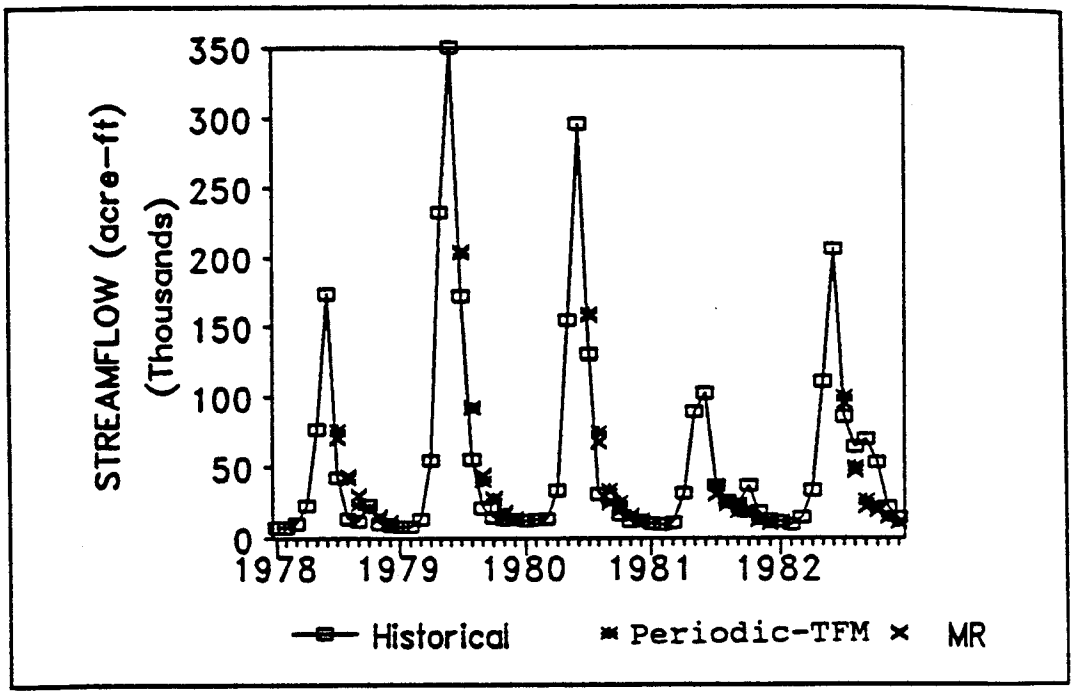


Fig. 7.102 July 1st forecasting of the monthly flow (Rio Grande River near Del Norte)

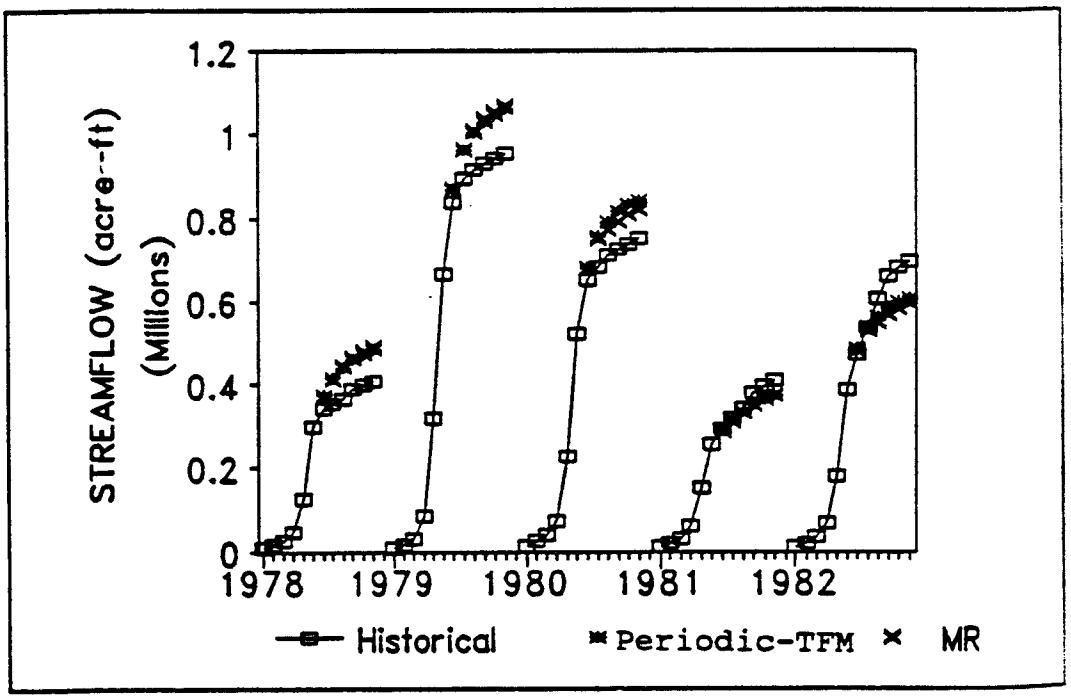


Fig. 7.103 July 1st forecasting of the cumulative flow (Rio Grande River near Del Norte)

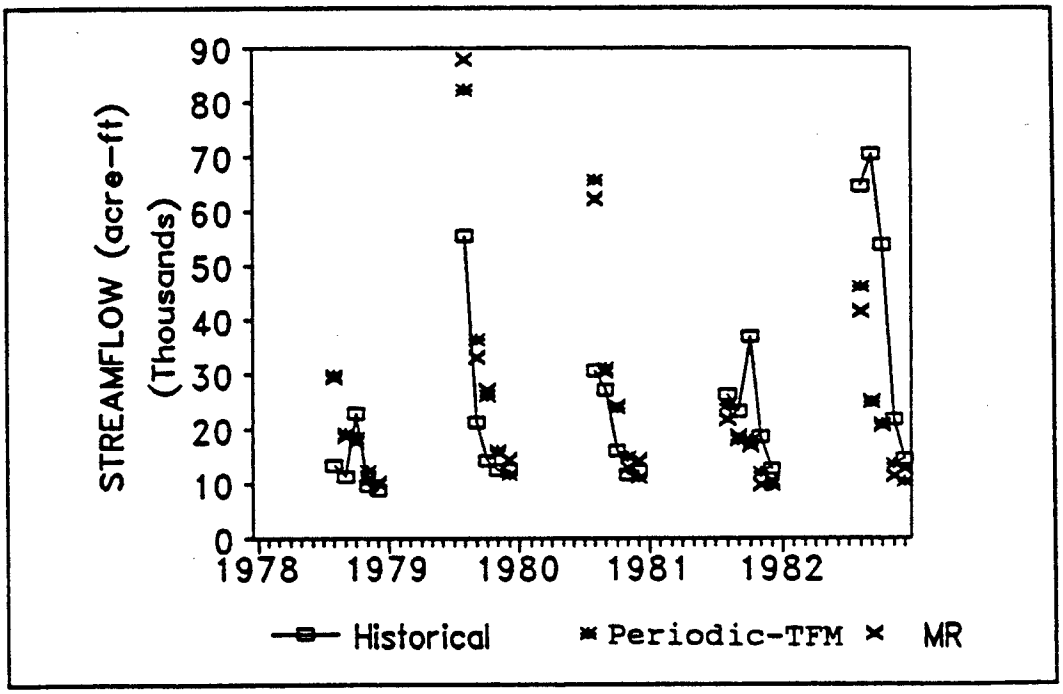


Fig. 7.104 August 1st forecasting of the monthly flow (Rio Grande River near Del Norte)

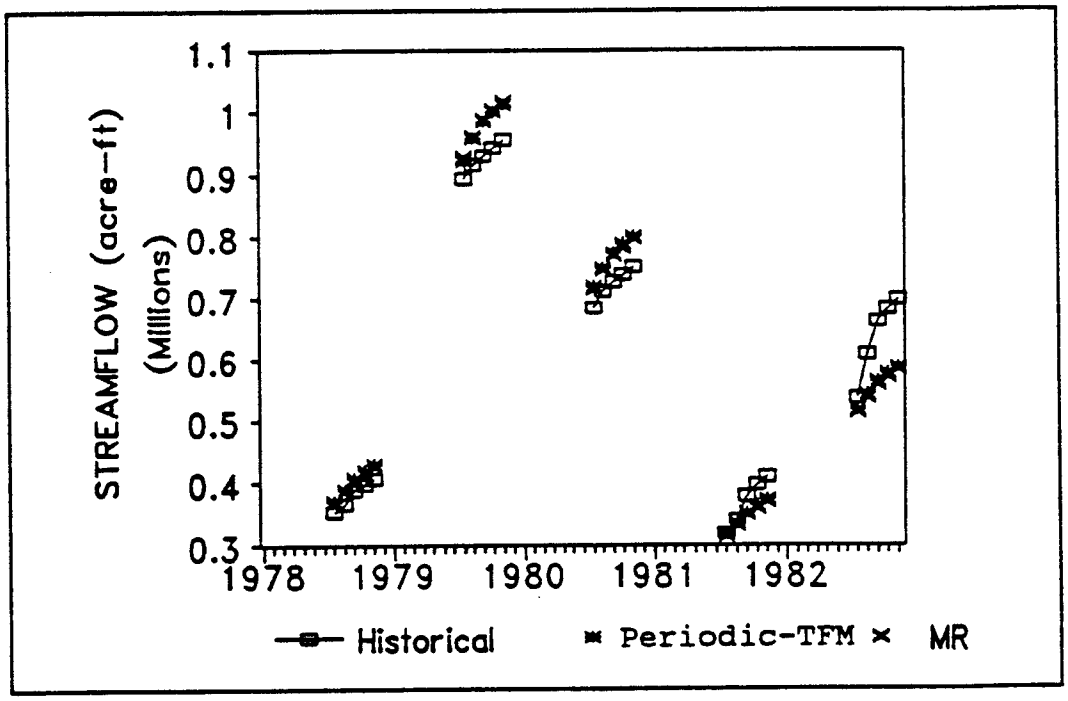


Fig. 7.105 August 1st forecasting of the cumulative flow (Rio Grande River near Del Norte)

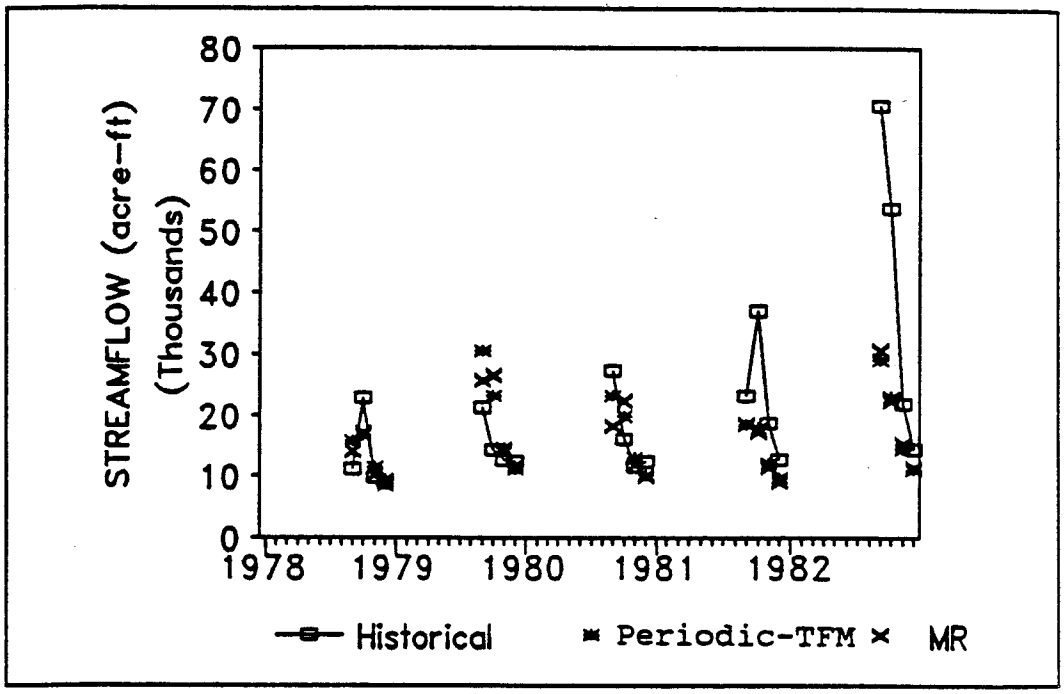


Fig. 7.106 September 1st forecasting of the monthly flow (Rio Grande River near Del Norte)

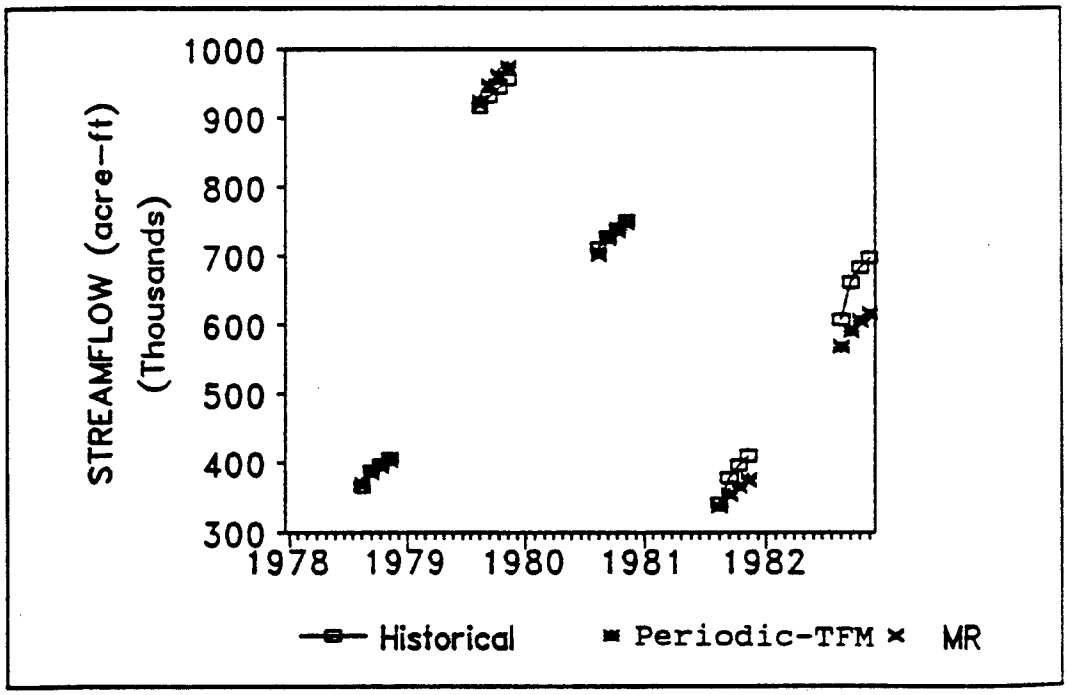


Fig. 7.107 September 1st forecasting of the cumulative flow (Rio Grande River near Del Norte)

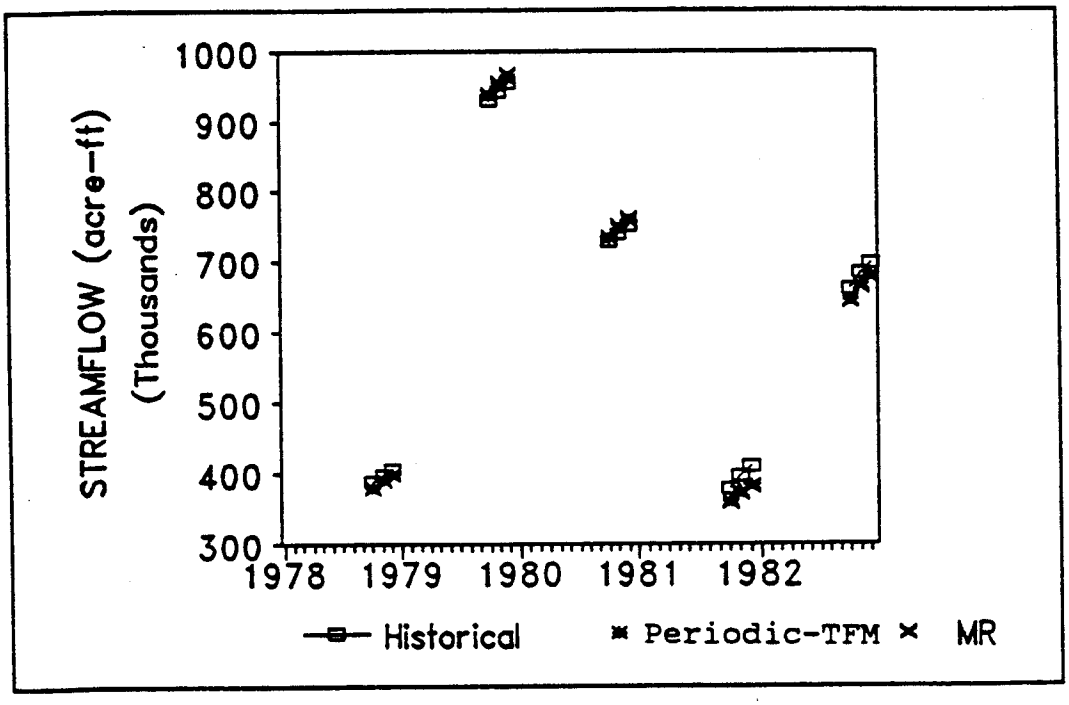


Fig. 7.108 October 1st forecasting of the monthly flow (Rio Grande River near Del Norte)

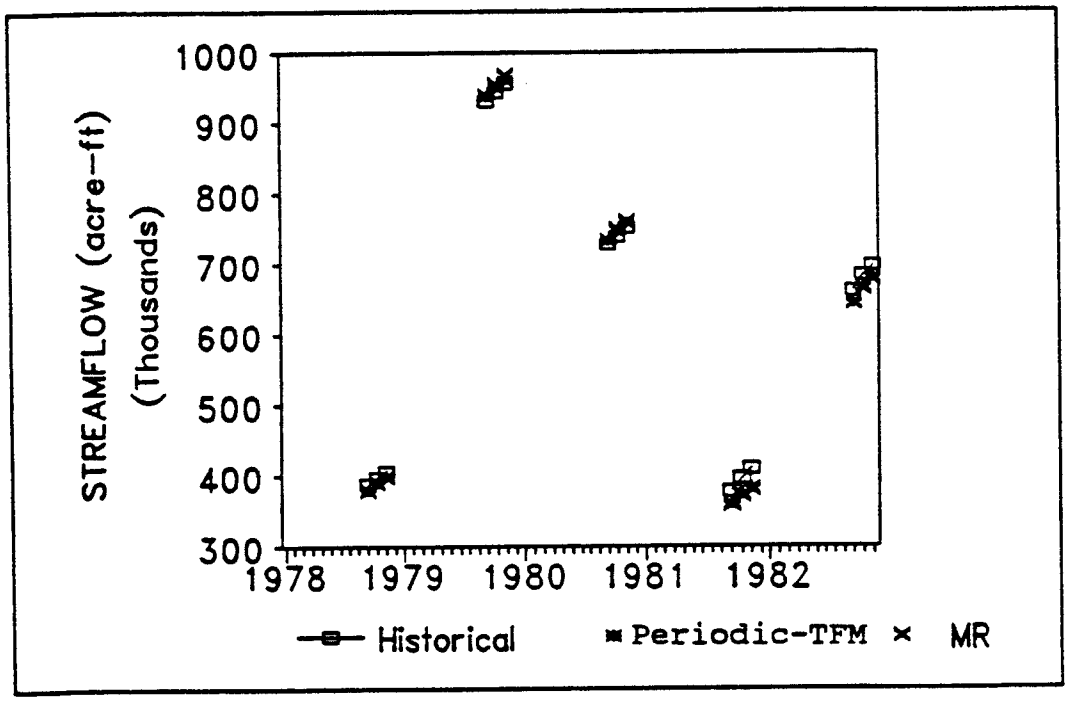


Fig. 7.109 October 1st forecasting of the cumulative flow (Rio Grande River near Del Norte)

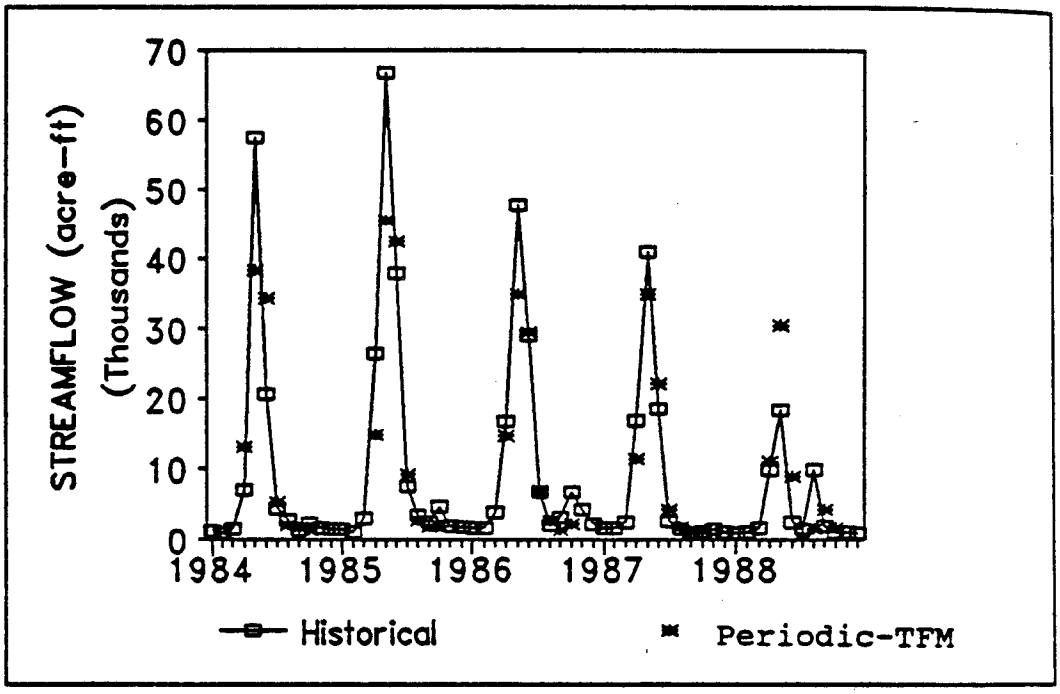


Fig. 7.110 One step ahead forecast of the monthly flow (Los Pinos River near Ortiz)

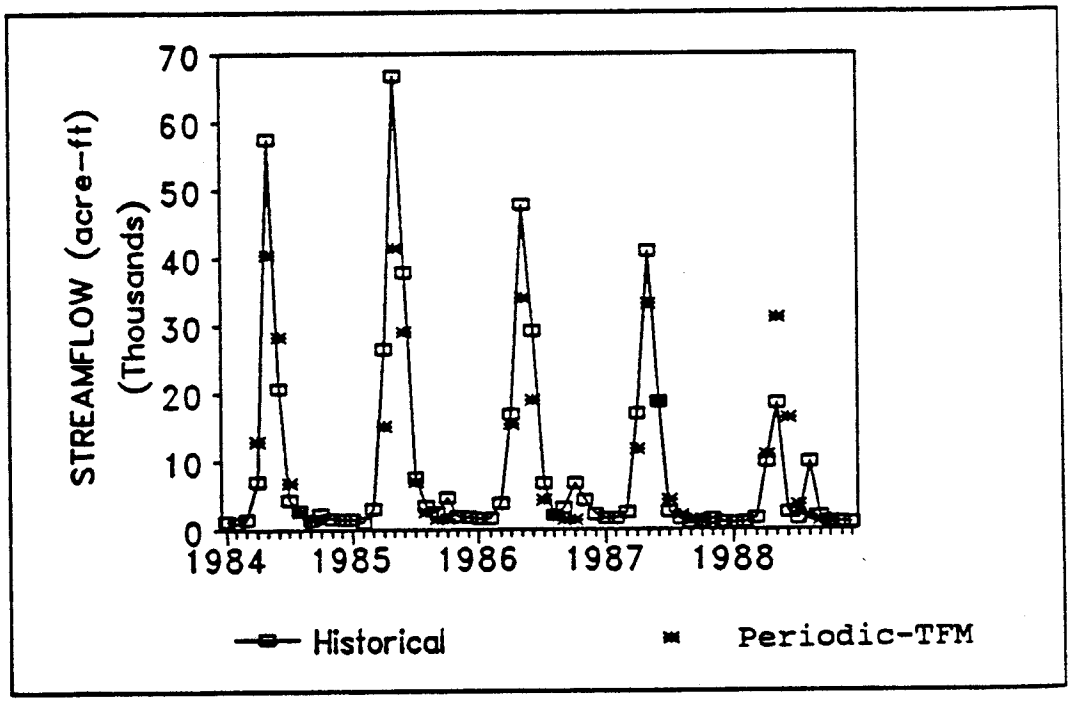


Fig. 7.111 April 1st forecasting of the monthly flow (Los Pinos River near Ortiz)

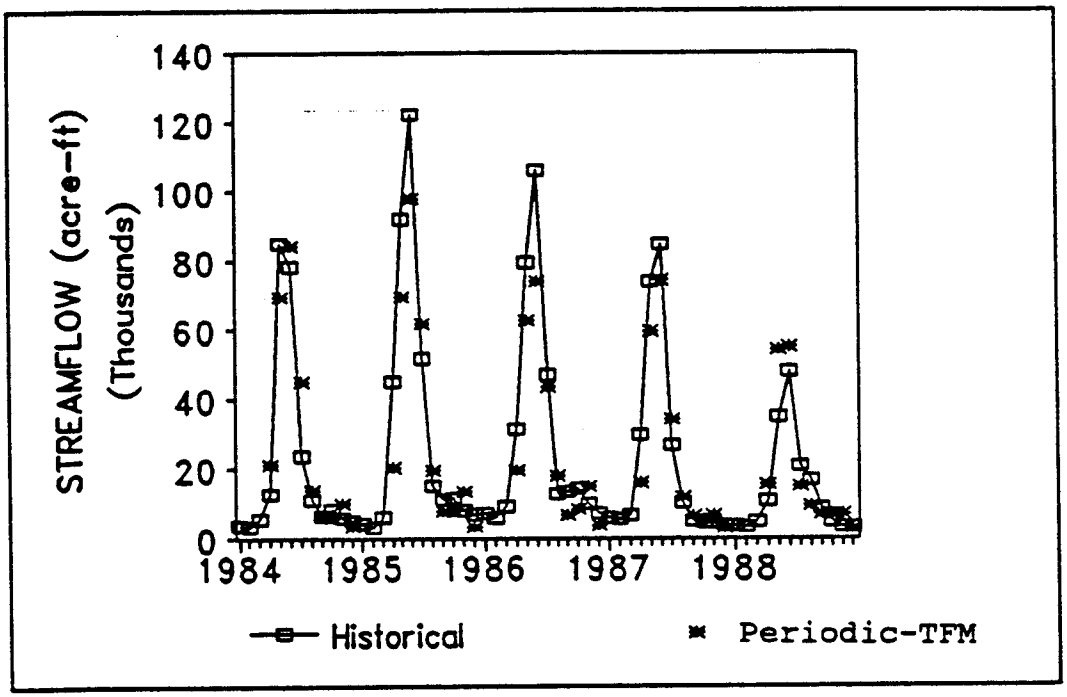


Fig. 7.112 One step ahead forecast of the monthly flow (Conejos River near Mogote)

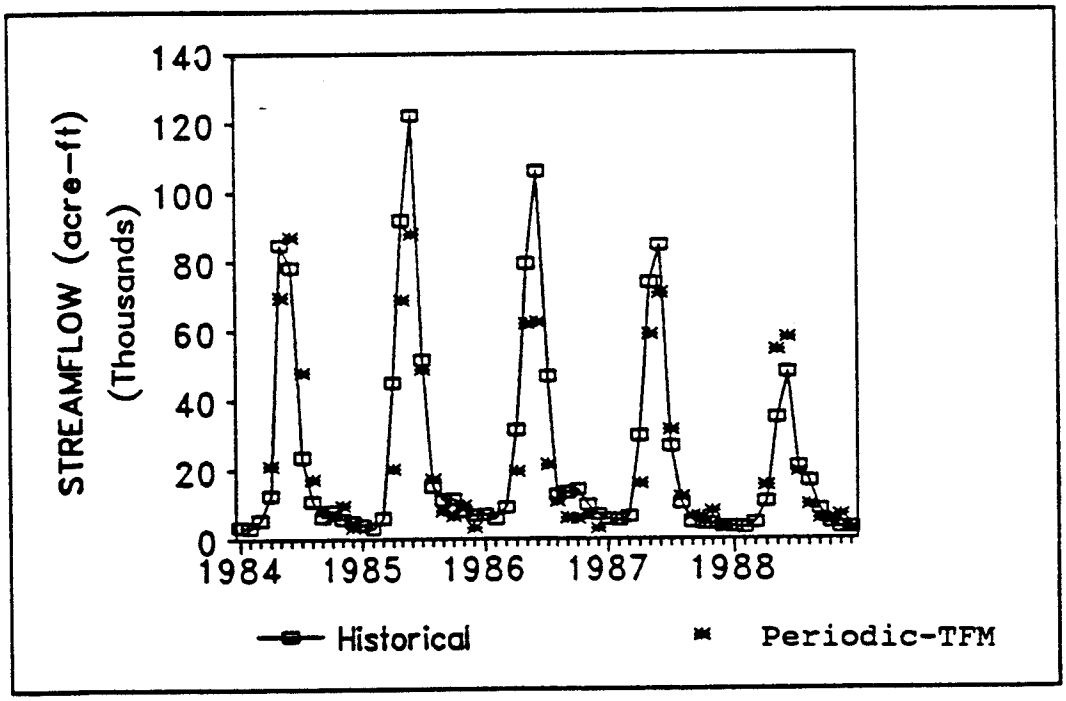


Fig. 7.113 April 1st forecasting of the monthly flow (Conejos River near Mogote)

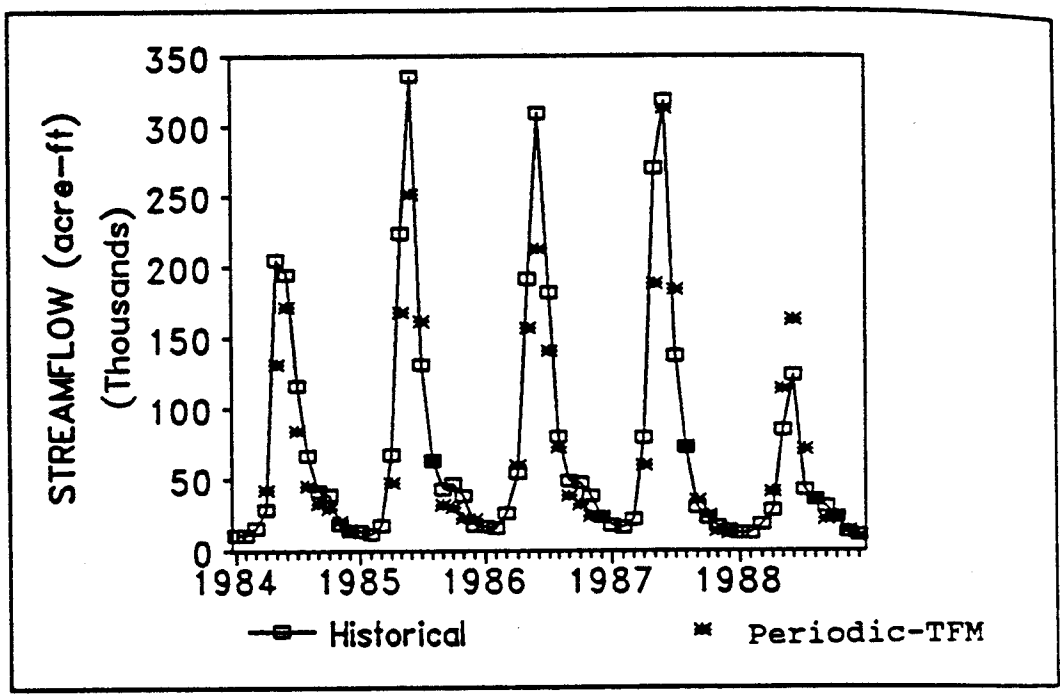


Fig. 7.114 One step ahead forecast of the monthly flow (Rio Grande River near Del Norte)

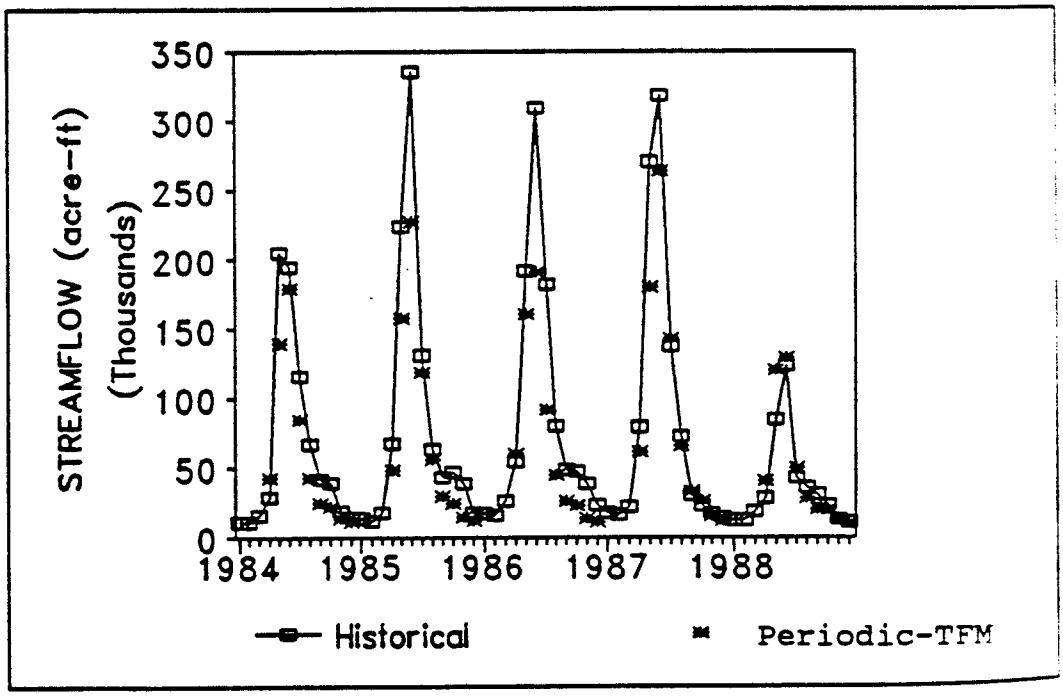


Fig. 7.115 April 1st forecasting of the monthly flow (Rio Grande River near Del Norte)

CHAPTER VIII

FORECASTING BASED ON THE MULTIPLE INPUT-SINGLE OUTPUT MODEL

8.1 Introduction

From the forecasting results shown in Chapter 7, we know that the single input-single output PAR(1) Residual-Transfer Function Model gives better forecasting result than other models. This was shown by comparing the plots of forecasts vs. historical values, root mean square deviation (RMSD) and the maximum deviation (MAD) between different models. In the last section of Chapter 7, we also found that the forecasts of streamflow in both 1985 and 1986 underestimated the high flow in the summer time. Since all three rivers in this research produced the same result for the summers of 1985 and 1986, the multiple input-multiple output model will be tried in an attempt to overcome this problem. We will assume that there are some factors influencing the streamflow in both 1985 and 1986 that we did not consider in our initial analysis. In this chapter, the climatic factors including temperature and precipitation will be considered.

The multiple input-single output model will now be discussed. The results show that the one step ahead forecast (with lead time $L=1$) of streamflow improved by using the multiple inputs - single output model, but the forecasts made with lead time $L=1,2,\ldots$ (stand at time t without moving) did not show significant improvement. The reason for this was that we need to forecast the multiple inputs before forecasting the output. When we forecast too far from the standing point, the inaccurate forecasting of the inputs is transferred to the output.

8.2 Multiple Input-Single Output Forecasting

8.2.1 Introduction

From the forecasting results shown in Chapter 7, we know that the no-log transformation of the original data gives better forecasting result then the log-transformed data. Therefore we will not do any data transformation in this section. The upper Rio Grande River Basin above Del Norte has more climatic gaging stations than the Los Pinos River Basin and the Conejos River Basin, hence the forecasting application of the multiple inputs, single output model will be applied in the upper Rio Grande River Basin.

The temperature and precipitation data of the gaging station Hermit 7 ESE and Del Norte discussed in section 6.4 will be used here. Both the temperature and precipitation data will be used jointly with the SWE data to forecast streamflow. Forecasts using the two input-single output model and the three input-single output model will be compared to the single input-single output model.

8.2.2 Using SWE and Temperature to Forecast Streamflow

The SWE data and the temperature data are first used as inputs to the system for forecasting streamflow. After using the Thiessen polygon method to compute the average SWE and temperature in the upper Rio Grande River above Del Norte, the average SWE $W_{v,\tau}^{(1)}$ and the temperature $W_{v,\tau}^{(2)}$ are deseasonalized using

$$x_{v,\tau}^{(1)} = \frac{W_{v,\tau}^{(1)} - \hat{\mu}_\tau(W_1)}{\hat{\sigma}_\tau(W_1)} \tag{8.2.1}$$

$$x_{v,\tau}^{(2)} = \frac{W_{v,\tau}^{(2)} - \hat{\mu}_\tau(W_2)}{\hat{\sigma}_\tau(W_2)} \tag{8.2.2}$$

Then the monthly flow $Q_{\nu,\tau}$ of the Rio Grande River at Del Norte is deseasonalized by using

$$z_{\nu,\tau} = \frac{Q_{\nu,\tau} - \hat{\mu}_\tau(Q)}{\hat{\sigma}_\tau(Q)} \tag{8.2.3}$$

The deseasonalized flow $z_{\nu,\tau}$ is then filtered by PAR(1) model using

$$z_{\nu,\tau} = \varphi_{1,\tau} z_{\nu,\tau-1} + y_{\nu,\tau} \tag{8.2.4}$$

with $\varphi_{1,\tau}$ shown below:

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.832	0.887	0.545	0.423	0.481	0.572
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.894	0.864	0.531	0.682	0.816	0.844

After $y_{\nu,\tau}$ is estimated, two input-single output transfer function models with inputs $x_{\nu,\tau}^{(1)}$ and $x_{\nu,\tau}^{(2)}$, and output $y_{\nu,\tau}$ can be written as

$$y_t = \sum_{j=0}^{\infty} a_j^{(1)} x_{t-j}^{(1)} + \sum_{j=0}^{\infty} a_j^{(2)} x_{t-j}^{(2)} + N_t \tag{8.2.5}$$

The impulse response functions are estimated and shown in Fig. 8.1-8.2. Then, Eq. (8.2.5) can be written as

$$y_t = -0.118x_t^{(1)} + 0.335x_{t-1}^{(1)} + 0.179x_t^{(2)} - 0.145x_{t-1}^{(2)} + N_t \tag{8.2.6}$$

The noise N_t was then estimated and shown in Fig. 8.3. The residual N_t was found to be white noise or

$$\hat{N}_t = \varepsilon_t \tag{8.2.7}$$

with $\varepsilon_t \sim WN(0,0.40)$. Then, Eq. (8.2.6) can be written as

$$y_t = -0.118x_{t,1} + 0.335x_{t-1,1} + 0.179x_{t,2} - 0.145x_{t-1,2} + \varepsilon_t \tag{8.2.8}$$

After substituting (8.2.8) into (8.2.4) and minimizing the sum of the square errors, a final model is obtained. This model can be written as

$$z_{\nu,\tau} = \hat{\varphi}_{1,\tau} z_{\nu,\tau-1} - 0.290x^{(1)}_{\nu,\tau} + 0.578x^{(1)}_{\nu,\tau-1} + 0.124x^{(2)}_{\nu,\tau} - 0.137x^{(2)}_{\nu,\tau-1} + \varepsilon^{*}_{\nu,\tau} \tag{8.2.9}$$

with $\varepsilon_{\nu,\tau}^{*} \sim \mathrm{WN}(0,0.38)$ and

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.832	0.887	0.545	0.488	0.507	0.319
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.893	0.835	0.568	0.630	0.776	0.873

The independence of the noise N_t and the inputs $x_t^{(1)}$, $x_t^{(2)}$ were checked and found to be uncorrelated. Hence, Eq. (8.2.9) is accepted.

Before forecasting the streamflow, we need to forecast the SWE and the temperature. The deseasonalized SWE $x_t^{(1)}$ is fitted by an ARMA model. It can be written as

$$x_t^{(1)} = 0.925\, x_{t-1}^{(1)} + \zeta_t^{(1)} \tag{8.2.10}$$

with $\zeta_t^{(1)} \sim \mathrm{WN}(0,0.105)$. The deseasonalized temperature $x_t^{(2)}$ is fitted by

$$x_t^{(2)} = 0.274\, x_{t-1}^{(2)} + \zeta_t^{(2)} \tag{8.2.11}$$

with $\zeta_t^{(2)} \sim \mathrm{WN}(0,0.898)$. By substituting (8.2.10) and (8.2.11) into (8.2.9), we have

$$z_{\nu,\tau} = \frac{-0.290+0.578B}{1-\hat{\varphi}_{1,\tau}B}\frac{1}{1-0.925B}\zeta^{(1)}_{\nu,\tau}$$

$$+\frac{0.124-0.137B}{1-\hat{\varphi}_{1,\tau}B}\frac{1}{1-0.274B}\zeta^{(2)}_{\nu,\tau} \qquad (8.2.12)$$

$$=\nu^{(1)}_{\tau}(B)\ \zeta^{(1)}_{\nu,\tau}+\nu^{(2)}_{\tau}(B)\ \zeta^{(2)}_{\nu,\tau}+\varphi_{\tau}(B)\varepsilon^{*}_{\nu,\tau}$$

Hence the forecast error can be written as

$$e_{\nu,\tau}(L) = \sum_{j=0}^{L-1}\nu^{(1)}_{j,\tau+L}\zeta^{(1)}_{\nu,\tau+L-j}+\sum_{j=0}^{L-1}\nu^{(2)}_{j,\tau+L}\zeta^{(2)}_{\nu,\tau+L-j}$$

$$+\sum_{j=0}^{L-1}\psi_{j,\tau+L}\varepsilon^{*}_{\nu,\tau+L-j} \qquad (8.2.13)$$

The cumulative frequency of the simulated forecast error with τ=3, and lead times L=1,..,7 are computed, and the quantiles of the 95% probability limit obtained. Then the forecast of the deseasonalized flow $z_{\nu,\tau}$ with lead time L=1,2,...,7 and the 95% confidence limits bound are shown in Fig 8.4.

The one step ahead forecast using Eqs. (8.2.9), (8.2.10) and (8.2.11) is shown in Fig. 8.5. In addition, the forecasting result using the single input-single output model of Eq. (7.7.5) is also shown in Fig. 8.5. The root mean square deviation (RMSD) and the maximum absolute deviation (MAD) of the forecast are shown in Table 8.1. From Fig. 8.5 and the Table 8.1, we can see that the two inputs-single output model gives better forecast result than the single input-single output model.

By standing at the end of March, the forecast with lead time L=1,...,7 using the two input-single output model and the single input-single output model are shown in Fig. 8.6. From Fig. 8.6, we can see that the forecast using the two input-single output model

with lead time far from the standing point does not improve the forecast result of the single input model. This is because we need to forecast the SWE and temperature before forecasting the streamflow. The forecast error of the SWE and temperature will transfer to the forecast values of the streamflow.

8.2.3 Using SWE and Precipitation to Forecast Streamflow

By using the same notations of Section 8.2.2, the SWE $W_{v,\tau}^{(1)}$, the precipitation $W_{v,\tau}^{(2)}$ and the streamflow $Q_{v,\tau}$ are deseasonalized using (8.2.1), (8.2.2) and (8.2.3). Then, the deseasonalized streamflow $z_{v,\tau}$ is fitted by the PAR(1) model of Eq. (8.2.4). A two input-single output transfer function model can then be formulated by using Eq. (8.2.5) and (8.2.6). From this point, the impulse response function is estimated and is shown in Figs. 8.7-8.8. Hence, Eq. (8.2.5) can be written as

$$y_t = -0.116x_t^{(1)} + 0.246x_{t-1}^{(1)} + 0.144x_{t-1}^{(1)} + 0.07x_t^{(2)} + 0.101x_{t-1}^{(2)} + N_t \qquad (8.2.14)$$

The noise N_t is estimated and shown in Fig. 8.9. Then, an ARMA model is fitted to the estimated noise. This can be written as

$$N_t = -0.13\ N_{t-1} + \varepsilon_t \qquad (8.2.15)$$

with $\varepsilon \sim \mathrm{WN}(0,0.413)$.

By substituting (8.2.15) into (8.2.14), we get

$$(1+0.13B)y_t = (1+0.13B)(-0.116x_t^{(1)} + 0.246x_{t-1}^{(1)} + 0.144x_{t-2}^{(1)} + 0.07x_t^{(2)} + 0.101x_{t-1}^{(2)}) + \varepsilon_t \qquad (8.2.16)$$

After substituting (8.2.4) into (8.2.16), Eq. (8.2.16) can be written as

$$
\begin{aligned}
z_{\nu,\tau} &= (\varphi_{1,\tau}-0.13)z_{\nu,\tau-1} + 0.13\varphi_{1,\tau-1}z_{\nu,\tau-2} \\
&-0.116x^{(1)}_{\nu,\tau}+0.231x^{(1)}_{\nu,\tau-1}+0.176x^{(1)}_{\nu,\tau-2}+0.019x^{(1)}_{\nu,\tau-3} \\
&+0.07x^{(2)}_{\nu,\tau}+0.11x^{(2)}_{\nu,\tau-1}+0.013x^{(2)}_{\nu,\tau-2}+\varepsilon_{\nu,\tau}
\end{aligned}
\tag{8.2.17}
$$

The non-linear least squares algorithm is then used. The algorithm suggests a final model can be written as

$$
\begin{aligned}
z_{\nu,\tau} &= (\varphi_{1,\tau}-0.186)z_{\nu,\tau-1} + 0.186\varphi_{1,\tau-1}z_{\nu,\tau-2} \\
&-0.259x^{(1)}_{\nu,\tau}+0.198x^{(1)}_{\nu,\tau-1}+0.184x^{(1)}_{\nu,\tau-2}+0.292x^{(1)}_{\nu,\tau-3} \\
&+0.106x^{(2)}_{\nu,\tau}+0.143x^{(2)}_{\nu,\tau-1}+0.09x^{(2)}_{\nu,\tau-2}+\varepsilon^{*}_{\nu,\tau}
\end{aligned}
\tag{8.2.18}
$$

with $\varepsilon_{\nu,\tau}^{*} \sim$ WN(0,0.408) and

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.832	0.887	1.000	0.511	0.392	0.302
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.622	0.709	0.552	0.660	0.815	0.897

The deseasonalized SWE $x_t^{(1)}$ is fitted to an ARMA model as shown in (8.2.10). We try fitting the deseasonalized precipitation $x_t^{(2)}$ with the ARMA model, but the analysis shows that it is white noise, i.e.

$$
x_t^{(2)} = \zeta_t^{(2)} \tag{8.2.19}
$$

with $\zeta_t^{(2)} \sim$ WN(0,0.971).

The one step ahead forecast of the streamflow using Eqs. (8.2.18), (8.2.10) and (8.2.19) is shown in Fig. 8.10 and Table 8.1. This graph shows that the two input-single

output model gives a good forecast result. In order to compare the forecasting ability of the model using temperature and SWE as the inputs and the model using precipitation and SWE as the inputs, the forecasting results of both models are shown in Fig. 8.11 . Results shown in Fig. 8.11 and Table 8.1 show that both model give similar forecasting results.

8.2.4 Using SWE, Temperature and Precipitation to Forecast Streamflow

In this section, the SWE, temperature and the precipitation data will be used as inputs to the upper Rio Grande River Basin system in order to forecast streamflow. Let the SWE $W_{\nu,\tau}^{(1)}$, temperature $W_{\nu,\tau}^{(2)}$ and the precipitation $W_{\nu,\tau}^{(3)}$ be deseasonalized using

$$x_{\nu,\tau}^{(1)} = \frac{W_{\nu,\tau}^{(1)} - \hat{\mu}_\tau(W_1)}{\hat{\sigma}_\tau(w_1)} \tag{8.2.19}$$

$$x_{\nu,\tau}^{(2)} = \frac{W_{\nu,\tau}^{(2)} - \hat{\mu}_\tau(W_2)}{\hat{\sigma}_\tau(w_2)} \tag{8.2.20}$$

$$x_{\nu,\tau}^{(3)} = \frac{W_{\nu,\tau}^{(3)} - \hat{\mu}_\tau(W_3)}{\hat{\sigma}_\tau(w_3)} \tag{8.2.21}$$

Suppose the monthly flow $Q_{\nu,\tau}$ of the Rio Grande River near Del Norte was deseasonalized by using

$$z_{\nu,\tau} = \frac{Q_{\nu,\tau} - \hat{\mu}_\tau(Q)}{\hat{\sigma}_\tau(Q)} \tag{8.2.22}$$

and the deseasonalized flow $z_{\nu,\tau}$ is filtered by a PAR(1) model of

$$z_{\nu,\tau} = \varphi_{1,\tau} z_{\nu,\tau-1} + y_{\nu,\tau} \tag{8.2.23}$$

Then the three input-single output transfer function model can be written as

$$y_t = \sum_{j=0}^{\infty} a_j^{(1)} x_{t-j}^{(1)} + \sum_{j=0}^{\infty} a_j^{(2)} x_{t-j}^{(2)} + \sum_{j=0}^{\infty} a_j^{(3)} x_{t-j}^{(3)} + N_t \tag{8.2.24}$$

The impulse response function of (8.2.24) is estimated and shown in Fig. 8.12-8.14. Then the Eq. (8.2.5) can be written as

$$\begin{aligned} y_t = & -0.169 x_t^{(1)} + 0.328 x_{t-1}^{(1)} + 0.226 x_t^{(2)} - 0.139 x_{t-1}^{(2)} \\ & + 0.118 x_t^{(3)} + 0.113 x_{t-1}^{(3)} + N_t \end{aligned} \tag{8.2.25}$$

The noise N_t in Eq. (8.2.25) is estimated and was fitted by an ARMA model. This can be written as

$$N_t = -0.118\, N_{t-1} + \varepsilon_t \tag{8.2.26}$$

with $\varepsilon_t \sim WN(0,0.362)$.

By substituting (8.2.23) and (8.2.26) into (8.2.25), Eq. (8.2.25) can be written as

$$\begin{aligned} z_{v,\tau} = & (\varphi_{1,\tau} - 0.118) z_{v,\tau-1} + 0.118 \varphi_{1,\tau-1} z_{v,\tau-2} + (1 + 0.118B) \\ & (-0.169 x_{v,\tau}^{(1)} + 0.328 x_{v,\tau-1}^{(1)} + 0.226 x_{v,\tau}^{(2)} - 0.139 x_{v,\tau-1}^{(2)} \\ & + 0.118 x_{v,\tau}^{(3)} + 0.113 x_{v,\tau}^{(3)}) + \varepsilon_{v,\tau} \end{aligned} \tag{8.2.27}$$

After minimizing the sum of square errors, a final model is obtained and is written as

$$\begin{aligned} z_{v,\tau} = & (\varphi_{1,\tau} - 0.212) z_{v,\tau-1} + 0.212 \varphi_{1,\tau-1} z_{v,\tau-2} - 0.283 x_{v,\tau}^{(1)} \\ & + 0.282 x_{v,\tau-1}^{(1)} + 0.366 x_{v,\tau-2}^{(1)} + 0.108 x_{v,\tau}^{(2)} + 0.144 x_{v,\tau-1}^{(2)} \\ & + 0.08 x_{v,\tau-2}^{(2)} + 0.118 x_{v,\tau}^{(3)} + 0.127 x_{v,\tau-1}^{(3)} + 0.013 x_{v,\tau-2}^{(3)} + \varepsilon_{v,\tau}^{*} \end{aligned} \tag{8.2.28}$$

with $\varepsilon_t^* \sim WN(0,0.443)$ and

τ	1	2	3	4	5	6
$\hat{\varphi}_{1,\tau}$	0.832	0.887	0.862	0.553	0.503	0.399
τ	7	8	9	10	11	12
$\hat{\varphi}_{1,\tau}$	0.749	0.922	0.524	0.680	0.828	0.927

The independence between the noise N_t and the inputs $x_t^{(1)}$, $x_t^{(2)}$ were checked and found to be uncorrelated. Hence, Eq. (8.2.28) is accepted.

The one step ahead forecast of the streamflow of the Rio Grande River near Del Norte is shown in Fig. 8.15 and Table 8.1. The forecast with lead time L=1,...,7 is shown in Fig. 8.16. From Figs. 8.15 and Fig. 8.16, we can see that the one step ahead forecast using three input-single output model is improved over the single input model. In order to compare the forecast ability of the three input-single output model to the two inputs-single output model, the results of the one step ahead forecast of both models are shown in Fig. 8.17. From Fig. 8.17, we can see that both models give similar forecasting results.

Table 8.1 Comparison of one step ahead forecast (multiple input-single output model)

Model		RMSD	MAD
1 in-1 out	[see Fig. 8.5]	31795	96447
2 in (SWE, Temp.)-1 out	[see Fig. 8.5]	27424	74693
2 in (SWE, Prec.)-1 out	[see Fig. 8.10]	29032	76530
3 in-1 out	[see Fig. 8.15]	30705	85088

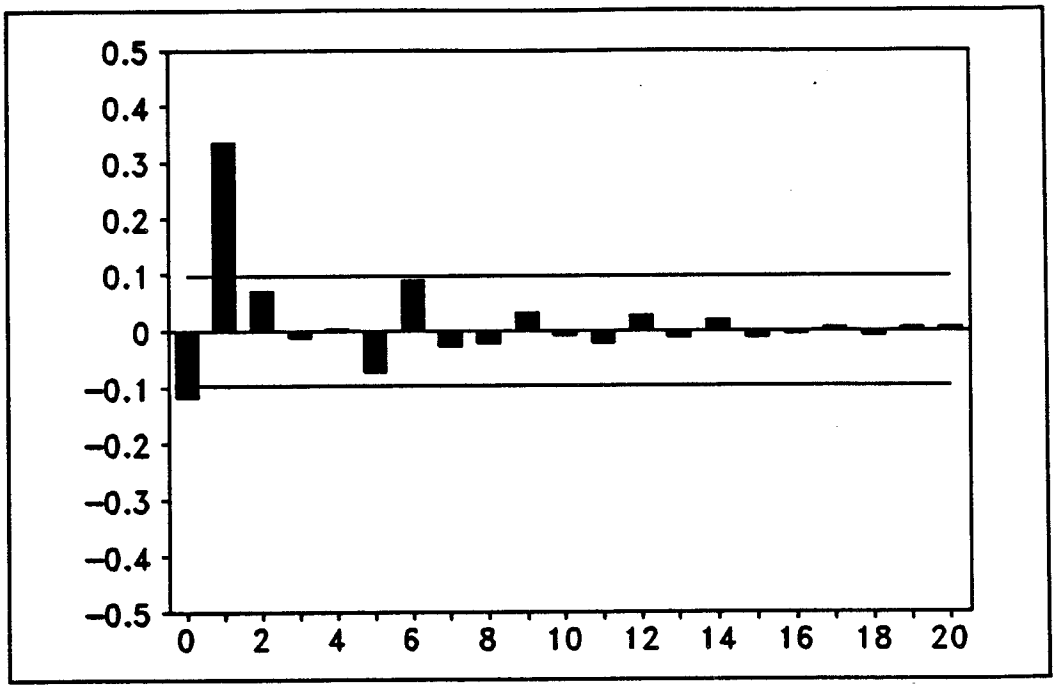


Fig. 8.1 Impulse response function $a_{j,1}$ (Rio Grande River; inputs: SWE & temperature)

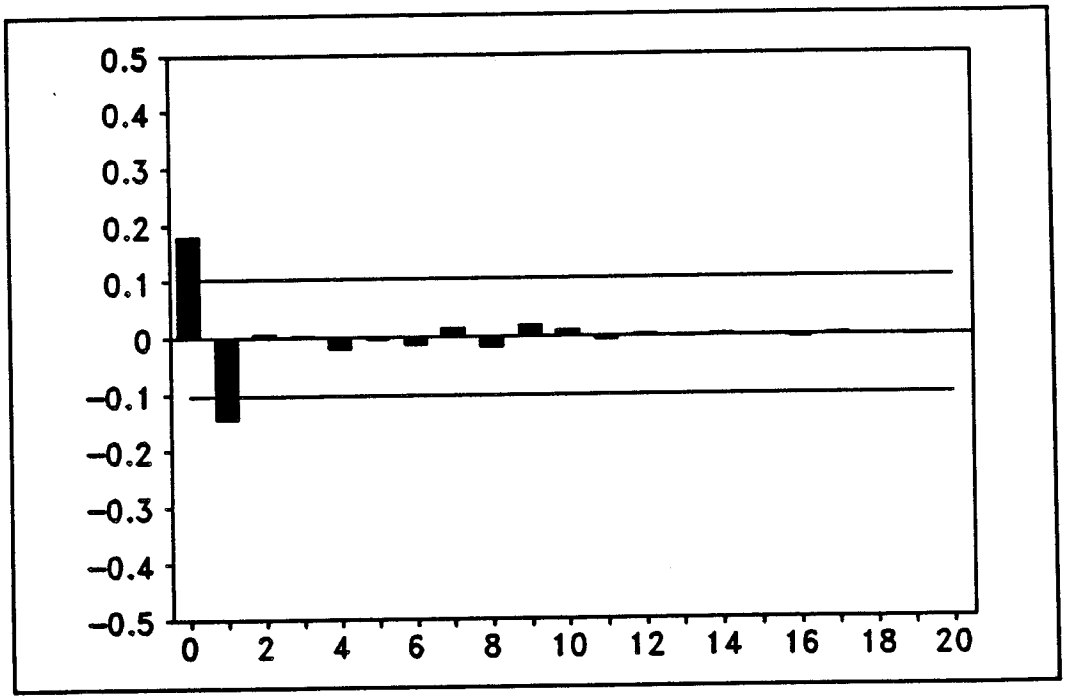


Fig. 8.2 Impulse response function $a_{j,2}$ (Rio Grande River; inputs: SWE & temperature)

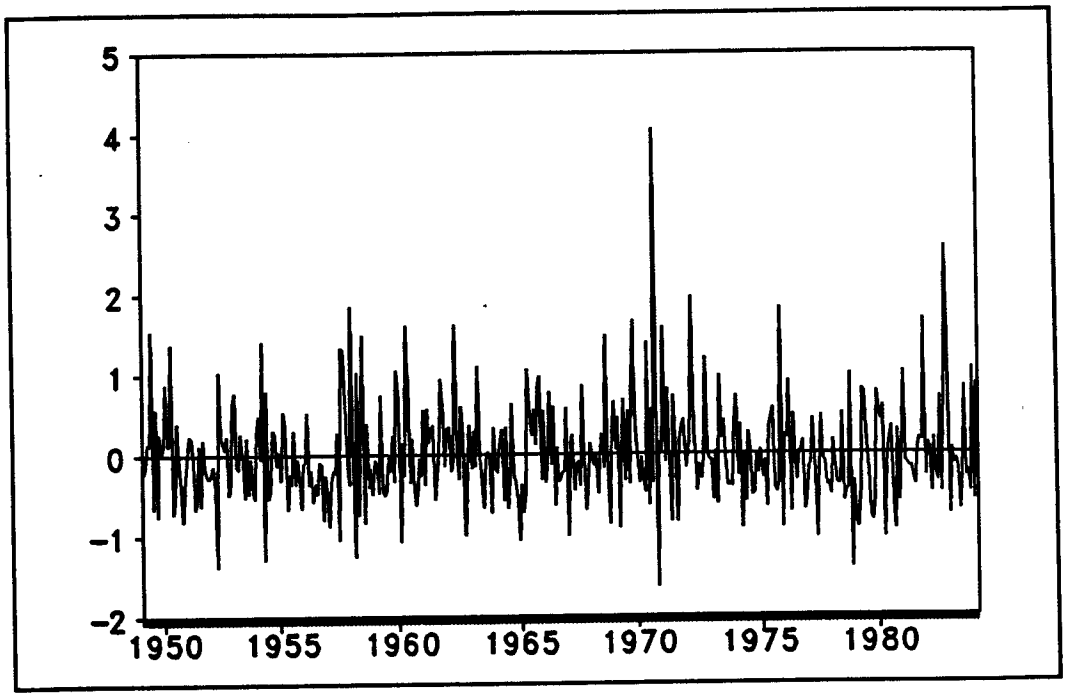


Fig. 8.3 The noise N_t of the 2 inputs-1 output model (Rio Grande River; inputs: SWE & temperature)

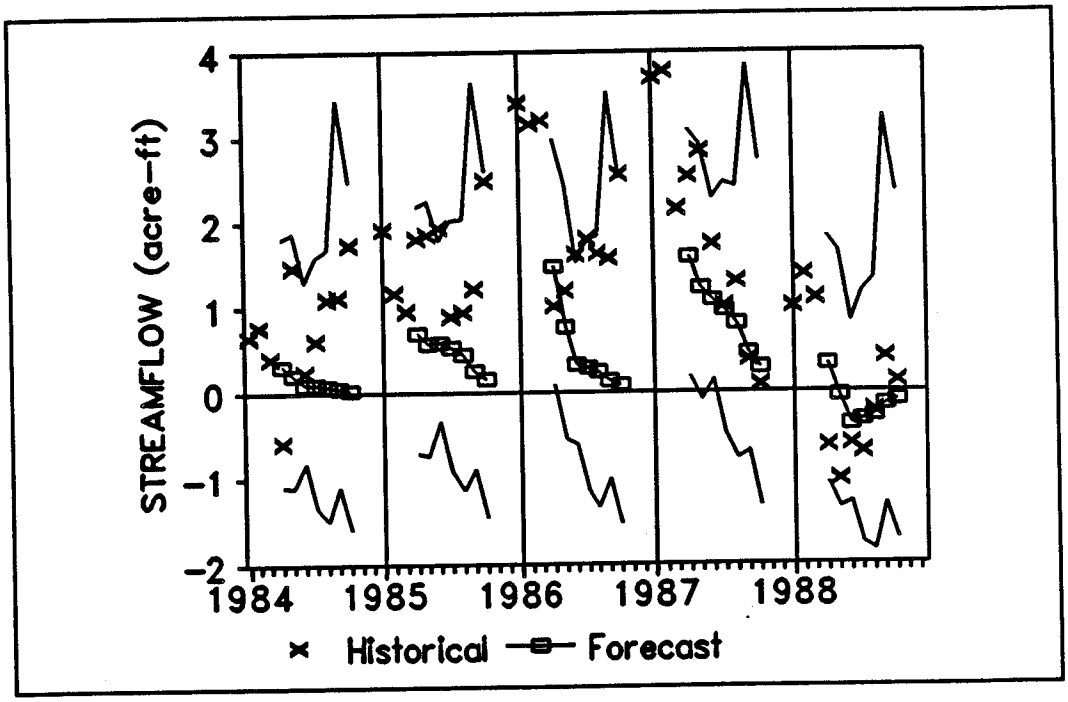


Fig. 8.4 Forecasting of the deseasonalized flow with L=1,...,7 (Rio Grande River; inputs: SWE & temperature)

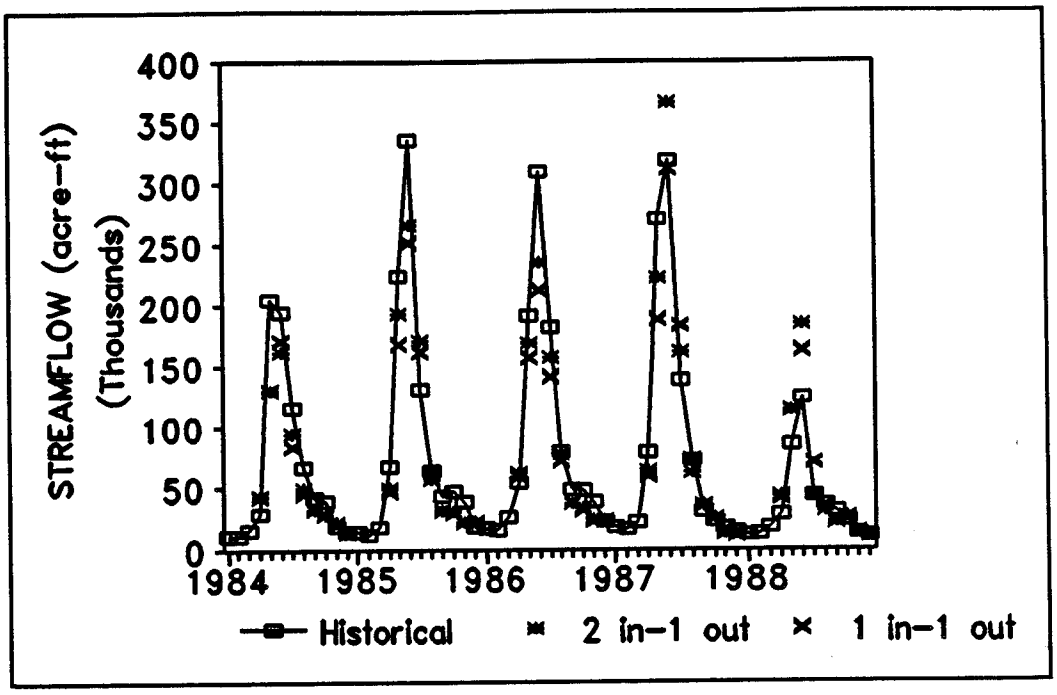


Fig. 8.5 One step ahead forecast of the monthly flow (Rio Grande River; inputs: SWE & temperature)

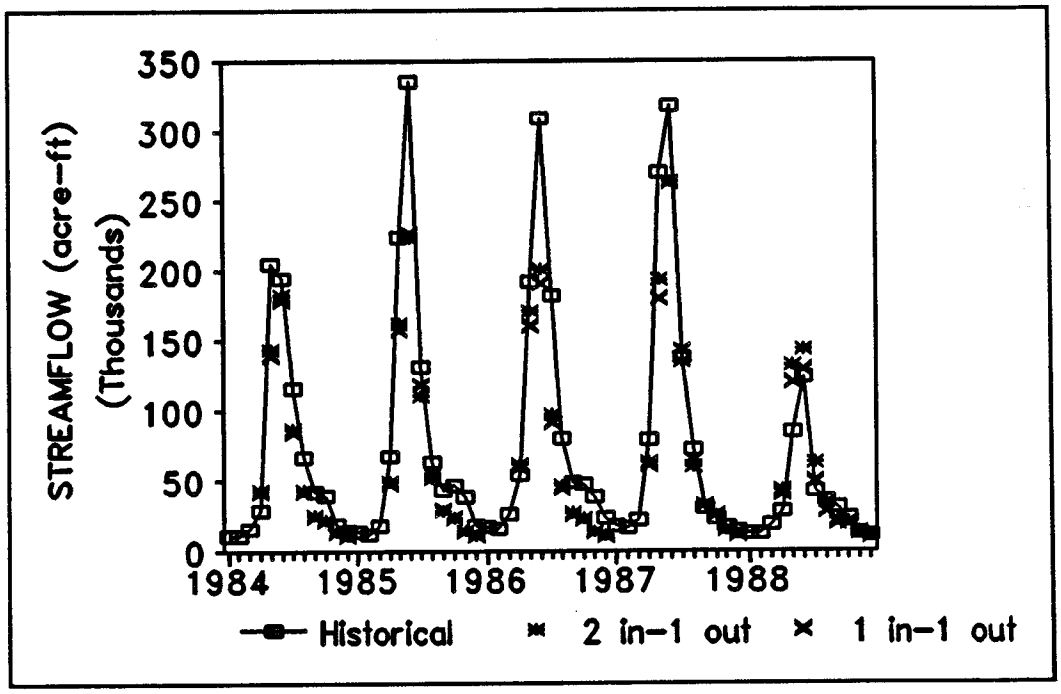


Fig. 8.6 April 1st forecasting with L=1,...,7 (Rio Grande River; inputs: SWE & temperature)

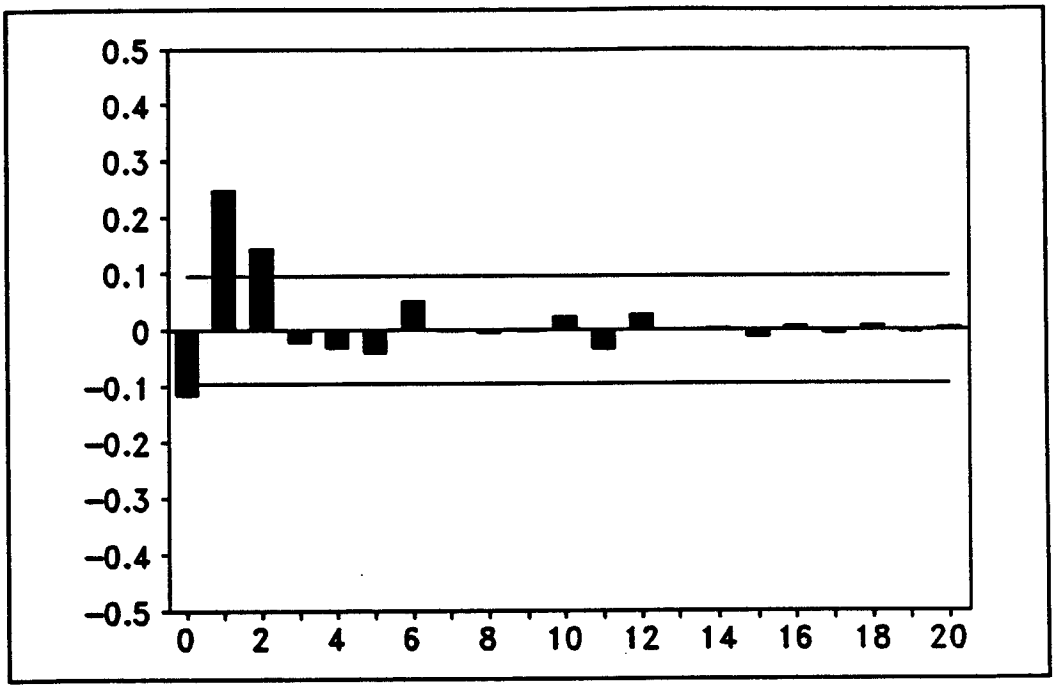


Fig. 8.7 The impulse response function $a_{j,1}$ (Rio Grande River; inputs: SWE & precipitation)

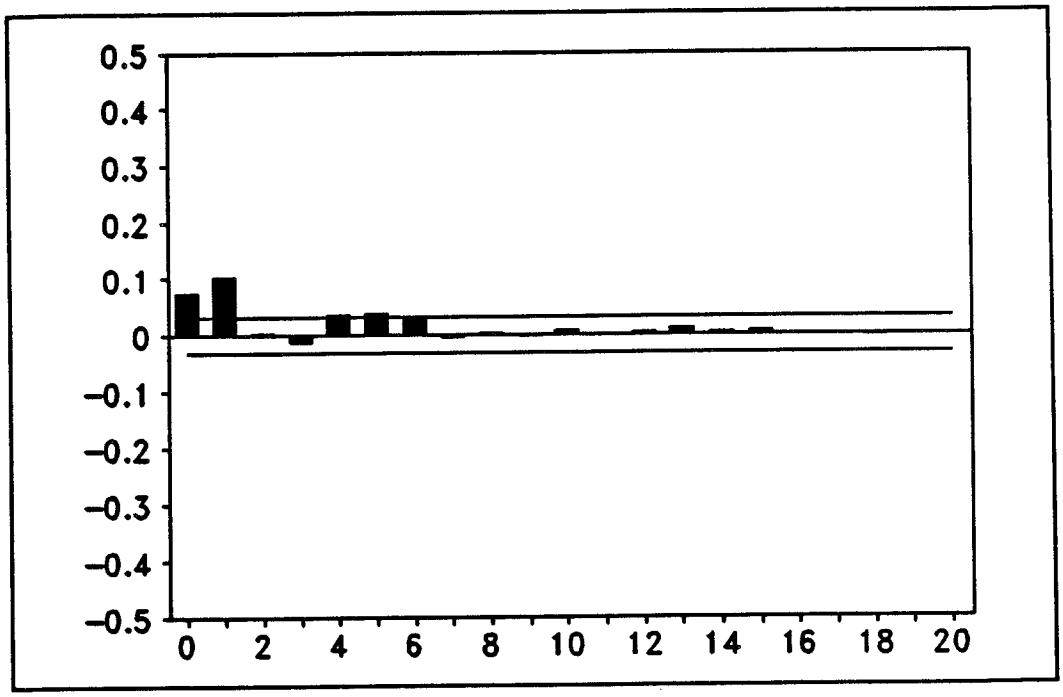


Fig. 8.8 The impulse response function $a_{j,2}$ (Rio Grande River; inputs: SWE & precipitation)

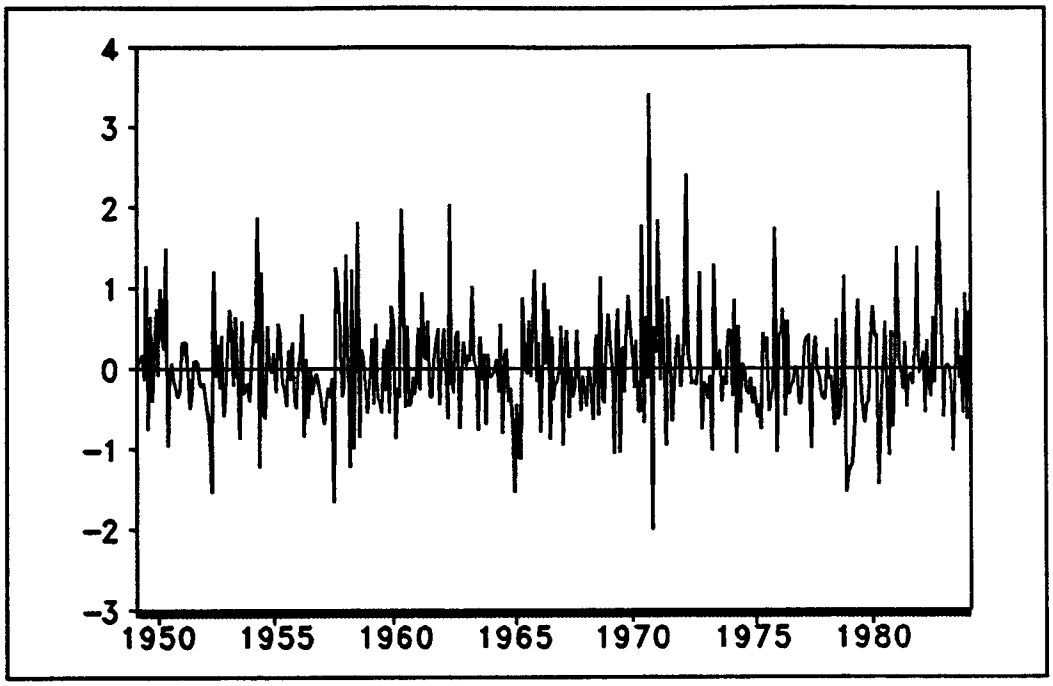


Fig. 8.9 The noise of the 2 inputs-1 output model (Rio Grande River; inputs: SWE & precipitation)

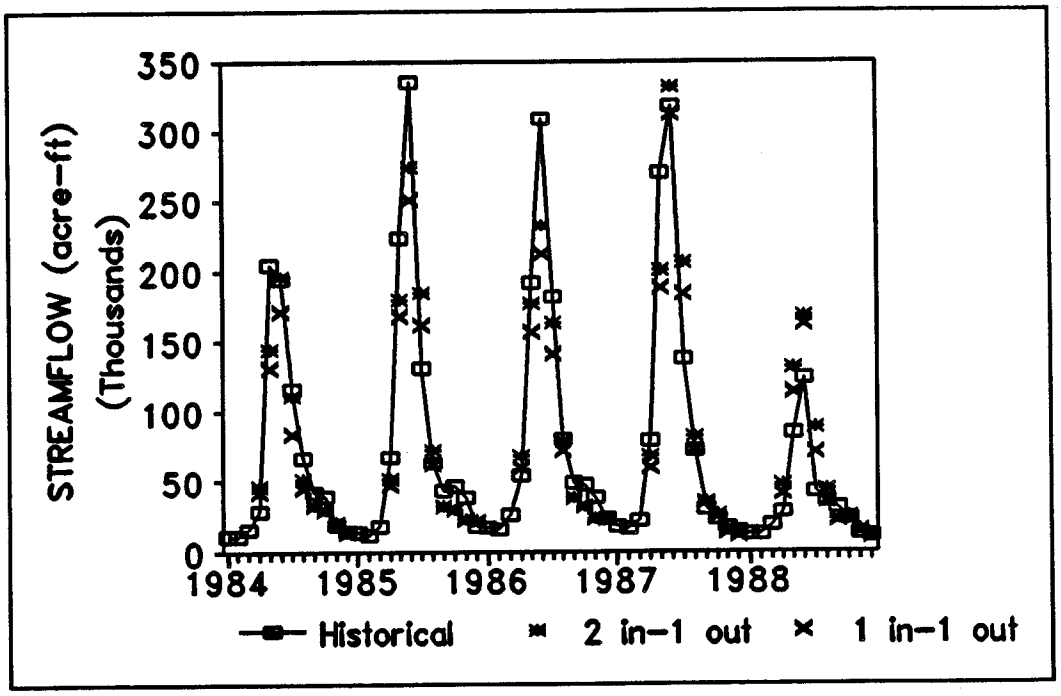


Fig. 8.10 One step ahead forecast of the streamflow (Rio Grande River; inputs: SWE & precipitation)

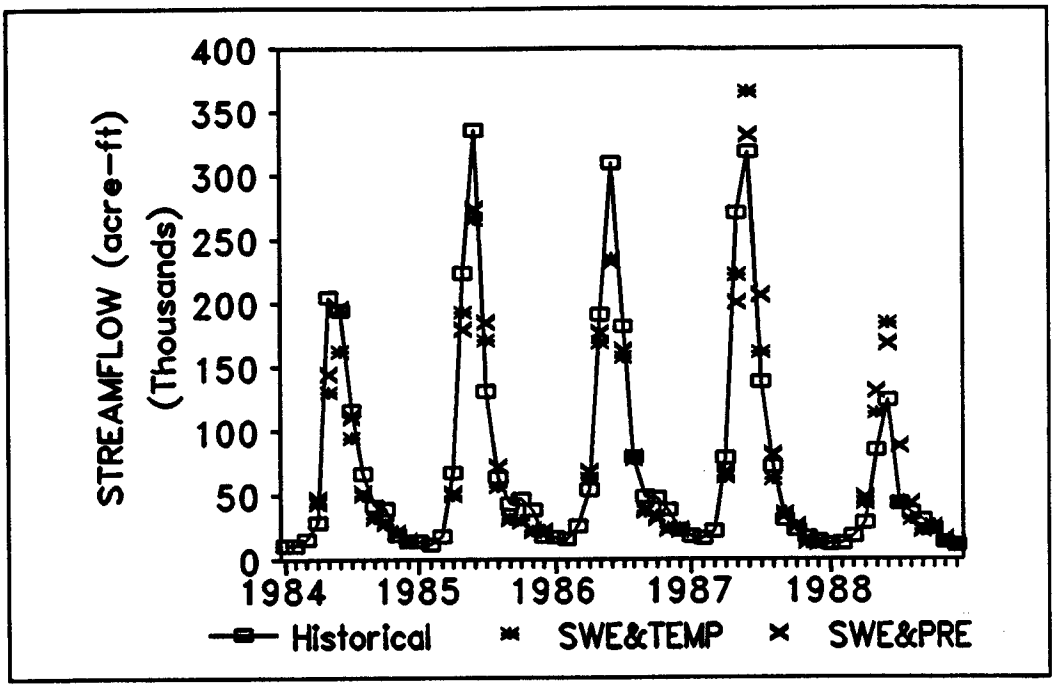


Fig. 8.11 Comparison of the forecasting with different inputs (Rio Grande River)

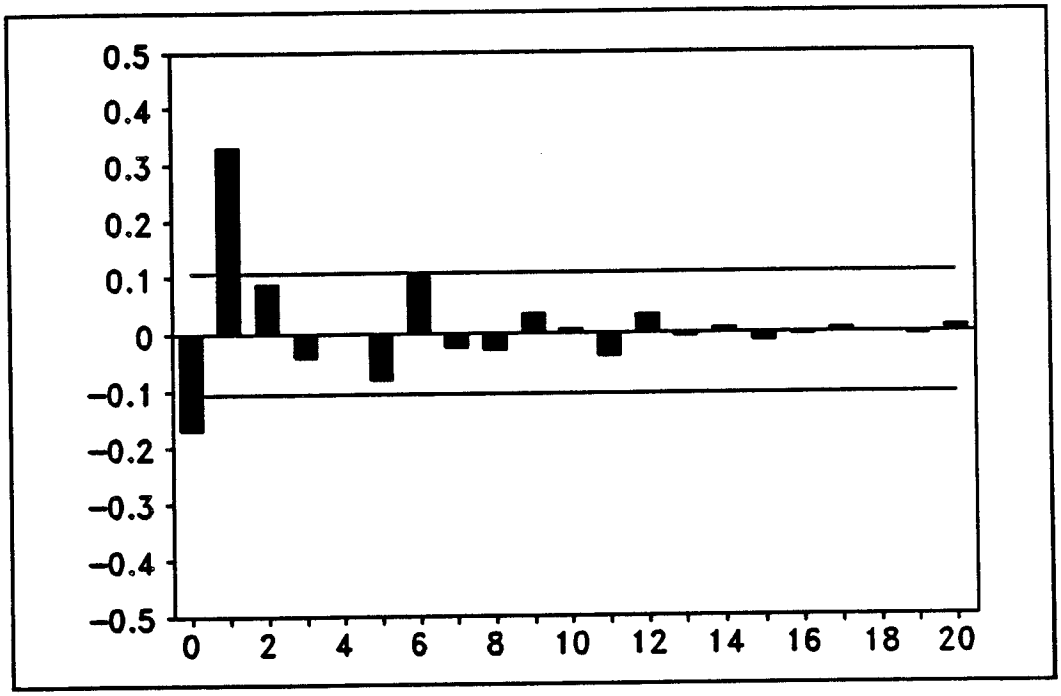


Fig. 8.12 The impulse response function $a_{j,1}$ (Rio Grande River; 3 inputs-1 output model)

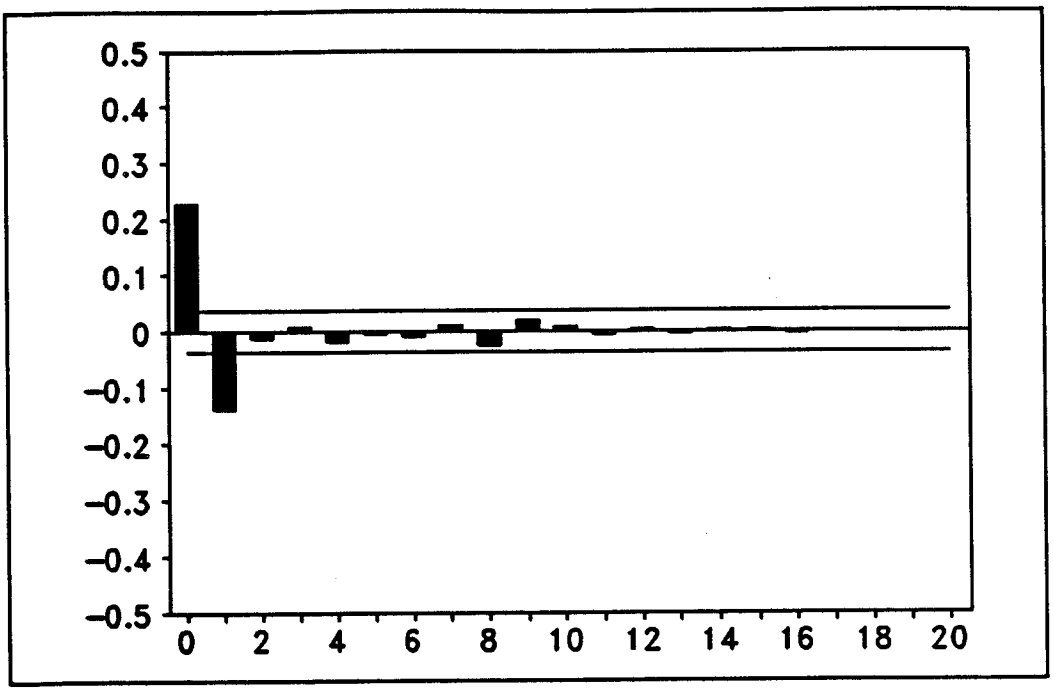


Fig. 8.13 The impulse response function $a_{j,2}$ (Rio Grande River; 3 inputs-1 output model)

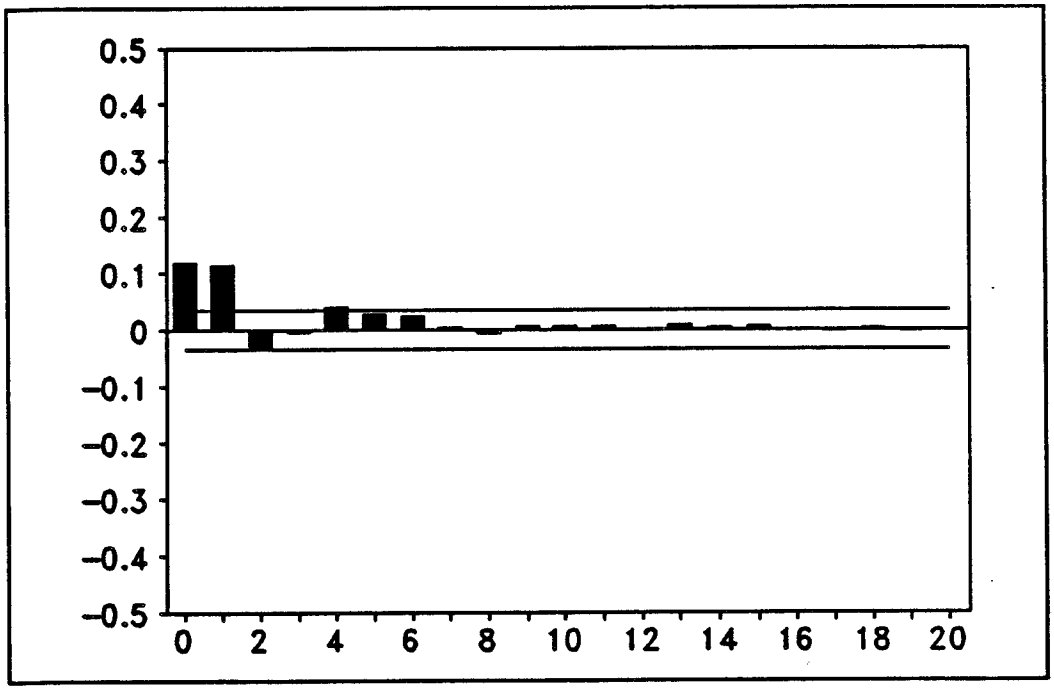


Fig. 8.14 The impulse response function $a_{j,3}$ (Rio Grande River; 3 inputs-1 output model)

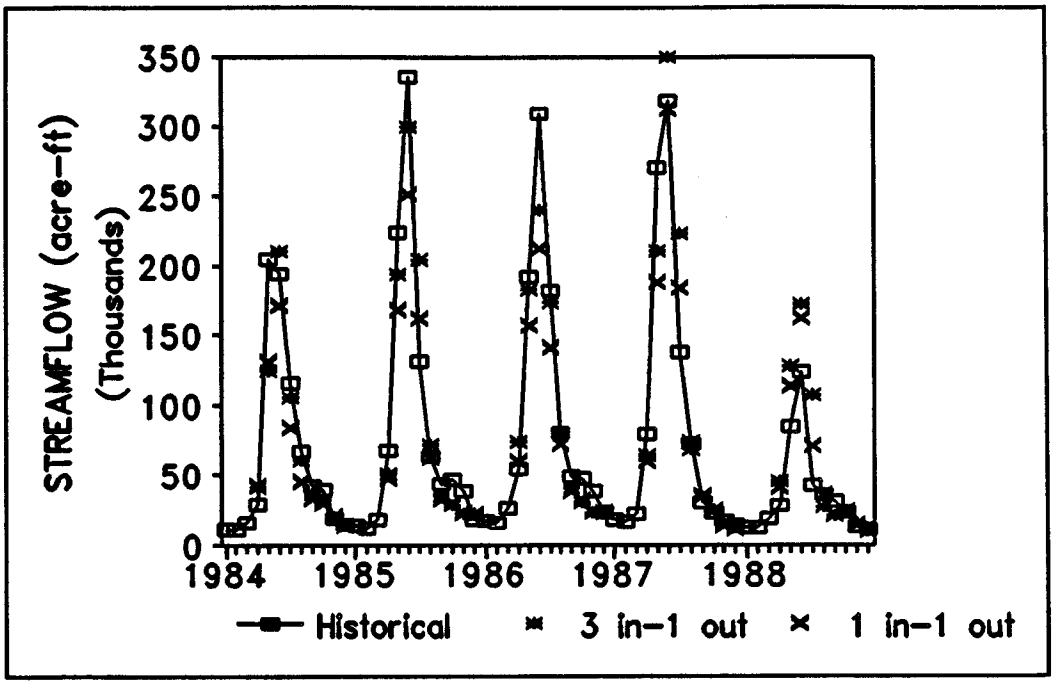


Fig. 8.15 The one step ahead forecast of the streamflow (Rio Grande River; 3 inputs-1 output model)

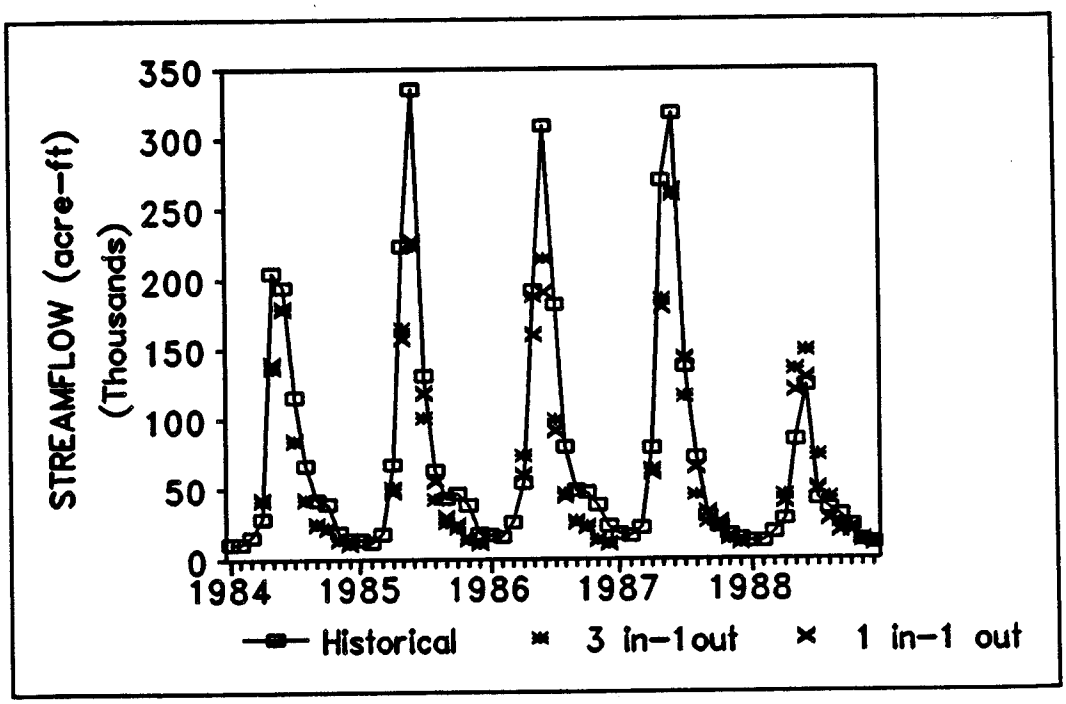


Fig. 8.16 April 1st forecasting with L=1,...,7 (Rio Grande River; 3 inputs-1 output model)

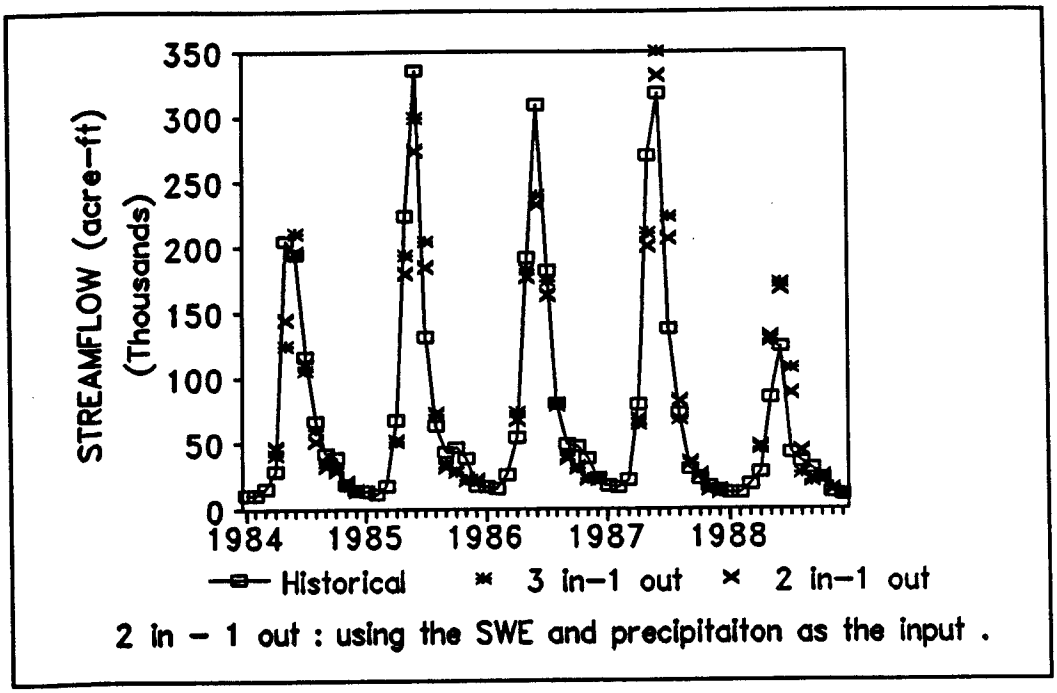


Fig. 8.17 One step ahead forecasting comparison (Rio Grande River)

CHAPTER IX

CONCLUSIONS

The purpose of this chapter is to summarize the information presented. This study has focused on the forecasting of monthly streamflow by using periodic transfer function models. The emphasis and application has been to examine the streamflow resulting from snowmelt, precipitation and temperature. The identification of the transfer function model with intermittent input data has been done by using spectral analysis without prewhitening the input data. The forecasting ability of the deseasonalized ARMA model was first examined. The monthly streamflow were transformed using log-transformation. The problems associated with forecasting a log-transformed series were discussed. The results show that the deseasonalized ARMA model always underestimated the high flows in the summer time.

The PARMA model was also used for modeling and forecasting the monthly streamflow. The residual of the PAR(1) or PARMA(1,1) was verified to be non-Gaussian distributed. This is true even if one takes the log-transformation of the original data. In order to find the confidence limits of the forecasts, the forecast error is simulated by randomly sampling the historical residual with replacement. The forecasting application showed that the PAR(1) model gives a better forecasting result than the deseasonalized ARMA model.

The single input-single output transfer function model with intermittent SWE as the input and with the streamflow as the output was then used. Models based on the original non-transformed data and based on log-transformation were compared. The

results indicate that the transfer function model gives a better forecast compared to the PAR(1) model. It also indicates that the model using the original non-transformed series gives a better forecasting result when compared to the model using log-transformed series.

The joint use of the PAR(1) model and transfer function model improved the forecast. By fitting a PAR(1) model to the deseasonalized streamflow series and obtaining the residual of the PAR(1) model, a single input-single output transfer function model, with deseasonalized SWE as the input and the residual of PAR(1) model as the output, can be formulated. Identification of the model is made by using spectral analysis. The non-linear least squares method is used for estimating the parameters. The forecast error of this modified model is derived. Since the forecast error has an unknown distribution, the cumulative relative frequency of the forecast error was simulated in order to find the 95% probability limit bound. The selection of better forecasting models is accomplished by comparing the plots of forecasts and historical values, root mean square deviation and the maximum absolute deviation for all models.

The forecasting method currently used for forecasting monthly flows in the Rio Grande River Basin is based on the multiple regression model. Data from 1948-1977 was used for comparing both the multiple regression model and the modified single input-single output model. The validation was done based on the years 1978-1982. After comparison, we found that the modified transfer function model gives a better forecast result. The modified single input-single output transfer function model with SWE as the input was then used to forecast the streamflow in the recent years (1983-1988).

The multiple input-single output model was also formulated by using the spectral

analysis technique. The upper Rio Grande River Basin above Del Norte has several climatic gaging stations and was used for formulating the multiple input-single output forecast. The forecasting study showed that using the SWE and the temperature data to forecast streamflow (1983-1988) improved the one step ahead forecasting as compared to the single input-single output model, but that it did not give a better forecasting result when forecasting with lead time far from the standing point. The reason for this result was the need to forecast the inputs before forecasting the streamflow. The forecasting error of the forecasted input series transfers to the forecasted output series.

Using SWE and precipitation to forecast the streamflow was also tried. This study showed that the one step ahead forecasting result improved. Forecasting the streamflow using three inputs (SWE, temperature and precipitation) was also tried. The results showed that the one step ahead forecasting also improved. In order to compare the forecasting ability of the multiple input-single output model with different inputs, the forecasting results were compared. The comparison showed that all the multiple input-single output models have a similar forecasting ability, and are better than the single input-single output models for the one step ahead forecasting.

From the discussion shown above, we can conclude:

(1) The deseasonalized ARMA model always underestimated the high monthly flows in the summer time.

(2) The PAR(1) model gives a better forecasting result when compared to the deseasonalized ARMA model, but it also underestimated the high monthly flows in the summer time.

(3) The rational transfer function model gives a better forecasting result when compared to both the deseasonalized ARMA model and the PARMA model.

(4) The periodic transfer function model improved the forecasting results further.

(5) The multiple input-single output model gives a forecast results that are better than those of the single input-single output model, especially for the one step ahead forecast.

REFERENCES

Akaike, H., "A New Look at the Statistical Model Identification." IEEE Trans. on Automatic Control, AS-19, 6, 716-723, 1974.

Anderson, R. L., "Distribution of the Serieal Correlation Coefficients." Annals of Math. Statistics, 8(1), 1-13, March 1941.

Baracos, P.C., W. Hipel and A. Ian Mcleod, "Modeling Hydrologic Time Series from the Arctic." Water Resources Bulletin, 17(3), June, 1981.

Bartlett, M. S., "STOCHASTIC PROCESSES", Cambridge Univ. Press, 1966.

Bloomfield, P. , "FOURIER ANALYSIS OF TIME SERIES : AN INTRODUCTION." New York : John Wiley , 1976.

Box, G. E. P. and G. M. Jenkins, "TIME SERIES ANALYSIS : FORECASTING AND CONTROL." San Francisco : Holden-Day, 1970.

Bowerman, B.L. and R.T. O'Connell, "TIME SERIES FORECASTING : UNIFIED CONCEPTS AND COMPUTER IMPLEMENTATION." Boston: Duxbury Press, 1987.

Brillinger, D. R., "TIME SERIES : DATA ANALYSIS AND THEORY." New York : John Wiley, 1981.

Brockwell, P. J. and R. A. Davis, "TIME SERIES : THEORY AND METHODS." New York : Springer-Verlag , 1987 .

Burn, D. H. and E. A. McBean, "River Flow Forecasting Model for Sturgeon River." ASCE, J. of Hydraulic Engineering, 111(2), Feb., 1985.

Canfield, R. V., "Hydrologic Series Generation from the Spectral Density Function." Hydrologic and Hydrology Series, UWRL/H-82/04, Utah Water Research Laboratory, Utah State University, Logan, UT., 1982.

Canfield, R. V. and D.S. Bowles, " Multivariate Series Generation from the Spectral Density Function." In : Shen et al., eds., MULTIVARIATE ANALYSIS OF HYDROLOGIC PROCESSES. Fort Collins, Colorado, 1985.

Chatfield, C. , "THE ANALYSIS OF TIME SERIES." 2nd ed. New York: Chapman and Hall, 1980.

Chiu, C. L., "APPLICATIONS OF KALMAN FILTERS TO HYDROLOGY." University of Pittsburgh, Pittsburgh, Pennsylvania, 1978.

Clarke, R. T., "MATHEMATICAL MODELS IN HYDROLOGY." Irrigation and Drainage Paper. Food and Agriculture Organization of the United Nations, Rome, 1973.

Cooley, J.W. and Tuckey, J.W., "An Algorithm for the Machine Calculation of Complex Fourier Series." Math. Comp. 19, 297-301, 1965.

Cooley, J. W., P. A. W. Lewis and P.D. Welch, "The Fast Fourier Transform and its Application to Time Series Analysis." In : K. Ensdein et al., eds., STATISTICAL METHODS FOR DIGITAL COMPUTERS. New York : John Wiley, 1977.

Cooper, D. M. and E. F. Wood, "Model Identification and Parameter Estimation for Hydrologic Input-Output Models." In: PROCEEDINGS OF THE JOINT AUTOMATIC CONTROL CONFERENCE, San Francisco, 1980.

Cooper, D. M. and E. F. Wood, "Identification of Multivariate Time Series and Multivariate Input-Output Models." Water Resour. Res., 18(4), 937-949, Aug., 1982a.

Cooper, D. M. and E. F. Wood, "Parameter Estimation of Multiple Input-Output Time Series Models: Application to Rainfall-Runoff Processes." Water Resour. Res., 18(5), 1352-1364, Oct., 1982b.

Delleur, J. W., P. C. Tao and M. L. Kavans, "An Evaluation of the Practicability and Complexity of Some Rainfall and Runoff Time Series Models." Water Resour. Res., 12(5), 957-970, 1976.

Granger, C. W. J. and P. Newbold, "Forecasting Transformed Series." J. of the Royal Stat. Soc. B., (38), 189-203, 1976 .

Haltiner J. P., "STOCHASTIC MODELING OF SEASONAL AND DAILY STREAMFLOW." Ph. D. Dissertation, Colorado State University, Fort Collins, Colorado, 1985.

Haltiner J. P. and J. D. Salas, " Short-Term Forecasting of Snowmelt Runoff Using ARMAX Model." Water Resources Bulletin, 24(5), 1083-1089, Oct., 1988.

Hipel, K. W., A. I. Mcleod and D. J. Noakes, "Fitting Dynamic Model to Hydrological Time Series." In : A. H. El-Shaarawi and S. R. Esterly, TIME SERIES METHODS IN HYDROSCIENCES. The Netherlands: Elsevier Scientific, 110-129, 1982.

Hoshi, K. and S. J. Burges, "Seasonal Runoff Volumes Conditional on Forecasted Total Runoff Volume." Water Resour. Res., 16(6), 1079-1084, Dec., 1980.

Jenkins, G. M. and D. J. Watts, "SPECTRAL ANALYSIS AND ITS APPLICATIONS." San Francisco : Holden-Day, 1968.

Koopmans, L. H., "THE SPECTRAL ANALYSIS OF TIME SERIES." New York : Academic Press, 1974.

Krzysztofowicz, R. and L. M. Watada, "Stochastic Model of Seasonal Runoff Forecasts." Water Resour. Res., 22(3), 296-302, 1986.

Kuester, J. L. and J. H. Mize, "Optimization Techniques with Fortran." McGraw-Hill, 1973.

Lang, H. , "Forecasting Meltwater Runoff from Snow-Covered Areas and from Glacier Basins. In : D. A. Kraijenhoff and J. R. Moll ed., RIVER FLOW MODELING AND FORECASTING , Boston: D. Reidel, 1986.

Liang, G. C., "Identification of a Multiple Input, Single Output, Linear, Time Invariant Model for Hydrological Forecasting." J. of Hydrology, 101, pp 251-262, 1988.

Marquardt, D. M., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters." J. Soc. Indust. Appl. Math., 11, 431-441, 1963.

Martinec, J. and A. Rango, "Discharge forecasts in mountain basins based on Satellite Snow Cover Mapping." Proceedings of a Workshop on Operational Applications of Satellite Snow-Cover Observations. 16-17, April 1979, Sparks, Nevada.

McLeod, A. I., D. J. Noakes, K. W. Hipel and R. M. Thompstone, "Combing Hydrologic Forecasts." ASCE, J. of Water Resources Planning and Management, 113(1), 29-41, Jan., 1987.

Mejia, J. M. and I. Rodriguez-Iturbe, "On the Synthesis of Random Field Sampling from the Spectrum : an Application to Generation of Hydrologic Spatial Process." Water Resour. Res., 10(4), 705-711, 1974.

Mizumura, K. and Chao-Lin Chiu, "Prediction of Combined Snowmelt and Rainfall Runoff." ASCE, J. of Hydraulic Engineering, 111(2), 179-193, 1985.

Mood, A. M., F. A. Graybill and D. C. Boes, "INTRODUCTION TO THE THEORY OF STATISTICS." New York : McGraw-Hill , 1974 .

Nicklin, M. E., "Bivariate Transfer Function Modeling of Irrigation Return Flows." In: Shen et al. eds., MULTIVARIATE ANALYSIS OF HYDROLOGIC PROCESSES. Fort Collins, Colorado, 1985.

Noakes, D. J., A. I. McLeod, and K. W. Hipel, "Forecasting Monthly Riverflow Time Series." J. of Forecasting, 179-190, North-Holland : Elsevier Science, 1985.

Otnes, R. K. and L. Enochson, "APPLIED TIME SERIES ANALYSIS." New York : John Wiley, 1978.

Padmanabhan, G and A. R. Rao, "Maximum Entropy Spectral Analysis of Hydrologic Data. Water Resour. Res., 24(9), 1519-1533, Sep., 1988.

Parzen, E., "Mathematical Considerations in the Estimation of Spectra." Technometrics 3, 167-190, 1961.

Pei, D., S. J. Burges and J. R. Stedinger, " Runoff Volume Forecasts Conditioned on a Total Seasonal Runoff Forecast." Water Resour. Res., 23(1), 9-14, Jan., 1987.

Prasad, R., "A Nonlinear Hydrologic System Response Model." ASCE, J. of Hydraulics, 93(4), 201-221, July, 1967.

Press, W. H., B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, "NUMERICAL RECIPES." New York : Cambridge University Press, 1986.

Priestley, M. B., "SPECTRAL ANALYSIS AND TIME SERIES." New York: Academic Press, 1981.

Rango, A. and J. Martinec, "Application of Snowmelt-Runoff Model Using Landsat Data. Nordic Hydrology, Vol.10, 225-238, 1979.

Rango, A., "Application of a Simple Snowmelt-Runoff Model to Large River Basins. Western Snow Conference Proceedings, 1983.

Robinson, E. A., "MULTICHANNEL TIME SERIES ANALYSIS WITH DIGITAL COMPUTER PROGRAMS. Texas : Goose Pond Press, 1983.

Robinson, P. M., "Review of Various Approaches to Power Spectrum Estimation." In : D. R. Brillinger and P. R. Krishnaish, eds., HANDBOOK OF STATISTICS, Vol. 3, New York : Elsevier Science, 1983.

Salas, J. D., "RANGE ANALYSIS FOR STORAGE PROBLEMS OF PERIODIC STOCHASTIC PROCESSES." Ph. D. Dissertation, Dep. of Civ. Eng., Colorado State University, Fort Collins, Co., 1972.

Salas, J. D. and V. Yevjevich, "Stochastic Structure of Water Time Series." Hydrology Paper 52, Colorado State University, Fort Collins, Colorado, 1972.

Salas, J. D., J. W. Delleur, V. Yevjevich and W. L. Lane, "APPLIED MODELING OF HYDROLOGIC TIME SERIES." Water Resources Publications, Littleton, Colorado, 1980.

Salas, J. D., D. C. Boes and R. A. Smith, "Estimation of ARMA Models with Seasonal Parameters." Water Resour. Res., 18(4), 1006-1010, 1982.

Shafer, B. A., E. B. Jones and D. M. Frick, "Snowmelt Runoff Modeling in Simulation and Forecasting Models with the Martinec-Rango Model." NASA Report CR 170452, 1982.

Shumway, R. H., "APPLIED STATISTICAL TIME SERIES ANALYSIS", New Jersey: Prentice Hall, 1988.

Sim, C. H., "A mixed Gamma ARMA(1,1) Model for River Flow Time Series." Water Resour. Res., 23(1), 32-36, Jan., 1987.

Tabios, G. Q. and J. D. Salas, "Forecasting Snowmelt Runoff." Proceedings of the International Symposium on Hydrometeorology, American Water Resources Association, June, 1982.

Thrall T., "Computer Programming of Spectrum Estimation." In : D. R. Brillinger and P. R. Krishnaish, eds., HANDBOOK OF STATISTICS, Vol.3, 409-437, New York : Elsevier Science, 1983.

Thompstone, R. M., "TOPICS IN HYDROLOGICAL TIME SERIES MODELING." Ph. D. Thesis, Dep. of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, 1983.

Thompstone, R. M., K. W. Hipel and A. I. McLeod, "Forecasting Quarter-Monthly Riverflow." Water Resources Bulletin, 21(5), 731-741, Oct., 1985.

Thomas, H. A. and Fiering, M. B., "Mathematical Synthesis of Streamflow Sequences for the Analysis of River Basins by Simulation." In : A. Mass et al. eds., DESIGN OF WATER RESOURCE SYSTEMS, 459-493, Massachusetts : Harvard University Press, 1962.

Tukey, J. W., "An Introduction to the Calculations of Numerical Spectrum Analysis." In : B. Harris eds., SPECTRAL ANALYSIS OF TIME SERIES, New York : Wiley, 1967.

U.S. Soil Conservation Service, "Soil Conservation Service Nation Engineering Handbook." Section 22 : Snow Survey and Water Supply Forecasting, Chapter 6, pp.6.1-6.13, 1970.

Vecchia, A. V., "AGGREGATION AND ESTIMATION FOR PERIODIC AUTOREGRESSIVE MOVING AVERAGE MODELS." Ph. D. Dissertation, Colorado State University, Fort Collins, Colorado, 1983.

Wahba, Grace, "Estimation of the Coefficients in Multidimensional Distributed Lag Model." Econometrica. 37: 398-407, 1969.

Yevjevich, V., "The Structure of Inputs and Outputs of Hydrologic Systems." In: V. Yevjevich ed., SYSTEMS APPROACH TO HYDROLOGY, Proceedings of the First Bilateral U.S.-Japan Seminar in Hydrology, Honolulu, Water Resources Publications, Littleton, Colorado, 1971.

Yevjevich, V., "STOCHASTIC PROCESSES IN HYDROLOGY." Water Resources Publications. Fort Collins, Colorado, 1972.