Stochastic Transfer Function Noise Model of the Smoky Hill River Inflow into the Kanopolis Reservoir

Joffre Castro
Marian Kemblowski
Al Macfarlane

Open-File Report 84-10
Kansas Geological Survey
University of Kansas
Lawrence, Kansas
December 1984

ABSTRACT

The hydrology of the Smoky Hill River basin between the cities of Bunker Hill and Ellsworth, in central Kansas, is studied by means of Time Series Analysis. The methodology is employed to investigate intrarelationships and interrelationships among components of the natural system. Thus, mean monthly riverflow time series is simulated as an ARMA(1,2) process. Mean monthly precipitation time series is found to have no statistical structure, and, therefore, to be a white noise process. The cross correlations of the phenomena are analyzed, modeled, and forecasted using transfer function methodology. Here, the hydrological system is treated as a stochastic process with inputs and outputs. The output series riverflow is estimated using a response function of the input series (precipitation) and an ARMA process, modeling the stochastic behavior of the output series.

TABLE OF CONTENTS

TABLE OF CONTENTS(continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

NOTATION

a(t)	white noise process
A(B)	autoregressive polynomial
ACF	autocorrelation function
AIC	Akaike Informatoin Criterion
ARMA	autoregressive moving average process
b	time lag
B	backward shift operator
C	constant
C(k)	covariance
CCF	cross correlation function
d(B)	lower polynomial of the transfer model
E	expectation
e	constant
F(B)	moving average polynomial
g(B)	white noise process of the Y(t) series
G1	skewness coefficient
G2	kurtosis
h(B)	white noise process of the X(t) series

NOTATION(continued)

I(B)	polynomial of the noise term
k	lag interval in months
l(B)	residual process of the filtered DIS series
m_y	monthly mean of the series Y(t)
N	number of observations
NID	normally independent distributed
p	number of autoregressive terms
PACF	partial autocorrelation function
$\bar{P}_k^*$	modified autocorrelation matrix
$\bar{P}_k$	autocorrelation matrix
q	number of moving average terms
Q(t)	transfer noise process
Q1	A Portmanteau lack of fit test for ARMA models
Q2	A Portmanteau lack of fit test for the $r_a(k)$ of the transfer model
Q3	A Portmanteau lack of fit test for the $r_{ha}(k)$ of the transfer function
r	number of terms of the $d_r(B)$ polynomial
r(k)	autocorrelation coefficient of order k

NOTATION(continued)

s	number of terms of the $w_s(B)$ polynomial
S_y^2	standard deviation of the $Y(t)$ series
t	time
$V(B)$	transfer function polynomial
V_k	k-th order weight of the $V(B)$ polynomial
	$VAR(k)$ varaiance of k
$w(B)$	upper polynomial of the transfer function
$Y(t)$	RAIN time series
u	maximum between r, and $b+s$
$X(t)$	DIS time series
$Z(t)$	time series

INTRODUCTION

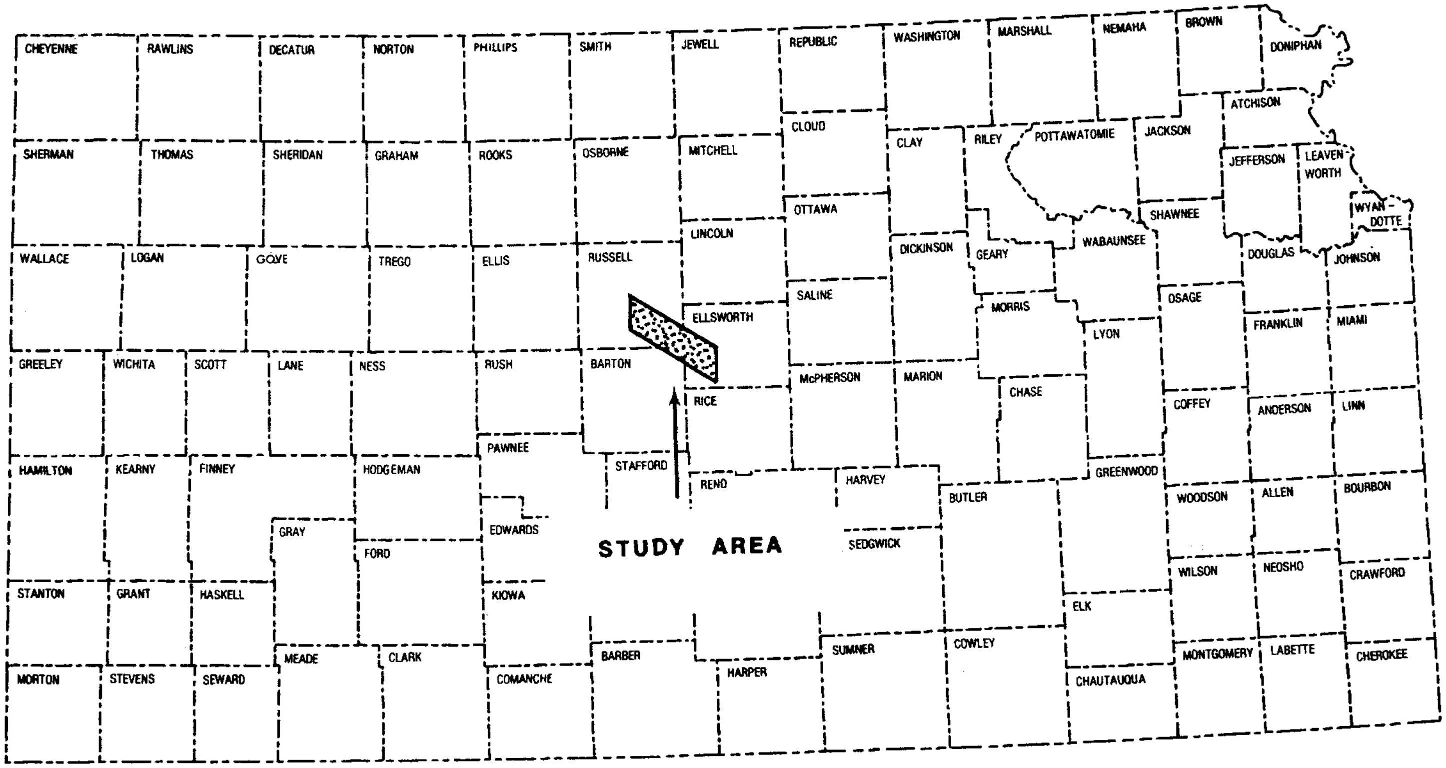


Fig. 1 Location map

Rural and urban development in the Smoky Hill River basin between Cedar Bluff Reservoir and Kanopolis Reservoir have brought about marked increase in water use. As a result there is a danger that over-development of surface- and subsurface-water will seriously diminish streamflow in the Smoky Hill River.

Reduction in the streamflow would disturb municipal, industrial, and irrigation water supply not only in quantity but also in quality. Water supply is one of the major concerns in the area, especially for the basin's two larger cities, Hays and Russell. This problem has become so serious that the Chief Engineer of the Water Resources Division, Kansas State Board of Agriculture, in 1984 declared a moratorium on new drilling in the Ellis County portion of the Smoky Hill River valley pending public hearings on the matter.

OBJECTIVES

The purpose of this study is to acquire a better understanding of the hydrology of the Smoky Hill River basin between the cities of Bunker Hill and Ellsworth (see figure 2) in order to foresee the hydrologic impact of increasing water development taking place upstream, and to develop a stochastic model to simulate and to forecast river inflows to Kanopolis Reservors, which could be useful in establishing the operation rules of the reservoir.

This study analyzes, models, and forecasts some of the hydrological components of the natural system, e.g. precipitation and streamflow. This task is accomplished by means of Time Series Analysis.

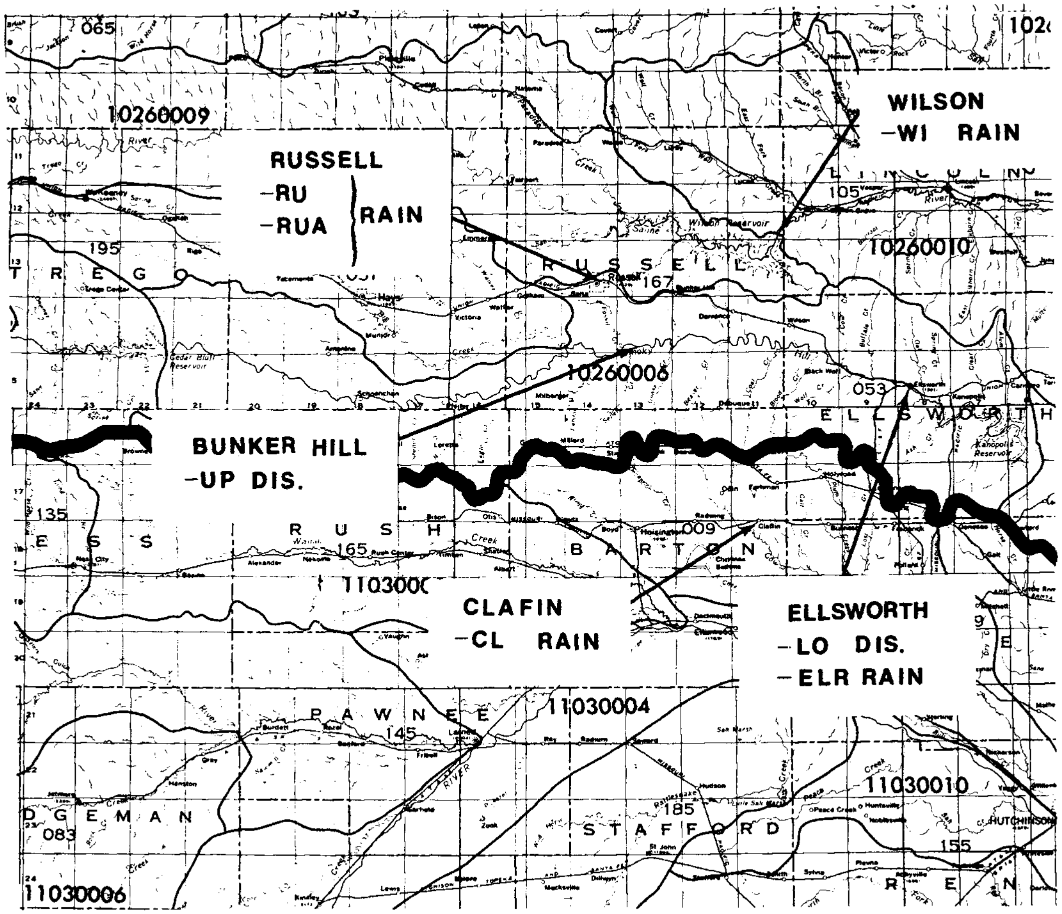


Fig. 2 Area of study and location of hydrological stations

METHODS

A literature search to collect data is a very important part of this study, which is to be complemented with field work. Additional information such as surface-water data is available through publications of the U.S. Geological Survey. Climatological data published by the Natural Oceanic and Atmospheric Administration are used in defining the distribution of precipitation over the area.

In the stochastic modeling, the region of study is treated as a hydrological system with inputs and outputs. The rainfall input and the streamflow output are treated as time series describing the stochastic behavior of the hydrological phenomena. In the next chapter a brief introduction to Time Series stochastic modeling is presented.

RESULTS

It is intended, that the results of this study will determine the importance and establish relationships among the hydrological components of the natural system. A model based on these relationships could be used as a predictive tool to investigate the effects of increased water use. Particularly, the model could simulate river inflow to Kanopolis Reservoir in order to perform a Reservoir Failure Analysis.

THEORETICAL MODEL

In general, a natural process can be decomposed into a deterministic component, which is one that can be determined for predictive purposes, and a stochastic component consisting of chance-dependent effects. The deterministic component may be made up of trends and/or periodicities, while the stochastic component consists of irregular oscillations and random effects which can not strictly be accounted for physically and requires stochastic methods for its description. It is this stochastic component, treated as a random variable, that will be studied and modeled throughout the course of this work.

A number of stochastic models have been proposed in the past. Some of them are (Salas et al., 1980):

- autoregressive models (AR)
- fractional Gaussian noise models (FGN)
- autoregressive-moving average models (ARMA)
- broken-line models (BL)
- shot-noise models
- model of intermittent processes
- disaggregation models
- general mixture models

In spite of the fact that each model has its own merit and some can be successfully applied in operational hydrology, all have limitations. The models selected to be used in this study are the so-called autoregressive moving-average models (ARMA). They have been extensively applied in the analysis of hydrological phenomena because of their efficiency in reproducing the mean, the variance, and also other higher-order moments of the parent population. These models are capable of preserving short-term and long-term persistences equally well. ARMA models are, in general, powerful, flexible, and simple in their practical applications. Finally, these models will be used as an introduction to more complex theory, e.g. transfer function methodology.

CONCEPTS AND DEFINITIONS

Before going into the description of ARMA models' theory, it is convenient to introduce some concepts and definitions relevant to the following discussion.

Stochastic process. A stochastic phenomenon is one that evolves in time according to probabilistic laws. A simple and useful way of describing it is to give the moments of the process, particularly the first and second moments, which are called the mean, variance, and autocovarinace functions (Spiegel, 1961).

The mean is defined by

$$m = E\left\{X\right\}$$

The variance is defined by

$$S^2 = E\left\{(X - m)^2\right\}$$

The autocovariance is defined by

$$C(t_1,t_2) = E\left\{(X_{t1} - m)*(X_{t2} - m)\right\}$$

where E represents expectation. For a discrete series it is

$$E\left\{f\right\} = (1/N)*\sum_{i=1}^{N} f_i$$

and N is the number of observations.

<u>Skewness</u> <u>G1</u>. Skewness is the degree of asymmetry of a distribution,

$$G1=((1/N)\sum_{t=1}^{N} a^3(t)) \;/\; ((1/N)\sum_{t=1}^{N} a^2(t))^{3/2}$$

G1 is approximately N(0,6/N).

<u>Kurtosis</u> <u>G2</u>. Kurtosis is the degree of peakedness of a distribution usually taken relative to a normal distribution,

$$G2=((1/N)\sum_{t=1}^{N} a^4(t)) \;/\; ((1/N)\sum_{t=1}^{N} a^2(t))^2 -3$$

G2 is approximately N(0,24/N).

Stationary process. A process is stationary when the moments of the process do not change over time. In practice, it is often useful to define a second-order (or weakly) stationary process, which implies that the mean as well as the covariance is constant and, therefore, finite.

One important class of processes is a stationary, normally distributed process, which is completely characterized by its first- and second-order moments, and hence it is a "strictly" stationary normally distributed process. Time series analysis is primarily concerned with this type of process and, for this reason, series are often transformed into stationary, normally distributed time series to use this theory.

Autocorrelation function (ACF), r(k). The autocorrelation function, ACF, measures the amount of linear dependence between observations of the same variable k units apart.

$$r(k) = C(k) / C(0)$$

where C(k) is the autocovariance function at lag k. It is important to note that

$$C(0) = S^2$$

and

$$r(0) = 1$$

as well as

$$r(k) < 1 \qquad k \neq 0$$

and the variance of the r(k) is given by (Bartlett, 1946)

$$Var(r(k)) = (1/N)* \sum_{i=k}^{K} r^2(i)$$

<u>Partial autocorrelation function</u>, <u>PACF</u>, <u>A_{kk}</u>. The partial autocorrelation function is defined as the set of partial autocorrelations at various lags A_{kk} :k=1,2,... and these are defined by

$$A_{kk}=|\bar{P}_k^*|/|\bar{P}_k|$$

where $\bar{P}_k$ is the k*k autocorrelation matrix,

$$\bar{P}_k = \begin{bmatrix} 1 & r_1 & \cdots & r_{k-1} \\ r_1 & 1 & \cdots & r_{k-2} \\ . & . & . & \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ r_{k-1} & r_{k-2} & \cdots & 1 \end{bmatrix}$$

$\bar{P}_k^*$ is $\bar{P}_k$ with the last column replaced by

$$\begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ . \\ . \\ . \\ r_k \end{bmatrix}$$

so

$$A_{11} = r_1$$

$$A_{22} = \frac{\begin{vmatrix} 1 & r_1 \\ r_1 & r_2 \end{vmatrix}}{\begin{vmatrix} 1 & r_1 \\ r_1 & 1 \end{vmatrix}} = \frac{r_2 - r_1^2}{1 - r_1^2}$$

For fairly large N, A_{kk} is approximately normally distributed, and its standard error is equal to (Anderson, 1976),

$$\text{VAR } (A_{kk}) = N^{-1/2}$$

<u>Cross correlation function (CCF), $r_{xy}(k)$</u>. Cross correlation function measures the linear relationship at different lags between two processes. Then, if X(t) and Y(t) are two discrete processes, the cross correlation function between X(t) and Y(t) is given by

$$r_{xy}(k) = (S_x *S_y)^{-1} *C_{xy}(k)$$

where the cross covariance function, $C_{xy}(k)$ between X(t) and Y(t), is equal to

$$C_{xy}(k) = E\ (X(t)-m_x)*(Y(t+k)-m_y)$$
$$k = 0,1,2.....$$

Pure random process. A discrete process a(t) is a random process if the random variables are a sequence of mutually independent, identically distributed variables. It is clear that this process is stationary and normally distributed since the mean and the autocorrelation function do not change over time.

The autocorrelation function is equal to

$$r(k) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

This process a(t) will be referred to as "white noise".

Moving average process (MA). Suppose that a(t) is a white noise process with mean zero and variance S^2. Then a process X(t) is said to be a moving average process of order q, MA(q), if

$$X(t) = a(t) - F_1*a(t-1) -....-F_q*a(t-q) \tag{1}$$

The notation in Eq. 1 can be simplified using the backshift operator B, which is defined as

$$B^n*X(t) = X(t-n)$$

then

$$X(t) = (1 - F_1*B - \ldots\ldots - F_q*B^q)*a(t)$$

or

$$X(t) = F(B)*a(t)$$

It is seen that

$$E\left\{X(t)\right\} = 0$$

and

$$VAR\ (X(t)) = S_a^2 * \sum_{i=0}^{q} F_i^2$$

also that

$$r(k) = \begin{cases} 1 & k = 0 \\ \sum_{i=0}^{q-k} F_i*F_{i+k} \Big/ \left(\sum_{i=0}^{q} F_i^2\right) & k = 1,2,\ldots.q \\ 0 & k > q \\ r(-k) & k < 0 \end{cases}$$

Note that the autocorrelation function (ACF) cuts off at lag q, which is a special feature of a moving average process of order q, MA(q).

Autoregressive process (AR). A process X(t) is said to be an autoregressive process of order p, AR(p), if

$$X(t) = A_1*X(t-1) + A_2*X(t-2)\ldots\ldots\ldots.A_p*X(t-p) + a(t)$$

or

$$(1 - A_1*B \;\; - A_2*B^2 \; \ldots\ldots - A_p*B^p \;)*X(t) = a(t)$$

also

$$A(B)*X(t) = a(t)$$

or

$$X(t) = A^{-1}(B)*a(t)$$

which means that an AR(p) process can be expressed as a MA process of infinite order; that is also true for a MA(q) process which can be expressed as an AR of infinite order as well.

<u>Mixed</u> <u>models</u> (<u>ARMA</u>). A useful class of models is that formed from a combination of MA and AR processes, say ARMA(p,q) given by

$$A(B)*X(t) = F(B)*a(t)$$

where A(B), and F(B) are polynomials of order p,q, respectively, such that

$$A(B) = 1 - A_1*B \;\; - A_2*B^2 \;\; \ldots\ldots\ldots - A_p*B^p$$

and

$$F(B) = 1 - F_1*B - F_2*B^2 \ldots\ldots\ldots\ldots - F_q*B^q$$

The importance of ARMA processes lies in the fact that a stationary time series may often be described by an ARMA model involving fewer parameters than a MA or an AR process by itself.

If an ARMA (p,q) has p=1, and q=1, then

$$(1 - A*B)*X(t) = (1 - F*B)*a(t)$$

or

$$X(t) = A*X(t-1) + a(t) - F*a(t-1) \qquad (2)$$

In order to analyze the behavior of the autocorrelation function ACF and the partial autocorrelation function PACF of an ARMA(1,1) process, both sides of Eq. 2 are multiplied by X(t-k), and the expectations are taken for example

$$E\left\{X(t)*X(t-k)\right\} = E\left\{(A*X(t-1) + a(t) - F*a(t-1))*X(t-k)\right\}$$

then

$$C(k) = A*C(k-1) + C_{xa}(k) - F*C_{xa}(k-1)$$

so

$$C(0) = A*C(1) + S_a^2 - F*C_{xa}(-1)$$

$$C(1) = A*C(0) + F*S_a^2$$

and the ACF for k > 1 is

$$C(k) = A*C(k-1)$$

and the autocorrelation function is,

$$r(k) = A*r(k-1) \qquad K > 1$$

This equation indicates that the autocorrelation function ACF decays geometrically from lag 1 onwards. The partial autocorrelation function PACF also decays in magnitude from an initial value $A_{11} = r_1$ (Anderson, 1976).

Box and Cox transformation. As was mentioned previously, a time series ought to be normally distributed to be modeled by stochastic processes. In practice, however, the majority of hydrological phenomena show non-normal distributions and must be transformed to normality. Box and Cox have suggested the following method of transforming an observed value Z (Box and Cox, 1964)

$$Y = (Z - 1)^{\lambda}/\lambda \qquad \lambda \neq 0$$
$$Y = \ln(Z) \qquad \lambda = 0$$

Sometimes it is known in advance that the time series observations of a given phenomenon require a certain type of transformation. For example, in practice, it is often necessary to transform average monthly river flow data using natural logarithms. If there are one or more zero values in the series a constant (e) can be included. The constant (e) is a location parameter which could be negative (Hipel et al., 1977)

$$Y = \ln(Z - e)$$

The value of the constant (e) is obtained from the following equation (Kottegoda, 1980)

$$m_z = EXP(m_y + S_y^2/2) + e \qquad (3)$$

and

$$S_z^2 = (EXP(2*m_y + S_y^2))*(EXP(S_y^2)-1) \qquad (4)$$

also,

$$G1 = (EXP(3*S_y^2) -3*(EXP(S_y^2) +2)*(EXP(S_y^2)-1)^{-3/2} \qquad (5)$$

where

G1 = coefficient of skewness of the Z(t) series

m_y = mean of the transformed series, Y(t)

S_y^2 = variance of the transformed series, Y(t)

The value of the constant (e) is easily obtained from Eq. 3 if the mean m_y and the variance S_y^2 of the transformed series Y(t) are known. These parameters, m_y and S_y^2, are implicitly expressed in Eqs. 4 and 5; therefore, an iterative procedure is needed to solve them.

<u>Periodicity</u>. Whereas observed trends are part of long-term fluctuations, periodic effects in hydrological time series are deterministic in nature with regard to their frequency of occurrence, because they are imposed by a cyclic phenomenon, that is, by one with a fixed period. This is known as the seasonal effect and is clearly evidenced in closely spaced data such as those from monthly river flows. In general, the periodic component in a hydrologic time series is removed by subtracting the estimated monthly mean and dividing by the estimated standard deviation for each data point (Kottegoda, 1980)

$$X_{jt} = (S_j)^{-1} * (Z_{jt} - m_j)$$

Summary

The time series methodology employed in this work is exclusively concerned with the analysis of stationary, normally distributed processes. Therefore, prior to the stochastic modeling, hydrologic phenomena need to be transformed into normally distributed processes. If there are evidences of non-stationarity, trends as well as seasonal effects need to be removed.

ARMA PROCESS

Let $X(1),X(2),\ldots\ldots X(t-1),X(t),X(t+1),\ldots\ldots X(N)$ be a discrete time series measured at equal time intervals. An ARMA model for $X(t)$ is written as

$$A(B)*X(t)=F(B)*a(t) \tag{6}$$

where

t = discrete time

$X(t)$ = appropriate transformed and normalized discrete variable

$a(t)$ = normally independent distributed white noise residual with mean zero and variance S_a^2; $NID(0,S_a^2)$

$A(B)$ = $1-A_1*B-A_2*B^2 \ldots\ldots A_p*B^p$, autoregressive polynomial of order p, AR(p)

$F(B)$ = $1-F_1*B-F_2*B^2\ldots\ldots F_q*B^q$, moving average polynomial of order q, MA(q)

B = backward shift operator defined by $B^n*Z(t)=Z(t-n)$

The modeling of a time series always reduces it to a noise or independent stochastic component such as in Eq. 6. It is ideal for a(t) to be a white noise or an independent, stationary normally distributed variable (Box and Jenkins, 1970).

Method of Analysis

When fitting a stochastic model to a given time series the recommended identification, estimation, and diagnostic check stages of model development should be followed. Descriptions of various techniques may be found in the statistical (Box, and Jenkins, 1970) and hydrologic (Hipel et al., 1977a; McLeod et al., 1977) literature.

Identification

The purpose of the identification stage is to determine the order of the autoregressive (AR) and moving average (MA) processes and to obtain initial guesses for the parameters. The tentative model so obtained provides a starting point for the application of the more formal and efficient estimation methods. The major diagnostic tools in the identification procedure are the sample autocorrelation (ACF) and partial autocorrelation (PACF) functions.

In practical applications, it is not always possible to identify non-stationary processes, nor is it possible to know the length of the seasonal effect present in the series. Two methods will be presented here to determine if the series is stationary and to evaluate the length of the seasonal effect.

a). Plot of the time series. A visual inspection of a plot of the original data gives an overall view of how the time series is generally behaving. It may reveal one or more of the following characteristics: seasonality, trends either in the mean level or in the variance of the series, persistence, long-term cycles, or extreme values and outliers.

b). Analysis of the ACF and the PACF. The autocorrelation function, ACF, and partial autocorrelation function, PACF, transform the given information into a format whereby it is possible to determine the presences of non-stationarity and seasonality, as well as the number of AR and MA terms required in the model.

The first step is to examine a plot of the autocorrelation function ACF to find any indication of non-stationarity in the series. When the data are non-seasonal, failure of the ACF to damp out indicates that the process is non-stationary and needs to be transformed (Box-Cox transformation). For seasonal data with length equal s, the ACF often follows a wave pattern peaking at s, 2s, 3s, and other integer multiples of s. If the ACF at lags that are integer multiples of the seasonal length s do not die out rapidly, this indicates that the process is non-stationary and the seasonal effect has to be removed (deseasonalization).

Once the data have been transformed and deseasonalized, the analysis of the ACF and PACF will give an indication of the number of AR and MA terms required in the model. Briefly, whereas the autocorrelation function ACF of an autoregressive process of order P tails off, its PACF has a cutoff after lag P. Conversely, the ACF of a moving average process of order Q has a cutoff after lag Q, while its PACF tails off. If both the ACF and PACF tail off, a mixed process is suggested. Furthermore, the ACF for a mixed process, containing a P-th order autoregressive component and a Q-th order moving average component, is a mixture of exponentials and damped sine waves after the first Q-P lags. Also the PACF is dominated by a mixture of exponential and damped sine waves after the first P-Q lags.

Estimation

Methods derived from mathematical statistics for estimating the parameters of models representing random variables are called estimation techniques. The most common estimation techniques are the method of moments, the method of least squares, and the method of maximum likelihood. Depending upon the estimation techniques, some estimators are better than others. The criteria for judging estimators are bias and mean square error. It is desirable to select a method with the smallest bias and the smallest mean square error. In addition to these characteristics, the properties of consistency and sufficiency are also important for describing estimators. A more thorough analysis of estimation techniques are presented in Box and Jenkins (1970) and Salas et al. (1980).

Diagnostic Check

The final stage of the model building methodology is to subject the identified and estimated model to diagnostic checks. When an inadequacy is detected, the checks should give an indication of how the model needs to be corrected, after which further fitting and checking take place. One class of diagnostic check is devised to test model adequacy by overfitting. However, most diagnostic tests deal with the residual assumptions in order to determine whether the a(t) in Eq. 6 are independent, homoscedastic, and normally distributed.

Overfitting. This involves fitting a more elaborated model. Extra parameters should be estimated for the complex model only where it is feared that the model may require more parameters. Careful thought should be given to the question of how the model should be augmented.

Tests for whiteness of the residuals. To determine whether the residuals a(t) in Eq. 6 are white noise, an appropriate procedure is to examine the residual autocorrelation function (ACF), r(k). A sensitive diagnostic check is first to plot the ACF along with asymptotic significance intervals, +/- 2 standard errors (SE). If some of the r(k) are significant different from zero, this may mean that the present model is inadequate. The important r(k) to examine are the r(k) at the first few lags.

A second but less sensitive test is to calculate and to perform a significance test for the Portmanteau statistic Q1.

$$Q1 = N * \sum_{k=1}^{k} r_a^2(k) \qquad (7)$$

Rather than to consider the r(k)`s individually (previous method), an indication is often needed of whether the first k autocorrelation of the residuals, taken as a whole, indicates inadequacy. The statistic Q1 is approximately a CHI-SQUARE (X^2) distribution with k-p-q degrees of freedom.

If $Q1 < X^2$ at a given significance level, a(t) is an independent series and the model is adequate. Otherwise the model is inadequate.

Test for normality of the residuals. The graph of the cumulative distribution of the residuals should appear as a straight line when plotted on normality paper if the residuals are normally distributed. Normally distributed data should possess neither significant skewness (G1) nor kurtosis (G2).

Akaike Information Criterion. A mathematical formulation which considers the principle of parsimony in model building is the Akaike Information Criterion (AIC) proposed by Akaike (1974). Under this criterion the model which gives the minimum AIC value is the one to be selected. The parameter AIC is defined by (Salas et al., 1980)

$$AIC = -N*\ln(S^2) + 2k \tag{8}$$

where

N = number of observations
K = number of AR and MA parameters
S^2 = maximum likelihood estimate of the residual variance
ln = natural logarithm

Whereas in previous paragraphs a brief description of univariate ARMA models has been given, in the next section one class of multivariate models which are a logical extension to univariate ARMA models is discussed. The specific branch of the time series methodology employed herein is referred to as transfer function modeling.

TRANSFER FUNCTION MODELS

One area of the time series analyses is the use of discrete linear transfer functions to model the interrelationships between input and output time series. For instance, precipitation obviously contributes to river flows; therefore, it is possible to identify a model which, given a set of precipitation data, will generate a set of river flows.

A simple model relating the two time series is

$$Y(t)=V_0*X(t) + V_1*X(t-1) + \ldots\ldots V_k*X(t-k) \qquad (9)$$

$$=V(B)*X(t) \qquad (10)$$

Eq. 10 is called a linear filter and the operator V(B) is called the transfer function of the filter. The unknown weights $V_0, V_1, \ldots.$ in Eq. 9 are called the impulse response function of the system. Eq. 9, however, is not a very satisfactory parameterization of the relationships between Y(t) and X(t), as it may contain an infinite number of unknown parameters V_0, V_1,..... . Furthermore, if there is not immediate response in Y(t) for a change in X(t), there may be a time lag present, so that X(t) influences observations Y(t+b), Y(t+b+1),......, but has no effect on Y(t),Y(t+1),...Y(t+b-1). Thus, the first b weights, say $V_0,V_1,\ldots,V_{b-1}$ in Eq. 9 will be zero. A convenient way to parameterize the model and to account for a time lag is to express the discrete dynamic system as a difference equation,

$$\begin{gathered}(1 - d_1*B - d_2*B^2 - \ldots\ldots d_r*B^r)*Y(t)= \\ (w_0 - w_1*B^1 - w_2*B^2 \ldots\ldots - w_s*B^s)*X(t - b)\end{gathered} \tag{11}$$

or

$$d_r(B)*Y(t)=w_s(B)*X(t - b) \tag{12}$$

$$Y(t)=d_r^{-1}(B)*w_s(B)*X(t - b) \tag{13}$$

If Eq. 10 is compared to Eq. 13, say

$$V(B)=d_r^{-1}(B)*w_s(B)*B^b \tag{14}$$

or

$$\begin{gathered}(1 - d_1*B- d_2*B^2\ldots\ldots - d_r*B^r)*(V_0 + V_1*B^1+ V_2*B^2\ldots)= \\ (w_0 - w_1*B^1 - w_2*B^2\ldots\ldots - w_s*B^s)*B^b\end{gathered} \tag{15}$$

then the impulse response function can be represented by the ratio of the two polynomials. The characteristics of the impulse response function are directly reflected in the choice of b, r, and s. If the coefficients of like power of B are equated in Eq. 15 the impulse response weights would form the following four groups (Montgomery and Weatherby, 1980):

1.- b zero weights $V_0, V_1, \ldots V_{b-1}$

2.- s-r+1 weights $V_b, V_{b+1}, \ldots V_{b+s-r}$ that follow no fixed pattern. These weights do not appear if $s < r$

3.- r startup or initial weights $V_{b+s-r+1}, \ldots V_{b+s-1}, V_{b+s}$

4.- weights $V_{b+s+1}, V_{b+s+2} \ldots$ that follow the pattern dictated by the following difference equation,

$$d_r(B)*V_j = 0 \qquad j > b+s \qquad (16)$$

In practical situations the values of r and s do not exceed 2. More comprehensive treatments of this analysis are in Box and Jenkins (1970), Granger and Newbold (1977), and Montgomery and Weatherby (1980).

Method of Analysis

When doing a simulation study, it is first necessary to thoroughly understand the problem and then to formulate a model which describes what is expected to occur based on knowledge of the known phenomena. This involves designing transfer functions plus choosing a tentative noise model. Then the parameters of the noise model and transfer function are estimated using the method of maximum likelihood. Finally, the model is checked for possible inadequacies. If discrepancies were observed the appropriate model modifications should be made.

Transfer Noise Model

The output series Y(t) will never exactly follow the pattern dictated by the transfer function model in Eq. 9. In general, there will be other influences that disturb the system caused mainly by variables not explicitly incorporated into the model, or by noise infecting the system. Denoting this noise component or disturbance term by N(t), the model for the single input case becomes:

$$Y(t)=d_r^{-1}(B)*w_s(B)*X(t-b) + N(t) \qquad (17)$$

Here X(t) and Y(t) represent input and output series. Assuming that the noise, N(t), can be modeled by an ARMA (p,q) process, say

$$A(B)*N(t)=F(B)*a(t) \qquad (18)$$

the transfer model becomes

$$Y(t)=d^{-1}(B)*w(B)*X(t-b) + A^{-1}(B)*F(B)*a(t) \qquad (19)$$

Identification

Two methods for identifying the appropriate form of the transfer function model, that is, determining b, r, and s, will now be introduced.

First method. For a given pair of time series, it is possible to identify interrelationships between the series and intrarelationships within the individual series. The interrelationships are summarized in the cross correlation function CCF, while the intrarelationships are summarized in the autocorrelation function ACF. The ACF will always affect the existing, relationships between the two series, e.g., the ACF will inflate the standard deviation of the CCF and will produce spurious cross correlations at different lags, which in many instances will be significantly correlated with each other. Therefore, the identification process is considerably simplified if the input series is reduced to a white noise. That is accomplished by pre-whitening the input series.

Assuming that X(t) is stationary and can be modeled as an ARMA process, say

$$A(B)*X(t)=F(B)*h(t) \tag{20}$$

where h(t) is a white noise; then the pre-whitened input series is

$$h(t)=F^{-1}(B)*A(B)*X(t) \tag{21}$$

Applying this same transformation to the output series Y(t), say

$$l(t)=F^{-1}(B)*A(B)*Y(t) \tag{22}$$

Then a transfer function model can be obtained relating the noise of the output series to the white noise of the input series, say

$$l(t)=V(B)*h(t) + Q(t) \tag{23}$$

where

$$Q(t)=F^{-1}(B)*A(B)*N(t) \tag{24}$$

and Q(t) is the transfer noise process. The pre-whitened input, h(t), and output series, l(t), are the key to the determination of the estimates of the impulse response weights. Box and Jenkins (1970) demonstrate that the impulse response weight V_k is simply a lag k cross correlation of the h(t) and l(t) series, $r_{hl}(k)$, multiplied by a scale factor. Specifically,

$$V_k= r_{hl}(k)*S_l/S_h \tag{25}$$

Thus, the CCF between the pre-whitened input and transformed output series is directly proportional to the impulse response function. The V(k) are not efficient estimators, but they are adequate to allow specification of appropriate values of b, r, and s in the transfer function model. After the form of the model has been tentatively identified, the noise component must be considered. An estimate of the noise series is obtained from

$$N(t)=Y(t) - V(B)*X(t) \tag{26}$$

The ACF and PACF of the series N(t) can be analyzed to produce a univariate model of the noise series.

Second method. Hipel et al. (1982) uses a double prewhitening procedure fitting ARMA models to each of the series, say

$$A(B)*X(t) = F(B)*h(t) \tag{27}$$

and

$$N(B)*Y(t) = M(B)*g(t) \tag{28}$$

where h(t) and g(t) are white noise processes. The relationships between X(t) and Y(t) are inferred from the cross correlation function between h(t) and g(t) (Haugh and Box, 1977). The transfer function linking the residuals is

$$g(t) = V'(B)*h(t) + Q'(B)*a(t) \tag{29}$$

and the transfer function weights are computed from

$$V'_k = r_{hg}(k)*(S_g/S_h) \tag{30}$$

In the present study, this method proved to be more sensitive and, therefore, more reliable.

Estimation

Once the form of the transfer function has been tentatively identified, it is necessary to estimate the parameters in the model. Then if the a(t) series is assumed to be normally distributed, a good approximation to the maximum likelihood estimates of the parameters is found by minimizing the conditional sum of the squares.

Generally, the conditional sum of squares function is nonlinear in the unknown parameters. The nonlinear estimation algorithm of Marquardt can be used to minimize the conditional sum of squares. The convergence properties of this algorithm are good, and the final estimates are not overly sensitive to the choice of starting values (Montgomery and Weatherby, 1980).

Diagnostic Checking

Following parameter estimation, the adequacy of the model should be examined. Most of the tests are aimed to verify the independence and the normality of the residuals. Other tests, however, are concerned directly with the amount of variance explained and the number of parameters employed by the models.

The autocorrelation function, ACF, of the residuals a(t), $r_a(k)$ should resemble that of a white noise. Structure in the ACF is an indication that the tentatively identified form of the model is incorrect. Then it is extremely helpful to examine the CCF between the pre-whitened input h(t) and the residuals, $r_{ha}(k)$. This procedure can assist in tracing the source of model inadequacy to either the noise model or the transfer function. If $r_a(k)$ exhibits structure and $r_{ha}(k)$ does not, the noise model is incorrect, while if both $r_a(k)$ and $r_{ha}(k)$ exhibit structure, the transfer function is incorrect.

An overall CHI-SQUARE test may also be applied to $r_a(k)$ and $r_{ha}(k)$. This test is useful in providing a measure of whether either the first k autocorrelations or cross correlations, considered as groups, indicate model inadequacy. The first k autocorrelations are conformable to the hypothesis that a(t) are white noise if

$$Q2 = m * \sum_{k=1}^{k} r_a^2(k) \ , \qquad m = N-u-p \qquad (31)$$

is less than an upper tail value of CHI-SQUARE with k-p-q degrees of freedom, u is the maximum of r and b+s (Box and Jenkins, 1970). Similarly the cross correlation $r_{ha}(k)$ indicates that there is not structure between the pre-whitened input and the residuals if

$$Q3 = m * \sum_{k=1}^{k} r_{ha}^2(k) \ , \qquad m = N-u-p \qquad (32)$$

is less than an upper tail value of CHI-SQUARE with k+1-(r+s+1) degrees of freedom (Box and Jenkins, 1970).

Overfitting is sometimes also useful as a model diagnostic technique. It means fitting additional parameters chosen to enlarge the model and testing their statistical significance. Parameters that are not statistically significant should be dropped from the model and the remaining parameters re-estimated.

The best dynamic model is then selected using the Akaike Informatiom Criterion (AIC). The AIC provides a combined measure of model parsimony and good statistical fit. The model which has the minimum AIC value in Eq. 8 should be selected when there are several competing models.

After following the three stages of model building as previously outlined, a stochastic model, representing a physical phenomenon, is obtained. This model could be now used to simulate series having the same statistical characteristics as the phenomenon under study or could be used to forecast future values of the sample series being analyzed. In the next section, the computation of the forecasts and their confidence prediction limits are presented.

FORECASTING

Frequently, forecasts of a time series Y(t), Y(t+1),...may be considerably improved by using information coming from an associated series X(t), X(t+1),... . This is particularly true if changes in Y tend to be anticipated by changes in X, called the "leading indicator" of Y. In order to obtain an optimal forecast using information from Y and X, a transfer function noise model is built connecting Y and X in the manner already outlined.

Referring to Eq. 19

$$d(B)*Y(t)=w(B)*X(t) + I(B)*a(t) \qquad (33)$$

where the X's and a's are stochastically independent. The minimum mean square error forecast, Y'(t+j), of Y(t+j) at origin t is the conditional expectation of the Y(t+j) at time t (Box and Jenkins, 1970). Representing the conditional expectation E as ",

$$Y'(t+j) = Y''(t+j) \qquad (34)$$

the lead-j forecast is,

$$\begin{aligned} Y''(t+j) &= d_1*Y''(t+j-1).......+d_{r+p}*Y''(t+j-r-p) \\ &+w_0*X''(t+j-b).........-w_{s+p}*X''(t+j-b-s-p) \\ &+ a''*(t+j) - I_1*a''(t+j-1)....... - I_{q+r}*a''(t+j-q-r) \end{aligned}$$

where the expectations at origin t are

$$Y''(t+j) = \begin{cases} Y(t+j) & j < 0 \\ \text{Forecast} & j > 0 \end{cases}$$

$$X''(t+j) = \begin{cases} X(t+j) & j < 0 \\ \text{Forecast} & j > 0 \end{cases}$$

$$a''(t+j) = \begin{cases} a(t+j) & j < 0 \\ 0 & j > 0 \end{cases}$$

It is also possible to obtain prediction intervals for the future observations Y(t+j). The variance S_{fj}^2 of the lead-j forecast error is found as

$$S_{fj}^2 = E\left\{(Y(t+j)-Y'(t+j))^2\right\}$$

$$S_{fj}^2 = S_h^2 * \sum_{j=b}^{j-1} V_j^2 + S_a^2 * \sum_{j=b}^{j-1} I_j^2 \qquad (35)$$

where

$Y(t+j)$ = actual value of output series

$Y'(t+j)$ = forecast of $Y(t+j)$

S_a^2 = variance of the white noise residual obtained from Eq. 33

S_h^2 = variance of the white noise residual obtained from Eq. 21

V_j = impulse response function Eq. 25

I_j = coeffs. of the linear filter representation of the noise Eq. 18

BASIN DESCRIPTION

The study area, in general terms, is a section of the watershed of the Smoky Hill River valley. It extends from the U.S. Geological Survey surface water-gaging station south of Bunker Hill (Russell Co.) up to the gaging station at Ellsworth City (Ellsworth Co.). This area lies within the Blue Hills and Smoky Hill physiographic subdivision of the Dissected High Plains Section of the Great Plains physiographic province (Schoewe, 1949).

GEOLOGY

The geology of the area is dominated by a northwest- to southeastwardly trending bedrock trough partially filled with alluvium, on top of which meanders the Smoky Hill River. The present river channel and the entire alluvial valley have cut or eroded 20 to 60 feet below a higher old valley approximately 2 to 3 miles wide following the same general course as the present valley. These old valley deposits, are called terrace deposits and are a moderate source of ground water. The bedrock is composed of Greenhorn Limestone, Graneros Shale, and Dakota and Kiowa sandstones of Cretaceous age. The bedrock is exposed at land surface for the most part, along the northern edge of the alluvial filled trough.

HYDROGEOLOGY

The alluvial valley consists mostly of well-sorted coarse sand and gravel of Pleistocene age. The alluvium is extremely permeable almost everywhere and is hydraulically connected to Smoky Hill River.

Ground water movement in the Smoky Hill River valley is toward the river and its major tributaries. Locally relief on the bedrock surface affects the direction of ground water movement (Leonard, 1975). The rate and direction of flow are also determined by the texture, composition, and distribution of sediments in the valley. Recharge to this system comes from precipitation falling on the alluvium and stream or from leakage from the bedrock units into the alluvium. Thus, the Smoky Hill River is a gaining stream along its entire length.

CLIMATE

Russell and Ellsworth counties have a semi-arid climate. During the summer the days are hot, wind velocity is moderate, humidity is low and the nights are generally cool. The winters usually are mild with occasional short periods of severe cold. The average annual precipitation at Russell, in central Russell Co., is 25.41 inches; meanwhile at Ellsworth, in Ellsworth Co., the average is 25.66 inches. Approximately 75% of this precipitation falls during the growing season of about 168 days, April to September.

JUSTIFICATION AND METHODS

In response to the increased rate of development in the Smoky Hill River valley, there is an intensified need for a broad understanding of the hydrogeology and geology of the area. The analysis and modeling of the hydrological system are essential for a proper management of the water resources, not only for preserving water rights of both surface and ground water users, but also for maintaining significant streamflow into Kanopolis Reservoir.

Since planning, designing, and operating water resource systems often involve several hydrological and water-use time series: multivariate stochastic analysis and multivariate models are used in this work to describe the inertia of the system. In the next chapter, univariate models are first applied to the system and used as an introduction to more elaborate processes such as transfer function models.

MATHEMATICAL MODEL

In this chapter, stochastic processes are applied to hydrologic and meteorologic data in order to analyze and simulate these natural phenomena. First, each data set is modeled as an autoregressive moving average process, and later, streamflow discharge is related to precipitation using the transfer function methodology.

Data collected for this study include two sets of mean monthly river flows and five sets of mean monthly precipitation series.

The river flow series were obtained from the U.S. Geological Survey surface-water gaging station #86405 (upper station) located at the westmost part of the study area and from station #8645 (lower station) located at the eastmost end. These hydrometric records range in length from 44 to 50 years.

The climatological information was obtained from the U.S. Department of Commerce, including records from the past 30 to 50 years. A general description of these series is given in Table 1 and their locations marked in Fig. 2.

DISCHARGE SERIES

The discharge series are made up of mean monthly values chronologically arranged according to water years. Since the length of the series is not the same, October 1940 is used as the starting record and September 1981 as the ending record.

Values of the upper station are subtracted from the corresponding values of the lower station, and a net gain series is obtained. This new series is called DIS and is a better representation of the interaction among the hydrological phenomena acting within the reach, because upstream and downstream effects have been eliminated. The DIS series thus obtained is next analyzed and modeled using autoregressive and moving average ARMA processes.

ARMA MODEL

According to the methodology introduced in previous chapters, the first stage in modeling a series is the identification stage.

Identification

The identification stage is divided into two steps: time series plot and ACF and PACF analyses.

1. <u>Time series plot</u>. The inspection of the DIS series plot (Fig. 3) shows this is a gaining stream with a well-defined yearly seasonality; high flows occur during the summer and low flows occur during the winter.

A long-term periodicity is also visible. This is an eleven-year periodicity. Although it is an interesting fact it is not surprising at all, since Wolf's sunspot numbers also have an eleven-year periodicity.

HYDROLOGICAL INFORMATON

PROCESS	STATION	CODE	YEARS	#OF YEARS
DISCHARGE	BUNKER HILL	UP	1940-1981	42
DISCHARGE	ELLSWORTH	LO	1930-1981	52
RAINFALL	RUSSELL	RU	1931-1958	26
RAINFALL	RUSSELL FA	RUA	1951-1981	30
RAINFALL	CLAFLIN	CL	1931-1981	50
RAINFALL	WILSON	WI	1951-1981	30
RAINFALL	ELLSWORTH	ELR	1931-1981	50
DISCHARGE		DIS	1940-1981	42
RAINFALL		RAIN	1940-1981	42

TABLE #1

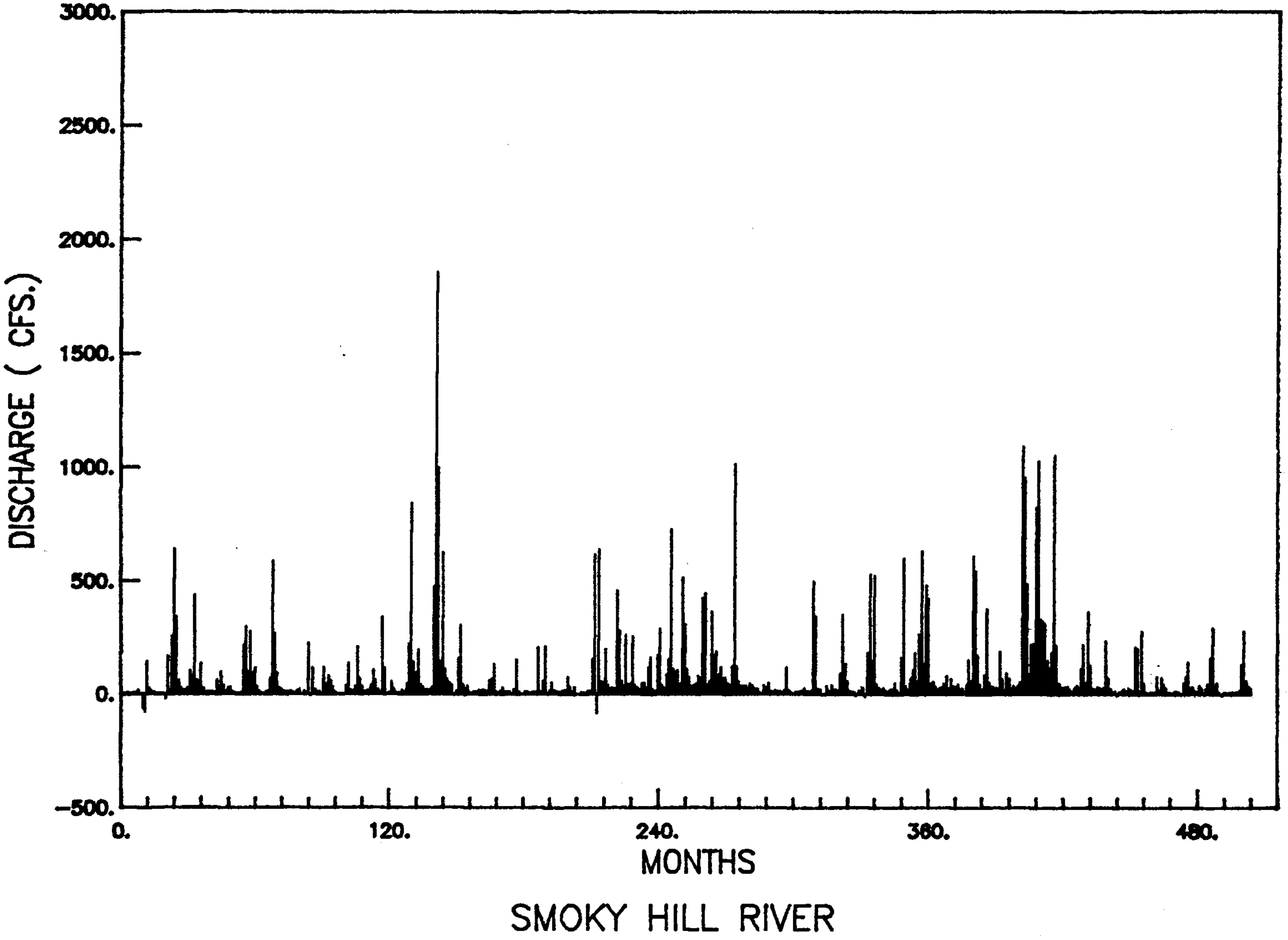


Fig. 3 Net gain discharge series

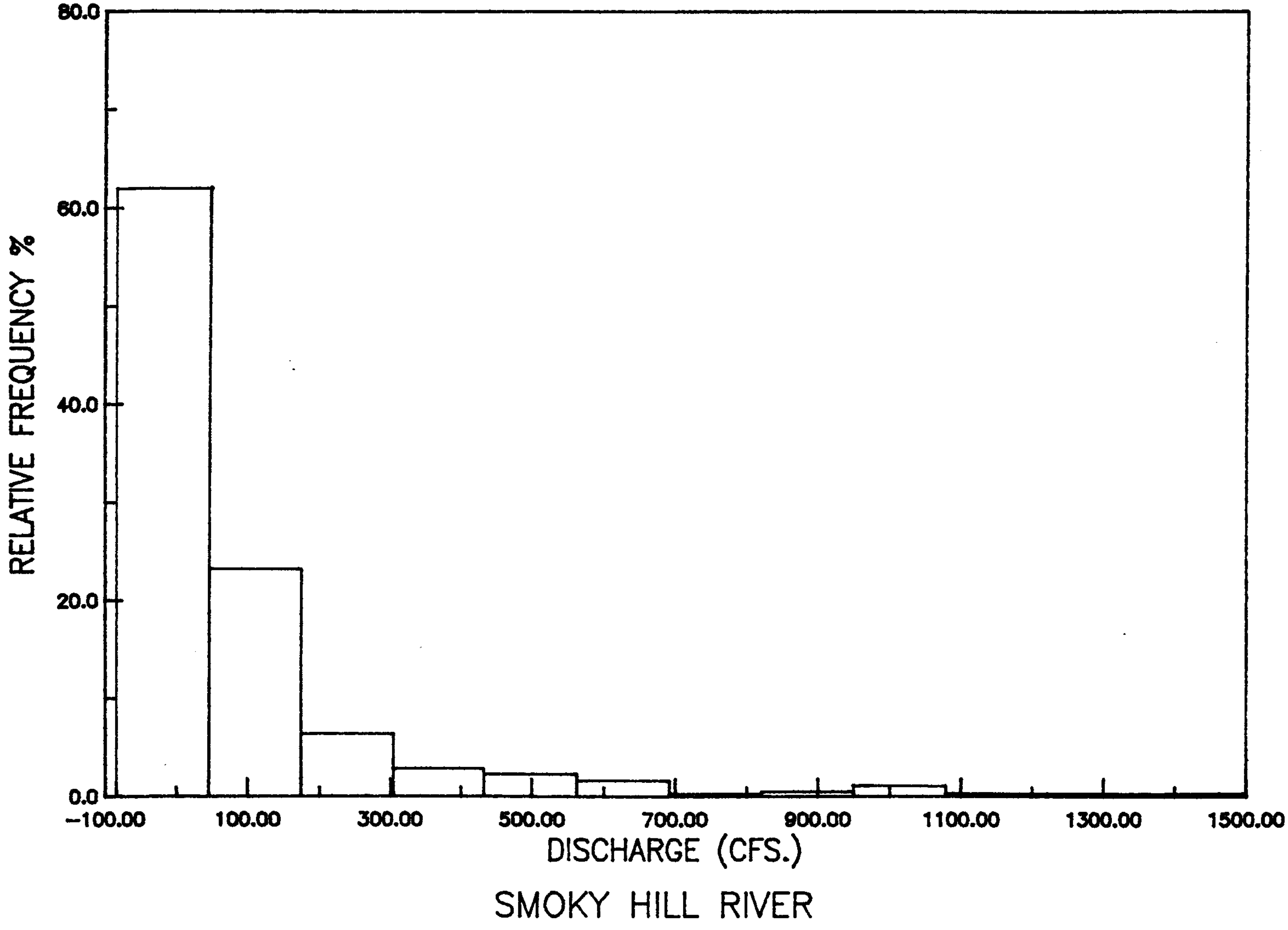


Fig. 4 Relative frequency of the transformed DIS series

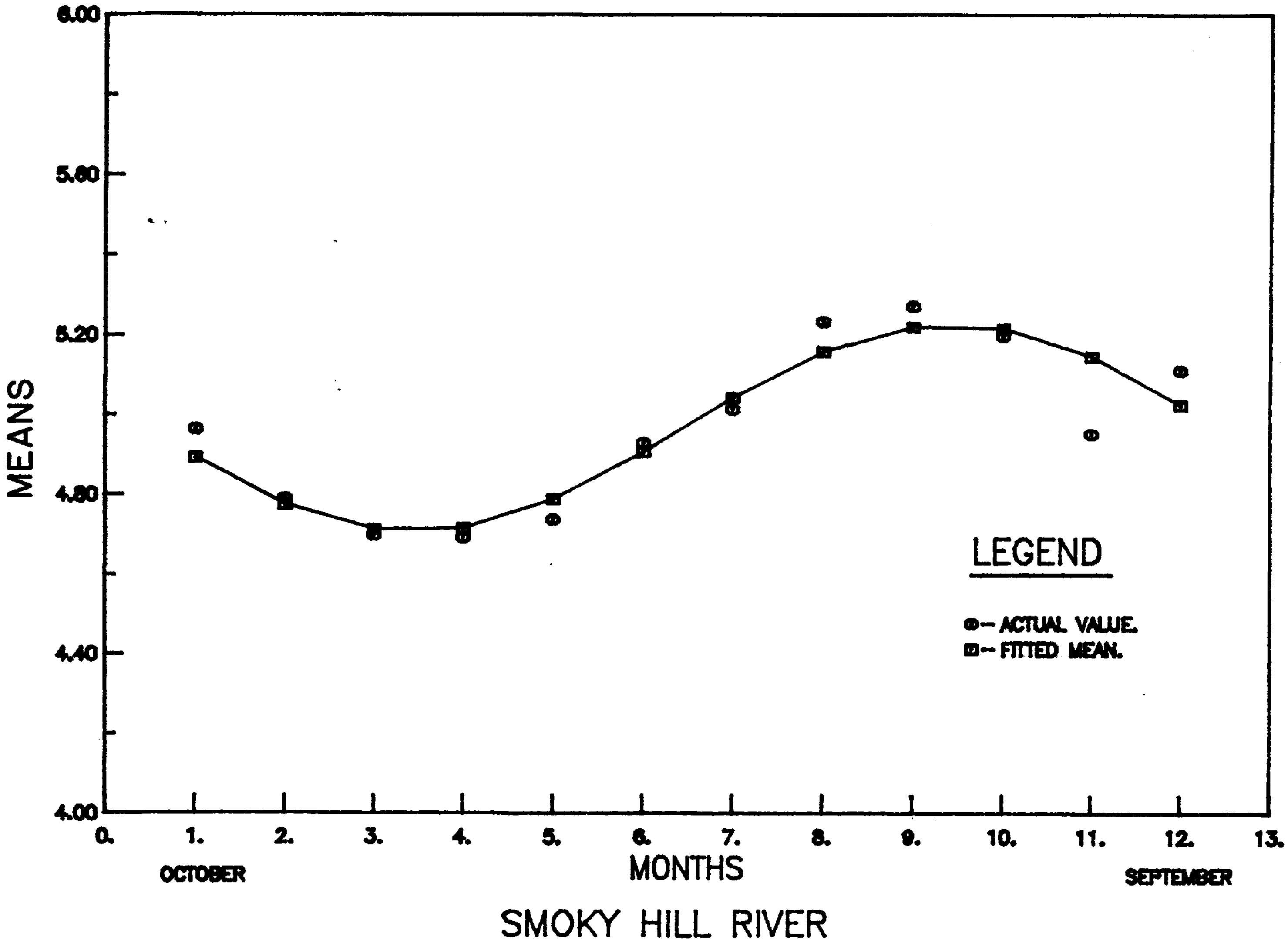
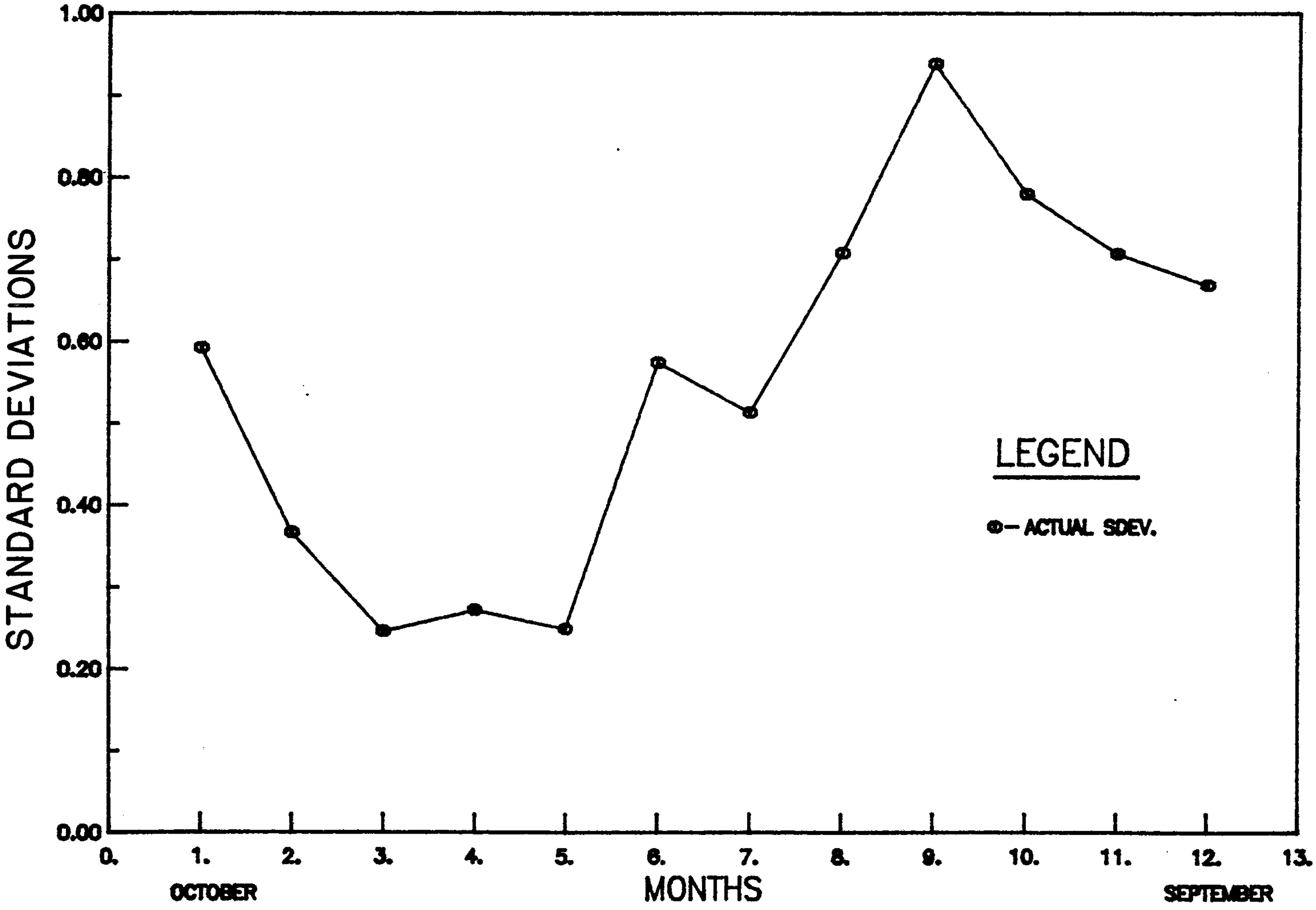
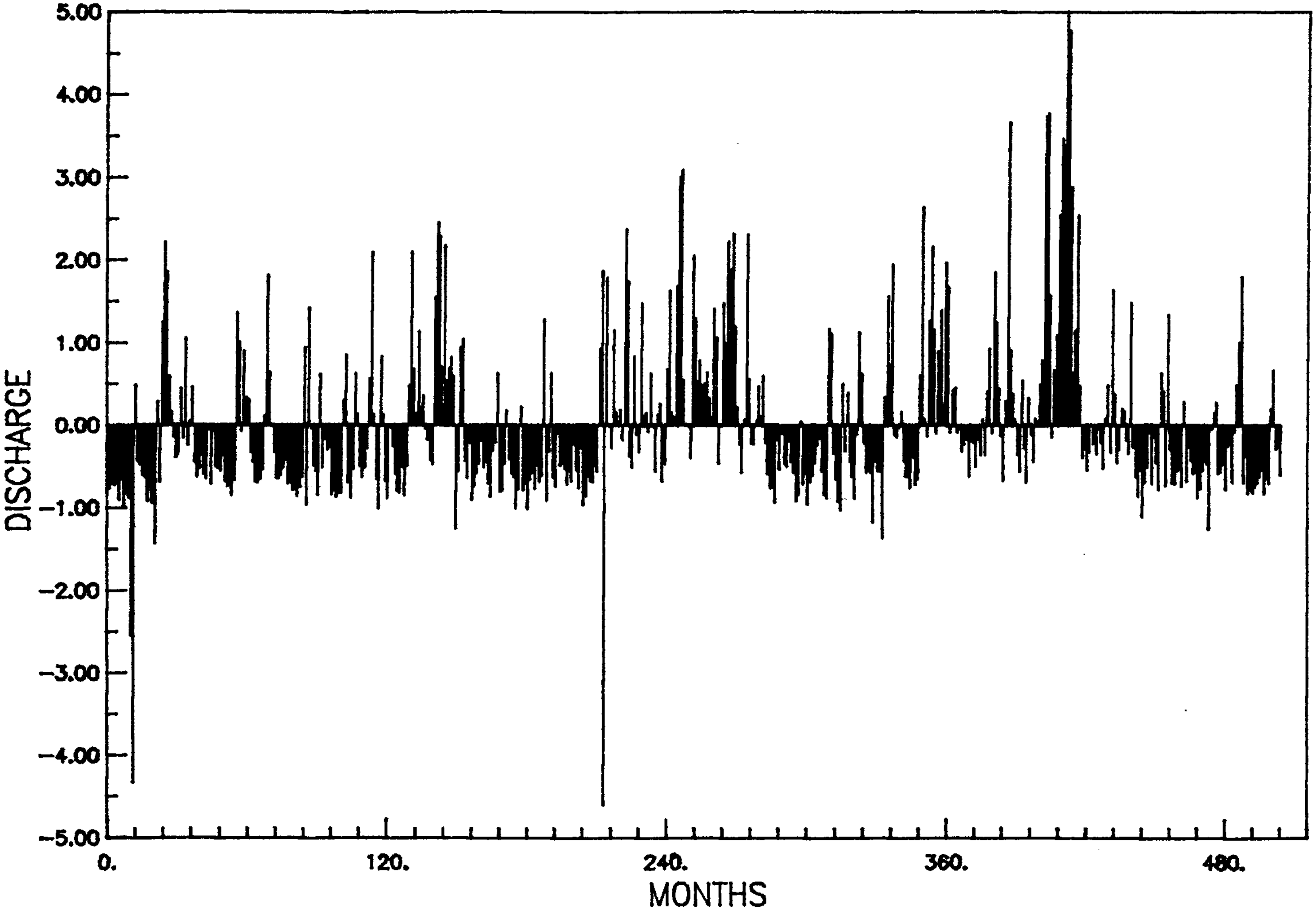


Fig. 5 Monthly means of the transformed DIS series



SMOKY HILL RIVER

Fig. 6 Monthly std. deviations of the transformed DIS series



SMOKY HILL RIVER

Fig. 7 Time series plot of the transformed and normalized DIS series

X(t) = original DIS series

Z(t) = transformed DIS series

and the coefficients of Skewness (G1) and Kurtosis (G2) are next calculated

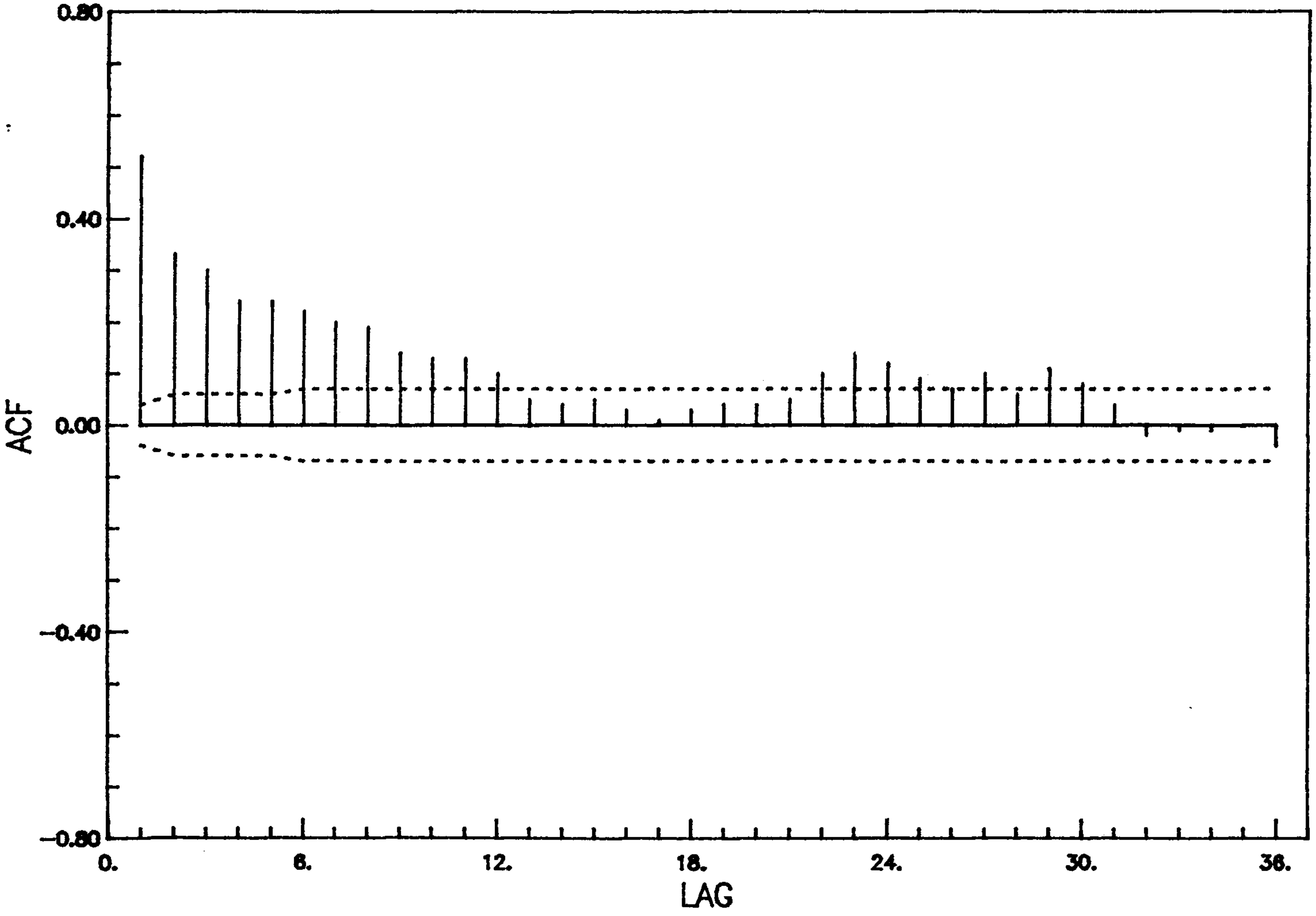
coef. Skewness (G1) = 0.6089

coef. Kurtosis (G2) = 5.8224

which suggest a symmetric distribution. Next, the periodicity of the DIS series is removed by subtracting out the monthly mean and dividing by the standard deviation. The twelve monthly means have been plotted in Fig. 5 and fitted with a sine wave using Fourier Analysis. Here the annual hydrological cycle of the DIS series is depicted with the low flows during December and January and the high flows during June and July. Fig. 6 shows the standard deviations of these monthly means conforming the hypothesis that large means are related to large standard deviations and, conversely, small means to small standard deviations.

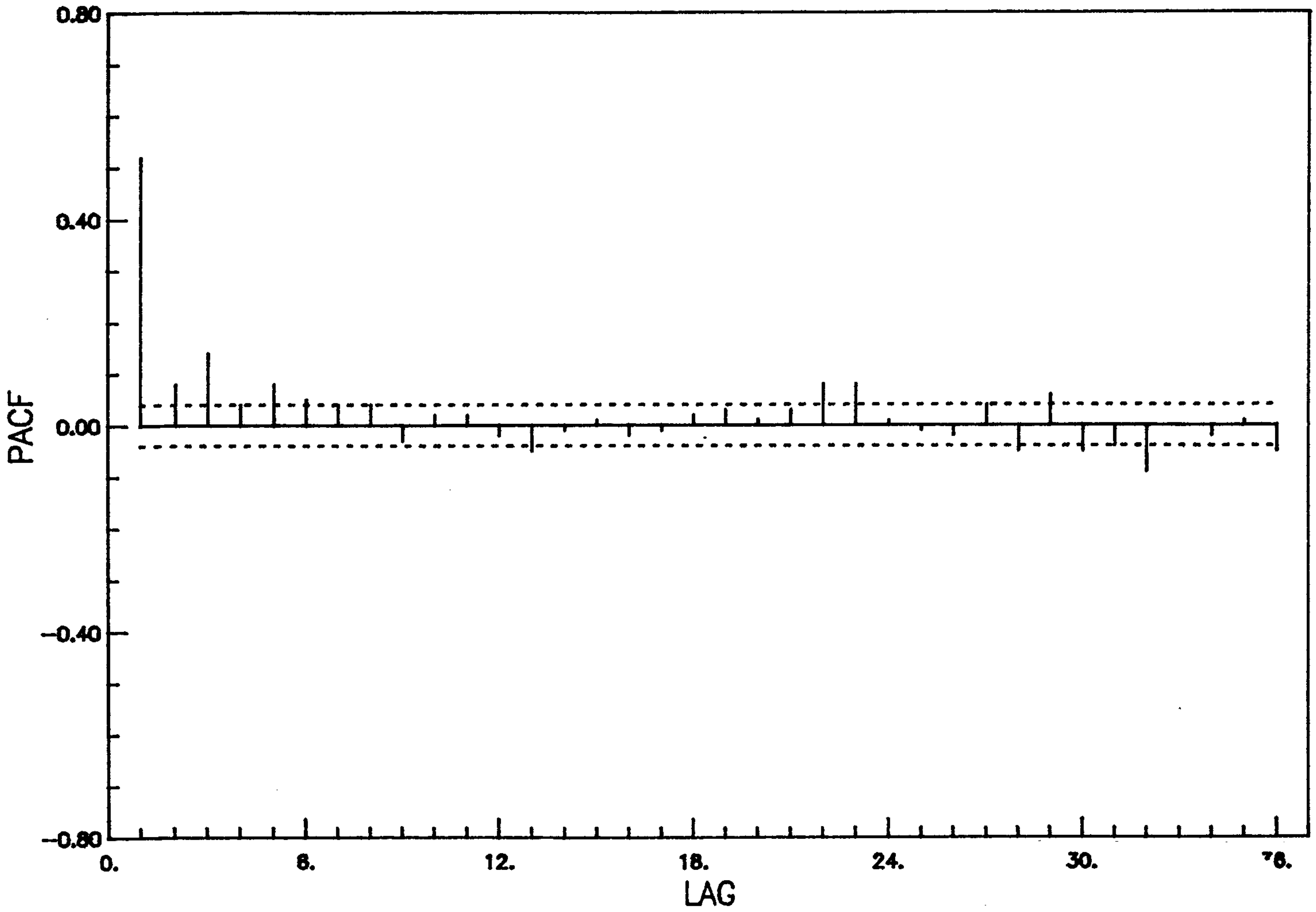
The transformed and deseasonalized DIS series is plotted in Fig. 7, and, it is evident, the mean is zero and about 2/3 of the values lie between +/- 1; moreover, neither trends nor periodicities are present. All these suggest that the DIS series is now a stationary, normally distributed process.

2. Autocorrelation and partial autocorrelation function. The ACF and the PACF plots present an alternative way to identify trends and periodicities and, what is even more valuable, permit in-



SMOKY HILL RIVER

Fig. 8 ACF of the transformed and normalized DIS series



SMOKY HILL RIVER

Fig. 9 PACF of the transformed and normalized DIS series

Estimation

The mathematical calculation of the parameters and of the tests have been all performed with the assistance of a computer frame. Although these calculations are not included the final results of these computations, either in a graph or in a table, are all presented. Tables containing the numerical values of the plots are included in the Appendix.

Several statistical programs available at the Computer Center of the University of Kansas were employed. However, one program, in particular, proved to be very efficient and easy to use in the Time Domain analysis of time series. This program is called P2T and is part of the Biomedical Computer Programs (BMDP).

Diagnostic Check

Several diagnostic checks tested the validity of the model in Eq. 38 and found it inadequate. There are strong autocorrelations which had not been accounted for. Table 2 lists a number of models and their corresponding AIC and Q1 values. As was pointed out during the analysis of the ACF, the inclusion of MA terms greatly improve the fit. As recommended in chapter #2, the model with the smallest Akaike Information Criterion (AIC) is selected (below):

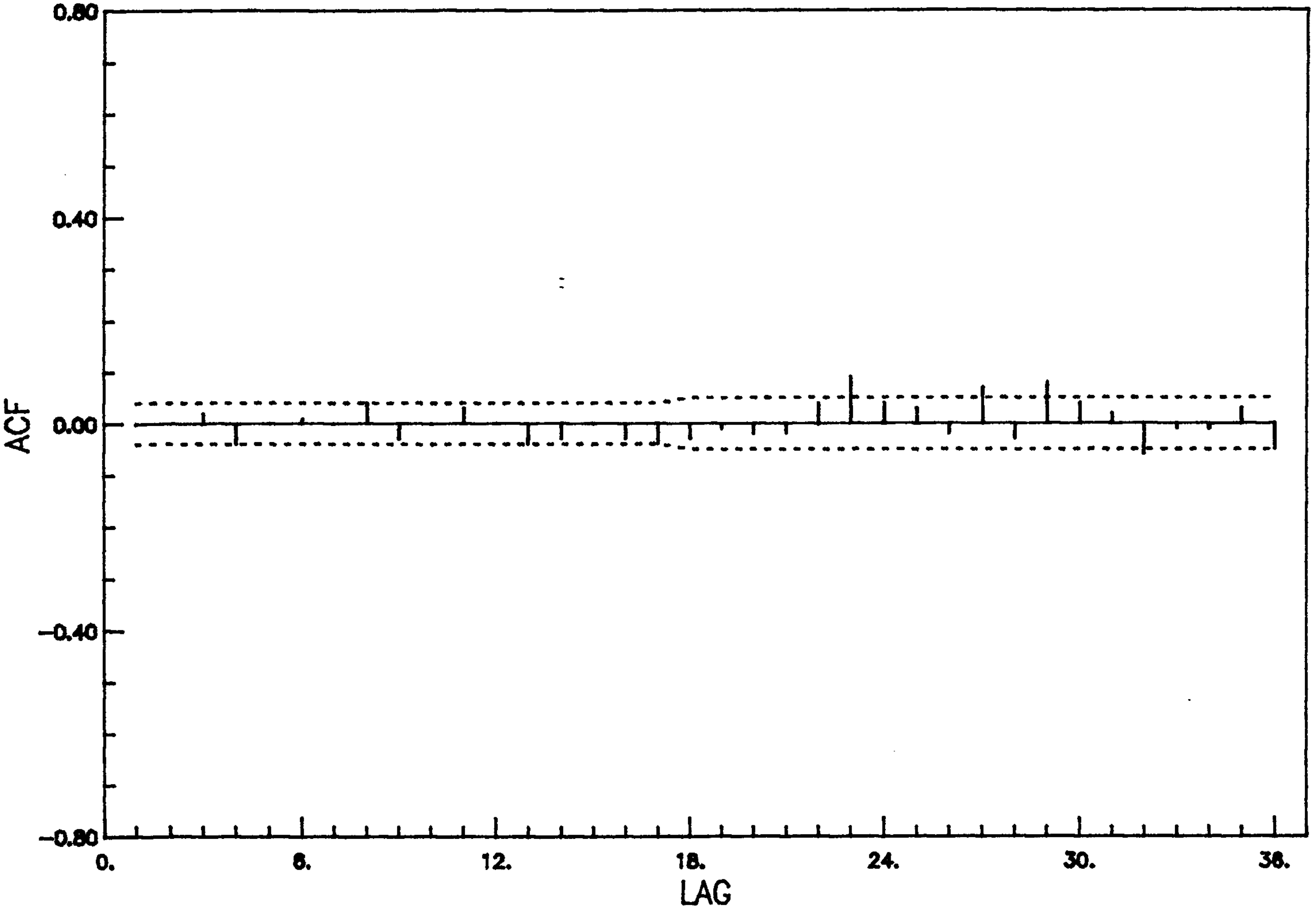
$$\begin{gathered}(1-0.9014*B)*Z(t) = \\ (1 -0.4429*B -0.2088*B^2)*a(t)\end{gathered} \tag{39}$$

DIS

ARMA MODELS

#	AR	MA	AIC	Q1	T-10%	T-5%
1	0	0	1.28	443.2	47.19	50.96
2	1,3	0	-165.18	27.22	44.88	48.57
3	1	1	-160.60	33.26	44.88	48.57
4	1	1,2	-169.82	22.08	43.73	47.37

TABLE #2



SMOKY HILL RIVER

Fig.10 ACF of the residuals from DIS ARMA(1,2) model represented by X1, X2, X3, X4, and X5, respectively, the spatial average equation becomes

$$X = ((A1/A)*X1 + (A2/A)*X2 + (A3/A)*X4 + (A5/A)*X5) \tag{41}$$

These values are in inches, and to convert them to CFS (cubic-feet-second) they are multiplied by the total area A (miles^2) and by C, a constant. That is

$$Y = X*A*C \tag{42}$$

Thus,

$$Y = C*(A1*X1 + A2*X2 + A3*X3 + A4*X4 + A5*X5) \tag{43}$$

where

$$C = 0.8963 \text{ (cubic-feet-second / inches-miles}^2\text{-month)}$$

Records only from October 1940 to September 1981 are used.

During the period of 1940-1951 there were three stations in the area of interest (Fig. 11). Therefore Eq. 41 is

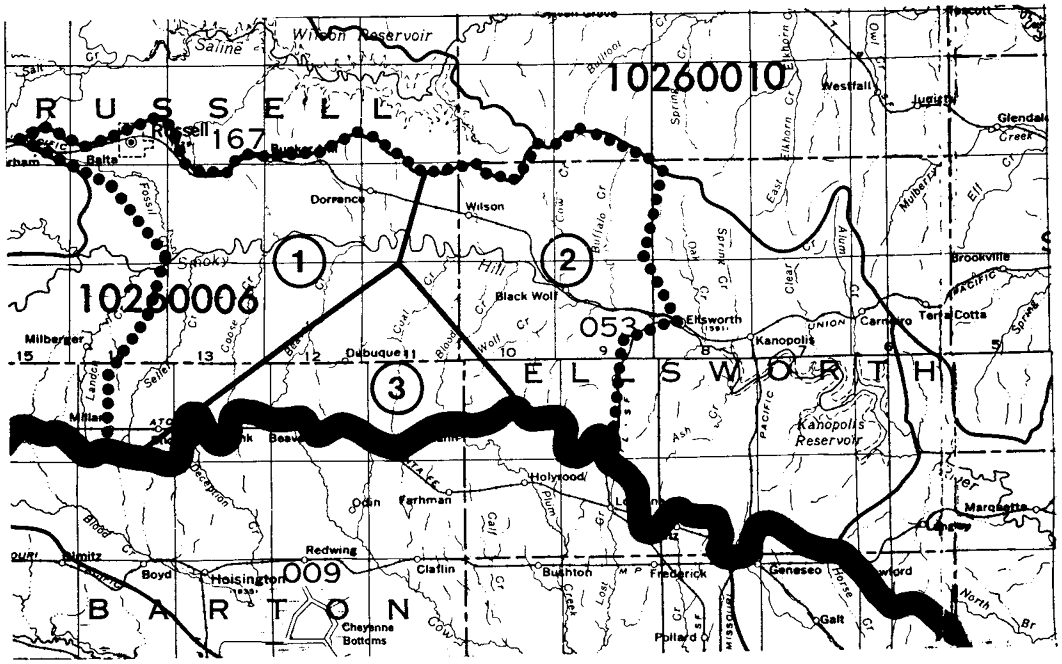


Fig.11 Thiessen polygons for the rainfall series, 1940-1951

$$Y=0.8963*(230.7*RU + 126.7*CL + 207.6*ELR) \qquad (44)$$

In 1952-1981, there were 4 stations (Fig. 12). This period is represented by

$$Y=0.8963*(182*RUA + 146*ELR + 112*CL + 125*WI) \qquad (45)$$

where

RU = Russell st.
CL = Claflin st.
ELR = Ellsworth st.
RUA = Russell st. (after 1952)
WI = Wilson st.

Eqs. 44 and 45 represent the value of each record for the new precipitation series for the periods of 1940-1951 and 1952-1981, respectively. This new series is called RAIN and is expressed in CFS. In the following sections this series is studied and modeled using moving average and autoregressive processes.

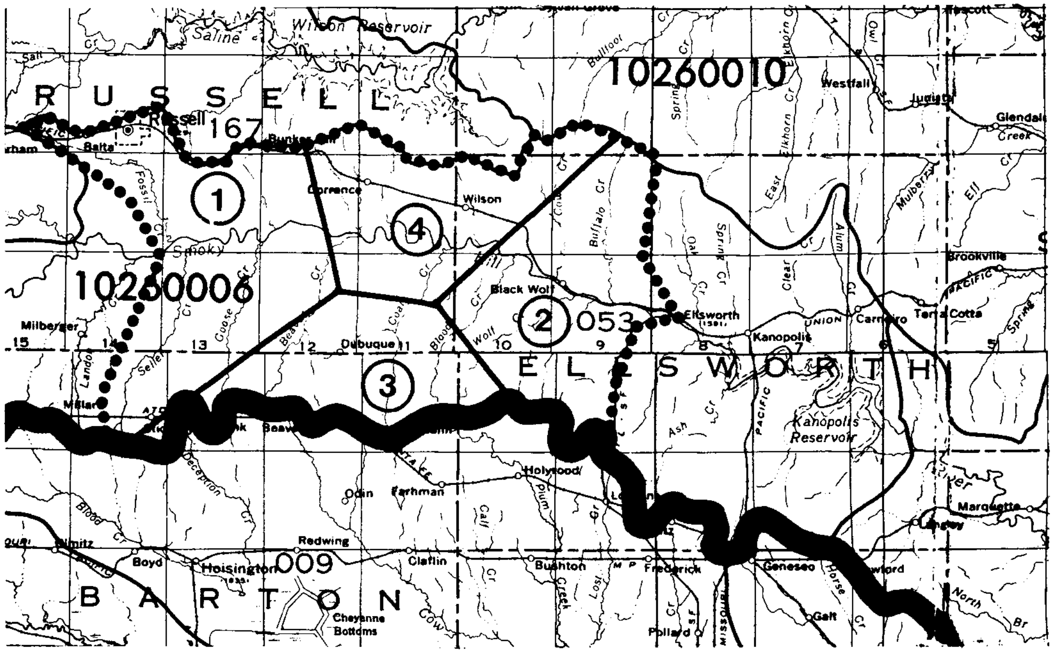


Fig.12 Thiessen polygons for the rainfall series, 1952-1981

ARMA MODEL

The purpose of this section is to develop a stochastic model which realistically describes the rainfall process in the study area. The three stages: identification, estimation, and diagnostic checking guide the construction of the model.

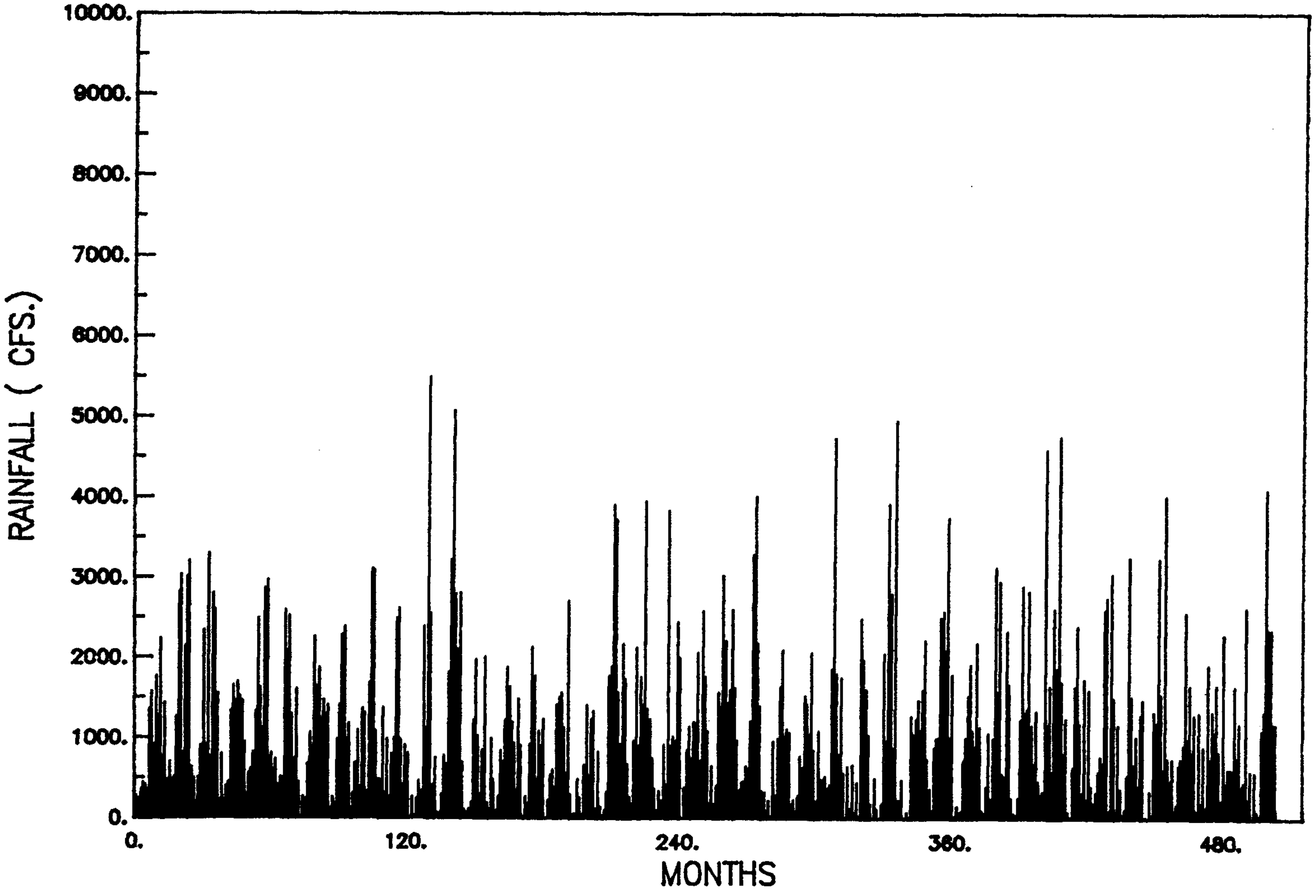
Identification

The selection of a model to fit the RAIN series is divided into two phases:

- the analysis of the time series plot and,
- the analysis of the autocorrelation and partial autocorrelation functions.

1. <u>Time series plot</u>. The analysis of the time series plot Fig. 13 shows a yearly seasonality, but long-term periodicities are not clearly visualized. It is evident the series is not stationary. Moreover, since the RAIN series does not have negative values, it is not a symmetrical distribution (Fig. 14). In summary, the RAIN series needs to be normalized and deseasonalized before proceeding with the selection of a stochastic model.

<u>Deseasonalization</u> <u>and</u> <u>normalization</u>. The RAIN series is normalized using the Box and Cox transformation. A constant (e) is first added to the series before taking the natural logarithms, say



SMOKY HILL RIVER

Fig.13 Time series plot of the average rainfall series

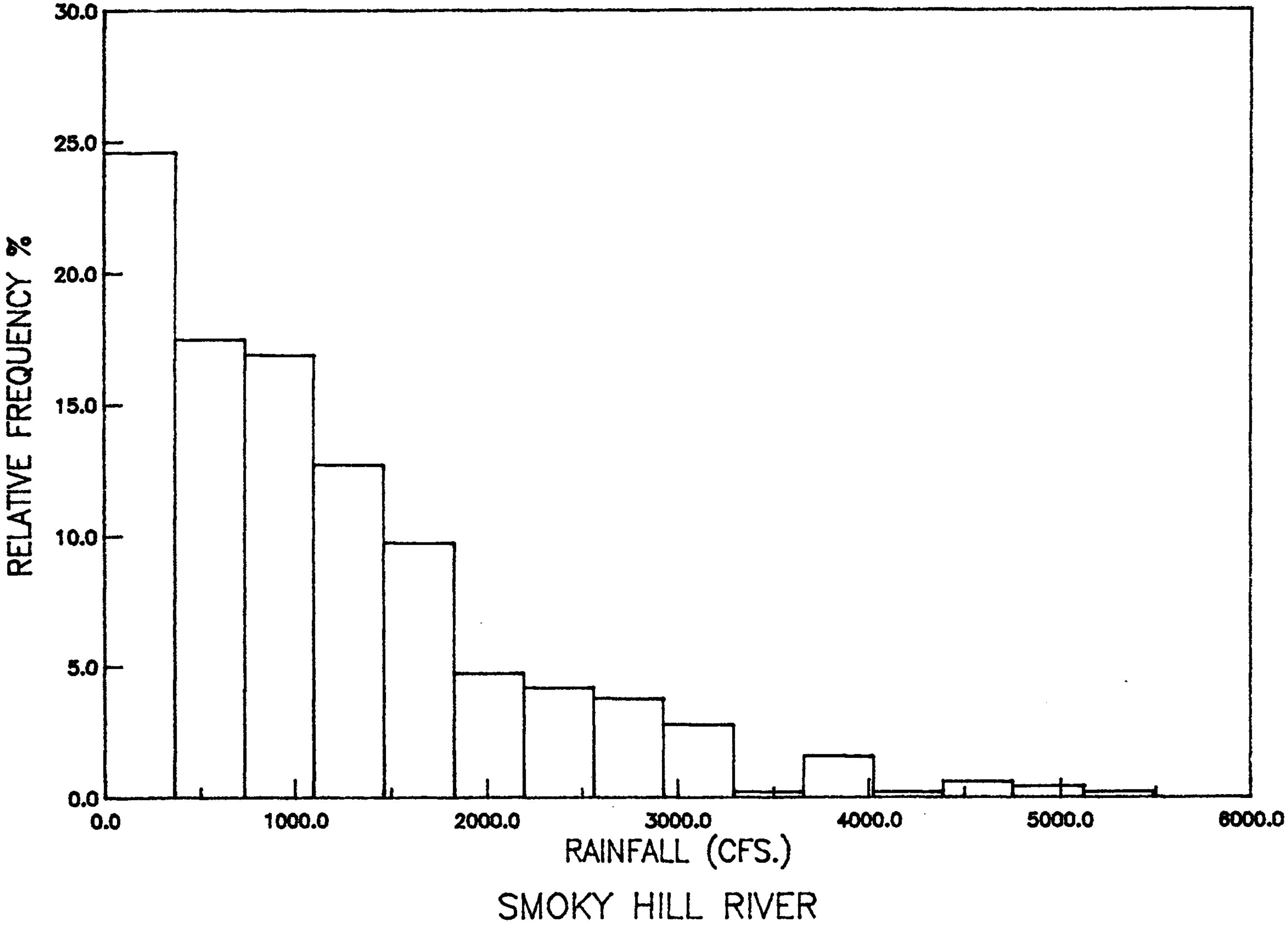


Fig.14 Relative frequency of the average rainfall series

$$X(t) = \ln(Y(t) - e) \tag{46}$$

Eqs 3, 4, and 5 are used to obtain the constant (e)

constant (e) = -1148.84
mean (m) = 7.649
standard deviation (S) = 0.3917

and the series is transformed by

$$X(t) = \ln (Y(t) + 1148.89) \tag{47}$$

where

$X(t)$ = transformed RAIN series
$Y(t)$ = original RAIN series
ln = natural logarithm

The seasonality (Figs. 15, 16) is removed by subtracting from each observation the corresponding monthly mean and then dividing by the corresponding standard deviation. A time series plot of the transformed and deseasonalized series is presented in Fig. 17, where it is evident that the mean is zero and the standard deviation is one.

It is interesting to study the annual cycle of the RAIN series Fig. 15; however, it is more valuable to analyze it in conjunc-

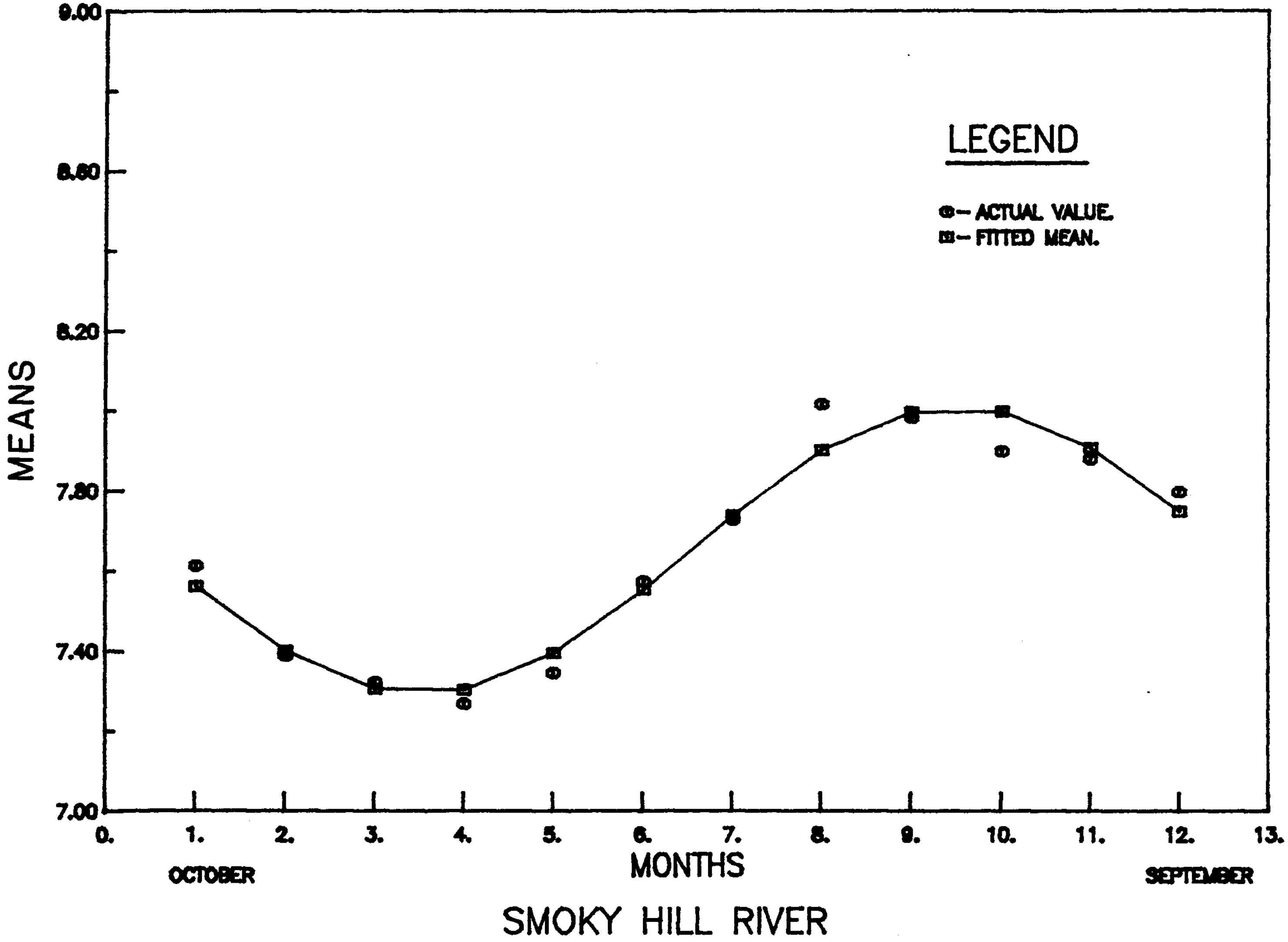
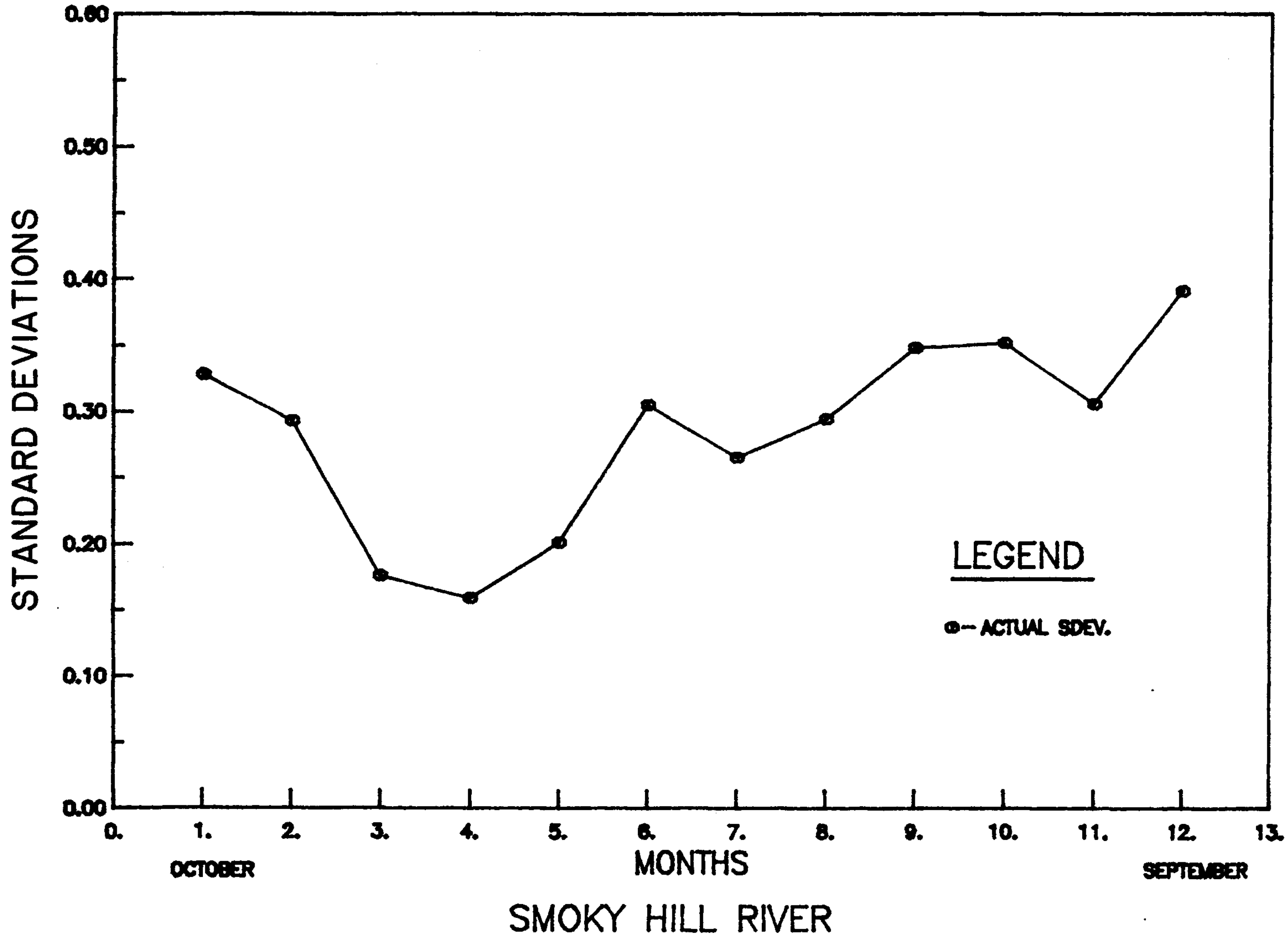
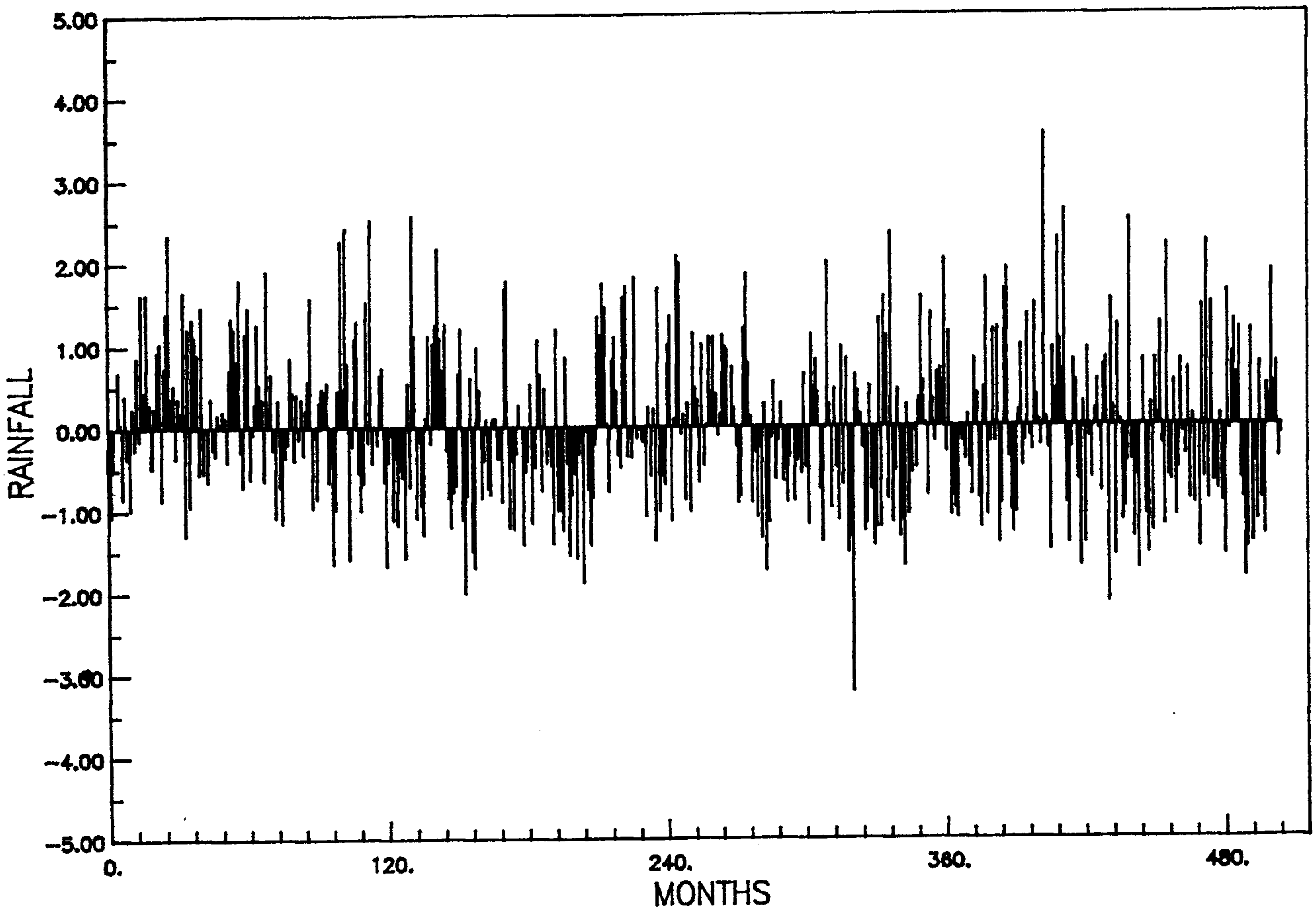


Fig.15 Monthly means of the transformed RAIN series



SMOKY HILL RIVER

Fig.16 Monthly std. deviations of the transformed RAIN series



SMOKY HILL RIVER

Fig.17 Time series plot of the transformed and normalized RAIN series tion with the annual cycle of the DIS series (Fig. 5):

- both the RAIN and the DIS series follow a sine wave pattern: high values and large standard deviations during the summer, and low values and low standard deviations during the winter,

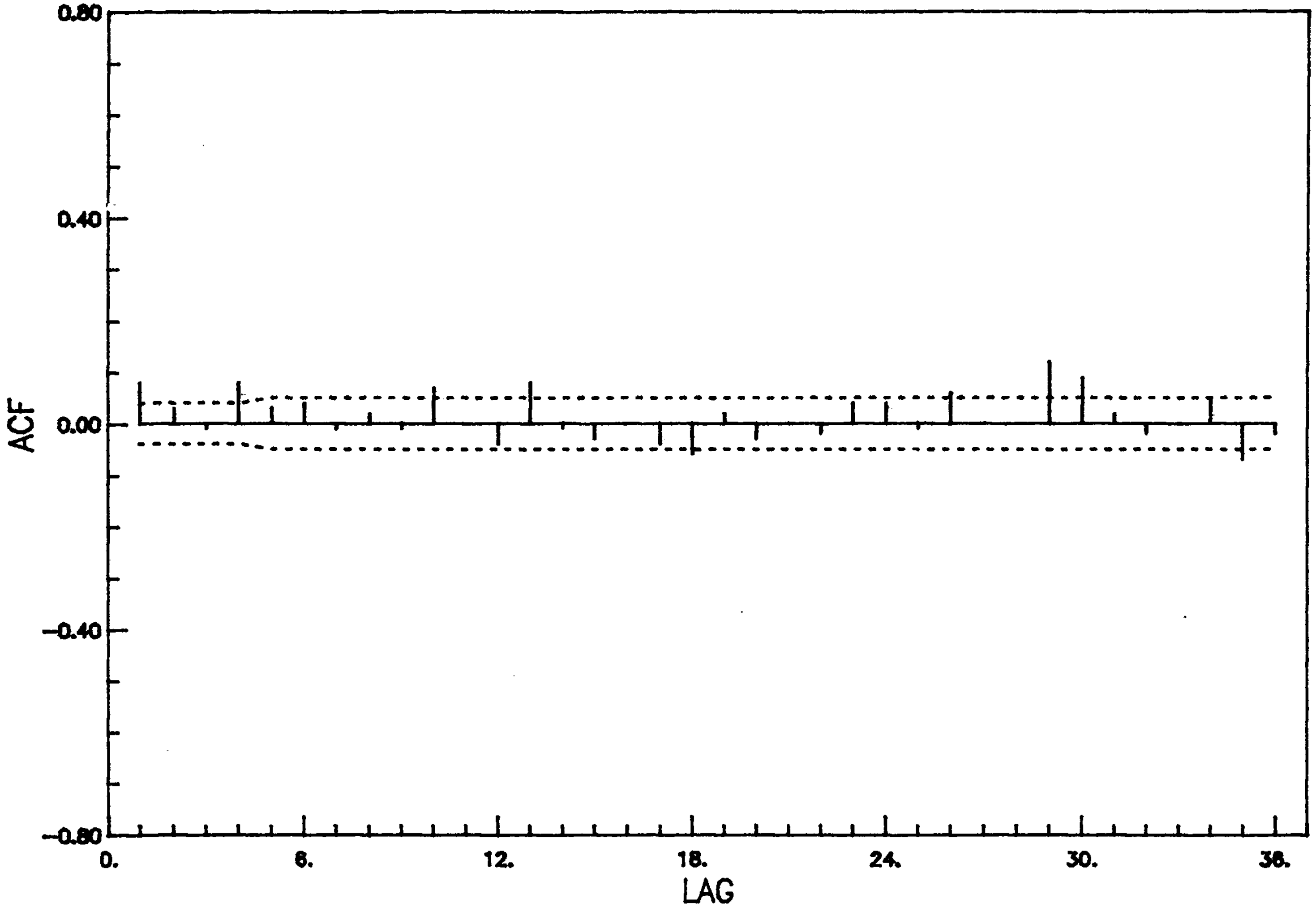
- the RAIN series shows larger values than the DIS series, suggesting the rainfall is the major source of recharge in the study area;

- the general trend and the relative monthly changes present in the RAIN series are well imitated by the DIS series, especially from October through May. This is an indication of a small and constant ground water contribution into the stream;

- from July through September, the DIS series does not parallel the behavior of the RAIN series, mainly due to influences of other climatologic phenomena, e.g. temperature. Nevertheless, the DIS series is constantly affected by the RAIN series.

In the following paragraphs the ACF and the PACF are investigated to provide inferences about the type of process required for this time series.

2. Autocorrelation and partial autocorrelation function. Having transformed and deseasonalized the RAIN series the analysis of the ACF and PACF permit the verification of the stationarity of the series, and inferences about the parameters and processes required in the modeling of the RAIN series. The ACF plotted in Fig. 18 shows that the series is indeed stationary. Because the ACF cuts off at lag 4, it suggests a moving average process of order 4. However, the PACF



SMOKY HILL RIVER

Fig.18 ACF of the transformed and normalized RAIN series

(Fig. 19) cuts off also at lag 4 and suggests an autoregressive process instead. This type of behavior is quite common in mixed models (Box and Jenkins, 1970). Therefore, the model selected for this series is an ARMA(1,1)

$$(1-A*B) = (1-F*B)*a(t)$$

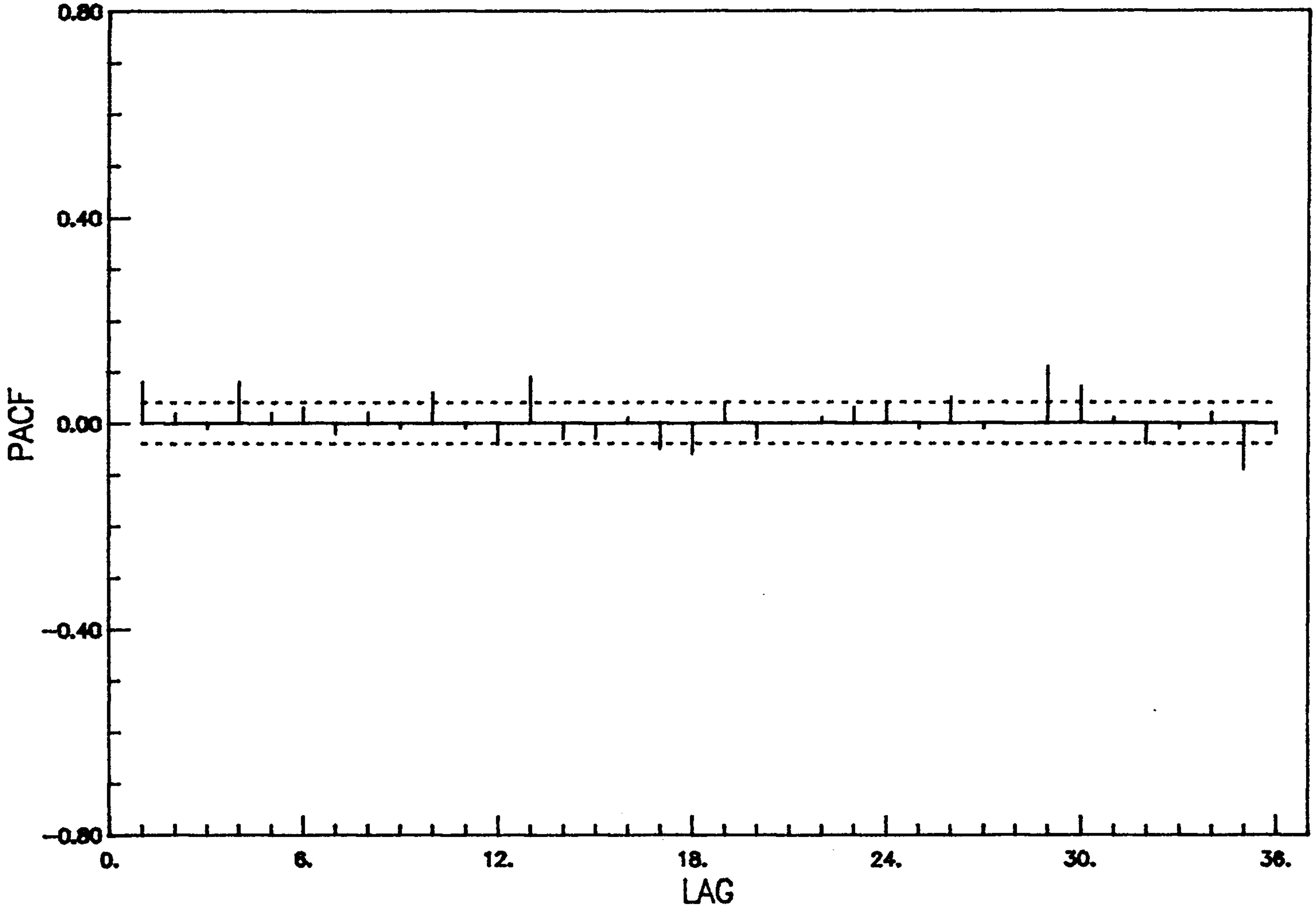
Estimation and Diagnostic Checking

This tentative model was estimated using the BMDP package, as has been explained, and was tested to verify its adequacy. Failure of this model to satisfy a test in the independence of the residuals demonstrated that it needed to be modified. Several other models were tried and tested (see Table 3).

According to the information in this table, the best model is a moving 4-th order average with parameters 1 through 3 constrained to zero. However, a t-test of the significance of this last parameter F_4 demonstrated it is not different from zero and should be dropped from the model.

In Table 4 there is a summary of the MA(4) model. Here, F_4 is equal to -0.074, and its standard deviation is equal to 0.0445. The F_4 parameter is less than 2 standard deviations, which means F_4 is equal to zero with a 95% probability.

The next best model is an ARMA(0,0). This means the RAIN series is a process with no statistical structure and can not be modeled. The RAIN series is a white noise process.



SMOKY HILL RIVER

Fig.19 PACF of the transformed and normalized RAIN series

RAIN

ARMA MODELS

#	AR	MA	AIC	Q1	T-10%	T-5%
1	0	0	0.64	35.68	47.19	50.96
2	1	1	2.06	32.76	44,88	48.57
3	1	4	2.11	32.56	44.88	48.57
4	0	4	0.11	26.41	46.04	49.77

TABLE #3

RAIN

SUMMARY OF THE MODEL

VARIABLE	TYPE	FACTOR	ORDER	ESTIMATE	STD.ERR.	T-RATIO
RAIN	MA	1	4	-0.0740	0.0445	-1.66

RESIDUAL SUM OF SQUARES = 501.11719

DEGREES OF FREEDOM = 503

RESIDUAL MEAN SQUARE = 0.996257

TABLE #4

$$Y(t) = a(t) \qquad (48)$$

Having selected a univariate ARMA model for each of the series, RAIN and DIS, the next part of this study is intended to obtain a multivariate model to simulate the relationships between precipitation and stream discharge.

TRANSFER FUNCTION NOISE MODEL

In the previous sections of this work an univariate model was used to relating the current realization of a phenomenon to previous realizations. The following sections are devoted to relate one time series to a second time series. Specifically, the DIS series is expressed as a linear function of the RAIN series.

Recalling Eq. 17,

$$Y(t) = d^{-1}(B)*w(B)*B^{b} *X(t) + N(t)$$
$$d(B) = 1-d_1*B-d_2*B^2-\dots\dots-d_r*B^r$$
$$w(B) = w_0-w_1*B -w_2*B^2-\dots\dots- w_s*B^s$$

where Y(t) and X(t) represent the DIS and the RAIN series respectively. The ratio of the two polynomials is the response function, and N(t) is a noise component.

The procedure of choosing a transfer function is considerably simplified if the three stages of model development are systematically met.

Identification

At this stage the general behavior of the series is studied. Special attention is focused in the stationary and normality conditions of each of the series. From previous analyses done in those series, it is known they need to be normalized and deseasonalized before proceeding with the identification process.

The second part in the identification process refers to the selection of a tentative response function. Two methods are employed.

<u>Single pre-whitening</u> (<u>Box and Jenkins</u>, <u>1970</u>). This method is as follows:

- The RAIN series is modeled with an ARMA process. Based on previous analyses it is known the RAIN series is a white noise process, that is

$$X(t) = a(t)$$

- The output series DIS is filtered through the RAIN model, and a residual is obtained. However, since there is no model for the RAIN series, the DIS series remains unchanged.

- The residuals obtained from both series are cross correlated. Because there are no residuals, the series themselves are used to compute the CCF. This cross correlation function is plotted in Fig. 20. The 95% confidence limits are included. The analysis of this CCF reveals that

- There is a strong cross correlation at lag zero, which indicates the time lag is either zero or less than one month and can not be properly identified with the current information.

- The CCF decays exponentially.

- Although the CCF eventually damps out, it does not converge fast enough to zero, suggesting a rather large number of terms in the w(B) polynomial of the transfer function. This slow decay of the CCF is probably due to the autocorrelation of the DIS series which has not been removed and is inflating and distorting the CCF.

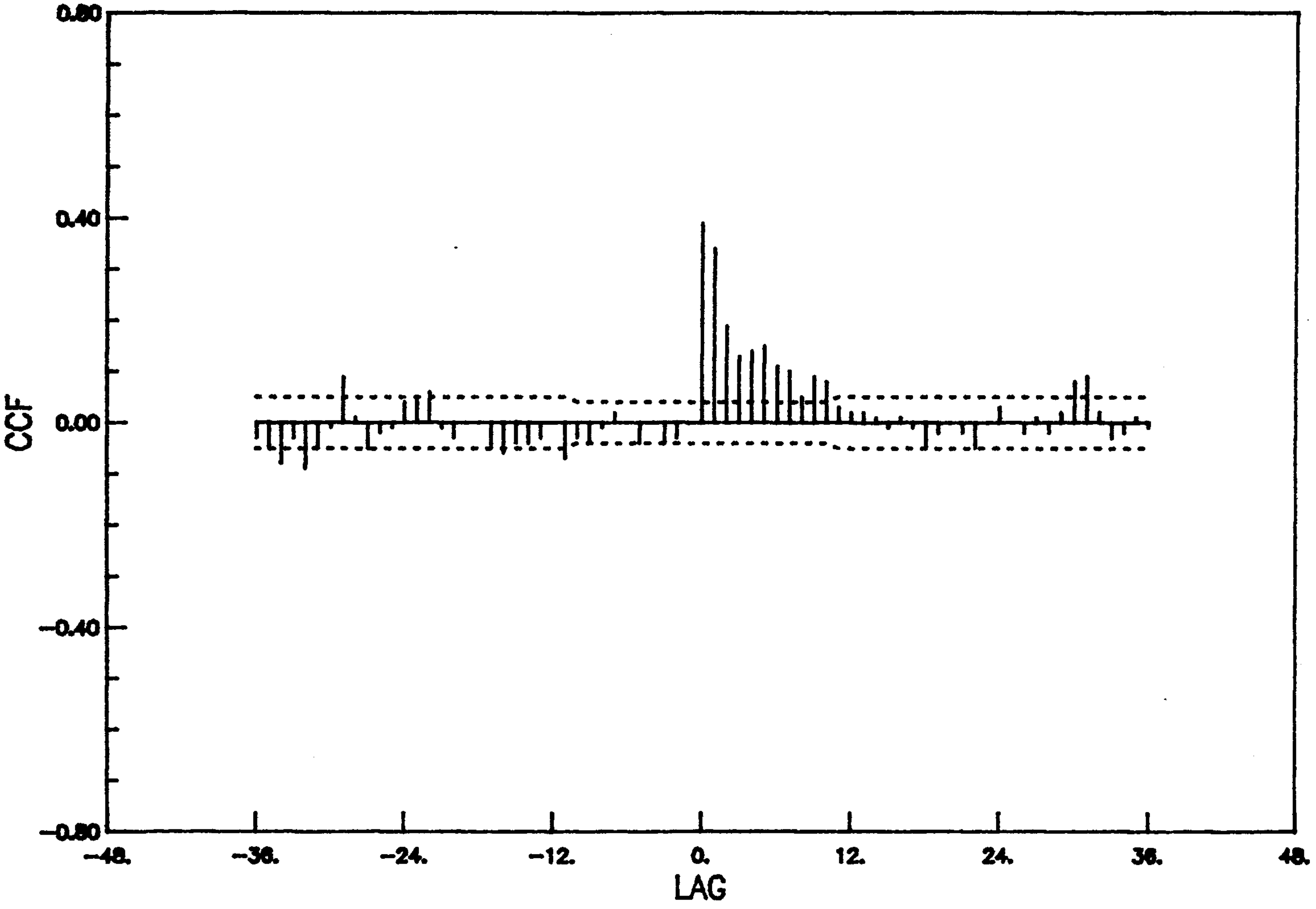
This method is not valid for this situation and is not used in the selection of the transfer function.

In the next method the intrastructure of the series is removed before computing the CCF.

Double pre-whitening (Hipel et al., 1977). This method is as follows:

- An appropriate ARMA model is fitted to each of the series. According to Eqs. 39, and 48

DIS series,



SMOKY HILL RIVER

Fig.20 CCF between Transformed and normalized RAIN and DIS series

$$Y(t) = 0.90*Y(t-1) + g(t) - 0.44*g(t-1) - 0.21*g(t-2) \qquad (49)$$

RAIN series,

$$X(t) = a(t) \qquad (50)$$

- The white noise residuals are then obtained and cross correlated.

The white residual for the DIS series is computed from Eq. 49:

$$g(t) = (1 - 0.44*B - 0.21*B^2)^{-1} *(1 - 0.90*B)*Y(t) \qquad (51)$$

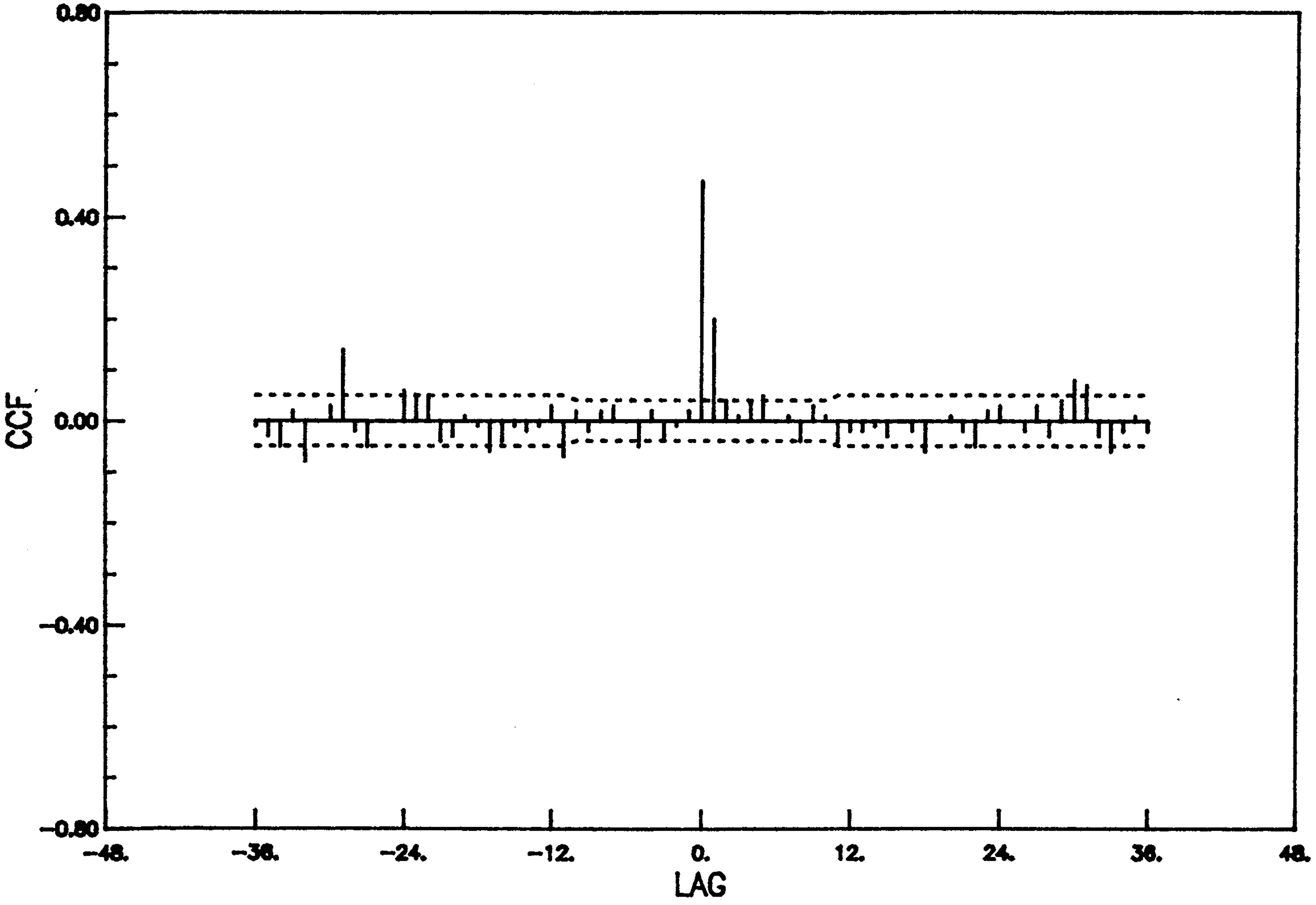
The white residual for RAIN series is computed from Eq. 50:

$$a(t) = X(t) \qquad (52)$$

The CCF between a(t) and g(t) is computed and plotted in Fig. 21. The 95% confidence limits are included.

From the analysis of the CCF it is possible to conclude that

- there is not a time lag, or it is less than one month, therefore,



SMOKY HILL RIVER

Fig.21 CCF between transformed and normalized RAIN series and DIS residuals from ARMA(1,2) model

$$b = 0$$

- the CCF has an exponential decay, suggesting

$$r = 1$$

- the CCF is significant at lags 0 and 1, also that

$$V_3 = d_1 * V_2 \tag{53}$$

$$V_2 = d_1 * V_1 \tag{54}$$

which suggests that V_1 is the first weight of the differential Eq. 16, therefore,

$$b + s + 1 = 1$$

and, hence

$$s = 0$$

V_0 becomes the only startup weight for the differential Eq. 16, and, therefore,

$$b + s - r + 1 = 0$$

confirming that,

$$r = 1$$

Having selected the values of b, s, and r, Eq. 15 can be written as

$$(1-d_1B)*(V_0+V_1*B) = w_0 \tag{55}$$

hence

$$w_0= V_0 \tag{56}$$

$$d_1= V_1/V_0 \tag{57}$$

The transfer function model now can be expressed as

$$Y'(t) = (1-d_1*B)^{-1}*(w_0)*X(t) \tag{58}$$

In general the output series Y(t) does not necessarily agree with the Y'(t) series. Discrepancies arise because of the numerous variables affecting the phenomenon which have been neglected, for obvious reasons, and also because the model, so far, does not account for the stochastic behavior of the process.

In the next section a noise term N(t) is added to the transfer function, thus incorporating into the model the effects of unknown phenomena. This noise term is

$$N(t) = Y(t) - Y'(t) \tag{59}$$

ARMA model of the residuals N(t). The residuals obtained from Eq. 59 are modeled using autoregressive and moving average processes. The study of the ACF and PACF reveals that the N(t) series is best modeled by an ARMA(1,1). This ARMA model and the already identified transfer function model Eq. 58 form the first tentative transfer function noise model

$$Y(t) = (1-d_1*B)^{-1} *(w_0)*X(t) + (1-A_1*B)^{-1} *(1-F_1*B)*a(t) \tag{60}$$

Estimation and Diagnostic Checking

Following estimation, several statistical tests were performed to verify the independence of the residuals. The analysis of the ACF of the residuals proved that the model selected above is inadequate and suggested the expansion of the MA polynomial, for example,

$$Y(t) = (1-d_1*B)^{-1}*w_0*X(t) + (1-A_1*B)^{-1}*(1-F_1*B -F_2*B^2)*a(t) \tag{61}$$

This and other models have been estimated and tested (Table 5). From the analysis of this table it is clear that the best model is

$$Y(t) = (1-d_1*B)^{-1}*w_0*X(t) + (1-A_1*B)^{-1}*(1-F_1*B-F_2*B^2)*a(t) \tag{62}$$

TRANSFER FUNCTION

NOISE MODELS

#	S	R	AR	MA	AIC	Q2	T-20%	Q3	T-20%
1	0	1	0	0	-183.46	615.17	41.70	18.66	40.78
2	0	1	1	1	-307.64	38.80	39.90		
3	0	1	1	1,2	-318.93	29.62	38.97	30.02	40.78

TABLE #5

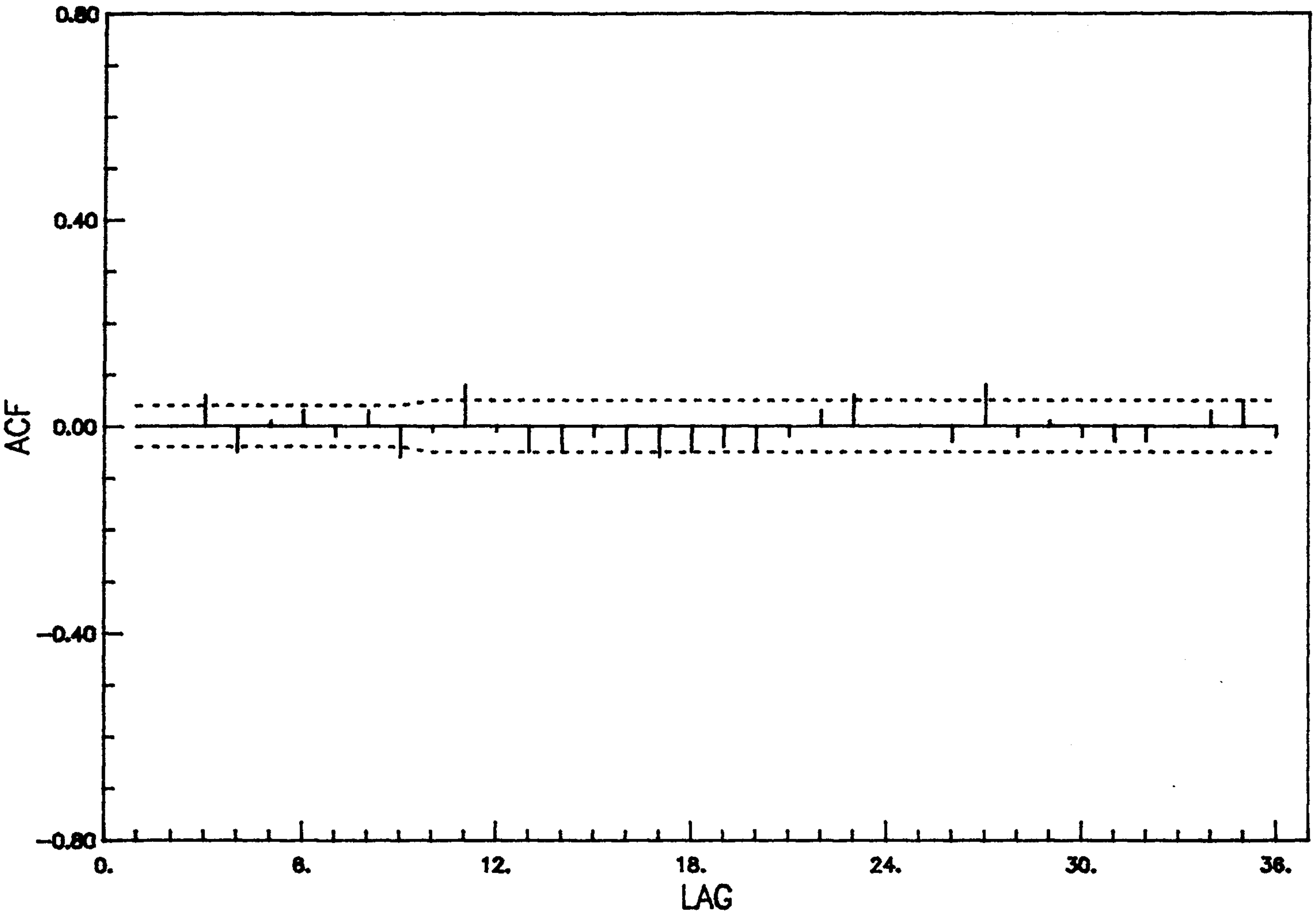
After substituting the values of the parameters and arranging some terms (below)

$$Y(t) = 0.75*Y(t-1) + 0.37*X(t) + (1-0.98*B)^{-1}*(1-0.75*B)*(1-0.66*B-0.19*B^{2})*a(t) \qquad (63)$$

this equation states that the present realization of the discharge is strongly correlated to the past realization of discharge and also to the present occurrence of rainfall.

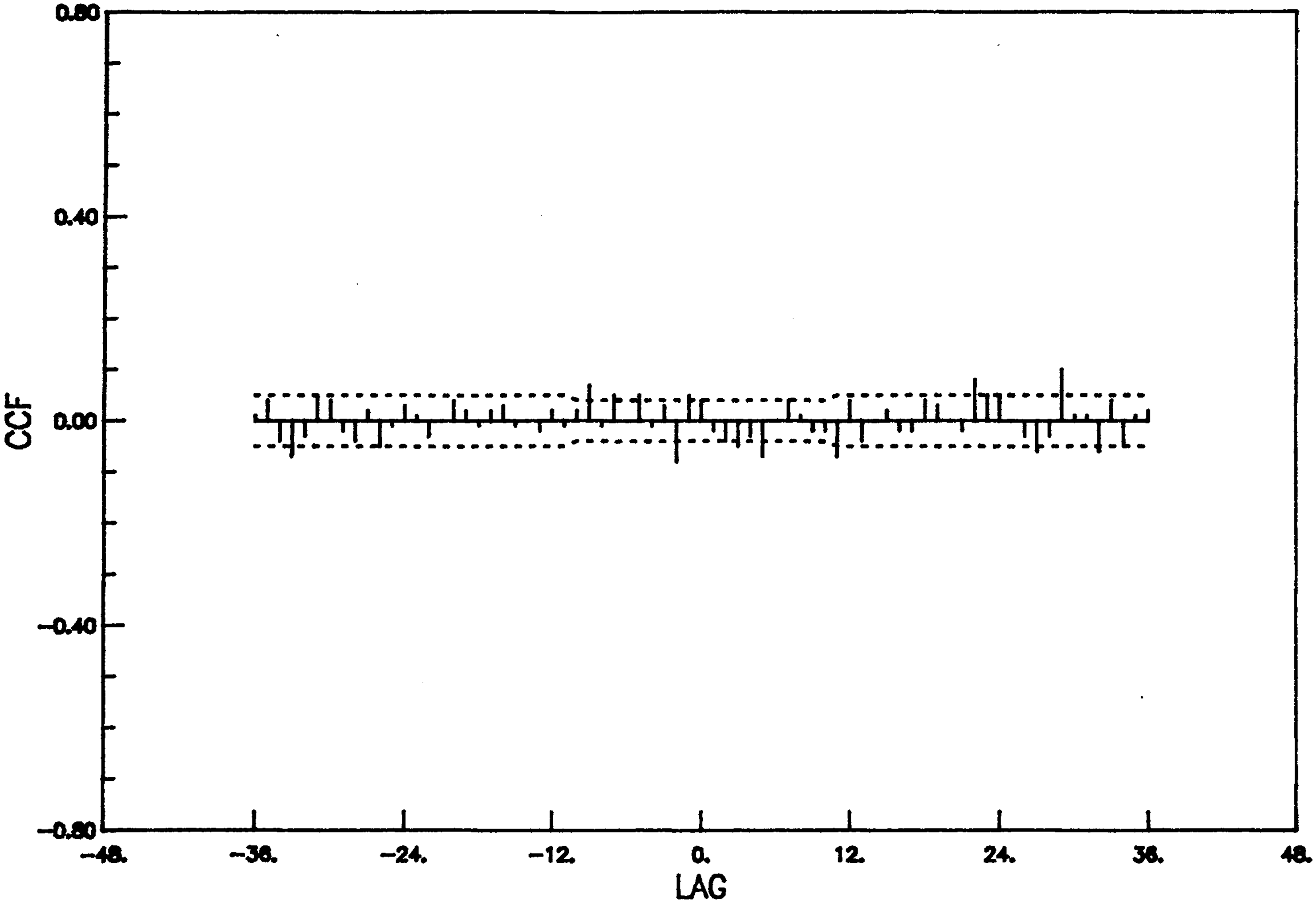
This model was subjected to different tests and satisfied them well. For example, the ACF of the residuals (Fig. 22) shows no evidence of statistical structure within the series and confirms the independence of the residuals. The Q2 test (Table 5) which studies the independence of the residuals considering them as a group also corroborates the whiteness of the residuals. The CCF between residuals of the transfer function and residuals of the covariant series, X(t), presented in Fig. 23, evidences no anomalies, signifying the residuals are uncorrelated and therefore independent. The Q3 test, which also analyzes the independence of the residuals of the transfer model and of the covariate series as a group, does not show anomalies. Finally, since this model has the smallest Akaike Information Criterion (AIC), this is the best model from the available selection.

Once a model has been developed, an immediate and logical application is to use it for forecasting. Thus, the last section of this chapter is entirely dedicated to the forecasting problem.



SMOKY HILL RIVER

Fig.22 ACF of the transfer function model's residuals



SMOKY HILL RIVER

Fig.23 CCF between transformed and deseasonalized RAIN series, and transfer function model's residuals

FORECASTING

The computation of the minimum square error forecast (Eq. 34) and the corresponding 50%, and 95% confidence limits (Eq. 35) are presented in the following section.

The transfer function developed in the previous sections is now used to forecast 12 values of the DIS series based on the information transmited by the RAIN series. Thus, Eq. 63 is first appropriately arranged for this computation as follows:

$$\begin{aligned} Y(t+j) = {} & 1.74*Y(t-1+j) - 0.75*Y(t-2+j) \\ & + 0.37*X(t+j) - 0.36*X(t-1+j) \\ & + a(t+j) - 1.42*a(t-1+j) \\ & + 0.31*a(t-2+j) + 0.14*a(t-3+j) \end{aligned} \tag{64}$$

In this equation, j is the number of units (months) of the forecast beyond the origin t. If t is fixed to an arbitrary starting point, say t=461, and j=1, Eq. 64 becomes

$$\begin{aligned} Y'(462) = {} & 1.74*Y(461) - 0.75*Y(460) \\ & + 0.27*X(462) - 0.26*X(461) \\ & + a(462) - 1.42*a(461) \\ & + 0.31*a(460) + 0.14*a(459) \end{aligned} \tag{65}$$

The starting point t has been selected to obtain 12 forecasted values which correspond to 12 known actual values. Therefore, in the forecasting equation (Eq. 65) the values with subscripts larger than 461 are assumed unknown and need to estimated.

For the input series X(t)

$X(t)=$ value of the RAIN series if, $t \leqslant 461$

$X(t)= 0$ if $t > 461$

This is a particular situation since the RAIN series is a white noise. In general, the values of X(t) for $t > 461$ are forecasted using ARMA models.

For the white noise series a(t)

$a(t)=$ values computed from Eq. 63 if, $t \leqslant 461$

$a(t)= 0$ if, $t > 461$

For the output series Y(t)

$Y'(j) =$ forecast for $j=1,2,....12$

$Y(t) =$ value of the DIS series if, $t \leqslant 461$

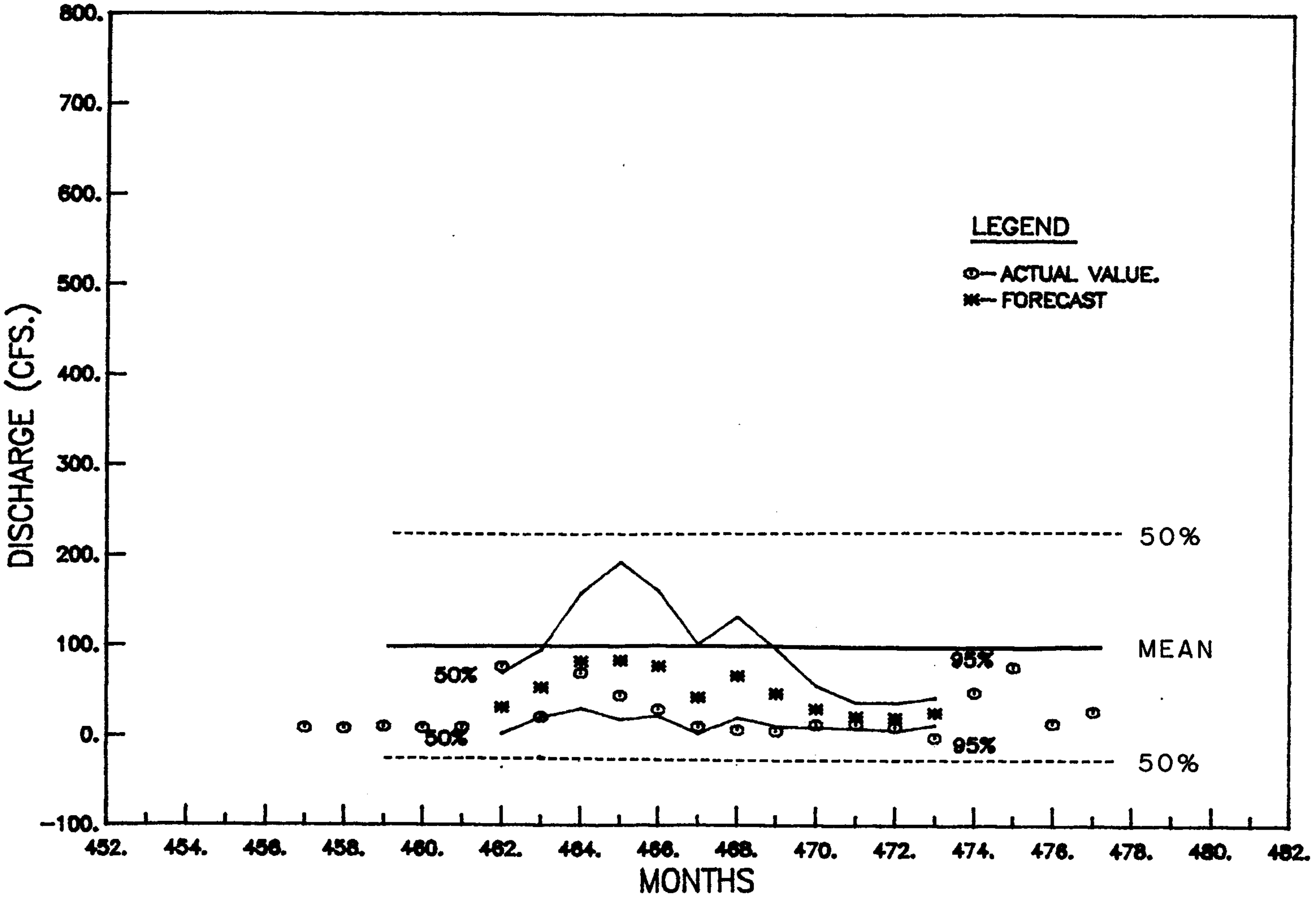
$Y(t) = Y'(t-461)$ if, $t > 461$

Table 6 lists these 12 forecasted values, their standard errors, and the actual observations. These values were converted back to their original form, which means the seasonality was added and the anti-logarithm was taken, before they were plotted in Fig. 24. This plot shows that the actual observations lie within +/- 0.67 standard deviations (50%) of the forecasted values, which is an indication of the quality of the forecasts. Furthermore, a band of +/- 0.67 standard deviations around the mean of the original DIS series is included, to compare with the one obtained from the forecasts. The former is indeed much smaller, showing the model is adequate and capable of reproducing the stochastic behavior of the DIS series based on the information transmitted by the covariate series.

FORECASTING

MONTH	FORECAST	STANDARD ERROR	ACTUAL VALUE
462	-0.29849	0.74085	0.28210
463	-0.16309	0.77731	-0.6823
464	-0.15477	0.78204	-0.2578
465	-0.14781	0.78654	-0.4259
466	-0.14189	0.79082	-0.5829
467	-0.13675	0.79488	-0.5515
468	-0.13222	0.79875	-0.8707
469	-0.12817	0.80242	-0.7649
470	-0.12447	0.80592	-0.5468
471	-0.12107	0.80925	-0.4232
472	-0.1179	0.81242	-0.4666
473	-0.11492	0.81544	-1.2465

TABLE #6



SMOKY HILL RIVER

Fig.24 Forecast of DIS series based on information transmitted by RAIN series

CONCLUSIONS

The stochastic analysis and modeling of the hydrological phenomena in the Smoky Hill River basin have permitted the formulation of the following conclusions and remarks

For studying complex water resources systems the use of Time Series methodology is relevant. It is a powerful technique for regional water resources planning. Furthermore, the analysis and modeling of the dynamics of the phenomena provide a better understanding of the complex processes and interactions of the system as well as regional projections which in turn can be used to establish guidelines for better planning and decision making.

Before proceeding, it is necessary to point out that the conclusions and information extracted from the records are no better than the quality and quantity of available data.

DISCHARGE SERIES. The Smoky Hill River is a gaining stream along the study area. The mean value of this net gaining is 57.70 CFS. This represents as much as 31% of the flow at the upstream station at Bunker Hill. The analysis of this series revealed a non-normal distribution and long-term as well as short-term seasonalities. The series has been transformed using the following equation

$$Y(t) = \ln(Z(t) + 85.608)$$

where Y(t) and Z(t) represent the transformed and untransformed discharge series, respectively. The long-term periodicity is similar to one present in the Wolfer sun spot numbers which follows a normal distribution with a mean equal to 11 years. However, the present series is not long enough to perform more thorough analyses. Nevertheless, ascending and descending trends are clearly visualized, e.g., the 1951, 1962, and 1973 water years are wet years, whereas the subsequent three or four years are dry years.

The short-term periodicity has a seasonality equal to 12 months and is responsible for the streamflow pattern. Low flows with small standard deviations occur during winter, and high flows with large standard deviations occur during summer. This annual cycle starts with a minimum equal to 23 CFS. in January and proceeds toward a maximum equal to 110 CFS. in June. July marks the beginning of the descending trend interrupted by a drastic decrease in August. In general, this annual cycle parallels the annual cycle of the rainfall series, suggesting the precipitation series as the driving force of the hydrological system except for during the months of June and August.

The precipitation series follows a descending trend during June, but the discharge series shows an ascending trend instead. This effect may be explained by the increase in stream recharge by the alluvial aquifer.

During August, the precipitation series does not evidence significant changes. However, the discharge series manifests a sharp decline, suggesting a loss of riverflow due to an increase in evapotranspiration during August, which is one of the hottest months of the year.

Considering precipitation as the sole source of recharge in the study area only 6% of this component becomes streamflow. This percentage is even further reduced to 4% during August, and most likely non of it is due to runoff. Evaporation is extremely large and the soil profile is quite dry during this month making very difficult any significant contribution from runoff . Based on this assumption, it is possible to establish that the minimum contribution of ground water flow is 4% of the total.

The ARMA model selected for the series is

$$Z(t)=0.90*Z(t-1) + a(t) - 0.44*a(t-1) - 0.21*a(t-2)$$

This equation reveals a first order autocorrelation of the discharge series, which maybe explained by the memory of the system related to the storage capacity of the underlying aquifer. The alluvial aquifer, over which the Smoky Hill River meanders, is constantly discharging to the stream, as has been previously stated.

Forecasting and simulation of riverflows could be readily pursued by employing the model developed. The main advantage of the model is its simplicity. Only three parameters are necessary for describing the stochastic behavior of the streamflow process.

RAINFALL SERIES. This series represents the volume of precipitation expressed as a mean monthly average over the study area (CFS). The analysis of this series revealed an asymmetric distribution and the presence of a short-term periodicity. The persistence of the values and the heteroscedasticity of the process make it difficult to identify long-term periodicities. The precipitation series has been normalized by the following equation

$$X(t) = \ln(Z(t) + 1148.84)$$

where X(t) and Z(t) represent the transformed and the untransformed precipitation series, respectively.

The annual cycle of the precipitation series has a minimum of 280 CFS. in January and a maximum of 1878 CFS. in May. Even though precipitation is the major source of recharge in the study area, 94% of the precipitation never reaches the stream.

The statistical analysis of the precipitation series did not reveal internal structure, suggesting that this series is a white noise process

$$X(t) = a(t)$$

which makes it very difficult to accurately forecast this phenomenon.

Transfer Function Noise Model. The relationship between precipitation and riverflow has been established by means of transfer functions. A linear dynamic model has been developed to link the input series precipitation to the output series discharge, as follows

$$Y(t) = 0.75*Y(t-1) + 0.37*X(t)$$
$$+ (1-0.98*B)^{-1}*(1-0.75*B)*(1-0.66*B-0.19*B^2)*a(t)$$

Inferences from this equation are broken up into three classes:

1. A first order autocorrelation of the discharge series. This autocorrelation represents the memory of the system, and states that the present value is a function of the discharge from the previous month. Physically, this is explained by the storage capacity of the alluvial aquifer.

2. A zero order cross correlation. It suggests there is no time lag response between the input series rainfall and the output series discharge. This means the time lag is less than one month and cannot be properly identified with the current information. The observations of precipitation and discharge are made on a monthly basis; therefore, the sampling interval of the observations is not small enough to detect such fluctuations.

3. A linear combination of white noise terms. The present value of discharge is a function of a linear combination of white noise terms, which accounts for other phenomena not directly incorporated into the analysis and for the stochastic behavior of the discharge series itself.

To demonstrate the statistical improvement as the complexity of the models is increased, the residual variance of various models is analyzed.

	RESIDUAL VARIANCE	AIC
Transformed and deseasonalized series	1	1.28
ARMA(1,2) MODEL	0.71	-169.82
Transfer Function Noise Model	0.52	-318.93

The total variance of the discharge series after being normalized and deseasonalized is 1.0. This is reduced by 29% to 0.71 by introducing an ARMA(1,2) process. A further reduction of 27% of the residual variance is due to the transfer function term. This final reduction represents 19% of the original variance. Thus, the transfer function component makes a statistical improvement to the final model.

Forecasting. In general, it is expected that transfer function models will greatly improve short-term forecasting performance by taking account of the additional information present in the input series. However, the input series precipitation is a white noise process. For this particular situation, there is no additional information obtained from the forecasted covariant series because the unconditional expected values of the precipitation series are zero, which means the forecasts of the precipitation series are zero. In spite of this drawback, forecasting performance of this model proved to be quite satisfactory. Actual values when compared to forecasted values lay inside of or near to the 50% confidence limits, showing that actual and forecast values are in clear agreement.

Two of the most important applications of the transfer function noise model developed for the Smoky Hill River basin are forecasting and simulation. The stochastic forecasting of a single datum of riverflow on a single step basis could be immediately applied in the actual scheduling of the real-time operation of the Kanopolis Reservoir. The stochastic simulation of sets of riverflows could provide several and various scenarios on which the efficacy of the decision rules of Kanopolis Reservoir can be tested. Moreover, these sets of riverflow could be come particularly helpful in performing a Reservoir Failure Analysis.

BIBLIOGRAPHY

Anderson,O.D., 1976, Time Series Analysis and Forecasting. The Box and Jenkins approach: Butterworths, London and Boston.

Baracos,P.C., Hipel,K.W., and McLeod,A.I., 1981, Modeling Hydrological Time Series from the Arctic: Water Resources Bulletin, 17(3), America Water Resources Association.

Bartlett,M.S., 1946, On The Theoretical Specification of Sampling Properties of Autocorrelated Time Series: J.Roy. Statisti. Soc.

Box,G.E.P. and Cox,D.R., 1964, An analysis of transformations: J. Roy. Statisti. Soc. Ser. B 26,pp. 211-252.

Box,G.E.P., and Jenkins,G.M., 1970, Time Series Analysis: Forecasting and Control, Holden-Day Inc., San Francisco, Calif.

Box,G.E.P., and Tio,G.C., 1977, Intervention Analysis with Applications to Economic and Environmental Problems: Journal of the America Statistical Association, 70(349),pp. 70-79.

D' Astous,F., and Hipel,K.W., 1979, Analyzing environmental time series: Journal of the Environmental Engineering Division, America Society of Civil Engineers, IOS(EE5),pp. 979-992.

Granger,C.W.J., 1969, Investigating causal relations by econometric models and cross-spectral methods: Econometrica, 37(3),pp. 424-438.

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

Hazen,A., 1914, Storage to be provided in impounding reservoirs for municipal water supply. Trans. Amer. Soc. Civil Engr., 77,pp. 1539-1669.

Hipel,K.W., McLeod,A.I., and Lennox,W.C., 1977 Advances in Box-Jenkins Modeling, 1. Model Construction: Water Resources Research, 13(3),pp. 567-575.

Hipel,K.W., McLeod,A.I., and Noakes,D.J., 1982 Fitting Dynamic Models to Hydrological Time Series, Time Series Methods in Hydrosciences: Elsevier Scientific Publishing Company, Amsterdam, 17,pp. 110-129.

Hurst,H.E., 1951, Long term storage capacity of reservoirs: Trans. Amer. Soc. Civil Engrs.,116, pp. 770-799.

Kottegoda, N.T., 1980, Stochastic Water Resources Technology: The Macmillan Press Ltd. pp. 1-172.

Marquardt,D.W., 1963, An Algorithm for Least Squares Estimation of Nonlinear Parameters: Journal of the Society of Industrial and App.lied Mathematics, 2,pp. 431-441.

McLeod,A.I., 1977, Improved Box-Jenkins Estimators: Biometrika 64(3), pp. 531-534.

McLeod,A.I., Hipel,K.W., and Lennox,W.C., 1977, Advances in Box-Jenkins modeling, 2, applications: Water Resources Research, 13(3),pp. 577-586.

Montgomery,D.C., and Weatherby,G., 1980, Modeling and Forecasting Time Series Using Transfer Function and Intervention Methods: AIIE Transactions, December.

Parzen,E., 1974, Some Recent Advances in Time Series Modeling: IEE Transactions on Automatic Control.

Salas,J.D., Delleur,J.W., Yevjevich,V., and Lane,W.L., 1980, Applied Modeling of Hydrologic Time Series: Water Resources Publications. Littleton, Colorado 80161,pp. 11-12.

Schoewe,W.H., 1949, The geography of Kansas, pt. 1, Political geography: Kansas Acad. Sci., Trans., V. 51, # 3,pp. 253-288.

Spiegel,M.R., 1961, Theory and problems of statistics: Schaum's outline series, McGraw-Hill Book.

APPENDIX

DIS

MONTHLY MEANS AND STANDARD DEVIATIONS

MONTH	MEAN	STD. DEVIATION
October	4.96180	0.59186
November	4.78984	0.36708
December	4.69727	0.24599
January	4.69074	0.27206
February	4.73590	0.24850
March	4.92824	0.57390
April	5.01179	0.51318
May	5.23255	0.70843
June	5.27230	0.93763
July	5.19648	0.78047
August	4.95112	0.70678
September	5.10985	0.66910

APPENDIX #1

DIS
AUTOCORRELATION AND PARTIAL AUTOCORRELATION

AUTOCORRELATION

1- 12	.52	.33	.30	.24	.24	.22	.20	.19	.14	.13	.13	.10
ST.ER	.04	.06	.06	.06	.06	.07	.07	.07	.07	.07	.07	.07
13-24	.05	.04	.05	.03	.01	.03	.04	.04	.05	.10	.14	.12
ST.ER	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07
25-36	.09	.07	.10	.06	.11	.08	.04	-.02	-.01	-.01	.00	-.04
ST.ER	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07	.07

PARTIAL AUTOCORRELATION

1-12	.52	.08	.14	.04	.08	.05	.04	.04	-.03	.02	.02	-.02
ST.ER	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
13-24	-.05	-.01	.01	-.02	-.01	.02	.03	.01	.03	.08	.08	.01
ST.ER	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
25-36	-.01	-.02	.04	-.05	.06	-.05	-.04	-.09	.00	-.02	.01	-.05
ST.ER	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04

APPENDIX #2

DIS RESIDUALS
AUTOCORRELATION AND PARTIAL AUTOCORRELATION

AUTOCORRELATION

1-12	0.0	0.0	.02	-.04	0.0	.01	.00	.04	-.03	.00	.03	.00
ST.ER	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
13-24	-.04	-.03	.00	-.03	-.04	-.03	-.01	-.02	-.02	.04	.09	.04
ST.ER	.04	.04	.04	.04	.04	.05	.05	.05	.05	.05	.05	.05
25-36	.03	-.02	.07	-.03	.08	.04	.02	-.06	-.01	-.01	.03	-.05
ST.ER	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05

PARTIAL AUTOCORRELATION

1-12	.00	.02	.02	-.04	.00	.01	.00	.04	-.03	.00	.03	.01
ST.ER	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
13-14	-.05	-.04	.00	-.03	-.05	-.03	-.01	-.02	-.02	.04	.09	.05
ST.ER	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
25-36	.03	-.02	.07	-.03	.09	.03	.02	-.07	-.01	-.02	.02	-.05
ST.ER	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04

APPENDIX #3

THIESSEN POLYGONAL METHOD

1940-1951

#	STATION	READING	AREA	WEIGHT	CODE
1	RUSSELL AP	0.2307	230.7	0.408	RU
2	ELLSWORTH	0.2076	207.6	0.367	ELR
3	CLAFLIN	0.1267	126.7	0.224	CL
			565	1	

Y(t)=0.8963*(230.7*RU+126.7*CL+207.6*ELR)

1952-1981

#	STATION	READING	AREA	WEIGHT	CODE
1	RUSSELL	0.182	182	0.322	RUA
2	ELLSWORTH	0.146	146	0.258	ELR
3	CLAFLIN	0.112	112	0.199	CL
4	WILSON	0.125	125	0.221	WI
			565	1	

Y=0.8963*(182*RUA+146*ELR+112*CL+125*WI)

APPENDIX #4

RAIN

MONTHLY MEANS AND STANDARD DEVIATIONS

MONTH	MEAN	STD. DEVIATION
October	7.60958	0.32816
November	7.38604	0.29303
December	7.31941	0.17574
January	7.26564	0.15933
February	7.34287	0.20062
March	7.57165	0.30543
April	7.72563	0.26478
May	8.01530	0.29431
June	7.98135	0.34781
July	7.89587	0.35200
August	7.87674	0.30562
September	7.79356	0.39091

APPENDIX #5

RAIN
AUTOCORRELATION AND PARTIAL AUTOCORRELATION

AUTOCORRELATION

1-12	.08	.03	-.01	.08	.03	.04	-.01	.02	-.01	.07	.00	-.04
ST.ER	.04	.04	.04	.04	.05	.05	.05	.05	.05	.05	.05	.05
13-24	.08	-.01	-.03	.00	-.04	-.06	.02	-.03	.00	-.02	.04	.04
ST.ER	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05
25-36	-.01	.06	.00	.00	.12	.09	.02	-.02	.00	.05	-.07	-.02
ST.ER	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05

PARTIAL AUTOCORRELATION

1-12	.08	.02	-.01	.08	.02	.03	-.02	.02	-.01	.06	-.01	-.04
ST.ER	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
13-24	.09	-.03	-.03	.01	-.05	-.06	.04	-.03	.00	.01	.03	.04
ST.ER	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04
25-36	-.01	.05	-.01	.00	.11	.07	.01	-.04	-.01	.02	-.09	-.02
ST.ER	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04

APPENDIX #6

RAIN-DIS
CROSS CORRELATION (HIPEL)

0	0.47											
1-12	.20	.04	.01	.04	.05	.00	.01	-.04	.03	.01	-.04	-.02
ST.ER	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.05	.05
13-24	-.02	-.01	-.03	.00	-.02	-.06	.00	.01	-.02	-.05	.02	.03
ST.ER	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05
25-36	.00	-.02	.03	-.03	.04	.08	.07	-.03	-.06	-.02	.01	-.02
ST.ER	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05

APPENDIX #7

RESIDUAL OF THE TRANSFER MODEL
AUTOCORRELATION

1-12	.00	.00	.06	-.05	.01	.03	-.02	.03	-.06	-.01	.08	-.01
ST.ER	.04	.04	.04	.04	.04	.04	.04	.04	.04	.05	.05	.05
13-24	-.05	-.05	-.02	-.05	-.06	-.05	-.04	-.05	-.02	.03	.06	.00
ST.ER	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05
25-36	.00	-.03	.08	-.02	.01	-.02	-.03	-.03	.00	.03	.05	-.02
ST.ER	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05

APPENDIX #8

INPUT AND TRANSFER FUNCTION RESIDUALS CROSS CORRELATION

0	0.04											
1-12	-.02	-.04	-.05	-.03	-.07	0.0	.04	.01	-.02	-.02	-.07	.04
ST.ER	.04	.04	.04	.04	.04	.04	.04	.04	.04	.04	.05	.05
13-24	-.04	.00	.02	-.02	-.02	.04	.03	.00	-.02	.08	.05	.05
ST.ER	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05
25-36	.00	-.03	-.06	-.03	.10	.01	.01	-.06	.04	-.05	.01	.02
ST.ER	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05

APPENDIX #9