A SYSTEMS APPROACH TO HYDROLOGIC TIME SERIES ANALYSIS

by

Timothy Chun Ta Hsu
B.E. (National Taiwan University)
M.Sc. (Queen's University)

A thesis submitted to the
School of Graduate Studies
in partial fulfilment of the
requirements for
the degree of
Doctor of Philosophy

Department of Civil Engineering
University of Ottawa
Ottawa, Canada
April 1977

SUMMARY

The objective of this study was a) to carry out extensive investigation concerning the relation between deterministic and stochastic modelling of the runoff component of the hydrologic cycle; and b) to establish a multivariate watershed hydrologic system model involving meteorological data as the input and river flow as the output of the system. Monthly hydrologic time series of runoff, temperature and precipitation were selected for analysis.

Analytical relationships between the linear stochastic differential equations (S.D.E.) and the difference equations of autoregressive-moving average (ARMA) type models were established. It was found that a first-order S.D.E. or its equivalent first-order autoregressive, ARMA(1,0), process was adequate to model the monthly runoff generating process. Through the regional multiple regression analysis, the constant coefficient in the first-order S.D.E. was found to bear the storage effect of the watershed system.

Multivariate system analysis of the monthly runoff, temperature and precipitation time series indicated that coordinated use of the meteorological and hydrometric data would enchance the accuracy of estimation of runoff characteristics. A first-order ARMA transfer function model was found adequate to describe the multivariate watershed hydrologic system for the monthly runoff and meteorological time series.

ACKNOWLEDGMENTS

The author is indebted to his supervisor, Dr. K. Adamowski, for his help and encouragement throughout the work.

He also wishes to express his gratitude to Dr. N. U. Ahmed, Department of Electrical Engineering, University of Ottawa, and Dr. V. Klemes, Hydrology Research Division, Department of the Environment, for their helpful suggestions on this study.

Further thanks are extended to Mr. A. S. Qureshi, Head, Hydrometric Network Planning and Forecasting Section, Department of the Environment, for the permission to use the hydro-meteorologic and physiographic data from the 'hydrologic data bank' files.

The financial support given by the National Research Council (Grant No. A7467) while the author was a full-time graduate student at the University of Ottawa is very much appreciated.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Background

The hydrologic cycle due to its tremendous complexity can not, at present, be described exactly. Therefore, scientists and engineers have developed many simplified approaches, i.e., models, mainly for the analysis of practical problems.

One approach, which is adopted in this study, is to consider the hydrologic cycle as a system with inputs and outputs in the form of time series. When considering surface water hydrology, a watershed can be assumed as an open system in which feedback action either does not exist or does not exert significantly any control of the system (Chow, 1969). A natural watershed receives meteorological inputs from the atmosphere and generates outputs in the form of evapotranspiration and runoff. Problems arising with systems usually consist of identification, prediction, detection and synthesis. Among the above, identification of the system is almost the pre-requirement to all the other system problems. This study is then confined to the identification, or modelling of the watershed hydrologic system.

In mathematical modelling of a watershed hydrologic system, the underlying hydrologic process is represented by a more or less simplified mathematical formation, mainly in the form of differential equations, with derivatives both in time and in space, which are solved not by calculus integration but by increments both in time and space, mainly using digital computers.

There are, in general, two major approaches in mathematical modelling: the first is to design a deterministic (or parametric) model whose response is equivalent to a physical system. This may be done by either through a black-box model (O'Donnell, 1960; Amorocho, 1967) or through a series of equations which are equivalent to the actual internal physical physics of the system (Crawford and Linsley, 1964; Lichty, Dawdy and Bergman, 1968; Hsu, 1970). The second major approach is to determine the statistical parameters describing the response of the system, and use these statistics to generate a record which is statistically indistinguishable from the measured record. This is called stochastic simulation. (Thomas and Fiering, 1962; O'Connell, 1971; Adamowski and Hsu, 1974).

Over centuries, however, there seemed to be a lack of communication between the two schools of thoughts, namely, determinism and probabilism. In hydrology this fact was also reflected by the dichotomy between deterministic and stochastic approaches to modelling. Although it is usually accepted among researchers in hydrology that best results would be obtained in research when hydrologic processes are simultaneously investigated by both deterministic and stochastic methods of analysis, traditional thinking of treating them separately has kept this most promising approach, namely, combined deterministic - stochastic approach, far from being shaped let alone well tested.

It would be difficult to find a pure deterministic hydrologic process in nature. Pure deterministic process can be realized only under strictly controlled conditions. When physical, chemical or biological deterministic process work under hydrologic natural conditions, they are nearly always superposed by various sources of stochastic variations. In continuously changing either the space or the time scale or both, transitions take place in describing an observed hydrologic phenomenon from a deterministic to a stochastic approach and the converse. The best illustrative example would be considering water movement from the deterministic laminar flow analysis in classic fluid mechanics to the stochastic annual flow analysis of a watershed. A comprehensive and thorough discussion on "Determinism and Stochasticity in Hydrology" was given in a recent paper by Yevjevich (1974).
Some attempts in establishing the structural relationship between deterministic and stochastic approaches to watershed modelling were given by Kisiel (1967), Adamowski (1969), Quimpo (1971, 1974), Dooge (1972), Solomon (1972a), and Spolia and Chander (1974). These attempts, however, are either preliminary investigations or of only theoretical consideration restricted to a particular type of conceptual model, i.e. Nash's (1960) linear reservoir catchment model.

Furthermore, in the past two decades, 'Joint Mapping' (Solomon, 1972b) or coordinated use of the meteorological and hydrometric data, from the simplest coordinated tracing of precipitation and runoff isolines (Coulson, 1967) to the most sophisticated deterministic watershed model (Crawford and Linsley, 1964) have been proved to greatly enhance the accuracy of estimation of river flow characteristics. This branch of deterministic (or parametric) response hydrology is in full progress at present. Its counterpart - stochastic system modelling, however, still remains in the state of investigating isolated hydrologic phenomenon, in most cases, precipitation or river flow series. Therefore, it is felt that in order to explore the physical implication of a stochastic system model and to improve its efficiency in describing river flow characteristics, it would be desirable to apply the same principle of joint mapping, i.e., coordinated use of meteorological and hydrometric data to stochastic modelling of a watershed hydrologic system.

Objectives

Therefore, the objective of this study is two-fold:

Firstly, to carry out an extensive investigation concerning the mathematical formation and its underlying physical interpretation of commonly used stochastic models so that more definite conclusions may be drawn with respect to the relation between deterministic and stochastic modelling of watershed hydrology.

Secondly, to establish a multivariate stochastic hydrologic system model involving meteorological data as the input and river flow as the output of the system so that understanding of the multivariate hydrologic system behaviour may be obtained and accuracy in estimating river flow characteristics may be improved.

Scope

Bearing in mind the above objectives, monthly time series of river flow, precipitation and temperature are considered in this study. Annual river flow series were found (Yevjevich, 1964), in most cases except river basins with large storage, to be nearly purely random process which can not be described by any deterministic approach. On the other hand, daily flow series may well be described by a deterministic model (Watt and Hsu, 1971), a combined deterministic-stochastic model (Solomon, 1972a) or a stochastic model (Quimpo, 1967). In daily flow modelling, however, system formation is more complicated. Behaviour of various subsystems must be properly modelled in order to reach more precise description of river flow characteristics. As the time interval increases, temporal and spacial variations, or transients, in watershed responses due to various subsystems are smoothed, a lumped linear time-invariant system may be assumed without losing its physical generality. Monthly flow model taking advantage of the above system property, therefore, is considered here as a simple yet heuristic tool for the demonstration of the relation between deterministic and stochastic system modelling. Operationally, they could also be used for flood control, irrigation and other water management purposes.

Since this study is confined to the analysis of monthly hydrologic time series, some basic assumptions in carrying out the time series analysis may be made, namely, stationary time series and linear time-invariant system, which are briefly outlined in Chapter 2. It also covers the literature review on the applications of time series analysis in hydrology. In Chapter 3, the relation between deterministic and stochastic modelling is studied through the examination of the structure of stochastic differential equations. Establishment of analytical relationships between stochastic differential and difference equations is attempted in terms of their system impulse response functions. The relationships are applied to a number of selected watersheds for the identification of a monthly runoff generating process. Physical implication of the proposed stochastic model is sought through the multiple regression analysis between model parameters and watershed physiographic characteristics. Chapter 4 is then devoted to the identification of a multivariate watershed hydrologic system which receives meteorological inputs of temperature and precipitation and generates output in the form of runoff. General theory in multivariate linear systems, multiple cross spectral analysis and transfer function system models are discussed. The application of the multivariate system approach is then carried out for some watershed where both meteorological and hydrometric observations are available. The procedures of time series analysis in both Chapters 3 and 4 are presented in the flowchart of Figure 1.1. Since this thesis was intended to be a comprehensive study on hydrologic time series analysis, tremendous computational and programming efforts were involved which resulted in a vast amount of computed information. It was impossible to tabulate and plot all the results in the text. Only pertinent results were listed. However; a typical example in graphical form was always given to illustrate the approaches used in the study. Finally, major findings of the study and recommendations for future research are presented in Chapter 5.

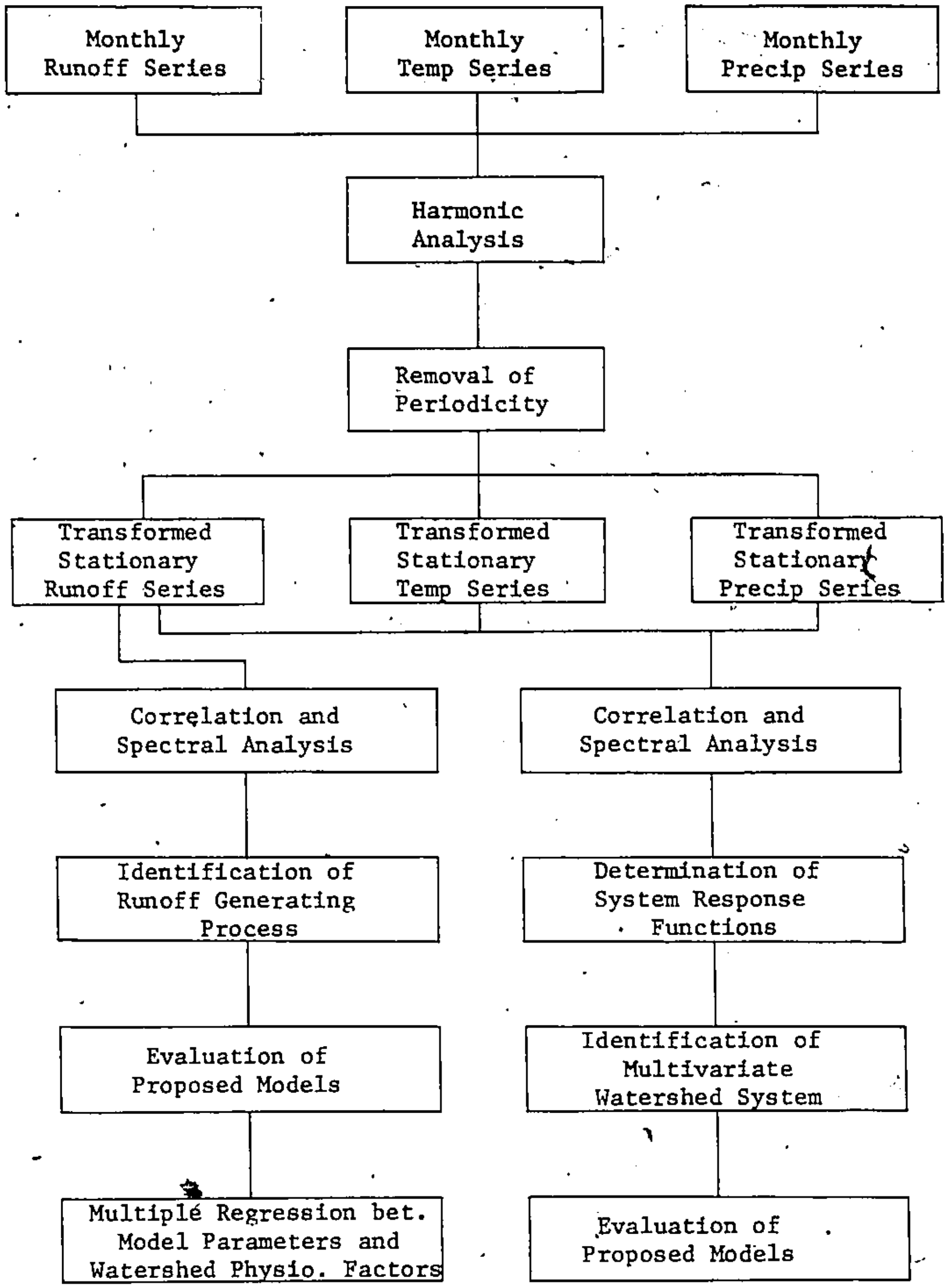


FIGURE 1.1 Procedures of Monthly Hydrologic Time Series Analysis

CHAPTER 2

LITERATURE REVIEW

Introduction

Observations of any hydrologic time series are finite. Therefore, it is extremely difficult to infer from these finite observations the nature of the hydrologic process from which the observed time series is regarded as being generated. Some simplifications have to be made in order to make the analysis of observed time series tractable and yet fruitful.

The most important assumptions made about a time series are that the corresponding generating process is stationary and that a stationary process may be adequately described by the lower moments of the time series, i.e., the mean, variance, covariance function in the time domain and the variance spectrum in the frequency domain. Another important assumption in studying the underlying generating process is to assume a time-invariant linear system. In this case, system characteristics such as impulse response functions and frequency response functions can also be determined from the lower moments of the involved time series. These simplifying assumptions are briefly described in the following sections.

Time Series Analysis

Stationary time series

A general way of defining a stochastic process $X(t)$ is by means of a joint distribution function $F(x)$ for any finite number of observations $x_1, x_2, \ldots, x_n$ made at time intervals $t_1, t_2, \ldots, t_n$; that is

$$F(x_1, x_2, \ldots, x_n;\ t_1, t_2, \ldots, t_n)$$

$$= P(X(t_1) \le x_1;\ X(t_2) \le x_2, \ldots, X(t_n) \le x_n) \qquad (2.1)$$

where $P(.)$ is the joint probability that $X(t_i) \le x_i$, $i = 1,2,\ldots,n$.

The stochastic process is said to be stationary in the strict sense if the distribution functions are uneffected by a shift 'u' in the time origin, i.e.,

$$F(x_1, x_2, \ldots, x_n;\ t_1, t_2, \ldots, t_n)$$

$$= F(x_1, x_2, \ldots, x_n;\ t_1+u, t_2+u, \ldots, t_n+u). \qquad (2.2)$$

In other words, for a stationary process the joint distribution function does not depend upon $t_1, t_2, \ldots, t_n$, but upon the differences $t_j - t_{j+1} = u_j$, $j = 1,2,\ldots,(n-1)$, so that

$$F(x_1, x_2, \ldots, x_n;\ t_1, t_2, \ldots, t_n)$$

$$= F(x_1, x_2, \ldots, x_n;\ u_1, u_2, \ldots, u_{n-1}). \qquad (2.3)$$

A special case, which is of particular interest in theory and practice, is when the stationarity is confined to the lower moments of the time series and is called weakly stationary process (Matalas, 1966). The lower moments include the mean, variance, covariance function in the time domain and the variance spectrum in the frequency domain.

Autocorrelation and variance spectrum

The mean and covariance function of a stationary process $X(t)$ are defined respectively as follows:

$$\mu_x = E(X(t)) = \int_{-\infty}^{\infty} x(t)dF(x_t), \tag{2.4}$$

$$\gamma_{xx}(u) = E((X(t)-\mu_x)(X(t+u)-\mu_x)) = \int_{-\infty}^{\infty} (x(t)-\mu_x)(x(t+u)-\mu_x)dF(x_t,x_{t+u};u), \tag{2.5}$$

where $E(.)$ denotes the expectation of a random variable or function of random variable(s).

From eq. (2.5), the variance σ_x^2 is obtained when $u = 0$, i.e.,

$$\sigma_x^2 = \gamma_{xx}(0). \tag{2.6}$$

Sample estimates for the mean and covariance in the best mse (mean square error) sense (Jenkins and Watts, 1968) are given by

$$\hat{\mu}_x = \bar{x} = \frac{1}{N}\sum_{n=1}^{N} x_n, \tag{2.7}$$

$$\hat{\gamma}_{xx}(u) = c_{xx}(k) = \frac{1}{N}\sum_{n=1}^{N-k} (x_n-\bar{x})(x_{n+k}-\bar{x}), \tag{2.8}$$

where N is the number of observations x_n at interval Δt and $u = k\Delta t$. In practice, Δt is usually taken as one time unit, i.e., $\Delta t = 1$.

The autocorrelation function is defined as

$$\rho_{xx}(u) = \frac{\gamma_{xx}(u)}{\gamma_{xx}(0)} \tag{2.9}$$

and its sample estimate is

$$r_{xx}(k) = \frac{c_{xx}(k)}{c_{xx}(0)} \ . \tag{2.10}$$

The covariance $\gamma_{xx}(u)$ is non-negative definite and can be represented in the form

$$\gamma_{xx}(u) = \int_{-\infty}^{\infty} e^{i2\pi fu} S_{xx}(f)df \tag{2.11}$$

and

$$\gamma_{xx}(0) = \sigma_x^2 = \int_{-\infty}^{\infty} S_{xx}(f)df \tag{2.12}$$

where the function $S_{xx}(f)$ is non-negative and is called variance spectrum (or simply autospectrum) since it shows how the variance of the X(t) process is distributed over the total frequency range.

If the X(t) process is real which is true for hydrologic time series, then $\gamma_{xx}(u)$ may be rewritten as

$$\gamma_{xx}(u) = 2 \int_{0}^{\infty} \cos 2\pi fu \ S_{xx}(f)df \tag{2.13}$$

which can be inversely transformed to give

$$S_{xx}(f) = \int_{-\infty}^{\infty} e^{-i2\pi fu} \gamma_{xx}(u)du \tag{2.14}$$

$$= 2 \int_{0}^{\infty} \cos 2\pi fu \gamma_{xx}(u)du. \tag{2.15}$$

The sample estimate of $S_{xx}(f)$ for a finite discrete time series can be obtained from

$$\hat{S}_{xx}(f) = c_{xx}(0) + 2 \sum_{k=1}^{N-1} c_{xx}(k)\cos 2\pi fk, \; -\frac{1}{2} \le f \le \frac{1}{2} \tag{2.16}$$

where the frequency $f_N = \frac{1}{2}$ is called Nyquist frequency and is the highest frequency which can be detected with data sampled at unit intervals.

Further details on spectral representations can be found in Bartlett (1966), Yaglom (1962) and Blackman and Tukey (1958).

In the subsequent sections sample estimate representation of various statistics will not be included until the discussion of computational procedures. Only continuous time series will be covered in most of the theoretical discussions.

Cross correlation and cross spectrum

The cross covariance between two stationary time series X(t) and Y(t) is defined as

$$\gamma_{xy}(u) = \int_{-\infty}^{\infty} \cdot \int_{-\infty}^{\infty} (x(t)-\mu_x)(y(t+u)-\mu_y)dF(x_t,y_{t+u};u) \tag{2.17}$$

where $F(x_t,y_{t+u};u)$ is the joint distribution function of the variates x_t and y_{t+u}.

The cross correlation is thus obtained by

$$\rho_{xy}(u) = \frac{\gamma_{xy}(u)}{\sqrt{\gamma_{xx}(0)\gamma_{yy}(0)}} . \tag{2.18}$$

For real processes, $\gamma_{xy}(u)$ has the spectral representation

$$\gamma_{xy}(u) = 2\int_0^{\infty} \cos 2\pi fu C_{xy}(f)df - 2\int_0^{\infty} \sin 2\pi fu Q_{xy}(f)df \tag{2.19}$$

$$= \int_{-\infty}^{\infty} e^{i2\pi fu} S_{xy}(f)df \tag{2.20}$$

where $S_{xy}(f) = C_{xy}(f) - iQ_{xy}(f)$. The function $S_{xy}(f)$ is called the cross spectrum; $C_{xy}(f)$ is the cospectrum which measures the contribution of different frequencies to the total covariance with zero lag of the two time series; $Q_{xy}(f)$ is the quadrature spectrum which measures the contribution when the waves of one time series are delayed by a quarter period. Now $S_{xy}(f)$ can be expressed in the form

$$S_{xy}(f) = A_{xy}(f)e^{i\phi_{xy}(f)}, \tag{2.21}$$

where the amplitude $A_{xy}(f) = \sqrt{C_{xy}^2(f) + Q_{xy}^2(f)}$, (2.22)

and the phase angle $\phi_{xy}(f) = \tan^{-1}\frac{-Q_{xy}}{C_{xy}}$. (2.23)

In order to measure the linear relationship in the frequency domain between two time series, the squared coherency is defined as

$$K_{xy}^2(f) = \frac{C_{xy}^2(f) + Q_{xy}^2(f)}{S_{xx}(f)S_{yy}(f)}, \tag{2.24}$$

where $S_{xx}(f)$ and $S_{yy}(f)$ are the spectrums of X(t) and Y(t) series respectively as defined in the previous section.

The residual spectrum then takes the analogy to the residual variance in the time domain, i.e.,

$$R_{xy}(f) = S_{yy}(f)(1-K^2_{xy}(f)), \tag{2.25}$$

where $R_{xy}(f)$ is the residual spectrum.

The inversion formulae for $C_{xy}(f)$ and $Q_{xy}(f)$ are

$$S_{xy}(f) = \int_{-\infty}^{\infty} e^{-i2\pi fu}\gamma_{xy}(u)du, \tag{2.26}$$

$$C_{xy}(f) = \int_{0}^{\infty} \cos 2\pi fu(\gamma_{xy}(u) + \gamma_{xy}(-u))du , \tag{2.27}$$

$$Q_{xy}(f) = \int_{0}^{\infty} \sin 2\pi fu(\gamma_{xy}(u) - \gamma_{xy}(-u))du . \tag{2.28}$$

Detailed descriptionsof cross spectral analysis can be found in Granger and Hatanaka (1964) and Jenkins and Watts (1968).

Linear time-invariant systems

Amorocho (1967) generalized the relationship between the input and the output of a time-variable non-linear system with lumped parameters as follows:

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} h_1(t;u)x(u)du + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_2(t;u_1,u_2)x(u_1)x(u_2)du_1du_2 \\ &+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_3(t;u_1,u_2,u_3)x(u_1)x(u_2)x(u_3)du_1du_2du_3 + \ldots \\ &= \sum_{n=1}^{\infty} \underset{(n)}{\int_{-\infty}^{\infty}} h_n(t;u_1,u_2,\ldots,u_n) \prod_{i=1}^{n} x(u_i)du_1du_2\ldots du_n \end{aligned} \tag{2.29}$$

where x(t) and y(t) are input and output respectively of the system; h_i is i^{th} order weighting function or kernel function of the system and $\prod_{i=1}^{n} x(u_i) = x(u_1)x(u_2)\ldots x(u_n)$. Equation (2.29) is also the well-known Volterra functional.

Each one of the terms in eq. (2.29) is a functional; it may be viewed as representing the operation of a separate component of the total system, and its nonlinearity is defined by the order of the integration. Thus the first term represents a linear or first order system, the second term represents a quadratic or second order system, and so on. These component systems are said to be pure n^{th} order systems operating on the input function and its product combinations. The sum of the outputs of these systems is equal to the output of the total, fully non-linear system.

In practice, a linear time-invariant system is usually assumed, i.e.,

$$y(t) = \int_{-\infty}^{\infty} h(u)x(t-u)du, \tag{2.30}$$

where the weighting function h(t) is also called the impulse response function and satisfies the condition $h(t) = 0$ for $t < 0$ in order to ensure the physical reality.

In the discrete-time case, the above input-output relationship can be written as

$$y_n = \sum_{r=0}^{\infty} h_r x_{n-r} \tag{2.31}$$

where h_r's are constants independent of time 't'.

Linear processes

It can be shown that a linear system in the form of eq. (2.30) or eq. (2.31) is in general a solution to a linear differential equation

$$(\alpha_m D^m + \ldots + \alpha_1 D + \alpha_0)y(t) = x(t) \text{ for continuous time} \qquad (2.32)$$

where $D^m = \frac{d^m}{d_t{}^m}$, $\alpha_0, \alpha_1, \ldots, \alpha_m$ are constants and the roots of the characteristic equation $\alpha_m D^m + \ldots + \alpha_1 D + \alpha_0 = 0$ have negative real parts for a stable system*; or a solution to a linear difference equation

$$(1 - a_1 B - \ldots - a_m B^m)y_n = x_n \quad \text{for discrete time} \qquad (2.33)$$

where $B^m y_n = y_{n-m}$, $a_1, a_2, \ldots, a_m$ are constants and the roots of the characteristic equation $1-a_1 B - \ldots - a_m B^m = 0$ must lie outside the unit circle for a stable system.

Equations (2.32) and (2.33) are called a stochastic differential equation and a stochastic difference equation respectively if the input $X(t)$ or X_n and/or the coefficients α's or a's are stochastic in nature. Consequently the output $Y(t)$ or Y_n is also a stochastic process. In this study, the coefficients α's and a's are assumed constant and the input and output are considered the only stochastic part of the system. Furthermore, if the input process $X(t)$ or X_n is a purely random process

---

*A system is said to be stable if a bounded input produces a bounded output, or in terms of the system expression (2.30)

$$\int_0^\infty |h(u)|du < K$$

where K is a finite constant. (Jenkins and Watts, 1968).

or white noise, i.e., $\rho_{xx}(0) = 1$ and $\rho_{xx}(u) = 0$ when $|u| > 0$, the output process $Y(t)$ or Y_n is called a linear process.

The linear processes $Y(t)$ or Y_n from equations (2.30) and (2.31) and their corresponding stochastic differential or difference equations (2.32) and (2.33) are called moving average processes and autoregressive processes respectively. The equivalence between equations (2.30) and (2.31) and equations (2.32) and (2.32) also show that an infinite moving average process may be represented by a finite autoregressive process. Sometimes it may be beneficial to write down a mixed autoregressive-moving average (ARMA) process (Box and Jenkins, 1970).

The autocorrelation function for a moving average process, i.e., equations (2.30) and (2.31), is

$$\rho_{yy}(u) = \frac{\int_0^\infty h(v)h(v+u)dv}{\int_0^\infty h^2(v)dv} \qquad \text{for continuous time} \tag{2.34}$$

$$\text{or } \rho_{yy}(k) = \frac{\sum_{j=0}^{\infty} h_j h_{j+k}}{\sum_{j=0}^{\infty} h_j^2} \qquad \text{for discrete time;} \tag{2.35}$$

whereas for an autoregressive process of order 'm', i.e., equations (2.32) and (2.33),

$$(\alpha_m D^m + \dots + \alpha_1 D + \alpha_0)\rho_{yy}(u) = 0, \qquad (u > 0)$$
$$\text{for continuous time} \tag{2.36}$$

$$\text{or} \quad \rho_{yy}(k) = a_1\rho_{yy}(-1) + \dots + a_m\rho_{yy}(k-m), \quad (k \geq 1),$$
$$\text{for discrete time} \tag{2.37a}$$

$$1 = a_1\rho_{yy}(1) + \ldots + a_m\rho_{yy}(m) + \frac{\sigma_x^2}{\gamma_{yy}(0)}. \qquad (k=0) \qquad (2.37b)$$

Detailed discussions on linear processes can be found in Jenkins and Watts (1968).

System response functions

A linear system can be described in the time domain by the impulse response function and in the frequency domain by the frequency response function.

As indicated in the previous section, the impulse response function is defined as the weighting function $h(u)$ or h_r in the convolution integral (2.30) or the weighted sum (2.31):

$$y(t) = \int_0^\infty h(u)x(t-u)du, \qquad (2.30)$$

and

$$y_n = \sum_{r=0}^{\infty} h_r x_{n-r} . \qquad (2.31)$$

It measures the output at time 't' of a system subject to a unit impulse at $t = 0$.

The frequency response function, on the other hand, measures the response of a system to a periodic signal $e^{i2\pi ft}$ at a specified frequency 'f' in the following way,

$$\begin{aligned} y(t) &= \int_0^\infty h(u)e^{i2\pi f(t-u)}du \\ &= e^{i2\pi ft} \int_0^\infty h(u)e^{-i2\pi fu}du \\ &= H(f)e^{i2\pi ft} \qquad (2.38) \end{aligned}$$

where $$H(f) = \int_0^\infty h(u)e^{-i2\pi fu}du \tag{2.39}$$

is called the frequency response function. Hence the frequency response function is the Fourier transform of the impulse response function. For discrete time, the frequency response is obtained from

$$H(f) = \sum_{r=0}^{\infty} h_r e^{-i2\pi fr}, \quad -\frac{1}{2} \leq f \leq \frac{1}{2}. \tag{2.40}$$

It can also be shown that

$$Y(f) = H(f)X(f), \tag{2.41}$$

where $Y(f)$ and $X(f)$ are Fourier transforms of $Y(t)$ and $X(t)$ respectively; and

$$S_{xy}(f) = H(f)S_{xx}(f). \tag{2.42}$$

The frequency response function can be rewritten in the form

$$H(f) = G(f)e^{i\phi(f)} \tag{2.43}$$

where $G(f)$ and $\phi(f)$ are called gain and phase respectively of the frequency response function. From equations (2.21) and (2.42), it can be seen that

$$A_{xy}(f) = G(f)S_{xx}(f), \tag{2.44a}$$

and $$\phi_{xy}(f) = \phi(f). \tag{2.44b}$$

The impulse response function can be obtained through inverse Fourier transformation of the frequency response function:

$$h(u) = \int_{-\infty}^{\infty} H(f)e^{i2\pi fu}df \quad \text{for continuous time} \tag{2.45}$$

and

$$h_r = \int_{-1/2}^{1/2} H(f)e^{i2\pi fr}df \quad \text{for discrete time} \tag{2.46}$$

Thorough theoretical discussions on linear systems and system response functions are given by Wiener (1949) and Jenkins and Watts (1968).

Some statistical considerations in estimation

Bartlett (1946) gave approximate variances and covariances of the estimated autocorrelations of a stationarity Normal process,

$$\text{Var}(r_k) \approx \frac{1}{N} \sum_{v=-\infty}^{\infty} \rho_v^2 \quad \text{for large k,} \tag{2.47}$$

$$\text{Cov}(r_k, r_{k+s}) \approx \frac{1}{N} \sum_{v=-\infty}^{\infty} \rho_v \rho_{v+s} \quad \text{for large k .} \tag{2.48}$$

From eq. (2.48), it can be seen that there will be strong correlation between the autocorrelations if the correlation in the original series is moderately strong.

Anderson (1942) showed that for an independent Normal process and k << N, the distribution of the estimated autocorrelations is approximately Normal with the variance

$$\text{Var}(r_k) \approx \frac{1}{N}, \tag{2.49}$$

which is the special case of eq. (2.47) when the original series is a purely random process.

For the sample spectral estimate, it can be shown (Jenkins and Watts, 1968) that

$$E(\hat{S}_{xx}(f)) \approx S_{xx}(f), \tag{2.50}$$

$$\mathrm{Cov}(\hat{S}_{xx}(f_1), \hat{S}_{xx}(f_2)) \approx 0(\frac{1}{N^2}), \tag{2.51}$$

and

$$\mathrm{Var}(\hat{S}_{xx}(f)) \approx S_{xx}^2(f). \tag{2.52}$$

Hence to a good degree of approximation, spectral estimates at spacing greater than $\frac{1}{N}$ may be regarded as uncorrelated. However, the variance of the estimate does not decrease with the sample size.

Furthermore, $\hat{S}_{xx}(f)$ has the property that $2\hat{S}_{xx}(f)/S_{xx}(f)$ is approximately distributed as χ_2^2, i.e., a chi-squared statistic with two degrees of freedom. Whereas at zero frequency and at the folding frequency, $f_N = \frac{1}{2}$ for discrete time, $\hat{S}_{xx}(f)/\hat{S}_{xx}(f)$ is approximately distributed as a χ_1^2.

In order to reduce the variance of $\hat{S}_{xx}(f)$, some smoothing procedures are introduced. In the time domain, a lag window is defined as follows (Jenkins and Watts, 1968):

$$w(0) = 1, \tag{2.53a}$$

$$w(u) = w(-u) > 0, \quad \text{for } |u| < M \tag{2.53b}$$

$$w(u) = 0, \quad \text{for } |u| \geq M,\ M < N. \tag{2.53c}$$

Then the smoothed spectral estimate for discrete time is obtained as

$$\bar{S}_{xx}(f) = c_{xx}(0) + 2 \sum_{k=1}^{M} w(k)c_{xx}(k)\cos 2\pi fk. \tag{2.54}$$

Using the convolution property, $\bar{S}_{xx}(f)$ may be rewritten

$$\bar{S}_{xx}(f) = \int_{-1/2}^{1/2} W(f'-f)S_{xx}(f')df', \qquad (2.55)$$

where $W(f) = \sum_{k=-M}^{M} e^{i2\pi fk} w(k)$ is called the spectral window in the frequency domain.

The variance for the smoothed spectral estimates then becomes

$$\mathrm{Var}(\bar{S}_{xx}(f)) \approx \frac{S_{xx}^2(f)}{N} \int_{-M}^{M} w^2(u)du. \qquad (2.56)$$

For example, by introducing the Bartlett lag window

$$w(u) = 1 - \frac{|u|}{M}, \qquad \text{for } |u| < M \qquad (2.57a)$$

$$w(u) = 0. \qquad \text{for } |u| \geq M \qquad (2.57b)$$

it can be shown that

$$\mathrm{Var}(\bar{S}_{xx}(f)) \approx 0.67(\frac{M}{N})S_{xx}^2(f). \qquad (2.58)$$

Thus the variance is significantly reduced to at least 67% of the variance of the sample spectrum. The bias of the smoothed estimate is approximately proportional to a certain order to the equivalent bandwidth of the spectral window, which is defined as

$$b = \frac{1}{\int_{-M}^{M} w^2(u)du} = \frac{1}{\int_{-1/2}^{1/2} W^2(f)df} . \qquad (2.59)$$

It can be seen that the variance of the spectral estimate is inversely proportional to the bandwith of the spectral window:

Variance x Bandwidth = Constant

Hence conflict exists between the variance and the bias of the spectral estimate, a compromise must be made in choosing the proper bandwidth, in other words the proper spectral window and the truncation point M.

The statistic $v\bar{S}_{xx}(f)/S_{xx}(f)$ is approximately distributed as a χ_v^2 with degrees of freedom 'v' given by

$$v = \frac{2N}{\int_{-M}^{M} w^2(u)du}. \tag{2.60}$$

The same smoothing techniques can be applied to the cross spectral analysis. The smoothed estimates for discrete time are as follows:

$$\bar{S}_{xy}(f) = \sum_{k=-M}^{M} e^{-i2\pi fk} w(k) c_{xy}(k), \tag{2.61}$$

$$\bar{C}_{xy}(f) = c_{xy}(0) + \sum_{k=1}^{M} w(k)(c_{xy}(k) + c_{yx}(k))\cos 2\pi fk, \tag{2.62}$$

$$\bar{Q}_{xy}(f) = \sum_{k=1}^{M} w(k)(c_{xy}(k) - c_{yx}(k))\sin 2\pi fk . \tag{2.63}$$

Hence

$$\bar{\phi}_{xy}(f) = \tan^{-1} \frac{-\bar{Q}_{xy}(f)}{\bar{C}_{xy}(f)}, \tag{2.64}$$

$$\bar{K}^2_{xy}(f) = \frac{\bar{C}^2_{xy}(f) + \bar{Q}^2_{xy}(f)}{\bar{S}_{xx}(f)\ \bar{S}_{yy}(f)}, \tag{2.65}$$

$$\bar{H}(f) = \frac{\bar{S}_{xy}(f)}{\bar{S}_{xx}(f)}, \tag{2.66}$$

and $$\overline{G}(f) = \frac{\sqrt{\overline{C}^2_{xy}(f) + \overline{Q}^2_{xy}(f)}}{\overline{S}_{xx}(f)}. \tag{2.67}$$

As suggested by Jenkins and Watts (1968), by applying Fisher's z-transformation

$$\overline{Z}_{xy}(f) = \tanh^{-1}(|\overline{K}_{xy}|) = \frac{1}{2}\ln \frac{1+|\overline{K}_{xy}|}{1-|\overline{K}_{xy}|}, \tag{2.68}$$

an approximate 100(1-a)% confidence interval for $\overline{Z}_{xy}(f)$ is

$$\overline{Z}_{xy}(f) \pm Z(1-\frac{a}{2})\sqrt{\frac{1}{2N}} \quad (\text{where } I = \int_{-M}^{M} w^2(u)du) \,. \tag{2.69}$$

For smoothed estimates $\overline{\phi}_{xy}(f)$ and $\overline{G}(f)$, the approximate 100(1-a)% confidence intervals are

$$\overline{\phi}_{xy}(f) \pm \sin^{-1}\sqrt{\frac{2}{v-2} F_{2,v-2}(1-a)\left(\frac{1-\overline{K}^2_{xy}(f)}{\overline{K}^2_{xy}(f)}\right)}, \tag{2.70}$$

and $$\overline{G}(f) \pm \overline{G}(f)\sqrt{\frac{2}{v-2} F_{2,v-2}(1-a)\left(\frac{1-\overline{K}^2_{xy}(f)}{\overline{K}^2_{xy}(f)}\right)}, \tag{2.71}$$

where $F_{2,v-2}$ is an F-statistic with degrees of freedom (2, v-2).

Applications of Time Series Analysis in Hydrology

Some detailed and heuristic reviews on hydrologic time series analysis have been given by Matalas (1966), Cavadias (1966) and Kisiel (1969).

Roesner and Yevjevich (1966) detected the periodic components in monthly precipitation series and monthly runoff series by performing the variance spectrum analysis. They found that the periodic component in both time series may be described by Fourier series, with the 12-month cycle and its subharmonics. By removing the periodic component and examining the remaining portion of the time series, they concluded that the monthly precipitation series was composed of a deterministic component of periodic movement and a nearly independent stochastic component; whereas the monthly runoff series was composed of a deterministic component of periodic movement and a highly time dependent stochastic component. The time dependence of the stochastic component could be described in most cases by a first-order Markov process, i.e., a first-order autoregressive process.

Using the same approach as above, Quimpo and Yevjevich (1967) analyzed daily streamflow series. First, periodic components were detected, in the daily means, and in standard deviations, and isolated by using spectral analysis. Then, autoregressive schemes were fitted to stationary stochastic series after the periodic components had been removed. Finally, stochastic components of these series were found to be well fitted by the second-order Markov linear model.

Adamowski (1969) applied the periodogram approach, which is equivalent to the spectral estimate using Bartlett's window given by eq. (2.57), to estimate the spectral density of the stochastic component of the daily river flow time series. By adopting Bartlett's test of homogeneity of variances, he found that sample spectral density estimates from fifteen to forty-five years of data did not differ significantly and suggested, for practical purposes, to use about fifteen years of daily records to obtain an estimation of the spectral density. Also, he found that the estimated spectra did not differ significantly from the spectrum of the second-order autoregressive representation. This confirms the applicability of the proposed second-order Markov model for the stochastic component of daily streamflow series. A second-order stochastic differential equation was obtained through the theoretical connection with the spectrum. Finally, a method of predicting the underlying time series was given based on the classical Wiener-Kolmogoroff theory. The computed prediction variance indicated that the stochastic component may be predicted only two days in advance at any reasonable level of prediction variance for the three rivers studied.

Rodriguez-Iturbe (1967) gave a detailed description on analysing a bivariate hydrologic process. He suggested that cross spectrum should be used for the construction of frequency response function whereas in prediction problems, it was natural to work in the time domain with auto- and cross correlation since prediction was done in time. No significant coherence were observed in the annual series of precipitation and runoff except in very close stations. Through partial cross spectral analysis was found that in the Pacific Coast of the United States the annual cycle in temperature was highly correlated with the annual cycle in precipitation, but the annual cycle in atmospheric pressure seemed to be related to the cycle in precipitation only through the cycle in temperature. Also the amplitude of the annual cycle in precipitation appeared to decrease when advancing in the northerly direction. On the other hand, no similar trend was observed in the runoff stations. A highly significant coherence between the annual cycle of hydrologic time series suggested the possibility to predict the annual oscillation in one station on the basis of the annual oscillation at another station. An example was given on a bivariate process involving two river stations with daily runoff data. Coherence analysis was found to be a powerful tool to determine the underlying process. The same concept of cross spectral analysis was applied to water and sediment discharge data in a later paper (Rodriguez-Iturbe and Nordin, 1968).

Chow and Ramaseshan (1965) modelled the hourly annual storm rainfalls by a first-order non-stationary Markov-chain model. From the historical data, 1,000 annual storms were generated sequentially by the Monte Carlo methods and then routed through a simulated basin system which was composed of a series of equal linear reservoirs to produce 1,000 generated floods for use in water resources planning and design. In a later study by Chow and Kareliotis (1970), a watershed was treated as the stochastic hydrologic system whose components of precipitation, runoff, storage and evapotranspiration were simulated as stochastic processes by time series models to be determined by correlogram and spectral analysis. The hydrologic system was then formulated on the basis of the simple water balance equation, and composed of deterministic periodic components which were detected from variance spectra and stochastic components which were merely assumed as mutually independent white noises. The above two investigations were further developed by Chow and Prasad (1972) by introducing the concept of storage reservoir theory to model the basin storage system.

In a pilot hydrologic forecasting study for southern Ontario, Solomon (1972a) demonstrated the relationship between the physical discharge recession equation and a lag-one serial correlation model, or first-order Markov-chain model. The recession characteristic was related to the lag-one serial correlation of the flow and the inflow was identified by a purely random process. Spatial transfer of the recession characteristic and the inflow was obtained through the correlation with basin physiographic characteristics. Therefore the serial correlation of flow was maintained due to the intrinsic serial correlation properties of the Markov model and the cross correlation of flows was also preserved by the way in which the inflow was synthesized.

An interesting study was carried out by Eriksson (1970a, 1970b) who investigated groundwater time series at 10-day intervals as well as streamflow, precipitation and temperature using spectral analysis. He found that the stochastic component of the groundwater level series might be well fitted by a second-order Markov process. The spectral density fluctuations of observation wells agreed fairly well at low frequencies but for higher frequencies there was hardly any similarity. He concluded that the representativity of groundwater fluctuations in a hydrologic region may be picked up at low frequencies while higher frequencies are more "local". The streamflow series was found to be fitted by a second-order Markov process as also found by many others. The precipitation series was fitted by a weak first-order Markov model

($\rho_{xx}(1) = 0.12$). The spectrum for the temperature was rather confounded due to some inherent property of the atmosphere. A simple stochastic model of groundwater reservoirs was developed based on a physical balance equation and an average infiltration rate was derived from the proposed generating process. A further study was made on records of precipitation, river water stage and groundwater levels using cross spectral analysis. The coherence between precipitation and the other variables was found extremely weak or absent. The river water level and groundwater levels showed appreciable coherence relatively independent of frequency, suggesting that variation in groundwater levels runs parallel to variations in river stage. The coherence between groundwater levels, however, was such that interaction was suggested. The phase shift between river water level and groundwater level was approximately 10-30 days with the groundwater levels lagging. The magnitude of phase shift seemed to depend on the distance between ground surface and groundwater level.

More rigorous mathematical treatment on the relation between the stochastic and the deterministic hydrologic system modelling was given in the studies by Quimpo (1971, 1973) and Spolia and Chander (1974). In the first study, the background of the mathematical formation of a linear hydrologic system was reviewed which disclosed the linkage between the solution of a linear differential equation and the discrete system representation of an infinite moving average process which in turn may be represented by a finite auto-regressive process. The effort was then concentrated in demonstrating the relation between

Nash's linear reservoir model and the stochastic streamflow model by examining the kernel functions, or impulse response functions in the two models and the resulting autocorrelation functions. Some field data in Nash's study (1960) were used to construct the correlograms for the purpose of validation. Along the same line, the second study dealt with a more general case, i.e., a mixed autoregressive-moving average type model in the form of differential equation for continuous time and difference equation (ARMA) for discrete time. The structural relationship between the parameters of the ARMA model and Nash's linear reservoir model was established. A second-order autoregressive model, or ARMA(2,0) was then proposed to model the surface runoff. It was shown that the proposed formulation corresponded to a conceptual model of two equal linear reservoir in series. The model was also investigated for its stability, nature of response function and range of parameter values.

The above reported studies dealt mostly with stationary time series and linear system models. For non-stationary time series, investigations have been carried out through time-variant reservoir theory (Chiu and Bittler, 1969; Mandeville and O'Donnell, 1973), or linear functional of derivatives of an exponentially smoothed function (Cavadias and Brisebois, 1971). Identification of non-linear system has been attempted by orthogonal functions (Amorocho and Brandstetter, 1971; Zand and Harder, 1973), modified stepwise regression analysis (Bidwell, 1971) or a constrained optimization scheme (Diskin and Boneh, 1973).

One characteristic common to most of the above mentioned studies is that the time series under study are either storm runoff or daily river flow which are generally recognized to be of more significant time-variant and non-linear in nature. For monthly hydrologic time series, however, it is thought that the effort of introducing the concept of non-stationarity and non-linearity would not be fully justified.

It can be observed from the above literature review that not many works have been accomplished in studying the possible relation between the deterministic (or parametric) and stochastic modelling of hydrologic time series, except for the studies by Quimpo (1971, 1973) and Spolia and Chander (1974) both of which are confined to the theoretical discussion of a special case of Nash's linear reservoir model (1960). Although the above topic was also slightly touched upon by some other authors (Adamowski, 1969; Eriksson, 1970a; Solomon, 1972a), their investigations in this respect generally lack sound theoretical development and the results can be considered only as preliminary. Therefore, it is intended in this study to devote more comprehensive theoretical discussion on the relation between the deterministic and stochastic modelling and to carry out more extensive data analysis to validate such relation.

Another observation from the literature review is that there is hardly anything done in the area of stochastic modelling of multivariate hydrologic system. In the work by Chow and Kareliotis (1970), a set of stochastic formulations of individual hydrologic time series was merely superimposed to a simple water balance equation. Multiple cross spectral analysis was introduced by Rodriguez-Iturbe (1967), but no further effort was made to model the watershed system. Therefore, it is also decided in this study to attempt a multivariate hydrologic system analysis making use of the available basic hydrologic time series, namely, runoff, temperature and precipitation. It is believed that this coordinated use of meteorological and hydrometric data should improve the accuracy in estimating runoff characteristics and shed more light upon the physical mechanism of the watershed system.

CHAPTER 3

STOCHASTIC DIFFERENTIAL EQUATION

vs.

STOCHASTIC DIFFERENCE EQUATION

Introduction

In deterministic hydrology, it is the presumption that the physics of each phenomenon in the hydrologic cycle are well-enough known to permit derivation of realistic dynamic equations in the form of differential equations in both time and space. Each derivative term in the dynamic equations bears its own physical significance and explanation to the conservation of mass, momentum and energy whichever suitable. To study the relation between deterministic and stochastic modelling, therefore, instead of merely investigating the stochastic difference equation of the autoregressive-moving average (ARMA) type, it is also necessary to look at its continuous equivalent - the stochastic differential equation.

The structural relationship between the stochastic differential equation and the stochastic difference equation can be established through examining the solutions of the impulse response functions in both cases. The attempt was first made by Bartlett (1946) to study the relation between a second-order stochastic differential equation model and a second-order stochastic difference equation model forced by a random impulse function. The theory was further extended by Box and Jenkins (1970) for models subject to a random step function input.

In this study, a generalized approach is attempted to establish the relationship between the stochastic differential equation model and the ARMA process. These relationships for some selected stochastic models are used to identify the underlying generating process of the observed monthly runoff series. The physical interpretation of the proposed stochastic model is then obtained through the correlation between the model parameters and the watershed physiographic characteristics.

Theoretical Developments

Relationship between stochastic differential and difference formulations

The relationship between the stochastic differential equation model and the stochastic difference equation, or ARMA type model can be obtained through proper mathematical treatment of their respective solutions in terms of impulse response functions.

For a linear stochastic differential equation model,

$$(D^M+a_1D^{M-1}+ \dots +a_{M-1}D+a_M)y(t) = x(t), \tag{3.1}$$

the impulse response function h(u) can be obtained as the solution to the linear differential equation

$$(D^M+a_1D^{M-1}+ \dots +a_{M-1}D+a_M)h(u) = 0 \tag{3.2a}$$

or

$$(D+\mu_1)(D+\mu_2) \dots (D+\mu_M)h(u) = 0 \tag{3.2b}$$

This leads to the following solution if all roots are distinct,

$$h(u) = c_1 e^{-\mu_1 u} + c_2 e^{-\mu_2 u} + \ldots + c_M e^{-\mu_M u}, \qquad u \geq 0. \tag{3.3}$$

where $-\mu$'s must have negative real parts for a stable system.

On the other hand, for a linear stochastic difference equation of ARMA type

$$(1-d_1B-d_2B^2 - \ldots - d_mB^m)y_n = (w_0-w_1B-w_2B^2 - \ldots - w_sB^s)x_n, \tag{3.4}$$

the impulse response function h_r satisfies the following relationship:

$$(1-d_1B-d_2B^2 - \ldots - d_mB^m)h_r = 0, \qquad r > s, \tag{3.5a}$$

or

$$(1-S_1B)(1-S_2B) \ldots (1-S_mB)h_r = 0, \qquad r > s. \tag{3.5b}$$

The solution to the above difference equation becomes

$$h_r = c_1'S_1^r + c_2'S_2^r + \ldots + c_m'S_m^r, \tag{3.6}$$

for 'm' distinct roots which must lie outside the unit circle for a stable system.

Taking into consideration the analogy between differential and difference equations and comparing equations (3.3) and (3.6), it is natural to assume that

$$M = m \quad \text{and} \quad S_i = e^{-\mu_i}, \quad i = 1,2, \ldots,m. \tag{3.7}$$

Then the transformation of a linear stochastic differential equation model into its equivalent discrete form can be accomplished in the following manner.

From the relationship (3.7) and the linear system formulation for continuous time

$$y(t) = \int_{-\infty}^{t} h(t-u)x(u)du, \tag{3.8}$$

the discrete time equivalent is obtained by the following linear operation

$$d(B)y(t) = d(B) \int_{-\infty}^{t} h(t-u)x(u)du, \tag{3.9a}$$

or $$\prod_{i=1}^{m} (1-S_i B)y(t) = \prod_{i=1}^{m} (1-S_i B) \int_{-\infty}^{t} h(t-u)x(u)du, \tag{3.9b}$$

where $d(B) = (1-d_1B-d_2B^2 - \ldots - d_mB^m)$ and $\prod_{i=1}^{m} (1-S_iB) = (1-S_1B)(1-S_2B)$ $\ldots (1-S_mB)$; or equivalently

$$\prod_{i=1}^{m} (1-e^{-\mu_i}B)y(t) = \prod_{i=1}^{m} (1-e^{-\mu_i}B) \int_{-\infty}^{t} h(t-u)x(u)du. \tag{3.9c}$$

By expanding the right-hand side of eq. (3.9a), it becomes

$$\begin{aligned}
& d(B) \int_{-\infty}^{t} h(t-u)x(u)du \\
&= (1-d_1B-d_2B^2 - \ldots - d_mB^m) \int_{-\infty}^{t} h(t-u)x(u)du \\
&= \int_{-\infty}^{t} h(t-u)x(u)du-d_1 \int_{-\infty}^{t-1} h(t-1-u)x(u)du-d_2 \int_{-\infty}^{t-2} h(t-2-u)x(u)du \\
&\qquad - \ldots - d_m \int_{-\infty}^{t-m} h(t-m-u)x(u)du \\
&= \int_{t-m}^{t} h(t-u)x(u)du + \int_{-\infty}^{t-m} h(t-u)x(u)du \\
&\qquad - d_1 \int_{t-m}^{t-1} h(t-1-u)x(u)du-d_1 \int_{-\infty}^{t-m} h(t-1-u)x(u)du \\
&\qquad - \ldots - d_m \int_{-\infty}^{t-m} h(t-m-u)x(u)du \\
&= \int_{t-m}^{t} h(t-u)x(u)du-d_1 \int_{t-m}^{t-1} h(t-1-u)x(u)du - \ldots \\
&\qquad + \int_{-\infty}^{t-m} d(B)h(t-u)x(u)du \\
&= \int_{t-m}^{t} h(t-u)x(u)du-d_1 \int_{t-m}^{t-1} h(t-1-u)x(u)du
\end{aligned}$$

$$- \dots - d_{m-1} \int_{t-m}^{t-m+1} h(t-m+1-u)x(u)du$$

$$= \int_{t-1}^{t} H_0(h,t-u)x(u)du + \int_{t-2}^{t-1} H_1(h,t-u)x(u)du + \dots +$$

$$+ \int_{t-m}^{t-m+1} H_{m-1}(h,t-u)x(u)du \qquad (3.10)$$

where $H_0, H_1, \dots, H_{m-1}$ are functions of h and t-u, to be determined by grouping the convolution integrals at time increments (t-m,t-m+1), (t-m+1,t-m+2),...,(t-1,t).

Thus it can be seen that the equivalent discrete form of a m^{th} order linear differential equation model is indeed a mixed auto-regressive-moving average (ARMA) type model of the order (m,m-1):

$$(1-d_1B-d_2B^2-\dots-d_mB^m)y_n = (w_0-w_1B-w_2B^2-\dots-w_{m-1}B^{m-1})x_n. \qquad (3.11)$$

It can also be observed, however, from equations (3.10) and (3.11) that the exact equivalence between $\int_{t-r-1}^{t-r} H_r(h,t-u)x(u)du$ and $w_r x_{n-r}$ depends on the nature of the input function X(u), which may be a random impulse function, step function or any other known processes.

Given an output time series and assuming the input process to be a random impulse function, the model parameters, or the coefficients of the linear stochastic differential equation may be determined from the autocorrelation functions of the given time series:

$$\rho_y(u) = \frac{\int_0^\infty h(v)h(v-u)dv}{\int_0^\infty h^2(v)dv} \qquad (3.12)$$

$$= f(\mu_1, \mu_2, \dots, \mu_m;\ u)$$

Analytical solution to the above equation is difficult to obtain. It can be solved,hhowever, by the method of iterative interpolation.

First and second-order stochastic differential equations

General formulation of the analytical relationship between the stochastic differential equation model and its equivalent difference equation model is very attractive. For most practical purposes of hydrologic time series analysis, however, it is sufficient to consider only some special cases, namely, first and second-order stochastic differential equations.

a) For a first-order stochastic differential equation, it is given by

$$(1+KD)y(t) = x(t) \tag{3.13}$$

which takes the form resembling Nash's conceptual linear reservoir model (Nash, 1960) subject to a random impulse function, or excess rainfall in Nash's case.

The impulse response function can be shown to be

$$h(u) = \frac{1}{K}e^{-u/K}, \qquad K > 0 \tag{3.14}$$

The linear system representation by this first-order process then becomes

$$y(t) = \frac{1}{K}\int_{-\infty}^{t} e^{-(t-u)/K}x(u)du. \tag{3.15}$$

By applying the linear operation (3.9), eq. (3.15) becomes

$$(1-e^{-1/K}B)y(t) = \frac{1}{K}\int_{t-1}^{t} e^{-(t-u)/K}x(u)du. \tag{3.16}$$

Assuming zero means for both input and output processes, the relationship between σ_y^2 and σ_x^2 in eq. (3.13) can be shown to be obtained by

$$\sigma_y^2 = \frac{\sigma_x^2}{2K} . \tag{3.17}$$

The autocorrelation function of Y(t) is

$$\rho_y(u) = e^{-u/K}. \tag{3.18}$$

If eq. (3.16) is rewritten in the form of

$$(1-d_1B)y_n = w_0x_n, \quad \text{where } d_1 = e^{-1/K} \text{ and } |d_1| < 1; \tag{3.19}$$

it can be shown that $\frac{1}{K}\int_{t-1}^{t} e^{-(t-u)/K}x(u)du$ may be approximated in the discrete form by w_0x_n, where

$$w_0 = \sqrt{(1-e^{-2/K})/2K}, \tag{3.20}$$

and X_n is an independent random process with zero mean and variance equal to σ_x^2. This approximation preserves the mean and variance of the increment $\frac{1}{K}\int_{t-1}^{t} e^{-(t-u)/K}x(u)du$. The proof of eq. (3.20) is given in Appendix B.

Alternatively, w_0 may be computed from

$$w_0 = \sqrt{\frac{\sigma_z^2}{\sigma_x^2}} , \tag{3.21}$$

where $z_n = w_0x_n$ is the random component for the ARMA(1,0) model, or first-order autoregressive model (3.19).

b) For a second-order stochastic differential equation, it is defined by

$$(D^2+a_1D+a_2)y(t) = x(t), \tag{3.22}$$

where X(t) is assumed to be a random impulse function.

The impulse response function becomes

$$h(u) = \frac{1}{\mu_2-\mu_1}\,(e^{-\mu_1 u} - e^{-\mu_2 u}), \tag{3.23}$$

where $-\mu_1$ and $-\mu_2$ are the distinct roots of the characteristic equation $D^2+a_1D+a_2 = 0$ and must have negative real parts for a stable system. When the roots are real, the system response is a double exponential decay function and is the familiar 'overdamped' motion in the terminology of mechanical or electrical circuits; whereas for the complex roots, a damped sine wave called 'underdamped' motion results.

Equation (3.23) leads to the following system representation,

$$y(t) = \int_{-\infty}^{t} \frac{e^{-\mu_1(t-u)} - e^{-\mu_2(t-u)}}{\mu_2-\mu_1}\, x(u)du. \tag{3.24}$$

The discrete time equivalent of eq. (3.22) is then obtained through the linear operation (3.9)

$$(1-(e^{-\mu_1}+e^{-\mu_2})B+e^{-(\mu_1+\mu_2)}B^2)y(t) = \int_{t-1}^{t} \frac{e^{-\mu_1(t-u)} - e^{-\mu_2(t-u)}}{\mu_2-\mu_1}\, x(u)du$$

$$-e^{-(\mu_1+\mu_2)}\int_{t-2}^{t-1} \frac{e^{-\mu_1(t-2-u)} - e^{-\mu_2(t-2-u)}}{\mu_2-\mu_1}\, x(u)du. \tag{3.25}$$

Assuming zero means for both input and output processes, the relationship between σ_y^2 and σ_x^2 can be shown to be

$$\sigma_y^2 = \frac{\sigma_x^2}{2a_1a_2}\ . \tag{3.26}$$

The autocorrelation function of Y(t) is

$$\rho_y(u) = \frac{\mu_2 e^{-\mu_1 u} - \mu_1 e^{-\mu_2 u}}{\mu_2 - \mu_1} . \tag{3.27}$$

Rewriting eq. (3.25) in the form of

$$(1-d_1B-d_2B^2)y_n = (w_0-w_1B)x_n , \tag{3.28}$$

where $d_1 = e^{-\mu_1}+e^{-\mu_2}$, $d_2 = -e^{-(\mu_1+\mu_2)}$ and $|d_2| < 1$, $d_2+d_1 < 1$, $d_2-d_1 < 1$; it can be shown that the random increments at the right-hand side of eq. (3.25) may be approximated in the discrete form by w_0x_n and w_1x_{n-1}, where

$$w_0 = \frac{1}{\mu_2-\mu_1} \sqrt{\frac{1}{2\mu_1}(1-e^{-2\mu_1}) - \frac{2}{\mu_1+\mu_2}(1-e^{-(\mu_1+\mu_2)}) + \frac{1}{2\mu_2}(1-e^{-2\mu_2})}, \tag{3.29a}$$

and $$w_1 = \frac{e^{-(\mu_1+\mu_2)}}{\mu_2-\mu_1} \sqrt{\frac{1}{2\mu_1}(e^{2\mu_1}-1) - \frac{2}{\mu_1+\mu_2}(e^{\mu_1+\mu_2}-1) + \frac{1}{2\mu_2}(e^{2\mu_2}-1)}; \tag{3.29b}$$

and X_n is an independent random process with zero mean and variance equal to σ_x^2. Derivation of eq. (3.29) is presented in Appendix B.

Alternatively, w_0 and w_1 may be obtained by

$$w_0 = \sqrt{\frac{\sigma_z^2}{\sigma_x^2}} , \tag{3.30a}$$

and

$$w_1 = c_1\sqrt{\frac{\sigma_z^2}{\sigma_x^2}} , \tag{3.30b}$$

from the fitted ARMA(2,1) process

$$(1-d_1B-d_2B^2)y_n = (1-c_1B)z_n, \tag{3.28a}$$

where $z_n = w_0x_n$ and $c_1z_{n-1} = w_1x_{n-1}$.

c) For a second-order stochastic differential equation with equal roots, it can be written as

$$(D^2+2\mu D+\mu^2)y(t) = x(t) \tag{3.31}$$

which is a special case of model (3.24) and is analogous to Nash's conceptual double linear reservoir model; X(t) is assumed to be a random impulse function.

The impulse response function in this case is

$$h(u) = ue^{-\mu u}, \tag{3.32}$$

where $-\mu$ is the double root of the characteristic equation $D^2+2\mu D+\mu^2=0$ and must be a negative real number for a stable system. This results in a 'critically damped' motion.

The system is then represented by

$$y(t) = \int_{-\infty}^{t} (t-u)e^{-\mu(t-u)}x(u)du. \tag{3.33}$$

Through the linear operation (3.9), the discrete time equivalent of eq. (3.31) then becomes

$$(1-2e^{-\mu}B+e^{-2\mu}B^2)y(t)$$
$$= \int_{t-1}^{t} (t-u)e^{-\mu(t-u)}x(u)du-e^{-2\mu}\int_{t-2}^{t-1} (t-2-u)e^{-\mu(t-2-u)}x(u)du. \tag{3.34}$$

The variance σ_y^2 can be shown to be

$$\sigma_y^2 = \frac{\sigma_x^2}{4\mu^3}, \tag{3.35}$$

where Y(t) is assumed to have zero mean.

The autocorrelation function of Y(t) can be obtained as follows,

$$\rho_y(u) = e^{-\mu u}(1+\mu u). \tag{3.36}$$

For the alternative form of eq. (3.34)

$$(1-d_1B-d_2B^2)y_n = (w_0-w_1B)x_n, \tag{3.37}$$

where $d_1 = 2e^{-\mu}$, $d_2 = -e^{-2\mu}$ and $|d_2| < 1$, $d_2 + d_1 < 1$, $d_2-d_1 < 1$; it can be shown that the discrete time equivalent of the random increment at the right-hand side of eq. (3.34) may be estimated in the weak statistical sense by w_0x_n and w_1x_{n-1}, where

$$w_0 = \sqrt{\frac{-1}{2\mu}e^{-2\mu} - \frac{1}{2\mu^2}e^{-2\mu} - \frac{1}{4\mu^3}e^{-2\mu} + \frac{1}{4\mu^3}}, \tag{3.38a}$$

and

$$w_1 = e^{-2\mu}\sqrt{\frac{1}{2\mu}e^{2\mu} - \frac{1}{2\mu^2}e^{2\mu} - \frac{1}{4\mu^3} + \frac{1}{4\mu^3}e^{2\mu}}; \tag{3.38b}$$

and X_n is an independent random process with zero mean and variance equal to σ_x^2. The proof of eq. (3.4) can be found in Appendix B.

Again, w_0 and w_1 may be alternatively estimated by

$$w_0 = \sqrt{\frac{\sigma_z^2}{\sigma_x^2}}, \tag{3.39a}$$

and $$w_1 = c_1 \sqrt{\frac{\sigma_z^2}{\sigma_x^2}} \, ; \tag{3.39b}$$

from the fitted ARMA (2,1) process

$$(1-d_1B-d_2B^2)y_n = (1-c_1B)z_n, \tag{3.39b}$$

where $z_n = w_0x_n$ and $c_1z_{n-1} = w_1x_{n-1}$. (3.37a)

Computational Procedures

Data

Monthly runoff* records were obtained from the so-called 'hydrological data bank' resulting from the hydrometric network planning study for western and northern Canada by the Department of the Environment (Shawinigan, 1970). Forty-two stations were selected from the data bank's monthly runoff data file for the period 1952-1966 which was also the study period for the above study. These stations were among those classified as 'natural regime' stations where man-made influences to the flow regime were kept to the minimum. The locations of these stations are shown in Table 3.1 and Figure 3.1. They are

---

*Runoff is defined here as the depth of the total flow volume uniformly distributed over the entire drainage area. This definition is used to reduce the noise variance introduced by the wide variation of the drainage area measurements, which accounts for a large portion of the flow variation and sometimes undermines the intrinsic statistical properties of the flow.

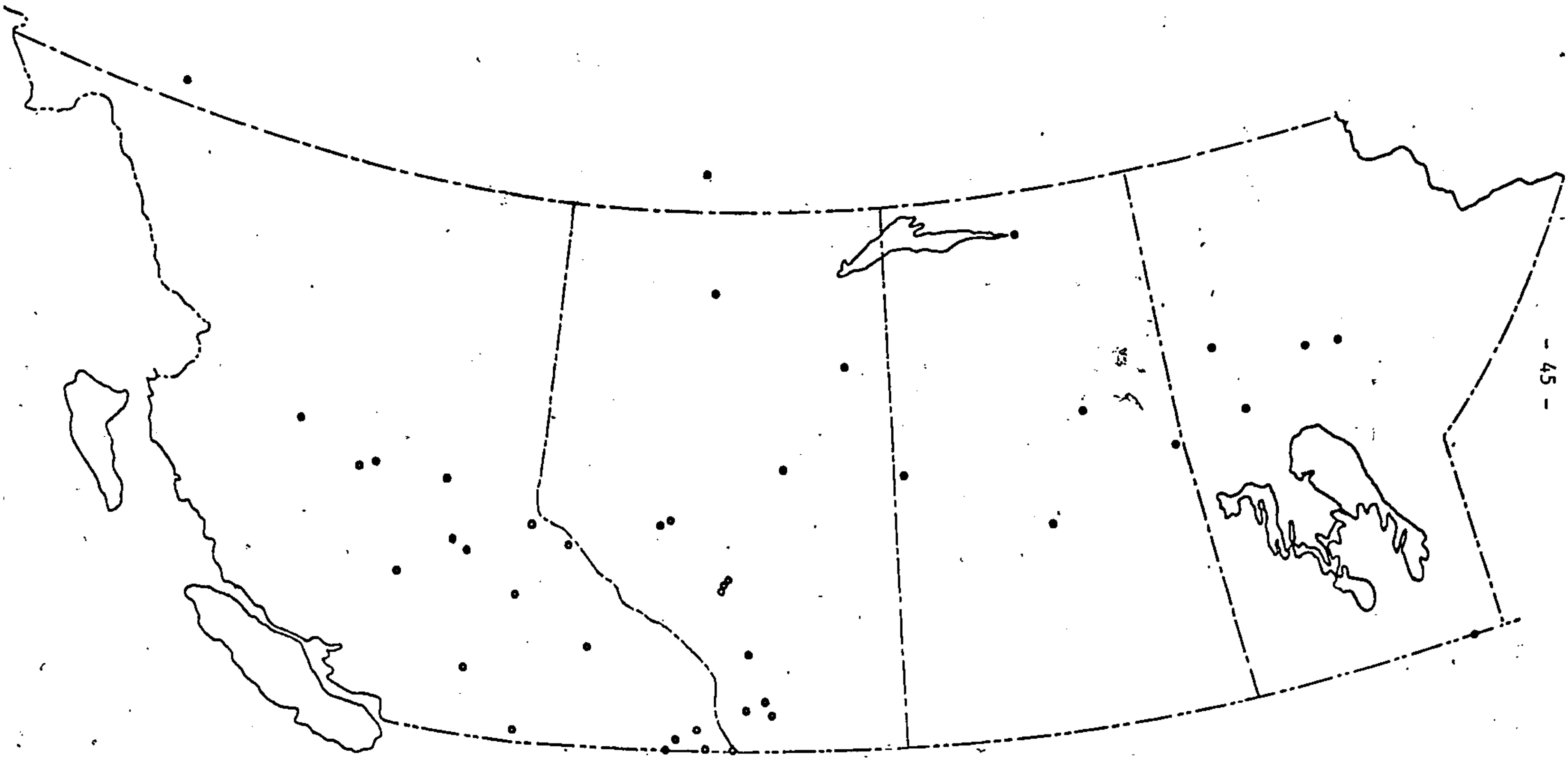


Figure 3.1 Stations Involved in Regional Regression Analysis

TABLE 3.1

Stations Involved in Regional Regression Analysis

Station ID	Latitude	Longitude	Name
05AA023	49 48 50	114 11 00	Oldman River near Waldron's Corner
05AB002	49 47 40	113 31 30	Willow Creek near Nolan
05AB021	50 01 05	113 42 50	Willow Creek near Claresholm
05BJ005	51 00 00	114 06 00	Elbow River above Glenmore Dam
05DB001	52 20 40	114 56 10	Clearwater River near Rocky Mountain House
05DB002	52 16 20	114 55 50	Prairie Creek near Rocky Mountain House
05DC001	52 22 51	114 56 21	North Saskatchewan River near Rocky Mountain House
05GG001	53 12 10	105 46 06	North Saskatchewan River near Prince Albert
05KG002	54 26 20	102 10 30	Sturgeon-Weir River at Relican Narrows
05OD032	48 59 35	95 55 04	Pine Creek near Pine Creek
05TB002	54 47 20	99 58 10	Grass River at Wekusko Falls
05TD001	55 44 35	97 00 00	Grass River above Standing Stone Falls
05TG001	55 44 30	97 54 00	Burntwood River near Thompson
06AF001	54 34 00	109 50 10	Cold River at Outlet of Cold Lake
06CB002	55 19 43	104 32 38	Rapid River at Outlet of Lac La Ronge
06EA006	56 09 00	100 27 00	Churchill River above Granville Falls
07AF002	53 28 10	116 37 45	McLeod River above Embarras River
07AG001	53 39 15	116 16 50	McLeod River near Wolf Creek
07BE001	54 43 20	113 17 10	Athabaska River at Athabaska
07DA001	56 46 53	111 24 09	Athabaska River below McMurray
07JD001	58 17 30	115 23 00	Wabaska River above Peace River

TABLE 3.1 (continued)

Station ID	Latitude	Longitude	Name
07LE001	59 16 00	105 48 00	Fond du Lac River at Stony Rapids
070B001	60 44 45	115 51 20	Hay River near Hay River
08EE004	54 37 05	126 53 55	Bulkley River at Quick
08JB002	54 00 31	125 00 18	Stellako River at Glenannan
08JB003	54 05 07	124 35 58	Nautley River near Fort Fraser
08KA005	53 18 45	120 08 40	Fraser River at McBride
08KA007	52 58 55	119 00 15	Fraser River at Red Pass
08KB001	54 00 40	125 37 00	Fraser River at Shelley
08KH003	52 40 00	121 40 20	Cariboo River below Kangaroo Creek
08KH006	52 51 13	122 14 07	Quesnel River near Quesnel
08LA008	51 55 55	120 14 30	Mahood River at Outlet of Mahood Lake
08LF051	52 21 14	121 23 33	Thompson River near Spences Bridge
08MA001	52 04 25	123 32 00	Chilko River near Redstone
08ND011	51 02 31	118 12 43	Columbia River above Steamboat Rapids
08NG005	49 25 13	115 25 10	Kootney River at Wardner
08NG042	49 00 52	116 10 45	Kootney River at Newgate
08NH006	48 59 55	116 10 45	Moyie River at Eastport
08NH034	49 15 30	115 51 00	Moyie River at Moyie
08NL004	49 11 30	120 00 00	Ashnola River near Keremeos
08NP001	49 00 08	114 28 31	Flathead River at Flathead
09AE001	60 29 07	133 18 04	Teslin River near Teslin

evenly distributed in British Columbia and across the Prairie Provinces, below the 60°N latitude. The drainage areas of these stations are ranging from 76 sq. mi. to 82,000 sq. mi. The average watershed physiographic characteristics were also obtained from the data bank's basin file and are given in Table 3.2, definition of which is given in Appendix A.

Approximation of second-order stationarity

Considering the entire length of the geophysical history, fifteen-years is a rather short time span within which no significant trend, geological, meteorological or hydrological, would be realized. Most of such geophysical time series could be assumed to be stationary or pseudo-stationary containing periodic components. Adamowski (1969) has shown that fifteen years of data are adequate for daily streamflow time series analysis and no significant trend was found in any of the time series under study. This same length of period was also adopted as the minimum acceptable record length for stationary monthly hydrologic time series analysis in various hydrometric network planning studies in Canada (Shawinigan, 1969, 1970). Therefore, trend analysis was not attempted for the monthly runoff series.

Several studies (Yevjevich, 1964; Rodriguez-Iturbe and Yevjevich, 1967) have shown that there is no statistical evidence that any cyclic (or deterministic) movement exists in river flow or precipitation time series except the astronomic cycle of the year. Therefore the following additive process was assumed for the monthly runoff series:

TABLE 3.2

Summary of Watershed Average Physiographic Characteristics of

Stations Involved in Regional Regression Analysis

Station Identification			05AA023	05AB002	05AB021	05BJ005	05DB001	05DB002
Drainage Area		sq.mi.	551	900	446	471	1210	318
Latitude Index			24	25	27	34	47	49
Average Elevation		ft	5830	4220	4600	5330	5160	4070
Average Slope		%	3.3	1.8	2.1	3.2	2.0	1.3
Slope Azimuth Index	N		3	3	4	4	4	4
	NW		2	2	3	3	3	3
	W		1	1	2	2	2	2
	SW		2	2	1	1	1	1
Distance to Sea	N	km	2090	2080	2060	1990	1870	1850
	NW		3365	3365	3351	3238	3068	3054
	W		970	1030	1040	990	1080	1150
	SW		1088	1159	1131	989	876	890
Area Lake and Swamp		%	0	2	3	3	4	3
Forest			65	32	48	51	76	95
Barrier Height	N	ft	660	330	280	350	170	60
	NW		3040	4920	4520	3430	2660	3440
	W		2260	4000	3770	3240	3880	4950
	SW		1270	2730	2550	2820	3790	4860
Shield Effect	N	ft	12800	11500	11400	13400	15600	12800
	NW		160100	157200	156800	145600	117400	106400
	W		69900	71200	71500	77300	52200	52000
	SW		111300	118500	112500	119700	106900	106200
Regional Slope	NE	ft/km	12	12	16	19	14	6
	E		16	15	21	30	21	15
	SE		7	10	14	17	15	12

TABLE 3.2 (continued)

Summary of Watershed Average Physiographic Characteristics of Stations Involved in Regional Regression Analysis

Station Identification			05DC001	05GG001	05KG002	05OD032	05TB002	05TD001
Drainage Area		sq.mi.	4220	46100	5540	76	1250	6030
Latitude Index			48	57	78	14	77	80
Average Elevation		ft	5720	2760	1260	1160	990	850
Average Slope		%	3.9	0.9	0.4	0.2	0.1	0.2
Slope Azimuth Index	N		4	4	3	1	2	3
	NW		3	3	2	2	1	2
	W		2	2	1	3	2	1
	SW		1	1	2	4	3	2
Distance to Sea	N	km	1870	1700	1470	2160	1470	1470
	NW		3054	2870	2333	3181	2375	2361
	W		1070	1440	1950	2200	2130	2220
	SW		820	1315	2248	2545	2488	2601
Area Lake and Swamp		%	3	6	22	9	33	30
Forest			66	34	77	69	65	69
Barrier Height	N	ft	730	340	360	30	430	390
	NW		2370	4330	780	1380	970	1050
	W		3290	5470	5110	6350	5610	5400
	SW		3010	5020	4060	8850	5340	6180
Shield Effect	N	ft	18400	11200	5400	5500	5800	4600
	NW		125500	75200	9500	17300	9800	9800
	W		51000	47700	41800	56100	41600	42900
	SW		96100	121200	193900	113200	188000	183800
Regional Slope	NE	ft/km	10	2	1	1	1	0
	E		16	4	2	2	2	1
	SE		10	2	1	1	1	0

TABLE 3.2 (continued)

Summary of Watershed Average Physiographic Characteristics of Stations Involved in Regional Regression Analysis

Station Identification			05TG001	06AF001	06CB002	06EA006	07AF002	07AG001
Drainage Area		sq.mi.	6310	2160	6080	82000	1000	2510
Latitude Index			85	79	76	83	61	62
Average Elevation		ft	910	2060	1560	1640	3700	3900
Average Slope		%	0.2	0.3	0.4	0.4	0.9	1.3
Slope Azimuth Index	N		3	2	4	4	4	4
	NW		2	1	3	3	3	3
	W		1	2	2	2	2	2
	SW		2	3	1	1	1	1
Distance to Sea	N	km	1420	1440	1580	1440	1750	1750
	NW		2304	2347	2290	2262	2898	2884
	W		2170	1540	1820	1600	1120	1110
	SW		2615	1456	2022	1753	777	777
Area Lake and Swamp		%	18	29	30	22	34	26
Forest			82	70	70	75	65	69
Barrier Height	N	ft	330	210	390	310	1000	910
	NW		930	3840	1330	2270	3840	3530
	W		5000	4160	5040	4910	4340	4020
	SW		5530	6770	4850	6120	4290	4060
Shield Effect	N	ft	4400	9200	6500	7500	16500	16800
	NW		9300	36500	15300	23800	102500	103600
	W		44800	41000	40200	42200	45400	44300
	SW		190000	117800	152800	128200	38400	39200
Regional Slope	NE	ft/km	0	1	0	0	7	9
	E		1	2	2	1	12	13
	SE		0	1	1	0	8	7

TABLE 3.2 (continued)

Summary of Watershed Average Physiographic Characteristics of Stations Involved in Regional Regression Analysis

Station Identification			07BE001	07DA001	07JD001	07LE001	07OB001	08EE004
Drainage Area		sq.mi.	29600	50000	14100	25800	18500	2800
Latitude Index			68	78	98	118	122	72
Average Elevation		ft	3670	2950	1990	1460	1620	3520
Average Slope		%	2.1	1.4	0.6	0.4	0.7	4.5
Slope Azimuth Index	N		4	4	5	5	3	1
	NW		3	3	4	4	2	2
	W		2	2	3	3	1	3
	SW		1	1	2	2	2	4
Distance to Sea	N	km	1680	1530	1350	1120	1180	1730
	NW		2799	2545	2177	1824	1852	1908
	W		1130	1230	1140	1960	1240	370
	SW		876	1032	1145	1838	1060	339
Area Lake and Swamp		%	15	20	26	15	38	4
Forest			67	68	73	84	60	82
Barrier Height	N	ft	770	630	910	180	820	3970
	NW		3620	3490	4920	400	4930	3480
	W		3660	4150	5100	6170	6190	2260
	SW		4550	5400	5870	7290	4520	1910
Shield Effect	N	ft	14500	12100	8700	4800	9000	64000
	NW		94300	70400	52200	8300	47000	140400
	W		42400	43300	44800	55000	54500	16800
	SW		46000	49100	35900	92000	26100	11300
Regional Slope	NE	ft/km	3	2	1	0	1	0
	E		6	4	2	1	2	2
	SE		3	2	1	0	1	2

TABLE 3.2 (continued)

Summary of Watershed Average Physiographic Characteristics of

Stations Involved in Regional Regression Analysis

Station Identification			08JB002	08JB003	08KA005	08KA007	08KB001	08KH003
Drainage Area		sq.mi.	1500	2430	2690	615	12500	1310
Latitude Index			69	70	58	57	66	57
Average Elevation		ft	3210	3150	5690	6320	4200	4770
Average Slope		%	2.2	2.3	7.7	6.8	4.7	4.6
Slope Azimuth Index	N		5	4	3	2	1	3
	NW		4	3	2	1	2	4
	W		3	2	1	2	3	5
	SW		2	1	2	3	4	4
Distance to Sea	N	km	1740	1730	1840	1850	1780	1860
	NW		2276	2380	2884	2912	2757	2842
	W		490	500	930	990	800	820
	SW		296	325	636	664	608	551
Area Lake and Swamp		%	10	9	1	1	2	2
Forest			88	90	51	49	76	72
Barrier Height	N	ft	4140	4180	1870	1590	2330	2000
	NW		4010	4040	1740	1650	2760	2030
	W		2930	2860	1850	1530	2180	1520
	SW		2420	2510	2190	1410	3130	3230
Shield Effect	N	ft	59600	58400	25600	24300	26000	29100
	NW		163800	169700	137700	138600	142500	161300
	W		19700	19400	35800	41200	27900	27200
	SW		11700	12000	31900	36000	24500	24700
Regional Slope	NE	ft/km	6	4	-2	-4	-2	-4
	E		12	9	-2	-22	-4	-9
	SE		7	5	0	-7	-2	-2

TABLE 3.2 (continued)

Summary of Watershed Average Physiographic Characteristics of

Stations Involved in Regional Regression Analysis

Station Identification			08KH006	08LA008	08LF051	08MA001	08ND011	08NG005
Drainage Area		sq.mi.	4690	1780	21600	3230	10300	5200
Latitude Index			54	45	37	40	40	26
Average Elevation		ft	4220	3890	4040	5400	5760	5650
Average Slope		%	4.7	3.2	5.7	4.8	8.5	6.3
Slope Azimuth Index	N		4	3	3	5	2	1
	NW		5	4	4	4	3	2
	W		4	5	5	3	4	3
	SW		3	4	4	2	5	4
Distance to Sea	N	km	1900	1980	2050	2090	1980	2090
	NW		2884	2983	2926	890	3124	3322
	W		780	620	640	340	820	870
	SW		523	466	466	311	692	890
Area Lake and Swamp		%	5	6	3	5	2	1
Forest			78	94	85	59	66	81
Barrier Height	N	ft	2490	3340	3510	1980	3020	3330
	NW		2730	3230	3210	1520	2350	2810
	W		2370	3890	3690	2480	2430	2450
	SW		3690	3640	2560	2190	1720	1770
Shield Effect	N	ft	30900	36900	40100	46100	31200	30100
	NW		168700	182000	175700	48000	157300	174600
	W		26300	26700	38700	24200	52300	63600
	SW		23700	22700	30300	18100	73400	112000
Regional Slope	NW	ft/km	-4	0	-2	1	0	-2
	E		-7	-2	-4	0	0	0
	SE		-7	-1	-2	1	0	-1

TABLE 3.2 (continued)

Summary of Watershed Average Physiographic Characteristics of Stations Involved in Regional Regression Analysis

Station Identification			08NG042	08NH006	08NH034	08NL004	08NP001	09AE001
Drainage Area		sq.mi.	7660	570	280	400	450	11700
Latitude Index			24	16	18	14	16	138
Average Elevation		ft	5380	4380	5470	6517	5550	3920
Average Slope		%	6.1	4.7	6.1	4.5	4.1	3.4
Slope Azimuth Index	N		1	5	4	5	2	3
	NW		2	4	3	4	3	4
	W		3	3	2	3	4	5
	SW		4	2	1	2	5	4
Distance to Sea	N	km	2090	1960	2180	1977	2170	1080
	NW		3308	3040	3393	995	3450	1753
	W		850	670	780	370	850	1220
	SW		904	820	904	339	1102	339
Area Lake and Swamp		%	1	1	1	0	0	5
Forest			83	98	98	66	92	69
Barrier Height	N	ft	3430	3770	3630	1322	1742	2520
	NW		2970	2440	2350	1579	3160	2250
	W		2490	2000	1670	735	1580	4930
	SW		1680	720	790	669	590	2170
Shield Effect	N	ft	29600	32800	35300	43960	19200	31200
	NW		175900	180200	198400	51320	184900	102400
	W		60400	39700	47600	18080	49400	81000
	SW		110800	94500	107100	67660	128200	17100
Regional Slope	NE	ft/km	-1	9	24	32	-12	0
	E		0	3	19	51	-20	-1
	SE		-1	2	12	28	-14	0

$$q(t) = m(t) + q'(t), \tag{3.40}$$

where q(t) is the observed runoff, m(t) is the long term monthly mean of the runoff and q'(t) is the stochastic component of the runoff. Sometimes, however, seasonal variation might still be found among monthly variances. In order to achieve the second-order stationarity for the runoff series, the transformation suggested by Roesner and Yevjevich (1966) was then applied:

$$y(t) = \frac{q'(t)}{s(t)} = \frac{q(t)-m(t)}{s(t)} \tag{3.41}$$

where s(t) is the long term monthly standard deviation; y(t) is the transformed stationary runoff component to the second order and is to be modelled by an autoregressive-moving average process.

Estimation of autocorrelations and autospectrums

The autocovariance and autocorrelation functions were estimated respectively from eqs. (2.8) and (2.10),

$$c_{yy}(k) = \frac{1}{N}\sum_{n=1}^{N-k} y_n y_{n+k}, \tag{2.8}$$

and

$$r_{yy}(k) = \frac{c_{yy}(k)}{c_{yy}(0)} \ . \tag{2.10}$$

For k << N, the distribution of $r_{yy}(k)$ can be taken as Normal with mean zero and its variance is given by eq. (2.49)

$$\mathrm{Var}(r_k) \approx \frac{1}{N}, \tag{2.49}$$

assuming that Y(t) is an independent Normal series (Anderson, 1942).

The sample estimate of autospectrum was then obtained from eq. (2.54),

$$\overline{S}_{yy}(f) = c_{yy}(0) + 2\sum_{k=1}^{M} w(k)c_{yy}(k)\cos 2\pi fk, \quad -\frac{1}{2} \le f \le \frac{1}{2} \tag{2.54}$$

where w(k) is the Parzen lag window defined as follows:

$$w(0) = 1, \tag{3.42a}$$

$$w(k) = 1-6\left(\frac{k}{M}\right)^2 + 6\left(\frac{|k|}{M}\right)^3, \qquad |k| \le \frac{M}{2} \tag{3.42b}$$

$$w(k) = 2\left(1 - \frac{|k|}{M}\right)^3, \qquad \frac{M}{2} < |k| \le M \tag{3.42c}$$

$$w(k) = 0, \qquad |k| > M,\ M < N. \tag{3.42d}$$

It can be shown that the Parzen window leads to smaller variance of spectral estimate (Jenkins and Watts, 1968), i.e.,

$$\mathrm{Var}(\overline{S}_{yy}(f)) \approx 0.539\left(\frac{M}{N}\right)S^2_{yy}(f). \tag{3.43}$$

The statistic $v\overline{S}_{yy}(f)/S_{yy}(f)$ is approximately distributed as a χ^2_v with degree of freedom 'v' given by

$$v = \frac{2N}{\int_{-\infty}^{\infty} w^2(u)\,du} \approx 3.71\left(\frac{N}{M}\right) \quad \text{for Parzen's lag window (Jenkins and Watts, 1968).} \tag{3.44}$$

Hence the 100(1-a)% confidence interval for $S_{yy}(f)$ is

$$\frac{v\overline{S}_{yy}(f)}{x_v(1-(a/2))} \le S_{yy}(f) \le \frac{v\overline{S}_{yy}(f)}{x_v(a/2)}, \tag{3.45}$$

where $x_v(1-(a/2))$ and $x_v(a/2)$ are χ^2_v values at the 100(1-(a/2))% and 100(a/2)% levels respectively.

It has been suggested (Jenkins and Watts, 1968) that $M = \frac{N}{4}$ is sufficient for reliable estimates of spectrums. Therefore in this study, it is decided that

$$M = \frac{N}{4} = \frac{180}{4} = 45 \approx 48\text{(months)}.$$

Identification of monthly runoff generating process

Identification of a system involves selecting a proper model, estimating parameters of the selected model and evaluating the efficiency of the fitted model. If the fitted model is found inadequate to describe the behavior of the system, alternative models must be considered.

Since it is generally recognized that the solution to any hydrologic problem is not unique, the identification of the transformed stationary monthly runoff series was carried out by fitting the time series with some preselected models. These models were selected based on the observation that the correlogram as well as the spectrum of the time series indicated an autoregressive-moving average scheme and on the literature review that most hydrologic time series may be fitted by a first-order autoregressive process (Roesner and Yevjevich, 1966; Solomon, 1972a), a second-order autoregressive process (Quimpo and Yevjevich, 1967; Adamowski, 1969) and an ARMA(1,1) process (O'Connell, 1971). The final consideration would then be given to the one with the simplest possible formation yet statistically adequate to represent the given time series.

a) First-order stochastic differential equation (S.D.E.) and its equivalent first-order autoregressive or ARMA(1,0) process

The mathematical forms of the first-order S.D.E. and the ARMA(1,0) process are given by equations (3.13) and (3.19):

$$(1+KD)y(t) = x(t), \qquad K > 0 \tag{3.13}$$

$$(1-d_1B)y_n = w_0x_n, \qquad |d_1| < 1. \tag{3.19}$$

It can be shown that d_1 may be estimated from

$$d_1 = r_{yy}(1), \tag{3.46}$$

and K was obtained from eq. (3.16),

$$K = -1/\ln(d_1). \tag{3.47}$$

The approximate variance of the estimate K is given by

$$\mathrm{Var}(K) \approx \frac{2K^3}{N}, \tag{3.48}$$

where N is the number of observations (Bartlett, 1946).

The variance of disturbance or noise to the model (3.13) was estimated using eq. (3.17)

$$\sigma_x^2 = 2K\sigma_y^2. \tag{3.17a}$$

The coefficient w_0 was then obtained from eq. (3.21)

$$w_0 = \sqrt{\frac{\sigma_z^2}{\sigma_x^2}}, \qquad \text{where } z_n = w_0x_n. \tag{3.21}$$

The theoretical autocorrelation function can be shown to be

$$r_{yy}(u) = e^{-u/K}, \qquad \text{for continuous time;} \qquad (3.49a)$$

$$r_{yy}(k) = d_1^k \qquad \text{for discrete time.} \qquad (3.49b)$$

The theoretical variance spectrums for the above process are

$$S_{yy}(f) = \frac{\sigma_x^2}{1+(2\pi fK)^2}, \quad -\infty \leq f \leq \infty \quad \text{for continuous time} \qquad (3.50a)$$

$$S_{yy}(f) = \frac{w_0^2\sigma_x^2}{1+d_1^2-2d_1\cos 2\pi f}, \quad -\frac{1}{2} \leq f \leq \frac{1}{2} \quad \text{for discrete time.} \qquad (3.50b)$$

The portion of the variance of the transformed runoff series or rather the one-step prediction variance, which could not be explained by the proposed model, was computed by taking the expectation of the squared deviation:

$$E(y_n-d_1y_{n-1})^2, \quad n = 1,2,\ldots,N.$$

b) Second-order S.D.E. and its equivalent ARMA(2,1) process

The second-order S.D.E. and the ARMA(2,1) process are given respectively by equations (3.22) and (3.28):

$$(D^2+a_1D+a_2)y(t) = x(t) \qquad (3.22)$$

with the roots of the characteristic equation having negative real parts†

and $$(1-d_1B-d_2B^2)y_n = (w_0-w_1B)x_n, \qquad (3.28)$$

with the condition that $|d_2| < 1$, $d_2+d_1 < 1$ and $d_2-d_1 < 1$.

Coefficients d_1, d_2 were estimated by

$$d_1 = \frac{r_{yy}(1)r_{yy}(2)-r_{yy}(3)}{r_{yy}^2(1)-r_{yy}(2)}, \tag{3.51}$$

$$d_2 = \frac{r_{yy}(1)r_{yy}(3)-r_{yy}^2(2)}{r_{yy}^2(1)-r_{yy}(2)}. \tag{3.52}$$

The relationship between the roots $-\mu_1$, $-\mu_2$ of the characteristic equation $D^2+a_1D+a_2 = 0$ and coefficients d_1, d_2 is given by eq. (3.25),

$$d_1 = e^{-\mu_1}+e^{-\mu_2} = e^{\frac{1}{2}(-a_1+\sqrt{a_1^2-4a_2})}+e^{\frac{1}{2}(-a_1-\sqrt{a_1^2-4a_2})}, \tag{3.53}$$

$$d_2 = -e^{-(\mu_1+\mu_2)} = -e^{-a_1}. \tag{3.54}$$

The determination of a_1, a_2 and computation of theoretical autocorrelations and variance spectrums were then carried out depending on the nature of the roots of the characteristic equation $1-d_1B-d_2B^2 = 0$, or alternatively on the value of $d_1^2 + 4d_2$.

i) Complex roots.

If the roots of the characteristic equation $1-d_1B-d_2B^2 = 0$ are complex, i.e., $d_1^2 + 4d_2 < 0$, it indicates implicitly complex roots for the characteristic equation $D^2+a_1D+a_2 = 0$. Subsequently eq. (3.51) may be rewritten as

$$d_1 = 2e^{-\frac{a_1}{2}} \cos \frac{1}{2} \sqrt{4a_2-a_1^2}. \tag{3.53a}$$

From equations (3.54) and (3.53a), a_1 and a_2 can be computed which in turn determine the real and imaginary parts of the complex roots $-\mu_1$ and $-\mu_2$ of the equation $D^2+a_1D+a_2 = 0$.

In order to reproduce the computed autocorrelations of the first two lags, which are considered reliable sample estimates of their parent values, the above computed values of a_1 and a_2 were used as the initial estimates to solve the following non-linear implicit equations derived from eq. (3.27):

$$r_{yy}(1) = \frac{e^{b\cos\theta}\sin(\theta-b\sin\theta)}{\sin\theta}, \tag{3.55a}$$

$$r_{yy}(2) = \frac{e^{2b\cos\theta}\sin(\theta-2b\sin\theta)}{\sin\theta}, \tag{3.55b}$$

where $b=2\sqrt{a_2}$ and $\theta=\tan^{-1}(-\sqrt{4a_2-a_1^2}/a_1)$.

It is difficult to solve eq. (3.55) analytically, a modified Newton's method for non-linear systems was developed. This is an iterative method of interpolation and extrapolation, which takes into consideration the variation of the geometry of the solution plane to locate the roots and improve the rate of convergence.

The final values of a_1 and a_2 from the above interpolation method were then used to determine d_1 and d_2 using equations (3.51) and (3.52) respectively. The approximate variances of the estimated

a_1 and a_2 are given respectively by

$$\mathrm{Var}(a_1) \approx \frac{2a_1}{N}, \tag{3.56}$$

and

$$\mathrm{Var}(a_2) \approx \frac{2a_1a_2}{N} \qquad \text{(Bartlett, 1946)}. \tag{3.57}$$

The noise variance was estimated using eq. (3.26)

$$\sigma_x^2 = 2a_1a_2\sigma_y^2 \ . \tag{3.26a}$$

The discrete time model was then fitted by

$$(1-d_1B-d_2B)y_n = (1-c_1B)z_n \tag{3.28a}$$

and coefficients w_0 and w_1 were computed using eq. (3.30):

$$w_0 = \sqrt{\frac{\sigma_z^2}{\sigma_x^2}}\ , \tag{3.30a}$$

$$w_1 = c_1\sqrt{\frac{\sigma_z^2}{\sigma_x^2}}\ . \tag{3.30b}$$

The theoretical autocorrelation for continuous time is

$$r_{yy}(u) = \frac{e^{ub\cos\theta}\sin(\theta-ub\sin\theta)}{\sin\theta}\ . \tag{3.58a}$$

For discrete time, theoretical autocorrelations may also be expressed in terms of the roots of the characteristic equation $1-d_1B-d_2B^2 = 0$. It is, however, easier to compute the autocorrelations through the recursive relationship

$$r_{yy}(k)=d_1r_{yy}(k-1)+d_2r_{yy}(k-2), \quad k \geq 2 \tag{3.58b}$$

with $$r_{yy}(1) = \frac{(d_1w_0-w_1)(w_0-w_1d_1)+w_0w_1d_2^2}{(w_0^2+w_1^2)(1-d_2)-2w_0w_1d_1} . \qquad (3.58c)$$

The theoretical variance spectrums become

$$S_{yy}(f) = \frac{\sigma_x^2}{(a_2-4\pi^2f^2)^2+(2\pi fa_1)^2}, \qquad -\infty \le f \le \infty$$

for continuous time; (3.59a)

$$S_{yy}(f) = \frac{(w_0^2+w_1^2-2w_0w_1\cos 2\pi f)\sigma_x^2}{1+d_1^2+d_2^2-2d_1(1-d_1)\cos 2\pi f-2d_2\cos 4\pi f}, \qquad -\frac{1}{2} \le f \le \frac{1}{2}$$

for discrete time. (3.59b)

ii) Equal roots.

If the roots of the characteristic equation $1-d_1B-d_2B^2 = 0$ are equal, i.e., $d_1+4d_2 = 0$, it also implies equal roots for the characteristic equation $D^2+a_1D+a_2 = 0$, i.e., $-\mu_1 = -\mu_2 = -\mu$.

Simple relationships exist among parameters d_1, d_2, a_1 and a_2:

$$d_2 = (\frac{d_1}{2})^2 = d^2; \qquad (3.60)$$

$$a_2 = (\frac{a_1}{2})^2 = a^2; \qquad (3.61)$$

$$d = e^{-\mu} = e^{-a}. \qquad (3.62)$$

Initial value of 'a' was obtained from eq. (3.62) and then adjusted to fit the sample estimate of lag-one autocorrelation

$$r_{yy}(1) = e^{-\mu}(1+\mu), \qquad (3.63)$$

where $\mu = a$.

Equation (3.63) was solved numerically and the final value of 'a' was used to determine 'd' using eq. (3.62). The approximate variance of the estimates a_1 and a_2 were computed using equations (3.56) and (3.57).

The procedures to determine the parameters w_0 and w_1, autocorrelations and variance spectrums are the same as for the case of complex roots except that the theoretical autocorrelation function for continuous time is expressed by,

$$r_{yy}(u) = e^{-\mu u}(1+\mu u). \tag{3.64}$$

iii) Real roots.

If the roots of the characteristic equation $1-d_1B-d_2B^2 = 0$ are real and distinct, i.e., $d_1^2+4d_2>0$, real roots are expected for the characteristic equation $D^2+a_1D+a_2 = 0$.

Equation (3.53) may be rewritten as

$$d_1 = 2e^{-\frac{a_1}{2}} \cosh\frac{1}{2}\sqrt{a_1^2-4a_2}. \tag{3.53b}$$

From equations (3.54) and (3.53b), initial values of a_1 and a_2 were obtained. The following non-linear implicit equations were then solved numerically to arrive at the final values of a_1 and a_2:

$$r_{yy}(1) = \frac{\mu_2 e^{-\mu_1} - \mu_1 e^{-\mu_2}}{\mu_2-\mu_1}, \tag{3.65a}$$

$$r_{yy}(2) = \frac{\mu_2 e^{-2\mu_1} - \mu_1 e^{-2\mu_2}}{\mu_2-\mu_1}. \tag{3.65b}$$

The approximate variances of the estimates were again obtained from equations (3.56) and (3.57). Adjusted values of d_1 and d_2 were obtained using equations (3.58) and (3.54) respectively.

In the same manner, w_0 and w_1 were computed using eq. (3.30). Theoretical autocorrelations and spectrums were computed using equations (3.58) and (3.59), with the exception that the theoretical autocorrelation function for continuous time may be directly computed from eq. (3.27)

$$r_{yy}(u) = \frac{\mu_2 e^{-\mu_1 u} - \mu_1 e^{-\mu_2 u}}{\mu_2 - \mu_1} . \tag{3.27}$$

The portion of the unexplained variance due to the proposed second-order S.D.E. or the equivalent ARMA(2,1) model was then computed by taking the expectation of the squared deviation:

$$E(y_n - d_1 y_{n-1} - d_2 y_{n-2} + w_1 x_{n-1})^2, \quad n=1,2,\ldots,N.$$

c) <u>ARMA(1,1) process</u>

The ARMA(1,1) process is represented by

$$(1-d_1 B)y_n = (1-w_1 B)x_n, \tag{3.66}$$

with the condition that $|d_1| < 1$.

It can be shown that

$$d_1 = r_{yy}(2)/r_{yy}(1), \tag{3.67}$$

and w_1 may be obtained by solving the implicit equation

$$r_{yy}(1) = \frac{(1-d_1w_1)(d_1-w_1)}{1+w_1^2-2d_1w_1} . \qquad (3.68)$$

The approximate variances of the estimates d_1 and w_1 are given respectively by

$$\text{Var}(d_1) \approx \frac{(1-d_1^2)(1-d_1w_1)^2}{N(d_1-w_1)^2} , \qquad (3.69)$$

and

$$\text{Var}(w_1) \approx \frac{(1-w_1^2)(1-d_1w_1)^2}{N(d_1-w_1)^2} \quad \text{(Box and Jenkins,1970).} \qquad (3.70)$$

The theoretical autocorrelations and spectrums for the ARMA(1,1) process are given by

$$r_{yy}(k) = d_1r_{yy}(k-1), \qquad k \geq 2 \qquad (3.71)$$

with $r_{yy}(1)$ ad defined by eq. (3.68); and

$$S_{yy}(f) = \frac{(1+w_1^2-2w_1\cos 2\pi f)\sigma_x^2}{1+d_1^2-2d_1\cos 2\pi f} , \quad -\frac{1}{2} \leq f \leq \frac{1}{2} . \qquad (3.72)$$

The unexplained variance due to the proposed model was then defined by taking the expectation

$$E(y_n-d_1y_{n-1}+w_1x_{n-1})^2, \quad n=1,2,\ldots,N.$$

d) Second-order autoregressive or ARMA(2,0) process

The ARMA(2,0) process is defined by

$$(1-d_1B-d_2B)y_n = x_n, \qquad (3.73)$$

with the condition that $|d_2| < 1$, $d_2+d_1 < 1$ and $d_2-d_1 < 1$.

It can be shown that

$$d_1 = \frac{r_{yy}(1)-r_{yy}(1)r_{yy}(2)}{1-r_{yy}^2(1)}, \tag{3.74}$$

$$d_2 = \frac{r_{yy}(2)-r_{yy}^2(1)}{1-r_{yy}^2(1)}. \tag{3.75}$$

The approximate variances of the estimates d_1 and d_2 are given respectively by

$$\mathrm{Var}(d_1) \approx \frac{1-d_2^2}{N}, \tag{3.76}$$

and $$\mathrm{Var}(d_2) \approx \frac{1-d_2^2}{N} \quad \text{(Box and Jenkins, 1970)}. \tag{3.77}$$

The theoretical autocorrelations and spectrums of an ARMA(2,1) process are

$$r_{yy}(k) = d_1 r_{yy}(k-1)+d_2 r_{yy}(k-2), \qquad k \geq 1; \tag{3.78}$$

$$S_{yy}(f) = \frac{\sigma_x^2}{1+d_1^2+d_2^2-2d_1(1-d_1)\cos 2\pi f-2d_2\cos 4\pi f}, \quad -\frac{1}{2} \leq f \leq \frac{1}{2}. \tag{3.79}$$

The unexplained variance due to the proposed model was obtained by taking the expectation

$$E(y_n-d_1y_{n-1}-d_2y_{n-2})^2, \qquad n=1,2,\ldots,N.$$

Evaluation of proposed models

Validation of the above proposed models was not carried out because of time and budgetary limitations. The efficiency of the models, however, was evaluated by following procedures:

i) Comparison between estimated and theoretical autocorrelation functions and variance spectrums of the transformed runoff series. Although the variance of sample autocorrelation at various lags might be estimated numerically through equations (2.47) or (2.48), this was not done because of the unstable nature of the sample estimate of autocorrelation function. Only smoothed variance spectrums were subject to further statistical analysis. Confidence limits of the variance spectrum were computed using equations (3.44) and (3.45).

ii) Examination of autocorrelations and variance spectrums of the noise or the residual terms of the proposed models. One basic assumption concerning the noise term of all the above models is that it is of the nature of a purely random process, or white noises.

iii) Examination of the unexplained portion of the variance or the one-step prediction variance due to the proposed models.

Correlation between model parameters and watershed physiographic characteristics

The next step was to seek the physical interpretation of the proposed stochastic model. This was attempted through the correlation regression between model parameters (criterion variables or the so-called dependent variables) and watershed physiographic characteristics (predictor variables or the so-called independent variables).

Both linear and non-linear relationships were attempted. In both cases only simple forms of relationships were assumed.

For the linear case, it was assumed that

$$Y_i = a_0+a_1X_1+a_2X_2+ \dots +a_kX_k, \tag{3.80}$$

where Y_i is the i^{th} model parameter; $X_1,X_2,\dots,X_k$ are the watershed physiographic characteristics; a_0 is the intercept constant and $a_1,a_2,\dots,a_k$ are regression coefficients.

For the non-linear case, a simple product relationship was assumed

$$Y_i = a_0X_1^{a_1}X_2^{a_2}\dots X_3^{a_k}, \tag{3.81a}$$

or
$$\log Y_i = \log a_0+a_1\log X_1+a_2\log X_2+ \dots +a_k\log X_k, \tag{3.81b}$$

where in the case of (3.81b) unity was added before logarithmic transformation to any variable which had equally likely chance to bear zero values.

Conventional multiple regression analysis was carried out except the following pre-screening procedure for possible reduction of the number of predictor variables:

i) Coefficients of correlation between criterion variables and predictor variables were first computed. Those predictor variables with less significant correlation with the criterion variables were then eliminated using Fisher's criterion, i.e. the statistic

$$z = \frac{1}{2}\ln\frac{1+r}{1-r} \tag{3.82}$$

is approximately normally distributed with variance $\frac{1}{N-3}$.

ii) The remaining predictor variables were then subject to the Varimax factor analysis (Kaiser, 1958; Harman, 1960) to identify the factor patterns or dominant predictor variables which were to be used for the final multiple regression analysis. This exercise would achieve the concept of 'simple structure' or 'parsimony' for the multiple regression model yet without introducing significant reduction or rather giving more meaningful indication of its ability in prediction of the behaviour of criterion variables. A brief description of the Varimax method of factor analysis is given in Appendix C. Application of Varimax factor analysis in hydrology can be found in Eiselstein (1967), Overton (1968), Wallis (1968), etc.

Along with the presentation and discussion of the results in the following section, the above procedures are illustrated in graphical forms by a typical example at 05AB002 - Willow Creek near Nolan, Alberta.

Results and Discussion

Seasonal and stochastic components of monthly runoff series

Before examining the seasonal variation in the monthly runoff series, the total mean and variance of the raw time series were first computed for each station and are listed in Table 3.3. It can be seen that runoffs in the mountainous areas generally have higher values with larger variance than those in the Prairies.

Harmonic analysis of the Fourier transformation of the runoff series is presented in Table 3.3 and it indicates that the annual cycle and its sub-harmonics, i.e., 12, 6, 4, 3, 2.4 and 2 (months) have accounted for approximately 24.5% to 93.9% of the total variance.

TABLE 3.3

Summary of Runoff Statistics for Stations Involved in Regional Regression Analysis

Station Identification			05AA023	05AB002	05AB021	05BJ005	05DB001	05DB002
Mean		in	0.96	0.20	0.30	0.84	0.84	0.59
Variance		in^2	1.59	0.14	0.29	0.58	0.58	0.35
Explained Seasonal Variance %	12 Month		47.83	25.68	27.16	52.08	53.33	38.71
	6		22.14	6.31	8.11	13.52	6.97	2.93
	4		7.58	1.03	1.34	4.36	1.64	1.65
	3		1.37	0.30	0.25	0.70	0.64	0.28
	2.4		0.19	0.11	0.09	0.16	0.19	0.03
	2		0.00	0.06	0.03	0.08	0.02	0.01
	Total		79.10	33.48	36.98	70.91	62.79	43.62
Residual Variance		in^2	0.33	0.09	0.18	0.17	0.22	0.20
1st-Order Autocorr. Coeff.			0.47	0.51	0.57	0.60	0.54	0.64
2nd			0.24	0.29	0.39	0.46	0.39	0.51

TABLE 3.3 (continued)

Summary of Runoff Statistics for Stations Involved in
Regional Regression Analysis

Station Identification		05DC001	05GG001	05KG002	05OD032	05TB002	05TD001
Mean	in	1.35	0.22	0.33	0.39	0.35	0.45
Variance	in^2	1.81	0.04	0.01	0.14	0.05	0.11
Explained Seasonal Variance %	12 Month	72.90	66.75	24.95	32.23	22.06	31.18
	6	14.22	4.26	1.28	26.68	4.15	6.28
	4	1.69	0.14	0.18	3.96	1.00	1.62
	3	0.30	1.02	0.02	2.46	0.43	0.87
	2.4	0.09	0.66	0.35	1.60	0.22	0.30
	2	0.04	0.16	0.05	0.39	0.01	0.06
	Total	89.24	73.01	26.84	67.32	27.87	40.31
Residual Variance	in^2	0.19	0.01	0.009	1.37	0.04	0.07
1st-Order Autocorr. Coeff.		0.56	0.64	0.91	0.58	0.76	0.79
2nd		0.40	0.42	0.80	0.44	0.64	0.69

TABLE 3.3 (continued)

Summary of Runoff Statistics for Stations Involved in Regional Regression Analysis

Station Identification			05TG001	06AF001	06CB002	06EA006	07AF002	07AG001
Mean		in	0.71	0.33	0.20	0.39	0.82	0.66
Variance		in^2	0.46	0.09	0.002	0.01	0.79	0.50
Explained Seasonal Variance %	12 Month		53.70	22.95	27.97	24.50	52.44	51.78
	6		16.24	0.76	3.21	3.25	12.14	10.32
	4		6.85	0.35	0.39	0.27	5.31	3.32
	3		1.39	0.29	0.11	0.15	0.68	0.13
	2.4		0.26	0.06	0.47	0.73	0.02	1.00
	2		0.00	0.08	0.19	0.01	0.08	0.64
	Total		78.44	24.49	32.33	28.92	70.66	67.18
Residual Variance		in^2	0.10	0.07	0.001	0.007	0.23	0.16
1st-Order Autocorr. Coeff.			0.67	0.71	0.89	0.90	0.52	0.55
2nd			0.43	0.59	0.78	0.76	0.34	0.38

TABLE 3.3 (continued)

Summary of Runoff Statistics for Stations Involved in Regional Regression Analysis

Station Identification		07BE001	07DA001	07JD001	07LE001	07OB001	08EE004
Mean	in	0.60	0.53	0.20	0.44	0.20	2.00
Variance	in^2	0.31	0.16	0.03	0.03	0.08	2.68
Explained Seasonal Variance %	12 Month	68.06	75.05	40.56	45.19	40.64	54.25
	6	7.37	6.92	13.40	9.09	13.43	24.43
	4	0.91	1.88	9.75	2.35	9.94	5.55
	3	0.05	0.03	7.02	0.60	10.23	1.07
	2.4	0.02	0.47	1.33	0.33	8.99	0.54
	2	0.00	0.33	0.93	0.01	2.86	0.03
	Total	76.40	84.68	72.98	57.56	86.08	85.87
Residual Variance	in^2	0.07	0.02	0.008	0.01	0.01	0.38
1st-Order Autocorr. Coeff.		0.64	0.65	0.52	0.86	0.58	0.36
2nd		0.53	0.51	0.34	0.76	0.47	0.05

TABLE 3.3 (continued)

Summary of Runoff Statistics for Stations Involved in Regional Regression Analysis

Station Identification		08JB002	08JB003	08KA005	08KA007	08KB001	08KH003
Mean	in	0.62	0.56	2.89	3.05	2.67	2.92
Variance	in^2	0.30	0.23	8.32	11.6	4.92	6.94
Explained Seasonal Variance %	12 Month	51.18	50.71	73.65	67.15	67.19	63.25
	6	19.37	19.11	16.06	18.65	15.03	16.27
	4	4.47	3.86	3.25	5.31	5.09	6.47
	3	1.23	1.33	0.23	0.97	0.37	0.76
	2.4	0.31	0.48	0.06	0.13	0.37	0.35
	2	0.01	0.04	0.01	0.05	0.01	0.00
	Total	76.56	75.53	93.26	92.25	88.05	87.09
Residual Variance	in^2	0.07	0.06	0.56	0.90	0.59	0.89
1st-Order Autocorr. Coeff.		0.85	0.84	0.33	0.38	0.43	0.50
2nd		0.67	0.67	0.19	0.16	0.17	0.25

TABLE 3.3 (continued)

Summary of Runoff Statistics for Stations Involved in Regional Regression Analysis

Station Identification		08KH006	08LA008	08LF051	08MA001	08ND011	08NG005
Mean	in	2.04	0.75	1.45	1.13	3.23	1.70
Variance	in^2	2.65	0.75	1.78	1.03	10.2	3.72
Explained Seasonal Variance %	12 Month	67.61	46.03	62.31	76.93	73.97	56.43
	6	15.94	24.28	21.21	14.70	16.80	23.01
	4	4.10	8.78	6.09	1.14	2.93	8.26
	3	0.32	2.21	0.89	0.04	0.17	1.71
	2.4	0.07	0.63	0.08	0.04	0.02	0.26
	2	0.01	0.08	0.00	0.03	0.00	0.02
	Total	88.05	82.01	90.59	92.88	93.89	89.69
Residual Variance	in^2	0.32	0.13	0.17	0.07	0.62	0.38
1st-Order Autocorr. Coeff.		0.59	0.67	0.61	0.46	0.39	0.61
2nd		0.35	0.46	0.33	0.13	0.19	0.42

TABLE 3.3 (continued)

Summary of Runoff Statistics for Stations Involved in Regional Regression Analysis

Station Identification		Units	08NG042	08NH006	08NH034	08NL004	08NP001	09AE001
Mean		in	1.67	1.49	1.27	0.82	2.49	1.03
Variance		in^2	3.46	5.03	3.94	1.51	12.4	0.98
Explained Seasonal Variance %	12 Month		55.91	45.51	42.72	39.37	45.70	53.21
	6		23.60	28.85	28.92	22.46	27.44	15.90
	4		8.29	10.72	11.84	9.25	11.19	9.89
	3		1.73	3.74	4.32	2.86	3.26	3.22
	2.4		0.28	1.13	1.18	0.54	0.78	0.78
	2		0.02	0.30	0.27	0.02	0.05	0.34
	Total		89.83	90.24	89.23	74.50	88.42	83.34
Residual Variance		in^2	0.35	0.49	0.42	0.38	1.44	0.16
1st-Order Autocorr. Coeff.			0.61	0.62	0.60	0.62	0.48	0.72
2nd			0.42	0.42	0.41	0.39	0.25	0.51

Seasonal influence also appears to be more profound in the mountainous areas.

Annual and seasonal components were removed by subtracting the long-term monthly means (Figure 3.2a) from the raw data using eq. (3.40). The remaining stochastic component then has variance varying from 6.1% to 75.5% of the total variance (Table 3.3). However, it is obvious from Figure 3.2b and the result of Bartlett's test of homogeneity of variances (Bartlett, 1937) also showed that there was significant difference among long-term monthly variances. In order to achieve the second-order stationarity, the effect of seasonal variation in monthly variances had to be removed using eq. (3.41):

$$y(t) = \frac{q(t)-m(t)}{s(t)}. \tag{3.41}$$

The resulting stochastic component y(t) would then yield zero mean and unit variance throughout the whole study period and between months.

The statistical analyses presented in the subsequent subsections are confined to this transformed stationary time series of y(t).

Autocorrelation function and variance spectrum

Sample estimates of the autocorrelation were computed using equations (2.8) and (2.10). It can be observed (Figure 3.3) that they follow generally the pattern of damped oscillation which is the characteristic belonging to the category of an autoregressive-moving average scheme (Jenkins and Watts, 1968). Autocorrelations of the first two lags are also listed in Table 3.3. It can be seen that large first-

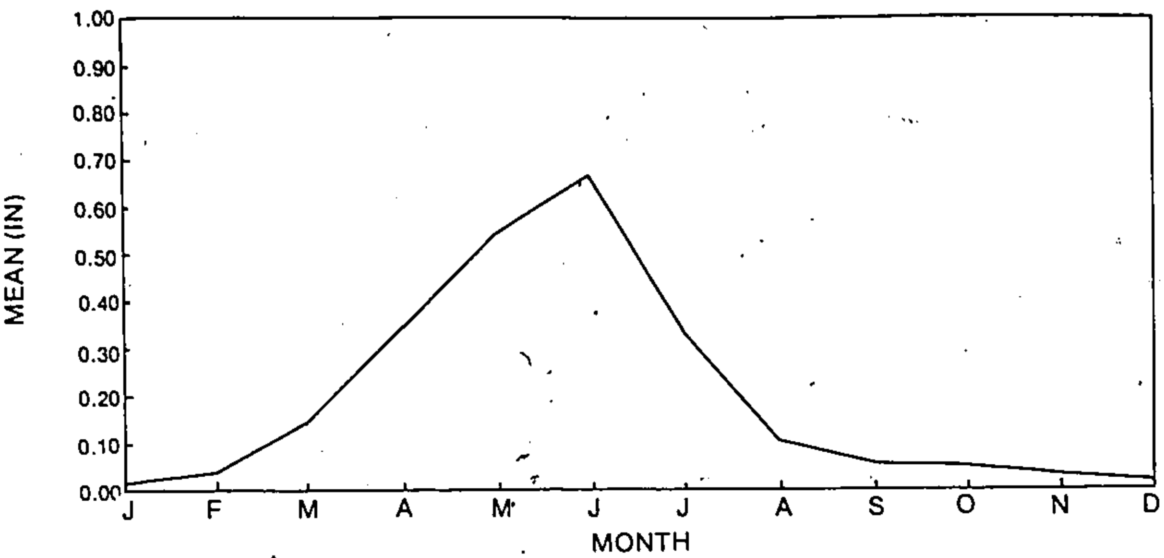


Figure 3.2a Monthly Means of Runoff - 05AB002 Willow Creek near Nolan

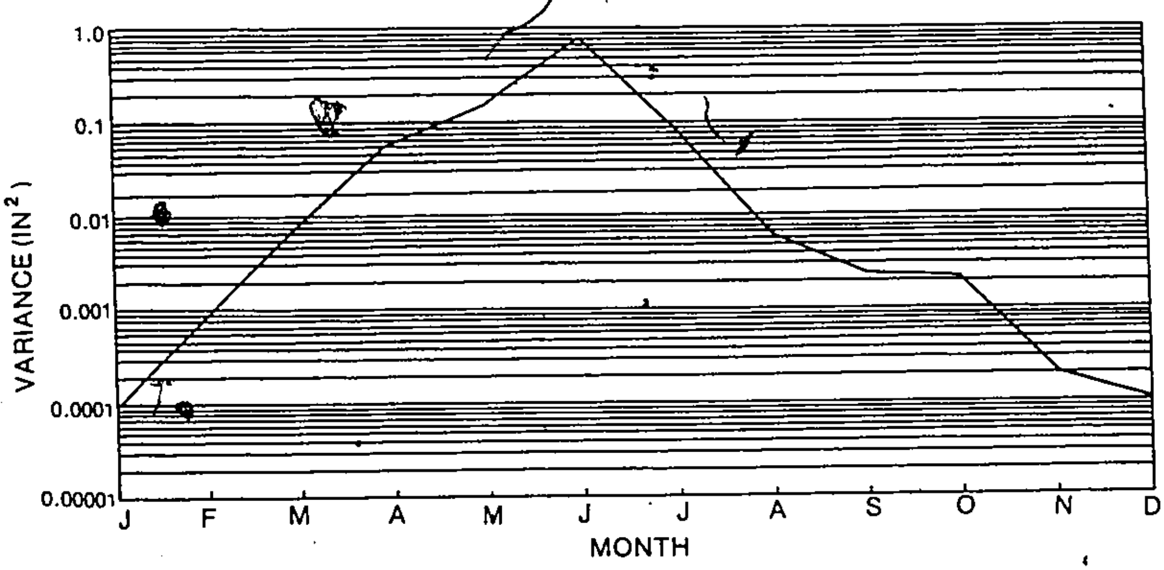


Figure 3.2b Monthly Variances of Runoff - 05AB002 Willow Creek near Nolan

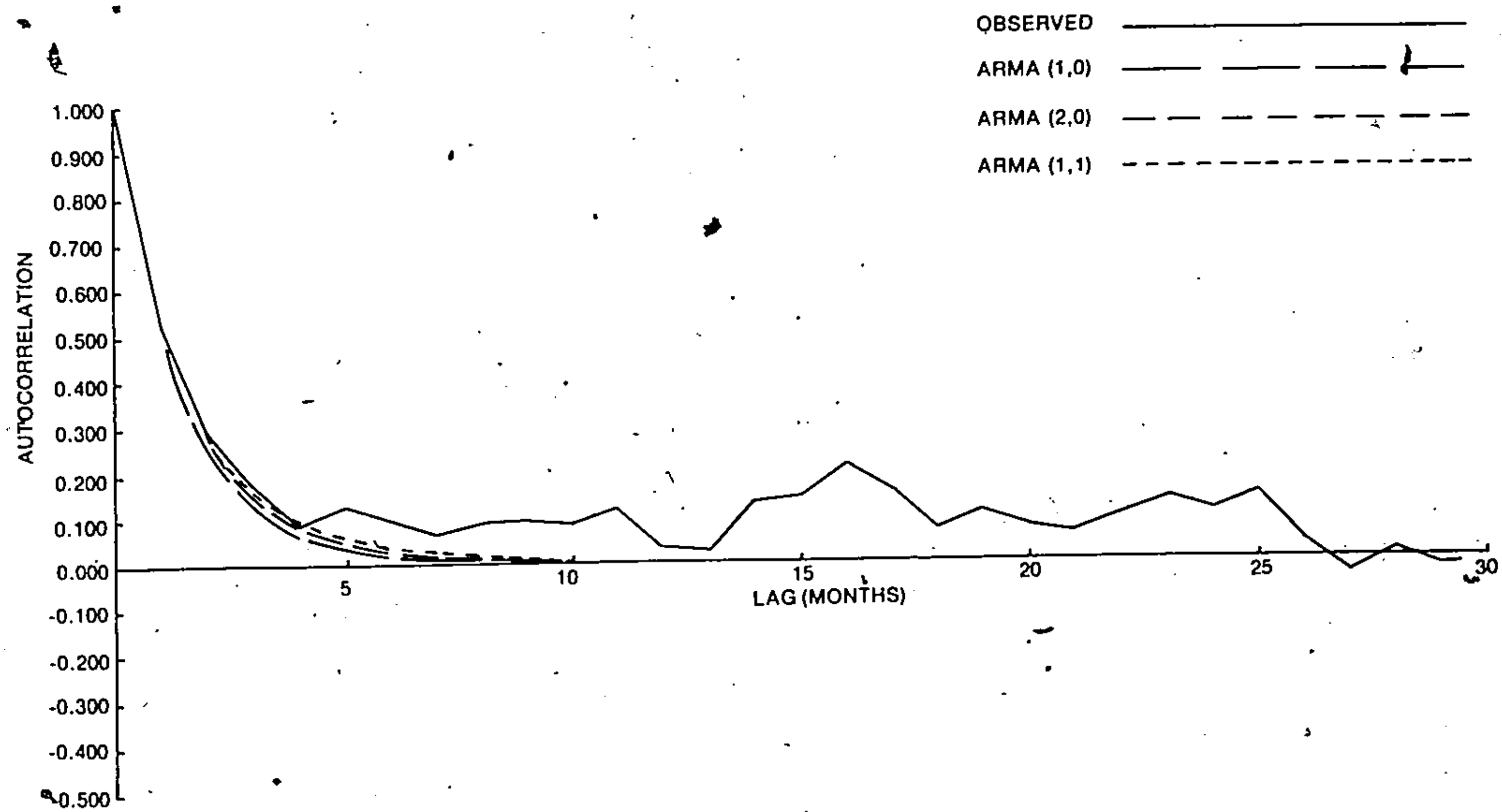


Figure 3.3 Autocorrelations of Stochastic Component of Runoff - 05AB002 Willow Creek near Nolan

order autocorrelations are usually associated with large storage areas (Table 3.2 and 3.3). Smoothed autospectrums were computed using eq. (2.54), whose pattern (Figure 3.4) further confirms that the underlying runoff generating process is of an autoregressive-moving average scheme (Jenkins and Watts, 1968).

Autoregressive-moving average representation of the stochastic process

The transformed series were fitted by the following autoregressive-moving average schemes, namely: first-order stochastic differential equations (S.D.E.) and its equivalent discrete ARMA(1,0) process, second-order S.D.E. and its equivalent ARMA(2,1) process, ARMA(1,1) process and ARMA(2,0) process. The results are given in Table 3.4*, which lists the computed model parameters and their standard deviations. In Table 3.4, two additional parameters were also computed, namely, 'unit noise variance' and 'gross noise variance'. The unit noise variance for the first-order S.D.E. is defined by eq. (3.17a), whereas that for the second-order S.D.E. is the value of eq. (3.26a) multiplied by a factor a_2^{-2} to assure the continuity condition of the runoff. The gross noise variance is then the product of the unit noise variance and the runoff residual variance which is defined as the variance of the stochastic component from eq. (3.40). The significance of these parameters would be discussed in the next sub-section.

It can be seen that most of the series could be represented by a first-order S.D.E. and its equivalent ARMA(1,0) process, and ARMA(1,1)

*In Table 3.4, 'Model Validity at 95% Level' was based on the various statistical tests at 95% confidence level on comparison of observed and theoretical autospectrums of the transformed runoff series, on randomness of the noise series to the proposed models. Question mark '?' indicates failure in any of the above tests to validate the proposed model. Standard deviations of estimates of coefficients in ARMA(1,0) and ARMA(2,1) processes are not computed because they were derived from the fitted first- and second-order S.D.E. models respectively.

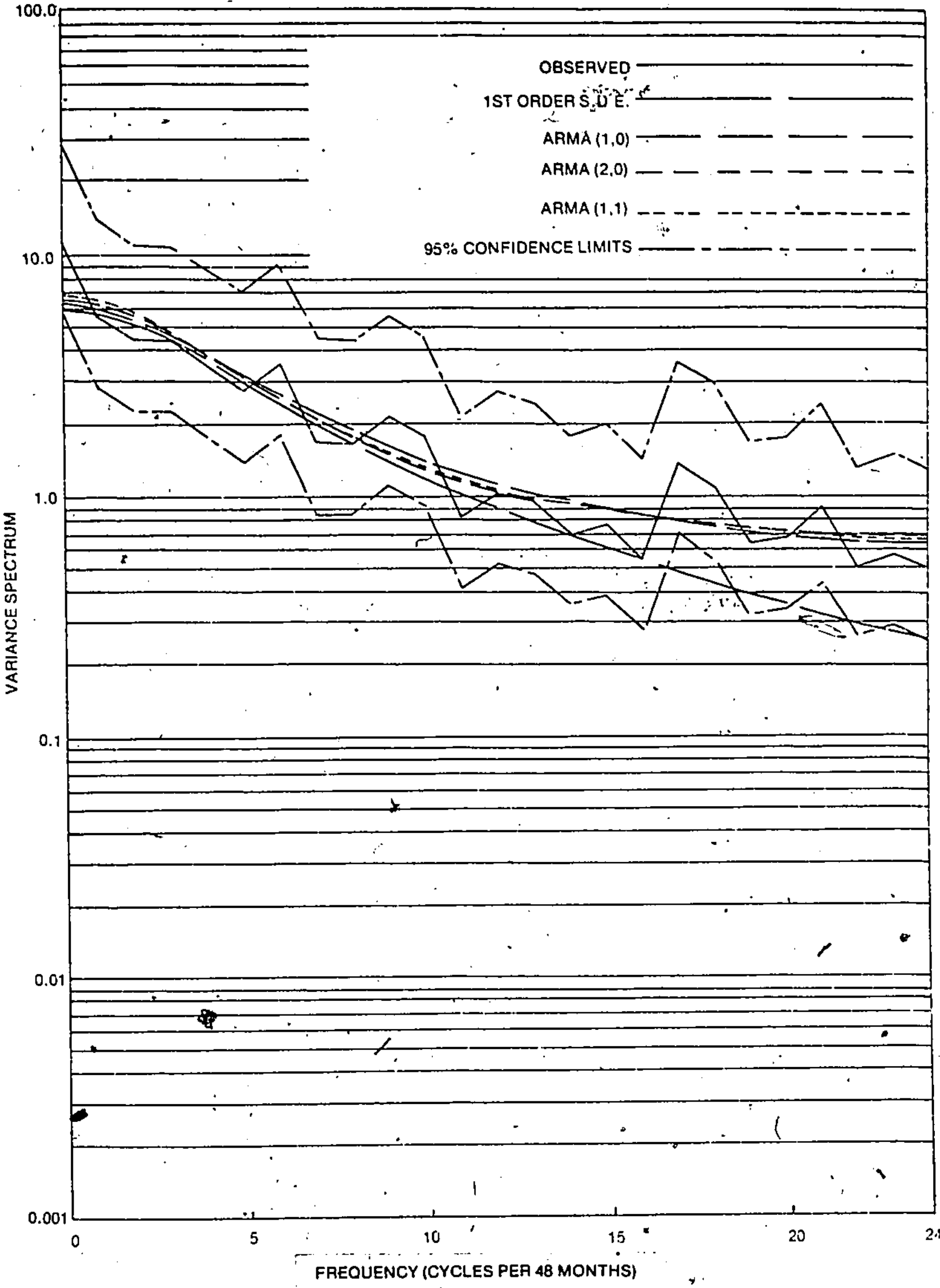


Figure 3.4 Variance Spectrums of Stochastic Component of Runoff - 05AB002 Willow Creek near Nolan

TABLE 3.4

Summary on Runoff Generating Process Identification

Station Identification		05AA023		05AB002		05AB021	
		Estimate	Std.Dev.	Estimate	Std.Dev.	Estimate	Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	1.32	0.16	1.50	0.19	1.79	0.25
$(1-d_1B)y_t=w_0x_t$	d_1	0.47		0.51		0.57	
	w_0	0.54		0.50		0.43	
Unit Noise Variance		2.65		2.99		3.58	
Gross Noise Variance	in^2	0.87		0.27		0.64	
Model Validity at 95% Level		yes		yes		yes	
Explained Residual Variance	%	25		30		33	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	3.55	0.20	3.27	0.19	2.92	0.18
	a_2	3.16	0.35	2.68	0.31	2.12	0.26
$(1-d_1B-d_2B^2)y_t=(w_0-w_1B)x_t$	d_1	0.34		0.39		0.47	
	d_2	-0.03		-0.04		-0.05	
	w_0	0.18		0.20		0.23	
	w_1	-0.03		-0.04		-0.05	
Unit Noise Variance		2.24		2.45		2.76	
Gross Noise Variance	in^2	0.74		0.22		0.50	
Model Validity at 95% Level		?		yes		?	
Explained Residual Variance	%	25		30		31	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.50	0.14	0.56	0.12	0.68	0.09
	w_1	0.04	0.16	0.06	0.14	0.17	0.13
Model Validty at 95% Level		yes		yes		yes	
Explained Residual Variance	%	25		31		34	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	0.46	0.07	0.50	0.07	0.52	0.07
	d_2	0.02	0.07	0.03	0.07	0.09	0.07
Model Validity at 95% Level		yes		yes		yes	
Explained Residual Variance	%	26		32		35	

TABLE 3.4 (continued)

Summary on Runoff Generating Process Identification

Station Identification		05BJ005 Estimate	05BJ005 Std.Dev.	05DB001 Estimate	05DB001 Std.Dev.	05DB002 Estimate	05DB002 Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	1.95	0.29	1.65	0.22	2.28	0.36
$(1-d_1B)y_t=w_0x_t$	d_1	0.60		0.54		0.64	
	w_0	0.40		0.46		0.36	
Unit Noise Variance	in^2	3.91		3.29		4.55	
Gross Noise Variance	in^2	0.66		0.72		0.91	
Model Validity at 95% Level		?		?		yes	
Explained Residual Variance	%	40		31		44	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	2.76	0.18	3.08	0.18	2.50	0.17
	a_2	1.90	0.24	2.37	0.28	1.56	0.21
$(1-d_1B-d_2B^2)y_t=(w_0-w_1B)x_t$	d_1	0.50		0.43		0.57	
	d_2	-0.06		-0.05		-0.08	
	w_0	0.24		0.22		0.27	
	w_1	-0.05		-0.04		-0.06	
Unit Noise Variance	in^2	2.91		2.60		3.21	
Gross Noise Variance	in^2	0.49		0.57		0.64	
Model Validity at 95% Level		?		?		?	
Explained Residual Variance	%	35		27		38	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.76	0.08	0.71	0.09	0.79	0.07
	w_1	0.26	0.12	0.24	0.13	0.25	0.11
Model Validty at 95% Level		?		?		yes	
Explained Residual Variance	%	43		33		45	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	0.51	0.07	0.47	0.07	0.54	0.07
	d_2	0.15	0.07	0.13	0.07	0.16	0.07
Model Validity at 95% Level		?		?		yes	
Explained Residual Variance	%	42		32		45	

TABLE 3.4 (continued)

Summary on Runoff Generating Process Identification

Station Identification		05DC001 Estimate	05DC001 Std.Dev.	05GG001 Estimate	05GG001 Std.Dev.	05KG002 Estimate	05KG002 Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	1.73	0.24	2.21	0.34	11.2	3.87
$(1-d_1B)y_t=w_0x_t$	d_1	0.56		0.64		0.91	
	w_0	0.44		0.37		0.09	
Unit Noise Variance	in^2	3.47		4.41		22.1	
Gross Noise Variance		0.66		0.04		0.20	
Model Validity at 95% Level		yes		yes		yes	
Explained Residual Variance	%	34		41		84	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	2.98	0.18	2.55	0.17	0.98	0.10
	a_2	2.22	0.27	1.63	0.21	0.24	0.05
$(1-d_1B-d_2B^2)y_t=(w_0-w_1B)x_t$	d_1	0.45		0.56		1.12	
	d_2	-0.05		-0.08		-0.37	
	w_0	0.22		0.26		0.50	
	w_1	-0.05		-0.06		-0.13	
Unit Noise Variance	in^2	2.68		3.13		7.99	
Gross Noise Variance		0.51		0.03		0.07	
Model Validity at 95% Level		?		yes		?	
Explained Residual Variance	%	28		38		81	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.71	0.09	0.66	0.09	0.88	0.04
	w_1	0.22	0.12	0.03	0.12	-0.20	0.08
Model Validty at 95% Level		?		yes		yes	
Explained Residual Variance	%	34		41		84	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	0.49	0.07	0.62	0.07	1.10	0.07
	d_2	0.12	0.07	0.02	0.07	-0.18	0.07
Model Validity at 95% Level		?		yes		yes	
Explained Residual Variance	%	34		41		84	

TABLE 3.4 (continued)

Summary on Runoff Generating Process Identification

Station Identification		05OD032		05TB002		05TD001	
		Estimate	Std.Dev.	Estimate	Std.Dev.	Estimate	Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	1.86	0.27	3.60	0.72	4.32	0.95
$(1-d_1B)y_t=w_0x_t$	d_1	0.58		0.76		0.79	
	w_0	0.42		0.24		0.21	
Unit Noise Variance	in^2	3.72		7.20		8.65	
Gross Noise Variance		5.10		0.22		0.61	
Model Validity at 95% Level		?		?		yes	
Explained Residual Variance	%	35		60		64	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	2.84	0.18	1.88	0.14	1.68	0.14
	a_2	2.02	0.25	0.88	0.13	0.71	0.12
$(1-d_1B-d_2B^2)y_t = (w_0-w_1B)x_t$	d_1	0.48		0.78		0.86	
	d_2	-0.06		-0.15		-0.19	
	w_0	0.24		0.34		0.37	
	w_1	-0.05		-0.08		-0.09	
Unit Noise Variance	in^2	2.82		4.30		4.74	
Gross Noise Variance		3.87		0.13		0.34	
Model Validity at 95% Level		?		?		?	
Explained Residual Variance	%	29		51		53	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.76	0.08	0.85	0.05	0.88	0.04
	w_1	0.27	0.12	0.21	0.10	0.23	0.09
Model Validty at 95% Level		?		yes		yes	
Explained Residual Variance	%	37		61		65	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	0.49	0.07	0.64	0.07	0.65	0.07
	d_2	0.16	0.07	0.16	0.07	0.18	0.07
Model Validity at 95% Level		?		yes		?	
Explained Residual Variance	%	36		61		65	

TABLE 3.4 (continued)

Summary on Runoff Generating Process Identification

Station Identification		05TG001		06AF001		06CB002	
		Estimate	Std.Dev.	Estimate	Std.Dev.	Estimate	Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	2.53	0.42	2.90	0.52	8.53	2.63
$(1-d_1B)y_t=w_0x_t$	d_1	0.67		0.71		0.89	
	w_0	0.33		0.29		0.11	
Unit Noise Variance	in^2	5.05		5.80		17.1	
Gross Noise Variance		0.50		0.35		0.02	
Model Validity at 95% Level		yes		?		?	
Explained Residual Variance	%	47		51		81	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	2.34	0.16	2.15	0.15	6.80	0.27
	a_2	1.37	0.19	1.15	0.17	0.92	0.26
$(1-d_1B-d_2B^2)y_t = (w_0-w_1B)x_t$	d_1	0.62		0.68		0.87	
	d_2	-0.10		-0.12		-0.001	
	w_0	0.28		0.30		0.13	
	w_1	-0.06		-0.07		-0.01	
Unit Noise Variance	in^2	3.42		3.75		14.8	
Gross Noise Variance		0.34		0.23		0.02	
Model Validity at 95% Level		yes		?		?	
Explained Residual Variance	%	45		41		81	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.64	0.08	0.83	0.06	0.87	0.04
	w_1	-0.06	0.11	0.26	0.10	-0.09	0.09
Model Validty at 95% Level		yes		?		yes	
Explained Residual Variance	%	46		52		81	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	0.70	0.07	0.58	0.07	0.96	0.07
	d_2	-0.04	0.07	0.17	0.07	-0.08	0.07
Model Validity at 95% Level		yes		?		yes	
Explained Residual Variance	%	47		52		81	

TABLE 3.4 (continued)

Summary on Runoff Generating Process Identification

Station Identification		06EA006 Estimate	06EA006 Std.Dev.	07AF002 Estimate	07AF002 Std.Dev.	07AG001 Estimate	07AG001 Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	9.96	3.31	1.52	0.20	1.67	0.23
$(1-d_1B)y_t=w_0x_t$	d_1	0.90		0.52		0.55	
	w_0	0.10		0.49		0.46	
Unit Noise Variance		19.9		3.05		3.34	
Gross Noise Variance	in^2	0.14		0.70		0.53	
Model Validity at 95% Level		?		?		?	
Explained Residual Variance	%	83		27		32	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	1.84	0.14	3.24	0.19	3.05	0.18
	a_2	0.33	0.08	2.62	0.31	2.32	0.28
$(1-d_1B-d_2B^2)y_t=(w_0-w_1B)x_t$	d_1	1.01		0.40		0.44	
	d_2	-0.16		-0.04		-0.05	
	w_0	0.36		0.21		0.22	
	w_1	-0.08		-0.04		-0.04	
Unit Noise Variance		11.2		2.46		2.64	
Gross Noise Variance	in^2	0.08		0.57		0.42	
Model Validity at 95% Level		yes		yes		?	
Explained Residual Variance	%	85		25		yes	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.84	0.04	0.66	0.10	0.69	0.10
	w_1	-0.44	0.07	0.20	0.14	0.20	0.13
Model Validty at 95% Level		yes		?		?	
Explained Residual Variance	%	84		29		33	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	1.21	0.07	0.47	0.07	0.49	0.07
	d_2	-0.34	0.07	0.10	0.07	0.11	0.07
Model Validity at 95% Level		yes		?		?	
Explained Residual Variance	%	85		29		32	

TABLE 3.4 (continued)

Summary on Runoff Generating Process Identification

Station Identification		07BE001 Estimate	07BE001 Std.Dev.	07DA001 Estimate	07DA001 Std.Dev.	07JD001 Estimate	07JD001 Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	2.23	0.35	2.29	0.37	1.53	0.20
$(1-d_1B)y_t=w_0x_t$	d_1	0.64		0.65		0.52	
	w_0	0.36		0.36		0.49	
Unit Noise Variance		4.46		4.59		3.05	
Gross Noise Variance	in^2	0.31		0.14		0.02	
Model Validity at 95% Level		?		?		yes	
Explained Residual Variance	%	42		42		28	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	2.53	0.17	2.49	0.25	3.24	0.19
	a_2	1.61	0.21	1.55	0.31	2.62	0.31
$(1-d_1B-d_2B^2)y_t=(w_0-w_1B)x_t$	d_1	0.56		0.58		0.40	
	d_2	-0.08		-0.08		-0.04	
	w_0	0.26		0.27		0.21	
	w_1	-0.06		-0.06		-0.04	
Unit Noise Variance		3.14		3.20		2.46	
Gross Noise Variance	in^2	0.22		0.10		0.02	
Model Validity at 95% Level		?		?		?	
Explained Residual Variance	%	34		35		25	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.83	0.06	0.79	0.07	0.66	0.11
	w_1	0.34	0.10	0.26	0.11	0.19	0.14
Model Validty at 95% Level		?		yes		yes	
Explained Residual Variance	%	45		44		29	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	0.51	0.07	0.54	0.07	0.47	0.07
	d_2	0.21	0.07	0.16	0.07	0.10	0.07
Model Validity at 95% Level		?		yes		yes	
Explained Residual Variance	%	44		44		29	

TABLE 3.4 (continued)

Summary of Runoff Generating Process Identification

Station Identification		07LE001		07OB001		08EE004	
		Estimate	Std.Dev.	Estimate	Std.Dev.	Estimate	Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	6.52	1.75	1.82	0.26	0.98	0.10
$(1-d_1B)y_t=w_0x_t$	d_1	0.86		0.58		0.36	
	w_0	0.14		0.43		0.66	
Unit Noise Variance	in^2	13.0		3.64		1.96	
Gross Noise Variance	in^2	0.17		0.04		0.74	
Model Validity at 95% Level		?		?		yes	
Explained Residual Variance	%	75		36		15	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	1.32	0.12	2.88	0.18	-	
	a_2	0.44	0.25	2.08	0.26	-	
$(1-d_1B-d_2B^2)y_t=(w_0-w_1B)x_t$	d_1	1.03		0.47		-	
	d_2	-0.27		-0.06		-	
	w_0	0.43		0.23		-	
	w_1	-0.11		-0.05		-	
Unit Noise Variance	in^2	5.99		2.77		-	
Gross Noise Variance	in^2	0.06		0.03		-	
Model Validity at 95% Level		?		?		-	
Explained Residual Variance	%	68		31		-	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.88	0.04	0.82	0.07	0.13	0.19
	w_1	0.08	0.09	0.38	0.11	-0.27	0.19
Model Validty at 95% Level		?		yes		yes	
Explained Residual Variance	%	75		40		16	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	0.80	0.07	0.46	0.07	0.40	0.07
	d_2	0.07	0.07	0.21	0.07	-0.10	0.07
Model Validity at 95% Level		?		yes		yes	
Explained Residual Variance	%	75		40		16	

TABLE 3.4 (continued)

Summary of Runoff Generating Process Identification

Station Identification		08JB002 Estimate	08JB002 Std.Dev.	08JB003 Estimate	08JB003 Std.Dev.	08KA005 Estimate	08KA005 Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	6.25	1.65	5.84	1.49	0.90	0.09
$(1-d_1B)y_t=w_0x_t$	d_1	0.85		0.84		0.33	
	w_0	0.15		0.16		0.70	
Unit Noise Variance		12.5		11.7		1.80	
Gross Noise Variance	in^2	0.88		0.70		1.01	
Model Validity at 95% Level		?		yes		?	
Explained Residual Variance	%	72		72		12	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	2.91	0.18	1.41	0.12	-	
	a_2	0.67	0.15	0.50	0.09	-	
$(1-d_1B-d_2B^2)y_t=(w_0-w_1B)x_t$	d_1	0.85		0.99		-	
	d_2	-0.05		-0.24		-	
	w_0	0.26		0.41		-	
	w_1	-0.05		-0.10		-	
Unit Noise Variance		8.64		5.60		-	
Gross Noise Variance	in^2	0.60		0.32		-	
Model Validity at 95% Level		yes		?		-	
Explained Residual Variance	%	74		69		-	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.78	0.05	0.79	0.05	0.57	0.17
	w_1	-0.26	0.08	-0.16	0.09	0.28	0.20
Model Validty at 95% Level		yes		yes		?	
Explained Residual Variance	%	74		72		13	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	1.03	0.07	0.95	0.07	0.30	0.07
	d_2	-0.21	0.07	-0.13	0.07	0.09	0.07
Model Validity at 95% Level		yes		?		?	
Explained Residual Variance	%	74		72		12	

TABLE 3.4 (continued)

Summary on Runoff Generating Process Identification

Station Identification		08KA007		08KB001		08KH003	
		Estimate	Std.Dev.	Estimate	Std.Dev.	Estimate	Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	1.03	0.11	1.17	0.13	1.45	0.18
$(1-d_1B)y_t=w_0x_t$	d_1	0.38		0.43		0.50	
	w_0	0.65		0.59		0.51	
Unit Noise Variance	in^2	2.05		2.35		2.90	
Gross Noise Variance	in^2	1.84		1.39		2.61	
Model Validity at 95% Level		yes		yes		yes	
Explained Residual Variance	%	16		20		26	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	–		3.85	0.21	3.35	0.19
	a_2	–		3.71	0.40	2.80	0.32
$(1-d_1B-d_2B^2)y_t=(w_0-w_1B)x_t$	d_1	–		0.29		0.38	
	d_2	–		-0.02		-0.04	
	w_0	–		0.17		0.20	
	w_1	–		-0.03		-0.04	
Unit Noise Variance	in^2	–		2.07		2.39	
Gross Noise Variance	in^2	–		1.23		2.12	
Model Validity at 95% Level		–		yes		yes	
Explained Residual Variance	%	–		19		25	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.42	0.18	0.40	0.16	0.50	0.13
	w_1	0.05	0.20	-0.04	0.17	-0.002	0.15
Model Validty at 95% Level		yes		yes		yes	
Explained Residual Variance	%	16		20		26	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	0.37	0.07	0.43	0.07	0.50	0.07
	d_2	0.02	0.07	-0.02	0.07	-0.001	0.07
Model Validity at 95% Level		yes		yes		yes	
Explained Residual Variance	%	16		20		26	

TABLE 3.4 (continued)

Summary on Runoff Generating Process Identification

Station Identification		08KH006		08LA008		08LF051	
		Estimate	Std.Dev.	Estimate	Std.Dev.	Estimate	Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	1.91	0.28	2.46	0.41	2.03	0.30
$(1-d_1B)y_t=w_0x_t$	d_1	0.59		0.67		0.61	
	w_0	0.41		0.34		0.39	
Unit Noise Variance		3.83		4.91		4.06	
Gross Noise Variance	in^2	1.23		0.64		0.69	
Model Validity at 95% Level		yes		yes		yes	
Explained Residual Variance	%	36		45		38	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	2.79	0.18	2.38	0.16	1.27	0.17
	a_2	1.95	0.25	1.42	0.19	1.18	0.23
$(1-d_1B-d_2B^2)y_t=(w_0-w_1B)x_t$	d_1	0.49		0.61		0.52	
	d_2	-0.06		-0.09		-0.07	
	w_0	0.24		0.28		0.25	
	w_1	-0.05		-0.06		-0.05	
Unit Noise Variance		2.87		3.36		2.15	
Gross Noise Variance	in^2	0.92		0.45		0.37	
Model Validity at 95% Level		?		?		yes	
Explained Residual Variance	%	33		40		38	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.59	0.10	0.69	0.08	0.54	0.10
	w_1	-0.01	0.13	0.04	0.11	-0.11	0.12
Model Validty at 95% Level		yes		yes		yes	
Explained Residual Variance	%	36		45		39	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	0.60	0.07	0.65	0.07	0.65	0.07
	d_2	-0.005	0.07	0.03	0.07	-0.07	0.07
Model Validity at 95% Level		yes		yes		yes	
Explained Residual Variance	%	36		45		39	

TABLE 3.4 (continued)

Summary on Runoff Generating Process Identification

Station Identification		08MA001		08ND011		08NG005	
		Estimate	Std.Dev.	Estimate	Std.Dev.	Estimate	Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	1.27	0.15	1.07	0.12	2.01	0.30
$(1-d_1B)y_t=w_0x_t$	d_1	0.46		0.39		0.61	
	w_0	0.56		0.63		0.40	
Unit Noise Variance		2.55		2.15		4.02	
Gross Noise Variance	in^2	0.18		1.33		1.53	
Model Validity at 95% Level		yes		yes		yes	
Explained Residual Variance	%	22		17		38	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	3.64	0.20	4.09	0.21	2.70	0.17
	a_2	3.32	0.37	4.18	0.44	1.83	0.23
$(1-d_1B-d_2B^2)y_t=(w_0-w_1B)x_t$	d_1	0.32		0.26		0.52	
	d_2	-0.03		-0.02		-0.07	
	w_0	0.18		0.16		0.25	
	w_1	-0.03		-0.03		-0.05	
Unit Noise Variance		2.20		1.96		2.95	
Gross Noise Variance	in^2	0.15		1.21		1.12	
Model Validity at 95% Level		yes		yes		?	
Explained Residual Variance	%	22		16		34	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.29	0.15	0.47	0.17	0.68	0.09
	w_1	-0.22	0.15	0.09	0.19	0.12	0.12
Model Validty at 95% Level		yes		yes		yes	
Explained Residual Variance	%	22		18		39	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	0.50	0.07	0.38	0.07	0.56	0.07
	d_2	-0.10	0.07	0.03	0.07	0.07	0.07
Model Validity at 95% Level		yes		yes		yes	
Explained Residual Variance	%	22		18		39	

TABLE 3.4 (continued)

Summary on Runoff Generating Process Identification

Station Identification		08NG042 Estimate	08NG042 Std.Dev.	08NH006 Estimate	08NH006 Std.Dev.	08NH034 Estimate	08NH034 Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	2.00	0.30	2.09	0.32	1.98	0.29
$(1-d_1B)y_t=w_0x_t$	d_1	0.61		0.62		0.60	
	w_0	0.40		0.38		0.40	
Unit Noise Variance	in^2	4.00		4.18		3.96	
Gross Noise Variance		1.40		2.05		1.66	
Model Validity at 95% Level		yes		?		?	
Explained Residual Variance	%	38		38		37	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	2.72	0.17	2.64	0.17	2.73	0.17
	a_2	1.84	0.24	1.74	0.23	1.87	0.24
$(1-d_1B-d_2B^2)y_t=(w_0-w_1B)x_t$	d_1	0.51		0.53		0.51	
	d_2	-0.07		-0.07		-0.06	
	w_0	0.25		0.25		0.24	
	w_1	-0.05		-0.06		-0.05	
Unit Noise Variance	in^2	2.95		3.04		2.92	
Gross Noise Variance		1.03		1.49		1.22	
Model Validity at 95% Level		?		?		?	
Explained Residual Variance	%	35		35		33	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.69	0.09	0.68	0.09	0.69	0.09
	w_1	0.14	0.12	0.10	0.12	0.13	0.12
Model Validdty at 95% Level		yes		?		?	
Explained Residual Variance	%	40		39		38	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	0.56	0.07	0.58	0.07	0.56	0.07
	d_2	0.08	0.07	0.06	0.07	0.08	0.07
Model Validity at 95% Level		yes		?		?	
Explained Residual Variance	%	39		39		38	

TABLE 3.4 (continued)

Summary on Runoff Generating Process Identification

Station Identification		08NL004 Estimate	08NL004 Std.Dev.	08NP001 Estimate	08NP001 Std.Dev.	09AE001 Estimate	09AE001 Std.Dev.
$(1+a_1D)y(t)=x(t)$	a_1	2.12	0.33	1.36	0.17	3.03	0.56
$(1-d_1B)y_t=w_0x_t$	d_1	0.62		0.48		0.72	
	w_0	0.38		0.53		0.28	
Unit Noise Variance	in^2	4.24		2.73		6.07	
Gross Noise Variance	in^2	1.61		3.93		0.97	
Model Validity at 95% Level		yes		yes		yes	
Explained Residual Variance	%	40		24		53	
$(D^2+a_1D+a_2)y(t)=x(t)$	a_1	2.61	0.17	3.48	0.20	2.09	0.15
	a_2	1.71	0.22	3.03	0.34	1.09	0.16
$(1-d_1B-d_2B^2)y_t=(w_0-w_1B)x_t$	d_1	0.54		0.35		0.70	
	d_2	-0.07		-0.03		-0.12	
	w_0	0.26		0.19		0.31	
	w_1	-0.06		-0.04		-0.07	
Unit Noise Variance	in^2	3.06		2.30		3.85	
Gross Noise Variance	in^2	1.16		3.31		0.61	
Model Validity at 95% Level		?		?		?	
Explained Residual Variance	%	38		23		50	
$(1-d_1B)y_t=(1-w_1B)z_t$	d_1	0.62	0.09	0.52	0.13	0.71	0.07
	w_1	-0.002	0.12	0.06	0.15	-0.01	0.10
Model Validty at 95% Level		yes		yes		yes	
Explained Residual Variance	%	40		24		53	
$(1-d_1B-d_2B^2)y_t=z_t$	d_1	0.62	0.07	0.47	0.07	0.72	0.07
	d_2	-0.001	0.07	0.03	0.07	-0.01	0.07
Model Validity at 95% Level		yes		yes		yes	
Explained Residual Variance	%	40		24		53	

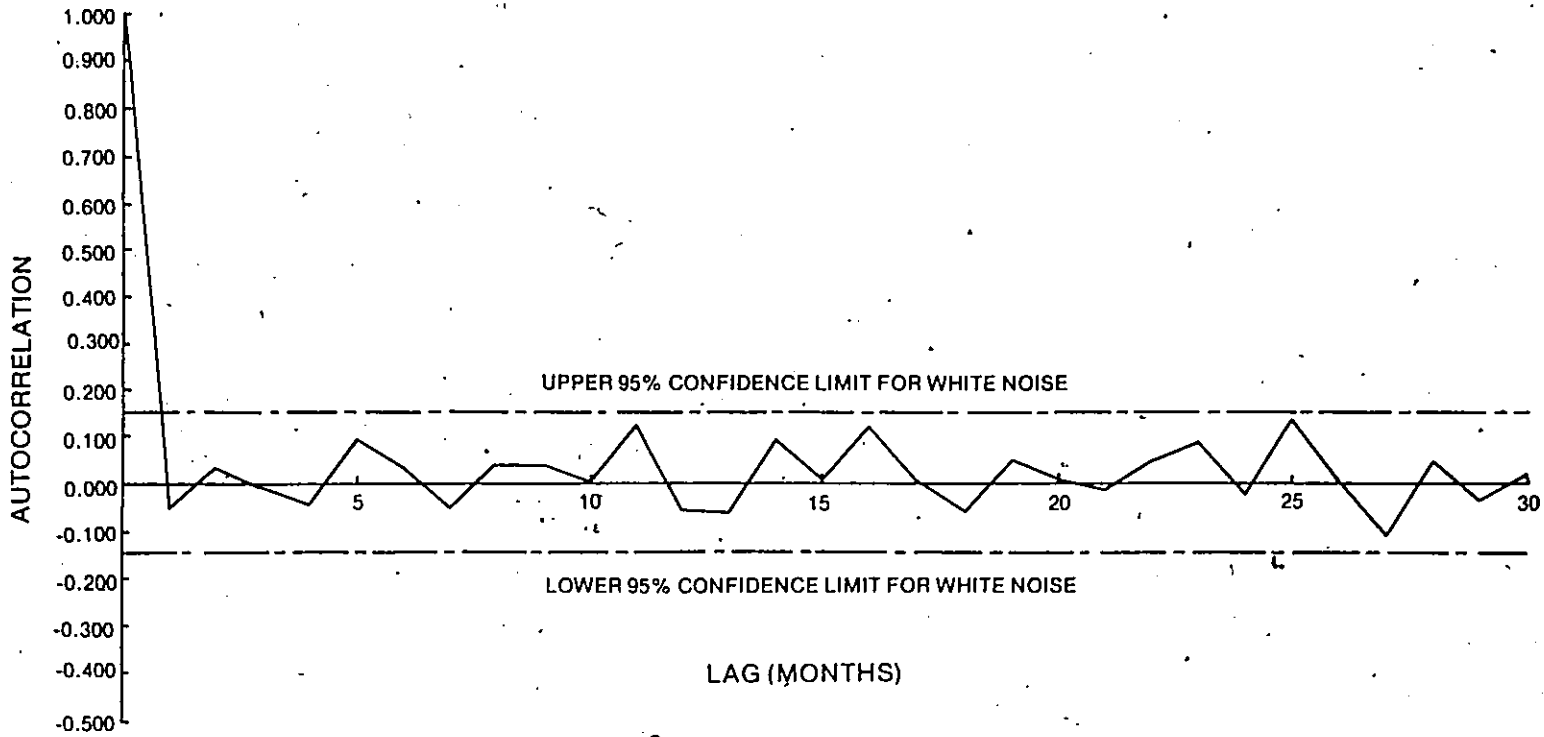


Figure 3.5 Autocorrelations of Noise to ARMA(1,0) Mod - 05AB002 Willow Creek near Nolan

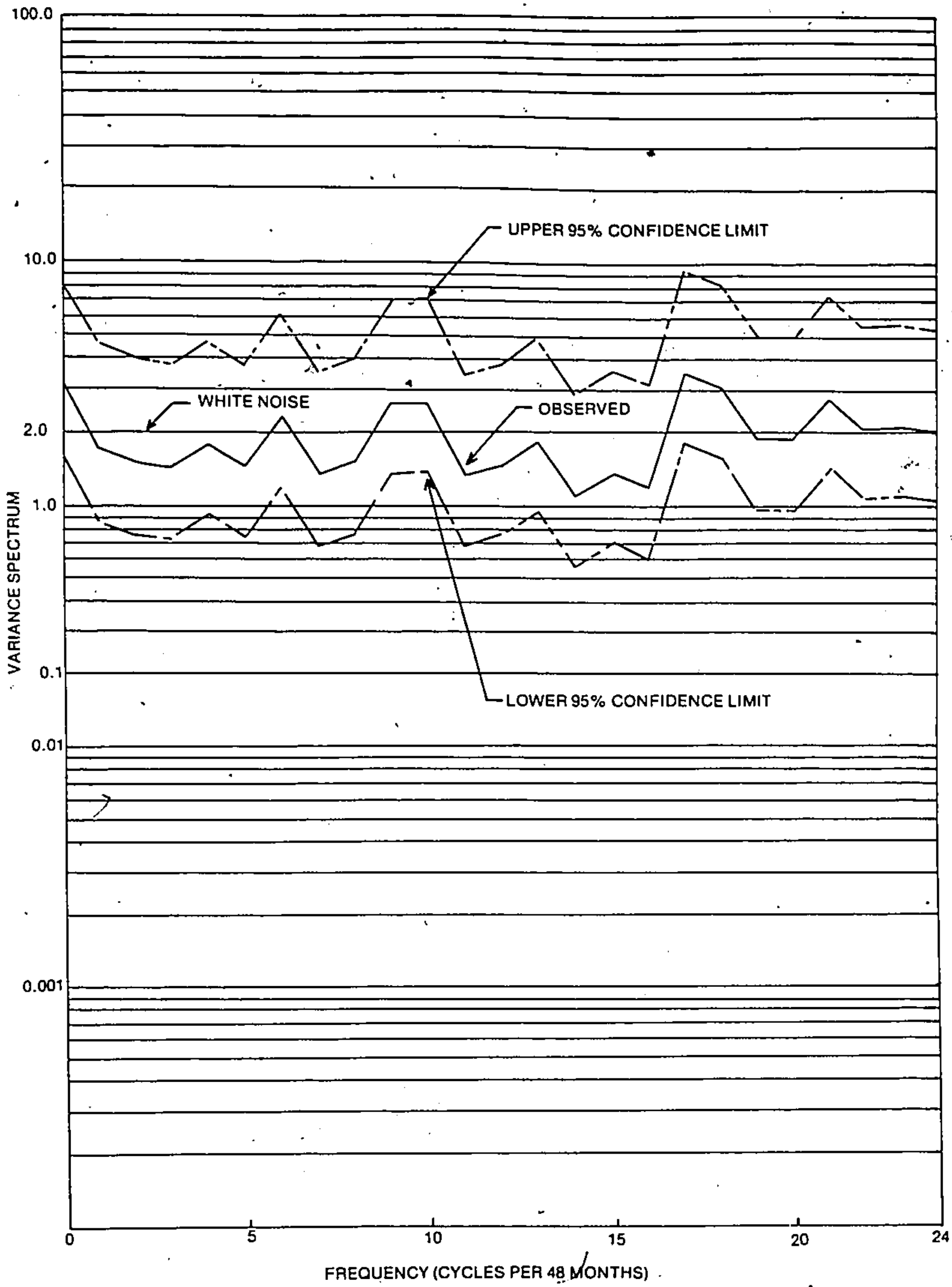


Figure 3.6 Variance Spectrums of Noise to ARMA(1,0) Model - 05AB002
Willow Creek near Nolan

process and an ARMA(2,0) process (Table 3.4 and Figures 3.3, 3.4, 3.5 and 3.6).

In the case of the ARMA(1,1) model, it is observed that the value of coefficient d_1 is of the same order of magnitude as that of the ARMA(1,0) model. By examining the standard error of estimate computed from eq. (3.70), it can be seen that the coefficient w_1 is of relatively small magnitude hardly to be statistically significant.

Similar observation could be made for the ARMA(2,0) or second-order autoregressive process model. The value of the coefficient d_1 is almost the same as that of the ARMA(1,0) model. The second coefficient d_2 is small with respect to the standard error of estimate computed from eq. (3.77) and can be considered not significant.

The second-order S.D.E. and the equivalent ARMA(2,1) process did not succeed to represent the transformed runoff series (Table 3.4) and were found in most cases being the special case, i.e., the solutions to the characteristic equation are of equal roots. It was also found that the percentage explained residual variances in all cases were lower than those from the first-order S.D.E. model. A possible explanation of this is when a fit of a second-order S.D.E. model is attempted to a time series which is in reality a first-order continuous autoregressive process described by a first-order S.D.E., it implies the following operation

$$(1+K'D)(1+KD)y(t) = (1+K'D)x(t). \tag{3.83}$$

The above equation contains a common factor (1+K'D) on both sides of the equation. This is the case of model overfitting, which leads to the instability of the model and explains the reduction in the percentage explained residual variance (Box and Jenkins, 1970).

Therefore it appears that the stochastic component of a monthly runoff series may well be represented by a first-order S.D.E. and its equivalent ARMA(1,0) or first-order autoregressive process as follows:

$$(1+KD)y(t) = x(t), \tag{3.13}$$

and

$$(1-d_1B)y_n = w_0x_n. \tag{3.19}$$

Any attempt to fit a monthly runoff series by a higher order S.D.E. or its equivalent ARMA process may lead to unstable model structure. On the other hand, the ARMA(1,1) and second-order autoregressive models can produce the same result but may be reduced to the simple ARMA(1,0) structure.

After identifying the transformed monthly runoff series by a first-order S.D.E. model, the next step is then to seek the physical implication of this proposed model (3.13). This was done indirectly through the establishment of the relationship between the model parameters and the watershed physiographic characteristics.

Physical implication of the proposed stochastic model

The proposed model (3.13) may be viewed as the result of linearization of some complicated time-variant non-linear monthly runoff generating system, by introducing the perturbation term y(t) defined by eq. (3.41). Therefore the solution obtained from this linearized equation should provide a certain degree of physical understanding of the original system provided the perturbation, or the deviation from the mean is not large. In other words, the proposed linear model can be considered as physically valid in the average sense.

To reconstruct the runoff series, the transformation (3.41) may be incorporated into the system (3.15) to yield the following system representation

$$q(t) = m(t) + \frac{s(t)}{K} \int_{-\infty}^{t} e^{-(t-u)/K} x(u)du. \qquad (3.84)$$

From eq. (3.84), it can be seen that the model parameters to be determined from the data are m(t), s(t) and K. The input x(t) is generated by a purely random process with zero mean and variance defined by eq. (3.17a). Monthly means and standard deviations, i.e., m(t) and s(t) are computed either by the non-parametric method, i.e., computing the long-term monthly means and standard deviations, or by the parametric method, i.e., determining the significant harmonics in monthly means and variances respectively (Yevjevich, 1972).

An attempt has beem made to correlate monthly means and standard deviations with watershed physiographic characteristics (Shawinigan, 1970). Therefore emphasis here would be to establish the relationship between the coefficient K and the watershed physiographic characteristics. By taking analogy to Nash's linear reservoir model (Nash, 1960), the coefficient K is presumed to bear the effect of watershed storage. It would also be of interest to see what are the significant physiographic parameters governing the behaviour of the input and the output of the system. This was done by investigating the effect of physiographic characteristics on the 'runoff residual variance' defined by the variance of the stochastic component from eq. (3.40) and the 'gross noise variance' defined as the product of the runoff residual variance and the unit noise variance given by eq. (3.17a). It is thought that the gross noise variance should be of the same order of magnitude as the variance of the precipitation (Matalas, 1966).

The correlation between model parameters and watershed physiographic characteristics was carried out through the Varimax factor analysis and the multiple regression analysis (which are described in detail in Appendix C). In a first attempt, all the stations were included in a single set of the above analyses and the results are given in Table 3.5. It was found, however, that the residuals from the various regression equations indicated obvious geographic pattern in their distributions. This is especially true for the estimation of the coefficient K. Overestimation (negative residuals) was found common in most of the prairie stations; whereas underestimation (positive residuals) in most mountainous areas (Figure 3.7). It was then decided to carry out the correlation regression analysis separately for two regions, namely, Pacific region (mountainous region) and Prairie region (Figure 3.8). The results are listed for the two regions in Tables 3.6 and 3.7 respectively. By comparing Table 3.5b with Tables 3.6b and 3.7b, it can be seen that the results of the estimation of K were improved to a significant extent* in the Pacific region but with no obvious improvement in the Prairie region. For the other two parameters, runoff residual variance (Tables 3.5c, 3.6c and 3.7c) and gross noise variance (Tables 3.5d, 3.6d and 3.7d), there is no significant improvement for both regions. However, there is no obvious geographic pattern in the distribution of the residuals from various regression equations (Figure 3.8). This preliminary delineation of model homogenous regions agrees with the statistical regions reported in the hydrometric network planning study for western and northern Canada (Shawinigan, 1970).

All the regression results are significant at the 95% confidence level.** There is no significant difference between the results

*Based on Fisher's criterion of z-transformation.

**Based on Fisher's F-test of difference in variances.

TABLE 3.5a

Summary on Regional Regression Analysis

(All Stations)

Criterion Variables and Predictor Variables			Original Form		Log Transformed		Variable No. in Regression		
			Mean	Std.Dev.	Mean	Std.Dev.	With 25 Physio. Factors	With 9 Physio. Factors	With 5 Physio. Factors
1st Coeff. of 1st-Order S.D.E.			2.80	2.38	0.35	0.27	X(1)	X(1)	X(1)
Residual Variance		in^2	0.25	0.30	-0.95	0.68	X(1)	X(1)	X(1)
Gross Noise Variance		in^2	0.84	0.78	-0.30	0.53	X(1)	X(1)	X(1)
Latitude Index			57.31	29.89	1.69	0.26	X(2)		
Average Elevation		ft	3773	1691	3.52	0.26	X(3)	X(2)	X(2)
Average Slope		%	2.93	2.34	0.01	0.01	X(4)	X(3)	X(3)
Slope Azimuth Index	N		3.29	1.22	1.48	0.21	X(5)	X(4)	
	NW		2.86	0.95	1.43	0.17	X(6)	X(5)	
	W		2.52	1.15	1.36	0.21	X(7)	X(6)	
	SW		2.33	1.28	1.30	0.25	X(8)	X(7)	
Distance to Sea	N	km	1771	295	3.24	0.08	X(9)		
	NW		2663	613	3.41	0.13	X(10)		
	W		1121	519	3.00	0.21	X(11)		
	SW		1064	669	2.95	0.27	X(12)		
Area Lake and Swamp		%	10.71	11.59	0.04	0.04	X(13)	X(8)	X(4)
Forest			71.60	15.49	0.23	0.04	X(14)	X(9)	X(5)
Barrier Height	N	ft	1542	1368	2.95	0.53	X(15)		
	NW		2769	1198	3.39	0.24	X(16)		
	W		3617	1490	3.52	0.21	X(17)		
	SW		3541	1960	3.47	0.29	X(18)		
Shield Effect	N	ft	22210	15930	4.23	0.33	X(19)		
	NW		107300	62810	4.89	0.44	X(20)		
	W		44980	15820	4.62	0.17	X(21)		
	SW		82010	54770	4.79	0.37	X(22)		
Regional Slope	NE	%	3.83	8.14	0.0005	0.001	X(23)		
	E		5.14	12.40	0.0007	0.002	X(24)		
	SE		3.19	7.21	0.0004	0.001	X(25)		
Drainage Area		sq.mi.	9449	16280	3.48	0.70	X(26)	X(10)	X(6)

TABLE 3.5b

Summary on Regional Regression Analysis

(All Stations)

Criterion Variable: Runoff Residual Variance

	Original Form					Log Transformed				
	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic
with 25 Physio. Factors	X(2)	0.39	0.023	0.013	1.77	X(2)	0.37	0.107	0.213	0.50
	X(10)	-0.31	-0.0006	0.0006	1.00	X(12)	0.35	0.043	0.197	0.22
	X(16)	-0.33	-0.0003	0.0003	1.00	X(13)	0.47	1.959	1.253	1.56
	X(18)	0.34	-0.00008	0.0002	0.40	X(16)	-0.39	-0.335	0.160	2.09
	X(22)	0.37	0.00002	0.000006	3.33	X(18)	0.38	0.030	0.192	0.16
	X(26)	0.36	0.00004	0.00002	2.00					
	Intercept Constant		2.428	1.929	1.26	Intercept Constant		0.988	1.013	0.98
	R = 0.659	R^2 = 0.434					R = 0.578	R^2 = 0.334		
	F = 4.472	ν_1 = 6	ν_2 = 35				F = 3.613	ν_1 = 5	ν_2 = 36	
with 9 Physio. Factors	X(3)	-0.47	-0.329	0.193	1.70	X(3)	-0.57	-13.500	5.117	2.64
	X(8)	0.41	0.024	0.039	0.62	X(8)	0.47	0.635	1.153	0.55
	X(10)	0.36	0.00004	0.00002	2.00					
	Intercept Constant		3.158	0.985	3.21	Intercept Constant		0.491	0.110	4.46
	R = 0.542	R^2 = 0.294					R = 0.577	R^2 = 0.333		
	F = 5.274	ν_1 = 3	ν_2 = 38				F = 9.735	ν_1 = 2	ν_2 = 39	
with 5 Physio. Factors	X(3)	-0.47	-0.329	0.193	1.70	X(3)	-0.57	-13.50	5.117	2.64
	X(4)	0.41	0.024	0.039	0.62	X(4)	0.47	0.635	1.153	0.55
	X(6)	0.36	0.00004	0.00002	2.00					
	Intercept Constant		3.158	0.985	3.21	Intercept Constant		0.491	0.110	4.46
	R = 0.542	R^2 = 0.294					R = 0.577	R^2 = 0.333		
	F = 5.274	ν_1 = 3	ν_2 = 38				F = 9.735	ν_1 = 2	ν_2 = 38	

$$^* \text{ t Statistic} = \frac{|\text{Regression Coeff.}|}{\text{Std.Dev.Regression Coeff.}}$$

TABLE 3.5c

Summary on Regional Regression Analysis

(All Stations)

Criterion Variable: Runoff Residual Variance

	Original Form					Log Transformed				
	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic
with 25 Physio. Factors	X(2)	-0.41	-0.002	0.001	2.00	X(2)	-0.51	0.382	0.391	0.98
	X(5)	-0.36	-0.025	0.025	1.00	X(8)	0.40	0.636	0.231	2.75
	X(7)	0.35	0.064	0.027	2.37	X(10)	0.31	1.634	0.504	3.24
	X(10)	0.35	0.00003	0.00005	0.60	X(12)	-0.49	-0.647	0.303	2.14
	X(15)	0.33	-0.0001	0.00003	3.33	X(13)	-0.70	-3.455	2.160	1.60
	X(18)	-0.60	-0.0001	0.00002	5.00	X(18)	-0.67	-0.721	0.309	2.33
	X(24)	-0.36	-0.011	0.003	3.67	X(26)	-0.48	-0.272	0.103	2.64
	X(26)	-0.30	-0.000002	0.000002	1.00					
	Intercept Constant		0.811	0.244	3.32	Intercept Constant		-2.498	1.777	1.41
	R = 0.843 F = 10.155	R^2 = 0.711 ν_1 = 8	ν_2 = 33				R = 0.854 F = 13.108	R^2 = 0.730 ν_1 = 7	ν_2 = 34	
with 9 Physio. Factors	X(2)	0.62	0.0001	0.00002	5.00	X(2)	0.72	1.675	0.221	7.58
	X(4)	-0.36	-0.077	0.026	2.96	X(4)	-0.32	-0.399	0.351	1.14
	X(6)	0.35	0.044	0.028	1.57	X(7)	0.40	0.730	0.298	2.45
	X(10)	-0.31	-0.000001	0.000002	0.50	X(10)	-0.48	-0.269	0.081	3.32
	Intercept Constant		0.042	0.134	0.31	Intercept Constant		-6.263	1.164	5.38
	R = 0.728 F = 10.429	R^2 = 0.530 ν_1 = 4	ν_2 = 37				R = 0.857 F = 25.579	R^2 = 0.734 ν_1 = 4	ν_2 = 37	
with 5 Physio. Factors	X(3)	0.65	0.069	0.021	3.29	X(2)	0.72	1.189	0.385	3.09
	X(4)	-0.53	-0.002	0.004	0.50	X(4)	-0.70	-4.099	2.311	1.77
	X(6)	-0.31	-0.000003	0.000002	1.50	X(6)	-0.48	-0.253	0.097	2.61
	Intercept Constant		0.101	0.108	0.94	Intercept Constant		-4.076	1.475	2.76
	R = 0.728 F = 10.429	R^2 = 0.530 ν_1 = 4	ν_2 = 37				R = 0.857 F = 25.579	R^2 = 0.734 ν_1 = 4	ν_2 = 37	

* $$t \text{ Statistic} = \frac{|\text{Regression Coeff.}|}{\text{Std.Dev.Regression Coeff.}}$$

TABLE 3.5d

Summary on Regional Regression Analysis

(All Stations)

Criterion Variable: Runoff Residual Variance

	Original Form					Log Transformed				
	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic
with 25 Physio. Factors	X(2)	-0.46	-0.0004	0.004	0.10	X(7)	0.31	-0.238	0.325	0.73
	X(3)	0.61	0.0003	0.00007	4.29	X(8)	0.37	0.477	0.223	2.14
	X(7)	0.43	-0.023	0.099	0.23	X(9)	0.52	-0.588	0.989	0.59
	X(8)	0.49	0.223	0.079	2.82	X(12)	-0.45	0.214	0.304	0.70
	X(10)	0.34	0.00005	0.0002	0.25	X(14)	0.32	3.208	1.260	2.55
	X(14)	0.34	0.019	0.006	3.17	X(17)	-0.70	-1.111	0.335	3.32
	X(19)	0.34	-0.000009	0.000007	1.29	X(20)	0.57	0.483	0.186	2.60
	X(26)	-0.35	-0.000003	0.000005	0.60	X(26)	-0.47	-0.184	0.083	2.22
	Intercept Constant		-1.931	0.699	2.76	Intercept Constant		2.125	3.470	0.61
	R = 0.808	R^2 = 0.652					R = 0.830	R^2 = 0.689		
	F = 7.742	ν_1 = 8	ν_2 = 33				F = 9.121	ν_1 = 8	ν_2 = 33	
with 9 Physio. Factors	X(2)	0.61	0.0003	0.00005	6.00	X(2)	0.62	1.297	0.216	6.00
	X(6)	0.43	-0.073	0.098	0.74	X(6)	0.31	-0.575	0.314	1.83
	X(7)	0.49	0.237	0.081	2.93	X(7)	0.37	0.769	0.220	3.50
	X(9)	0.34	0.017	0.005	3.40	X(9)	0.32	4.705	1.309	3.59
	X(10)	-0.45	-0.000002	0.000005	0.40	X(10)	-0.47	-0.188	0.071	2.65
	Intercept Constant		-1.715	0.441	3.89	Intercept Constant		-5.521	0.898	6.15
	R = 0.792	R^2 = 0.627					R = 0.822	R^2 = 0.676		
	F = 12.122	ν_1 = 5	ν_2 = 36				F = 15.032	ν_1 = 5	ν_2 = 36	
with 5 Physio. Factors	X(3)	0.61	0.132	0.053	2.49	X(3)	0.66	24.680	7.443	3.32
	X(4)	-0.55	-0.014	0.011	1.27	X(4)	-0.65	-2.321	1.768	1.31
	X(5)	0.34	0.014	0.005	2.80	X(5)	0.32	3.365	1.188	2.83
	X(6)	-0.35	-0.000008	0.000005	1.60	X(6)	-0.47	-0.248	0.074	3.35
	Intercept Constant		-0.292	0.476	0.61	Intercept Constant		-0.430	0.385	1.12
	R = 0.714	R^2 = 0.509					R = 0.812	R^2 = 0.660		
	F = 9.594	ν_1 = 4	ν_2 = 37				F = 17.927	ν_1 = 4	ν_2 = 37	

* $t \text{ Statistics} = \dfrac{|\text{Regression Coeff.}|}{\text{Std.Dev.Regression Coeff.}}$

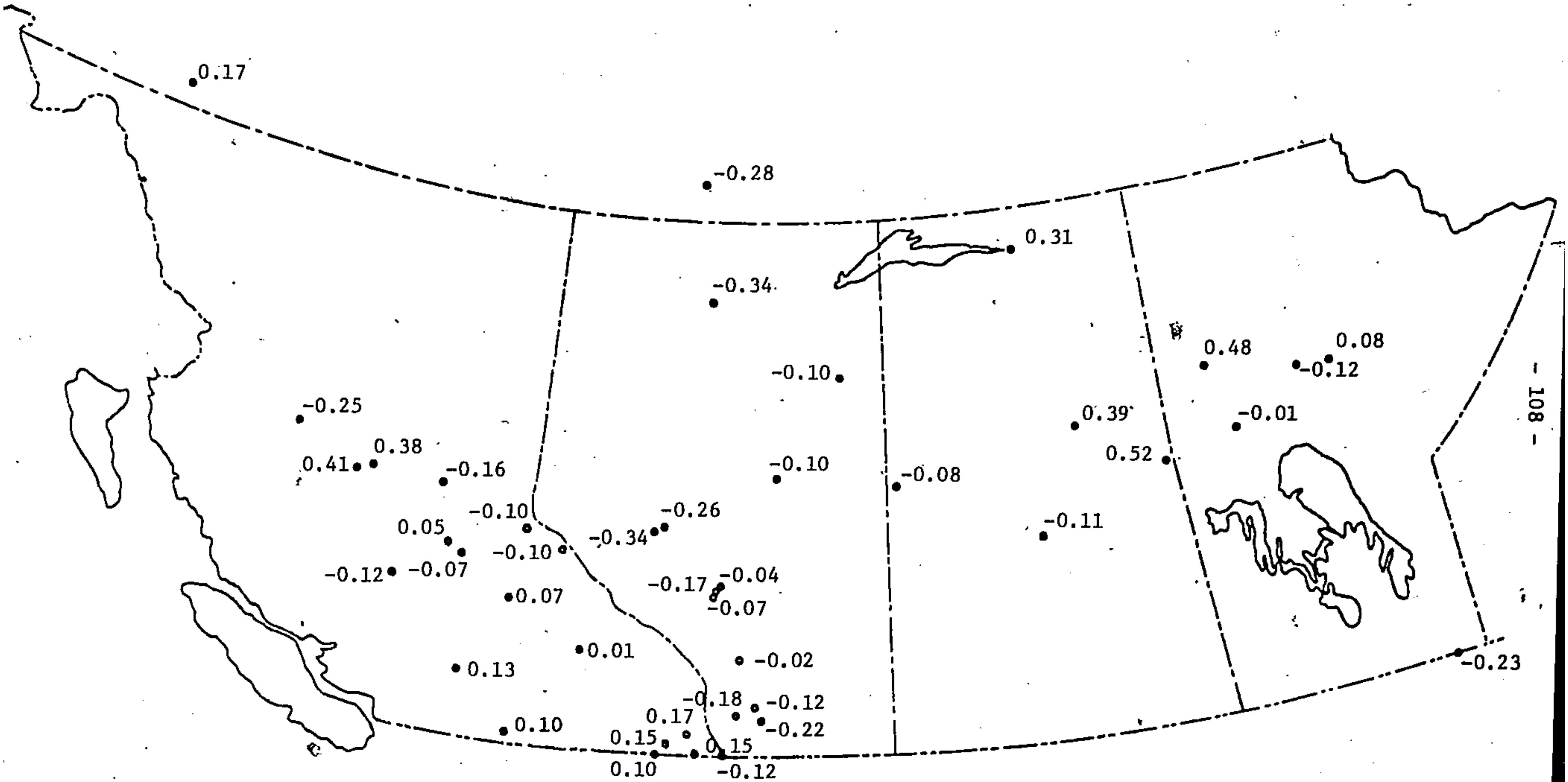


Figure.3.7 Geographic Distribution of Residuals in Estimation of Coefficient K before Statistical Regionalization

TABLE 3.6a

Summary on Regional Regression Analysis

(Pacific Region)

Criterion Variables and Predictor Variables			Original Form Mean	Original Form Std.Dev.	Log Transformed Mean	Log Transformed Std.Dev.	Variable No. in Regression: With 25 Physio. Factors	With 9 Physio. Factors	With 5 Physio. Factors
1st Coeff. of 1st-Order S.D.E.			2.16	1.48	0.26	0.23	X(1)	X(1)	X(1)
Residual Variance		in^2	0.44	0.35	-0.50	0.39	X(1)	X(1)	X(1)
Gross Noise Variance		in^2	1.39	0.84	0.07	0.28	X(1)	X(1)	X(1)
Latitude Index			48.26	29.48	1.61	0.27	X(2)		
Average Elevation		ft	4791	1037	3.67	0.10	X(3)	X(2)	X(2)
Average Slope		%	4.99	1.67	0.02	0.007	X(4)	X(3)	X(3)
Slope Azimuth Index	N		3.00	1.45	1.42	0.26	X(5)	X(4)	
	NW		3.16	1.07	1.47	0.18	X(6)	X(5)	
	W		3.32	1.16	1.49	0.18	X(7)	X(6)	
	SW		3.16	1.26	1.46	0.21	X(8)	X(7)	
Distance to Sea	N	km	1899	245	3.27	0.07	X(9)		
	NW		2633	755	3.40	0.17	X(10)		
	W		721	232	2.83	0.16	X(11)		
	SW		588	245	2.73	0.18	X(12)		
Area Lake and Swamp		%	3.10	2.90	0.01	0.01	X(13)	X(8)	X(4)
Forest			77.74	14.78	0.25	0.04	X(14)	X(9)	X(9)
Barrier Height	N	ft	2851	929	3.43	0.15	X(15)		
	NW		2648	766	3.41	0.13	X(16)		
	W		2413	967	3.35	0.18	X(17)		
	SW		2052	960	3.25	0.25	X(18)		
Shield Effect	N	ft	36550	12590	4.54	0.14	X(19)		
	NW		150200	41750	5.15	0.17	X(20)		
	W		37680	17720	4.53	0.20	X(21)		
	SW		50400	40760	4.56	0.37	X(22)		
Regional Slope	NE	%	2.26	10.13	0.0003	0.001	X(23)		
	W		1.32	15.32	0.0002	0.002	X(24)		
	SE		1.05	84.95	0.0001	0.001	X(25)		
Drainage Area		sq.mi.	4827	5630	3.39	0.56	X(26)	X(10)	X(6)

TABLE 3.6b

Summary on Regional Regression Analysis

(Pacific Region)

Criterion Variable: 1st Coeff. of S.D.E.

	Original Form					Log Transformed				
	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic
with 25 Physio. Factors	X(4)	-0.70	-0.299	0.154	1.94	X(4)	-0.72	-8.727	7.321	1.19
	X(13)	0.79	0.241	0.106	2.27	X(6)	0.47	0.094	0.211	0.45
	X(15)	0.55	0.0004	0.0002	2.00	X(13)	0.69	8.154	4.419	1.85
	X(19)	0.58	-0.000008	0.00002	0.40	X(14)	0.53	1.778	1.023	1.74
						X(19)	0.51	0.011	0.310	0.04
	Intercept Constant		2.016	1.180	1.71	Intercept Constant		-0.287	1.407	0.20
	R = 0.843, F = 8.581	R^2 = 0.710, ν_1 = 4	ν_2 = 14			R = 0.811, F = 4.989	R^2 = 0.657, ν_1 = 5	ν_2 = 13		
with 9 Physio. Factors	X(3)	-0.70	-0.262	0.162	1.62	X(3)	-0.72	-8.717	7.316	1.19
	X(8)	0.79	0.299	0.093	3.22	X(5)	0.47	0.094	0.211	0.45
						X(8)	0.69	8.240	3.687	2.23
						X(9)	0.53	1.782	1.016	1.75
	Intercept Constant		2.536	1.048	2.42	Intercept Constant		-0.240	0.462	0.52
	R = 0.817, F = 16.014	R^2 = 0.667, ν_1 = 2	ν_2 = 16			R = 0.811, F = 6.716	R^2 = 0.657, ν_1 = 4	ν_2 = 14		
with 5 Physio Factors	X(3)	-0.70	-0.262	0.162	1.62	X(3)	-0.72	-9.383	7.201	1.30
	X(4)	0.78	0.299	0.093	3.22	X(4)	0.69	8.430	3.682	2.29
						X(5)	0.53	1.883	0.996	1.89
	Intercept Constant		2.536	1.048	2.42	Intercept Constant		-0.115	0.371	0.31
	R = 0.817, F = 16.014	R^2 = 0.667, ν_1 = 2	ν_2 = 16			R = 0.809, F = 9.441	R^2 = 0.654, ν_1 = 3	ν_2 = 15		

* t Statistic = $\dfrac{|\text{Regression Coeff.}|}{\text{Std.Dev.Regression Coeff.}}$

TABLE 3.6c

Summary on Regional Regression Analysis
(Pacific Region)

Criterion Variable: Runoff Residual Variance

	Original Form					Log Transformed				
	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic
with 25 Physio. Factors	X(3)	0.51	-0.0002	0.00006	3.33	X(4)	0.57	11.649	10.11	1.15
	X(12)	0.66	0.001	0.0002	5.00	X(11)	0.51	0.662	0.401	1.65
	X(15)	-0.54	-0.0003	0.00006	5.00	X(15)	-0.47	-0.878	0.373	2.35
	X(17)	-0.58	-0.0001	0.00005	2.00	X(22)	0.60	0.372	0.186	2.00
	Intercept Constant		1.741	0.434	4.01	Intercept Constant		-1.306	1.807	0.72
	R = 0.889, F = 13.251	R^2 = 0.791, ν_1 = 4	ν_2 = 14			R = 0.785, F = 5.619	R^2 = 0.616, ν_1 = 4	ν_2 = 14		
with 9 Physio Factors	X(2)	0.51	-0.00001	0.00009	0.11	X(3)	0.57	-1.479	9.726	0.15
	X(8)	-0.66	-0.082	0.034	2.41	X(8)	-0.84	-28.24	5.553	5.09
	Intercept Constant		0.748	0.540	1.39	Intercept Constant		-0.097	0.265	0.37
	R = 0657, F = 6.083	R^2 = 0.432, ν_1 = 2	ν_2 = 16			R = 0.845, F = 19.904	R^2 = 0.713, ν_1 = 2	ν_2 = 16		
with 5 Physio Factors	X(2)	0.51	-0.00001	0.00009	0.11	X(3)	0.57	-1.479	9.726	0.15
	X(4)	-0.66	-0.082	0.034	2.41	X(4)	-0.84	-28.24	5.553	5.09
	Intercept Constant		0.748	0.540	1.39	Intercept Constant		-0.097	0.265	0.37
	R = 0.657, F = 6.083	R^2 = 0.432, ν_1 = 2	ν_2 = 16			R = 0.845, F = 19.904	R^2 = 0.713, ν_1 = 2	ν_2 = 16		

* $t \text{ Statistic} = \dfrac{|\text{Regression Coeff.}|}{\text{Std.Dev.Regression Coeff.}}$

TABLE 3.6d

Summary on Regional Regression Analysis

(Pacific Region)

Criterion Variable: Gross Noise Variance

	Original Form					Log Transformed				
	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic
with 25 Physio. Factors	X(10)	0.47	0.0003	0.0002	1.50	X(10)	0.57	0.880	0.229	3.84
	X(17)	-0.53	-0.0003	0.0001	3.00	X(17)	-0.51	-0.641	0.226	2.84
	X(18)	-0.46	-0.0002	0.0001	2.00	X(18)	-0.50	-0.258	0.158	1.63
	X(19)	-0.59	-0.00003	0.00001	3.00	X(19)	-0.63	-0.477	0.278	1.72
	Intercept Constant		2.789	0.838	3.33	Intercept Constant		2.232	1.682	1.33
	$R = 0.792$, $F = 5.873$	$R^2 = 0.627$, $\nu_1 = 4$	$\nu_2 = 14$			$R = 0.875$, $F = 11.484$	$R^2 = 0.766$, $\nu_1 = 4$	$\nu_2 = 14$		
with 9 Physio Factors	X(8)	-0.60	-0.172	0.053	3.25	X(8)	-0.61	-14.25	4.274	3.33
	Intercept Constant		1.925	0.226	8.52	Intercept Constant		0.254	0.076	3.34
	$R = 0.595$, $F = 9.328$	$R^2 = 0.354$, $\nu_1 = 1$	$\nu_2 = 17$			$R = 0.608$, $F = 9.949$	$R^2 = 0.369$, $\nu_1 = 1$	$\nu_2 = 17$		
with 5 Physio Factors	X(4)	-0.60	-0.172	0.053	3.25	X(4)	-0.61	-14.25	4.274	3.33
	Intercept Constant		1.925	0.226	8.52	Intercept Constant		0.254	0.076	3.34
	$R = 0.595$, $F = 9.328$	$R^2 = 0.354$, $\nu_1 = 1$	$\nu_2 = 17$			$R = 0.608$, $F = 9.949$	$R^2 = 0.369$, $\nu_1 = 1$	$\nu_2 = 17$		

* t Statistic = $\frac{|\text{Regression Coeff.}|}{\text{Std.Dev.Regression Coeff.}}$

TABLE 3.7a

Summary on Regional Regression Analysis

(Prairie Region)

Criterion Variables and Predictor Variables			Original Form		Log Transformed		Variable No. in Regression		
			Mean	Std.Dev.	Mean	Std.Dev.	With 25 Physio. Factors	With 9 Physio. Factors	With 5 Physio. Factors
1st Coeff. of 1st-Order S.D.E.			3.34	2.84	0.42	0.28	X(1)	X(1)	X(1)
Residual Variance		in^2	0.10	0.09	-1.32	0.64	X(1)	X(1)	X(1)
Gross Noise Variance		in^2	0.39	0.29	-0.60	0.50	X(1)	X(1)	X(1)
Latitude Index			64.78	28.72	1.76	0.24	X(2)		
Average Elevation		ft	2931	1678	3.39	0.28	X(3)	X(2)	X(2)
Average Slope		%	1.22	1.10	0.005	0.005	X(4)	X(3)	X(3)
Slope Azimuth Index	N		3.52	0.95	1.53	0.15	X(5)	X(4)	
	NW		2.61	0.78	1.39	0.16	X(6)	X(5)	
	W		1.87	0.63	1.25	0.16	X(7)	X(6)	
	SW		1.65	0.83	1.17	0.20	X(8)	X(7)	
Distance to Sea	N	km	1666	296	3.22	0.08	X(9)		
	NW		2687	482	3.42	0.08	X(10)		
	W		1452	453	3.14	0.13	X(11)		
	SW		1457	654	3.12	0.19	X(12)		
Area Lake and Swamp		%	17.00	12.34	0.07	0.05	X(13)	X(8)	X(4)
Forest			66.52	14.46	0.22	0.04	X(14)	X(9)	X(5)
Barrier Height	N	ft	460	284	2.55	0.37	X(15)		
	NW		2869	1475	3.38	0.31	X(16)		
	W		4612	1041	3.65	0.11	X(17)		
	SW		4771	1705	3.65	0.18	X(18)		
Shield Effect	N	ft	10370	4334	3.98	0.20	X(19)		
	NW		71900	55030	4.66	0.47	X(20)		
	W		51000	11210	4.70	0.09	X(21)		
	SW		108100	51540	4.97	0.26	X(22)		
Regional Slope	NE	%	5.13	59.72	0.0007	0.0008	X(23)		
	E		8.30	84.50	0.001	0.001	X(24)		
	SE		4.95	55.39	0.0007	0.0007	X(25)		
Drainage Area		sq.mi.	13270	20850	3.55	0.81	X(26)	X(10)	X(6)

TABLE 3.7b

Summary on Regional Regression Analysis

(Prairie Region)

Criterion Variable: 1st Coeff. of S.D.E.

	Original Form					Log Transformed				
	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic
with 25 Physio. Factors	X(4)	-0.45	0.039	0.711	0.05	X(2)	0.48	0.497	0.213	2.33
	X(9)	-0.45	-0.003	0.002	1.50	X(17)	0.44	0.273	0.413	0.66
	X(22)	0.49	0.00003	0.000009	3.33	X(21)	-0.49	-0.172	0.620	0.28
	X(24)	-0.46	-0.045	0.102	0.44	X(22)	0.49	0.621	0.141	4.40
						X(25)	-0.53	-63.89	72.68	0.88
	Intercept Constant		6.625	3.237	2.05	Intercept Constant		-3.687	3.510	1.05
	$R = 0.669$	$R^2 = 0.448$				$R = 0.804$	$R^2 = 0.646$			
	$F = 3.651$	$\nu_1 = 4$	$\nu_2 = 18$			$F = 6.203$	$\nu_1 = 5$	$\nu_2 = 17$		
with 9 Physio. Factors	X(3)	-0.45	-1.147	0.482	2.38	X(3)	-0.53	-26.53	10.11	2.62
						X(10)	0.42	0.105	0.059	1.78
	Intercept Constant		4.739	0.793	5.98	Intercept Constant		0.186	0.233	0.80
	$R = 0.445$	$R^2 = 0.198$				$R = 0.606$	$R^2 = 0.367$			
	$F = 5.177$	$\nu_1 = 1$	$\nu_2 = 21$			$F = 5.789$	$\nu_1 = 2$	$\nu_2 = 20$		
with 5 Physio. Factors	X(3)	-0.45	-1.147	0.482	2.38	X(3)	-0.53	-26.53	10.11	2.62
						X(6)	0.42	0.105	0.059	1.78
	Intercept Constant		4.739	0.793	5.98	Intercept Constant		0.186	0.233	0.80
	$R = 0.445$	$R^2 = 0.198$				$R = 0.606$	$R^2 = 0.367$			
	$F = 5.177$	$\nu_1 = 1$	$\nu_2 = 21$			$F = 5.789$	$\nu_1 = 2$	$\nu_2 = 20$		

$$^* \; t \text{ Statistic} = \frac{|\text{Regression Coeff.}|}{\text{Std.Dev.Regression Coeff.}}$$

TABLE 3.7c

Summary on Regional Regression Analysis

(Prairie Region)

Criterion Variable: Runoff Residual Variance

	Original Form					Log Transformed				
	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t* Statistic	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t* Statistic
with 25 Physio. Factors	X(2)	-0.62	-0.0006	0.0005	1.20	X(2)	-0.53	-0.920	0.629	1.46
	X(12)	-0.51	-0.00003	0.00002	1.50	X(12)	-0.48	-0.821	0.650	1.26
	X(17)	-0.74	-0.00004	0.00001	4.00	X(13)	-0.49	0.204	3.422	0.06
	X(21)	0.50	0.0000006	0.000001	0.60	X(17)	-0.62	-1.966	1.239	1.59
	X(26)	-0.49	-0.000001	0.0000005	2.00	X(21)	0.47	0.236	1.676	0.14
	Intercept Constant		0.362	0.095	3.81	Intercept Constant		8.915	9.405	0.95
	R = 0.849, F = 8.773	R^2 = 0.721, ν_1 = 5		ν_2 = 17		R = 0.708, F = 3.427	R^2 = 0.502, ν_1 = 5		ν_2 = 17	
with 9 Physio. Factors	X(3)	0.70	0.050	0.015	3.33	X(2)	0.59	1.034	0.413	2.50
	X(8)	-0.50	-0.0004	0.001	0.40	X(8)	-0.49	-0.357	2.604	0.14
	X(10)	-0.49	-0.000002	0.0000006	3.33	X(10)	-0.61	-0.379	0.119	3.18
	Intercept Constant		0.065	0.041	1.59	Intercept Constant		-3.455	1.568	2.20
	R = 0.792, F = 10.675	R^2 = 0.628, ν_1 = 3	ν_2 = 19			R = 0.752, F = 8.247	R^2 = 0.566, ν_1 = 3		ν_2 = 19	
with 5 Physio. Factors	X(3)	0.70	0.050	0.015	3.33	X(2)	0.59	1.034	0.413	2.50
	X(4)	-0.50	-0.0004	0.001	0.40	X(4)	-0.49	-0.357	2.604	0.14
	X(6)	-0.49	-0.000002	0.0000006	3.33	X(6)	-0.61	-0.379	0.119	3.18
	Intercept Constant		0.065	0.041	1.59	Intercept Constant		-3.455	1.568	2.20
	R = 0.792, F = 10.675	R^2 = 0.628, ν_1 = 3	ν_2 = 19			R = 0.752, F = 8.247	R^2 = 0.566, ν_1 = 3		ν_2 = 19	

$$* \ t \text{ Statistic} = \frac{|\text{Regression Coeff.}|}{\text{Std.Dev.Regression Coeff.}}$$

TABLE 3.7d

Summary on Regional Regression Analysis

(Prairie Region)

Criterion Variable: Gross Noise Variance

	Original Form					Log Transformed				
	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic	Significant Predictor Variables	Corr. with Criterion Variable	Regression Coeff.	Std. Dev. Regression Coeff.	t^* Statistic
with 25 Physio. Factors	X(2)	-0.53	-0.0007	0.002	0.35	X(13)	-0.43	0.147	2.194	0.07
	X(18)	-0.52	-0.000006	0.00003	0.20	X(17)	-0.55	-1.644	0.910	1.81
	X(19)	0.53	0.00001	0.00001	1.00	X(25)	0.59	124.9	168.4	0.74
	X(25)	0.71	0.021	0.013	1.62	X(26)	-0.55	-0.221	0.114	1.94
	X(26)	-0.52	-0.000004	0.000002	2.00					
	Intercept Constant		0.303	0.278	1.09	Intercept Constant		6.097	3.434	1.78
	$R = 0.762$	$R^2 = 0.581$				$R = 0.700$	$R^2 = 0.490$			
	$F = 4.709$	$\nu_1 = 5$	$\nu_2 = 17$			$F = 4.332$	$\nu_1 = 4$	$\nu_2 = 18$		
with 9 Physio Factors	X(3)	0.58	0.109	0.052	2.10	X(2)	0.43	0.429	0.381	1.13
	X(8)	-0.44	-0.003	0.005	0.60	X(8)	-0.43	-1.125	2.403	0.47
	X(10)	-0.52	-0.000006	0.000002	3.00	X(10)	-0.55	-0.275	0.110	2.50
	Intercept Constant		0.380	0.142	2.68	Intercept Constant		-1.006	1.447	0.70
	$R = 0.716$	$R^2 = 0.512$				$R = 0.623$	$R^2 = 0.388$			
	$F = 6.652$	$\nu_1 = 3$	$\nu_2 = 19$			$F = 4.020$	$\nu_1 = 3$	$\nu_2 = 19$		
with 5 Physio. Factors	X(3)	0.58	0.109	0.052	2.10	X(2)	0.43	0.429	0.381	1.13
	X(4)	-0.44	-0.003	0.005	0.60	X(4)	-0.43	-1.125	2.403	0.47
	X(6)	-0.52	-0.000006	0.000002	3.00	X(6)	-0.55	-0.275	0.110	2.50
	Intercept Constant		0.380	0.142	2.68	Intercept Constant		-1.006	1.447	0.70
	$R = 0.716$	$R^2 = 0.512$				$R = 0.623$	$R^2 = 0.388$			
	$F = 6.652$	$\nu_1 = 3$	$\nu_2 = 19$			$F = 4.020$	$\nu_1 = 3$	$\nu_2 = 19$		

* t Statistic = $\dfrac{|\text{Regression Coeff.}|}{\text{Std.Dev.Regression Coeff.}}$

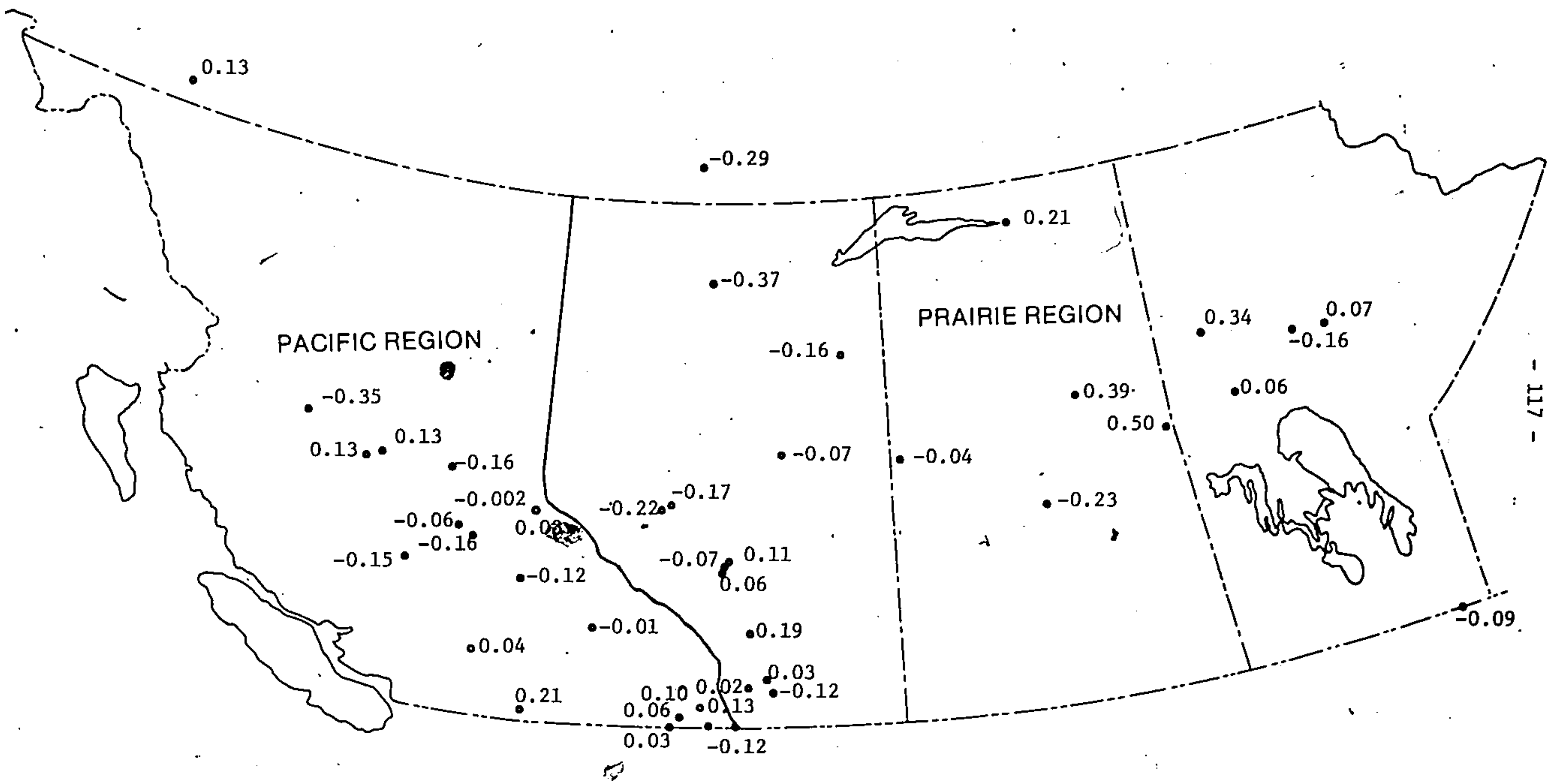


Figure 3.8 Geographic Distribution of Residuals in Estimation of Coefficient K after Statistical Regionalization

from linear and non-linear analyses for all the cases.* In the Pacific region non-linear regression gives slightly better results in most cases (Table 3.6) whereas in the Prairie region linear regression seems to be more desirable except for the estimation of the coefficient K (Table 3.7).

In carrying out Varimax factor analysis, caution must be exercised in selecting predictor variables for the analysis. Taking all possible predictor variables for the analysis may still introduce spurious correlation and the resulting multiple regression relationship may be misleading. Physical consideration must be given to the predictor variables to be included for the analysis with respect to their possible influence to the criterion variable. Therefore in the case of the coefficient K, it is thought that this parameter is an intrinsic characteristic of the system and should be related only to the local watershed physiographic characteristics such as average elevation, average slope, lake storage, etc. On the other hand, the gross noise variance pertains to the external disturbance which is governed dominantly by the regional physiographic characteristics such as distance to sea, barrier height and shield effect, etc. The runoff residual variance which results from the interaction between the system characteristic and the external disturbance will then be subject to the influences from both local and regional physiographic characteristics. For the purpose of comparison, however, analyses were carried out with all physiographic factors, with local factors including watershed aspect indices (slope azimuth indices)** and with local factors excluding watershed aspect indices.

*Based on Fisher's criterion of z-transformation.

**These watershed aspect indices were later found not significant in all cases.

Results from the regression analyses (Tables 3.6b and 3.7b) for the coefficient K indicate that there is no significant difference between the analyses involving all physiographic characteristics and the analyses involving only local physiographic characteristics. For the Pacific region, results are quite consistent (Table 3.6b). Significant physiographic factors governing the estimation of K are found to be average slope, area of lakes and swamps and area of forests. For the Prairie region, results are not as good and consistent (Table 3.7b). The significant physiographic parameters are average slope and drainage area. In general, all the above physiographic parameters are the dominant factors which govern the watershed storage effect. In other words, the coefficient K is a characteristic reflecting the storage effect of the watershed system. The difference between the regression models for the two regions lies in that a) the Prairie equation seems to be less accurate with respect to the percentage explained variance, i.e., R^2 and b) drainage area becomes a dominant factor in the Prairie region (Table 3.7b) instead of area of lakes and swamps and area of forests which are found significant for the Pacific region (Table 3.6b). This is generally true for the mountainous areas in the Pacific region that surface runoff and interflow are dominant components of total runoff and forest is the major land-use in the area which affects the runoff generating process (Ingledow, 1969). On the other hand, groundwater plays an important role in the Prairie hydrology (Meyboom, 1962) and agricultural activities are significant in the area (Shawinigan, 1972). None of the present available physiographic characteristics gives good indication concerning soil type, geology and vegetative cover, etc. This, of course, leads to the less desirable regional equation for the storage coefficient and also explains the drainage area being a dominant factor in the relationship because it is the only available factor which associates with the gross effect of surface storage (lakes and swamps), vegetation, soils and geology.

For the runoff residual variance, it is also observed that in both regions it is influenced more by the local watershed physiographic factors, i.e., area of lakes and swamps for the Pacific region (Table 3.6c) and average slope and drainage area for Prairie region (Table 3.7c). This confirms the common hydrological assumption of viewing any natural watershed as a 'damped' system that the high frequency variations of input disturbances are filtered out through its storage effect (Eagleson and Shack, 1965; Matalas, 1966). The inclusion of the drainage area as a dominant factor once more indicates the inadequacy of the regression model for the Prairie region.

Contrary to the above observations, it is found that in the case of the gross noise variance the regression equation based on all physiographic factors gives better results than those based on only local factors (Tables 3.6d and 3.7d). For the Pacific region the difference is significant. The dominant factors are regional physiographic parameters in all directions, i.e., N, NW, W and SW (Table 3.6d). It has been reported that this region is under the influence of various air masses of Arctic, Cool Pacific and Mild Pacific (Bryson, 1966). For the Prairie region, the difference is not significant but the general weather pattern influences are still picked up by the regression equation (Table 3.7d). Along with the significant regional parameters, drainage area again appears as a dominant factor in all equations for the Prairie region. This may be attributed to the high evapotranspiration in the Prairie region (McKay and Stickling, 1961; Meyboom, 1962) and also gives indication of the need for additional physiographic parameters in the area.

In conclusion, the first-order S.D.E. model which adequately describes the transformed monthly runoff series is physically equivalent to Nash's linear reservoir model. The coefficient K bears the effect of watershed storage. The input disturbances are controlled by the meteorological and physiographic factors. The output is then a damped response from the watershed, which is the result of the interaction between the watershed storage and the external disturbances from the atmosphere.

CHAPTER 4

MULTIVARIATE HYDROLOGIC TIME SERIES ANALYSIS

Introduction

In modelling the stochastic behaviour of a multivariate watershed hydrologic system, meteorological and runoff series are considered as the input and the output, respectively, of the system under study. Temperature and precipitation are the only meteorological series under consideration in this thesis because they are the most commonly available meteorological data in Canada. It is reasonable to assume that the effect of other hydrologic variables such as soil moisture content and evaporation are also taken care of by the dynamic nature of the system, i.e., they depend on the current and previous states of temperature and precipitation conditions.

A linear system can be described in the time domain by the impulse response function and in the frequency domain the frequency response function. These system response functions for a multivariate system can be determined through multiple cross correlation and multiple cross spectral analyses. Multiple cross spectral analysis bears analogy to the conventional multiple regression analysis except that for the former regression goes on separately at each frequency and that regression coefficients take complex values rather than real values, which enables one to learn more about the underlying relationship among various time series.

The other advantage of the analysis in the frequency domain is that the smoothing in the computation of spectrums reduces excessive sampling variations. Finally, the system is represented in the form of transfer function models. System transfer functions are constructed from the impulse response functions.

The theory and approach to be presented in the subsequent sections are mainly based on the works by Jenkins and Watts (1968) and Box and Jenkins (1970).

Theoretical Developments

Multivariate linear systems

A time-invariant multivariate linear system may be written in the matrix form

$$(x_{q+i}(t)) = \int_0^\infty (h_{q+i,j}(u))(x_j(t-u))du + (z_{q+i}(t)),$$

$$(i = 1,2,\ldots,r; \quad j = 1,2,\ldots,q) \qquad (4.1)$$

where $(x_{q+i}(t))$ is the column vector or 'r' output variables, $(x_j(t))$ is a column vector of 'q' input variables, $(z_{q+i}(t))$ is a column vector of 'r' residual variables and $(h_{q+i,j}(u))$ is an r x q impulse response matrix.

Thus for the $(q+i)^{th}$ output

$$x_{q+i}(t) = \int_0^\infty h_{q+i,1}(u)x_1(t-u)du + \ldots$$

$$+ \int_0^\infty h_{q+i,q}(u)x_q(t-u)du + z_{q+i}(t). \qquad (4.2)$$

Furthermore, all the above variables are assumed to have means equal to zero.

For simplicity, only a two-input and two-output system is considered, which is frequently met in practical applications (Jenkins and Watts, 1968). Thus in eq. (4.1), $q = 2$ and $r = 2$ and the equation becomes

$$x_3(t) = \int_0^\infty h_{31}(u)x_1(t-u)du + \int_0^\infty h_{32}(u)x_2(t-u)du + z_3(t),$$

and

$$x_4(t) = \int_0^\infty h_{41}(u)x_1(t-u)du + \int_0^\infty h_{42}(u)x_2(t-u)du + z_4(t). \qquad (4.3)$$

<u>Multiple cross spectral analysis</u>

In the time domain, the minimum mse (mean square error) estimates of the impulse response functions $h_{ij}(u)$ are obtained by separately minimizing the mean square errors

$$E \int_{-\infty}^{\infty} z_i^2(t)dt. \qquad (i=3,4)$$

This procedure leads to a set of Wiener-Hopf equations in the form

$$(\gamma_{ji}(u)) = \int_0^\infty (\gamma_{jk}(u-v))(h_{ik}(v))'dv, \qquad (i=3,4;\ j,k=1,2) \qquad (4.4)$$

where $\gamma_{ji}(u)$ is the lagged cross covariance between the output $X_i(t)$ and the input $X_j(t-u)$, $\gamma_{jk}(u)$ is the lagged covariance between the inputs $X_j(t-u)$ and $X_k(t)$ and $(.)'$ is the transpose of a matrix.

By the Fourier transformations of eq. (4.4), the estimation equation for the frequency response functions may be written

$$(S_{ji}(f)) = (S_{jk}(f))(H_{ik}(f))', \quad (i=3,4;\ j,k=1,2) \tag{4.5}$$

where $S_{ji}(f)$ is the cross spectrum between the output $X_j(t)$ and the input $X_i(t)$, $S_{jk}(f)$ is the cross spectrum between the inputs $X_j(t)$ and $X_k(t)$; and $H_{ik}(f)$ is the frequency response function corresponding to the impulse response function $h_{ik}(u)$.

In addition to estimating the frequency response matrix, it is necessary to characterize the properties of the residual variable $Z_i(t)$. This is done by determining the residual spectral matrix (S^z_{il}) whose elements are the cross spectrums between the processes $Z_i(t)$ and $Z_l(t)$. The residual cross covariance matrix is

$$(\gamma^z_{il}(u)) = (\gamma_{il}(u)) - \int_0^\infty (\gamma_{ij}(u-v))(h_{lj}(v))'dv, \quad (i,l=3,4;\ j=1,2) \tag{4.6}$$

where $\gamma_{il}(u)$ is the lagged covariance between the outputs $X_i(t-u)$ and $X_l(t)$. On transforming eq. (4.6), the cross spectrum matrix is given by

$$(S^z_{il}(f)) = (S_{il}(f)) - (S_{il}(f))(H_{lj}(f))', \quad (i,l=3,4;\ j=1,2) \tag{4.7}$$

where $S_{il}(f)$ is the cross spectrum between the outputs $X_i(t)$ and $X_l(t)$.

The residual spectrum $S^z_{ii}(f)$ may be written

$$S^z_{ii}(f) = S_{ii}(f)(1 - K^2_{i.12}(f)), \quad (i=3,4) \tag{4.8}$$

where

$$K^2_{i.12}(f) = (H_{i1}(f)S_{i1}(f) + H_{i2}(f)S_{i2}(f))/S_{ii}(f) \tag{4.9}$$

is called the squared multiple coherency spectrum of the i^{th} output process and the two input processes. The multiple coherency spectrum measures the proportion of the output spectrum which can be predicted from the inputs. Hence eq. (4.9) is the frequency domain analogue of the multiple correlation coefficient in the time domain.

As in multiple regression analysis, it is useful to be able to measure the cross spectrum between the output and one of the input processes after allowance is made for the effect of the other input processes. This leads to the partial cross spectrum, which is given by

$$S_{13.2}(f) = S_{13}(f) - \frac{S_{23}(f)S_{12}(f)}{S_{22}(f)} \quad , \tag{4.10}$$

and

$$\phi_{13.2}(f) = \tan^{-1}\left(\frac{Q_{12}C_{23}+C_{12}Q_{23}-Q_{13}S_{22}}{Q_{12}Q_{23}-C_{12}C_{23}+C_{13}S_{22}}\right). \tag{4.11}$$

It can also be shown that the squared partial coherency spectrum can be obtained from the relationship

$$1 - K^2_{13.2}(f) = \frac{1-K^2_{3.12}(f)}{1-K^2_{23}(f)} \; . \tag{4.12}$$

<u>Multivariate system response functions and transfer functions</u>

From eq. (4.4), the frequency response function may be explicitly rewritten as

$$(H_{ij}(f)) = (S_{ji}(f))'(S_{jk}(f))'^{-1}, \tag{4.13}$$

where $(.)^{-1}$ is the inverse of a matrix.

The impulse response function is then obtained by

$$h_{ij}(u) = \int_{-\infty}^{\infty} e^{i2\pi fu} H_{ij}(f)\,df, \qquad (i=3,4;\quad j=1,2) \tag{4.14}$$

for continuous time

or $$h_{ij,r} = \int_{-1/2}^{1/2} e^{i2\pi fr} H_{ij}(f)\,df, \qquad (i=3,4;\quad j=1,2) \tag{4.15}$$

for discrete time.

This, in turn leads to the constitution of the following time-invariant linear system

$$x_i(t) = \sum_{j=1}^{2} \int_{-\infty}^{t} h_{ij}(t-u)x_j(u)\,du, \qquad (i=3,4;\quad j=1,2) \tag{4.16}$$

for continuous time

or $$x_{i,n} = \sum_{j=1}^{2} \sum_{r=0}^{\infty} h_{ij,r}x_{j,n-r}, \qquad (i=3,4;\quad j=1,2) \tag{4.17}$$

for discrete time.

In practice for discrete-time model building, it is usually desirable to design an adequate but 'parsimonious' model, i.e., to employ the smallest possible number of parameters yet adequately still representing the real system. This could be accomplished through an ARMA type transfer function model

$$d_i(B)x_{i,n} = \sum_{j=1}^{2} w_{ij}(B)x_{j,n}, \qquad i=3,4 \tag{4.18a}$$

or more explicitly

$$(1-d_{i,1}B-\ldots-d_{i,q}B^{q})x_{i,n} = \sum_{j=1}^{2} (w_{ij,0}-w_{ij,1}B-\ldots w_{ij,r}B^{r})x_{j,n}, \tag{4.18b}$$

where d's and w's are constants; 'q' and 'r' are the orders the ARMA transfer function and are usually relatively small compared to the terms required for the impulse response functions.

With given impulse response functions, the transfer functions may be obtained from the relationship

$$h_{ij}(B) = \frac{w_{ij}(B)}{d_i(B)}, \qquad (i=3,4; \quad j=1,2) \tag{4.19}$$

and vice versa.

For systems with superimposed noise, eq. (4.17) may be modified as

$$x_{i,n} = \sum_{j=1}^{2} \sum_{r=0}^{\infty} h_{ij,r} x_{j,n-r} + y_{i,n}, \qquad i=3,4 \tag{4.17a}$$

and the transfer function model, eq. (4.18) then becomes

$$d_i(B) x_{i,n} = \sum_{j=1}^{2} w_{ij}(B) x_{j,n} + z_{i,n}, \qquad i=3,4 \tag{4.18c}$$

where $Y_{i,n}$ and $Z_{i,n}$ are residuals in the form of either white noises or some other known processes.

Computational Procedures

Data

Because of time and budgetary limitations, only eight out of the 42 natural regime watersheds used for univariate runoff analysis in Chapter 3 were arbitrarily selected for the multivariate hydrologic time series analysis. Four stations are in the Pacific region: three in the Fraser River Basin (08KA007, 08KA005 and 08KB001) and one lake outlet station (08LA008). The other four are from Prairie stations: two in the mountainous Prairies (05AB021 and 05AB002) and two in the central and eastern Prairies (05GG001 and 05TD001). Both monthly runoff data and monthly meteorological data, i.e., temperature and precipitation, for the period 1952-1966 were obtained from the hydrological data bank resulting from the hydrometric network planning study for western and northern Canada by the Department of the Environment (Shawinigan, 1970).

Estimation of cross correlations, cross spectrums and coherencies

Approximation to second-order stationarity (eq. 3.46), estimation of autocorrelations (eq. 2.10) and variance spectrums (eq. 2.54) for temperature and precipitation series were carried out in the same manner as for runoff series in Chapter 3.

Cross covariance functions were estimated by

$$c_{j3}(k) = \frac{1}{N} \sum_{n=1}^{N-k} x_{j,n} x_{3,n+k}, \qquad j=1,2 \tag{4.20}$$

where $X_{1,n}$, $X_{2,n}$ and $X_{3,n}$ are transformed temperature, precipitation and runoff series respectively.

Cross correlation functions were then obtained by

$$r_{j3}(k) = \frac{c_{j3}(k)}{\sqrt{c_{jj}(0)\, c_{33}(0)}}, \qquad j=1,2 \tag{4.21}$$

where $c_{11}(0)$, $c_{22}(0)$ and $c_{33}(0)$ are sample variances for transformed temperature, precipitation and runoff series respectively.

It has been shown (Jenkins and Watts, 1968) that the variance of the sample cross correlation may also be approximated by

$$\mathrm{Var}(r_{xy}(k)) \approx \frac{1}{N}, \tag{4.22}$$

for two uncorrelated discrete processes with either or both being white noises.

The smoothed cross spectrum may be expressed by

$$\overline{S}_{j3}(f) = \overline{C}_{j3}(f) - i\overline{Q}_{j3}(f), \qquad j=1,2 \ (-1/2 \le f \le 1/2) \tag{4.23}$$

where
$$\overline{C}_{j3}(f) = c_{j3}(0) + \sum_{k=1}^{M} w(k)(c_{j3}(k) + c_{3j}(k))\cos 2\pi fk; \tag{4.24}$$

$$\overline{Q}_{j3}(f) = \sum_{k=1}^{M} w(k)(c_{j3}(k) - c_{3j}(k))\sin 2\pi fk; \tag{4.25}$$

and w(k) is the Parzen lag window defined by eq. (3.47).

The phase spectrum was then obtained by

$$\overline{\phi}_{j3}(f) = \tan^{-1}\frac{-\overline{Q}_{j3}(f)}{\overline{C}_{j3}(f)} . \tag{4.26}$$

The squared coherency was computed by

$$\overline{K}^2_{j3}(f) = \frac{\overline{C}^2_{j3}(f) + \overline{Q}^2_{j3}(f)}{\overline{S}_{jj}(f)\overline{S}_{33}(f)} , \tag{4.27}$$

where $\overline{S}_{11}(f)$, $\overline{S}_{22}(f)$ and $\overline{S}_{33}(f)$ are autospectrums for transformed temperature, precipitation and runoff series respectively.

By taking analogy to the multiple correlation coefficient, a z-transformation was carried out for the cross coherency spectrum (Jenkins and Watts, 1968),

$$\overline{Z}(f) = \frac{1}{2} \ln \frac{1+|\overline{K}|}{1-|\overline{K}|} , \tag{4.28}$$

and the approximate 100(1-a)% confidence interval for $\overline{Z}(f)$ is

$$\overline{Z}(f) \pm z(1 - \frac{a}{2}) \sqrt{\frac{1}{v-2q}} , \tag{4.29}$$

where v is defined by eq. (3.49) and q=1.

In order to reduce the bias in estimating the cross spectral characteristics, the two time series under consideration sometimes have to be aligned so that the peak of the cross correlation function always occurs at the zero lag. This is called 'alignment' (Jenkins and Watts, 1968).

The above statistics were also computed between input series, i.e., temperature and precipitation, denoted by subscripts '12'.

<u>Estimation of system response functions, partial and multiple coherencies</u>

The frequency response function of the system, eq. (4.3a) were obtained by solving eq. (4.13)

$$\overline{H}_{31}(f) = \frac{\overline{S}_{13}(f)\overline{S}_{22}(f)-\overline{S}_{23}(f)\overline{S}_{12}(f)}{\overline{S}_{11}(f)\overline{S}_{22}(f)-|\overline{S}_{12}(f)|^2} , \tag{4.30a}$$

$$\overline{H}_{32}(f) = \frac{\overline{S}_{23}(f)\overline{S}_{11}(f)-\overline{S}_{13}(f)\overline{S}_{21}(f)}{\overline{S}_{11}(f)\overline{S}_{22}(f)-|\overline{S}_{12}(f)|^2} . \tag{4.30b}$$

To obtain expressions for the gains and phases, it is necessary to take the modulus and argument of eq. (4.30). Then

$$\overline{G}_{31}(f)=\sqrt{\overline{A}^2_{31}(f) + \overline{B}^2_{31}(f)}, \tag{4.31a}$$

and

$$\phi_{31}(f) = \tan^{-1} \frac{-\overline{B}_{31}}{\overline{A}_{31}}, \tag{4.31b}$$

where

$$\overline{A}_{31}(f) = \frac{\overline{C}_{13}\overline{S}_{33}+\overline{Q}_{23}\overline{Q}_{12}-\overline{C}_{23}\overline{C}_{12}}{\overline{S}_{11}\overline{S}_{22}-|\overline{S}_{12}|^2}, \tag{4.32a}$$

$$\overline{B}_{31}(f) = \frac{\overline{Q}_{13}\overline{S}_{22}-\overline{Q}_{23}\overline{C}_{12}-\overline{C}_{23}\overline{Q}_{12}}{\overline{S}_{11}\overline{S}_{22}-|\overline{S}_{12}|^2}. \tag{4.32b}$$

The gain and phase for $\overline{H}_{32}(f)$ may also be obtained from equations (4.31) and (4.32) by exchanging '$_1$' and '$_2$'.

Approximate confidence limits for gain and phase spectrums are given by

$$\overline{G}(f) \pm \overline{G}(f) \sqrt{\frac{2q}{v-2q} F_{2q,v-2q} \left(\frac{1-\overline{K}^2}{\overline{K}^2}\right)}, \tag{4.33}$$

and

$$\overline{\phi}(f) \pm \sin^{-1} \sqrt{\frac{2q}{v-2q} F_{2q,v-2q} \left(\frac{1-\overline{K}^2}{\overline{K}^2}\right)}, \tag{4.34}$$

where $F_{2q,v-2q}$ is an F-statistic with degrees of freedom (2q, v-2q) and q=2 (Jenkins and Watts, 1968).

The impulse response functions in the time domain were then obtained by taking the Fourier transformation of the corresponding frequency response functions. Therefore

$$h_{3j,r} = 2 \int_0^{1/2} \overline{A}_{3j}(f)\cos 2\pi frdf + 2 \int_0^{1/2} \overline{B}_{3j}(f)\sin 2\pi frdf, \quad j=1,2. \tag{4.35}$$

The multiple squared coherency between the runoff and the meteorological series was obtained from eq. (4.9)

$$\overline{K}^2_{3.12}(f) = \frac{\overline{S}_{22}|\overline{S}_{31}|^2+\overline{S}_{11}|\overline{S}_{32}|^2-2\mathrm{Re}(\overline{S}_{12}\overline{S}_{23}\overline{S}_{31})}{\overline{S}_{33}(\overline{S}_{11}\overline{S}_{22}-|\overline{S}_{12}|^2)}, \tag{4.36}$$

where

$$\mathrm{Re}(\overline{S}_{12}\overline{S}_{23}\overline{S}_{31}) = \overline{C}_{12}\overline{C}_{23}\overline{C}_{13}+\overline{C}_{12}\overline{Q}_{23}\overline{Q}_{13}-\overline{Q}_{12}\overline{Q}_{23}\overline{C}_{13}+\overline{Q}_{12}\overline{Q}_{13}\overline{C}_{23}. \tag{4.37}$$

The residual spectrum was computed using eq. (4.8)

$$\overline{S}^z_{33}(f) = \overline{S}_{33}(f)(1 - \overline{K}^2_{3.12}(f)). \tag{4.38}$$

The partial squared coherency between the output series and an input series after allowing for the other input series was obtained from eq. (4.12)

$$\overline{K}^2_{13.2}(f) = 1 - \frac{1 - \overline{K}^2_{3.12}(f)}{1 - \overline{K}^2_{23}(f)}, \tag{4.39a}$$

$$\overline{K}^2_{23.1}(f) = 1 - \frac{1 - \overline{K}^2_{3.12}(f)}{1 - \overline{K}^2_{13}(f)}. \tag{4.39b}$$

The approximate variance for system impulse response functions has been shown (Hannan, 1970) to be

$$\mathrm{Var}(h_{3j,r}) = \frac{1}{N}\int_{-1/2}^{1/2} \frac{\overline{S}^z_{33}}{\overline{S}_{jj}(1-\overline{K}^2_{12})}\,df, \quad j=1,2. \tag{4.40}$$

The approximate confidence limits for multiple and partial coherency spectrums were obtained in the same manner using equations (4.28) and (4.29) except that in this case q=2 in eq. (4.29).

Identification of system transfer functions

It has already been shown in Chapter 3 that most monthly runoff series may be successfully modelled by a first-order autoregressive process or ARMA(1,0) process. Therefore only two ARMA type transfer function schemes of first-order and second-order respectively were attempted for the identification of the underlying multivariate hydrologic system between monthly runoff and monthly meteorological time series, i.e., temperature and precipitation.

a) First-order ARMA transfer function

The first-order ARMA transfer function is defined by

$$(1-d_1B)x_{3,n} = \sum_{j=1}^{2} (w_{j,0}-w_{j,1}B)x_{j,n-b_j}+z_n, \tag{4.41}$$

where b_j is an integer denoting lag time in months, and Z_n is a residual term to be represented by an ARMA(1,0) process

$$(1-\theta_1B)z_n = a_n. \tag{4.41a}$$

The transfer function parameters d_1, b_j, $w_{j,0}$ and $w_{j,1}$ were obtained from the impulse response function system representation in the following manner.

The discrete-time representation of the system (4.3) can be written as

$$x_{3,n} = \sum_{j=1}^{2} \sum_{r=b_j}^{K} h_{3j,r}x_{j,n-r} , \tag{4.42}$$

where $h_{3j,r} \approx 0$ based on the criterion (4.40) when $r < b_j$ and K is the number of significant terms.

After applying the linear operator $(1-d_1B)$ to eq. (4.42), it becomes

$$(1-d_1B)x_{3,n} = (1-d_1B)\sum_{j=1}^{2}\sum_{r=b_j}^{K} h_{3j,r}x_{j,n-r}$$

$$= \sum_{j=1}^{2} (h_{3j,b_j}x_{j,n-b_j}+(h_{3j,b_j+1}-d_1h_{3j,b_j})x_{j,n-b_j-1}$$

$$+ (h_{3j,b_j+2}-d_1h_{3j,b_j+1})x_{j,n-b_j-2}+\ldots). \tag{4.43}$$

Comparing equations (4.41) and (4.43), the following relationships are obtained,

$$w_{j,0} = h_{3j,b_j}, \qquad j=1,2 \tag{4.44a}$$

$$w_{j,1} = -(h_{3j,b_j+1}-d_1h_{3j,b_j}), \tag{4.44b}$$

$$h_{3j,r+1}-d_1h_{3j,r} = 0, \qquad r>b_j. \tag{4.44c}$$

Applying least squares fitting to eq. (4.44c), d_1 was obtained

$$d_1 = \frac{\sum_{j=1}^{2}\sum_{r=b_j+1}^{K-1} h_{3j,r+1}h_{3j,r}}{\sum_{j=1}^{2}\sum_{r=b_j+1}^{K-1} h^2_{3j,r}} \tag{4.45}$$

which was inserted in eq. (4.44b) to estimate the values for $w_{j,0}$ and $w_{j,1}$. The value of θ_1 was estimated by fitting the ARMA(1,0) process to the residual series Z_n.

The final values of d_1, b_j, $w_{j,0}$, $w_{j,1}$ and θ_1 were those which minimized the variance of the noise A_n, i.e.,

$$E(x_{3,n}-d_1^* x_{3,n-1}-\sum_{j=1}^{2}(w_{j,0}^*-w_{j,1}^* B)x_{j,n-b_j^*}-\theta_1^* z_{n-1})^2 \rightarrow \text{Min.}$$

The above function minimization was carried out using Rosenbrock's algorithm of rotating coordinates (Rosenbrock, 1960; Wilde, 1964), which allows the principal axis of optimum seeking always in the line of the steepest descendent and hence expedites the rate of convergence to some required accuracy.

Denoting by $(p_1, p_2, \ldots, p_K)$ all the parameters d's, b's, w's and θ's defined in eq. (4.44), standard deviations of estimates of model parameters are obtained from the approximate covariance matrix of estimates (Box and Jenkins, 1970):

$$V = (X'X)^{-1}\sigma_a^2 \tag{4.46}$$

where σ_a^2 is the variance of white noises $(a_1, a_2, \ldots, a_N)$ defined in eq. (4.41a);

$V = (\text{cov}(p_i, p_j))$ with $i, j=1,2,\ldots,K$ is the covariance matrix of estimates;

and $X = (x_{kn}) = (\frac{\partial a_n}{\partial p_k})$ with $k=1,2,\ldots,K$ and $n=1,2,\ldots,N$.

b) Second-order ARMA transfer function

The second-order ARMA transfer function is defined by

$$(1-d_1B-d_2B^2)x_{3,n} = \sum_{j=1}^{2} (w_{j,0}-w_{j,1}B-w_{j,2}B^2)x_{j,n-b_j}+z_n, \tag{4.47}$$

where Z_n is a residual term to be modelled by an ARMA(2,0) process

$$(1-\theta_1B-\theta_2B^2)z_n = a_n. \tag{4.48}$$

The transfer function parameters d_1, d_2, b_j, $w_{j,0}$, $w_{j,1}$ and $w_{j,2}$ were again obtained from the impulse response function system representation (4.42)

$$x_{3,n} = \sum_{j=1}^{2} \sum_{r=b_j}^{K} h_{3j,r}x_{j,n-r}, \tag{4.42}$$

where $h_{3j,r} \approx 0$ based on the criterion (4.40) when $r < b_j$ and K is the number of significant terms.

After applying the linear operator $(1-d_1B-d_2B^2)$ to eq. (4.42), it becomes

$$\begin{aligned}(1-d_1B-d_2B^2)x_{3,n} &= (1-d_1B-d_2B^2) \sum_{j=1}^{2} \sum_{r=b_j}^{K} h_{3j,r}x_{j,n-r} \\ &= \sum_{j=1}^{2} (h_{3j,b_j}x_{j,n-b_j}+(h_{3j,b_j+1}-d_1h_{3j,b_j})x_{j,n-b_j-1} \\ &\quad + (h_{3j,b_j+2}-d_1h_{3j,b_j+1}-d_2h_{3j,b_j})x_{j,n-b_j-2} \\ &\quad + (h_{3j,b_j+3}-d_1h_{3j,b_j+2}-d_2h_{3j,b_j+1})x_{j,n-b_j-3}+\dots).\end{aligned} \tag{4.49}$$

Comparing equations (4.47) and (4.49), the following relationships are established,

$$w_{j,0} = h_{3j,b_j}, \qquad j=1,2 \tag{4.50a}$$

$$w_{j,1} = -(h_{3j,b_j+1} - d_1 h_{3j,b_j}), \tag{4.50b}$$

$$w_{j,2} = -(h_{3j,b_j+2} - d_1 h_{3j,b_j+1} - d_2 h_{3j,b_j}), \tag{4.50c}$$

$$h_{3j,r+2} - d_1 h_{3j,r+1} - d_2 h_{3j,r} = 0. \tag{4.50d}$$

Applying least squares fitting to eq. (4.50d), d_1 and d_2 were obtained by

$$\begin{aligned} d_1 = (& \sum_{j=1}^{2} \sum_{r=b_j+1}^{K-2} h_{3j,r+2} h_{3j,r+1} \sum_{j=1}^{2} \sum_{r=b_j+1}^{K-2} h^2_{3j,r} \\ & - \sum_{j=1}^{2} \sum_{r=b_j+1}^{K-2} h_{3j,r+1} h_{3j,r} \sum_{j=1}^{2} \sum_{r=b_j+1}^{K-2} h_{3j,r+2} h_{3j,r}) \\ & /(\sum_{j=1}^{2} \sum_{r=b_j+1}^{K-2} h^2_{3j,r+1} \sum_{j=1}^{2} \sum_{r=b_j+1}^{K-2} h^2_{3j,r} \\ & -(\sum_{j=1}^{2} \sum_{r=b_j+1}^{K-2} h_{3j,r+1} h_{3j,r})^2), \end{aligned} \tag{4.51a}$$

$$d_2 = \Big(\sum_{j=1}^{2}\sum_{r=b_j+1}^{K-2} h^2_{3j,r+1}\sum_{j=1}^{2}\sum_{r=b_j+1}^{K-2} h_{3j,r+2}h_{3j,r}$$

$$-\sum_{j=1}^{2}\sum_{r=b_j+1}^{K-2} h_{3j,r+2}h_{3j,r+1}\sum_{j=1}^{2}\sum_{r=b_j+1}^{K-2} h_{3j,r+1}h_{3j,r}\Big)$$

$$/\Big(\sum_{j=1}^{2}\sum_{r=b_j+1}^{K-2} h^2_{3j,r+1}\sum_{j=1}^{2}\sum_{r=b_j+1}^{K-2} h^2_{3j,r}$$

$$-\Big(\sum_{j=1}^{2}\sum_{r=b_j+1}^{K-2} h_{3j,r+1}h_{3j,r}\Big)^2\Big); \qquad (4.51b)$$

which were then inserted in equations (4.50b) and (4.50c) to obtain the estimates of $w_{j,0}$, $w_{j,1}$ and $w_{j,2}$. The values of θ_1 and θ_2 were estimated by fitting the ARMA(2,0) process to the residual series Z_n.

The final values of d_1, d_2, b_j, $w_{j,0}$, $w_{j,1}$, $w_{j,2}$, θ_1 and θ_2 were those which minimized the variance of the noise A_n, i.e.,

$$E(x_{3,n}-d_1^*x_{3,n-1}-d_2^*x_{3,n-2}-\sum_{j=1}^{2}(w^*_{j,0}-w^*_{j,1}B-w^*_{j,2}B^2)x_{j,n-b_j^*}-(\theta_1^*-\theta_2^*B)z_{n-1})^2 \rightarrow \text{Min.}$$

Approximate standard deviations of estimates of model parameters were obtained in the same manner using eq. (4.46).

Evaluation of proposed models

Validation of the above proposed models was not carried out because of time and budgetary limitations. The efficiency of the models was then evaluated by the following procedures:

i) Comparison between the estimated impulse response function from eq. (4.35) and the theoretical impulse response function using eq. (4.19). Approximate standard deviations of the respective estimated system impulse response functions were computed by taking square root of the variances defined by eq. (4.40).

ii) Comparison between estimated and theoretical frequency response functions, i.e., estimated and theoretical gain and phase spectrums. Confidence limits of gain and phase spectrums were computed using equations (4.33) and (4.34) respectively.

iii) Examination of autocorrelations of noise series A_n and its cross correlations with input series $X_{1,n}$ and $X_{2,n}$. One basic assumption concerning the noise term of the above schemes is that it is of the nature of a purely random process, or white noises and it is uncorrelated with the input series.

iv) Examination of the unexplained portion of the variance or the variance of noise series A_n due to the proposed models.

Along with the presentation and discussion of the results in the following section, the above procedures are illustrated in graphical forms by a typical example at 05AB002 - Willow Creek near Nolan, Alberta.

Results and Discussion

Seasonal and stochastic components of monthly runoff, temperature and precipitation series

Means and variances of the raw time series were computed for runoff, temperature and precipitation respectively for each watershed and are listed in Table 4.1. It can be observed that the temperature is warmer in the Pacific region and tends to gradually decrease towards north or east. On the other hand, it is observed that the Prairie region has larger variation in temperature than the Pacific region. Runoff and precipitation are generally higher in the mountainous area and also exhibit larger variation.

Harmonic analysis was carried out for runoff, temperature and precipitation series and is also presented in Table 4.1. It can be observed that the annual cycle and its sub-harmonics, i.e., 12, 6, 4, 3, 2.4 and 2 (months) have accounted for 92.0% to 97.3% of the total variance of the temperature; whereas only 25.7% to 66.2% for the precipitation series and in this case 33.5% to 93.2% for the runoff series. Furthermore, the seasonal influence on the precipitation is more dominant in the Prairie region.

Annual seasonal components for runoff, temperature and precipitation were removed by subtracting the long-term monthly means (Figure 4.1a) from the raw data using eq. (3.40). The remaining stochastic components then have variances varying from 2.7% to 8.0% of the total variance for the temperature and 33.8% to 74.3% for the precipitation, and in this case 6.8% to 66.5% for the runoff (Table 4.1). It can be seen from Figure 4.1b that there are significant differences among long-term monthly variances, which was also confirmed by the Bartlett test of homogeneity of variances (Bartlett, 1937). It was then necessary to remove the seasonal variation in monthly variances to achieve the second-order stationarity by dividing the individual value of the

TABLE 4.1

Summary of General Statistics of Hydrologic Time Series Involved in Multivariate Watershed System Identification

Station Identification			05AB002			05AB021		
Drainage Area			900			446		
Met. Stations Available			12			3		
Statistical Parameters \ Hydrological Parameters			Runoff	Temp	Precip	Runoff	Temp	Precip
Mean			0.20 in	38.6 °F	1.65 in	0.30 in	37.3 °F	1.92 in
Variance			0.14 in^2	255.2 $(°F)^2$	1.60 in^2	0.29 in^2	229.5 $(°F)^2$	2.10 in^2
Explained Seasonal Variance %		12 month	25.68	91.15	27.44	27.16	90.97	29.93
		6	6.31	0.14	3.37	8.11	0.34	2.57
		4	1.03	0.06	1.16	1.34	0.31	0.22
		3	0.30	0.11	1.19	0.25	0.17	0.14
		2.4	0.11	0.31	2.92	0.09	0.36	3.69
		2	0.06	0.24	1.34	0.03	0.39	0.78
		Total	33.48	92.01	37.42	36.98	92.53	37.34
Residual Variance			0.09 in^2	20.4 $(°F)^2$	1.00 in^2	0.18 in^2	17.1 $(°F)^2$	1.31 in^2
Auto and Cross Corr. Coeffs.	Runoff	Zero Order	1.00	-0.06	0.14	1.00	-0.10	0.12
		1st	0.51	-0.32	0.32	0.57	-0.26	0.27
		2nd	0.29	-0.19	0.20	0.39	-0.23	0.22
	Temp	Zero Order	-0.06	1.00	-0.55	-0.01	1.00	-0.50
		1st	0.09	0.05	0.01	0.06	0.03	0.02
		2nd	0.08	0.03	-0.05	0.05	0.04	-0.01
	Precip	Zero Order	0.14	-0.55	1.00	0.12	-0.50	1.00
		1st	-0.15	0.05	-0.01	-0.14	0.04	-0.01
		2nd	-0.17	0.08	0.03	-0.10	0.10	-0.07

TABLE 4.1 (continued)

Summary of General Statistics of Hydrologic Time Series

Involved in Multivariate Watershed System Identification

Station Identification			05GG001			05TD001		
Drainage Area				46100			6030	
Met. Stations Available				182			7	
Statistical Parameters \ Hydrological Parameters			Runoff	Temp	Precip	Runoff	Temp	Precip
Mean			0.22 in	35.8 °F	1.38 in	0.45 in	30.5 °F	1.43 in
Variance			0.04 in^2	413.0 $(°F)^2$	1.11 in^2	0.11 in^2	620.5 $(°F)^2$	0.82 in^2
Explained Seasonal Variance %		12 month	66.75	94.90	51.37	31.18	96.44	45.69
		6	4.26	0.28	13.24	6.28	0.46	3.59
		4	0.14	0.04	0.03	1.62	0.09	0.35
		3	1.02	0.20	0.46	0.87	0.05	0.83
		2.4	0.66	0.11	0.97	0.30	0.08	1.27
		2	0.16	0.15	0.10	0.06	0.17	2.25
		Total	73.01	95.68	66.17	40.31	97.29	53.98
Residual Variance			0.01 in^2	17.9 $(°F)^2$	0.38 in^2	0.07 in^2	16.8 $(°F)^2$	0.38 in^2
Auto and Cross Corr. Coeffs	Runoff	Zero Order	1.00	0.07	0.01	1.00	-0.10	0.13
		1st	0.64	-0.20	0.30	0.79	-0.18	0.16
		2nd	0.42	-0.24	0.25	0.69	-0.13	0.09
	Temp	Zero Order	0.07	1.00	-0.44	-0.10	1.00	-0.16
		1st	0.04	0.09	-0.13	-0.07	0.11	-0.12
		2nd	-0.01	0.06	0.03	-0.02	-0.01	-0.01
	Precip	Zero Order	0.01	-0.44	1.00	0.13	-0.16	1.00
		1st	-0.11	-0.12	0.06	0.11	-0.09	0.01
		2nd	0.04	-0.14	-0.05	0.07	-0.09	-0.03

TABLE 4.1 (continued)

Summary of General Statistics of Hydrologic Time Series

Involved in Multivariate Watershed System Identification

Station Identification			08KA005			08KA007		
Drainage Area			2690			615		
Met. Stations Available			3			1		
Statistical Parameters \ Hydrological Parameters			Runoff	Temp	Precip	Runoff	Temp	Precip
Mean			2.90 in	38.0 °F	2.04 in	3.05 in	35.5 °F	2.58 in
Variance			8.32 in^2	254.4 $(°F)^2$	0.93 in^2	11.6 in^2	262.9 $(°F)^2$	2.08 in^2
Explained Seasonal Variance %		12 month	73.65	92.10	19.47	67.15	92.54	24.82
		6	16.06	0.68	4.21	18.65	0.90	4.25
		4	3.25	0.09	0.47	5.31	0.15	0.03
		3	0.23	0.05	0.25	0.97	0.10	0.03
		2.4	0.06	0.34	0.42	0.13	0.34	0.72
		2	0.01	0.14	0.88	0.05	0.21	0.60
		Total	93.26	93.41	25.70	92.25	94.23	30.46
Residual Variance			0.56 in^2	16.8 $(°F)^2$	0.69 in^2	0.90 in^2	15.1 $(°F)^2$	1.45 in^2
Auto and Cross Corr. Coeffs.	Runoff	Zero Order	1.00	0.31	0.02	1.00	0.27	0.01
		1st	0.33	-0.01	0.15	0.38	0.001	0.14
		2nd	0.19	-0.05	0.10	0.16	-0.10	0.08
	Temp	Zero Order	0.31	1.00	-0.33	0.27	1.00	-0.33
		1st	-0.001	0.27	-0.16	0.001	0.29	-0.20
		2nd	-0.02	0.12	-0.11	-0.001	0.12	-0.11
	Precip	Zero Order	0.02	-0.33	1.00	0.01	-0.33	1.00
		1st	-0.02	-0.13	-0.03	0.05	-0.10	-0.001
		2nd	0.05	-0.001	-0.05	0.08	-0.05	0.02

TABLE 4.1 (continued)

Summary of General Statistics of Hydrologic Time Series

Involved in Multivariate Watershed System Identification

Station Identification			08KB001			08LA008		
Drainage Area			12500			1780		
Met. Stations Available			20			4		
Statistical Parameters \ Hydrologic Parameters			Runoff	Temp	Precip	Runoff	Temp	Precip
Mean			2.68 in	37.8 °F	2.24 in	0.75 in	40.2 °F	1.84 in
Variance			4.92 in^2	252.6 $(°F)^2$	8.01 in^2	0.75 in^2	222.1 $(°F)^2$	0.95 in^2
Explained Seasonal Variance %		12 month	67.19	92.68	16.46	46.03	92.42	12.86
		6	15.03	0.75	11.85	24.28	0.28	12.77
		4	5.09	0.08	2.38	8.78	0.02	3.07
		3	0.37	0.03	0.08	2.21	0.09	2.07
		2.4	0.37	0.24	4.05	0.63	0.37	7.02
		2	0.01	0.17	1.31	0.08	0.34	1.31
		Total	88.05	93.95	36.13	82.01	93.52	39.10
Residual Variance			0.59 in^2	15.3 $(°F)^2$	0.51 in^2	0.13 in^2	14.4 $(°F)^2$	0.58 in^2
Auto and Cross Corr. Coeffs.	Runoff	Zero Order	1.00	0.20	0.19	1.00	0.04	0.003
		1st	0.43	-0.09	0.29	0.66	-0.09	0.25
		2nd	0.17	-0.06	0.10	0.46	-0.14	0.15
	Temp	Zero Order	0.20	1.00	-0.35	0.04	1.00	-0.35
		1st	0.03	0.29	-0.14	0.04	0.34	-0.28
		2nd	-0.02	0.13	-0.12	0.07	0.16	-0.09
	Precip	Zero Order	0.19	-0.35	1.00	0.003	-0.35	1.00
		1st	0.07	-0.17	-0.03	-0.10	-0.10	0.04
		2nd	-0.06	-0.09	-0.04	-0.15	-0.09	0.003

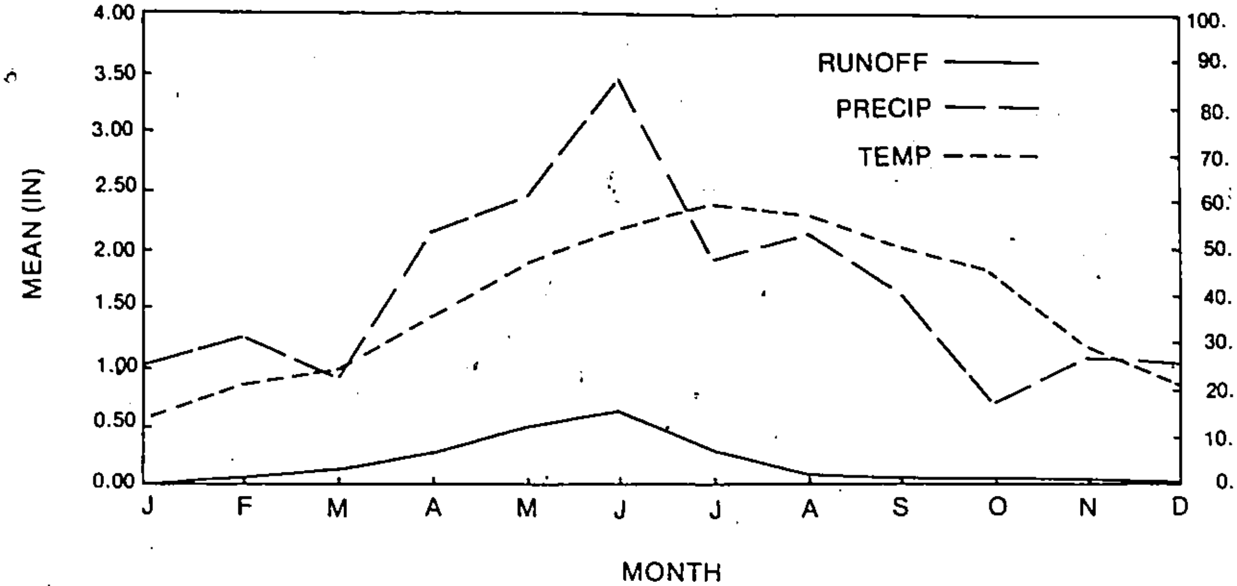


Figure 4.1a Monthly Means of Hydrologic Time Series - 05AB002 Willow Creek near Nolan

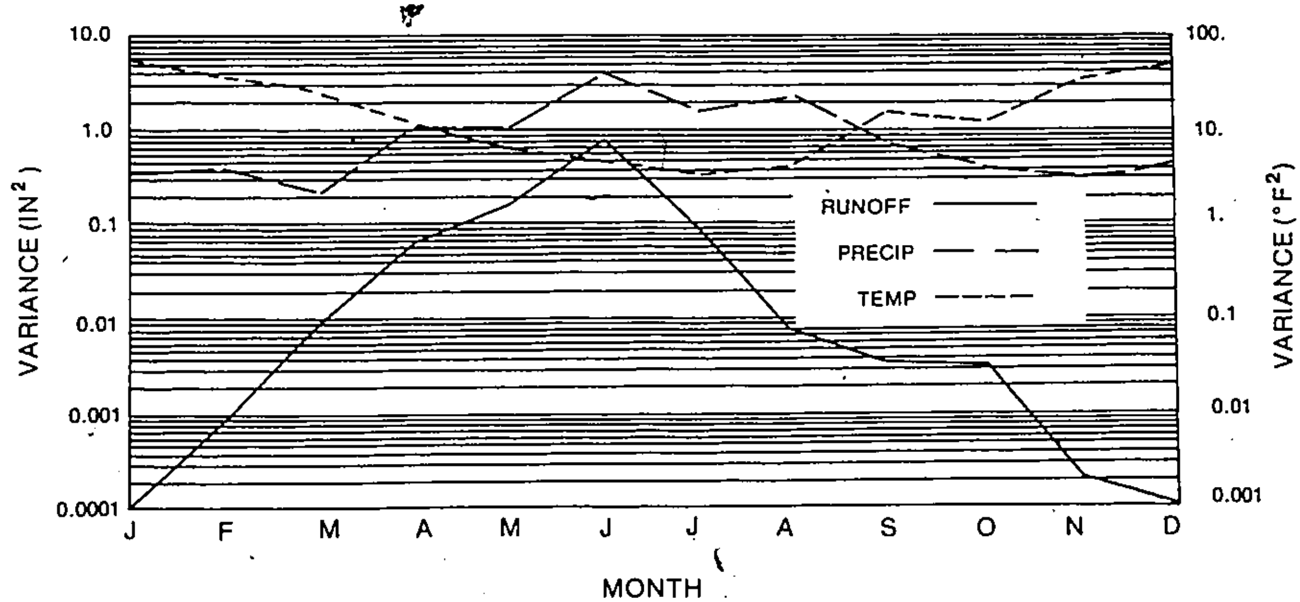


Figure 4.1b Monthly Variances of Hydrologic Time Series - 05AB002 Willow Creek near Nolan

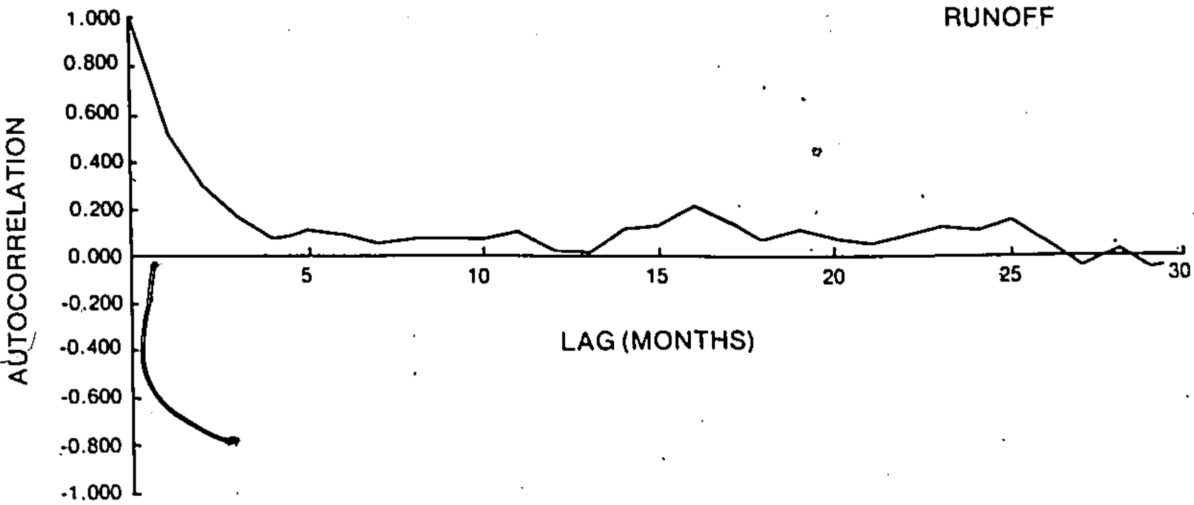
seasonal mean removed time series by the corresponding monthly standard deviation. The resulting stochastic component would yield zero mean and unit variance throughout the whole study period and between months.

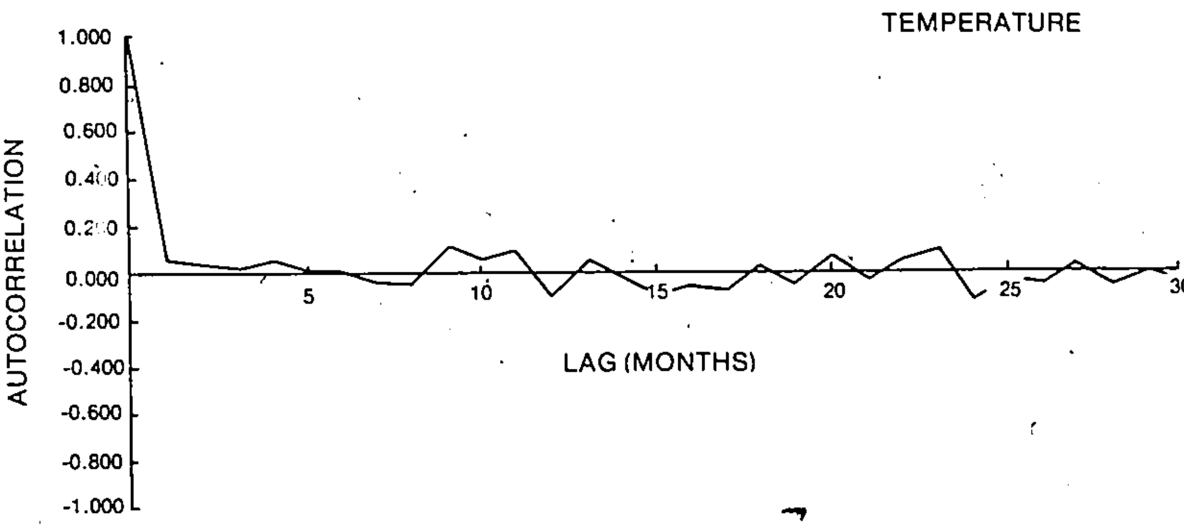
The following statistical analyses are confined to these transformed stationary runoff, temperature and precipitation time series.

Autocorrelation functions and autospectrums

Sample estimates of the autocorrelations for the transformed runoff, temperature and precipitation series were computed using equations (2.8) and (2.10) and are illustrated in Figure 4.2. Autocorrelations up to the first two lags are listed in Table 4.1. It is observed that all the runoff series fall into the autoregressive-moving average scheme. However, temperature and precipitation showed no obvious deviation from a purely random process except for temperatures in the Pacific mountainous region indicating persistence in the time series. This latter observation along with the previous finding that temperatures in this same region showed less variation all indicate that the Pacific region is under the influence of marine climate and should be treated as a separate statistical or hydrological entity from the Prairies which is under the influence of continental climate. This supports the conclusion from Chapter 3 that the whole study area should be divided into two regions, namely, Pacific region and Prairie region.

Smoothed autospectrums for the transformed runoff and meteorological series were computed using eq. (2.54) and are illustrated in Figure 4.3, whose patterns further confirm the above observation.





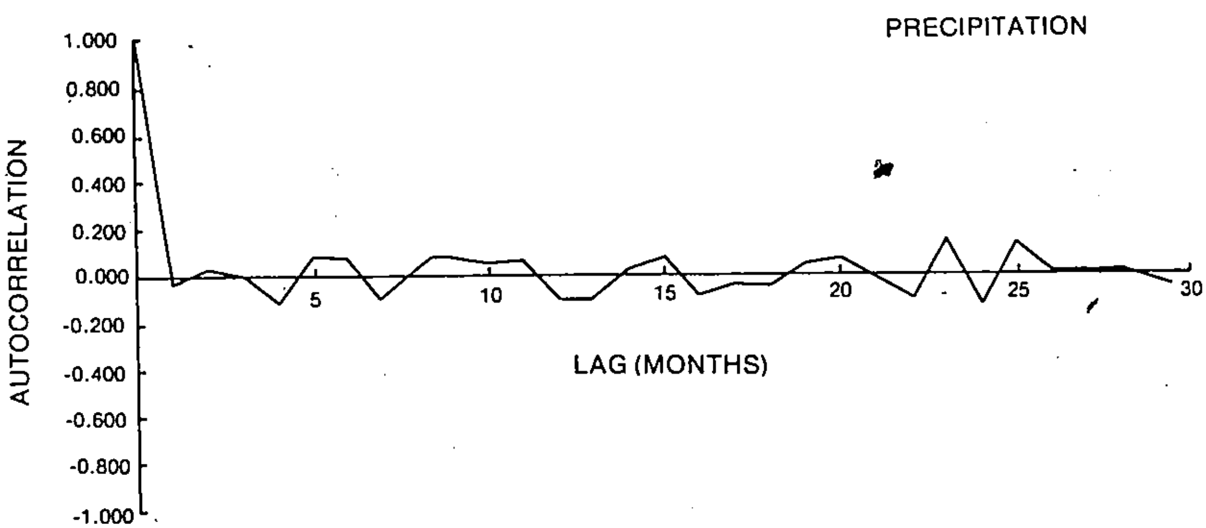


Figure 4.2 Autocorrelations of Stochastic Components of Hydrologic Time Series - 05AB002 Willow Creek near Nolan

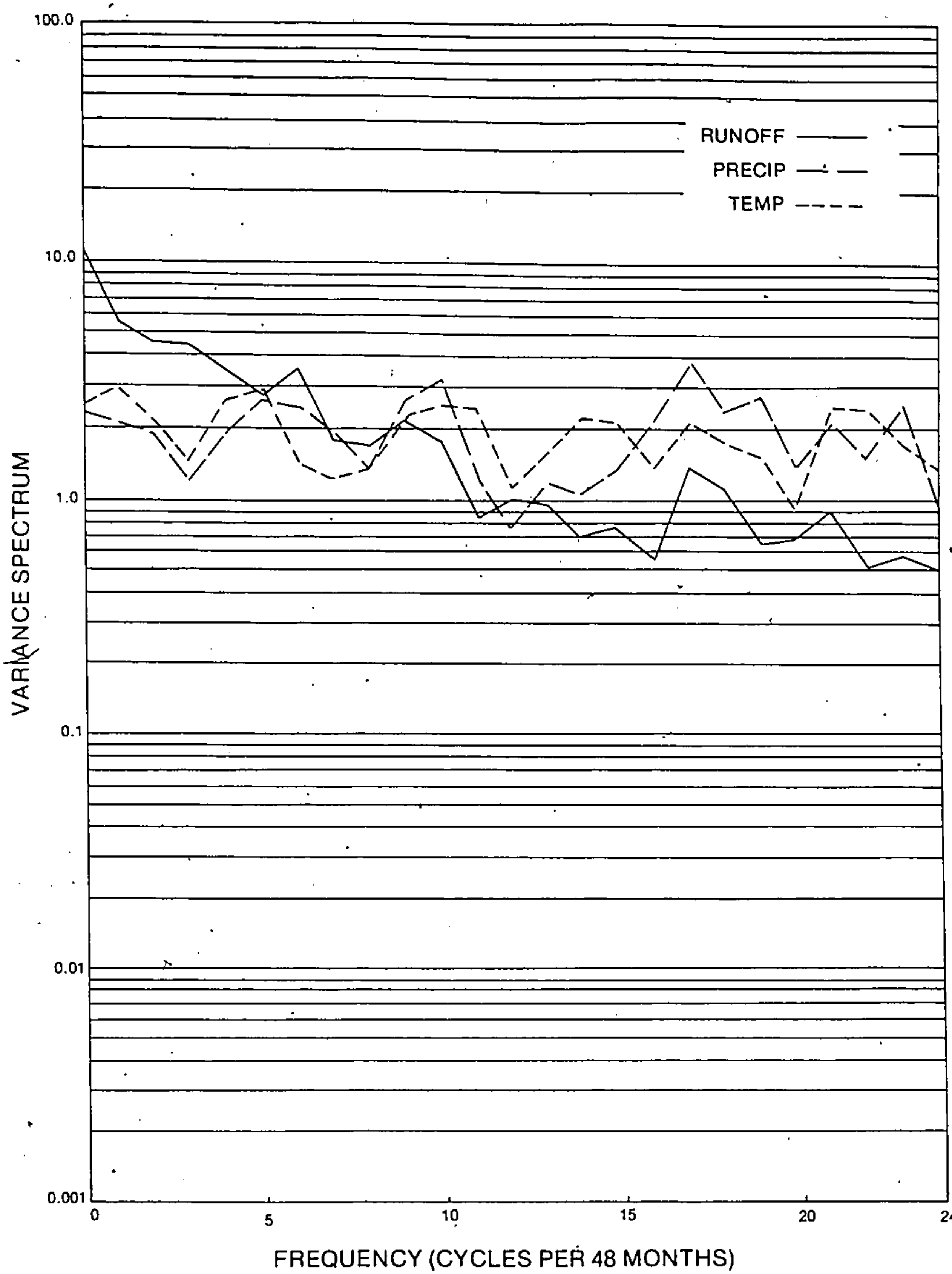


Figure 4.3 Variance Spectrums of Stochastic Components of Hydrologic Time Series - 05AB002 Willow Creek near Nolan

Cross correlations, cross spectrums and coherencies

Cross correlations among the transformed runoff, temperature and precipitation series were estimated using equations (4.20) and (4.21) and are illustrated in Figure 4.4. Cross correlations up to the first two lags are given in Table 4.1. It is observed that the cross correlations between the runoff and the meteorological series are not stable except at the three stations 05AB002, 05AB021 and 05GG001. These are the three watersheds with denser meteorological network (Table 4.1), which provides more representative pictures of the meteorological condition in the watershed. It is also observed from Table 4.1 and Figure 4.4, however, that the meteorological series generally lead the runoff series, i.e., peaks occur at positive lags. Cross correlations between the temperature and precipitation series are stable. This is expected because both temperature and precipitation are measured at the same station. All the meteorological time series have high cross correlations at the zero lag. There are no very distinct peaks observed in the computed cross correlation functions (Figure 4.4). Therefore 'alignment' of the time series was not considered necessary in performing the cross spectral analysis and the subsequent multiple cross spectral analysis.

Cospectrums, quadrature spectrums, phase functions and coherencies were computed using equations (4.24), (4.25), (4.26) and (4.27) respectively and are illustrated in Figures 4.5 and 4.6. They generally support the findings from the above discussion on cross correlation functions, i.e., meteorological series generally lead the runoff series (negative phase spectrum over most frequencies),

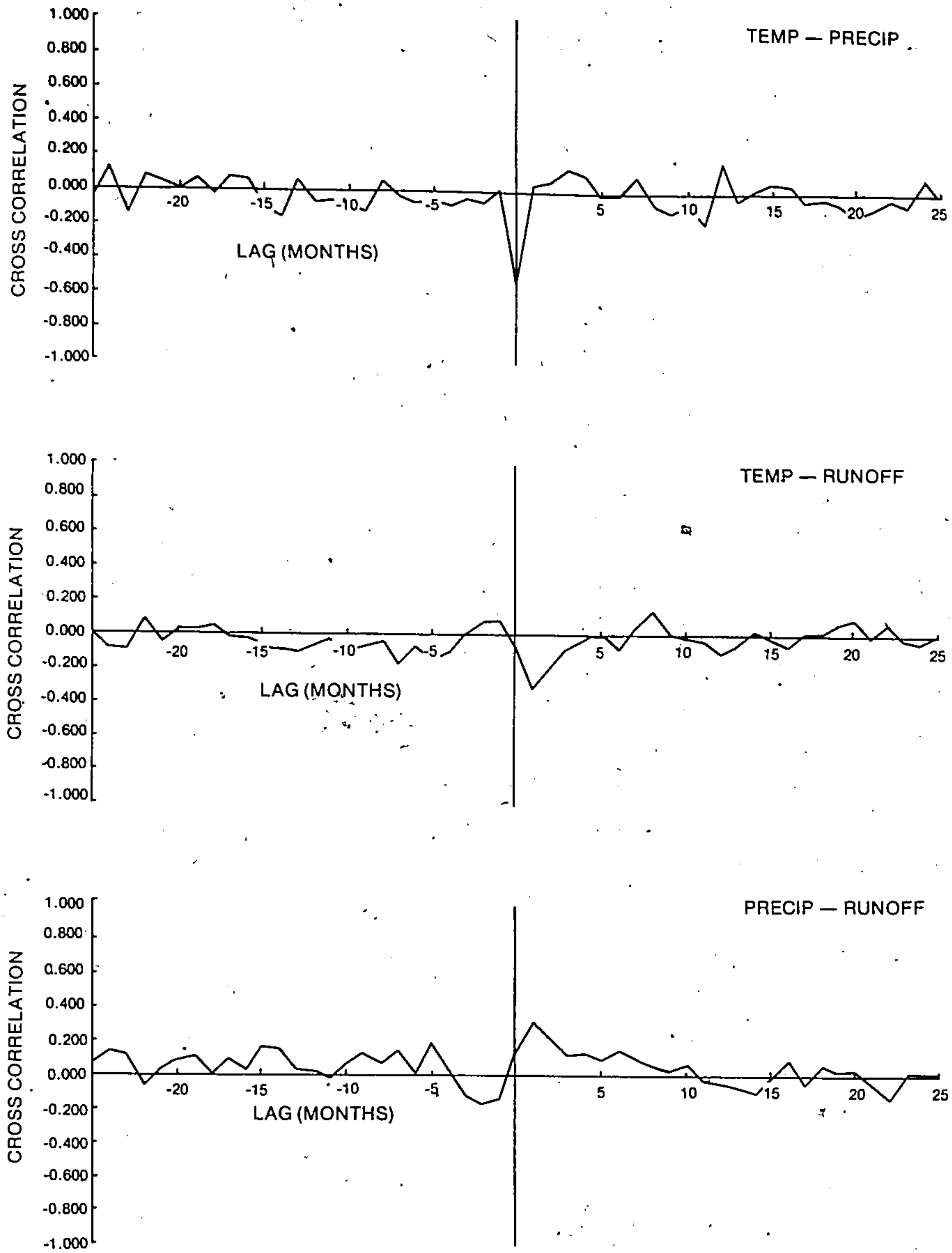


Figure 4.4 Cross Correlations Among Stochastic Components of Hydrologic Time Series - 05AB002 Willow Creek near Nolan

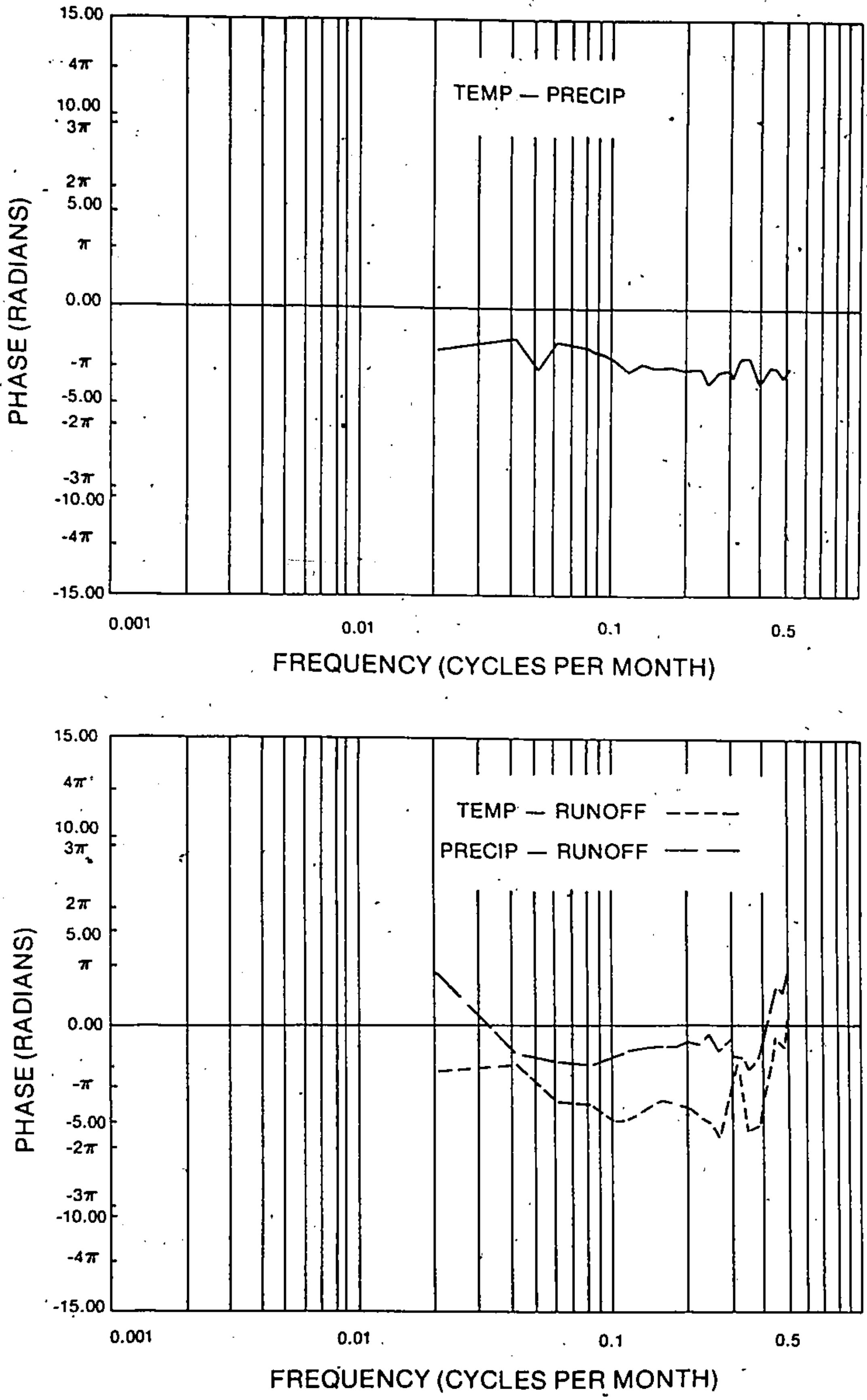
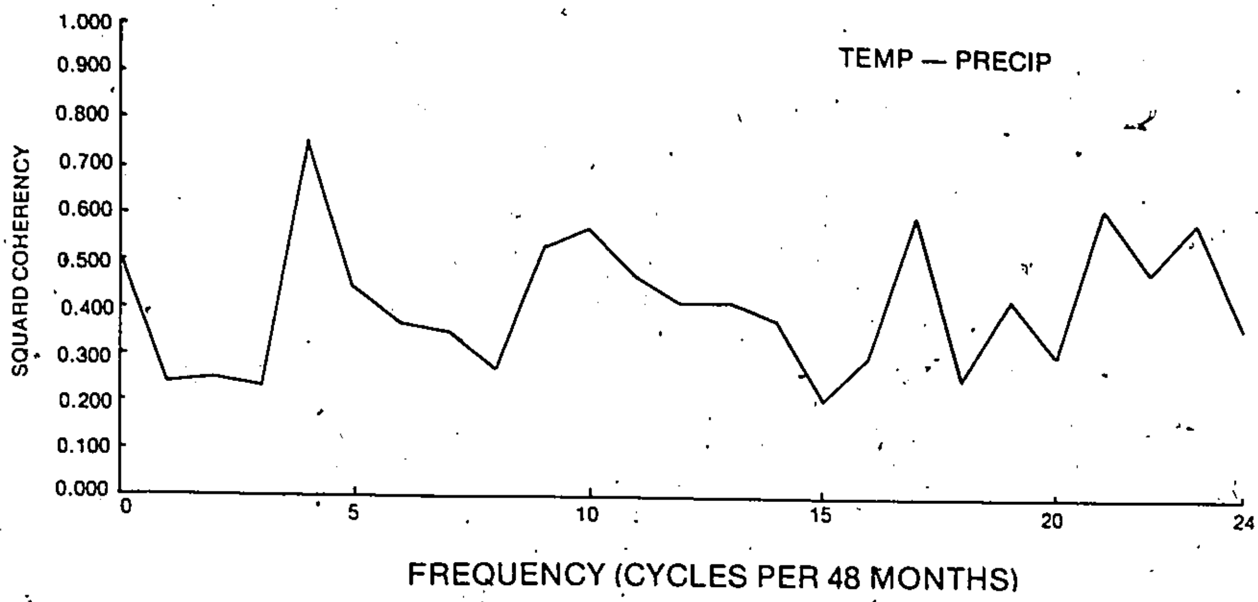


Figure 4.5 Phase Functions among Stochastic Components of Hydrologic Time Series - 05AB002 Willow Creek near Nolan



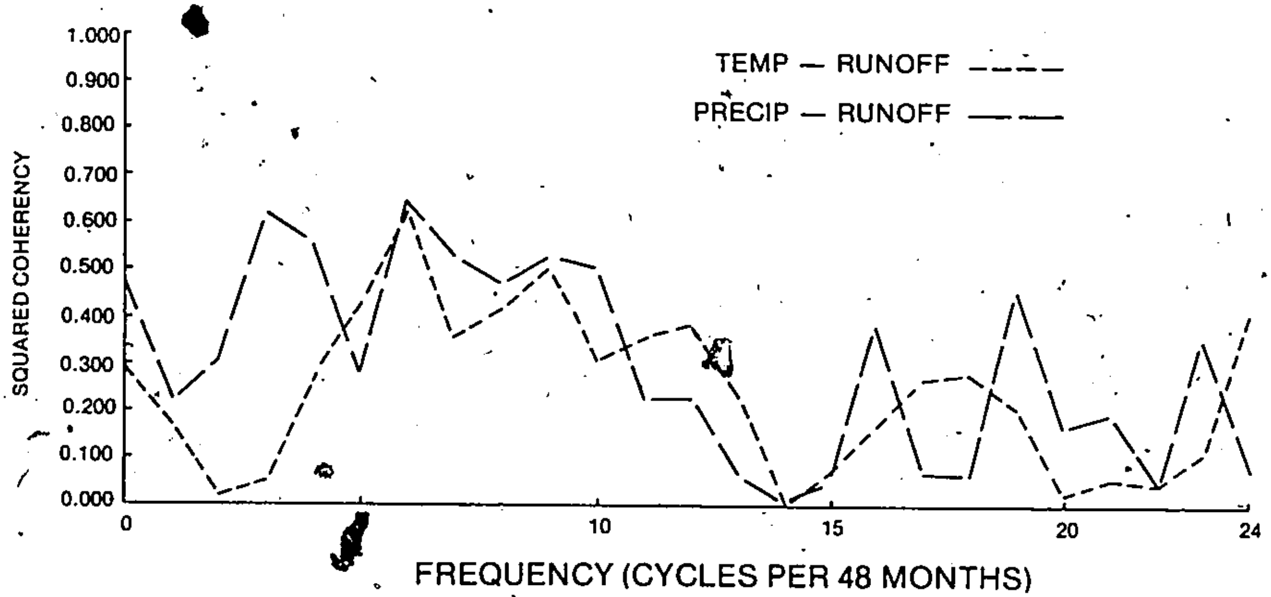


Figure 4.6 Squared Coherencies among Stochastic Components of Hydrologic Time Series - 05AB002 Willow Creek near Nolan

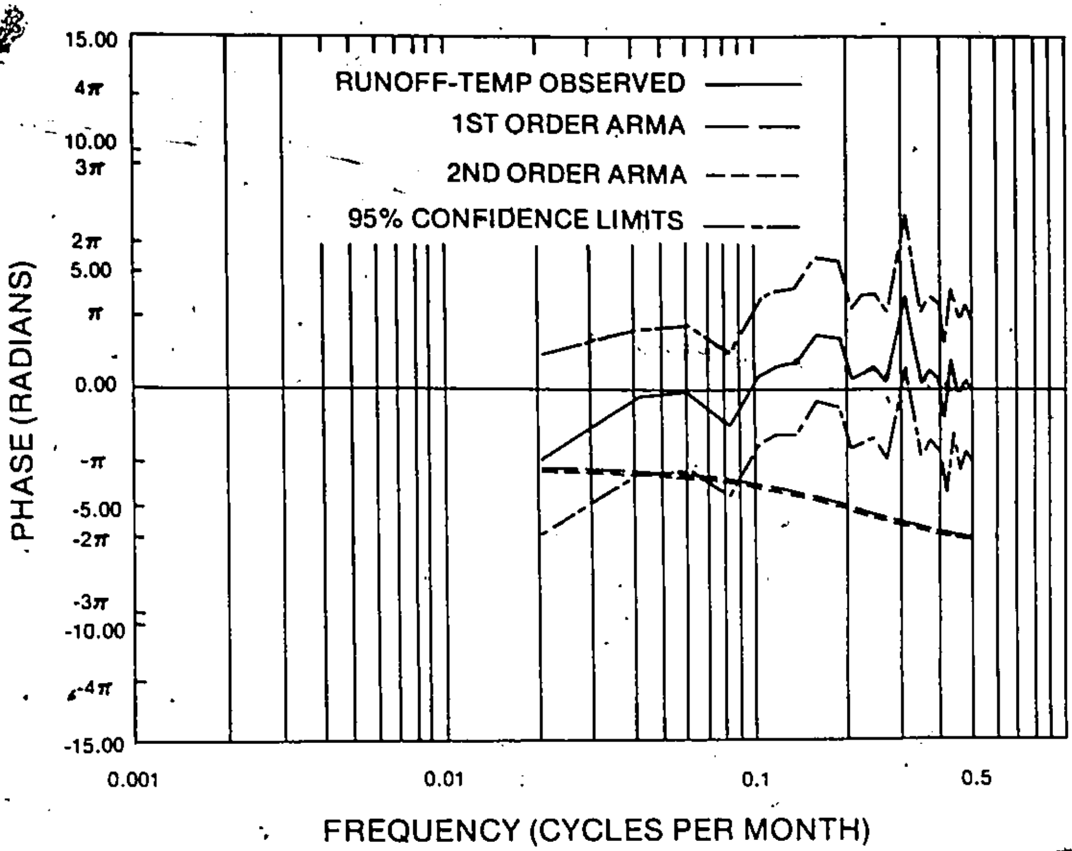
significant coherencies exist at low frequencies between runoff and meteorological series and among meteorological series.

Frequency response functions, partial and multiple coherencies

As in multiple regression analysis, frequency response functions serve as the regression coefficients in the frequency domain. They measure the relative contribution of variance to the output process from one of the input processes allowing for the effect of other input processes at various frequencies.

Gain and phase for the frequency response functions were estimated from equations (4.31) and (4.32) and the results are illustrated in Figures 4.7 and 4.8. Most of the gain functions revealed a system of exponential decay or damped sine wave, which is the characteristic of an ARMA transfer function. Phase functions, however, were hardly resembling any of the known systems, let alone to detect absolute delay in physical time. This was also confirmed by Dowling (1968) and a time domain system response function is therefore required for the determination of physical time lags and for the construction of a time domain transfer function model which is to be discussed in the following sub-section.

Partial coherency spectrums were computed using eq. (4.39) and are illustrated in Figure 4.9. As expected, they were lower than the corresponding cross coherency spectrums (Figure 4.6) but not significantly different. Multiple coherency spectrums between the runoff and the meteorological series were obtained from equations (4.36) and (4.37) and are also illustrated in Figure 4.9. It can be seen from Figure 4.9 that



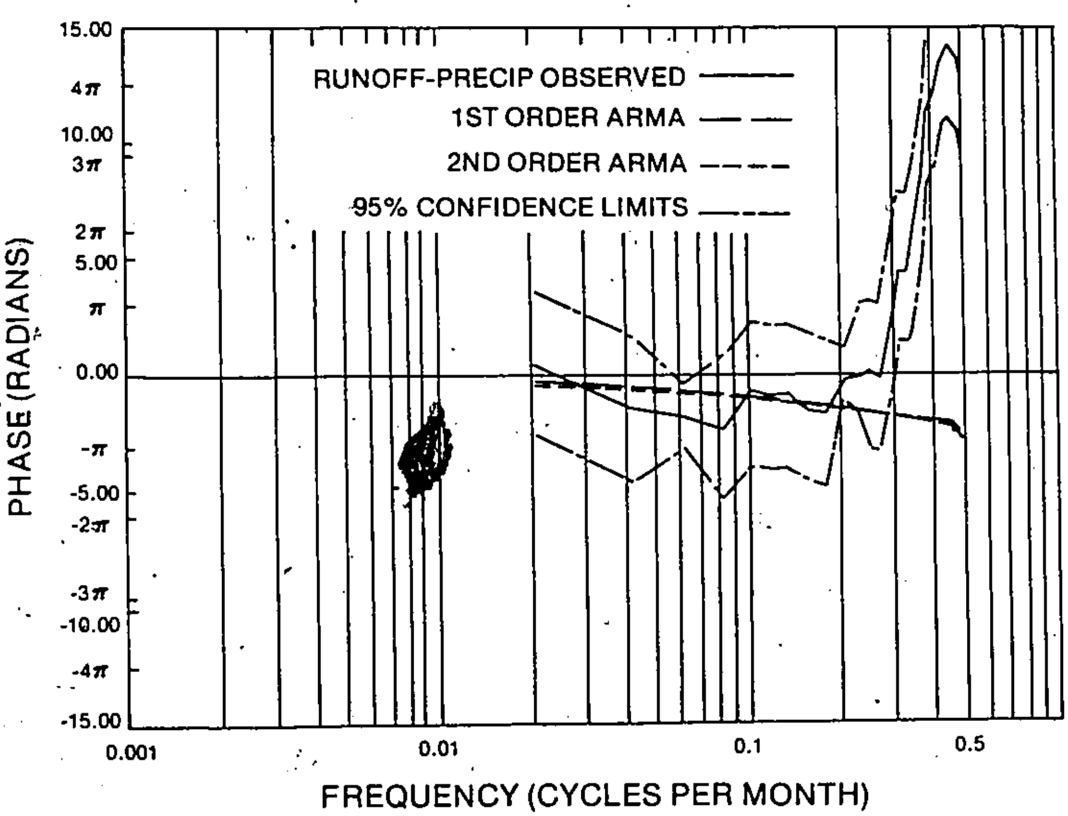


Figure 4.7 Phases of Frequency Response Functions of Watershed Hydrologic System - 05AB002 Willow Creek near Nolan

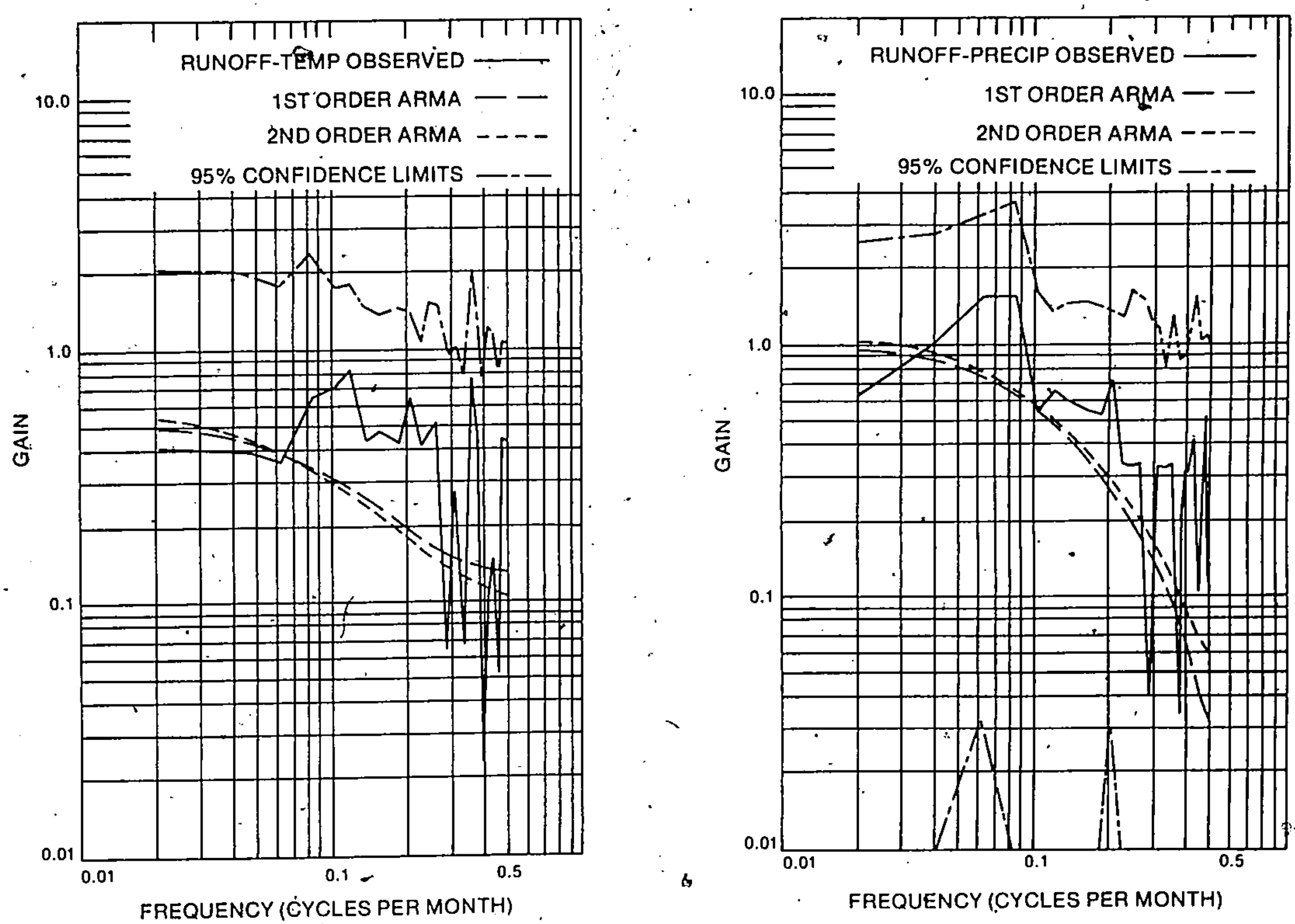
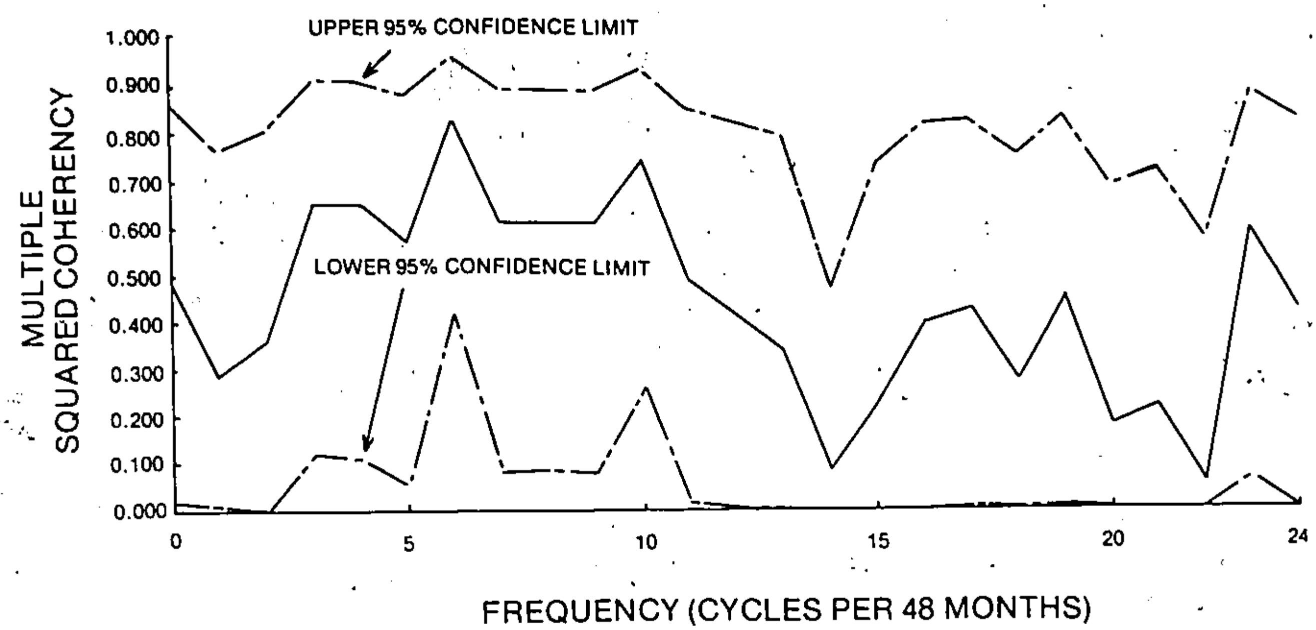


Figure 4.8 Gains of Frequency Response Functions of Watershed Hydrologic System - 05AB002 Willow Creek near Nolan



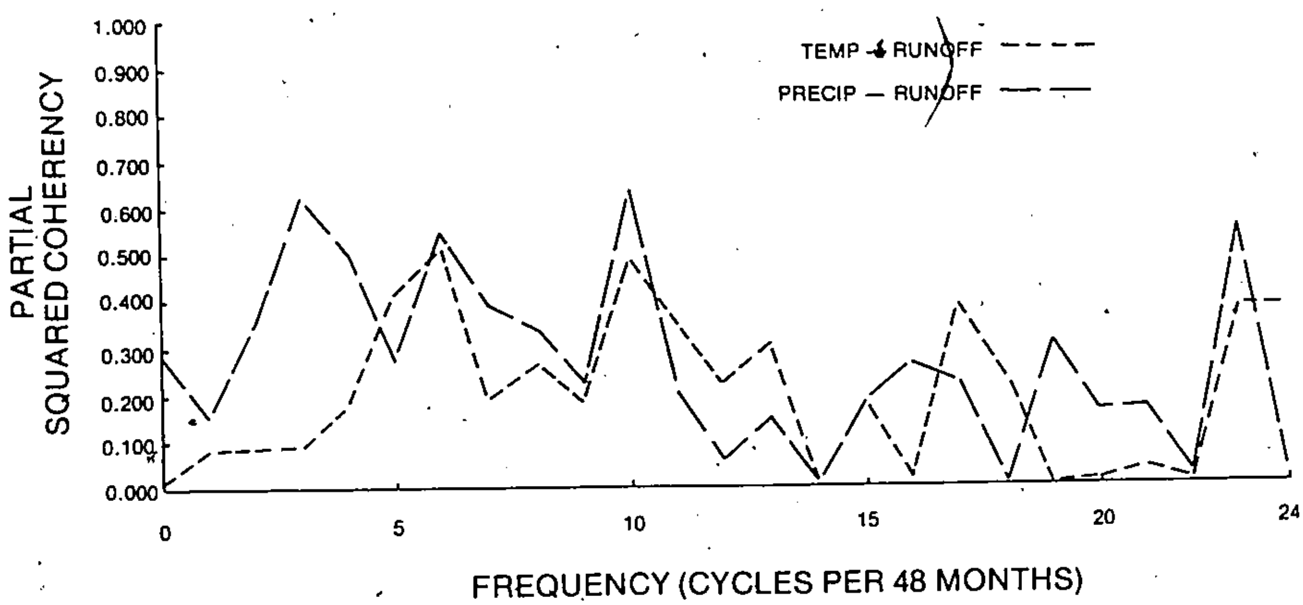


Figure 4.9 Multiple and Partial Squared Coherencies among Stochastic Components of Hydrologic Time Series - 05AB002 Willow Creek near Nolan

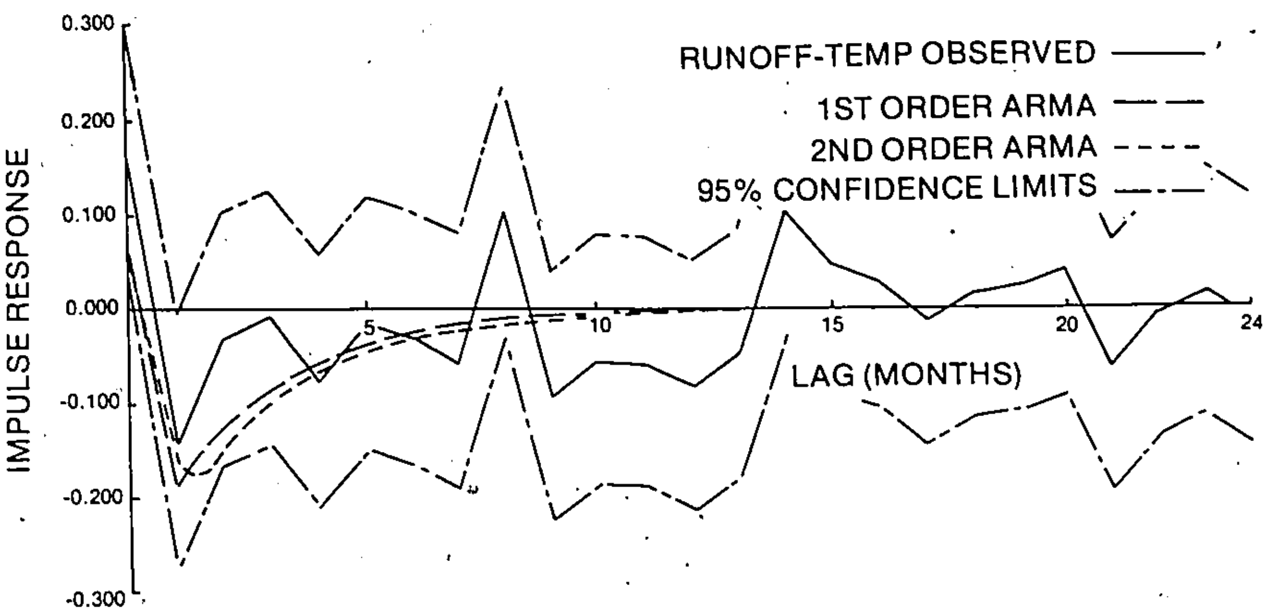
the coherency was generally improved through the multiple correlation regression of the runoff on the meteorological time series, i.e., temperature and preceipitation.

It can also be observed that gain spectrums of the frequency response functions and multiple coherencies all indicate high values at low frequencies around the annual cycle (Figures 4.7, 4.8 and 4.9). Recalling the fact that monthly runoff series fall into a first-order autoregressive scheme which has rich variance at low frequencies (Figure 3.4), the above observation further confirms the finding in Chapter 3 that a natural watershed is a 'damped' system which dampens out the high frequency variations of the meteorological inputs through its storage effect.

Time domain impulse response functions and identification of system transfer functions

Impulse response functions were obtained by taking the Fourier transformation of the corresponding frequency response functions using eq. (4.35) and the results are illustrated in Figure 4.10. Their patterns did follow the form of exponential decay or damped sine wave. This justifies the postulation of an ARMA type transfer function model.

Physical time lags were determined based on the criterion (4.40). Two types of ARMA transfer functions were attempted, namely, first-order and second-order. Initial values of the transfer function parameters were estimated using equations (4.44), (4.45) and (4.50), (4.51) respectively for the two models. The values for the noise model



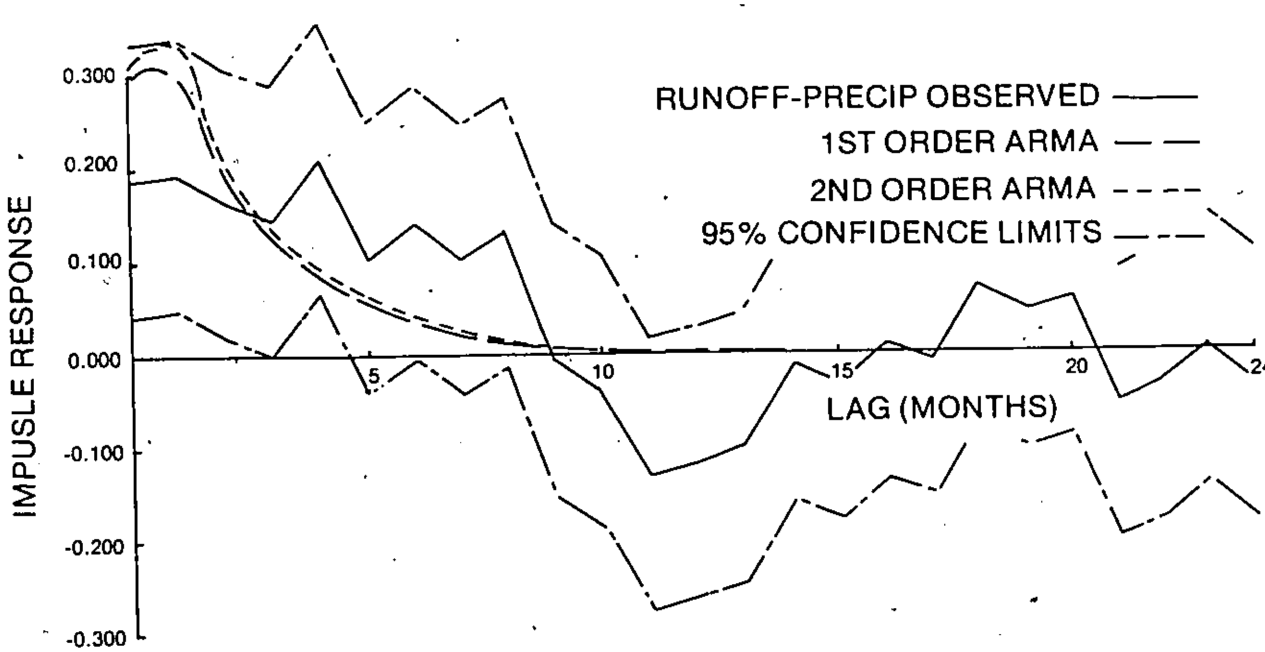


Figure 4.10 Impulse Response Functions of Watershed Hydrologic System - 05AB002 Willow Creek near Nolan

parameters were obtained by fitting ARMA(1,0) and ARMA(2,0) processes respectively to the two noise series. These initial model parameters are listed in Table 4.2*.

The final values of the model parameters were then obtained through model parameter optimization based on Rosenbrock's hill-climbing techniques. The final model parameters and their approximate standard deviations are also given in Table 4.2. It can be seen that most of the watershed systems could be represented by both first-order and second-order ARMA transfer function models (Table 4.2 and Figures 4.7, 4.8, 4.10 and 4.11).

Furthermore, from the results in Table 4.2 it can be observed that for the Prairie region the initial parameters generally agree remarkably well with the final parameters after parameter optimization. The estimates of time lags are also consistent in both cases. Therefore system impulse response function seems to be an efficient estimator for the estimation of transfer function model parameters. The estimation of model parameters was also attempted using least square fitting directly on the data, which equivalently is to solve the simultaneous Wiener-Hopf equations using auto and cross correlations of the involved time series. The results, however, were not as good and consistent and therefore are not presented here.

---

*In Table 4.2, 'Model Validity at 95% Level' was based on various statistical tests at 95% confidence level on comparison between observed and theoretical impulse response functions and frequency response functions, on cross correlations between the noise and input series and on randomness of the noise series. Question mark '?' indicates failure in any of the above tests to validate the proposed model. Standard deviations of estimates of time lags 'b' are not computed because they can take only integer values.

TABLE 4.2

Summary on Multivariate Watershed System Identification

Station Identification		05AB002			05AB021		
x_3: Runoff x_2: Precip x_1: Temp		Initial Estimate	Final Estimate	Std.Dev. Estimate	Initial Estimate	Final Estimate	Std.Dev. Estimate
$(1-d_1B)x_{3t} = (w_{10}-w_{11}B)x_{1t-b_1}$	d_1	0.58	0.65	0.05	0.80	0.73	0.05
	w_{10}	0.16	0.06	0.06	0.20	- 0.10	0.06
	w_{11}	0.24	0.23	0.06	0.28	0.24	0.06
$+ (w_{20}-w_{21}B)x_{2t-b_2}$	b_1	0	0		0	0	
	w_{20}	0.19	0.30	0.06	0.20	0.27	0.06
$+ \theta_1 z_{t-1} + a_t$	w_{21}	-0.08	-0.11	0.06	-0.06	-0.09	0.06
	b_2	0	0		0	0	
	θ_1	-0.24	-0.30	0.08	-0.36	-0.32	0.08
Model Validity at 95% Level			yes			yes	
Explained Residual Variance %			50			50	
$(1-d_1B-d_2B^2)x_{3t} =$	d_1	0.43	0.43	0.23	0.75	0.69	0.10
	d_2	0.21	0.16	0.17	0.06	0.02	0.07
$(w_{10}-w_{11}B-w_{12}B^2)x_{1t-b_1}$	w_{10}	0.16	0.06	0.06	0.20	0.10	0.06
	w_{11}	0.22	0.20	0.07	0.26	0.21	0.06
	w_{12}	0.002	0.08	0.08	0.007	0.06	0.06
$+ (w_{20}-w_{21}B-w_{22}B^2)x_{2t-b_2}$	b_1	0	0		0	0	
	w_{20}	0.19	0.31	0.06	0.20	0.30	0.06
$+ (\theta_1B+\theta_2B^2)z_t + a_t$	w_{21}	-0.11	-0.19	0.09	-0.07	-0.09	0.07
	w_{22}	-0.04	-0.005	0.02	-0.04	-0.03	0.06
	b_2	0	0		0	0	
	θ_1	-0.10	-0.10	0.24	-0.37	-0.32	0.12
	θ_2	-0.15	-0.08	0.10	-0.15	-0.10	0.09
Model Validity at 95% Level			yes			yes	
Explained Residual Variance %			52			52	

TABLE 4.2 (continued)

Summary on Multivariate Watershed System Identification

Station Identification		05GG001			05TD001		
x_3: Runoff x_1: Temp x_2: Precip		Initial Estimate	Final Estimate	Std.Dev. Estimate	Initial Estimate	Final Estimate	Std.Dev. Estimate
$(1-d_1B)x_{3t} = (w_{10}-w_{11})x_{1t-b_1}$	d_1	0.65	0.72	0.05	0.74	0.85	0.04
	w_{10}	0.13	0.10	0.05	-0.11	-0.09	0.04
	w_{11}	0.17	0.16	0.06	-0.06	-0.02	0.04
$+ (w_{20}-w_{21}B)x_{2t-b_2}$	b_1	0	0		1	1	
	w_{20}	0.24	0.25	0.06	0.16	0.05	0.04
	w_{21}	0.02	0.02	0.03	-0.08	-0.03	0.04
$+ \theta_1 z_{t-1} + a_t$	b_2	1	1		0	0	
	θ_1	-0.11	-0.18	0.09	-0.09	-0.22	0.08
Model Validity at 95% Level			yes			yes	
Explained Residual Variance %			52			67	
$(1-d_1B-d_2B^2)x_{3t} =$	d_1	0.69	0.57	0.09	0.69	0.78	0.20
	d_2	-0.02	0.20	0.09	0.09	0.10	0.20
$(w_{10}-w_{11}B-w_{12}B^2)x_{1t-b_1}$	w_{10}	0.13	0.08	0.05	-0.11	-0.09	0.04
	w_{11}	0.18	0.11	0.06	-0.05	-0.02	0.03
	w_{12}	0.08	0.13	0.06	0.01	0.06	0.04
$+ (w_{20}-w_{21}B-w_{22}B^2)x_{2t-b_2}$	b_1	0	0		1	1	
	w_{20}	0.24	0.23	0.06	0.16	0.06	0.04
$+ (\theta_1B+\theta_2B^2 z_t + a_t$	w_{21}	0.03	-0.02	0.05	-0.08	-0.05	0.04
	w_{22}	0.10	0.11	0.05	0.04	0.06	0.04
	b_2	1	1		0	0	
	θ_1	-0.13	0.01	0.05	-0.06	-0.16	0.21
	θ_2	0.004	-0.16	0.09	-0.09	-0.19	0.08
Model Validity at 95% Level			yes			yes	
Explained Residual Variance %			55			68	

TABLE 4.2 (continued)

Summary on Multivariate Watershed System Identification

Station Identification		08KA005			08KA007		
x_3: Runoff x_1: Temp x_2: Precip		Initial Estimate	Final Estimate	Std.Dev. Estimate	Initial Estimate	Final Estimate	Std.Dev. Estimate
$(1-d_1B)x_{3t} = (w_{10}-w_{11}B)x_{1t-b_1}$	d_1	0.35	0.54	0.08	0.27	0.48	0.10
	w_{10}	0.49	0.44	0.07	0.40	0.38	0.07
$+ (w_{20}-w_{21}B)x_{2t-b_2}$	w_{11}	0.12	0.22	0.08	0.03	0.19	0.07
	b_1	0	0		0	0	
$+ \theta_1 z_{t-1} + a_t$	w_{20}	0.20	0.15	0.07	0.13	0.09	0.07
	w_{21}	-0.21	-0.15	0.07	-0.19	-0.15	0.07
	b_2	0	0		0	0	
	θ_1	-0.02	-0.21	0.10	0.14	-0.08	0.12
Model Validity at 95% Level			yes			?	
Explained Residual Variance %			30			29	
$(1-d_1B-d_2B^2)x_{3t} =$	d_1	0.17	0.29	0.08	0.29	0.44	0.15
	d_2	0.10	0.40	0.09	-0.08	0.15	0.13
$(w_{10}-w_{11}B-w_{12}B^2)x_{1t-b_1}$	w_{10}	0.49	0.44	0.07	0.40	0.37	0.07
	w_{11}	0.03	0.12	0.07	0.04	0.14	0.08
$+ (w_{20}-w_{21}B-w_{22}B^2)x_{2t-b_2}$	w_{12}	-0.02	0.15	0.08	0.03	0.13	0.08
	b_1	0	0		0	0	
$+ (\theta_1B+\theta_2B^2)z_t + a_t$	w_{20}	0.20	0.14	0.07	0.13	0.09	0.06
	w_{21}	-0.24	-0.18	0.07	-0.19	-0.17	0.07
	w_{22}	-0.15	-0.05	0.06	-0.03	0.0005	0.004
	b_2	0	0		0	0	
	θ_1	0.14	-0.02	0.06	0.10	0.06	0.15
	θ_2	-0.06	-0.41	0.09	0.11	-0.21	0.10
Model Validity at 95% Level			yes			yes	
Explained Residual Variance %			35			32	

TABLE 4.2 (continued)

Summary on Multivariate Watershed System Identification

Station Identification		08KB001			08LA008		
x_3: Runoff x_2: Precip x_1: Temp		Initial Estimate	Final Estimate	Std.Dev. Estimate	Initial Estimate	Final Estimate	Std.Dev. Estimate
$(1-d_1B)x_{3t} = (w_{10}-w_{11}B)x_{1t-b_1}$	d_1	0.52	0.41	0.09	0.74	0.72	0.07
	w_{10}	0.41	0.36	0.07	0.17	0.10	0.06
$+ (w_{20}-w_{21}B)x_{2t-b_2}$	w_{11}	0.14	0.15	0.07	0.10	0.09	0.06
	b_1	0	0		0	0	
$+ \theta_1 z_{t-1} + a_t$	w_{20}	0.40	0.28	0.06	0.26	0.25	0.06
	w_{21}	-0.23	-0.22	0.07	0.02	0.05	0.05
	b_2	0	0		1	1	
	θ_1	-0.18	-0.05	0.11	-0.12	-0.10	0.09
Model Validity at 95% Level			yes			yes	
Explained Residual Variance %			38			52	
$(1-d_1B-d_2B^2)x_{3t} =$	d_1	0.34	0.03	0.04	0.44	0.39	0.13
	d_2	0.13	0.42	0.07	0.35	0.33	0.12
$(w_{10}-w_{11}B-w_{12}B^2)x_{1t-b_1}$	w_{10}	0.41	0.39	0.06	0.18	0.10	0.05
	w_{11}	0.06	-0.02	0.02	0.51	0.05	0.05
$+ (w_{20}-w_{21}B-w_{22}B^2)x_{2t-b_2}$	w_{12}	-0.05	0.08	0.06	0.04	0.05	0.05
	b_1	0	0		0	0	
$+ (\theta_1 \cancel{B+\theta_2} B^2)z_t + a_t$	w_{20}	0.40	0.31	0.06	0.26	0.27	0.06
	w_{21}	-0.30	-0.39	0.06	-0.05	-0.06	0.07
	w_{22}	-0.07	-0.13	0.07	0.03	-0.03	0.05
	b_2	0	0		1	1	
	θ_1	-0.02	0.23	0.07	0.20	0.23	0.13
	θ_2	-0.25	-0.50	0.08	-0.26	-0.26	0.11
Model Validity at 95% Level			yes			yes	
Explained Residual Variance %			44			53	

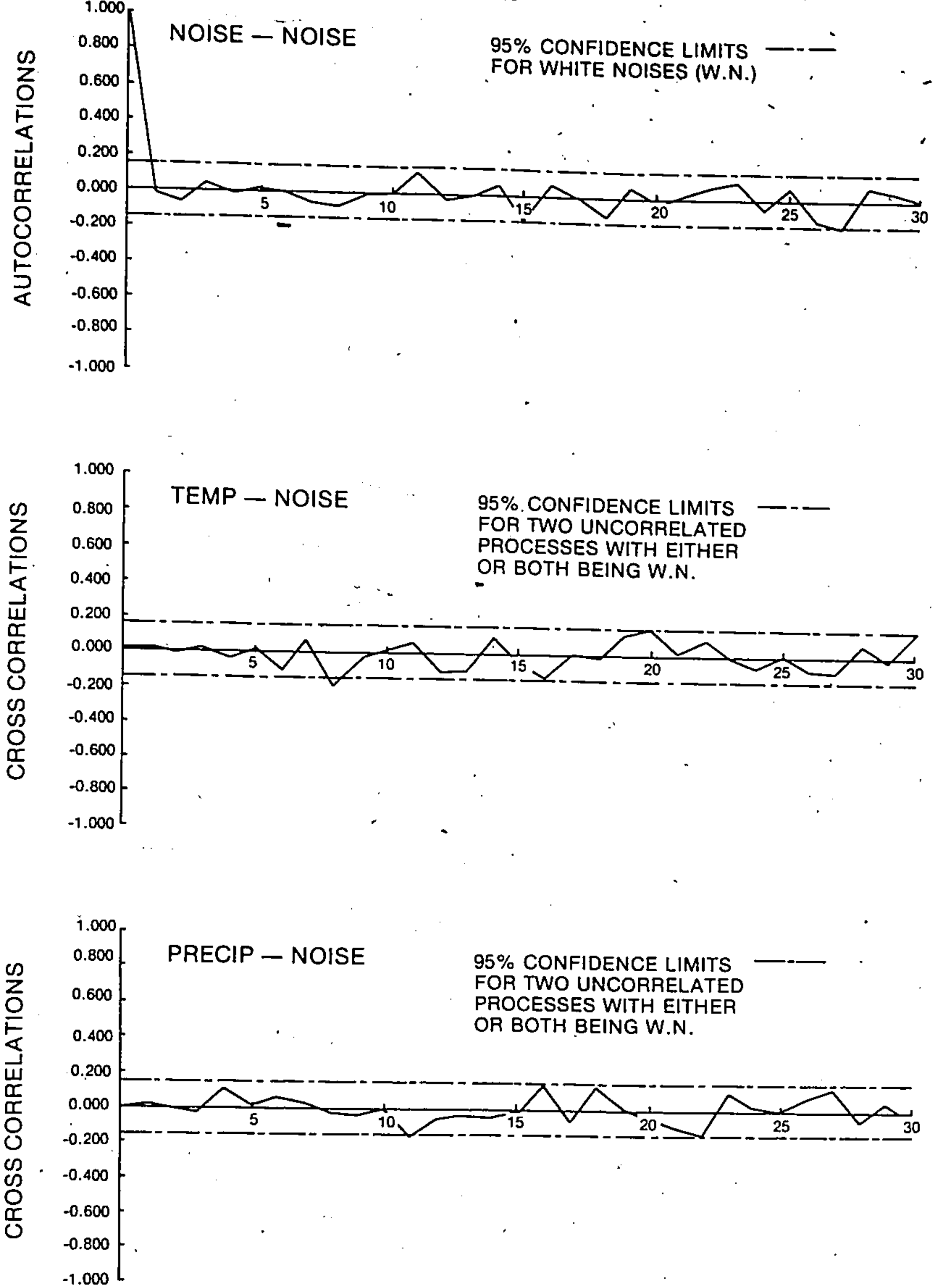


Figure 4.11 Auto- and Cross Correlations of Noise and between Noise and Inputs to 1st-Order ARMA Transfer Function Model - 05AB002 Willow Creek near Nolan

It is also observed from Table 4.2 that the improvement from fitting a second-order ARMA transfer function model instead of a first-order model to the watershed system of monthly runoff, temperature and precipitation series is not significant. By examining standard deviations of estimates obtained from eq. (4.46), higher order weights in the second-order model are relatively small and hardly to be of any statistical significance.

However, for the Pacific mountain stations, the initial and final estimates of transfer function parameters as well as the two sets of model parameters are not consistent. This may be attributed to the scarcity of the available meteorological data in these mountainous basins, with the meteorological network density from 445 to 897 (sq. mi. per station) and most of meteorological data sampled at lower to medium band of the elevation (1,465 ft. to 4,180 ft.).

It is further observed that the weight for the previous state of runoff (Table 4.2) is of the same order of magnitude as the autoregressive coefficient for the univariate runoff model (Table 3.4). This leads to the postulation that the disturbance term in the univariate model may be modelled by the terms at the right-hand side of the ARMA transfer function model. By comparing the explained variances in Tables 3.4 and 4.2, the improvement from coordinated use of the meteorological and hydrometric data is obvious. For watershed with large storage, i.e., 05TD001, however, the improvement is not pronounced. In this case, the large storage effect plays a dominant role in the generating process of system response and the contribution from external disturbances will be damped out before it becomes significant.

Under the assumption that the multivariate ARMA transfer function model is just the extension of a univariate ARMA process by introducing the contribution from meteorological inputs, some interesting findings can be made. Considering the case of ARMA(1,0) process fitted directly from the transformed runoff series, standard deviations* of estimates of autoregressive coefficients for the eight stations included in the multivariate system analysis were found respectively to be 0.0[illegible] 0.06, 0.06, 0.04, 0.07, 0.07, 0.07 and 0.06. Comparing these values with the corresponding values given in Table 4.2 for the first-order ARMA transfer function model, it can be observed that for Prairie stations the autoregressive coefficients obtained from the multivariate system analysis have smaller standard errors than those obtained from the univariate runoff analysis except for the large storage watershed, i.e., 05TD001, where the improvement is not significant. In contrast, the Pacific mountain stations all indicate higher standard errors in the multivariate case due to the noise introduced by insufficient sampling of meteorological data in these basins. Thus, with reasonable coverage of meteorological network in the area, the coordinated use of meteorological and hydrometric data also improves the estimation of parameters in the univariate runoff model.

In conclusion, a first-order ARMA transfer function model is considered adquate to model a watershed system of monthly runoff, temperature and precipitation time series. This coordinated use of the meteorological and hydrometric data has been shown to enhance the accuracy of estimation of runoff characteristics.

---

*The approximate variance of estimate of the coefficient of ARMA(1,0) process is given by

$$\mathrm{Var}(d_1) \approx \frac{1-d_1^2}{N},$$

where N is the number of observations (Box and Jenkins, 1970).

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

1. A generalized approach was demonstrated to establish the relationships between the linear stochastic differential equations (S.D.E.) in the continuous time and the autoregressive-moving average (ARMA) models in the discrete time. In general, for a m^{th} order S.D.E. model there exists an equivalent ARMA(m,m-1) discrete process.

2. For monthly runoff, temperature and precipitation time series, cyclic (or deterministic) components are effectively removed by subtracting the monthly means and variances from the raw time series.

3. For the transformed stationary runoff time series given by eq. (3.41), a first-order S.D.E. model or its equivalent first-order autoregressive or ARMA(1,0) model is adequate to model the underlying generating process.

4. Any attempt to fit a monthly runoff series by a higher order S.D.E. model or its equivalent ARMA process may lead to unstable model structure. On the other hand, the discrete ARMA(1,1) and second-order autoregressive, or ARMA(2,0), models can produce the same result but may be reduced to the simple ARMA(1,0) structure.

5. The results of the regional multiple regression analysis indicate that the coefficient K of the first-order S.D.E. bears the effect of watershed storage, which in turn is related to the local watershed physiographic factors. Residual runoff variance was found being also related to the watershed storage effect, whereas the gross noise variance of external disturbances was found being more influenced by the regional physiographic factors. Therefore, a watershed can be considered as a 'damped' system whereby the high frequency variations of input disturbances are filtered out through its storage effect.

6. The results of the regional multiple regression analysis also indicates the necessity to divide the whole study area into two statistical or model homogeneous regions, namely, Pacific region and Prairie region. The regression models were found not adequate in the Prairie region indicating the need for some additional physiographic factors concerning geology, soil type and vegetative cover, etc.

7. Varimax factor analysis is useful in reducing the number of predictor variables, detecting the dominant common factors and achieving the 'simple structure' for the subsequent multiple regression analysis. However, caution must be exercised in selecting the predictor variables to be included for the analysis on the basis of their possible physical influence to the criterion variable.

8. The transformed stationary temperature and precipitation time series showed no significant difference from a purely random process except for temperatures in the Pacific mountainous region indicating persistence in the time series.

9. Significant cross correlation exists between temperature and precipitation at the zero lag. Cross correlations between the runoff and the meteorological time series are in some cases not stable because of the scarcity of the available meteorological data. In general, meteorological series were found to lead the runoff series.

10. Most of the gain functions revealed a system of exponential decay or damped sine wave, which is the characteristic of an ARMA transfer function. Phase functions, however, did not resemble any of the known systems. Thus, a time domain system response function is required for the determination of physical time lags and for the construction of a time domain transfer function model.

11. Most of the significant coherencies and gains of the frequency response functions occur at low frequencies around the annual cycle. This suggests a 'damped' watershed system which transforms the meteorological inputs into the runoff with rich variance at low frequencies.

12. Impulse response function was found to be an efficient estimator for the estimation of transfer function parameters and for the determination of physical time lags.

13. The multivariate watershed system of monthly runoff, temperature and precipitation time series may be adequately represented by a first-order ARMA transfer function model. However, model instability may be introduced due to the insufficient spatial sampling of meteorological data.

14. The weight for the previous state of runoff in the multivariate system transfer function is of the same order of magnitude as the autoregressive coefficient in the univariate runoff model. This leads to the postulation that the disturbance term in the univariate model may be modelled by the terms at the right-hand side of the ARMA transfer function model.

15. Coordinated use of the meteorological and hydrometric data would enhance the accuracy of estimation of runoff characteristics. For watershed with large storage, the improvement was not pronounced. In this case, the large storage effect plays a dominant role in the generating process of system response and the contribution from external disturbances will be damped out before it becomes significant.

Recommendations

1. The approach for the identification of monthly runoff process may well be applied to the daily runoff time series which might reveal more the nature of the underlying runoff generating process.

2. Validation of the model was not carried out in this study because of time and budgetary limitations. It is always advisable to

perform the model validation by split sampling in both temperal and spatial sense.

3. The results of the regional multiple regression analysis indicate that additional physiographic factors concerning geology, soil type and vegetative cover, etc., need to be sampled especially in the Prairie region.

4. More rigorous statistical analysis and good-quality stations are required for more refined statistical regionalization.

5. For multivariate watershed hydrologic system modelling, a combined deterministic-statistical approach could be developed. For example, instead of considering temperature and precipitation as separate input variables, a transformed variable of net precipitation obtained by subtracting from the precipitation the amount of evapotranspiration computed from some known physical or empirical equations such as Penman's equation (Penman, 1940), Turc's formula (Turc, 1954), etc., ~~may~~ be used as the input variable to the watershed system model. This will establish more meaningful link between the univariate and the multivariate system models and provide more insight in the physical significance of the proposed stochastic system model.

6. The identification of the watershed hydrologic system should not be considered as the ultimate goal of the hydrologic system analysis. The practical application of the proposed models is at least two-fold: to estimate the runoff characteristics or synthesize the run-

off time series at the ungaged site through the correlation regression between model parameters and watershed physiographic characteristics (Pentland and Cuthberd, 1971), and to make short-term runoff forecast from the proposed system models following the procedures demonstrated by Box and Jenkins (1970).

LIST OF REFERENCES

Adamowski, K., 1969. Stochastic analysis of daily river flow time series. Ph.D. Thesis, Univ. of Guelph.

Adamowski, K., and C.T. Hsu, 1974. Stochastic linear prediction of hydrologic time series. Presented at the AGU Fall Annual Meeting, San Francisco, December.

Amorocho, J., 1967. The non-linear prediction problem in the study of the runoff cycle. Water Resources Res., Vol. 3, No. 3, pp.861-880.

Amorocho, J., and A. Brandstetter, 1971. Determination of non-linear functional response functions in rainfall-runoff processes. Water Resources Res., Vol. 7, No. 5, pp.1087-1101.

Anderson, R.L., 1942. Distribution of the serial correlation coefficient. Annals of Math. Stat., Vol. 13, pp.1-13.

Bartlett, M.S., 1937. Properties of sufficiency and statistical tests. Proc. Royal Soc., A160, pp.268-282.

Bartlett, M.S., 1946. On the theoretical specification and sampling properties of autocorrelated time series. J. Royal Stat. Soc., B8, pp.27-41.

Bartlett, M.S., 1950. Tests of significance in factor analysis. British J. Psych. (Stat. Sec.) III, pp.77-85.

Bartlett, M.S., 1966. Stochastic processes. Cambridge Univ. Press, Cambridge.

Bidwell, V.G., 1971. Regression analysis of non-linear catchment system. Water Resources Res., Vol. 7, No. 5, pp.1118-1126.

Blackman, R.B., and J.W. Tukey, 1958. The measurement of power spectra. Dover, N.Y.

Box, G.E.P., and G.M. Jenkins, 1970. Time series analysis, forecasting and control. Holden Day, San Francisco.

Bryson, R.A., 1966. Air masses, streamlines and the Boreal forest. Geograph. Bull., Vol. 8, No. 3, pp.228-269.

Carroll, J.B., 1953. An analytical solution for approximating simple structure in factor analysis. Psychometrika, Vol. 18, No. 1, pp.23-38.

Cavadias, G.S., 1964. Method of analysis and interpretation. Proc. Hydrology Sym., No. 4, pp.275-284.

Cavadias, G.S., 1966. River flow as a stochastic process. Proc. Hydrology Sym., No. 5, pp.315-359.

Cavadias, G.S., and R. Brisebois, 1971. Long-term forecasting of snow-melt runoff. Proc. Hydrology Sym., No. 8, Vol. 1, pp.238-257.

Chiu, C.L., and R.P. Bitter, 1969. Linear time-varying model of rainfall-runoff relation. Water Resources Res., Vol. 5, No. 2, pp.426-437.

Chow, V.T., and S. Rameseshan, 1965. Sequential generation of rainfall and runoff data. J. Hydraulics Div., ASCE, Vol. 91, No. HY4, pp.205-233.

Chow, V.T., 1969. Systems approaches in hydrology and water resources. Proc. the 1st Intern. Seminar for Hydrology Professors, Urbana, Vol. 1, pp.490-509.

Chow, V.T., and S.J. Kareliotis, 1970. Analysis of stochastic hydrologic systems. Water Resources Res., Vol. 6, No. 6, pp.1569-1582.

Chow, V.T., and T. Prasad, 1972. Theory of stochastic modelling of water-shed systems. J. Hydrology, Vol. 15, No. 4, pp.261-284.

Coulson, A., 1967. Estimating runoff in southern Ontario. Tech. Bull. No. 7, Inland Waters Branch, Department of Energy, Mines and Resources, Ottawa.

Crawford, N.H., and R.K. Linsley, 1964. A conceptual model of the hydrologic cycle. IASH Publ., No. 63, pp.573-587.

Diskin, M.H., and A. Boneh, 1973. Determination of optimal kernels for second-order stationary surface runoff systems. Water Resources Res., Vol. 9, No. 2, pp.311-325.

Dooge, J.C.I., 1972. Mathematical models of hydrologic systems. Proc. Intern. Sym. Modelling Techniques in Water Resources Systems, Vol. 1, pp.171-[illegible]89.

Dowling, J.M., 1974. A note on the use of spectral analysis to detect leads and lags in annual cycles of water quality. Water Resources Res., Vol. 10, No. 2, pp.343-344.

Eagleson, P.S., and W.J. Shack, 1966. Some criteria for the measurement of rainfall and runoff. Water Resources Res., Vol. 2, No. 3, pp.427-436.

Eiselstein, L.M., 1967. A principal component analysis of surface run-off data from a New Zealand alpine watershed. Proc. Intern. Hydrology Sym., Fort Collins, Vol. 1, pp.479-489.

Eriksson, E., 1970a. Groundwater time series, an exercise in stochastic hydrology. Nordic Hydrology, Vol. 1, No. 3, pp.181-205.

Eriksson, E., 1970b. Cross-spectrum analysis of groundwater levels in an esker. Nordic Hydrology, Vol. 1, No. 4, pp.245-259.

Fiering, M.B., 1974. Multivariate technique for synthetic hydrology. J. Hydraulics Div., ASCE, Vol. 90, No. HY5, pp.43-59.

Granger, C.W.J., and M. Hatanaka, 1964. Spectral analysis of economic time series. Princeton Univ. Press, Princeton.

Hald, A., 1952. Statistical theory with engineering applications. John Wiley, New York.

Hannan, E.J., 1962. Time series analysis. Methuen, London.

Hannan, E.J., 1970. Multiple time series. John Wiley, N.Y.

Harman, H.H., 1960. Modern factor analysis. Univ. Chicago Press, Chicago.

Hsu, C.T., 1970. A mathematical model for continuous basin runoff simulation. M.Sc. Thesis, Queen's Univ.

Ingledow, T. and Associates, Ltd., 1969. British Columbia hydrometric network study. Report for the Canada Department of Energy, Mines and Resources.

Jenkins, G.M., and D.G. Watts, 1968. Spectral analysis and its application. Holden Day, San Francisco.

Kaiser, H.F., 1958. The Varimax criterion for analytic rotation in factor analysis. Psychometrika, Vol. 23, No. 3, pp.187-200.

Kendall, M.G., 1961. A course in multivariate analysis. Charles Griffin, London.

Kisiel, C.C., 1967. Transformation of deterministic and stochastic processes in hydrology. Proc. Intern. Hydrology Sym., Fort Collins, Vol. 1, pp.600-607.

Kisiel, C.C., 1969. Time series analysis of hydrologic data. Advances in Hydroscience, Vol. 5, pp.1-119.

Lawley, D.N., and A.E. Maxwell, 1963. Factor analysis as a statistical method. Butterworths, London.

Matalas, N.C., 1966. Some aspects of time series analysis in hydrologic studies. Proc. Hydrology Sym., No. 5, pp.271-313.

McKay, G.A., and W. Stichling, 1961. Evaporation computations for Prairie reservoirs. Proc. Hydrology Sym., No. 2, pp.135-167.

Meyboom, P., 1962. Patterns of groundwater flow in the Prairie profile. Proc. Hydrology Sym., No. 3, pp.5-33.

Nash, J.E., 1960. A unit hydrograph study, with particular reference to British catchments. Proc. Inst. Civil Eng., Vol. 17, pp.249-271.

O'Connell, P.E., 1971. A simple stochastic modelling of Hurst's Law. IASH Intern. Sym. Math. Models in Hydrology, Warsaw, pp.327-358.

O'Donnell, T., 1960. Instantaneous unit hydrograph derivation by harmonic analysis. IASH Publ., No. 51, pp.546-557.

Overton, D.E., 1968. Eigenvector analysis of geomorphic interrelations of small agricultural watersheds. Presented at the ASAE Winter Meeting, Chicago, December.

Penman, H.L., 1948. Natural evaporation from open waters, bare soil and grass. Proc. Royal Soc., A193, pp.120-145.

Pentland, R.L., and D.R. Cuthbert, 1971. Operational hydrology for ungauged streams by the grid squares technique. Water Resources Res., Vol. 7, No. 2, pp.283-291.

Quimpo, R.G., and V. Yevjevich, 1967. Stochastic description of daily river flows. Proc. Intern. Hydrology Sym., Fort Collins, Vol. 1, pp.290-297.

Quimpo, R.G., 1971. Kernels of stochastic linear hydrologic systems. Proc. First U.S.-Japan Bilateral Seminar in Hydrology, Honolulu, pp.163-185.

Quimpo, R.G., 1973. Link between stochastic and parametric hydrology. J. Hydraulics Div., ASCE, Vol. 99, No. HY3, pp.461-470.

Rice, R.M., 1967. Multivariate methods useful in hydrology. Proc. Intern. Hydrology Sym., Fort Collins, Vol. 1, pp.471-477.

Rodriguez-Iturbe, I., 1967. The application of cross-spectral analysis to hydrologic time series. Hydrology Paper No. 24, Colorado State Univ.

Rodriguez-Iturbe, I., and V. Yevjevich, 1967. Sunspots and hydrologic time series. Proc. Intern. Hydrology Sym., Fort Collins, Vol. 1, pp.397-405.

Rodriguez-Iturbe, I., and C.F. Nordin, 1968. Time series analysis of water and sediment discharges. IASH Bull., Vol. 13, No. 2, pp.69-84.

Roesner, L.A., and V. Yevjevich, 1966. Mathematical models for time series of monthly precipitation and monthly runoff. Hydrology Paper No. 15, Colorado State Univ.

Rosenbrock, H.H., 1960. An automatic method for finding the greatest or least value of a function. Computer J., Vol. 3, pp.173-184.

Sharps, A.L., A.E. Gibbs, W.J. Owen and B. Harris, 1960. Application of the multiple regression approach in evaluating parameters affecting water yields of river basins. J. Geophys. Res., Vol. 65, pp.1273-1286.

Shawinigan Eng. Co. Ltd., 1970. Hydrometric network planning study for western and northern Canada. Report for the Canada Department of Energy, Mines and Resources.

Snyder, W.M., 1962. Some possibilities for multivariate analysis in hydrologic studies. J. Geophys. Res., Vol. 67, No. 2, pp.721-729.

Solomon, S.I., 1972a. Pilot hydrologic forecasting study for southern Ontario. Report for the Canada Department of the Environment.

Solomon, S.I., 1972b. Joint mapping. Sec. III-2.1, WMO Casebook on Hydrological Network Design Practice.

Spolia, S.K., and S. Chander, 1974. Modelling of surface runoff systems by an ARMA model. J. Hydrology, Vol. 22, No. 3/4, pp.317-332.

Stammers, W.N., 1966. The application of multivariate technique in hydrology. Proc. Hydrology Sym., No. 5, pp.255-269.

Stoddart, R.B.L., and W.E. Watt, 1970. Flood frequency prediction for intermediate drainage basins in southern Ontario. C.E. Report No. 66, Queen's Univ.

Thomas, H.A., Jr., and M.B. Fiering, 1962. Mathematical synthesis of streamflow sequences. Chap. 12, Design of Water Resources Systems by Maass et al., Harvard Univ. Press.

Thurnston, L.L., 1947. Multiple factor analysis. Univ. Chicago Press, Chicago.

Turc, L., 1954. Le bilan d'eau des sols. Relations entre les precipitations, l'evaporation, et l'ecoulement. Ann. Agron., Vol. 5, No. 4, pp.491-595.

Wallis, J.R., 1965. Multivariate statistical methods in hydrology ... a comparison using data of known functional relationship. Water Resources Res., Vol. 1, No. 4, pp.447-461.

Wallis, J.R., 1968. Factor analysis in hydrology - an agnostic view. Water Resources Res., Vol. 4, No. 3, pp.521-527.

Watt, W.E., and C.T. Hsu, 1971. Continuous basin snowmelt runoff simulation. Proc. Hydrology Sym., No. 8, Vol. 2.

Wiener, N., 1949. The extrapolation, interpolation and smoothing of stationary time series. MIT Press, Cambridge.

Wilde, D.J., 1964. Optimum seeking method. Prentice-Hall, N.J.

Wong, S.T., 1963. A multivariate statistical model for predicting mean annual flood in New England. Annals of Assoc. Amer. Geograph., Vol. 53, No. 3, pp.298-311.

Yaglom, A.M., 1962. Theory of stationary random functions. Prentice-Hall, N.J.

Yevjevich, V., 1964. Fluctuations of wet and dry years, part II: analysis by serial correlation. Hydrology Paper No. 4, Colorado State Univ.

Yevjevich, V., 1972. Structural analysis of hydrologic time series. Hydrology Paper No. 56, Colorado State Univ.

Yevjevich, V., 1974. Determinism and stochasticity in hydrology. J. Hydrology, Vol. 22, No. 3/4, pp.225-238.

Zand, S.M., and J.A. Harder, 1973. Application of non-linear system identification to the lower Mekong River, Southeast Asia. Water Resources Res., Vol. 9, No. 2, pp.290-297.

APPENDIX A

WATERSHED PHYSIOGRAPHIC CHARACTERISTICS FOR REGIONAL REGRESSION ANALYSIS

The watershed physiographic characteristics were computed as the area weighted average of the physiographic characteristics of all 10x10 km squares enclosed in a watershed drainage boundary. These square-grid physiographic factors were derived from the information contained on the 1:250,000 Universal Transverse Mercator map sheets.

The factors included in the presented study were:

1. Latitude index defined as the coordinate of the square in the south-north direction starting from the most southern line of squares (stored as the number of squares). The origin of the array is located at the grid square with a military grid reference of 7UDD00.
2. Average elevation of each square estimated as the average of the elevations of its four corners and stored as feet.
3. Local slope estimated as the average slope of two plans: one defined by the elevations of the two southern corners and the left-hand northern corner and the other defined by the elevations of the two southern corners and the right-hand northern corner; stored as the tangent of the angle of inclination in percentage.
4. Slope azimuth index in the north direction defined as an index related to the angle (in the clockwise sense) between the north

direction and the horizontal projection of the line of inclination of the local slope plan and expressed in integers from 1 to 5 as follows:

$= 1$ for $337^\circ30' < \alpha < 360^\circ00'$, $0^\circ00' < \alpha < 22^\circ30'$;

$= 2$ for $22^\circ30' < \alpha < 67^\circ30'$, $292^\circ30' < \alpha < 337^\circ30'$;

$= 3$ for $67^\circ30' < \alpha < 112^\circ30'$, $247^\circ30' < \alpha < 292^\circ30'$;

$= 4$ for $112^\circ30' < \alpha < 157^\circ30'$, $202^\circ30' < \alpha < 247^\circ30'$,

$= 5$ for $157^\circ30' < \alpha < 202^\circ30'$

5, 6 and 7. Slope azimuth indices in NW, W and SW directions defined similarly as the azimuth index in the north direction.

8. Distance to sea in the north direction defined as the distance from the ocean in the north to the given square, was computed as a weighted average of the distances to the sea of the given square and of two pairs of adjacent squares on each side in a direction perpendicular to that for which the value of the distance is sought (stored as kilometers).

9, 10 and 11. Distances to sea in NW, W and SW directions defined in a similar manner as the distance to sea in the north direction.

12. Lake and swamp area index defined as the percentage area of lakes and swamps in a given square.

13. Forest area index defined as the percentage area of forests in a given square.

14. Barrier height in the north direction defined as the difference between the average elevation of the square and the highest

elevation encountered in the north direction from the given square until the ocean is reached, was determined for each square as the weighted average of the barrier heights for the given square and of the two pairs of adjacent squares on each side in a direction perpendicular to that for which the value of the barrier height is sought (stored as feet).

15, 16 and 17. Barrier heights in NW, W and SW directions defined similarly as the barrier height in the north direction.

18. Shield effect to the north defined as the sum of the elevation differential of all ascending stretches of terrain encountered when travelling from the ocean shore at the north to the given square, was computed as a weighted average of shield effects of the given square and two pairs of adjacent squares on each side in a direction perpendicular to that for which the value of the shield effect is sought (stored as feet).

19, 20 and 21. Shield effects to NW, W and SW directions defined in the same manner as the shield effect to the north.

22. Regional slope in the SE-NW direction defined as a slope which takes into account the general configuration of the terrain by computing the slope in the given direction for the total reach for which the elevation is increasing, or decreasing, continuously in vertical increments of 1000 feet. When going in the SE-NW direction, slopes were considered positive for increases in elevation and negative for decreases in elevation (stored as feet/km).

23 and 24. Regional slopes in E-W and NE-SW directions defined similarly as the regional slopes in SE-NW direction.

25. In addition to the above square-grid based physiographic characteristics, drainage area of a watershed was obtained directly from the Water Survey of Canada publication (stored as sq. mi.).

APPENDIX B

VARIANCE OF DISCRETIZED SYSTEM INTEGRAL

First-Order S.D.E.

For a first-order S.D.E. (3.13), namely,

$$(1+KD)y(t) = x(t), \tag{B.1}$$

where x(t) is a random impulse function, it has the system representation (3.15),

$$y(t) = \frac{1}{K}\int_{-\infty}^{t} e^{-(t-u)/K}x(u)du. \tag{B.2}$$

The relationship between σ_y^2 and σ_x^2 and the autocorrelation function of Y(t) are given by equations (3.17) and (3.18) respectively,

$$\sigma_y^2 = \frac{\sigma_x^2}{2K} \tag{B.3}$$

and

$$\rho_y(u) = e^{-u/K}. \tag{B.4}$$

The discretized form of eq. (B.2) is given by eq. (3.16),

$$(1-e^{-1/K}B)y(t) = \frac{1}{K}\int_{t-1}^{t} e^{-(t-u)/K}x(u)du. \tag{B.5}$$

For the discretized equation (B.5), the variance of the term at the right-hand side of the equal sign can be shown to be

$$E\left(\frac{1}{K^2}\int_{t-1}^{t}\int_{t-1}^{t} e^{-(t-u)/K}e^{-(t-v)/K}x(u)x(v)\,dudv\right)$$

$$= \frac{1}{K^2}\int_{t-1}^{t}\int_{t-1}^{t} e^{-(2t-u-v)/K}E(x(u)x(v))\,dudv$$

$$= \frac{1}{K^2}\int_{t-1}^{t} e^{-2(t-u)/K}\sigma_x^2\,du$$

$$= \frac{\sigma_x^2}{2K}(1-e^{-2/K}). \qquad \text{(B.6)}$$

If eq. (B.5) is rewritten in the form of

$$(1-d_1B)y_n = w_0x_n, \quad \text{where } d_1 = e^{-1/K} \text{ and } |d_1| < 1; \qquad \text{(B.7)}$$

it can be shown and from eq. (B.3) that the following relationship exists:

$$w_0^2\sigma_x^2 = \sigma_y^2(1-d_1^2)$$

$$= \frac{\sigma_x^2}{2K}(1-e^{-2/K}). \qquad \text{(B.8)}$$

Therefore, from equations (B.6) and (B.8), $\frac{1}{K}\int_{t-1}^{t} e^{-(t-u)/K}x(u)du$ may be approximated in the discrete form by w_0x_n, where

$$w_0 = \sqrt{(1-e^{-2/K})/2K}, \qquad \text{(B.9)}$$

and X_n is an independent process with zero mean and variance equal to σ_x^2. This approximation preserves the mean and variance of the increment $\frac{1}{K}\int_{t-1}^{t} e^{-(t-u)/K}x(u)du$.

Second-Order S.D.E. with Distinct Roots

For a second-order S.D.E. (3.22) with distinct roots,

$$(D^2+a_1D+a_2)y(t) = x(t) \tag{B.10}$$

where X(t) is assumed to be a random impulse function, it has the system representation given by eq. (3.24)

$$y(t) = \int_{-\infty}^{t} \frac{e^{-\mu_1(t-u)}-e^{-\mu_2(t-u)}}{\mu_2-\mu_1} x(u)du, \tag{B.11}$$

$-\mu_1$ and $-\mu_2$ being the distinct roots of the characteristic equation $D^2+a_1D+a_2 = 0$.

The relationship between σ_y^2 and σ_x^2 and the autocorrelation function $\rho_y(u)$ are given by equations (3.26) and (3.27) respectively,

$$\sigma_y^2 = \frac{\sigma_x^2}{2a_1a_2} \tag{B.12}$$

and

$$\rho_y(u) = \frac{\mu_2e^{-\mu_1u}-\mu_1e^{-\mu_2u}}{\mu_2-\mu_1}. \tag{B.13}$$

The discrete time equivalent of eq. (B.11) is given by eq. (3.25)

$$(1-(e^{-\mu_1}+e^{-\mu_2})B+e^{-(\mu_1+\mu_2)}B^2)y(t)$$

$$= \int_{t-1}^{t} \frac{e^{-\mu_1(t-u)}-e^{-\mu_2(t-u)}}{\mu_2-\mu_1} x(u)du$$

$$-e^{-(\mu_1+\mu_2)} \int_{t-2}^{t-1} \frac{e^{-\mu_1(t-2-u)} - e^{-\mu_2(t-2-u)}}{\mu_2-\mu_1} x(u)du. \qquad (B.14)$$

The variance of the term at the right-hand side of the discretized equation (B.14) can be shown to be

$$E(\int_{t-1}^{t} \frac{e^{-\mu_1(t-u)} - e^{-\mu_2(t-u)}}{\mu_2-\mu_1} x(u)du$$

$$-e^{-(\mu_1+\mu_2)} \int_{t-2}^{t-1} \frac{e^{-\mu_1(t-2-u)} - e^{-\mu_2(t-2-u)}}{\mu_2-\mu_1} x(u)du)^2$$

$$= \frac{1}{(\mu_2-\mu_1)^2} (\int_{t-1}^{t} \int_{t-1}^{t} (e^{-\mu_1(t-u)} - e^{-\mu_2(t-u)})(e^{-\mu_1(t-v)} - e^{-\mu_2(t-v)})$$

$$E(x(u)x(v))dudv$$

$$-2e^{-(\mu_1+\mu_2)} \int_{t-1}^{t} \int_{t-2}^{t-1} (e^{-\mu_1(t-u)} - e^{-\mu_2(t-u)})$$

$$(e^{-\mu_1(t-2-v)} - e^{-\mu_2(t-2-v)})E(x(u)x(v))dudv$$

$$+e^{-2(\mu_1+\mu_2)} \int_{t-2}^{t-1} \int_{t-2}^{t-1} (e^{-\mu_1(t-2-u)} - e^{-\mu_2(t-2-u)})$$

$$(e^{-\mu_1(t-2-v)} - e^{-\mu_2(t-2-v)})E(x(u)x(v))dudv)$$

$$= \frac{\sigma_x^2}{(\mu_2-\mu_1)^2} (\int_{t-1}^{t} (e^{-\mu_1(t-u)} - e^{-\mu_2(t-u)})^2 du$$

$$+e^{-2(\mu_1+\mu_2)}\int_{t-2}^{t-1}(e^{-\mu_1(t-2-u)}-e^{-\mu_2(t-2-u)})du)$$

$$=\frac{\sigma_x^2}{(\mu_2-\mu_1)^2}\left(\frac{1}{2\mu_1}(1-e^{-2\mu_1})-\frac{2}{\mu_1+\mu_2}(1-e^{-(\mu_1+\mu_2)})+\frac{1}{2\mu_2}(1-e^{-2\mu_2})\right.$$

$$+\frac{e^{-2(\mu_1+\mu_2)}}{2\mu_1}(e^{2\mu_1}-1)-\frac{2e^{-2(\mu_1+\mu_2)}}{\mu_1+\mu_2}(e^{\mu_1+\mu_2}-1)$$

$$\left.+\frac{e^{-2(\mu_1+\mu_2)}}{2\mu_2}(e^{2\mu_2}-1)\right)$$

$$=\frac{\sigma_x^2}{2\mu_1\mu_2(\mu_1+\mu_2)(\mu_2-\mu_1)^2}\left((\mu_2-\mu_1)^2-\mu_1\mu_2(e^{-\mu_1}-e^{-\mu_2})^2\right.$$

$$-(\mu_2e^{-\mu_1}-\mu_1e^{-\mu_2})^2-e^{-2(\mu_1+\mu_2)}(\mu_2-\mu_1)^2$$

$$\left.+\mu_1\mu_2(e^{-\mu_1}-e^{-\mu_2})^2+(\mu_1e^{-\mu_1}-\mu_2e^{-\mu_2})^2\right)$$

$$=\frac{\sigma_x^2}{2\mu_1\mu_2(\mu_1+\mu_2)(\mu_2-\mu_1)}\left(\mu_2-\mu_1-e^{-2(\mu_1+\mu_2)}(\mu_2-\mu_1)\right.$$

$$\left.-(\mu_1+\mu_2)(e^{-2\mu_1}-e^{-2\mu_2})\right). \qquad (B.15)$$

Rewriting eq. (B.14) in the form of

$$(1-d_1B-d_2B^2)y_n = (w_0-w_1B)x_n\ , \qquad (B.16)$$

where $d_1=e^{-\mu_1}+e^{-\mu_2}$, $d_2=-e^{-(\mu_1+\mu_2)}$ and $|d_2|<1$, $d_2+d_1<1$, $d_2-d_1<1$; it can be shown from equations (B.16), (B.12) and (B.13) that

$$
\begin{aligned}
(w_0^2+w_1^2)\sigma_x^2 &= (1+d_1^2-d_2^2-2d_1\rho_1)\sigma_y^2 \\
&= \Big(1+(e^{-\mu_1}+e^{-\mu_2})^2-e^{-2(\mu_1+\mu_2)} \\
&\qquad + 2(e^{-\mu_1}+e^{-\mu_2})\,\frac{\mu_2 e^{-\mu_1}-\mu_1 e^{-\mu_2}}{\mu_2-\mu_1}\Big)\sigma_y^2 \\
&= \frac{\sigma_y^2}{\mu_2-\mu_1}\Big(\mu_2-\mu_1-e^{-2(\mu_1+\mu_2)}(\mu_2-\mu_1) \\
&\qquad - (\mu_1+\mu_2)(e^{-2\mu_1}-e^{-2\mu_2})\Big) \\
&= \frac{\sigma_x^2}{2\mu_1\mu_2(\mu_1+\mu_2)(\mu_2-\mu_1)}\Big(\mu_2-\mu_1-e^{-2(\mu_1+\mu_2)}(\mu_2-\mu_1) \\
&\qquad - (\mu_1+\mu_2)(e^{-2\mu_1}-e^{-2\mu_2})\Big). \qquad \text{(B.17)}
\end{aligned}
$$

From equations (B.15) and (B.17), the random increments at the right-hand side of eq. (B.14) may be approximated in the discrete form by w_0x_n and w_1x_{n-1}, where

$$
w_0 = \frac{1}{\mu_2-\mu_1}\sqrt{\frac{1}{2\mu_1}(1-e^{-2\mu_1}) - \frac{2}{\mu_1+\mu_2}(1-e^{-(\mu_1+\mu_2)}) + \frac{1}{2\mu_2}(1-e^{-2\mu_2})}, \qquad \text{(B.18a)}
$$

and $w_1 = \frac{e^{-(\mu_1+\mu_2)}}{\mu_2-\mu_1} \sqrt{\frac{1}{2\mu_1}(e^{2\mu_1}-1) - \frac{2}{\mu_1+\mu_2}(e^{\mu_1+\mu_2}-1) + \frac{1}{2\mu_2}(e^{2\mu_2}-1)}$; (B.18b)

and X_n is an independent random process with zero mean and variance equal to σ_x^2.

Second-Order S.D.E. with Equal Roots

For a second-order S.D.E. (3.31) with equal roots,

$$(D^2+2\mu D+\mu^2)y(t) = x(t) \tag{B.19}$$

where X(t) is assumed to be a random impulse function, its system representation is given by eq. (3.33),

$$y(t) = \int_{-\infty}^{t} (t-u)e^{-\mu(t-u)}x(u)du. \tag{B.20}$$

The relationship between σ_y^2 and σ_x^2 and the autocorrelation function $\rho_y(u)$ are given respectively by equations (3.35) and (3.36) as follows:

$$\sigma_y^2 = \frac{\sigma_x^2}{4\mu^3} \tag{B.21}$$

and

$$\rho_y(u) = e^{-\mu u}(1+\mu u) \tag{B.22}$$

The discretized form of eq. (B.20) is given by eq. (3.34),

$$(1-2e^{-\mu}B+e^{-2\mu}B^2)y(t)=\int_{t-1}^{t}(t-u)e^{-\mu(t-u)}x(u)du-e^{-2\mu}\int_{t-2}^{t-1}(t-2-u)e^{-\mu(t-2-u)}x(u)du. \tag{B.23}$$

The variance of the random increments at the right-hand side of the discretized equation (B.23) is

$$E\left(\int_{t-1}^{t}(t-u)e^{-\mu(t-u)}x(u)du-e^{-2\mu}\int_{t-2}^{t-1}(t-2-u)e^{-\mu(t-2-u)}x(u)du\right)^2$$

$$=\int_{t-1}^{t}\int_{t-1}^{t}(t-u)(t-v)e^{-\mu(t-u)}e^{-\mu(t-v)}E(x(u)x(v))dudv$$

$$-2e^{-2\mu}\int_{t-1}^{t}\int_{t-2}^{t-1}(t-u)(t-2-v)e^{-\mu(t-u)}e^{-\mu(t-2-v)}E(x(u)x(v))dudv$$

$$+e^{-4\mu}\int_{t-2}^{t-1}\int_{t-2}^{t-1}(t-2-u)(t-2-v)e^{-\mu(t-2-u)}e^{-\mu(t-2-v)}E(x(u)x(v))dudv$$

$$=\sigma_x^2\left(\int_{t-1}^{t}(t-u)^2e^{-2\mu(t-u)}du+e^{-4\mu}\int_{t-2}^{t-1}(t-2-u)^2e^{-2\mu(t-2-u)}du\right)$$

$$=\sigma_x^2\left(\frac{-1}{2\mu}e^{-2\mu}-\frac{1}{2\mu^2}e^{-2\mu}-\frac{1}{4\mu^3}e^{-2\mu}+\frac{1}{4\mu^3}+\frac{e^{-4\mu}}{2\mu}e^{2\mu}-\frac{e^{-4\mu}}{2\mu^2}e^{2\mu}-\frac{e^{-4\mu}}{4\mu^3}+\frac{e^{-4\mu}}{4\mu^3}e^{2\mu}\right)$$

$$=\frac{\sigma_x^2}{4\mu^3}(1-4\mu e^{-2\mu}-e^{-4\mu}). \tag{B.24}$$

For the alternative form of eq. (B.23)

$$(1-d_1B-d_2B^2)y_n = (w_0-w_1B)x_n, \tag{B.25}$$

where $d_1=2e^{-\mu}$, $d_2=e^{-2\mu}$ and $|d_2|<1$, $d_2+d_1<1$, $d_2-d_1<1$; it can be shown from equations (B.25), (B.21) and (B.22) that

$$(w_0^2+w_1^2)\sigma_x^2 = (1+d_1^2-d_2^2-2d_1\rho_1)\sigma_y^2$$

$$= (1+4e^{-2\mu}-e^{-4\mu}-4e^{-2\mu}(1+\mu)\sigma_y^2$$

$$= \frac{\sigma_x^2}{4\mu^3}(1-4\mu e^{-2\mu}-e^{-4\mu}). \qquad (B.26)$$

From equations (B.24) and (B.26), the discrete time equivalent of the random increments at the right hand side of eq. (B.23) may be estimated in the weak statistical sense by $w_0 x_n$ and $w_1 x_{n-1}$, where

$$w_0 = \sqrt{\frac{-1}{2\mu} e^{-2\mu} - \frac{1}{2\mu^2} e^{-2\mu} - \frac{1}{4\mu^3} e^{-2\mu} + \frac{1}{4\mu^3}}\ , \qquad (B.27a)$$

and $$w_1 = e^{-2\mu}\sqrt{\frac{1}{2\mu} e^{2\mu} - \frac{1}{2\mu^2} e^{2\mu} - \frac{1}{4\mu^3} + \frac{1}{4\mu^3} e^{2\mu}}; \qquad (B.27b)$$

and X_n is an independent random process with zero mean and variance equal to σ_x^2.

APPENDIX C

VARIMAX ROTATION FACTOR ANALYSIS

Introduction

In operational hydrology, problems often arise in predicting the effect on runoff characteristics from changes in watershed physiographic characteristics, in constructing a rainfall-runoff model to synthesize a unit hydrograph for an ungaged watershed, etc. Suitable statistical techniques would be needed to correlate the runoff characteristics or the model parameters with many physiographic characteristics of watersheds. In this connection, multivariate statistical methods which describe the statistical behaviour of a number of individuals or objects based on multiple measurements have been used by hydrologists to achieve the above purposes. General description of multivariate methods and their possible applications were given by Wallis (1965), Stammers (1966) and Rice (1967). In the following sections, some commonly used multivariate methods for hydrologic analysis, namely, multiple regression analysis, principal component analysis and factor analysis are briefly discussed.

Multiple Regression Analysis

Multiple regression analysis has been most widely used by hydrologists for many years (Nash, 1960; Sharp, Gibbs, Owen and Harris, 1960; Stoddart and Watt, 1970). However, this statistical technique is often unsuitable for application in surface water hydrologic research because of assumptions inherent in the mathematical development of multiple regression analysis (Sharp, Gibbs, Owen and Harris, 1960; Fiering, 1964). In the mathematical development of multiple regression theory for hydrologic data analysis, it is assumed that the so-called 'independent' variables are truely independent. This is seldom the case in surface water hydrologic research. In a multiple regression equation, the regression coefficient of each independent variable is the weight given to each in order to minimize the sum of the squares of the residual errors. If these residual errors are small, the multiple correlation is high and a good prediction equation results. If multicollinearity exists among independent variables, these regression coefficients cannot be given any physical interpretation, and the result is a poor prediction equation. To illustrate this effect, a simple example is given as follows:

If a multiple linear regression model is assumed,

$$x_0 = \beta_1 x_1 + \beta_2 x_2 + \varepsilon, \tag{C.1}$$

then

$$\hat{x}_0 = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \tag{C.2}$$

where x_0, $\hat{x}_0$ are the true value and the estimated value of the dependent variable respectively, $x_{i\neq 0}$ are measurable independent variables, β_i, $\hat{\beta}_i$ are population parameter and its estimate respectively, ε is the error term, and all x's and ε are standardized variable, i.e., with zero mean and unit variance.

It can be found from any statistical textbook (Hald, 1952) that

$$\hat{\beta}_1 = \frac{\gamma_{01}-\gamma_{02}\gamma_{12}}{1-\gamma_{12}^2}, \qquad (C.3)$$

$$\hat{\beta}_2 = \frac{\gamma_{02}-\gamma_{01}\gamma_{21}}{1-\gamma_{21}^2}. \qquad (C.4)$$

If x_1, x_2 are statistically independent, then $\gamma_{12} = \gamma_{21} = 0$ and $\hat{\beta}_1 = \gamma_{01}$, $\hat{\beta}_2 = \gamma_{02}$. If x_1, x_2 are of perfect collinearity, $\gamma_{12} = \gamma_{21} = 1$, $\gamma_{01} = \gamma_{02}$ and $\hat{\beta}_1 = \frac{0}{0}$, $\hat{\beta}_2 = \frac{0}{0}$. This indeterminate solution implies that there are infinite linear combinations of $\hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$ which give the same maximum multiple correlation, sometimes ridiculous value or sign of $\hat{\beta}_i$ may be obtained.

Eq. (C.2) can be rewritten as

$$\Delta\hat{x}_0 = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2. \qquad (C.5)$$

It can be seen that β_i are supposed to measure the relative change of $\hat{x}_0$ with respect to individual changes of $x_{i\neq 0}$. This is the only property which gives $\hat{\beta}_i$ their physical meaning. This is true for truly independent variables, whereas these coefficients are meaningless at all for prediction problems in terms of individual effect of changes of independent variables highly correlated among themselves.

From the above argument one can observe that multiple measurements may duplicate the same information and may cause trouble in prediction problem if they are treated as independent predictors. Hence hydrologist wants to condense the number of variables by expressing them in terms of a relatively small number of linearly independent factors, in other words to determine a minimum set of factors containing most of the original information.

Principal Component Analysis

An alternative to multiple regression analysis is principal component regression analysis. Excellent discussions of principal component regression analysis and its application in hydrology can be found in Wong (1963), Cavadias (1964), Fiering (1964), Eiselstein (1967) and Overton (1968). The principal component analysis breaks down a correlation matrix into a set of uncorrelated (orthogonal) components equal or less in number to the original set of variates. The method of principal components extracts the roots of the characteristic equation in descending order of magnitude so that only a few components will summarize the data. However, problem arises in assigning physical meaning to each component. Hence the rotating factor analysis is receiving a good deal of attention by hydrologist in the context that it seeks to explain the correlation matrix with a minimum of variables. The components are scaled to correlation coefficients and factor matrix is rotated to detect redundant variables, if indicated (Eiselstein, 1967; Overton, 1968).

Factor Analysis

Factor analysis starts generally with a correlation matrix of 'n' variables. The object of a factor analysis is to account for the observed intercorrelations (off-diagonal element of the correlation matrix) among the 'n' variables with a lesser number of 'm' of common factors. These factors are mathematically derived variables which are linear combinations of the original variables.

In addition to the common factors there are also 'n' unique factors which do not help to account for the intercorrelations but which do explain the fact that the intercorrelations are less than unity. The unique factor is associated with the independent portion of the variable. Each unique factor consists of two parts: an error factor and a specific factor. The error factor results from errors of measurement, whereas the specific factor is the non-error portion of the unique factor. A factor pattern is a set of equations relating the variables and the factors. Each variable is expressed as a weighted sum (linear combination) of the 'm' common factors and one unique factor. The weights are called factor loadings and are unknowns to be determined.

The general factor pattern of the linear model is

$$X_i = \sum_{j=1}^{m} a_{ij}F_j + u_iU_i \ , \quad i = 1,2,\ldots,n \qquad \text{(C.6)}$$

where X_i are the standardized variables, F_j are the common factors, U_i are the unique factors, a_{ij} are factor loadings for common factors and u_i are the weights for the unique factors.

Each u_iU_i can be written as $s_iS_i+e_iE_i$ if reliability coefficients can be established. This is usually done by repetition of measurements to assess the error of measurement. As implied by the model, the 'n' unique factors are not correlated with one another nor with the 'm' common factors. This is necessarily true; for if they were correlated with one another, then the factor would not be unique. The 'm' common factors are usually assumed to be uncorrelated, i.e., orthogonal.

A common factor which has 'n' non-zero loadings is termed a general factor. One which has two or more but less than 'n' non-zero loadings is termed a group factor.

The communality (h_i^2) of a variable is its common factor variance, i.e., that portion of the total variance of a variable that can be explained by the 'm' common factors. The uniqueness (u_i^2) of a variable then is that portion of its total variance that is not shared with any other variables in the set. Because the variables are first standardized before a factor analysis, the total variance of each variable is unity. Therefore, $u_i^2+h_i^2=1$ or the unique variance plus the common factor variance equals the total variance. If only the common factor variance is to be factor-analyzed, the 1's on the principal diagonal of the correlation matrix should be reduced to the communalities of the variables. If the total variance is to be factor-analyzed, the 1's are used unless reliability estimates are available.

Estimation of Factor Loadings

The first approximation of factor loadings was carried out using principal component technique as follows:

The sum of squares of factor loadings across a variable (a_{kj}^2), gives the explained variance of a particular variable, while any particular (a_{kp}^2) indicates the contribution of the factor F_p to the variance of that variable (X_k). The principal component method chooses initially the loadings a_{i1} that make the sum of the contributions of that factor to the total variance a maximum, i.e.,

$$V_1 = a_{11}^2+a_{21}^2+\ldots+a_{n1}^2 \qquad (C.7)$$

subject to the conditions

$$\gamma_{ik} = \Sigma a_{ij}a_{kj}, \qquad \begin{array}{l} i,k = 1,2,\ldots,n \\ j = 1,2,\ldots,m \end{array} \qquad (C.8)$$

where $\gamma_{ik} = \gamma_{ki}$, $\gamma_{ii} = 1$ (or the communalities, if used).

The maximization of eq. (C.7) subject to the conditions of eq. (C.8) results in the following 'n' equations:

$$\begin{aligned} (1-\lambda)a_{11}+\gamma_{12}a_{21}+\ldots+\gamma_{1n}a_{n1} &= 0 \\ \gamma_{21}a_{11}+(1-\lambda)a_{21}+\ldots+\gamma_{2n}a_{n1} &= 0 \\ &\vdots \\ \gamma_{n1}a_{11}+\gamma_{n2}a_{21}+\ldots+(1-\lambda)a_{n1} &= 0 \end{aligned} \qquad (C.9)$$

In matrix form the system of 'n' equations can be written as

$$(R-\lambda I)A = 0. \qquad (C.10)$$

If this system of equations has a non-trivial solution ($A\neq0$), then the determinant, $|R-\lambda I|$, must be zero, i.e.,

$$|R-\lambda I| = 0. \tag{C.11}$$

Equation (C.11) is known as the characteristic equation and λ is the characteristic root of the equation. This will lead to 'n' or less different roots λ_p and therefore 'n' or less corresponding combinations of $(a_{1p}, a_{2p},\ldots,a_{np})$ which again correspond to 'n' or less different factors F_p. It can be shown that

$$\Sigma a_{ip}^2 = \lambda_p \quad \text{and} \quad \Sigma a_{ip}a_{ij} = 0, \qquad \begin{array}{l} i = 1,2,\ldots,n; \\ p,j = 1,2,\ldots,m; p \neq j. \end{array} \tag{C.12}$$

This implies that the resulting components or factors are uncorrelated or orthogonal. Usually the first factor accounts for the largest portion of the total variance and the second factor for the second largest portion and so on.

There are also other techniques available to estimate the factor loadings such as centroid method (Thurnston, 1947), maximum likelihood estimation (Lawley and Maxwell, 1963).

Varimax Method of Factor Rotation

The factor pattern resulting from the principal component solution or other techniques is such that all 'm' factors are usually general, i.e., non-zero loadings for all variables. In this solution the complexity of a variable is said to be 'm' since 'm' factors are necessary to describe the variable. This type of solution is not amenable to physical interpretation.

Thurnston's principle of "simple structure" was an attempt to define the multiple-factor type of solution that could be interpreted. His criteria of simple structure are (Thurnston, 1947):

1. Each row of the factor matrix should have at least one zero;
2. If there are 'm' common factors, each column of the factor matrix should have at least 'm' zeros;
3. For every pair of columns of the factor matrix there should be several variables whose entries vanish in one column but not in the other;
4. For every pair of columns of the factor matrix, a large proportion of the variables should have vanishing entries in both columns where there are four or more factors;
5. For every pair of columns of the factor matrix there should be only a small number of variables with non-vanishing entries in both columns.

In general, the rotation of axes (factors) to a simple structure is an attempt to reduce the complexity of the variables. The ultimate objective would be a uni-factor solution, in which each variable is of complexity one. The uni-factor solution is an ideal but it is practically impossible with empirical data.

If the multiple-factor solution satisfies Thurnston's simple structure criteria, the graphical plot in the plane of each pair of factors will exhibit the following: (1) many points near the two final two-factor axes; (2) a large number of points near the origin; and (3) only a small number of points removed from the origin and between the two axes. These three characteristics are the basis for several graphical rotation methods. However, any graphical method is subjective and hence the solution can vary from individual to individual.

In order to reduce subjectivity in rotation to simple structure, Carroll (1953) proposed an unique solution by minimizing the sums of cross-products (across factors) of squares of factor loadings, i.e., minimizing

$$f = \sum_{p<q} w_{pq} \qquad p,q = 1,2,\ldots,m \tag{C.13}$$

where

$$w_{pq} = \sum_{j=1}^{n} \Big(\sum_{k=1}^{m} a_{jk}\lambda_{kp}\Big)^2 \Big(\sum_{k=1}^{m} a_{jk}\lambda_{kq}\Big)^2, \tag{C.14}$$

and $\sum_{k=1}^{m} a_{jk}\lambda_{kp}$ is the element v_{jp} of the matrix

$$V = F\Lambda \tag{C.15}$$

where F is the initial factor matrix and Λ is the required directional cosine matrix such that V minimizes 'f'. The equations involved appear to be capable only of iterative solution. This analytic solution is especially good for oblique factor analysis which, however, is not yet commonly used in hydrologic data research.

An alternative and more stable solution, called the normal varimax rotation, suggested by Kaiser (1958) is given as follows:

The Varimax method aims at simplifying the columns, or factors, of the factor matrix while at the same time the simplicity of each variable may be attained concurrent with a large loading on a factor.

The simplicity of a factor 'p' is defined as the variance of its squared loadings, i.e.,

$$\sigma_p^2 = \frac{1}{n}\Sigma(b_{ip}^2)^2 - \frac{1}{n^2}(\Sigma b_{ip}^2)^2, \quad \begin{array}{l} i = 1,2,\ldots,n; \\ p = 1,2,\ldots,m. \end{array} \tag{C.16}$$

When the variance is maximum, the factor has the greatest interpretability or simplicity since factor loadings (the b's) tend toward unit and zero. Therefore, the complete factor matrix possess maximum simplicity when the sum of the simplicities of the individual factor is a maximum:

$$\sigma^2 = \sum_p \sigma_p^2 = \frac{1}{n}\sum_p\sum_i b_{ip}^4 - \frac{1}{n^2}\sum_p(\sum_i b_{ip}^2)^2. \tag{C.17}$$

The maximization of eq. (C.17), called the 'raw' varimax criterion, leads to biased solutions because the communalities implicitly influence the weight given to each variable in the rotations. Kaiser modified the 'raw' varimax criterion to the so called 'normal' varimax criterion such that the final factor loadings after rotation should be determined so as to maximize the function:

$$v = n\sum_p\sum_i (b_{ip}/h_i)^4 - \sum_p(\sum_i b_{ip}^2/h_i^2)^2. \tag{C.18}$$

In matrix equation form, the rotation of the factors subject to the normal varimax criterion, eq. (C.18) is equivalent to finding a transformation matrix T such that

$$B = AT \tag{C.19}$$

and B maximizes 'v', where B is the transformed factor matrix and A is the initial factor matrix. The factors are rotated two at a time and the complete cycle of $m(m-1)/2$ pairings of factors is repeated until the value of 'v' to the specified number of decimal places no longer increases.

Kaiser developed an equation for the required angle of rotation. Let the 'normalized' factor loadings of a variable X_i for a particular pair of factors 'p' and 'q' be designated by

$$y_i = a_{ip}/h_i, \quad z_i = a_{iq}/h_i \tag{C.20}$$

and the rotated loadings by Y_i, Z_i. The transformation matrix can then be put in the form:

$$(Y_i, Z_i) = (y_i, z_i) \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \tag{C.21}$$

where ϕ is the angle of rotation in the plane of the factors 'p' and 'q'. Then if

$$u_i = y_i^2 - z_i^2$$

$$v_i = 2y_i z_i \tag{C.22}$$

and

$$A = \Sigma u_i$$

$$B = \Sigma v_i$$

$$C = \Sigma(u_i^2 - v_i^2)$$

$$D = 2\Sigma u_i v_i$$

where all sums are on 'i' from 1 to 'n', the required angle of rotation is

$$\tan 4\phi = \frac{nD-2AB}{nC-(A^2-B^2)} . \tag{C.23}$$

The normal varimax method not only provides an analytical solution for simple structure, but the solution is according to Kaiser, "invariant under changes in the composition of the test battery" (set of variables), i.e., the angle of rotation does not depend on the number of tests in such factor cluster. The obvious advantage of this invariance property is that it permits inferences to be drawn from the factors derived from a sample to the universe varimax factors. (Note: The above normal varimax criterion (C.18) is applicable only to the orthogonal faster analysis. For the oblique case, either Carroll's criterion or the following alternative normal varimax criterion can be used, minimizing

$$C = \sum_{s<t} \{[n\sum_j (a_{js}^2/h_j^2)(a_{jt}^2/h_j^2) - (\sum_j a_{js}^2/h_j^2)(\sum_j a_{jt}^2/h_j^2)]/n^2\}$$

where s, t = 1,2,...,m, j = 1,2,...,n.)

Some Statistical Considerations

The model development in factor analysis does not require a knowledge of the variate distributions, but the development of the sample theory, in almost all cases, requires that the variates be Normally distributed. As pointed out by Kendall (1961), in multivariate analysis, especially in those branches which deal with components, factors and canonical correlations, exact inferences from sample to parent offer some exceedingly difficult problems in distribution theory. Practically everything that is known about exact distributions in the multivariate case depends on the assumption that the parent distribution is multivariate

Normal. In fact, this first place requirement imposed on original variables is usually not met in most hydrologic measurements. Sometimes proper transformation, such as logarithm has to be taken on original variates in order to get approximately normal distributions.

Approximate statistical tests of significance to determine the number of principal factors based on the examination of residual correlation matrix have been developed by several authors (Bartlett, 1950; Lawley and Maxwell, 1963). However, for the sake of simplicity, it was decided to chose the first few principal components which contribute to a high percentage of explained variance, say, 95% (Wallis, 1965).

The final factor pattern after the Varimax rotation was obtained by determining the predictor variables which give highest factor loadings in each column of the factor matrix. The selection of the dominant predictor variables was also based on the following criterion of minimum acceptable factor loading (Harman, 1960):

$$s_a = \frac{1}{2} \sqrt{(\frac{3}{R} -2-5R+4R^2)/N}, \qquad (C.24)$$

where s_a is the standard error of a factor loading 'a'; R is the average value in the correlation matrix and N is the number of observations.

CURRICULUM VITAE

Name: Timothy Chun Ta Hsu

Place and Year of Birth: Shanghai, Rep. of China
1946

Education:

National Taiwan University
Taiwan, Rep. of China
B.E. (Civil Engineering)
1963-1967

Queen's University
Kingston, Ontario
M.Sc. (Civil Engineering)
1968-1970

University of Ottawa
School of Graduate Studies
Ph.D. (Civil Engineering)
1970-1977

Experience:

Network Planning Engineer
Hydrometric Network Planning and Forecasting Section
Department of the Environment
Ottawa, Ontario
1972-1976

Project Methodology Officer
System Analysis and Statistics Division
Transportation Development Agency
Montreal, Quebec
1976-

Publications:

A mathematical model for continuous basin runoff simulation. C.T. Hsu, M.Sc. Thesis, Queen's University, 1970.

Continuous basin runoff simulation for intermediate drainage basins in southern Ontario. C.T. Hsu and W.E. Watt, C.E. Research Report No. 68, Queen's University, 1970.

A relation between peak discharge and twenty-four hour flow for rainfall floods. W.E. Watt and C.T. Hsu, C.E. Research Report No. 70, Queen's University, 1971.

Continuous basin snowmelt runoff simulation. W.E. Watt and C.T. Hsu, Proc. Hydrology Sym., No. 8, Vol. 2, 1971.

Stochastic prediction of hydrologic time series. K. Adamowski and C.T. Hsu, presented at the AGU Fall Annual Meeting, San Francisco, December 1974.

Stochastic generation of runoff. K. Adamowski and C.T. Hsu, presented at the 10th Canadian Hydrology Sym., Winnipeg, August 1975.

Linear stochastic prediction. K. Adamowski and C.T. Hsu, to be presented at the 3rd Intern. Hydrology Sym., Fort Collins, June 1977.