STOCHASTIC MODEL OF DAILY RIVER FLOW SEQUENCES

By

Rafael G. Quimpo

February 1967

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

18

STOCHASTIC MODEL OF DAILY RIVER FLOW SEQUENCES

By

Rafael G. Quimpo

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO

February 1967 No. 18

ACKNOWLEDGMENTS

Research on this paper was initiated and carried out while the author was studying towards his Doctor of Philosophy degree at Colorado State University under the guidance of his major professor, Dr. V. Yevjevich, to whom he would like to express his sincere appreciation. The writer also wishes to acknowledge the helpful advice of Dr. M. M. Siddiqui, Professor of Statistics, during the course of this investigation. In addition, thanks are due to Mr. Raymond Ing-Song Jeng who carefully read the manuscript and did the layout.

The financial support of the U. S. National Science Foundation in the research leading to this hydrology paper is greatfully acknowledged. Last of all, the National Center for Atmospheric Research kindly permitted free use of a computer while this research was in progress.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ABSTRACT

A search for a structural model for the time series of daily river flows is undertaken by the author. First, records of daily runoff from 17 river basins chosen on the postulated absence of trends induced by manmade improvements are analyzed. As a result, the model envisaged is a superposition of a cyclic deterministic process and a stochastic component. In the analysis of records, spectral methods are used to detect cycles which are then removed by subtracting from the original series a periodic function obtained by harmonic analysis. To alleviate the effect of a changing variance during the course of the year, the series of standard deviations is simularly fitted with a harmonic function which is used to standardize the series. After standardization, all the residual series are found to satisfy the second order autoregressive representation:

$$Z_t + a_1 Z_{t-1} + a_2 Z_{t-2} = \eta_t$$

where a_1 and a_2 are the autoregressive coefficients and η_t is an independently distributed random variable. The adequacy of fit is judged on the agreement between the theoretical and explained variances.

STOCHASTIC MODEL OF DAILY RIVER FLOW SEQUENCES

by Rafael G. Quimpo*

CHAPTER I

INTRODUCTION

An important facet in the planning of a water resource project is the prediction of characteristics of future water supply, the most common of which are rainfall and runoff. For prediction, use of the latter seems to be more practical since it is less susceptible to intermediate processes which could radically change final quantitative estimates of available water and since it, at the same time, admits to facility in measurement. It is, thus, natural that records of river flows should be the subject of an intensive investigation of structural characteristics most suitable for mathematical modeling and prediction purposes.

With the ever-increasing demand for water, estimates of water needs, expressed in lumped quantities averaged over a year or more, are no longer sufficient. Specification of river runoff in terms of mean annual flow or flow duration is gradually giving way to description of flow in terms of time sequence and distribution, the knowledge of which allows for better regulation and control of water, e. g., by storage reservoirs.

Runoff from a basin is the combined effect of variables which may be deterministic or stochastic in nature. The interaction of these variables has thus far defied a complete mathematical analysis. Therefore, the engineer who has to make a projected estimate of flow must rely on statistics. At this time, prediction procedures properly belong to this discipline.

Objective - The purpose of this study is to investigate the structure of the time series of daily river flows, to detect and isolate the deterministic component from the stochastic of the time series, and to reconstruct the underlying process in terms of a mathematical model which will adequately describe the structure of the underlying mechanisms which generate the process.

Approach - River runoff is a continuous process. To be able to closely examine the properties of the continuous time process, one must work with the shortest time interval possible. Although it has been measured by continuous recorders in many sites, problems in information retrieval has limited published data to equi-spaced records of average values. The analysis of daily river flows is offered as an attempt to improve the accuracy of predictions.

Definition - Daily river flow as defined in this study is the average daily runoff at a section of a river, the averaging being done either from a continuous record of an automatic recorder or from river stage measurements taken at representative time intervals to make interpolation and averaging consistent.

* Former graduate student of Colorado State University, and at present Assistant Professor, University of Pittsburgh.

CHAPTER II

MATHEMATICAL TECHNIQUES*

The statistical analysis of a time series is best approached as a study of random functions. In that case, a random function X(t) is defined as a function whose values are random variables. If X(t) is taken as the result of an experiment or the record of a process, then X(t) is called a realization or a sample function. If t is allowed to represent time, one gets a probabilistic representation of an observational time series.

A time series is called stationary if it is temporarily homogeneous, i. e., if its statistical properties do not vary with time. Otherwise, the time series is called non-stationary or evolutive. Actually, complete temporal homogeneity is difficult to show except in very special cases. Because of this, most statistical investigations have been limited to a less rigid definition of stationarity, based on the first two moments of the distribution function:

$$\mu(t) = E\,X(t)$$

$$C(t, s) = E[X(t) - \mu(t)]\;[X(s) - \mu(t)] \qquad (2.1)$$

where $E(\cdot)$ denotes the mathematical expectation.

Thus, X(t) is said to be weakly stationary or stationary in the wide sense if its mean is equal to a constant and its autocovariance function C(t, s) is a function only of the difference (t - s), i. e.,

$$EX(t) = m$$

$$E[X(t) - m]\;[X(s) - m] = C(t - s). \qquad (2.2)$$

This study will be confined to time series which may be assumed stationary, as defined above, or to those that can be reduced to such stationarity.

It should be noted that the use of probability theory, in the above definition of terms, implies an ensemble of time series. Although in practice usually only a single realization is available, it is still possible to calculate the characteristics of a stationary random function because of the ergodic theorem which generally applies when the conditions of stationarity are met. According to the ergodic theorem, the mathematical expectation of both X(t) and X(t) X(s) obtained by averaging over an ensemble of time series can be replaced by the time average of the same quantities over a realization.

2.1 The General Model. Consider a realization of a random process taken at equal intervals of time and denote it by X_t, the discrete equivalent of X(t). A general heuristic model used to describe such a sequence is given by

$$X_t = R_t + P_t + \epsilon_t \qquad (2.3)$$

where

R_t is a trend component,

P_t is a periodic or cyclic component, and

ϵ_t is a nondeterministic or stochastic component.

Let it be assumed that there is no trend. Such an assumption does not necessarily limit the applicability of later results since, when a trend does exist, there are ways of isolating it and subtracting it from the time series. However, the procedures for doing so are outside the scope of the present study which must, therefore, proceed on the above assumption. To continue, eq. 2.3 may then be written as the sum of a periodic component and a stochastic component

$$X_t = P_t + \epsilon_t. \qquad (2.4)$$

The main problem in applying scheme 2.4 is the separation of the periodic component from the nondeterministic part.

2.2 The Periodic Component. The classical approach to the detection of periods is by use of Schuster's periodogram obtained through a harmonic analysis. In this approach, if X_t is assumed to have a period p, then the function X_t must satisfy the relation

$$X_t - X_{t-p} = 0 \qquad (2.5)$$

for any t. The solution of eq. 2.5 may be written as

$$X_t = \overline{X} + \Sigma\, I_k \cos\left(\frac{2\pi kt}{L} + \phi_k\right)$$

$$= \overline{X} + \Sigma\left(A_k \cos\frac{2\pi kt}{L} + B_k \sin\frac{2\pi kt}{L}\right) \qquad (2.6)$$

where

k runs from 1 to $\frac{L}{2}$ or $\frac{L-1}{2}$

$I_k^2 = A_k^2 + B_k^2 \qquad A_k, B_k$ and $I_k > 0$

$\phi_k = \arctan\left(\frac{-B_k}{A_k}\right),\ \left(-\frac{\pi}{2} \le \phi_k \le \frac{\pi}{2}\right).$

On the basis of eq. 2.6, an approach function may be considered, composed of superposed harmonics each having an amplitude I_k^2, a phase ϕ_k, and an angular frequency λ_k given by

* The reader who is well familiar with the autocorrelation and spectral analysis techniques in the investigation of time series may delete the reading of this chapter.

$$\lambda_k = \frac{2\pi k}{L} \quad (2.7)$$

Thus, one obtains a general function, a composed harmonic of the form

$$X(t) = \overline{X} + \sum_{k=1}^{s} I_k \cos(\lambda_k t + \phi_k)$$

$$= \overline{X} + \sum_{k=1}^{s} (A_k \cos\lambda_k t + B_k \sin\lambda_k t) . \quad (2.8)$$

The basic difference between eq. 2.6 and eq. 2.8 is that in eq. 2.6 the periods $p_i = \frac{2\pi}{\lambda_k}$ are true fractions of p, whereas in eq. 2.8 they are not restricted. Thus, limiting λ_k between 0 and π in eq. 2.8 would not affect its generality. The essential problem in applying scheme 2.8 is evaluating the frequency numbers λ_k which would give the best possible fit to empirical data. For a sample of size n, the formulas required are:

$$\left.\begin{aligned} A_k &= \frac{2}{n} \sum_{t=1}^{n} X_t \cos\lambda_k t, \\ B_k &= \frac{2}{n} \sum_{t=1}^{n} X_t \sin\lambda_k t, \\ I_k^2 &= A_k^2 + B_k^2 . \end{aligned}\right\} \quad (2.9)$$

For $\lambda_k = \pi$, $B_k = 0$ and $A_k = \frac{1}{2} \sum_{t=1}^{n} (-1)^t X_t$.

A widely used description of this scheme is the empirical periodogram in which I_k^2 is plotted as ordinate against λ_k. A variant of this plot is the integrated periodogram wherein I_k^2 is replaced by S_k defined by

$$s_k = \frac{\sum_{\lambda_k \le \lambda} I_k^2(\lambda)}{\sum_{\lambda_k = 1}^{s} I_k^2(\lambda)} . \quad (2.10)$$

The periodogram is analogous to the spectrum of light in optics in the sense that, as the lines in the spectrum of light correspond to the power of the respective wave component, the periodogram ordinate corresponds to the contribution to the variance by the harmonic of a given frequency.

2.3 The Stochastic Component. Assuming that the periodic component P_t has been detected and subtracted from eq. 2.4, the residual series,

$$\epsilon_t = X_t - P_t , \quad (2.11)$$

will now be considered. According to Doob's terminology [8], ϵ_t belongs to a class of nondeterministic processes which include autoregressive, moving average, and other schemes of linear regressions. In these schemes of linear regression, the interdependence between successive terms is assumed to be probabilistic rather than functional.

2.3.1 The Schemes of Moving Averages and Linear Autoregressions. If it is assumed that each term in the series is the combined effect of a primary sequence of identically distributed variables η_t, then ϵ_t can be represented by

$$\epsilon_t = \sum_{i=0}^{\infty} b_i \eta_{t-i} \quad (2.12)$$

It is possible that the summation extends only over a finite number of terms. This representation is called, in the literature, the scheme of moving averages.

Another type of linear regression which uses the same set of primary variables η_t, is written in the form,

$$\eta_t = \sum_{i=0}^{\infty} a_i \epsilon_{t-i} . \quad (2.13)$$

This representation is called the scheme of linear autoregression.

While the infinite representations are theoretically sound, the fact is that in all processes the effects of past values on the present decrease with time distance and a finite summation very often suffices. Further, if the degree of required precision is fixed, only a finite number of terms will be required and hence the alternative representations are:

$$\epsilon_t = \sum_{i=1}^{n} b_i \eta_{t-i} \quad (2.14)$$

$$\sum_{i=0}^{m} a_i \epsilon_{t-i} = \eta_t . \quad (2.15)$$

Wold [21] shows that a necessary condition for the existence of an autoregressive scheme (2.15) is the inclusion of the roots of the characteristic equation,

$$X^m + a_1 X^{m-1} + \ldots + a_{m-1} X + a_m = 0 , \quad (2.16)$$

all inside the unit circle of the complex plane.

Wold shows that the coefficients (a) of eq. 2.15 are related to the coefficients (b) of eq. 2.14 and that each set defines the other.

2.3.2 Estimation of Parameters. Consider a finite-ordered autoregressive representation of the form 2.15:

$$X_t + a_1 X_{t-1} + \ldots + a_h X_{t-h} = \eta_t \quad (2.17)$$

and the moving average scheme:

$$X_t = \eta_t + b_1 \eta_{t-1} + b_2 \eta_{t-2} + \cdots . \quad (2.18)$$

Eq. 2.18 may be rewritten for any k as

$$X_{t+k} = \eta_{t+k} + b_1\eta_{t+k-1} + b_2\eta_{t+k-2} + \cdots \qquad (2.19)$$

Multiplying eq. 2.17 by eq. 2.19 and taking expectations, one obtains

$$r_k + a_1 r_{k-1} + a_2 r_{k-2} + \ldots + a_h r_{k-h} \qquad (2.20)$$

since the η's are uncorrelated. Repeating this procedure for h consecutive values of k ($k = 1, 2, \ldots, h$) the system of linear equations is obtained:

$$\begin{aligned} r_1 + a_1 + a_2 r_1 + \ldots + a_h r_{h-1} &= 0 \\ r_2 + a_1 r_1 + a_2 r_2 + \ldots + a_h r_{h-2} &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ r_h + a_1 r_{h-1} + a_2 r_{h-2} + \ldots + a_h &= 0 \end{aligned} \qquad (2.21)$$

Since the system 2.21 is linear with respect to the a's, it may be solved for $[a] = [a_1, a_2, \ldots, a_h]$ and thus the coefficients of the autoregressive scheme are obtained [21]. Further, after solving for the set [a], the existence of a stationary scheme 2.17 may be verified by substituting the a's in eq. 2.16 and checking whether the roots of the characteristic equation all lie in the unit circle of the complex plane.

There are other methods of estimating the parameters [a]. Yule [24] also used the equivalent procedure of minimizing residuals, a very convenient method for lower-ordered schemes and one which will be used in a later section.

The use of the system 2.21 employing empirical values of the autocorrelation coefficients yields a set of a's which minimizes the variance of η_t in eq. 2.17. Further, the first h autocorrelation coefficient of the derived process will coincide with the empirical values. However, Wold warns that the hypothetical correlogram will not in its whole range coincide with the empirical correlogram.

Consider the first order autoregressive scheme (taking n = 1 in eq. 2.14). With a slight change in notation, it may be rewritten as:

$$X_t = \rho X_{t-1} + \epsilon_t \;. \qquad (2.22)$$

Thus, the least square estimator of ρ may be shown to be equal to r_1.

For the second order autoregressive scheme (taking n = 2 in eq. 2.14), consider:

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} = \epsilon_t \;. \qquad (2.23)$$

The use of system 2.21 or the method of least squares yields estimates for a_1 and a_2:

$$a_1 = \frac{r_1(1 - r_2)}{1 - r_1^2}$$

and $\qquad\qquad (2.24)$

$$a_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$$

2.4 <u>Autocorrelation Analysis and the Correlogram.</u> In detecting patterns of movement, a logical question is whether or not successive values of a time series are interdependent. A measure of this dependence is given by the autocorrelation coefficient--the equivalent of the correlation coefficient between two variables. For a discrete time series, it is defined as:

$$\rho_k = \frac{E X_t X_{t+k} - E X_t E X_{t+k}}{E X_t^2 - (E X_t)^2} \;. \qquad (2.25)$$

If ρ_k is plotted against k and if the plotted points are connected by straight lines, one obtains a correlogram. It is obvious that if the terms of the series are uncorrelated, the correlogram will have a value of 1 at k = 0 and an expected value of zero at all other points. Kendall [16] further shows that if the series has a single sinusoidal component, the correlogram, given a large sample, will have the same period as the sinusoidal component and the correlogram will be an undamped cosine function. If the series has a composed harmonic component, however, while the correlogram will still show sinusoidal characteristics due to amplification and cancellation of amplitudes, the correlogram may be so distorted as to prevent visual detection of cycles present.

In the case of the scheme of moving averages,

$$\begin{aligned} \rho_k &= \frac{E(\epsilon_j \epsilon_{j+k})}{E\epsilon_j^2} = \frac{(b_1 b_{k+1} + \cdots + b_{n-k} b_n)\ \mathrm{var}\ \epsilon}{\mathrm{var}\ \epsilon \sum_{i=1}^{n} b_i^2} \\ &= \frac{\sum_{i=1}^{n-k} b_i b_{i+k}}{\sum_{i=1}^{n} b_i^2} \qquad \text{for } n > k \\ &= 0 \qquad\qquad \text{for } n < k \end{aligned} \qquad (2.26)$$

In the case where $b_i = b_j$ ($i = j = 1, 2, \ldots,$), eq. 2.26 becomes

$$\begin{aligned} \rho_k &= \frac{(n-k)\, b_i^2}{n b_i^2} = \frac{n-k}{n} = 1 - \frac{k}{n} \qquad n > k \\ &= 0 \qquad\qquad n < k \end{aligned} \qquad (2.27)$$

so that the correlogram consists of a broken line composed of the segment joining (0, 1) and (k, 0) and the x-axis from (k, 0) onwards.

For autoregressive schemes, the correlograms may take a number of characteristic shapes depending on the order of the scheme. The correlogram for the first order autoregressive model, also known as the first order Markov model, is

$$\rho_k = \rho^k \tag{2.28}$$

and the correlogram for the second order autoregressive, or second order Markov model, is

$$\rho_k = \frac{a_2^{k/2} \sin(k\theta + \psi)}{\sin\psi} \tag{2.29}$$

where

$$\theta = \arccos \frac{-a_1}{2\sqrt{a_2}} \text{ , and}$$

$$\tan\psi = \frac{1 - a_2}{1 + a_2} \tan\theta$$

It can be seen that the correlogram is a harmonic function with frequency θ damped by a factor $(a_2)^{k/2}$ [16].

Theoretically, the characteristic shapes of the correlograms reviewed above furnish the discriminating tools in the determination of the scheme which is applicable in a particular problem. Thus, an undamped sinusoidal correlogram would signify a cyclic series, while a correlogram which goes to zero at the kth lag would signify a moving average scheme. Further, a damped sinusoidal correlogram would indicate an autoregressive representation.

The cases encountered in practice, however, present some difficulties. Sampling fluctuations, especially for relatively short series, often obscure the asymptotic behavior of a first order autoregressive scheme on the correlogram, which may vanish at a lag k. The complication caused by the superposition of two or more harmonics has been previously cited as another source of confusion. This will be dealt with in more detail in a later section. A more perplexing situation arises when the series under study is generated by a linear combination of the above mentioned schemes. This is almost always the case encountered in practice.

2.5 Aspects of Spectral Theory. The transition from autocorrelation analysis to spectral methods may be followed in Wold's work [21]. On the basis of a theorem due to Khintchine for the continuous case, Wold shows that, given a sequence ρ_k ($k = 0, \pm 1, \pm 2, \ldots$), a necessary and sufficient condition for the existence of a discrete stationary process with ρ_k for correlation coefficients is that ρ_k values are the Fourier constants of a non-decreasing function $W(x)$ such that $W(0) = 0$ and $W(\pi) = \pi$, with

$$\rho_k = \frac{1}{\pi}\int_0^{\pi} \cos kx \, dW(x) \text{ .} \tag{2.30}$$

Reverting to the continuous case, Cramer [7] has shown that any stationary process $X(t)$ has the spectral representation

$$X(t) = \int_{-\pi}^{\pi} e^{itf} \, dz(f) \tag{2.31}$$

where $z(f)$ is an orthogonal set function with

$$E\left|dz(f)\right|^2 = dG(f) \text{ .}$$

Assume an actual process, roughly stationary, with $EX(t) = 0$ and $EX(t)\,X(t+k) = C(k)$. It might also be noted that the covariance function $C(k)$ is related to the autocorrelation coefficient by the expression

$$\rho_k = \frac{C(k)}{C(o)} \text{ .} \tag{2.32}$$

Using Cramer's representation, it can be shown that

$$C(k) = \int_{-\infty}^{\infty} e^{2\pi ifk} \, dG(f) = \int_{-\infty}^{\infty} e^{2\pi ifk} \, P(f) \, df \tag{2.33}$$

with the eq. 2.33 being valid if $G(f)$ is differentiable. Here, $G(f)$ is the spectral distribution function, and $P(f)$, the spectral density function of the process.

It may further be shown that eq. 2.33 may be inverted to give the spectral density function as the Fourier transform of the covariance function:

$$P(f) = \int_{-\infty}^{\infty} e^{-2\pi ifk} \, C(k) \, dk \tag{2.34}$$

Since $X(t)$ is real, $C(k)$ is symmetric about $k = 0$ so that $C(k)$ and $P(f)$ may be represented simply as cosine transforms:

$$\left.\begin{aligned} C(k) &= \int_{-\infty}^{\infty} P(f) \cos 2\pi fk\,df = 2\int_0^{\infty} P(f) \cos 2\pi fk\,df \\ P(f) &= \int_{-\infty}^{\infty} C(k) \cos 2\pi fk\,dk = 2\int_0^{\infty} C(k) \cos 2\pi fk\,dk \text{ .} \end{aligned}\right\} \tag{2.35}$$

The equivalence of Wold's representation 2.30 and the second half of eq. 2.35 is at once apparent.

According to Cramer's representation, the process $X(t)$ may be considered an integral over the frequency interval $(-\pi, \pi)$. Thus, the function,

$$dG(f) = G(f_2) - G(f_1)$$

may be interpreted as the part of the total variance attributable to the frequency band $(f_2 - f_1)$. The integral over the whole range gives the total variance and hence the equally appropriate name--variance spectrum. In a stretched context, the spectral distribution function is essentially an integrated periodogram of eq. 2.10.

In retrospect, it may be recalled that Wold's theorem and Cramer's spectral representation are both premised on the stationarity of $X(t)$. When $X(t)$ is not stationary and if the covariance $C(k)$ is used to estimate the spectrum, then, as Granger and Hatanaka [9] show, the result is not the true spectrum but an average spectrum. Although this presents certain theoretical difficulties, it should not cause too much concern since, in practice, one will be dealing with discrete values of X_t which are, to begin with, averages over the sampling time interval. Thus, one can only work with, at most, the precision of the averaging procedure in the data assembly. An average spectral estimate should suffice for practical purposes.

In the expression for $dG(f)$, if the band $(f_2 - f_1)$ is narrow enough or, to use the communication engineer's terminology, with a fine enough resolution, the spectral density may be used to detect cycles in the time series in exactly the same way that significant ordinates in the periodogram are used. In the analysis of discrete data, one is never able to obtain a continuous density function. Instead, he obtains the average power smudged over a frequency interval. Then, the identification of powers attributable to adjacent frequencies is attained by resolving the power into narrow bands of frequencies. The resolution is a measure of the concentration of a spectral estimate expressed in units of frequency. Following Tukey, this study accepts it as equal to the reciprocal of the maximum number of lags taken in computing autocovariances. Thus, increasing the lag improves the resolving power. There is, however, a limit to the fineness of resolution practicable because of the effect on the variability of estimates. Grenander and Rosenblatt [10] have shown that, given a length of data n, the variance of Tukey's estimate is proportional to the maximum number of lags m taken in computing autocovariances. Although an increase in variance requires a compromise between resolution and variability.

The analysis of equi-spaced records immediately imposes an upper bound on the spectrum. Since at least two points are necessary in order to fit one frequency, the fastest sine wave that can be detected, given a sampling interval Δt is $\frac{1}{2\Delta t}$ cycles per unit time. Since 2π radians equal one cycle, all frequencies above a "folding frequency," $f_n = \frac{2\pi}{2\Delta t} = \frac{\pi}{\Delta t}$, will not be detected. The wave that oscillates three times during the sampling interval cannot be distinguished from the one that oscillates only once from the one which oscillates twice. Thus, the density which is measured at f is actually:

$$\hat{P}(f) = \sum_{k=o}^{\infty} P(f \pm 2kf_n) \quad (k = 1, 2, \dots n) . \qquad (2.36)$$

Therefore, before any analysis is made, one must be sure that $P(f)$ is negligible for $f > f_n$. The choice of the sampling interval is decided on the basis of how much power remains above f_n and the largest frequency that is of interest.

2.5.1 Estimation of the Spectra. For the discrete case, an expression equivalent to the second part of eq. 2.35 yields the spectral density,

$$P(f) = \frac{1}{2\pi}[C_o + 2 \sum_{k=1}^{\infty} C_k \cos 2\pi fk] \quad (-\tfrac{1}{2} \le f \le \tfrac{1}{2}), \quad (2.37)$$

if the power is confined in a band $(-\pi, \pi)$.

Given a finite sample of observations, $x_1, \dots, x_n$, one can at most obtain only an estimate of spectral density since only n - 1 autocovariances can be calculated. Thus, one finds that:

$$\hat{P}(f) = \frac{1}{2\pi}[C_o + 2 \sum_{k=1}^{n-1} C_k \cos 2\pi fk] . \qquad (2.38)$$

In practice, even a lesser number of lags is sometimes used for reasons which will be discussed later. As regards the estimate in eq. 2.38, Jenkin [12] has shown that it is related to the ordinates of Schuster's periodogram defined by eq. 2.6. However, it can be shown that Schuster's periodogram ordinates do not give a consistent estimate of $P(f)$ [11]. This has led to estimates of the form arrived at by means of the following equation,

$$\hat{P}(f) = \frac{1}{2\pi}[C_o D_o(f) + 2 \sum_{k=1}^{n-1} C_k D_k(f) \cos 2\pi fk], \quad (2.39)$$

This is equivalent to the application of a filter or kernel function D_k on the covariance function C_k. Blackman and Tukey [6] have explained that the multiplication of C_k by a suitable even function D_k makes the transform of the product a respectable estimate of the smoothed values of the spectral density. Various forms of the kernel function have been suggested. This study used the function,

$$D_k(f) = \tfrac{1}{2}\left(1 + \frac{\cos \pi k}{m}\right) \qquad |k| < m$$

$$= 0 \qquad |k| > m \qquad (2.40)$$

whose transform takes the form,

$$Q_k = \tfrac{1}{2} Q(f) + \tfrac{1}{4}\left[Q\left(f + \frac{1}{2T_m}\right) + Q\left(f - \frac{1}{2T_m}\right)\right], \qquad (2.41)$$

where m is the maximum number of lags used in computing autocovariances and where m is also equal, for unit time interval, to the maximum time lag T_m used in estimating the spectrum.

In the actual analysis of discrete time data, there are two ways of arriving at estimates $\hat{P}(f)$. One is to multiply the covariances by the chosen kernel function before performing the Fourier transformation; another is to make the Fourier cosine transformation first and then form linear combinations of the results according to eq. 2.41. These two approaches are both possible because of the equivalence of multiplication and convolution under Fourier transformation [20].

To summarize, the estimation of the spectrum of a process involves:

(1) Choice of the maximum number of lags m necessary in calculating autocovariances. This decision is governed by the desired resolution (which for unit time interval, is equal to $\frac{1}{m}$), but only after consideration has been given to the adverse effect of an increased m on variability.

(2) Computation of m + 1 autocovariances, $C_o, C_1, \ldots, C_m$.

(3) Application of a finite cosine series transform to the sequences of autocovariances. This takes the form,

$$V_k = \frac{1}{m}\left[C_o + 2\sum_{j=1}^{m-1} C_j \cos\frac{kj\pi}{m} + C_m \cos k\pi\right] \qquad (2.42)$$

where the values of V_o and V_m are taken as half of their computed values.

(4) Formation of linear combinations of V_k using the coefficients in eq. 2.41 as weights.

2.6.1 Significance Test for Spectral Estimates. Jenkins [12] has demonstrated that the distribution of P(f) may be approximated by a Chi-square distribution with an equivalent number of degrees of freedom which depend on the kernel function used. In the case of the kernel in eq. 2.42, the equivalent number of degrees of freedom ν is approximately $\frac{2n}{m}$. In place of n, Tukey [20] suggested the substitution of an effective length of record $T_n = n - \frac{m}{3}$ so that

$$\nu = \frac{2(n - \frac{m}{3})}{m} = \frac{2n}{m} - \frac{2}{3} .$$

This then provides the basis for establishing confidence limits and for testing for significance. Tukey [20] has also tabulated values for the confidence limits for the ratio $\frac{P(f)}{\hat{P}(f)}$.

2.6.2 Test for Autoregressive Schemes. The statistical test used for testing the adequacy of finite ordered autoregressive schemes is derived from Quenouille [17]. According to this test, the kth ordered autoregressive representation may be written thus:

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} + \ldots + a_k X_{t-k} = \epsilon_t . \qquad (2.43)$$

Quenouille defines a function $\phi(x)$, two sets of constants [A] and [P], and a test parameter R_j by the equations:

$$\phi(x) = \sum_{j=o}^{k} a_j x^j ,$$

$$\phi^2(x) = \sum_{j=-\infty}^{\infty} A_j x^j = \sum_{j=o}^{2k} A_j x^j ,$$

$$P_j = \sum_{i=-\infty}^{\infty} \rho_1 \rho_{j-i} ,$$

$$R_j = \sum_{i=-\infty}^{\infty} A_i \rho_{j-i} = \sum_{i=o}^{2k} A_i \rho_{j-i} ,$$

where ρ's are the autocorrelation coefficients. He has shown that, if X_t satisfies eq. 4. 23, then $R_{k+1}, \ldots, R_{k+f}$ follow asymptotically normal distribution with mean zero, and variance

$$\frac{1}{n}\sum_{j=o}^{2k} A_j P_j .$$

This provides a test for the hypothesis that the equation,

$$\chi^2_f = \frac{n}{\sum_{j=o}^{2k} A_j P_j}\left(R^2_{k+1} + R^2_{k+2} + \ldots + R^2_{k+f}\right) ,$$

has a Chi-square distribution with f degrees of freedom. The empirical autocorrelation coefficients may be used for the ρ's.

For the first order Markov model, direct substitution into the above equation yields

$$\sum_{j=k+1}^{k+f} \frac{nR_j^2}{1-\rho^2}$$

as the test parameter with f degrees of freedom where

$$R_j = r_j - 2r_1 r_{j-1} + r_1^2 r_{j-2} .$$

Similarly, for the second order Markov model, the Chi-square distributed parameter is

$$\sum_{j=k+1}^{k+f} \left\{\frac{(1+a_2)}{(1-a_2)\left[(1+a_2)^2 - a_1^2\right]}\right\}^2 nR_j^2$$

where

$$R_j = r_j + 2a_1 r_{j-1} + (a_1^2 + 2a_2) r_{j-2} + 2a_1 a_2 r_{j-3} + a_2 r_{j-4} .$$

2.6.3 Test for Independence of Residual Series. In the use of autoregressive schemes, a measure of the adequacy of the representation adopted is the independence of the residual series obtained after subtracting the autoregressive scheme from the non-deterministic component ϵ_t. Anderson [1] has given a two-tailed test for $\rho_L = 0$ at a given significance level α. His test was derived for a circular time series but may also be applied to an open series--if due consideration is given to its limitations. For the present case, the value of ρ_1 is of interest.

The confidence limits at 95% level of significance are

$$S(\alpha) = \frac{-1 \pm 1.96\sqrt{N - 3}}{N - 2}$$

where N is the number of observed values.

CHAPTER III

RESEARCH DATA ASSEMBLY

3.1 Source, Reliability, and Accuracy of Data. Runoff records of daily flows for most of the rivers and their tributaries in the United States have been compiled by the U. S. Geological Survey. At regular intervals, these records have been published in the Water Supply papers. Initially, only figures for the more important rivers were compiled. Then, the list was gradually expanded to include data from all of the present network of several thousand gaging sites.

During the early stages of compilation, the mean daily flows were computed from daily mean gage heights as obtained from staff gages. This immediately made the early records subject to the frequency of observations taken during a day. The advent of continuous water stage recorders which gradually replaced the staff gages alleviated this situation somewhat. Conversion from the daily mean gage heights to flows was made by the use of stage-discharge rating curves.

Where the stage-discharge relation is subject to change due to frequent or continual alterations in the physical features of the control, the mean daily discharge is determined by the shifting control method which involves the application of correction factors based on individual measurements. This method is also used to correct for temporary changes in the control section due to debris or aquatic growth.

The crudeness of instrumentation in the early period was further aggravated by the lack of sufficient personnel to make frequent observations. This necessitated, in some instances, the estimation of unmeasured flows using correlation procedures before actual data were published. The perennial problem of ice reducing the area of the control section during winter was another source of error.

Because of such difficulties, the records published by the U. S. Geological Survey are classified as excellent, good, fair, or poor depending on whether the errors in them are less than 5, 10, or 15 per cent, or greater than 15 per cent, respectively.

3.2 Criteria for the Selection of Stations. The limited scope of this study precluded the analysis of non-homogeneous records. Gaging stations whose flows have been significantly altered by man-made diversions or flow regulation upstream through the construction dams and reservoirs were automatically excluded. Minor diversions, up to a maximum of one per cent of the average annual flow, were, however, tolerated.

Stations were selected basically on the virginity of their flows. Ideally, radical changes in the consumptive use in a basin should have been considered, but because of the imposing if not impossible task that would have entailed, homogeneity in this aspect was assumed. The absence of short term trends was also postulated--in spite of the fact that it has been demonstrated [14, 23] that extensive agricultural exploitation, among other things, can cause perceptible trends in river runoff.

3.3 Stations Selected for Analysis. With the above limitations in mind, a rough survey of the records published by the U. S. Geological Survey yielded 17 runoff gaging stations which satisfied the above criteria and which had records of sufficient length. An arbitrary minimum of 34 years was assumed. The approximate geographic locations of the stations selected are shown in fig. 1.

The mean annual hydrographs obtained by taking the average flow for each day of the year over the total number of years of record, for the sites in fig. 1, is shown in fig. 2. Also plotted are the standard deviations about the mean daily values of each hydrograph.

It may be noted that the limitation caused by the homogeneity requirement imposed a restriction on the size of the drainage basin. The location, drainage area, mean flow, and other pertinent information for each of the stations selected are tabulated in Table 1.

Although some of the records were relatively long, for reasons of possible inconsistencies due to improved instrumentation or more frequent observations during the later years, only records taken after 1921 were used. In addition, a cut-off year of 1960 was used. The choice of both of these dates was again arbitrary.

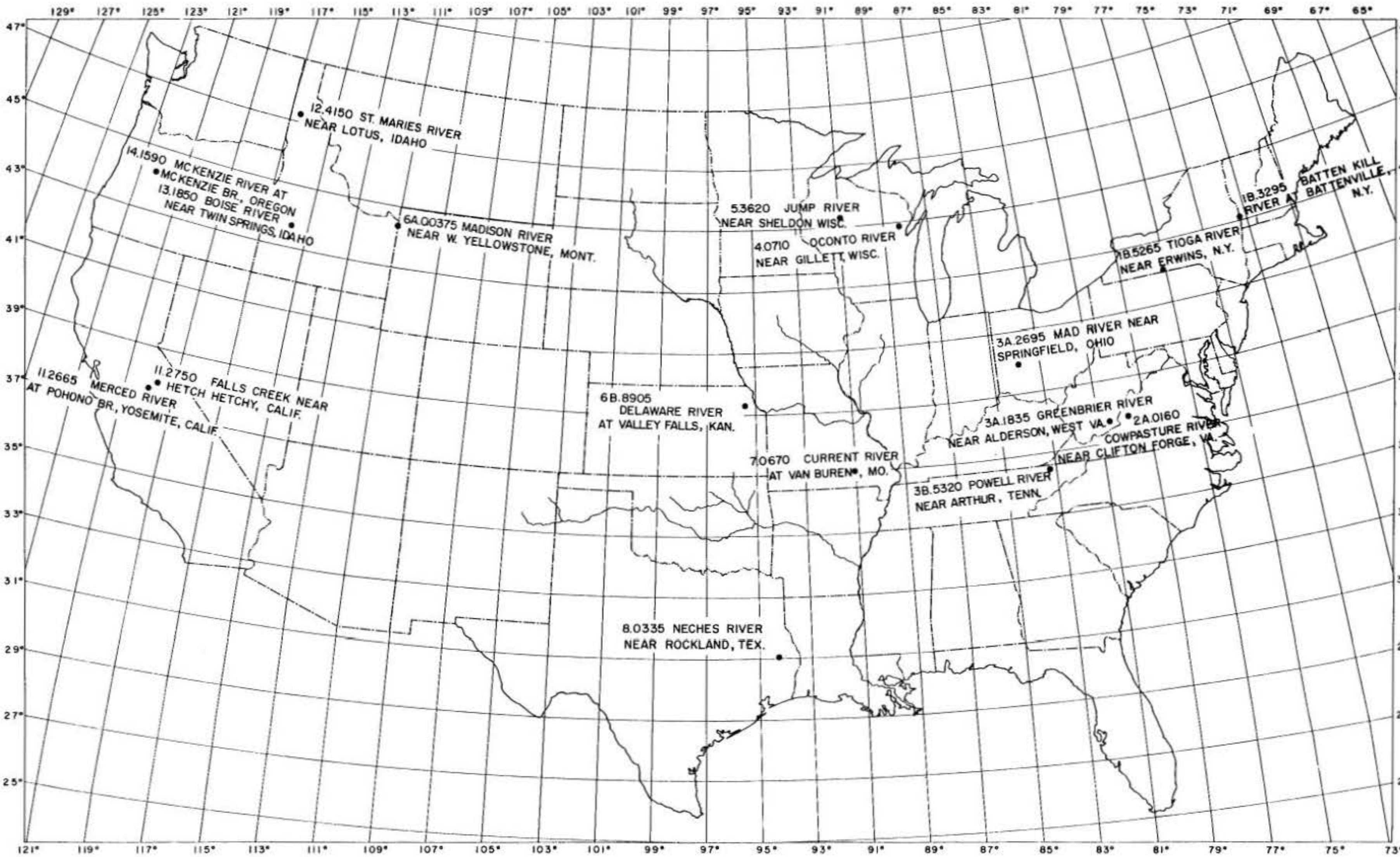


Fig. 1 Geographic distribution of stations selected

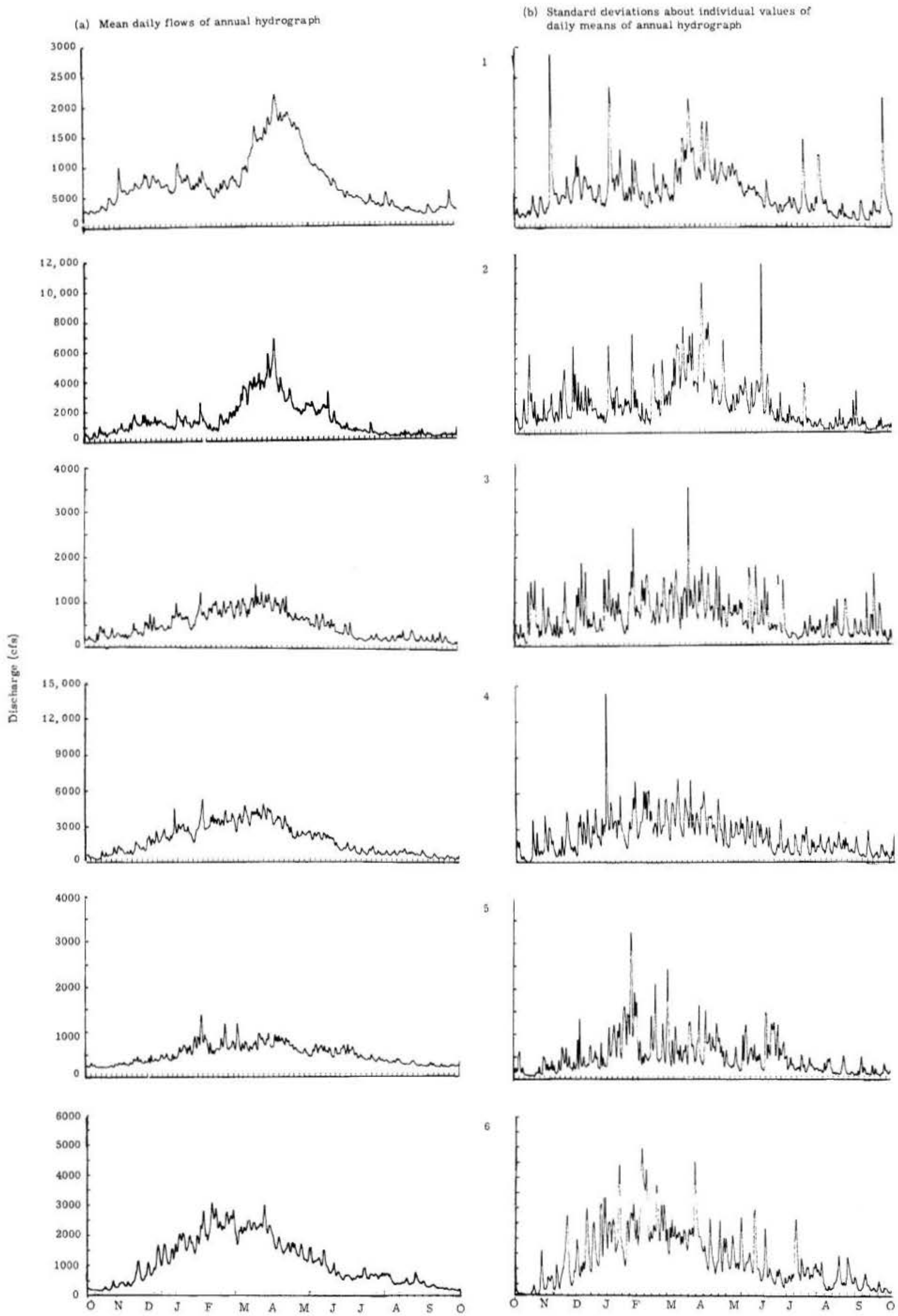


Fig. 2 Daily means and standard deviations about daily means for stations selected - continued

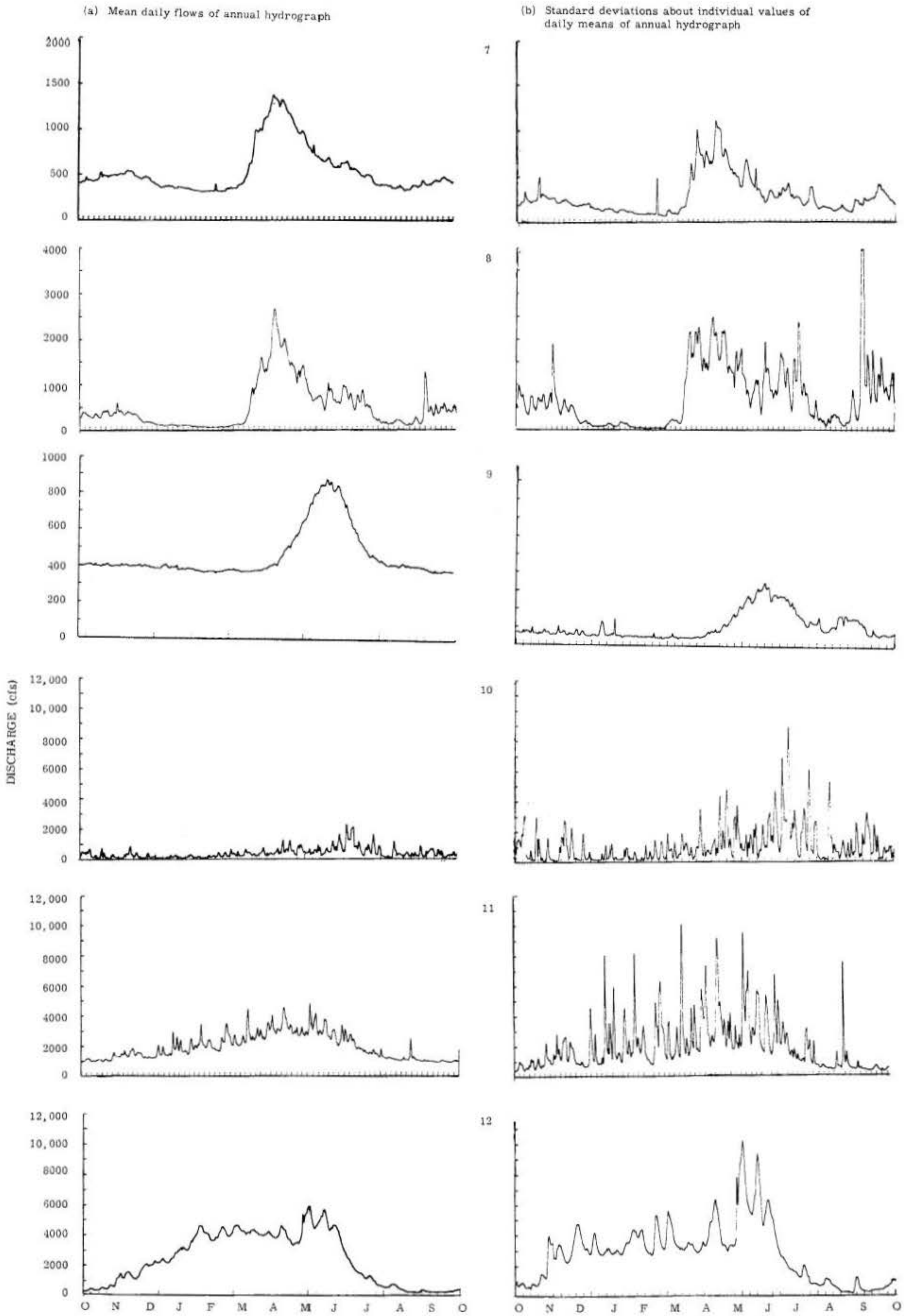
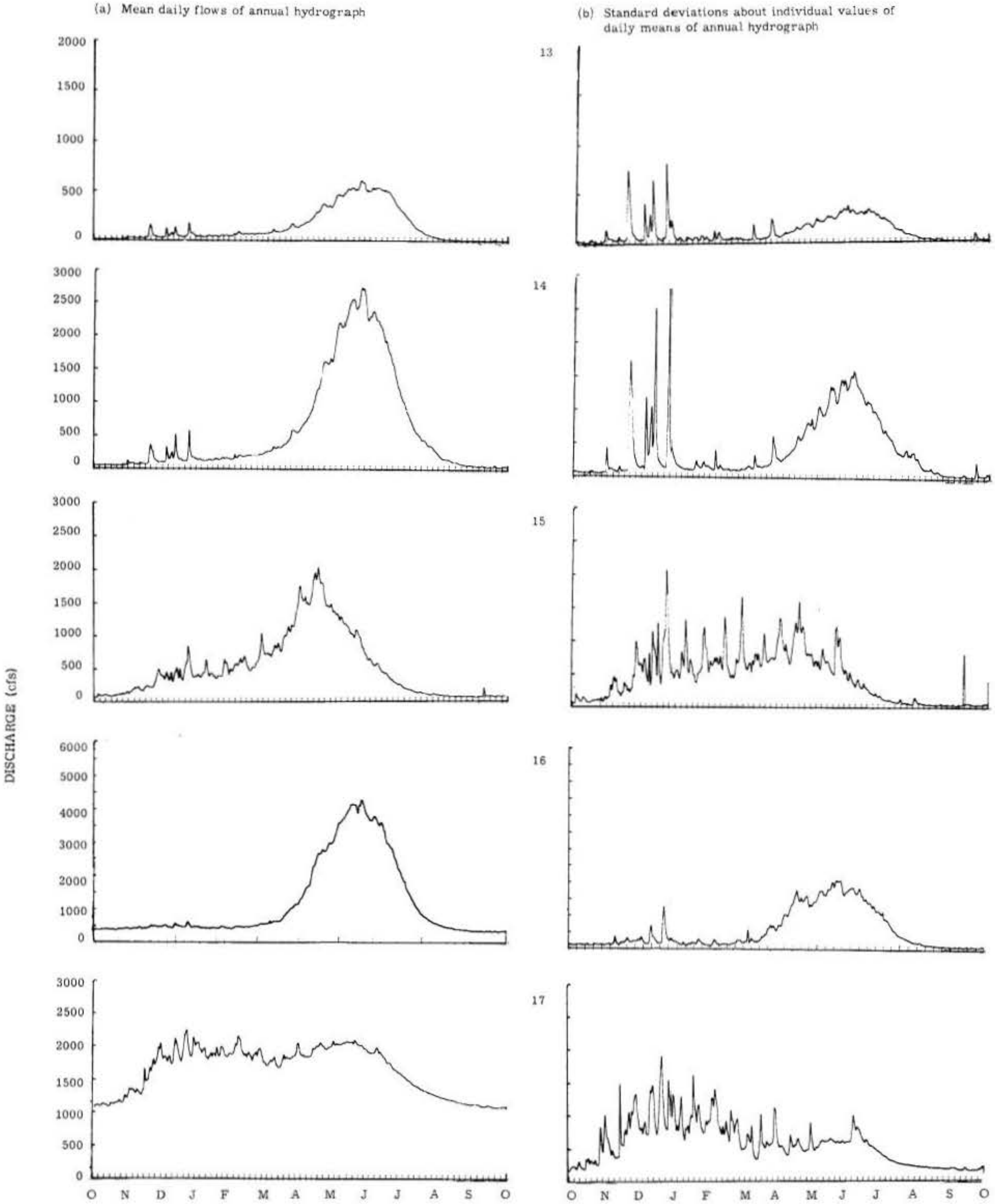


Fig. 2 - continued



(1) Batten Kill River
(2) Tioga River
(3) Cowpasture River
(4) Greenbrier River
(5) Mad River
(6) Powell River
(7) Oconto River
(8) Jump River
(9) Madison River

(10) Delaware River
(11) Current River
(12) Neches River
(13) Falls Creek
(14) Merced River
(15) St. Maries River
(16) Boise River
(17) McKenzie River

Fig. 2 - continued

TABLE 1 STATIONS SELECTED FOR ANALYSIS

Station Number	River	Location		Area (Sq. Mi.)	Records Available	Mean Daily Flow	Standard Deviation	Remarks on Accuracy of Record
		Latitude	Longitude					
1B.3295	Batten Kill at Battenville, N. Y.	$43^{\circ}06'$	$73^{\circ}25'$	394.0	1923 - 1960	722.9	722.9	Good. Fair during periods of ice effect.
1B.5265	Tioga near Erwins, N. Y.	$42^{\circ}07'$	$77^{\circ}08'$	1370.0	1921 - 1960	1378.6	2777.8	Excellent. Fair during periods of ice effect.
2A.0160	Cowpasture near Clifton Forge, Va.	$37^{\circ}48'$	$79^{\circ}46'$	456.0	1926 - 1960	515.6	762.3	Good
3A.1835	Greenbrier near Alderson, W. Va.	$37^{\circ}44'$	$80^{\circ}38'$	1357.0	1921 - 1960	1885.5	3053.4	Good. Poor during periods of ice effect.
3A.2695	Mad near Springfield, Ohio	$39^{\circ}55'$	$83^{\circ}52'$	1474.0	1921 - 1960	487.2	686.7	Good
3B.5320	Powell near Arthur, Tenn.	$36^{\circ}32'$	$83^{\circ}38'$	683.0	1921 - 1960	1116.1	1739.0	Good
4.0710	Oconto near Gillett, Wisconsin	$44^{\circ}52'$	$88^{\circ}18'$	678.0	1921 - 1960	543.5	441.0	Good. Fair during periods of ice effect.
5.3620	Jump near Sheldon, Wisconsin	$45^{\circ}18'$	$90^{\circ}57'$	574.0	1921 - 1960	505.0	1162.0	Good. Fair during periods of ice effect.
6A.0375	Madison near W. Yellowstone, Mont.	$44^{\circ}39'$	$111^{\circ}04'$	419.0	1924 - 1960	458.6	190.7	Excellent. Good during periods of ice effect.
6B.8905	Delaware at Valley Falls, Kansas	$39^{\circ}21'$	$95^{\circ}27'$	922.0	1923 - 1960	375.9	1617.7	Good. Fair during periods of ice effect.
7.0670	Current at Van Buren, Mo.	$37^{\circ}00'$	$91^{\circ}01'$	1667.0	1922 - 1960	1921.0	2694.3	Good. Poor during periods of ice effect.
8.0335	Neches near Rockland, Tex.	$31^{\circ}02'$	$94^{\circ}24'$	3539.0	1924 - 1960	2385.2	3813.0	Good
11.2750	Falls Creek near Hetch-hetchy, Cal.	$37^{\circ}58'$	$119^{\circ}46'$	45.2	1923 - 1960	141.2	234.2	Good. Fair during periods of ice effect.
11.2665	Merced at Pohono Br., Yosemite, Cal.	$37^{\circ}43'$	$119^{\circ}40'$	321.0	1921 - 1960	595.7	979.4	Good
12.4150	St. Maries near Lotus, Idaho	$47^{\circ}15'$	$116^{\circ}38'$	437.0	1923 - 1960	515.0	762.3	Good. Poor during periods of ice effect.
13.1850	Boise near Twin Springs, Idaho	$43^{\circ}40'$	$115^{\circ}44'$	830.0	1921 - 1960	1172.7	1458.6	Excellent. Good during periods of ice effect.
14.1590	McKenzie at McKenzie Br., Ore.	$44^{\circ}11'$	$122^{\circ}08'$	345.0	1924 - 1960	1638.2	744.4	Excellent

CHAPTER IV

DATA PROCESSING AND RESULTS

4.1 Aims in Representation. The analysis was designed to obtain a mathematical representation which would fit the time series of daily flows. The model envisaged is of the general representation 2.4 since the choice of the gaging sites was made in such a way that the trend component is absent or relatively insignificant. The adequacy of the model was based on the amount of the variance explained and the independence of the residual series upon subtraction of deterministic, and stochastic dependence components found. Thus, a relatively small explained variance would be acceptable if it could be shown that after removing the periodic and autoregressive model, the residual series could be approximated by an independently distributed random series.

4.2 Spectral Analysis. The detection of the harmonic component of the time series was achieved by estimating the variance spectrum of the process and inspecting the spectrum for prominent peaks. The general procedure outlined in Sec. 2.5.1 was followed.

Since it was almost a certainty that the annual cycle would show up in the spectrum, provision was made to obtain estimates of the corresponding frequency and those of its harmonics. In order to have estimates at the frequency 0.00273 cycles per day (corresponding to one cycle per year), the maximum number of lags m was narrowed down to a choice over multiples of 365. A value of $m = 365$ gives estimates at intervals of 0.00137, but since this would smudge the estimates over too wide a band of frequencies, a value of $m = 365$ was deemed inadequate. Similarly, $m = 730$ was rejected as not giving enough resolution. Although not entirely satisfactory, a maximum lag of 1095, giving a resolution of 0.000457, was selected as a compromise in consideration of computation time. As to its variability, even that number was found to be too high.

The sampling interval of one day yielded a folding frequency of one cycle every two days, which is a much higher frequency than had previously been thought to be of interest, and the problem of aliasing did not arise.

The sample autocovariances were computed up the value of $m = 1095$ using the formula,

$$C_k = \frac{1}{N-k}\sum_{t=1}^{N-k} X_t X_{t+k} - \frac{1}{(N-k)^2}\sum_{t=1}^{N-k} X_t \sum_{t=1}^{N-k} X_{t+k}, \quad (4.1)$$

which is equivalent to the second part of eq. 2.2.

In so doing, the Fourier cosine series transform was applied by using eq. 2.42, and then linear combinations of the raw estimates V_j's were formed using the coefficients in eq. 2.41 to find the estimate of the variance spectrum.

The spectra of the original time series for the seventeen rivers selected are shown in fig. 3. It can be seen that generally the most conspicuous peaks fall on the frequencies corresponding to the annual cycle and its subharmonics.

For comparative purposes, the covariances obtained were divided by the variance, and the correlograms were also plotted in fig. 4. A cursory inspection of the two sets of figures (i. e., the spectra and the correlograms) provide a graphic illustration of the features of both methods. Whereas the harmonics are distinctly obvious in the spectra, in the correlograms sinusoidal components are indeed present with the fundamental cycle so confounding all other harmonics that distinction between them becomes quantitatively impossible. A further illustration is presented in figs. 5 and 6. For Boise River at Twin Springs, Idaho, the cycles were gradually removed starting with the fundamental harmonic down to its fourth harmonic. The effect on the spectrum was just the obliteration of the peak corresponding to the sinusoid. In the correlogram, however, the removal of harmonics gave prominence to the least-ordered harmonic remaining which was not at all perceptible before removal of cycles with longer periods.

A Chi-square test was made but was not used to test for significance of the peaks because of the high variability in the estimates, compromised earlier to give a good resolution. The alternative was to determine the variances explained by the cycles corresponding to the peaks in the spectrum.

4.3 Harmonic Analysis and Removal of the Periodic Component. An analysis of variance due to the annual cycle was made. This was compared to the explained variances corresponding to the annual cycle and its subharmonics, these being equal to the ordinates of Schuster's periodogram obtained by harmonic analysis.

The harmonic analysis was done using formula 2.9. A maximum of six harmonics was fitted. The results and the variance explained by the analysis of variance are shown in Table 2.

From the last two rows of Table 2, it may be seen that, except for the Delaware River, the six harmonics accounted for more than 80 per cent of the explained variance due to the annual cycle and its harmonics.

It is noted further that in some rivers the use of more sub-harmonics did not increase the variance explained by the cyclic component appreciably. In the interest of representing the series by as few parameters as possible, two methods of removing the cyclic component were attempted--one using all six harmonics and another using just the harmonics which contributed significantly to the explained variance. The number of harmonics used to represent the periodic component for each river is shown in Table 4. The removal of the cyclic component was achieved by subtracting from the original series those cyclic functions whose parameters were obtained from the harmonic analysis.

4.4 Approximation of Wide-Sense Stationarity. Subtraction of the periodic component from the series leaves a stationary residual,

$$\epsilon_t = X_t - P_t , \tag{4.2}$$

where P_t is the periodic component characterized by parameters determined through harmonic analysis. Fig. 2, however, shows that the variances of the seventeen series are not constant with time. The harmonic representation of the time series itself suggests that the series of standard deviations may be similarly fitted by a periodic function. A harmonic analysis of the series of standard deviations was made yielding harmonic function S_t. The results are tabulated in Table 3.

As for the cyclic component, two methods were used to represent S_t, one using all six harmonics and another using only those harmonics which contribute significantly to the variance of S_t as shown in Table 4.

With harmonic representations for the periodic component of the time series and for the series of standard deviations, a standardized residual was obtained:

$$y_t = \frac{\epsilon_t}{S_t} = \frac{X_t - P_t}{S_t} . \tag{4.3}$$

While this transformation does not necessarily insure a wide-sense stationary series, its merits can only be judged by the final results after the complete mathematical fitting procedure has been made and a check on the properties of the residual series is done. It might be noted that this standardization does not result in a series with mean zero and variance unity. To achieve these convenient properties, another standardization was necessary, and was done by subtracting the mean of the new series from each term and dividing each result by the new standard deviation; i.e.,

$$Z_t = \frac{y_t - \bar{y}}{S_y} . \tag{4.4}$$

4.5 Autoregressive Representation. The choice of the autoregressive scheme to be applied on the residual series 4.3 was based on the fulfillment of the requirement that the roots of the characteristic equation 2.16 must all lie in the unit circle. A preliminary check showed that various possible schemes for each of the stations analyzed met this condition. The logical choice is the scheme which requires the least number of parameters and yet adequately describes the series. Thus, lower ordered schemes were first investigated, and, if they proved unsatisfactory, the order was progressively increased until a satisfactory scheme was found.

In fitting the first order linear Markov model, ρ was estimated by its least square estimator r_1. Using this estimate, a primary series,

$$\eta_t = Z_t - \hat{\rho} Z_{t-1} , \tag{4.5}$$

was obtained. If the representation 4.5 is adequate, then the η's should be uncorrelated with the Z's and

$$\text{var } \eta_t = (1-\hat{\rho}^2) \text{ var } Z_t , \tag{4.6}$$

Fulfillment of this requirement was the test of the validity of a scheme. The results are summarized in Table 5.

The second order autoregressive scheme,

$$\eta_t = Z_t + a_1 Z_{t-1} + a_2 Z_{t-2} , \tag{4.7}$$

was then tried. Values of a_1 and a_2 were estimated using eq. 2.21. For all rivers, the roots of the resulting characteristic equations were within the unit circle of the complex plane.

The variance of the residual was then obtained and compared with the value given by Kendall:

$$\frac{\text{var } \eta}{\text{var } Z} = \frac{1 - a_2}{1 + a_2}\left\{(1 + a_2)^2 - a_1^2\right\}. \tag{4.8}$$

The results are tabulated in Table 6 for the residual series after removing the harmonics in Table 4 for the residual series after removing six harmonics are tabulated in Table 7.

The correlograms of the primary series η_t were also obtained and plotted in fig. 7. Anderson's test for significance for r_L was made, but because of the size of the sample, the confidence band was found to be too narrow for practical use.

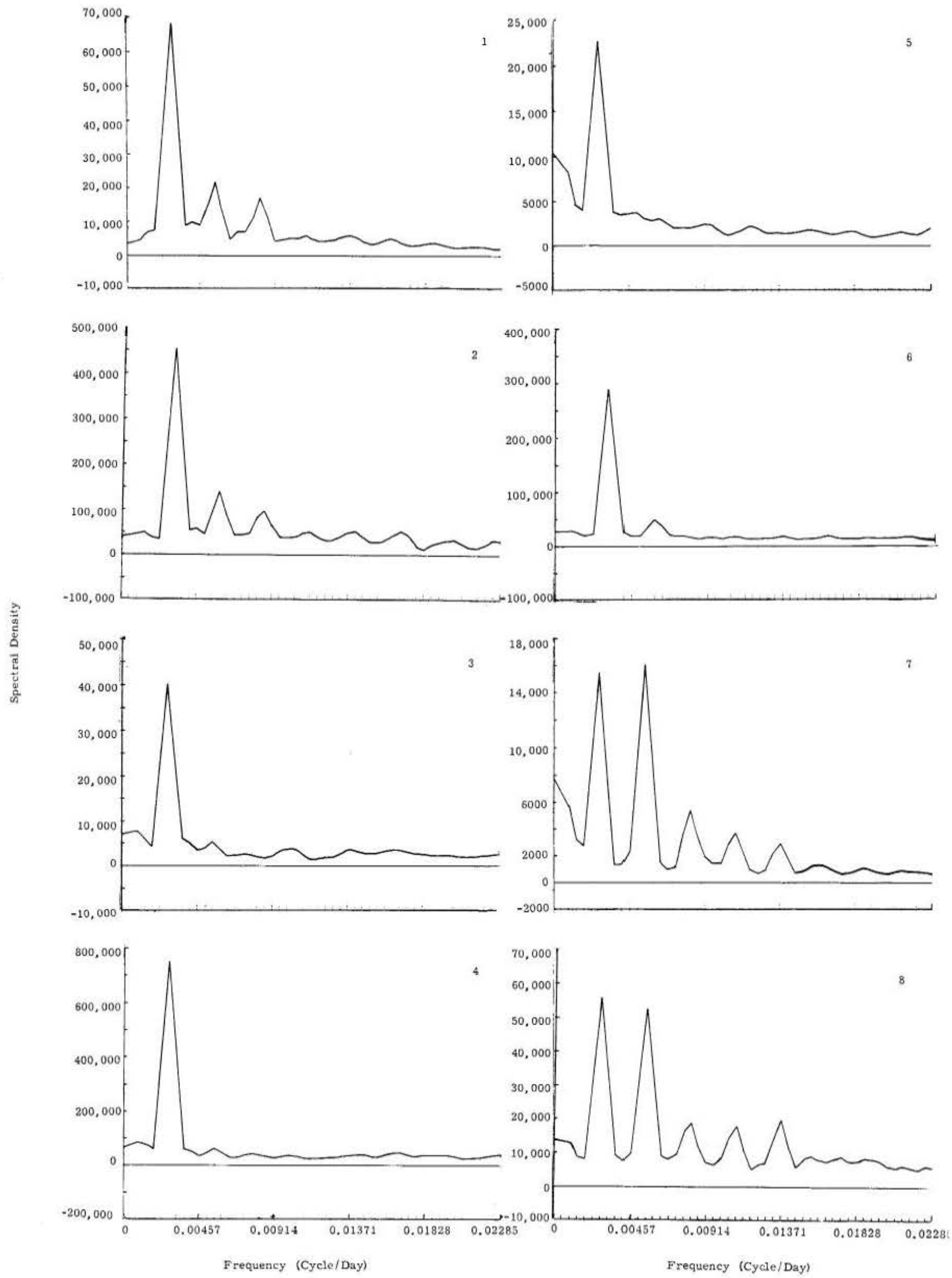
Although the results of the second order autoregressive fitting were quite satisfactory, a third order autoregressive scheme was tried for the series which satisfied the conditions imposed on the roots of the characteristic equation. Wold's general procedure was used in estimating the paramters a_1, a_2, and a_3. On checking the roots of the characteristic equation, only Neches River failed to satisfy the requirement. The variance of the series,

$$\eta_t = Z_t + a_1 Z_{t-1} + a_2 Z_{t-2} + a_3 Z_{t-3} \tag{4.9}$$

was then obtained and compared with the explained variance assuming a multiple regression. Results are tabulated in Table 8.

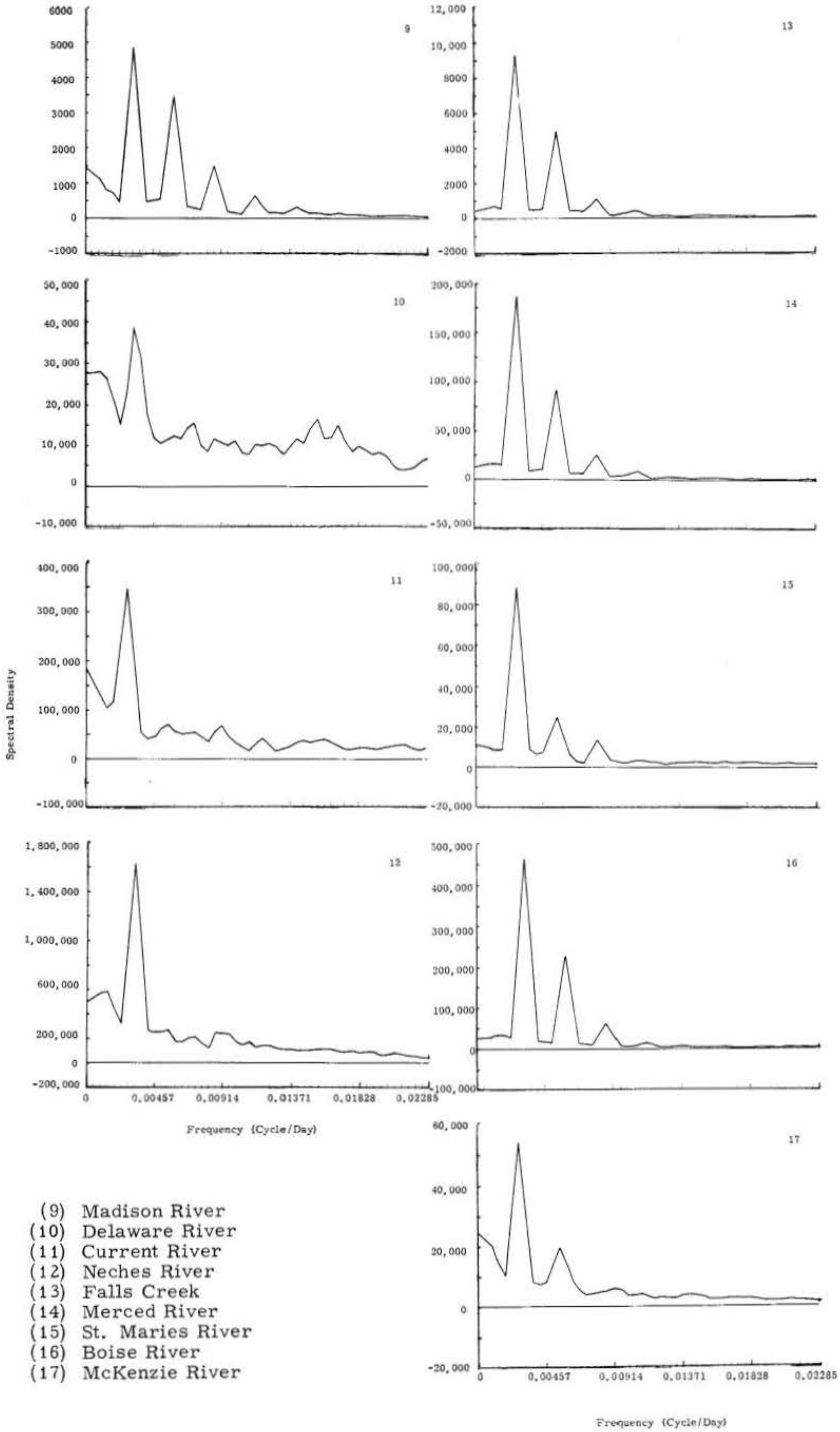
4.6 Quenouille's Tests. The tests suggested by Quenouille were applied to the first and second order autoregressive schemes. Despite the validity of the models as evidenced by Tables 5 and 6, Quenouille's test rejected the model. Similar results were observed for all the other stations which were found to satisfy the second order autoregressive representation.

4.7 Computer Program. It is quite apparent that with the volume of data used and the calculations invoked, the task of making statistical inferences would be impossible without the aid of a computer. All but the simplest arithmetic calculations were done at first on a high speed digital computer and at a later stage on a faster high speed computer. Where it was convenient and practicable, the plotting of figures was also done on a data plotter, a cathode-ray tube device hooked up to the computer used and from which photographic prints were made.



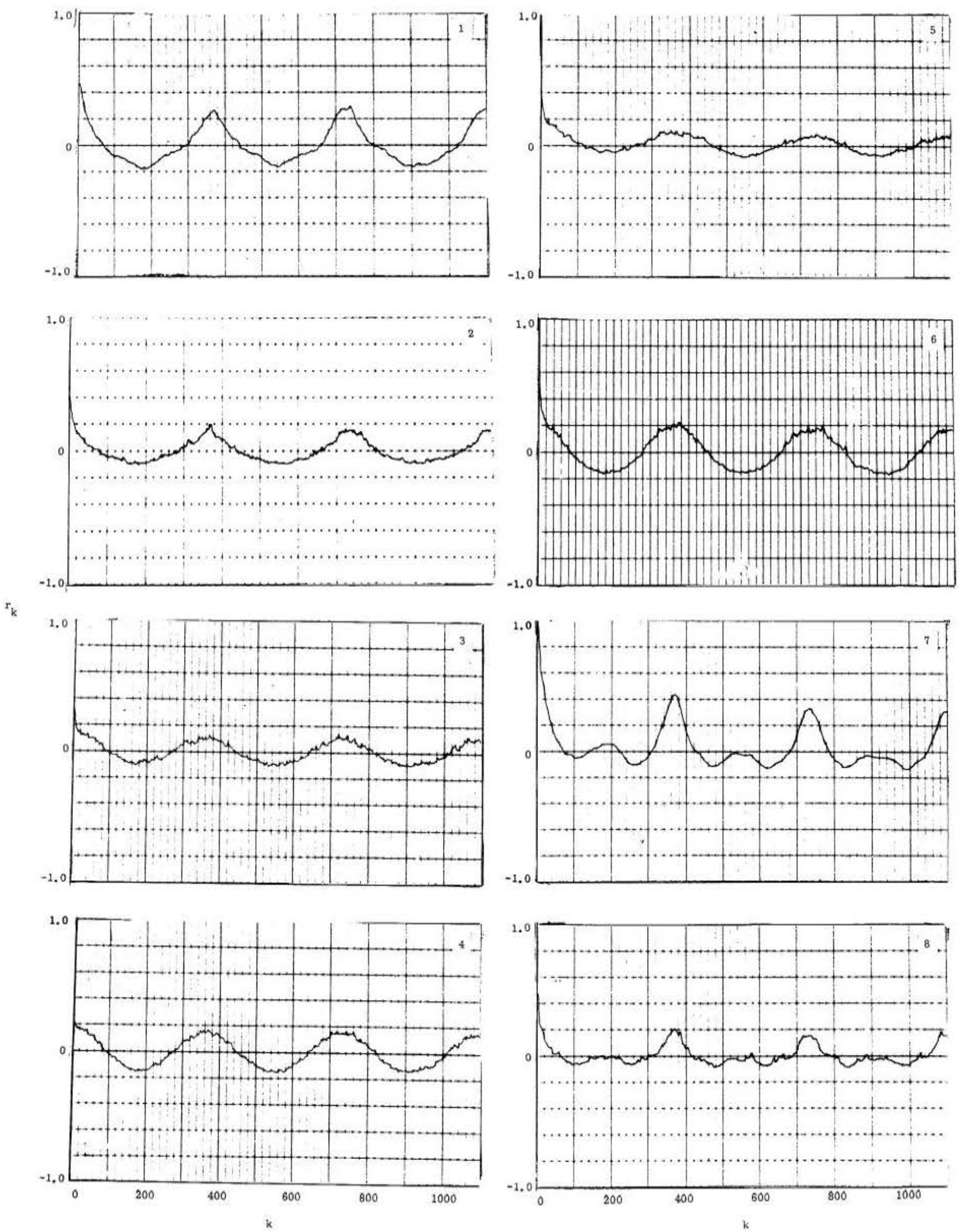
(1) Batten Kill River
(2) Tioga River
(3) Cowpasture River
(4) Greenbrier River
(5) Mad River
(6) Powell River
(7) Oconto River
(8) Jump River

Fig. 3 Spectral density of daily flows of stations selected - continued



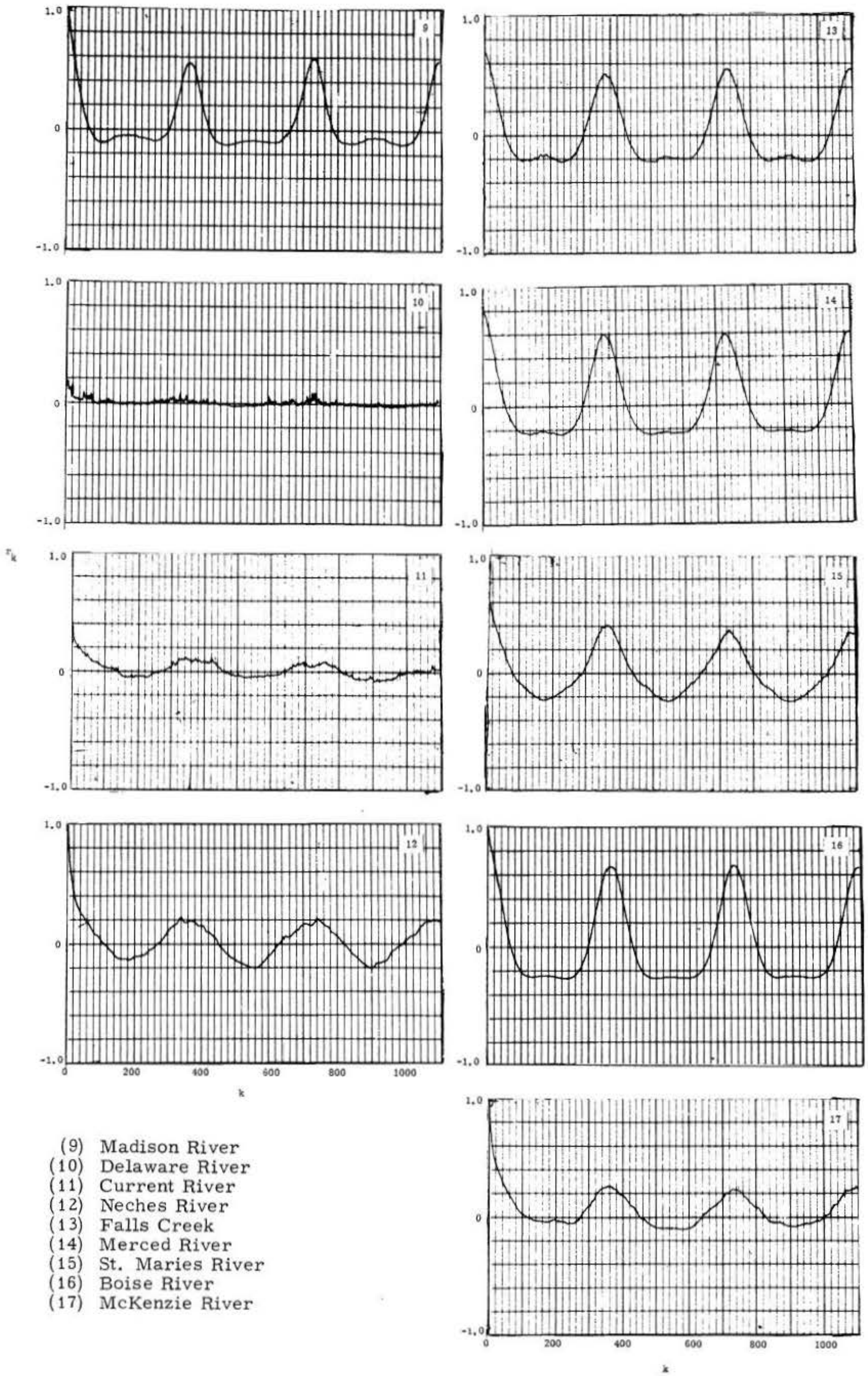
(9) Madison River
(10) Delaware River
(11) Current River
(12) Neches River
(13) Falls Creek
(14) Merced River
(15) St. Maries River
(16) Boise River
(17) McKenzie River

Fig. 3 - continued



(1) Batten Kill River
(2) Tioga River
(3) Cowpasture River
(4) Greenbrier River
(5) Mad River
(6) Powell River
(7) Oconto River
(8) Jump River

Fig. 4 Correlogram of daily flows of stations selected - continued



(9) Madison River
(10) Delaware River
(11) Current River
(12) Neches River
(13) Falls Creek
(14) Merced River
(15) St. Maries River
(16) Boise River
(17) McKenzie River

Fig. 4 - continued

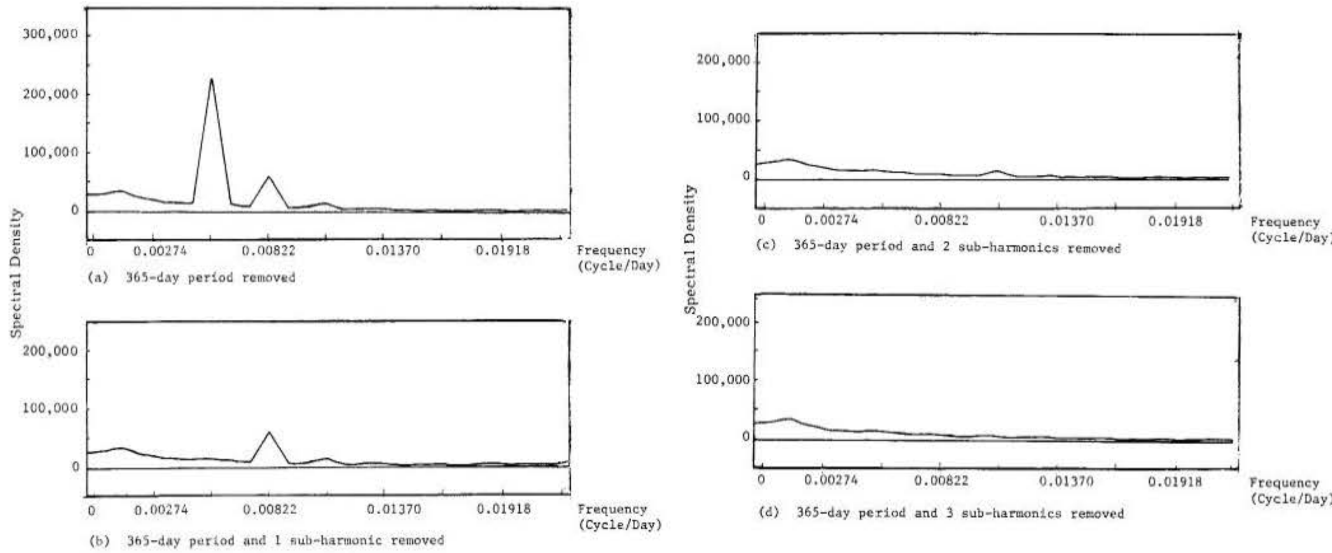


Fig. 5 Obliteration of peaks in the spectrum of Boise River with gradual removal of lower-ordered harmonic components

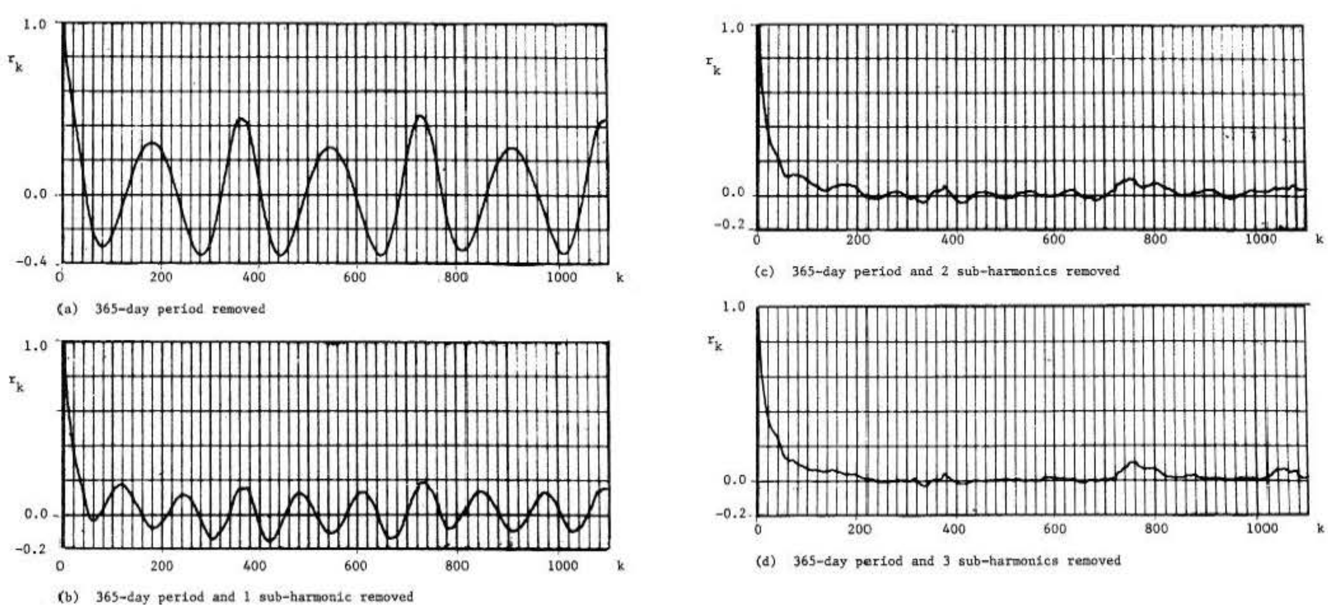


Fig. 6 Emergence of short-period cyclicity in the correlograms of Boise River with gradual removal of lower-ordered harmonic components

TABLE 2 RESULTS OF HARMONIC ANALYSIS AND ANALYSIS OF VARIANCE OF ORIGINAL TIME SERIES OF DAILY FLOW

Harmonic		Batten Kill River	Tioga River	Cow-pasture River	Green-brier River	Mad River	Powell River	Oconto River	Jump River	Madi-son River	Dela-ware River	Cur-rent River	Neches River	Falls Creek	Merced River	St. Maries River	Boise River	Mc-Kenzie River
	A	-503.1	-1274.7	-344.0	-1482.0	-261.8	-888.3	-200.3	-338.0	-60.2	-98.6	-1014.0	-2225.4	-133.0	-563.5	-563.8	-867.9	-414.0
1	B	71.1	245.3	147.8	753.2	62.8	510.0	-112.4	-259.4	-116.1	-252.4	-126.8	477.0	-135.1	-625.9	-6.8	-1014.0	125.2
	Explained Variance	0.1843	0.1092	0.0975	0.1483	0.0769	0.1734	0.1356	0.0672	0.2350	0.0140	0.0743	0.1781	0.3274	0.3697	0.2736	0.4187	0.1689
	A	193.6	613.5	79.0	140.8	-8.7	-20.3	145.4	310.8	-48.5	-46.3	112.1	-33.0	-485	-166.4	198.8	-228.3	-164.0
2	B	199.8	268.6	-46.1	-276.1	-31.8	-311.0	185.0	286.7	99.7	92.5	303.2	349.0	125.3	558.0	193.7	900.5	130.9
	Explained Variance	0.0552	0.0291	0.0058	0.0052	0.0012	0.0161	0.1423	0.0661	0.1691	0.0020	0.0072	0.0042	0.1646	0.1767	0.0663	0.2028	0.0379
	A	-192.8	-501.4	-31.5	-140.4	17.8	-17.6	-85.5	-137.1	69.3	56.8	49.5	418.9	62.4	278.2	-100.9	422.2	12.9
3	B	-128.5	-42.1	-12.2	-29.2	2.6	-1.5	-79.9	-132.8	-9.7	71.8	-87.4	-218.5	-9.8	-114.9	-158.4	-180.1	-79.9
	Explained Variance	0.0383	0.0164	0.0008	0.0011	0.0003	0.0001	0.0352	0.0135	0.0674	0.0016	0.0007	0.0077	0.0364	0.0472	0.0304	0.0495	0.0059
	A	80.6	238.8	4.4	24.5	17.1	-6.2	58.0	174.2	-30.9	65.0	-39.2	-342.1	-14.3	-124.6	28.1	-145.5	-18.6
4	B	72.8	-194.4	-9.1	-66.1	11.5	11.0	65.6	83.5	-33.3	-60.0	8.6	-96.0	-29.5	-95.3	63.2	-113.0	-43.8
	Explained Variance	0.0084	0.0046	0.0001	0.0003	0.0005	0.0000	0.0197	0.0138	0.0284	0.0015	0.0001	0.0043	0.0098	0.0128	0.0041	0.0008	0.0020
	A	-47.6	-219.6	-27.3	-58.9	-43.4	13.6	-39.8	-146.4	-8.9	-49.0	-88.2	17.0	-7.1	9.7	3.1	-15.9	22.8
5	B	-57.6	129.9	16.9	121.1	-7.8	5.4	-72.5	-173.0	26.9	-77.9	13.5	188.2	1.0	56.9	-53.8	36.6	18.8
	Explained Variance	0.0040	0.0042	0.0007	0.0010	0.0021	0.0000	0.0176	0.0190	0.0111	0.0016	0.0005	0.0012	0.0005	0.0017	0.0025	0.0004	0.0008
	A	15.5	89.8	13.2	41.0	13.5	7.9	7.4	13.8	5.8	-22.2	8.8	67.3	0	10.1	6.2	2.2	9.7
6	B	-5.5	-136.0	11.2	-26.5	-4.6	23.9	27.8	66.9	1.6	63.2	2.6	-192.0	7.4	-2.8	47.2	34.9	24.4
	Explained Variance	0.0002	0.0017	0.0002	0.0001	0.0002	0.0001	0.0021	0.0017	0.0005	0.0009	0.0000	0.0014	0.0005	0.0001	0.0020	0.0003	0.0006
Variance Explained by 6 Harmonics		0.2904	0.1632	0.1051	0.1560	0.0812	0.1897	0.3525	0.1813	0.5115	0.0216	0.0828	0.1969	0.5387	0.6082	0.3789	0.6725	0.2161
Total Variance Explained by Annual Cycle		0.3019	0.1840	0.1210	0.1709	0.1003	0.2029	0.3584	0.1970	0.5152	0.0445	0.0994	0.2029	0.5463	0.6120	0.3894	0.6818	0.2256

TABLE 3 RESULTS OF HARMONIC ANALYSIS OF THE TIME SERIES OF STANDARD DEVIATIONS ABOUT INDIVIDUAL VALUES OF DAILY MEANS OF ANNUAL HYDROGRAPH

Harmonic		Batten Kill River	Tioga River	Cow-pasture River	Green-brier River	Mad River	Powell River	Oconto River	Jump River	Madi-son River	Dela-ware River	Cur-rent River	Neches River	Falls Creek	Merced River	St. Maries River	Boise River	Mc-Kenzie River
	A	-269.3	-1286.6	-278.5	-1222.8	-302.5	-798.3	-123.3	-194.4	-18.6	-137.8	-1372.9	-1938.8	-49.5	-229.3	-383.0	-462.8	-186.0
1	B	143.3	485.0	180.3	925.2	154.4	648.3	-85.6	-406.3	-79.4	-739.7	-63.6	398.1	-47.6	-259.7	202.7	-480.5	280.4
	Explained Variance	0.3081	0.3574	0.2799	0.4822	0.3228	0.6009	0.2706	0.1862	0.5578	0.2065	0.3375	0.5104	0.1759	0.2916	0.6556	0.6102	0.5157
	A	78.8	565.1	77.2	-79.2	-93.6	-235.9	141.6	439.5	-30.9	-127.5	-11.5	-167.5	-61.5	-214.2	11.4	-133.6	-203.6
2	B	67.0	287.5	-15.1	-246.4	-51.2	-267.6	105.7	226.7	41.3	240.0	458.6	1179.0	81.1	342.3	91.7	455.7	78.6
	Explained Variance	0.0354	0.0760	0.0157	0.0137	0.0318	0.0723	0.3750	0.2244	0.2233	0.0269	0.0376	0.1848	0.3855	0.3961	0.0298	0.3092	0.2170
	A	-142.3	-386.3	-4.6	-133.8	91.6	11.8	-66.4	-75.7	29.9	107.7	96.0	373.2	9.9	82.3	-50.1	142.2	13.0
3	B	-10.8	168.1	-6.1	-44.5	14.5	-47.8	-46.2	-105.2	-11.8	199.7	-143.4	-371.6	4.9	-42.1	-108.9	-92.2	-25.4
	Explained Variance	0.0674	0.0336	0.0001	0.0041	0.0240	0.0014	0.0786	0.0154	0.0868	0.0188	0.0053	0.0361	0.0046	0.0208	0.0502	0.0394	0.0037
	A	91.7	168.5	-23.4	-14.6	27.4	-47.3	75.7	236.0	-20.5	199.1	-77.0	-755.7	2.8	-40.4	8.0	-1.5	-7.6
4	B	15.3	-267.0	-27.2	-1.7	24.8	31.4	31.7	-107.4	-17.7	-121.2	8.8	-116.5	-31.4	-125.3	-20.1	-48.9	-48.5
	Explained Variance	0.0286	0.0188	0.0033	0.0000	0.0038	0.0018	0.0809	0.0617	0.0616	0.0198	0.0011	0.0762	0.0370	0.0421	0.0016	0.0033	0.0110
	A	-18.3	-243.6	-48.0	15.0	-124.0	2.4	-47.2	-259.6	-9.6	-109.5	-235.2	200.6	2.9	26.0	29.3	-11.6	-11.7
5	B	13.9	320.3	51.9	187.9	-4.7	-30.7	-43.2	-206.6	14.4	-269.8	87.0	110.2	-10.2	13.2	-12.2	-15.6	-23.5
	Explained Variance	0.0017	0.0306	0.0127	0.0073	0.0431	0.0005	0.0492	0.1010	0.0250	0.0309	0.0112	0.0068	0.0042	0.0021	0.0035	0.0005	0.0031
	A	6.6	117.5	28.4	3.8	38.5	3.0	8.6	-41.7	-0.5	-75.1	12.0	90.6	1.8	25.0	10.9	-3.1	10.9
6	B	-115.8	-207.2	-6.4	-10.1	-30.1	-52.5	4.3	-59.0	6.8	138.5	-0.5	-288.2	5.7	24.1	39.6	55.3	44.7
	Explained Variance	0.0446	0.0107	0.0022	0.0000	0.0067	0.0016	0.0011	0.0048	0.0039	0.0091	0.0000	0.0119	0.0013	0.0029	0.0059	0.0042	0.0097
Variance Explained by 6 Harmonics		0.4858	0.5271	0.3139	0.5073	0.4322	0.6785	0.8554	0.5935	0.9584	0.3120	0.3927	0.8262	0.6085	0.7556	0.7466	0.9668	0.7602

TABLE 4 NUMBER OF HARMONICS USED TO REPRESENT P_t and S_t

River	P_t			S_t		
	Number of Harmonics	Percent of Explained Variance	Total Exp. Variance due to Annual Cycle	Number of Harmonics	Percent of Explained Variance	Total Exp. Variance due to Six Harmonics
Batten Kill	3	0.2778	0.3019	6	0.4858	0.4858
Tioga	3	0.1547	0.1840	6	0.5271	0.5271
Cowpasture	2	0.1033	0.1210	5	0.3117	0.3139
Greenbrier	1	0.1483	0.1709	2	0.4959	0.5073
Mad	1	0.0769	0.1003	5	0.4255	0.4322
Powell	1	0.1734	0.2029	2	0.6732	0.6785
Oconto	5	0.3504	0.3584	5	0.8543	0.8554
Jump	5	0.1796	0.1970	5	0.5887	0.5935
Madison	5	0.5110	0.5152	5	0.9535	0.9584
Delaware	1	0.0140	0.0445	6	0.3120	0.3120
Current	1	0.0743	0.0994	5	0.3927	0.3927
Neches	1	0.1781	0.2029	4	0.8075	0.8262
Falls Creek	4	0.5382	0.5463	4	0.6030	0.6085
Merced	4	0.6064	0.6120	4	0.7406	0.7556
St. Maries	3	0.3703	0.3894	3	0.7356	0.7466
Boise	3	0.6710	0.6818	3	0.9588	0.9668
McKenzie	2	0.2086	0.2256	2	0.7327	0.7602

TABLE 5 RESULTS OF FITTING FIRST ORDER AUTOREGRESSIVE SCHEME TO STANDARDIZED STOCHASTIC COMPONENT
(after removing harmonics of Table 4)

RIVER	$-a_1 = \hat{\rho} = r_1$	Variance of Residual Series	
		Theoretical	Computed
Batten Kill	0.78956	0.37660	1.27026
Tioga River	0.59174	0.64984	0.87947
Cowpasture	0.65989	0.56455	1.03938
Greenbrier	0.70010	0.50986	1.09418
Mad River	0.63625	0.59519	0.85853
Powell River	0.84002	0.29437	1.45685
Oconto River	0.79593	0.36650	0.73138
Jump River	0.67083	0.54999	0.59837
Madison River	0.93719	0.12168	0.65934
Delaware River	0.58911	0.65295	0.61152
Current River	0.71192	0.49317	0.92779
Neches River	0.96823	0.06253	0.74477
Falls Creek	0.80647	0.34961	0.72066
Merced River	0.82632	0.32720	0.14579
St. Maries River	0.55836	0.68824	0.84098
Boise River	0.95457	0.08880	1.06296
McKenzie River	0.93866	0.11892	0.72816

TABLE 6 RESULTS IN FITTING SECOND ORDER AUTOREGRESSIVE SCHEMES TO STANDARDIZED STOCHASTIC COMPONENTS.
(After removing harmonics of Table 4)

River	a_1	a_2	Absolute Value of the Roots of the Characteristic Equation	Variance of Residual Series	
				Theoretical	Computed
Batten Kill	-0.90850	+0.15065	0.69023 and 0.21827	0.3642	0.3681
Tioga	-0.60930	+0.02967	0.55593 and 0.05337	0.6493	0.6486
Cowpasture	-0.71559	+0.08441	0.56661 and 0.14897	0.5605	0.5606
Greenbrier	-0.77385	+0.10534	0.59755 and 0.17629	0.5042	0.5043
Mad	-0.60875	-0.04321	0.67295 and 0.06421	0.5941	0.5941
Powell	-1.07380	+0.27830	0.63671 and 0.43709	0.2716	0.2714
Oconto	-0.54176	-0.31933	0.89754 and 0.35578	0.3291	0.3292
Jump	-0.58300	-0.13094	0.75617 and 0.17317	0.5405	0.5406
Madison	-0.92528	-0.01270	0.95069 and 0.02559	0.1217	0.1218
Delaware	-0.65798	+0.11690	0.34192*	0.6440	0.6448
Current	-0.73531	+0.03285	0.77756 and 0.04224	0.4926	0.4927
Neches	-1.32541	+0.36890	0.92780 and 0.39760	0.0540	0.0543
Falls Creek	-0.94554	+0.17244	0.69876 and 0.24678	0.3392	0.3393
Merced	-0.80397	-0.02704	0.83630 and 0.03234	0.3175	0.3176
St. Maries	-0.42437	-0.023998	0.78219 and 0.35783	0.6486	0.6487
Boise	-1.03034	+0.07937	0.94648 and 0.08386	0.0882	0.0882
McKenzie	-0.98179	+0.04594	0.93252 and 0.04926	0.1186	0.1188

* Roots are complex conjugates.

TABLE 7 RESULTS OF FITTING SECOND ORDER AUTOREGRESSIVE SCHEMES TO STANDARDIZED STOCHASTIC COMPONENTS
(After removing 6 harmonics)

River	a_1	a_2	Absolute Values of the Roots of the Characteristic Equation	Variance of Residual Series	
				Theoretical	Computed
Batten Kill	-0.93028	+0.16011	0.70231 and 0.22797	0.3477	0.3479
Tioga	-0.62523	+0.04096	0.25873 and 0.21777	0.6289	0.6375
Cowpasture	-0.71181	+0.08103	0.56956 and 0.14226	0.5627	0.5628
Greenbrier	-0.78026	+0.12609	0.62167 and 0.15859	0.5124	0.5125
Mad	-0.60685	-0.05580	0.68828 and 0.08144	0.5851	0.5852
Powell	-1.08724	+0.33729	0.58075*	0.3001	0.3002
Oconto	-0.56671	-0.30560	0.90355 and 0.33683	0.3027	0.3028
Jump	-0.51370	-0.11271	0.67543 and 0.16973	0.6564	0.6566
Madison	-0.92798	-0.01065	0.91636 and 0.01162	0.1250	0.1252
Delaware	-0.62514	+0.11192	0.33450*	0.6750	0.6758
Current	-0.69157	+0.00178	0.68898 and 0.00260	0.5233	0.5234
Neches	-1.45329	+0.50655	0.87315 and 0.56014	0.0520	0.0521
Falls Creek	-0.92855	+0.15132	0.71772 and 0.11084	0.3415	0.3421
Merced	-0.78877	-0.02332	0.82730 and 0.02854	0.3475	0.3477
St. Maries	-0.32864	-0.21764	0.65894 and 0.33030	0.7845	0.7846
Boise	-0.99132	+0.09326	0.88606 and 0.10526	0.1760	0.1761
McKenzie	-0.96222	+0.04026	0.91865 and 0.04357	0.1442	0.1443

* Roots are complex conjugates

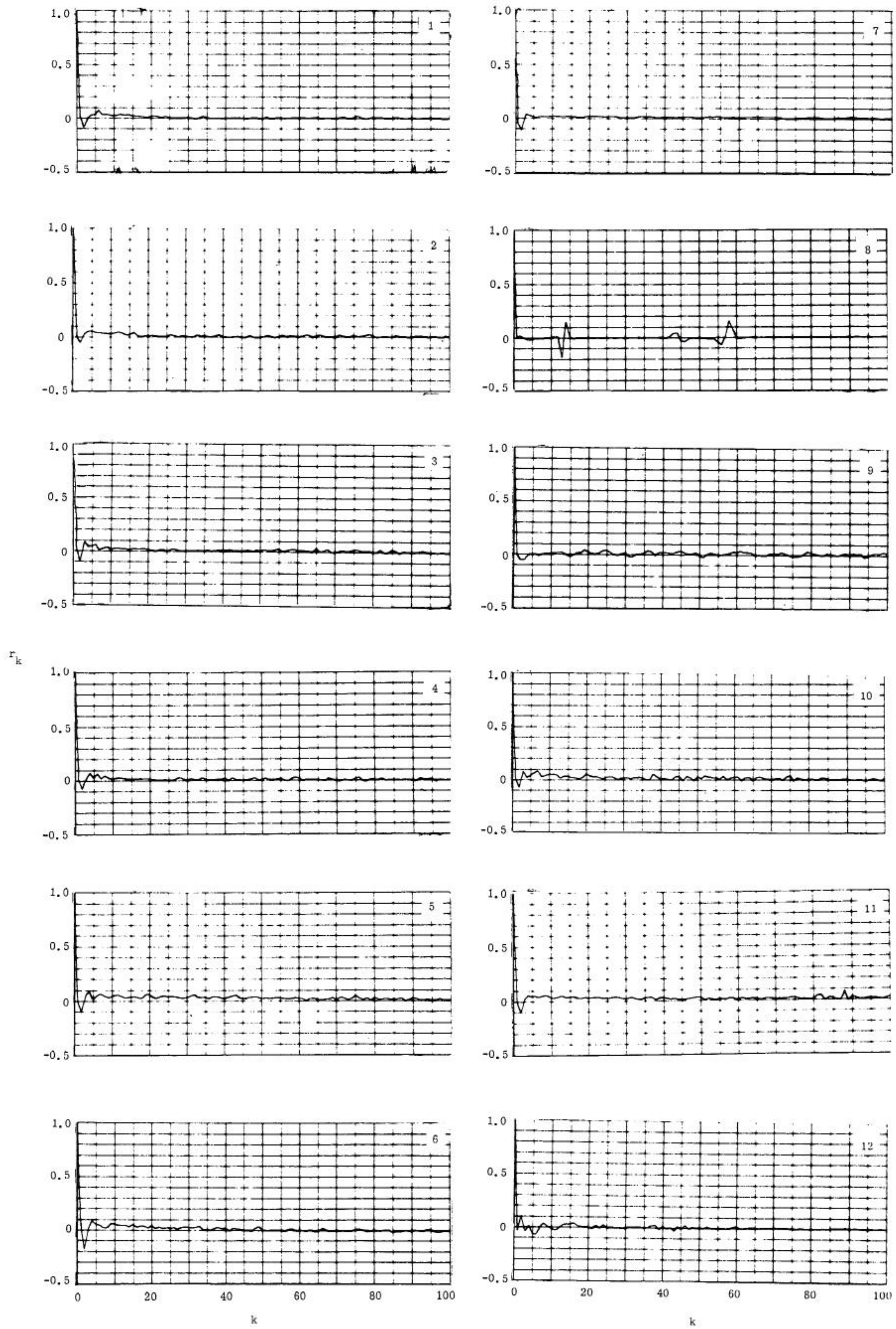


Fig. 7 Correlograms of residual series - continued

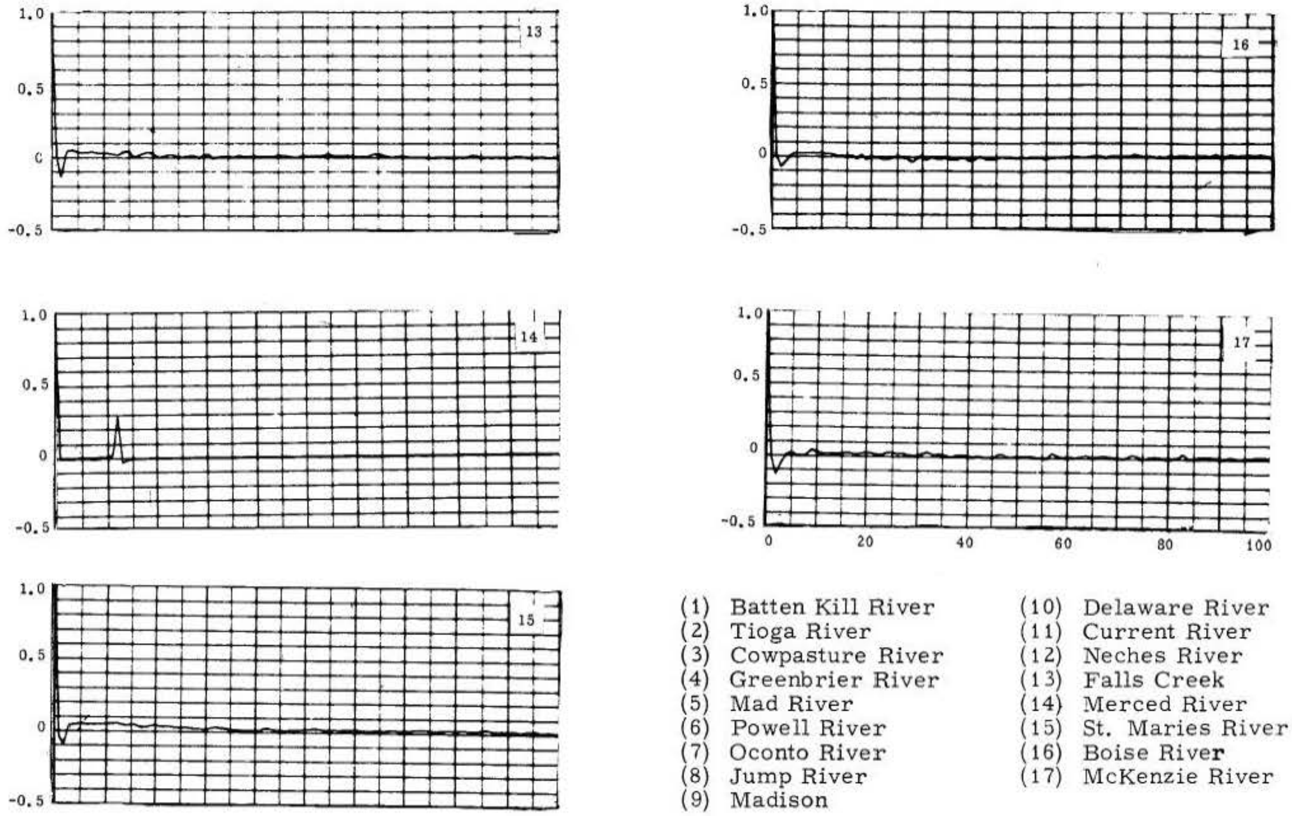


Fig. 7 - continued

TABLE 8 RESULTS OF FITTING THIRD ORDER AUTOREGRESSIVE SCHEMES TO STANDARDIZED STOCHASTIC COMPONENTS
(after removing harmonics of Table 4)

River	a_1	a_2	a_3	Absolute Value of Roots of Characteristic Equation	Variance of Residuals	
					Theoretical	Computed
Batten Kill	-0.92708	+0.26266	-0.12328	0.79198 and 0.39454	0.36805	1.46139
Tioga	-0.61216	+0.08856	-0.09666	0.68775 and 0.37490*	0.64927	0.89314
Cowpasture	-0.72772	+0.18718	-0.14361	0.73783 and 0.44118*	0.56052	1.10120
Greenbrier	-0.78669	+0.19979	-0.12208	0.73971 and 0.40625*	0.50420	1.17584
Mad	-0.60199	+0.05223	-0.15677	0.78811 and 0.44600*	0.59408	0.84960
Powell	-1.13961	+0.53231	-0.23658	0.84115 and 0.53034*	0.27157	2.04490
Oconto	-0.49715	-0.24365	-0.13970	0.92428 and 0.388769*	0.32912	0.61660
Jump	-0.58630	-0.14566	+0.02525	0.73740 and 0.27542*	0.54056	0.56842
Madison	-0.92460	+0.03762	-0.05438	0.94563 and 0.23980*	0.12166	0.65124
Delaware	-0.67158	+0.19351	-0.11645	0.64964 and 0.42338*	0.64403	0.64000
Current	-0.74003	+0.13817	-0.14322	0.79340 and 0.42487*	0.49262	0.95634
Neches	+0.28416	-0.13857	-1.36961	1.06161 and 1.13584*	----	----
Falls Creek	-0.97591	+0.33889	-0.17601	0.82388 and 0.46220*	0.33921	0.85846
Merced	-0.83044	+0.24541	-0.29145	0.91193 and 0.54904*	0.31754	0.14579
St. Maries	-0.39148	-0.18184	-0.13702	0.81831 and 0.40920*	0.64860	0.76772
Boise	-0.95637	+0.00686	+0.00199	0.94690, 0.05088 & 0.04140	0.08824	1.06502
McKenzie	-0.98834	+0.18620	-0.14287	0.95057 and 0.38768*	0.11867	0.78011

* Roots are complex conjugates.

CHAPTER V

DISCUSSION OF RESULTS

5.1 On the Reality of Short-Period Cycles. The reality of cycles corresponding to 4-month and 3-month periods as suggested in the spectra is not explainable in physical terms. It appears that since a single harmonic does not fit the correlograms in fig. 4, the fitting of a harmonic function to the time series needs sub-harmonics that do not necessarily occur in nature. It is quite possible that the emergence of short period cycles with the gradual removal of lower-ordered harmonics has been induced by the choice of a harmonic function to describe its behavior.

5.2 Number of Harmonics Used to Represent Periodic Component. While the higher harmonics may not be justifiable in physical terms, their use in representing the periodic component has to be considered in the light of the adequacy of fit.

It appears from Table 10 that it is not necessary to use all six harmonics to adequately describe the cyclic component. Thus, with the use of only those sub-harmonics which contribute significantly towards the explanation of variance, a reduction of parameters is effected in the equation,

$$y_t = \frac{\epsilon_t}{S_t} = \frac{X_t - P_t}{S_t} .$$

If n harmonics are used to describe P_t, $(2n+1)$ parameters will have to be estimated (i.e., $\overline{X}$, $A_1, \ldots, A_n$, $B_1, \ldots, B_n$). If m harmonics are used to describe S_t, other $(2m+1)$ parameters will be introduced so that $2(n+m+1)$ parameters will have to be estimated exclusive of the coefficients of the autoregressive scheme and the parameters characterizing the distribution of η_t.

5.3 General Applicability of Results. A primary object of this investigation is to seek for a general model--in this case, to determine the order of the autoregressive scheme which may be applied to the residual series. While the results of Tables 6 and 7 strongly support the second order autoregressive model, the cases of the Delaware and Jump Rivers present two exceptions. Tables 5 and 8 show that a first order scheme (with $\rho = r_1$) and a third order representation are also valid in addition to the second order scheme which has been found to be generally applicable.

Thorough inspection of Tables 5, 6, and 8 will reveal the heartening similarity in the parameters describing the three applicable schemes for the Delaware River. The pertinent information from the above tables is summarized in Table 9. In comparison with the generally applicable second order representation, the parameter ρ for the first order scheme is not much different from a_1 in the second order scheme. The slight change seems to be attributable to the reduction in the number of parame ters (i.e., from two to one). Similarly, in the third order autoregressive model, the addition of another parameter a_3 necessitated the adjustment of the values of a_1 and a_2 in the second order Markov model to allow for the effect of the third parameter.

One result shown in Table 9 is the progressive increase in explained variance with the increased order of the autoregressive scheme. The rate of increase seems to give the answer as to what scheme one may settle for. If the addition of another parameter (i.e., increasing the order of the scheme by one), is not accompanied by a significant increase in explained variance, then one may as well decide on the next lower ordered representation.

A summary of the results is shown in Table 10 wherein two schemes representing S_t and P_t are composed.

5.4 Difficulties in Statistical Inference. A recurring problem throughout this investigation was the difficulty met in the application of statistical tests where the volume of data was brought to bear on. In the test for significance of peaks in the spectrum, failure to use a Chi-square test suggested by Tukey may be justified since it was previously decided to sacrifice increased variability for better resolution.

In the application of Anderson's test for $\rho_1 = 0$, however, despite a value of $r_1 = 0.05$, the size of the statistical sample was such that even this was, according to the test, still significantly different from zero. The same difficulty was encountered in applying Quenouille's test to the first and second order autoregressive schemes.

In view of this, all inferences were based on the agreement of the computed variance of the residual series with the theoretical variance of residuals if the representation was valid. In all cases where the representations were accepted, these variances differed only in the third decimal place.

TABLE 9 COMPARISON OF PARAMETERS OF 3 POSSIBLE SCHEMES FOR DELAWARE AND JUMP RIVERS

River	Parameter	Order of autoregressive scheme 1	2	3
Delaware	a_1	-0.58911	-0.65798	-0.67158
	a_2	---	+0.11690	+0.19351
	a_3	---	----	-0.11645
	Explained Variance	0.347	0.356	0.357
Jump	a_1	-0.67083	-0.58300	-0.58630
	a_2	---	-0.13094	-0.14566
	a_3	---	---	+0.02525
	Explained Variance	0.450	0.460	0.461

TABLE 10 VARIANCES EXPLAINED BY STOCHASTIC MODELS

River	P_t and S_t Fitted with Harmonics in Table 4				P_t and S_t Each Fitted with 6 Harmonics			
	Periodic Component	Autoregressive Component	Total	No. of Parameters Used*	Periodic Component	Autoregressive Component	Total	No. of Parameters Used*
Batten Kill	0.2778	0.4564	0.7342	22	0.2904	0.4627	0.7531	28
Tioga	0.1547	0.2970	0.4515	22	0.1632	0.3033	0.4665	28
Cowpasture	0.1033	0.3940	0.4973	18	0.1051	0.3912	0.4964	28
Greenbrier	0.1483	0.4222	0.5705	10	0.1560	0.4121	0.5681	28
Mad	0.0769	0.3747	0.4516	16	0.0812	0.3811	0.4623	28
Powell	0.1734	0.6021	0.7755	10	0.1897	0.5670	0.7567	28
Oconto	0.3504	0.4358	0.7862	24	0.3525	0.4514	0.8039	28
Jump	0.1796	0.2769	0.4565	24	0.1813	0.2813	0.4626	28
Madison	0.5110	0.4294	0.9404	24	0.5115	0.4273	0.9388	28
Delaware	0.0140	0.3502	0.3642	18	0.0216	0.3170	0.3386	28
Current	0.0743	0.4696	0.5139	16	0.0828	0.4371	0.5199	28
Neches	0.1781	0.7773	0.9554	14	0.1969	0.7613	0.9582	28
Falls Creek	0.5382	0.3501	0.8433	20	0.5387	0.3035	0.8422	28
Merced	0.6064	0.2412	0.8476	20	0.6082	0.2556	0.8638	28
St. Maries	0.3703	0.2212	0.5915	16	0.3787	0.1338	0.5127	28
Boise	0.6710	0.3000	0.9710	16	0.6725	0.2699	0.9424	28
McKenzie	0.2086	0.6974	0.9060	12	0.2161	0.6708	0.8869	28

* Exclusive of parameters describing residual series η_t

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

After consideration of the foregoing results and uncertainties, the writer ventures the following conclusions:

1. In detecting periodicity in a hydrologic time series, spectral method complements autocorrelation analysis. If the periodic component is a composed harmonic, the sub-harmonics are easily identified by the characteristic peaks that they induce in the spectrum, but if the component is not a composed harmonic, the fundamental cycle confounds all other harmonics in the correlogram. It may, however, be possible that the secondary peaks are due to the intrinsic property of the harmonic function used.

2. In general, no other periodicities are perceptible in the time series of daily river runoff except that corresponding to the annual astronomic cycle and its sub-harmonics.

3. In removing the periodic component from a time series, the amount of variance explained by each harmonic may be used as a criterion in determining how many harmonics are necessary to compose the periodic component.

4. For a trend-free time series of daily river flows, after the deterministic periodic function has been sufficiently removed, the residual series may in general be represented by a second order autoregressive scheme. The inverse relation, corresponding to eq. 4.3 and applied to eq. 4.7, yields:

$$X_t + a_1 \frac{S_t}{S_{t-1}} X_{t-1} + a_2 \frac{S_t}{S_{t-1}} X_{t-2} = P_t + a_1 \frac{S_t}{S_{t-1}} P_{t-1} + + a_2 \frac{S_t}{S_{t-2}} P_{t-2} + S_t \eta_t$$

where a_1 and a_2 are the parameters of the second order autoregressive scheme, P_t is a periodic function representative of the periodic component of the series, and S_t is another periodic function used to describe the series of standard deviations.

5. Where the first or third order autoregressive schemes are applicable, the variances explained by these schemes do not differ appreciably from the explained variance obtained by the second order autoregressive representation.

In view of the difficulties encountered, two major areas where more intensive study would serve to shed light on certain aspects of this investigation are:

1. Determination of how the volume of data could be reduced and yet still give consistent results for statistical inference. This might be approached in the light of obtaining an optimum sampling time interval that would yield maximum information.

2. Analysis of time series like that of the Delaware River, which is amenable to several autoregressive representations. This might involve looking into the basin's physical characteristics and would be a significant departure from the purely statistical approach, serving to bridge the two seemingly divergent viewpoints in looking at the same problem.

BIBLIOGRAPHY

1. Anderson, R. L. 1941, Distribution of the serial correlation coefficient. Annals of Mathematical Statistics, Vol. 13, pp. 1-13.

2. Bartels, J., and S. Chapman, 1962, Geomagnetism. Vol. 2. Oxford, Clarendon Press, pp. 545-605.

3. Bartlett, M. S., 1946, On the theoretical specification and sampling properties of autocorrelated time series. Suppl. Journal Royal Statistical Society, Vol. 8, pp. 27-41.

4. Bartlett, M. S., 1950, Periodogram analysis and continuous spectra. Biometrika, Vol. 37, pp. 1-16.

5. Bartlett, M. S., and J. Mehdi, 1955, On the efficiency of procedures for smoothing periodogram of time series of continuous spectra. Biometrika, Vol. 42, pp. 143-150.

6. Blackman, R. B., and J. W. Tukey, 1958, The measurement of power spectra from the point of view of communications engineering, New York, Dover Publications Inc., 190 p.

7. Cramer, H., 1940, On the theory of stationary random processes. Annals of Mathematics, Vol. 41, pp. 215-230.

8. Doob, J. L., 1953, Stochastic processes. New York, Wiley and Sons, 654 p.

9. Granger, C. W. J., and M. Hatanaka, 1964, The spectral analysis of economic time series. Princeton, University Press, 299 p.

10. Grenander, U., and M. Rosenblatt, 1957, Statistical analysis of stationary time series. New York, Wiley and Sons, 300 p.

11. Hannan, E. J., 1960, Time series analysis. London, Methuen, 152 p.

12. Jenkins, G. M., 1961, General consideration in the analysis of spectra. Technometrics, Vol. 3, pp. 133-166.

13. Jenkins, G. M., and M. B. Priestley, 1957, The spectral analysis of time series. Journal Royal Statistical Society B., Vol. 19, pp. 1-12.

14. Julian, P. R., 1961, A study of the statistical predictability of stream runoff in the upper Colorado River Basin. Research paper of the High Altitude Observatory. University of Colorado. Boulder, Colorado, 98 p.

15. Kendall, M. G., 1944, On autoregressive time series. Biometrika, Vol. 33, pp. 105-122.

16. Kendall, M. G., 1948, The advanced theory of statistics. Vol. 2. London, Charles Griffin and Co., 521 p.

17. Quenouille, M. H., 1949, A large sample test for the goodness of fit in autoregressive schemes. Journal Royal Statistical Society, Vol. 110, pp. 123-129.

18. Quimpo, R. G., 1966, Stochastic analysis of daily river flows. Ph. D. Dissertation, Colorado State University, 96 p.

19. Slutsky, E., 1937, The summation of random causes as a source of cyclic process. Econometrica, Vol. 5, pp. 105-146.

20. Tukey, J. W., 1959, An introduction to the measurement of spectra. Probability and statistics, Edited by U. Grenander. Stockholm, Almquist and Widsell, pp. 300-330.

21. Wold, H., 1954, A study in the analysis of stationary time series. Uppsala, Almquist and Wiksell, 236 p.

22. Yaglom, A. M., 1962, An introduction to the theory of stationary random functions. Englewood Cliffs, Prentice Hall, 223 p.

23. Yevdjevich, V. M., 1961, Some general aspects of fluctuations of annual runoff in the Upper Colorado River Basin. Colorado State University, Fort Collins, CER61VMY54.

24. Yule, G. U., 1962, On a method of investigating periodicities in distrubed series with special reference to Wolfer's sunspot numbers. Philosophical Transactions, Royal Society A. Vol. 226, pp. 267-298.

Key Words: Hydrology, Time Series, Daily River Flows, Autoregressive Schemes, Spectral Analysis, Stochastic Model.

Abstract: A search for a structural model for the time series of daily river flows is undertaken by the author. First, records of daily river runoff from 17 river basins chosen on the postulated absence of trends induced by manmade improvements are analyzed. As a result, the model envisaged is a superposition of a cyclic deterministic process and a stochastic component. In the analysis of records, spectral methods are used to detect cycles which are then removed by subtracting from the original series a periodic function obtained by harmonic analysis. To alleviate the effect of a changing variance during the course of the year, the series of standard deviations is simularly fitted with a harmonic function which is used to standardize the series. After standardization, all the residual series are found to satisfy the second order autoregressive representation:

$$Z_t + a_1 Z_{t-1} + a_2 Z_{t-2} = \eta_t$$

where a_1 and a_2 are the autoregressive coefficients and η_t is an independently distributed random variable. The adequacy of fit is judged on the agreement between the theoretical and explained variances.

Key Words: Hydrology, Time Series, Daily River Flows, Autoregressive Schemes, Spectral Analysis, Stochastic Model.

Abstract: A search for a structural model for the time series of daily river flows is undertaken by the author. First, records of daily river runoff from 17 river basins chosen on the postulated absence of trends induced by manmade improvements are analyzed. As a result, the model envisaged is a superposition of a cyclic deterministic process and a stochastic component. In the analysis of records, spectral methods are used to detect cycles which are then removed by subtracting from the original series a periodic function obtained by harmonic analysis. To alleviate the effect of a changing variance during the course of the year, the series of standard deviations is simularly fitted with a harmonic function which is used to standardize the series. After standardization, all the residual series are found to satisfy the second order autoregressive representation:

$$Z_t + a_1 Z_{t-1} + a_2 Z_{t-2} = \eta_t$$

where a_1 and a_2 are the autoregressive coefficients and η_t is an independently distributed random variable. The adequacy of fit is judged on the agreement between the theoretical and explained variances.

Key Words: Hydrology, Time Series, Daily River Flows, Autoregressive Schemes, Spectral Analysis, Stochastic Model.

Abstract: A search for a structural model for the time series of daily river flows is undertaken by the author. First, records of daily river runoff from 17 river basins chosen on the postulated absence of trends induced by manmade improvements are analyzed. As a result, the model envisaged is a superposition of a cyclic deterministic process and a stochastic component. In the analysis of records, spectral methods are used to detect cycles which are then removed by subtracting from the original series a periodic function obtained by harmonic analysis. To alleviate the effect of a changing variance during the course of the year, the series of standard deviations is simularly fitted with a harmonic function which is used to standardize the series. After standardization, all the residual series are found to satisfy the second order autoregressive representation:

$$Z_t + a_1 Z_{t-1} + a_2 Z_{t-2} = \eta_t$$

where a_1 and a_2 are the autoregressive coefficients and η_t is an independently distributed random variable. The adequacy of fit is judged on the agreement between the theoretical and explained variances.

Key Words: Hydrology, Time Series, Daily River Flows, Autoregressive Schemes, Spectral Analysis, Stochastic Model.

Abstract: A search for a structural model for the time series of daily river flows is undertaken by the author. First, records of daily river runoff from 17 river basins chosen on the postulated absence of trends induced by manmade improvements are analyzed. As a result, the model envisaged is a superposition of a cyclic deterministic process and a stochastic component. In the analysis of records, spectral methods are used to detect cycles which are then removed by subtracting from the original series a periodic function obtained by harmonic analysis. To alleviate the effect of a changing variance during the course of the year, the series of standard deviations is simularly fitted with a harmonic function which is used to standardize the series. After standardization, all the residual series are found to satisfy the second order autoregressive representation:

$$Z_t + a_1 Z_{t-1} + a_2 Z_{t-2} = \eta_t$$

where a_1 and a_2 are the autoregressive coefficients and η_t is an independently distributed random variable. The adequacy of fit is judged on the agreement between the theoretical and explained variances.