EFFECT OF MISESTIMATING HARMONICS IN PERIODIC HYDROLOGIC PARAMETERS

by

K. L. Bullard, V. Yevjevich
and
N. Kottegoda

November 1976

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

88

EFFECT OF MISESTIMATING HARMONICS IN PERIODIC HYDROLOGIC PARAMETERS

by

K. L. Bullard, V. Yevjevich and N. Kottegoda

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523

November 1976 No. 88

TABLE OF CONTENTS

Chapter		Page
	ACKNOWLEDGMENTS	iv
	ABSTRACT	iv
I	INTRODUCTION	1
	1-1 Hydrologic Simulation	1
	1-2 Study Objectives	1
	1-3 Study Approach	1
II	MODELING TECHNIQUE AND ITS APPLICATIONS	2
	2 1 Brief Overview of Time Series Analysis	2
	2-2 The Periodic-Stochastic Model	2
	2-3 Past Applications of the Periodic-Stochastic Model to Hydrologic Time Series	4
III	PRACTICAL CONSIDERATIONS IN ANALYSIS OF PERIODIC TIME SERIES	6
	3-1 Approximate or Temporary Stationarity	6
	3-2 Error Propagation	6
	3-3 Effect of Unremoved Harmonics on Autoregressive Models	6
	3-4 Distribution of Independent Residuals	8
IV	INVESTIGATION BY USING THE EXPERIMENTAL METHOD OF GENERATED SAMPLES	10
	4-1 Method of Analysis	10
	4-2 Series with Overremoved and Underremoved Harmonic Terms	11
V	EFFECT OF REMAINING HARMONICS AND AUTOREGRESSIVE COEFFICIENTS ON FREQUENCY DISTRIBUTIONS OF RESIDUALS	13
	5-1 Methodology	13
	5-2 Series with Harmonics Remaining in the Mean	13
	5-3 Series with some Harmonics Unremoved in both the Mean and the Standard Deviation	14
	5-4 Discussion of Experimental Results	18
VI	APPLICATIONS OF EXPERIMENTAL METHOD IN USING HARMONIC COEFFICIENTS FROM HISTORICAL DATA, NORMAL AND GAMMA NOISE AND EFFECT OF PERIODIC AUTOCORRELATION	22
	6-1 Methodology	22
	6-2 Periodicities in the Mean and Standard Deviation	22
	6-3 Periodicities in the Autoregressive Coefficients	22
VII	CONCLUSIONS	24
	REFERENCES	25

ACKNOWLEDGMENT

The financial support received at Colorado State University from the U.S. National Science Foundation, under the Grant ENG-74-17396, "Stochastic Water Resources Processes," with Dr. Vujica Yevjevich as the principal investigator, is acknowledged. The M.S. thesis by K. L. Bullard served as the basic material from which this paper is shaped by additional investigations.

ABSTRACT

The periodic-stochastic model, used to describe the structure of historic hydrologic time series, is examined in the light of finding that it distorts the distribution of residuals, or of the time-independent stochastic component under certain circumstances. In general, this distortion caused the distribution of an independent stochastic component to have a sharper peak, and sometimes it appears to follow a double branch exponential distribution function. The apparent cause for distortions is the failure of the least-squares method to accurately estimate amplitudes and phases of harmonics in periodic parameters of historic data. The unremoved and/or misestimated dependence of the autoregressive type in stochastic component was found to also affect the inferred distribution of residuals. The investigation method used is by generating new samples of given sizes and properties by the classical Monte Carlo method.

Chapter I
INTRODUCTION

1-1 Hydrologic Simulation

Simulation of new samples of streamflow, precipitation and other hydrologic stochastic processes has become beneficial for water resource planners, often used in conjunction with design or policy decisions in water utilization. The possible courses of action may be evaluated in a more thorough, statistically acceptable manner than is possible using only the observed hydrologic data as inputs and outputs. The usefulness of generated samples is limited by the statistical accuracy of inferred models and estimated parameters from the observed data. In defining mathematical models for such simulation the analyst must retain statistical properties of the original data, with some of the more obvious being: (1) The overall mean, variance, and parameters related to higher-order moments of the historical records; (2) Trends and periodicities displayed by the original data; (3) Serial correlation properties; and (4) The distribution of the independent stochastic component as the noise underlying the composed process. If all of these statistical characteristics of the original data could be retained in generated samples, then the water resources planner would have a useful tool for evaluating different courses of action in water resources planning and management.

Past attempts to model such hydrologic processes as streamflow and precipitation have met with varied success. In particular, some attempts to model daily series of precipitation and runoff have not in general been able to retain reasonably well the distribution of inferred independent stochastic components in the modeling process. Unusual departures of distributions of independent stochastic components by the modeling process in comparison with the inferred distribution, led some investigators to question the completeness of the models, and to search for sources of errors in modeling techniques.

1-2 Study Objectives

The purpose of this study is to investigate the simulation of periodic-stochastic processes as applied to daily series, and to examine the sources of errors involved in the modeling process. By examining sources of such errors, reasons for arriving at distorted distributions by the modeling process should be evident. The results are expected to lead to remedies and an improved modeling technique.

1-3 Study Approach

Since statistical characteristics of daily series of precipitation and streamflow are not easily expressed in mathematical terms, it is necessary to simulate samples of complex hydrologic processes in a computer. One can start with the assumption that the true population properties of a series are known prior to simulation. The periodic-stochastic structural analysis is then applied to generated samples, and differences determined between the properties that are incorporated and the estimated statistical properties of generated samples. By varying the complexity of a population process the sources of errors in the technique of structural analysis may become apparent and could be quantified.

Chapter II
MODELING TECHNIQUE AND ITS APPLICATION

2-1 Brief Overview of Time Series Analysis

In general, currently known techniques for modeling a given historic time series may be divided into three basic phases: (1) Identification and removal of trends, periodicities or almost-periodic long-term movements in parameters of the process in order to isolate an approximate second-order or higher-order stationary time dependent process; (2) Identification and removal of time dependence of inferred second-order or higher-order stationary process in order to obtain an inferred independent, second-order or higher-order stationary stochastic residual process; and (3) Analysis of the probability distribution of the random residual variables, which are assumed to be independent and identically distributed, that constitute the resulting stationary stochastic process.

A temporarily stationary time series may be defined as one in which the statistical properties of the series do not vary with time for sufficient periods of the immediate past and future. Because it is extremely difficult to prove the stationarity either for the historical series of random variables or for unobserved past and future periods, a less rigid definition of stationarity is often used in practice. A series X_t is said to be weakly stationary, stationary in a wide sense, or to possess second-order stationary if its expected mean at any time position is equal to a constant and its autocorrelation function $\rho_x(\ell)$ is a function of a finite time difference $\ell = \left| t_i - t_{i+\ell} \right|$ only. The concept of stationarity is important in examining the effects of unremoved periodicities between the above first and second phases of analysis of a time series.

2-2 The Periodic-Stochastic Model

Several modeling techniques are extensively discussed in various publications and dissertations [8], [12], [16], [23]. Since it is the stated purpose of this study to review one such technique, the periodic-stochastic model, the equations and assumptions are briefly summarized below. For a detailed discussion of the technique see Yevjevich [23].

The model being considered is parametric, with amplitudes as constants for the inferred significant harmonics in periodic parameters. The basic form of the periodic portion of this model requires that each periodic hydrologic parameter in the model be modeled as a sum of several independent sine and cosine functions. Thus, if ν_τ is the value of any periodic parameter (such as the mean, variance, skewness, etc.) at time τ within a fundamental period ω of a given hydrologic series to be modeled and mathematically described, then

$$\nu_\tau = \bar{\nu} + \sum_{j=1}^{l} C_j \cos\left(\frac{2\pi j\tau}{\omega} + \theta_j\right) . \qquad (2\text{-}1)$$

A mathematically equivalent form of Eq. (2-1) is

$$\nu_\tau = \bar{\nu} + \left[\sum_{j=1}^{l} A_j \cos\left(\frac{2\pi j\tau}{\omega}\right) + B_j \sin\left(\frac{2\pi j\tau}{\omega}\right) \right] \qquad (2\text{-}2)$$

where $\bar{\nu} = \frac{1}{\omega} \sum_{t=1}^{\omega} \nu_\tau$, the mean of ν_τ series, (2-3)

with ν_τ the parameter values along the period ω, and l = number of significant harmonics.

The Fourier coefficients A_j and B_j of significant harmonics, are obtained from the estimates V_τ of ν_τ by using the least squares estimation method:

$$\hat{A}_j = \frac{2}{\omega} \sum_{t=1}^{\omega} V_\tau \cos\left(\frac{2\pi j\tau}{\omega}\right) , \qquad (2\text{-}4)$$

and

$$\hat{B}_j = \frac{2}{\omega} \sum_{t=1}^{\omega} V_\tau \sin\left(\frac{2\pi j\tau}{\omega}\right) . \qquad (2\text{-}5)$$

The amplitude C_j and the phase angle θ_j of Eq. (2-1) are related to the coefficients A_j and B_j by

$$C_j = \sqrt{A_j^2 + B_j^2} \qquad (2\text{-}6)$$

and

$$\theta_j = \tan^{-1}\left(\frac{A_j}{B_j}\right) . \qquad (2\text{-}7)$$

The major problem encountered at this point in the modeling process of periodic parameters is determining the number of significant harmonics l to be included in Eq. (2-1). Analytical methods of infering the value of l have been suggested by Schuster [19], Walker [21] and Fisher [9]. All of these have limitations when applied to hydrological data. Yevjevich [24] discusses the following practical method. Let $S^2(V_\tau)$ be the estimate of the variance of the ω estimates V_τ of ν_τ, and let the variance of the j-th harmonic be

$$\text{Var } h_j = (A_j{}^2 + B_j{}^2)/2. \qquad (2\text{-}8)$$

The amount of the variance $S^2(V_\tau)$ explained by the j-th harmonic is estimated by the ratio

$$\Delta P_j = \frac{\text{Var } \hat{h}_j}{S^2(V_\tau)} = \frac{(\hat{A}_j^2 + \hat{B}_j^2)}{2 \quad S^2(V_\tau)} \qquad (2\text{-}9)$$

with

$$S^2(V_\tau) = \frac{1}{\omega-1} \sum_{\tau=1}^{\omega} (V_\tau - \bar{V})^2 \qquad (2\text{-}10)$$

where $\bar{V} = \frac{1}{\omega} \sum_{\tau=1}^{\omega} V_\tau$. (2-11)

Once the portion of explained variance by each harmonic is estimated, those harmonics which are significant may be chosen by setting a significance level based on experience and choosing as significant those harmonics whose explained variance exceeds the chosen significance level. A significance level of 0.01 or 0.02 may be chosen for this purpose. Previous studies [16] have indicated that for hydrologic parameters of daily time series, such as mean, standard deviation, and autocorrelation coefficients, that harmonics computed beyond the first six are rarely significant.

Another empirical method is suggested by Yevjevich [23] using two criteria, P_{min} and P_{max} as given below, for the inference of l, with

$$P_{min} = k[\omega/cn]^{1/2} \tag{2-12}$$

$$P_{max} = 1 - P_{min} \tag{2-13}$$

where k = a constant, n = the number of years of record, c = the highest moment used in computing a parameter and ω = the period. A value of $k = 0.033$ may be initially chosen but it could be changed to suit the particular application. Using Eq. (2-9), if $\sum_{j=1}^{6} \Delta P_j < P_{min}$ the process is not considered to be periodic. If this inequality is reversed, then l harmonics are considered to be significant where $l = 1,2,\ldots,6$, is the minimum number of harmonics which satisfies $\sum_{j=1}^{l} \Delta P_j > P_{max}$. A value of $l = 6$ is chosen in case the second inequality is not satisfied for $l = 6$.

The model as applied by Quimpo [16] and Tao [20] to series of daily river flows considers as periodic parameters the mean μ_τ and the standard deviation σ_τ, thus the model for daily river flow series $Q_{p,\tau}$ becomes

$$Q_{p,\tau} = \mu_\tau + \sigma_\tau \, \varepsilon_{p,\tau} , \tag{2-14}$$

where p represents the year, τ represents the day within a year and $\varepsilon_{p,\tau}$ = the approximate second-order stationary, stochastic dependent series. The series $\hat{\varepsilon}_{p,\tau}$ is obtained as follows by using $\hat{\mu}_\tau$ and $\hat{\sigma}_\tau$, the periodic means and standard deviations evaluated through inferred numbers and estimated parameters of harmonics

$$Y_{p,\tau} = \frac{Q_{p,\tau} - \hat{\mu}_\tau}{\hat{\sigma}_\tau} . \tag{2-15}$$

The $Y_{p,\tau}$ series is then standardized to zero mean and unit variance, to become the $\hat{\varepsilon}_{p,\tau}$ series,

$$\hat{\varepsilon}_{p,\tau} = \frac{Y_{p,\tau} - \overline{Y}}{S(Y)} , \tag{2-16}$$

where $\overline{Y}$ = the mean of $Y_{p,\tau}$ series, and $S(Y)$ = its standard deviation. With the inferred second-order stationary series $\hat{\varepsilon}_{p,\tau}$, the analysis of time dependence begins. Common dependence models for stationary time series are the autoregressive schemes, moving average schemes, and mixed autoregressive-moving average schemes (ARMA). Box and Jenkins [4] and Jenkins and Watts [11] provide the complete analytical treatments for these modeling schemes. Yevjevich [22] has demonstrated some physical evidence for hydrologic dependent stationary series to be represented by autoregressive schemes. In this study the autoregressive models will be used exclusively. The general m-th order autoregressive model of a non-periodic process is given by

$$\varepsilon_i = \sum_{j=1}^{m} \alpha_j \, \varepsilon_{i-j} + \xi_i , \tag{2-17}$$

in which α_j = the autoregressive coefficient, m = the order of the model, and ξ_i = the assumed underlying independent stationary stochastic process.

For discrete time series used in this study, the population autocorrelation coefficient ρ_k for lag k is defined by

$$\rho_k = \frac{\mathrm{cov}(X_i, X_{i+k})}{[\mathrm{var}(X_i)\mathrm{var}(X_{i+k})]^{1/2}} = \frac{E(X_i X_{i+k}) - E(X_i)E(X_{i+k})}{[E(X_i^2)-(EX_i)^2]^{1/2}[E(X_{i+k}^2)-(EX_{i+k})^2]^{1/2}} \tag{2-18}$$

in which X_i and X_{i+k} are observations at times i and $i+k$, respectively, $\mathrm{cov}(X_i, X_{i+k})$ is the autocovariance function, and $\mathrm{var}(X_i)$ and $\mathrm{var}(X_{i+k})$ are variances at lag 0 and lag k, respectively.

For the open series approach, the ρ_k are estimated by the sample serial correlation coefficients, r_k, as follows

$$r_k = \frac{\frac{1}{N-k}\sum_{i=1}^{N-k} x_i x_{i+k} - \frac{1}{(N-k)^2}\left(\sum_{i=1}^{N-k} x_i\right)\left(\sum_{i=1}^{N-k} x_{i+k}\right)}{\left[\frac{1}{N-k}\sum_{i=1}^{N-k} x_i^2 - \frac{1}{(N-k)^2}\left(\sum_{i=1}^{N-k} x_i\right)^2\right]^{1/2}\left[\frac{1}{N-k}\sum_{i=1}^{N-k} x_{i+k}^2 - \frac{1}{(N-k)^2}\left(\sum_{i=1}^{N-k} x_{i+k}\right)^2\right]^{1/2}} \tag{2-19}$$

with N = the total number of sample observations. From estimates of serial correlation coefficients, r_k, it is possible to obtain estimates of autoregressive coefficients, α_j, of Eq. (2-17) for any order m by using the Yule-Walker equations [4]:

$$\begin{aligned}
\rho_1 &= \alpha_1 + \alpha_2 \rho_1 + \cdots + \alpha_m \rho_{m-1} \\
\rho_2 &= \alpha_1 \rho_1 + \alpha_2 + \cdots + \alpha_m \rho_{m-2} \\
&\vdots \\
\rho_m &= \alpha_1 \rho_{m-1} + \alpha_2 \rho_{m-2} + \cdots + \alpha_m
\end{aligned} \tag{2-20}$$

Since these equations are linear in α_j, $j = 1, 2, \ldots, m$, they may be solved efficiently for large values of m on a computer by Gaussian elimination technique [10]. For smaller values of m it is often more efficient to use a direct algebraic approach.

Estimates a_j of the autoregression coefficients α_j for the first three order models are:

First-order model, $m = 1$

$$a_1 = r_1 . \tag{2-21}$$

Second-order model, $m = 2$

$$a_1 = \frac{r_1 - r_1 r_2}{1 - r_1^2} , \tag{2-22}$$

and

$$a_2 = \frac{r_2 - r_1^2}{1 - r_1^2} \tag{2-23}$$

Third-order model, $m = 3$

$$a_1 = \frac{(1-r_1^2)(r_1-r_3) - (1-r_2)(r_1 r_2 - r_3)}{(1-r_2)(1-2r_1^2 + r_2)} , \tag{2-24}$$

and

$$a_2 = \frac{(1-r_2)(r_2 + r_2^2 - r_1^2 - r_1 r_3)}{(1-r_2)(1-2r_1^2 + r_2)} , \tag{2-25}$$

$$a_3 = \frac{(r_1-r_3)(r_1^2 - r_2) - (1-r_2)(r_1 r_2 - r_3)}{(1-r_2)(1-2r_1^2 + r_2)} . \tag{2-26}$$

The estimated autocorrelation coefficients may be considered to be periodic and expressed as $r_{k,\tau}$, $k=1,2,\ldots$, $\tau=1,2,\ldots,\omega$. They are estimated in the same way as in Eq. (2-19) except that N is replaced by n, the number of years of data if $\omega-\tau \geq k$ and by $n-1$ if $\omega-\tau < k$. If periodicities in the autocorrelation are significant, the autoregression coefficients are also significantly periodic, in which case the Yule-Walker equations, Eq. (2-20), are modified to suit, and solutions to the periodic coefficients are given by Salaz-la Cruz and Yevjevich [18].

Selecting the order of the autoregressive model is the first step, in order to insure that the residual series, ξ_i, is close to an independent series. A simplified method to accomplish the task of selecting the proper order is proposed by Yevjevich [23]. This method uses the coefficient of determination, R_j^2, as the criterion for selecting the order of the model. The coefficient of determination is the part of the variance of series ε_i, which is explained by the selected model. The remaining portion of the variance is that of the independent component, ξ_i. The estimates of coefficients of determination R_j^2 for linear autoregressive models are

$$R_1^2 = r_1^2 , \tag{2-27}$$

$$R_2^2 = \frac{r_1^2 + r_2^2 - 2r_1^2 r_2}{1 - r_1^2} , \tag{2-28}$$

$$R_3^2 = \frac{r_1^2+r_2^2+r_3^2+2r_1^3r_3+2r_1^2r_2^2+2r_1r_2^2r_3}{1-2r_1^2-r_2^2+2r_1^2r_2} - \frac{\left(2r_1^2r_2+4r_1r_2r_3+r_1^4+r_2^4+r_1^2r_3^2\right)}{1-2r_1^2-r_2^2+2r_1^2r_2} \tag{2-29}$$

The selection criteria are as follows:

For the first-order model

$$R_2^2 - R_1^2 \leq 0.01 , \quad \text{and} \quad R_3^2 - R_2^2 \leq 0.02, \tag{2-30}$$

For the second-order model

$$R_2^2 - R_1^2 > 0.01 , \quad \text{and} \quad R_3^2 - R_2^2 \leq 0.01, \tag{2-31}$$

and for the third-order model

$$R_2^2 - R_1^2 > 0.01 , \quad \text{and} \quad R_3^2 - R_2^2 > 0.01 . \tag{2-32}$$

Once the order of the autoregressive model is chosen, the independent stochastic series is computed by

$$\hat{\xi}_i = \frac{\hat{\varepsilon}_i - \sum_{j=1}^{m} a_j \hat{\varepsilon}_{i-j}}{\left[1 - \sum_{k=1}^{m} \sum_{j=1}^{m} a_k a_j r_{k-j}\right]^{1/2}} . \tag{2-33}$$

The $\hat{\xi}_i$ series is then tested for independence and if shown as such, is accepted as a second-order stationary independent stochastic process, assuming that autoregression coefficients are not periodic. Values of the $\hat{\xi}_i$ series are used to obtain the probability distribution function of the best fit to their empirical frequency distribution. Tao [20] reviews several possible probability distribution functions as distributions for $\hat{\xi}_i$ component.

2-3 Past Applications of the Periodic-Stochastic Model to Hydrologic Time Series

Applications of the periodic-stochastic modeling technique of time series in hydrology are found in the literature for such cases as riverflow, long-term climatic changes, etc. Adamowski and Smith [2] studied the technique as a means of simulating the daily rainfall over Bark Lake, Ontario. The model used was one in which only the harmonics in the mean were removed.

The correlogram of the residual series was found to have persistence at lag one with a periodic appearance for larger lags, though this part of the correlogram is well defined within the tolerance limits drawn for the independent series. A linear autoregressive model of the first order ($\alpha_1 = 0.097$) was chosen because the coefficient of determination is not significantly increased for higher-order models.

The distribution of the independent series $\hat{\xi}_i$ was not reported, but the conclusions of the study did indicate results other than what the authors expected. Some of the conclusions of the Bark Lake study were: (1) Several short-term periodicities appeared at the 5 to 6 day, 8 to 10 day and at 16 day intervals in the resulting independent series; and (2) The short-term periodicities, if they exist at all, are difficult to justify and model by existing generation techniques.

In studies on rainfall series Landsberg [14] found short-term periodicities in the calculated independent component of daily rainfall series at Woodstock, Maryland. Computed periodicities were of 3 day, 5 to 7 day, 15 days and 25 days length. Dickenson [6] showed that anticyclones located north of 60^oN latitude can block the normal flow of air over the earth and possibly be correlated to a 10 day or a lesser cycle in rainfall patterns. This anticyclone theory was in part discarded because it was of little value in explaining the calculated periodicities greater than 10 days and is of a questionable value for explaining the short-term periodicities of 3 days or less. The probability distributions of the assumed independent stochastic series $\hat{\xi}_i$ was not reported in any of these studies of the structure of daily rainfall series.

Roesner and Yevjevich [17] modeled the monthly precipitation and runoff series by using the periodic-stochastic model. It was found that the resulting correlograms of dependent series, after the removal of significant harmonics in the mean $\hat{\mu}_\tau$ and in the standard deviation $\hat{\sigma}_\tau$ of the series, were strongly periodic unless a large number of harmonic terms were removed. Also, the resulting dependent series for some runoff series may have periodicity in the autoregressive coefficients. Periodic autoregressive coefficients invalidate the assumption of the second-order stationarity of the series. Jones and Brelsford [12] found the autoregressive coefficients periodic in modeling several kinds of meteorological data. The periodicity in these coefficients is explained as being related to changing physical mechanisms from season to season. No description of the probability distribution of the inferred independent series was reported either by Roesner and Yevjevich or by Jones.

In the area of streamflows, Quimpo [16] applied the periodic-stochastic modeling to 17 daily river flow series, which had no man-made disturbances in records. Some of the major conclusions of that study are as follows: (1) Both the estimated daily means and the daily standard deviations of river flow have the annual periodicity; (2) The estimated correlation coefficients $r_{k,t}$ are often shown to be periodic; and (3) The resulting dependent stochastic process, after removing the estimated periodic functions of the mean $\hat{\mu}_\tau$, the standard deviation $\hat{\sigma}_\tau$, and the autocorrelation coefficients $\alpha_{j,\tau}$, could in general be acceptable as stationary of the second-order autoregressive models.

Tao [20] used the same model and the same 17 daily river series for investigating the distributions of the resulting independent series, $\hat{\xi}_i$. The major conclusions of his thesis are: (1) Errors in estimating the number and amplitudes of significant harmonics in the mean and standard deviation greatly affect the accuracy of the resulting dependent stochastic series, $\hat{\varepsilon}_i$; (2) Tails of distributions of independent $\hat{\xi}_i$ series are generally long, found to be well approximated by the simple exponential functions; and (3) Tested probability distribution function did not fit well the frequency distributions of $\hat{\xi}_i$; none passed the chi-square test of goodness of fit; among the several distribution functions tested, the smallest chi-square values were for the double-branch gamma distribution, exemplified by a sharp peak at the mode, high kurtosis, and a long exponential tail.

Chin [5] used the periodic-stochastic model to describe the long-term climatic changes. He found an indication of the overall temperatures of the earth for many thousands of years in the long series of oxygen isotope data taken from deep sea sediment cores. In this study four separate series were modeled by deterministic-stochastic methods. In all but the longest series, almost periodic terms were replaced by nonlinear or linear trends. In the longest oxygen isotope data series, which extends approximately 126,000 years into the past, the long-term almost-periodic effects due to various astronomical cycles, first described by Milankovich, were modeled and removed. The independent residuals found in the three series after the trend was removed were found to be approximately normally distributed, while the one series which was modeled with an inferred almost-periodic component showed a sharp peak in the distribution of the independent residual series; it was fitted by a double branch gamma distribution.

Chapter III
PRACTICAL CONSIDERATIONS IN ANALYSIS OF PERIODIC TIME SERIES

3-1 Approximate or Temporary Stationarity

It has been found necessary for most structural models of hydrologic time series to remove the periodicity in at least the mean and standard deviation. The deterministic portion of the periodic-stochastic process should describe periodicities which exist in nature. After the periodic parameters of historic data have been inferred and removed, the residual series, Eq. (2-16), is assumed to be stationary up to the highest order parameters which are shown to be periodic. This assumption is true only if the number, amplitudes and phases of the inferred harmonics in periodic parameters are equal to the true state of nature which is unknown. If the number or amplitudes of the removed harmonics differ from the true state of nature, the resulting residual series will in fact contain some harmonics in its parameters. This will produce errors (which have been apparently ignored in the past) in any further analysis of the series, since the analysis of the residual dependent series assumes its stationarity. Errors resulting from the estimation of numbers and parameters of harmonics in the residual dependent series are explored in this chapter.

3-2 Error Propagation

Tao [20] noted in his dissertation that errors involved in estimating the harmonic components and the order of the autoregressive model had the potential of seriously altering the respective residual series produced at each stage of the analysis. In his analysis the errors were assumed to be so small as to be negligible. However, special attention is given to the effects of such errors in this study. For example of possible serious effects of errors in the estimate of harmonics in periodic parameters of the time series, see Tao [20].

Errors made in the autoregressive modeling of a true autoregressive process in general do not have such a profound effect on residuals as the errors in estimation of the number and the amplitudes of harmonics. If the dependent stochastic component includes the errors, it can be written as

$$\varepsilon_i \pm \Delta\varepsilon_i = e_\sigma \varepsilon_i + e_\mu \qquad (3\text{-}1)$$

If e_μ and e_σ represent errors made in estimating the harmonic terms in the mean μ_τ and the standard deviation series σ_τ, respectively, then it can be demonstrated that if e_σ and e_μ are constants the correlogram $\rho_\varepsilon(l)$ will not be affected. The real problem is when e_μ and e_σ are time dependent errors (or in other words they are periodic) and have a significant serial correlation structure.

The remaining aspect is the overestimation or underestimation of the autoregressive coefficients in the dependent stochastic process of the model. The question is one that pertains to mathematical statistics, but a few generalizations are offered here on various aspects of the problem. The estimates commonly used for autoregressive coefficients are often based in the estimates by serial correlation coefficients, $\rho_\varepsilon(k)$, of the particular realization of a stochastic process investigated. Depending on the computational formula used to estimate the serial correlation coefficients, the estimates of autoregressive coefficients will be either biased, less biased, or will be the minimum variance estimates [11]. In either case the estimates of autoregressive coefficients increase in accuracy as the length of the series of stochastic process being investigated increases. The real problem is not in overestimation or underestimation of autoregressive coefficients but in determining the number of coefficients to be estimated. Quenouille [15], Jenkins and Watts [11], and Yevjevich [23], suggest methods for estimating the order of the autoregressive model. As the number of autoregressive terms increases, they generally have a corrective influence on each other, and the actual overestimation or underestimation of autoregressive coefficients tend to decrease.

Relating errors in estimating either the number or the magnitude of autoregressive coefficients to errors which will be propagated into the distribution of independent stochastic components, a small effect is found. If the true distribution of the resulting independent stochastic process is normal, the dependent stochastic process ε_i, Eq. (2-17), represents a sum of normally distributed random variables, which is normal regardless of constant multipliers involved. Thus a change in the number or magnitude of autoregressive coefficients results only in a change of the variance of the normal distribution of the ξ_i variable, but the basic form of the distribution will stay the same.

3-3 Effect of Unremoved Harmonics on Autoregressive Models

A linear autoregressive equation may be represented by

$$\varepsilon_i = \sum_{j=1}^{m} \alpha_j \varepsilon_{i-j} + \xi_i \qquad (3\text{-}2)$$

with m = an unknown value to be determined, where α_j = the true autoregressive coefficients, and ξ_i = a normally distributed and uncorrelated variable, with unknown mean and variance. The problem in this approach is that the model may have harmonic terms left in the estimated residual series. This residual series is now designated by X_i.

The computed dependent series X_i may be represented as a function of remaining harmonics in the mean, μ_τ, the standard deviation, σ_τ, and the true autoregressive process ε_i, namely as

$$X_i = \mu_\tau + \sigma_\tau \varepsilon_i, \qquad (3\text{-}3)$$

where μ_τ and σ_τ have the form of Eq. (2-1) and ε_i is defined by Eq. (3-2). It should be assumed that the amplitude of any harmonic included in μ_τ or σ_τ is so small, that it would not be included as significant in any previous test for significance of harmonic terms. However, the sum of all unremoved or over-removed harmonic terms may explain a relatively large portion of the variance of the X_i series.

The first step in analyzing the serial dependence structure of the nonstationary X_i series is to find the expected correlogram of the X_i series, denoted by $\rho_X(k)$. In order to accomplish this task it is useful to simplify the X_i series and examine it in various stages. As a first approximation, it is assumed that the computed residual series X_i has the remaining harmonic terms in its mean only, with Eq. (3-3) then given in the form

$$X_i = \mu_\tau + \varepsilon_i \ . \tag{3-4}$$

It is assumed that μ_τ and ε_i are independent. The expected correlogram $\rho_X(k)$ for lag k, k = 1,2,..., of the X series can now be computed

$$\rho_X(k) = \frac{\text{Cov}\ (X_i,\ X_{i+k})}{\text{Var}\ (X_i)} \ . \tag{3-5}$$

and by substituting for μ_τ then

$$\rho_X(k) = \frac{\text{Cov}\ (\mu_\tau + \varepsilon_i,\ \mu_{\tau+k} + \varepsilon_{i+k})}{\text{Var}\ (\mu_\tau + \varepsilon_i)} \ , \tag{3-6}$$

$$\rho_X(k) = \frac{\text{Cov}\left\{\left[\sum_{j=1}^{m}\left(C_j\ \cos[\lambda_j'(\tau) + \theta_j]\right) + \varepsilon_i\right] \cdot \left[\sum_{j=1}^{m}\left(C_j\ \cos[\lambda_j'(\tau+k) + \theta_j]\right) + \varepsilon_{i+k}\right]\right\}}{\text{Var}\left\{\sum_{j=1}^{m}\left[C_j\ \cos[\lambda_j'(\tau) + \theta_j]\right] + \varepsilon_i\right\}} \tag{3-7}$$

where m = the number of harmonics, $\lambda_j' = 2\pi j/\omega$, k = the lag (in days), and ω = the length of fundamental harmonic (ω = 365 days for daily series), and τ = the point on the periodic time scale. Since all of the cosine terms in summations of Eq. (3-7) are orthogonal, they are mutually independent. Also, they are independent of the ε_i series. Thus, the covariance operator may be taken inside the summations resulting in the expression

$$\rho_X(k) = \frac{\sum_{j=1}^{m}\left\{\text{Cov}\left[C_j\ \cos[\lambda_j'(\tau) + \theta_j]\right],\ \left[C_j\ \cos\ (\lambda_j'(\tau+k) + \theta_j)\right]\right\} + \text{Cov}\ [\varepsilon_i,\ \varepsilon_{i+k}]}{\sum_{j=1}^{m} \text{Var}[C_j\ \cos\ (\lambda_j'(\tau) + \theta_j)] + \text{Var}\ \varepsilon_i} \tag{3-8}$$

Substitution can be made as

$$\text{Var}\ [C_j\ \cos\ (\lambda_j'\ (\tau) + \theta_j)] = C_j^2/2, \tag{3-9}$$

$$\text{Var}\ \ \varepsilon_i = \sigma_\varepsilon^2 \ , \tag{3-10}$$

and

$$\sum_{j=1}^{m} \text{Cov}[C_j\ \cos\ (\lambda_j'(\tau) + \theta_j),\ C_j\ \cos\ (\lambda_j'(\tau+k) + \theta_j)] =$$

$$\sum_{j=1}^{m} \frac{C_j^2}{2}\ \cos\ (\lambda_j'(k)) \tag{3-11}$$

By definition

$$\rho_\varepsilon(k) = \frac{\text{Cov}\ (\varepsilon_i,\ \varepsilon_{i+k})}{\text{Var}\ (\varepsilon_i)} \tag{3-12}$$

or

$$\text{Cov}\ (\varepsilon_i,\ \varepsilon_{i+k}) = \rho_\varepsilon(k) \cdot \text{Var}\ (\varepsilon_i) = \rho_\varepsilon(k) \cdot \sigma_\varepsilon^2 \tag{3-13}$$

By combining all of the above terms the following expression for the expected correlogram of the nonstationary X_i series is

$$\rho_X(k) = \frac{\sum_{j=1}^{m} C_j^2/2\ \cos\ [\lambda_j'(k)] + \rho_\varepsilon(k) \cdot \sigma_\varepsilon^2}{\sum_{j=1}^{m} (C_j^2/2) + \sigma_\varepsilon^2} \ . \tag{3-14}$$

A further simplifying assumption is made, that the time dependent series, ε_i of Eq. (3-2), is in fact of the order one, i.e.,

$$\varepsilon_i = \alpha_1\ \varepsilon_{i-1} + \xi_i \ . \tag{3-15}$$

Using this assumption expressions for $\rho_\varepsilon(k)$ and σ_ε^2 can be used [24] as follows:

$$\rho_\varepsilon(k) = (\alpha_1)^k \tag{3-16}$$

and

$$\sigma_\varepsilon^2 = (1 - \alpha_1^2)^{-1} \ . \tag{3-17}$$

For these expression substituted into Eq. (3-14) then

$$\rho_X(k) = \frac{\sum_{j=1}^{m} (C_j^2/2)\ \cos[\lambda_j'(k)] + \alpha_1^k(1-\alpha_1^2)^{-1}}{\sum_{j=1}^{m} (C_j^2/2) + (1-\alpha_1^2)^{-1}} \tag{3-18}$$

For small lags (k = 1,2,3), the sum of the terms cos $[\lambda_j'(k)]$ could be approximated by unity. With this assumption there are only two unknown quantities in the expected correlogram. These are $\sum C_j^2/2$ and α_1.

It is now possible to set up a system of simultaneous equations and solve for $\sum C_j^2$ and α_1, using only $\rho_X(1)$ and $\rho_X(2)$.

$$\rho_X(1) = \frac{\alpha_1(1-\alpha_1^2)^{-1} + \sum_{j=1}^{m} C_j^2/2}{(1-\alpha_1^2)^{-1} + \sum_{j=1}^{m} C_j^2/2} \tag{3-19}$$

and

$$\rho_X(2) = \frac{\alpha_1^2(1-\alpha_1^2)^{-1} + \sum_{j=1}^{m} C_j^2/2}{(1-\alpha_1^2)^{-1} + \sum_{j=1}^{m} C_j^2/2} \ . \tag{3-20}$$

Solving these two simultaneous equations for the two unknowns yields

$$\alpha_1 = \frac{\rho_X(1) - \rho_X(2)}{1 - \rho_X(1)} \quad (3\text{-}21)$$

and

$$\sum_{j=1}^{m} C_j^2/2 = \left(\frac{\rho_X(1) - \alpha_1}{1 - \rho_X(1)}\right) (1-\alpha_1^2)^{-1} \quad (3\text{-}22)$$

Thus, from the estimated correlogram of the X_i series it is possible to estimate α, the true lag one autoregressive coefficient, and $\sum_{j=1}^{m} C_j^2/2$, the variance of the remaining harmonic terms of the X_i series. Without such estimates, common practice is to assume that $\sum C_j^2/2 = 0$, for $\alpha_1 = \rho_X(1)$, which may lead to misleading results.

If the model used for investigation of the raw data has periodicity in both the mean and the standard deviation, the calculated residual series X_i may have periodicity in the mean and stancard deviation as given by Eq. (3-3). The influence of the term $\sigma_\tau \varepsilon_i$ is not simple to analyze mathematically. However, its presence must be accounted for in the modeling process. Experimental analysis by Jones and Brelsford [12] suggests that the time series with the periodic structure in the standard deviation may also have the periodic autoregressive coefficients. Both Jones [12] and Yevjevich [23] suggest similar modeling techniques to handle periodicity in the autoregressive coefficients, the influence of the periodic standard deviation in the calculated X_i series is presumably accounted for. Tao [20] used the periodic autoregressive coefficients. However, Chin [5] did not use the periodic autoregressive coefficients in the analysis of the almost-periodic series in his investigation. Because both Tao and Chin found the double branch gamma distribution as the best fits for the residual series, after removing the estimated time dependence of the X_i series, it suggests that the modeling of the periodic autoregressive coefficients may not entirely account for the presence of the remaining periodicity in the standard deviation of the X_i series. This subject is investigated in the experimental part of this study.

If the true nature of the ε_i series of Eq. (3-2) is of a high-order (two or greater), the analysis becomes more complex. Kendall and Stuart [13] give the following form of the expected correlogram of an autoregressive process of the order two

$$\varepsilon_i = \alpha_1 \varepsilon_{i-1} + \alpha_2 \varepsilon_{i-2} + \xi_i , \quad (3\text{-}23)$$

$$\rho_\varepsilon(k) = \frac{\alpha_2^{k/2} \sin (k\theta + \psi)}{\sin (\psi)} \quad (3\text{-}24)$$

where

$$\theta = \text{arc cos} \left(\frac{-\alpha_1}{2\sqrt{\alpha_2}}\right) \quad (3\text{-}25)$$

and

$$\tan \psi = \frac{1 + \alpha_2}{1 - \alpha_2} \tan \theta . \quad (3\text{-}26)$$

The variance of the ε_i series of the order two is [24]

$$\text{Var}(\varepsilon_i) = \sigma_\varepsilon^2 = \left[1 - \alpha_1^2 - \alpha_2^2 - 2\alpha_1 \alpha_2 \cdot \rho_\varepsilon(1)\right]^{-1} . \quad (3\text{-}27)$$

Theoretically, with the harmonics in the mean only, it should be possible to substitute Eqs. (3-24) through (3-27) into the expected correlogram of Eq. (3-24), and solve it for the unknown values of α_1, α_2, and $\sum C_j^2/2$ in terms of values of $\rho_X(1)$, $\rho_X(2)$ and $\rho_X(3)$, which can be estimated by Eq. (2-19). This computation is mathematically intractable, and past investigations have resorted to modeling the periodic autoregressive coefficients. It can be concluded at this moment that the X_i series has become complex, and investigators must resort to simplifying assumptions and techniques.

3-4 Distribution of Independent Residuals

If it is assumed here that the modeling procedure used for generating samples of the assumed independent series of Eq. (2-33) allows the errors in estimated harmonic terms of the mean (μ_τ) and the standard deviation (σ_τ) to be carried into the estimated series from the sample as the independent stochastic component, then the effects of the unremoved or overremoved harmonic terms on the distribution of the estimated independent series can be investigated. The series X_i represents now the residual series after the autoregressive dependence has been removed, or

$$X_i = \mu_\tau + \sigma_\tau \xi_i . \quad (3\text{-}28)$$

The difference between Eq. (3-28) and Eq. (3-23) is in the term ξ_i, which is assumed to be a series of uncorrelated, normally distributed random variables. The ideal statistical approach would be to mathematically describe the exact theoretical distribution of the X_i series of Eq. (3-28). Theoretical approaches such as the use of the characteristic functions or the convolution formulas [7] are available to describe distributions of sums and products of independent random variables. These approaches are mathematically complex and when combined with the complex expressions for μ_τ and σ_τ, they become intractable.

Most investigators resort to simpler solutions by fitting the residual series by several known distribution functions, and accept as the result that distribution which has the smallest least chi-square value. This simplifying procedure is used here also. The two probability density functions used to investigate distributions of independent residual series here are the normal and the double-branch exponential probability functions.

The Normal Probability Function. The probability density function of the normal distribution used is

$$f(\hat{\xi}) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-(\hat{\xi} - \mu)^2/2\sigma^2] , \quad (3\text{-}29)$$

in which μ is the expected value and σ standard deviation of ξ.

The maximum liklihood estimated of parameters of the normal density function are

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} \hat{\xi}_i \qquad (3\text{-}30)$$

and

$$\hat{\sigma} = \left[\frac{1}{N} \sum_{i=1}^{N} (\hat{\xi}_i - \hat{\mu}) \right]^{1/2} \qquad (3\text{-}31)$$

in which N is the sample size.

The Bilateral or Double-Branch Exponential Probability Density Function. The probability density function of the bilateral exponential distribution is

$$f(\hat{\xi}) = \frac{1}{2\beta} \exp[- \frac{1}{\beta} |\hat{\xi} - \mu|] \qquad (3\text{-}32)$$

in which μ is the expected value of the random variable ξ_i and $2\beta^2$ its variance.

If a value of either μ or β is obtained from an apriori knowledge of the process, then a maximum liklihood estimator of the remaining parameter can be derived analytically. If neither μ nor β is known, an empirical approach using the numerical methods may be required to maximize the likelihood function for estimating the two parameters. For the purpose of this study, it was assumed that an apriori knowledge for μ is available, with μ=0, because $\hat{\xi}_i$ is an independent random series derived from $\hat{\varepsilon}_i$ which has been standardized as a (0,1)-series. Thus, the maximum likelihood estimator of β is given by

$$\hat{\beta} = \frac{1}{N} \sum_{i=1}^{N} |\hat{\xi}_i| \quad . \qquad (3\text{-}33)$$

By simulating the X_i series of Eqs. (3-3) and (3-28) distributions, can be experimentally determined and compared to distributions found by Tao and Chin.

Chapter IV
INVESTIGATION BY USING THE EXPERIMENTAL METHOD OF GENERATED SAMPLES

4.1 Method of Analysis

The method of experimental analysis is through data simulation, using random numbers generated in a computer. Several samples of the form of Eq. (3-3) were generated using the Monte Carlo technique. Each sample was then analyzed using the various steps of the general periodic-stochastic model as described in Chapter II. The end result of each generation and analysis of samples is a frequency distribution curve of the assumed random independent variable denoted by $\hat{\xi}_i$. This frequency distribution represents the results of either an insufficient or excessive estimate of harmonic analysis of the periodic-stochastic process. The sequence of steps in the generation and analysis is:

Step 1: Generate ξ_i sample, $\xi_i \approx N(0,1)$, $i = 1, 2, \ldots, N$, with $N = 365\,n$, n = the number of years of simulated daily data.

Step 2: Generate the autoregressive process when needed,

$$\varepsilon_i = \sum_{j=1}^{m} \alpha_j e_{i-j} + \xi_i, \quad i = 1, 2, \ldots, N, \text{ with}$$

m = the order of autoregressive model, and α_j, $j = 1, 2, \ldots, m$, the estimated constants.

Step 3: Introduce periodic dependence as deviations in the mean and standard deviation, $\mu_\tau = f_\mu(\tau)$, $\sigma_\tau = f_\sigma(\tau)$, with $f_\mu(\tau)$, $f_\sigma(\tau)$ the periodic functions of any form, with the exact equations as inputs.

Step 4: Generate (daily) series with errors, $X_i = \mu_\tau + \sigma_\tau\, \varepsilon_i$, $i = 1, 2, \ldots, N$.

Step 5. Estimate experimentally periodic functions of mean and standard deviation, $\hat{\mu}_\tau = \hat{f}_\mu(\tau)$, $\hat{\sigma}_\tau = \hat{f}_\sigma(\tau)$, with $\hat{f}_\mu(\tau)$, and $\hat{f}_\sigma(\tau)$ estimated by Fourier analysis, with number of harmonic terms m'.

Step 6. Remove estimated periodic parameters from X_i series by

$$y_i = \frac{X_i - \hat{\mu}_\tau}{\hat{\sigma}_\tau}, \quad i = 1, 2, \ldots, N, \text{ and}$$

$$\hat{\varepsilon}_i = \frac{y_i - \overline{y}_i}{\hat{\sigma}_y}, \quad i = 1, 2, \ldots, N, \text{ with } \overline{y}_i = \text{the mean}$$

of y_i, $\hat{\sigma}_y$ = the standard deviation of y_i, and $\overline{\varepsilon} = 0$, and $S(\varepsilon) = 1$.

Step 7. Estimate α_j coefficients and remove the dependence of autoregressive process from $\hat{\varepsilon}_i$ by

$$\hat{\xi}_i = \hat{\varepsilon}_i - \sum_{j=1}^{m} \alpha_j\, \hat{\varepsilon}_{i-j}, \text{ with } m' \text{ and } m \text{ input constants}$$

α_j, $j = 1, 2, \ldots, m'$, by using Eqs. (2-21) through (2-26).

Step 8. Estimate frequency distributions of $\hat{\xi}_i$, and estimate mean, variance, skewness, kurtosis, mode, and chi-square in test for normal distribution (0,1).

Each simulated sample corresponds to approximately ten years of daily time series data (365 x 10 = 3650 values). The mode, skewness and kurtosis of the frequency distribution of $\hat{\xi}_i$ were initially computed. The mean and variance of the frequency distribution and of the fitted normal function are 0 and 1, respectively, since the sample was standardized in the modeling process.

It was desirable to remove sampling autocorrelation by the computer in the generated random numbers, ξ_i, in each sample. A computer subroutine was used to obtain several samples of initially standard normal but uncorrelated random variables. Distributions of ξ_i were then compared to the standardized normal distribution by the chi-square goodness of fit test. A check for independence of the ξ_i variable was made by computing and plotting the correlogram for each sample generated. The tolerance limits given by Anderson [3] for correlograms of normal independent random variables were computed by

$$\frac{-1 - t_\alpha \sqrt{N-2}}{N-1} < \rho_x(k) < \frac{-1 + t_\alpha \sqrt{N-2}}{N-1}, \qquad (4\text{-}1)$$

in which t_α is the normal deviate corresponding to the probability α of rejection of a hypothesis, and N is the number of observations. For the first stage of this study, N was fixed at 3650. These tolerance limits are plotted on correlograms of generated standard normal ξ_i variables, Fig. 4-1.

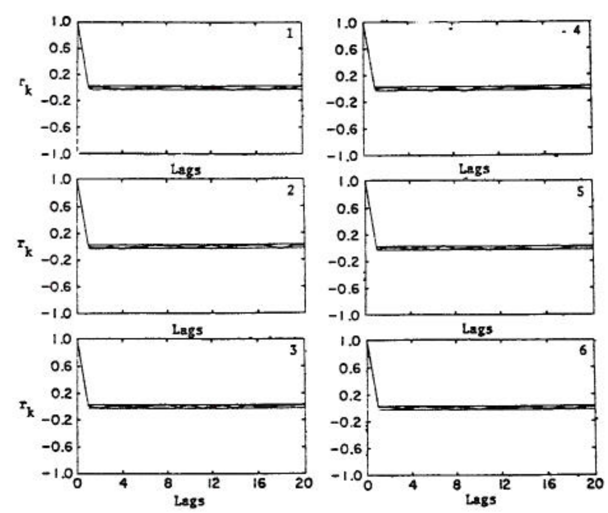

Fig. 4-1. Correlograms for the Six Generated Samples of Standard Normal Independent Variable, ξ_i, Used to Generate Study Samples. All Six Series of ξ_i are Serially Independent.

The chi-square test for goodness of fit is widely used in fitting distributions and is described in many standard text books of statistics. In executing the chi-square test the range of sample observations is divided into k mutually exclusive class intervals each having an observed frequency O_i. If the expected class frequencies from probabilities of the distribution being tested are denoted as E_i then a chi-square statistic is

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} , \tag{4.2}$$

This statistic is asymptotically distributed as a chi-square distribution with k-1-λ degrees of freedom, where λ is the number of distribution parameters estimated from data. Values of the chi-square distribution for various degrees of freedom and probability levels are found in standard tables. If the computed chi-square statistic is larger than the critical value from the table for a given probability level, then the tested distribution is rejected.

The generated normally distributed random deviates, ξ_i, may be transformed to standardized gamma variates, ξ'_i, with skewness λ on the basis of the Wilson-Hilferty approximation to the χ^2- distribution as follows,

$$\xi'_i = (2/\lambda)\ (1+\lambda\xi_i/6-\lambda^2/36)^3-2/\lambda \tag{4-3}$$

4-2 Series with Overremoved and Underremoved Harmonic Terms

The first set of samples was analyzed to test whether the over-removal and under-removal of harmonic terms in the mean, $\hat{\mu}_\tau$, and in the standard deviation, $\hat{\sigma}_\tau$, will change significantly the underlying distribution of the residual independent variable $\hat{\xi}_i$, namely to show sharp peaks similar to those found by Tao and Chin. The samples of the process

$$X_i = \mu_\tau + \sigma_\tau\ \varepsilon_i , \tag{4-4}$$

in which the periodic parameters are assumed to be

$$\mu_\tau = \sigma_\tau = 10 + \sum_{j=1}^{5} (A_j \cos \frac{2\pi j\tau}{\omega} + B_j \sin \frac{2\pi j\tau}{\omega}) , \tag{4-5}$$

with

$$A_j = B_j = 3.0, \text{ for } j=1, 2, \ldots, 5, \tag{4-6}$$

and

$$\varepsilon_i = 0.8\ \varepsilon_{i-1} + \xi_i , \tag{4-7}$$

with ξ_i = the independent standardized normal process, were generated and then analyzed as a periodic-stochastic process as described in Chapter II, including also the use of the periodic autoregressive coefficients.

Figure 4-2 shows the frequency distribution of the generated sample and the estimated independent residual series, after one through ten harmonics were removed. A first-order autoregressive scheme was used in generating the dependent stochastic process $\hat{\varepsilon}_i$. The coefficient of determination was used to measure the effect of the dependence model, with Eqs. (2-27) through (2-32) specifying the approach applied consistently to all samples. The frequenty distributions of Fig. 4-2 show that the peaks of density curves are a direct result of the number of harmonic terms used, being either overestimated or underestimated in comparison with those of the original data. When the right number of harmonics is inferred and removed, the

Table 4-1. Statistics of Distributions of Estimated $\hat{\xi}_i$ Series After Removal of One Through Ten Harmonics. Series were Generated by Using Five Harmonics for the Periodic Mean and Periodic Standard Deviation.

(1) Sample	(2) Number of Harmonics Removed	(3) Total Explained Variance by Removed Harmonics: Mean	(4) Total Explained Variance by Removed Harmonics: Standard Deviation	(5) Mode	(6) Skewness	(7) Kurtosis	(8) Chi-Square For Normal	(9) Beta Parameter (Equation 3-33)	(10) Chi-Square For Double-Branch Exponential
1	1	15.7	37.2	-0.1	0.78958	22.95	592.84	0.3769	52.99
2	2	36.6	51.5	-0.1	-0.11897	16.94	344.31	0.4825	33.60
3	3	61.8	59.9	-0.1	1.10923	16.36	310.97	0.4733	32.91
4	4	77.4	64.7	-0.1	0.59519	8.62	43.98	0.5061	30.69
5	5	90.3	73.3	-0.1	1.11643	37.84	544.59	0.4066	38.25
6	6	91.0	85.2	-0.1	0.87061	35.89	486.80	0.3864	36.99
7	7	91.3	87.8	-0.1	0.14577	63.77	403.22	0.4058	33.81
8	8	91.7	88.8	0.0	-0.33920	27.73	431.32	0.5055	31.02
9	9	92.1	89.4	-0.1	-1.011954	22.51	174.65	0.5579	30.16
10	10	92.2	89.8	-0.1	-0.07283	28.82	181.01	0.5365	30.46

resulting peaks in frequency distributions of $\hat{\xi}_i$ should be attributed to inaccuracy in estimating the amplitudes and phases of the removed harmonics.

Table 4-1 gives statistics associated with frequency distributions shown in Fig. 4-2. This table indicates that almost all of these frequency distributions have slightly negative modes. Such properties were also found by Tao [20]. It can also be seen that a double-branch exponential distribution fits the estimated frequency distributions well enough to pass a chi-square test in all cases. The standard normal distribution is rejected in all cases, except when only four harmonics were removed, but in that case the double-branch exponential function gives a much better fit. It can also be seen that the kurtosis coefficient is always high, as it is expected for the sharp-peaked frequency distributions illustrated in Fig. 4-2.

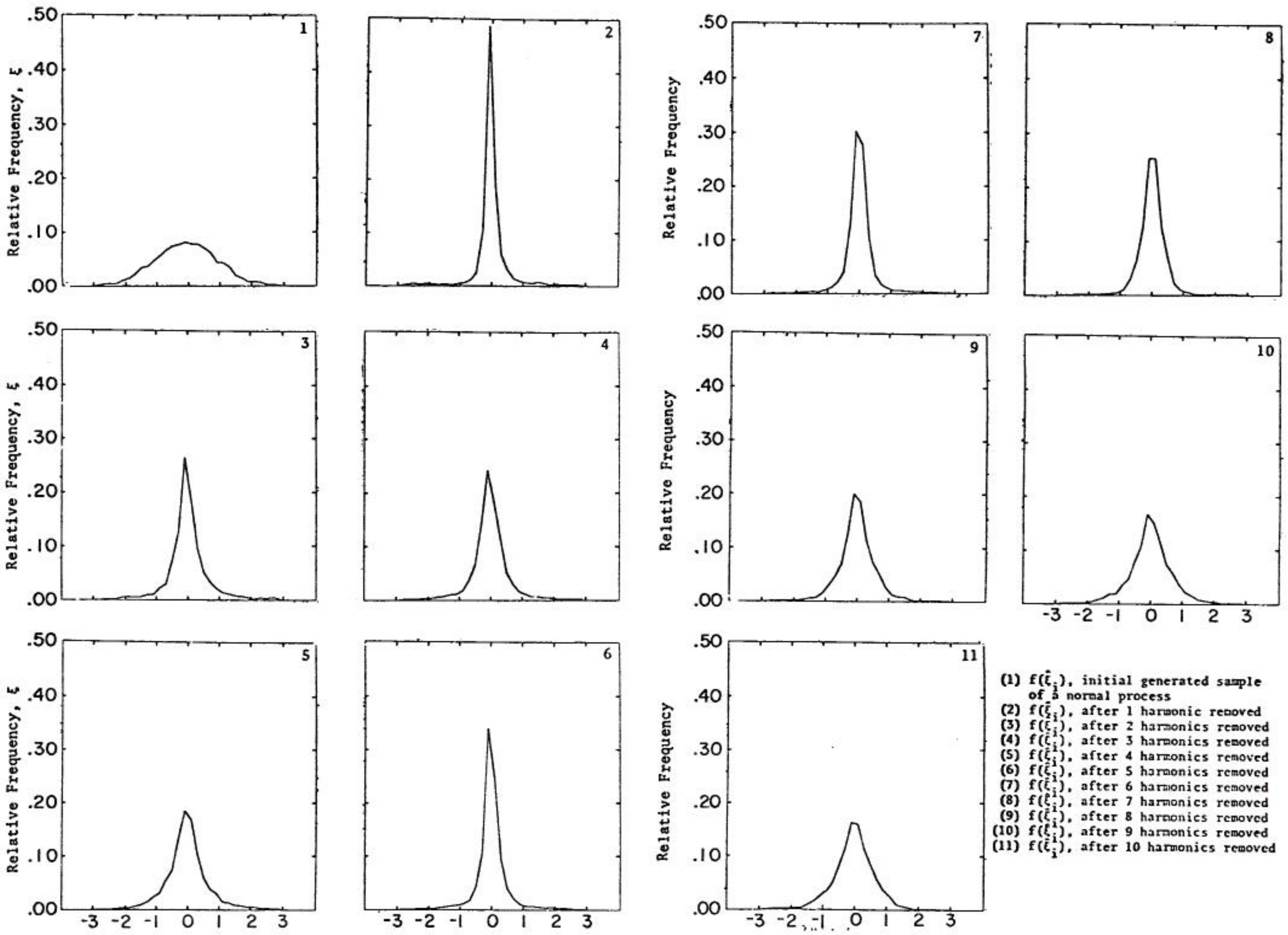

Fig. 4-2. Frequency Distributions of Estimated Independent $\hat{\xi}_i$ Process After One to Ten Harmonics in Parameters of the Periodic-Stochastic Process are Removed Sequentially, with an Initial Standard Normal Process, with Added Periodicities in the Mean and Standard Deviation, Each Represented by 5 Harmonics.

Chapter V
EFFECT OF REMAINING HARMONICS AND AUTOREGRESSIVE COEFFICIENTS ON FREQUENCY DISTRIBUTIONS OF RESIDUALS

5-1 Methodology

It was demonstrated in the preceeding chapter that mis-estimating the number and amplitudes of harmonics in parameters of a hydrologic time series would lead to peaked frequency distributions for the inferred independent $\hat{\xi}_i$. The effects of the remaining harmonic terms will be studied in more detail in this chapter. To perform this analysis the generated series X_i are analogous to the inferred $\hat{\xi}_i$ series of the previous chapter. This is done so that the exact mathematical composition of the X_i series may be known before its frequency distribution is found, whereas the exact mathematical composition of the inferred $\hat{\xi}_i$ series of the previous chapter was difficult to control and predict. By studying the frequency distribution of the X_i series with different known mathematical compositions, more may be learned regarding the effects of remaining harmonics and autoregressive terms on the frequency distribution of the inferred independent series $\hat{\xi}_i$ in the modeling process.

5-2 Series with Harmonics Remaining in the Mean

In this section, the study of generated series is in the form,

$$X_i = \mu_\tau + \xi_i \qquad (5\text{-}1)$$

with μ_τ defined by Eq. (2-2) and ξ_i by Eq. (3-2).

Data generated using Eq. (5-1) have harmonics in the mean only. The first set of samples was generated with no autoregressive terms. The frequency distributions of the X_i series were estimated as the amplitudes in the remaining harmonics in μ_τ were increased. The variable used to describe the proportion of the variance of X_i explained by the remaining number of harmonics in μ_τ is

$$P_m = \sum_{j=1}^{m} \Delta P_j \qquad (5\text{-}2)$$

where ΔP_j is defined in Eq. (2-9). Figure 5-1 shows distributions of six of the ten cases investigated under this approach. It is seen that the deviations are small for low values of P_m. The influence of μ_τ is large for frequency distributions with double peaks. Table 5-1 gives the main distribution parameters for all ten cases investigated under this approach. The only noticeable trends shown by this set of samples is that as P_m increases the kurtosis coefficient decreases, while, as expected, the chi-square statistic obtained in fitting the standard normal function increases.

The next set of generated samples was also based on the approach of Eq. (5-1), but for this set a first order autoregressive scheme was superposed on the generated X_i series. In order to examine the influence of the remaining autoregressive dependence in the process of eliminating it by the estimation procedure, the P_m value was held constant at 0.95 and the assumed population values of α_1 were allowed to vary over the range 0.1 to 0.9, the range within which values of α_1 are commonly found in hydrologic time series.

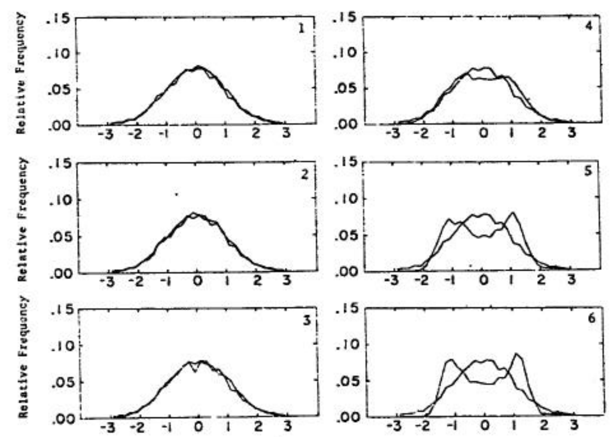

Fig 5-1. Frequency of Estimated X_i Series When Harmonics are Left in the Mean, the Initial Normal Distribution or ξ_i Series For: (1) $P_m = 0.1$, (2) $P_m = 0.3$, (3) $P_m = 0.5$, (4) $P_m = 0.7$, (5) $P_m = 0.9$, and (6) $P_m = 0.95$.

Table 5-1. Distribution Statistics for X_i Series from Samples with Harmonics Remaining in the Mean. No Autoregressive Terms Included.

(1) Sample	(2) P_m	Modes (3) Right	Modes (4) Left	(5) Skewness Coefficient	(6) Kurtosis Coefficient	(7) Chi-Square for Fitting N(0,1)
1	0.1	0.1	--	0.05	2.95	5.69
2	0.2	0.1	--	0.05	2.93	7.89
3	0.3	0.5	-0.3	0.05	2.87	13.52
4	0.4	0.3	-0.5	0.05	2.78	25.00
5	0.5	0.5	-0.5	0.05	2.65	41.12
6	0.6	0.3	-0.5	0.04	2.49	53.24
7	0.7	0.5	-0.7	0.04	2.30	109.85
8	0.8	0.7	-0.9	0.03	2.07	238.62
9	0.9	0.9	-1.1	0.02	1.81	396.74
10	0.95	1.3	-1.1	0.01	1.67	595.45

Figure 5-2 shows the distribution of the estimated X_i series for six of the ten samples generated. Comparing the results of different series, it can be seen that the presence of a large positive autoregression coefficient has a smoothing effect on the estimated distribution of X_i of these samples. Table 5-2 gives the values of distribution parameters

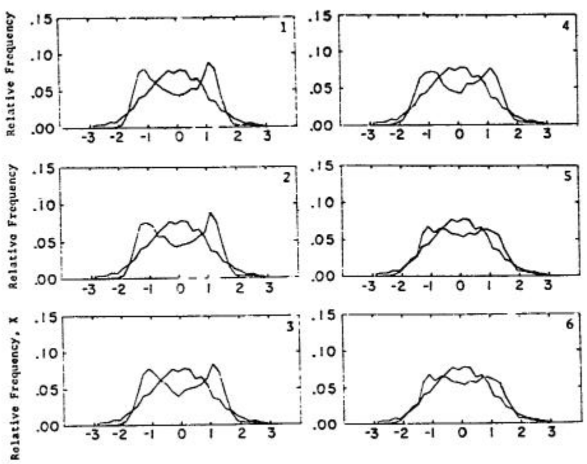

Fig. 5-2. Distribution of the X_i Series for Samples with Harmonics in the Mean Remaining and an Autoregressive Series of Order-One for:

(1) $P_m = 0.95$, $\alpha_1 = 0.1$, (2) $P_m = 0.95$, $\alpha_1 = 0.3$, (3) $P_m = 0.95$, $\alpha_1 = 0.5$, (4) $P_m = 0.95$, $\alpha_1 = 0.7$, (5) $P_m = 0.95$, $\alpha_1 = 0.8$, (6) $P_m = 0.95$, $\alpha_1 = 0.9$.

gives a correlation coefficient of 0.97. If Eq. (3-21) is not used, it may be assumed that the values r_1 of column 8 in Table 5-2 are the best estimates of the true values of α_1, given in column 2. This assumption would obviously lead to erroneous conclusions regarding the true state of an autoregressive process found in nature.

5-3 Series with some Harmonics Unremoved in both the Mean and the Standard Deviation

In this section the independent residual series is assumed to follow the complex form of Eq. (4-4), which allows the number of remaining harmonics in the mean and the standard deviation to be simulated. The first set of samples under this approach is generated under the assumption that no autoregressive dependence is present; however, the amplitude of the remaining harmonic terms in the mean μ_τ and in the standard deviation σ_τ were allowed to vary. The proportion of variance explained by the remaining harmonic terms in μ_τ, and σ_τ was set equal, but increased for each new sample. Again, the explained variance is measured by P_m. Figure 5-3 shows the estimated fre-

Table 5-2. Distribution Statistics for X_i Series with Harmonics in the Mean Held Constant ($P_m = 0.95$) and an Autoregressive Scheme of Order-One.

(1) Sample	(2) α_1	(3) Modes Right	(4) Modes Left	(5) Skewness	(6) Kurtosis	(7) Chi-Square N(0,1) Fit	(8) r_1	(9) r_2	(10) $\hat{\alpha}_1$
1	0.1	1.1	-1.1	-0.003	1.65	617.54	0.955	0.948	0.143
2	0.2	1.1	-1.1	-0.002	1.65	573.76	0.958	0.948	0.243
3	0.3	1.1	-1.1	-0.002	1.66	531.39	0.962	0.949	0.343
4	0.4	1.1	-1.1	-0.001	1.68	491.05	0.965	0.949	0.444
5	0.5	1.1	-1.1	-0.005	1.69	452.12	0.967	0.950	0.545
6	0.6	1.1	-1.1	0.001	1.73	442.10	0.970	0.951	0.646
7	0.7	1.1	-1.0	0.006	1.78	396.74	0.972	0.952	0.749
8	0.8	1.1	-1.1	0.014	1.87	297.03	0.975	0.954	0.850
9	0.9	1.0	-1.1	0.047	2.08	252.61	0.979	0.960	0.951

for ten samples used under this approach. The conclusions from this table are that with an increase of α_1 the chi-square value of the standard normal distribution fit decreases, while the skewness coefficient increases marginally.

The generated samples under this approach allow Eq. (3-21) to be checked. Using the estimates r_1 and r_2, obtained by using Eq. (2-19), the values of $\hat{\alpha}_1$ are estimated by Eq. (3-21) and given in the last column of Table 5-2. The values of $\hat{\alpha}_1$ are close to the true value of α_1, given in the second column of Table 5-2. A linear regression between the first nine values in column 2 and column 10 of Table 5-2

quency distributions of X_i in four of the ten samples. The sharp-peaked curves in all cases represent frequency distributions of the X_i series, while the lower curves correspond to the frequency distributions of the generated independent normal process ξ_i. Table 5-3 gives the distribution parameters for all ten samples.

The chi-square values in Table 5-3 indicate that the largest differences between the generated sample and the estimated frequency curve of X_i are obtained when the remaining harmonics in μ_τ and σ_τ explain between 40 percent and 60 percent of the total variance of these parameters. With the 40 percent or 60 percent explained variance, the mode has the largest negative values, the skewness and kurtosis coefficients are the largest, and the chi-square values are also the largest.

In the next set of samples generated, the harmonics in the μ_τ are held constant while the amplitudes of harmonics in the σ_τ were allowed to vary. Again no autoregressive term was used to produce the X_i series from the generated ξ_i series.

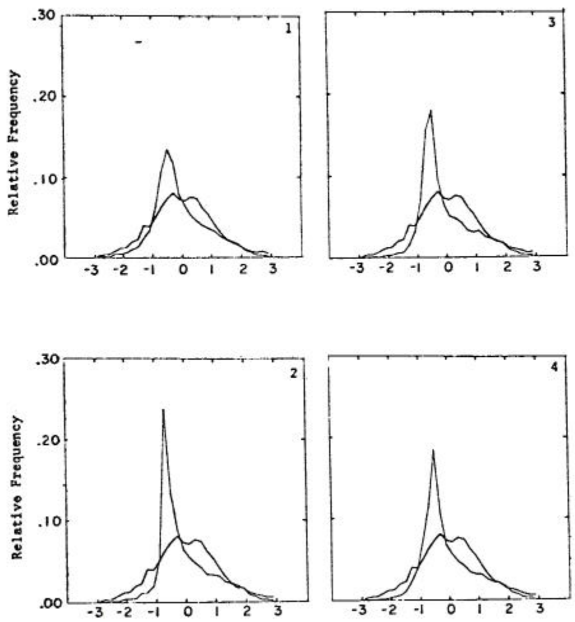

Fig. 5-3. Distribution of the X_i Series with Equal Amplitudes and Phases of Unremoved Harmonics in the Mean and Standard Deviation for: (1) $P_m = 0.2$, (2) $P_m = 0.4$, (3) $P_m = 0.5$, (4) $P_m = 0.6$.

Table 5-3. Distribution Statistics for the X_i Series with Equal Amplitude and Phases in Unremoved Harmonics in the Mean and the Standard Deviation. No Autoregressive Term is Present.

(1) Sample	(2) P_m	(3) Mode	(4) Skewness	(5) Kurtosis	(6) Chi-Square N(0,1) Fit
1	0.1	-0.3	0.442	3.63	20.83
2	0.2	-0.5	0.759	4.07	99.25
3	0.3	-0.7	0.946	4.28	172.24
4	0.4	-0.7	1.052	4.36	236.48
5	0.5	-0.5	1.087	4.33	236.84
6	0.6	-0.5	1.065	4.25	223.12
7	0.7	-0.5	0.988	4.12	185.55
8	0.8	-0.3	0.851	3.97	139.91
9	0.9	-0.3	0.629	3.81	52.75
10	0.95	-0.1	0.454	3.72	30.45

The ratio C_s^2/C_m^2 measures the differences in amplitudes of the remaining harmonics in μ_τ. Values of C_s^2 and C_m^2 are computed by Eq. (2-6) and the ratio is given the range from 0.111 to 19.0 in the samples tested. The larger the ratio C_s^2/C_m^2, the larger is the proportion of the variance explained by the unremoved harmonics in σ_τ as compared to μ_τ. Figure 5-4 gives the distribution of the estimated X_i and the generated ξ_i series in four of the samples. Table 5-4 gives the distribution parameters for all ten samples. It is difficult to determine from Table 5-4 when the worst conditions occur. It appears that equal variance explained by the unremoved harmonics in μ_τ and σ_τ will produce largest differences when the variance of σ_τ explained by unremoved harmonics is either much greater or much smaller than the variance explained by the unremoved harmonics in μ_τ. No definite trends can be noticed in the distribution parameters as the ratio of C_s^2/C_m^2 increases.

The next set of generated samples uses the same approach of Eq. (4-4). The third sequences ξ_i and the equal explained variances of μ_τ and σ_τ by the unremoved harmonics were used. However, the phase angle θ_1 of Eq. (2-7) varied from 45 degrees to 405

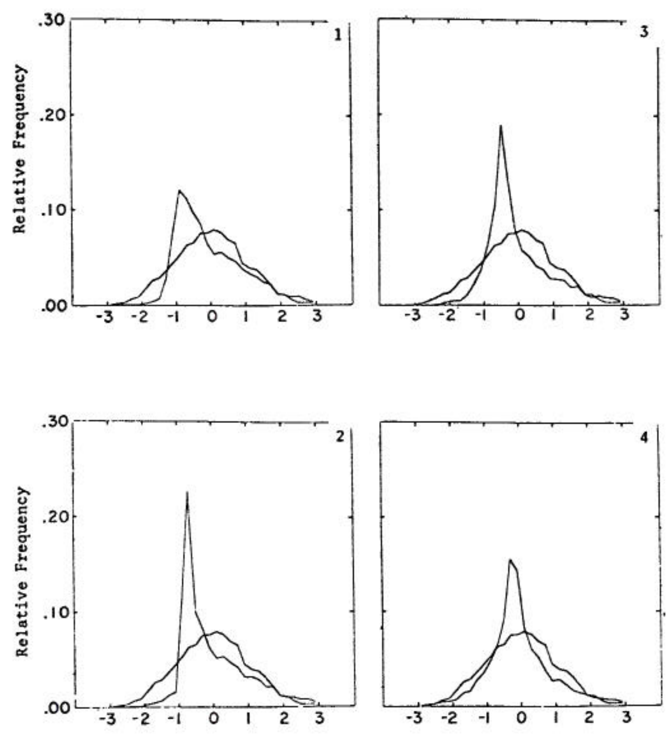

Fig. 5-4. Distribution of the X_i Series When the Amplitude of Unremoved Harmonics in the Mean are Held Constant ($P_m = 0.5$), and the Amplitudes of the Unremoved Harmonics in the Standard Deviation are Allowed to Vary for: (1) $C_s^2/C_m^2 = 0.25$, (2) $C_s^2/C_m^2 = 0.667$, (3) $C_s^2/C_m^2 = 1.50$, (4) $C_s^2/C_m^2 = 4.00$

Table 5-4. Distribution Statistics for the X_i Series with Unremoved Harmonics in the Mean are Held Constant (P_m = 0.5), and the Unremoved Harmonics in the Standard Deviation are Allowed to Vary.

(1) Sample	(2) C_s^2/C_m^2	(3) Mode	(4) Skewness	(5) Kurtosis	(6) Chi-Square N(0,1) Fit
1	0.111	-0.7	0.719	3.159	34.64
2	0.250	-0.9	0.962	3.614	132.78
3	0.429	-0.9	1.116	4.051	255.13
4	0.667	-0.7	1.207	4.461	297.99
5	1.000	-0.5	1.240	4.830	396.43
6	1.500	-0.5	1.212	5.132	431.95
7	2.333	-0.3	1.110	5.331	450.20
8	4.000	-0.2	0.915	5.336	376.74
9	9.000	-0.1	0.592	5.145	253.20
10	19.000	-0.1	0.364	4.894	188.39

degrees for the fundamental harmonic of σ_τ was held constant at the 45 degrees. The difference between the fundamental phase angles of μ_τ and σ_τ series, $\Delta\theta$, is defined by

$$\Delta\theta = \theta_{1\ (\text{for } \sigma_\tau)} - 45 . \qquad (5\text{-}3)$$

The amplitudes of the remaining harmonic terms in μ_τ and σ_τ are kept equal, and no autoregressive term was present for the ε_i series. Figure 5-5 shows six of the 13 samples generated under this approach. Table 5-5 gives the distributions parameters for all 13 samples. The interesting aspect of this approach is the reappearance of the double-peak distribution, when the phase angles of the fundamental harmonics of μ_τ and σ_τ were out of phase. The phase angles are nearly equal in most cases of observed series, and distributions are usually with single sharp peak, negative mode, high kurtosis and positive skewness. No distribution in these samples had a single-peaked and a positive mode distribution.

The next set of generated samples was used to detect the influence of the unremoved autoregressive dependence in the ε_i series, while retaining the unremoved harmonics in μ_τ and σ_τ of Eq. (4-4). The fourth sample of ξ_i was used in this generation of samples. In the first set, an autoregressive scheme of the order one was used in generating the ε_i series. The remaining harmonic terms in μ_τ and σ_τ explained each 40 percent of the variances, while the autoregressive lag-one coefficient α_1 was allowed to vary between 0.1 and 0.9. Figure 5-6 shows four cases in this category. Table 5-6 gives the parameters of X_i distributions for all ten cases. These data sets show a surprising influence of the unremoved part of the autoregressive dependence. As α_1 in the ε_i series increases, the kurtosis of the distribution of the X_i series increases, while the positive skewness and a negative mode have smaller values. The fact that the chi-square values computed for fitting the standard normal distribution decrease as α_1 increases was reflected in the decrease of skewness. This shows a lesser influence of the large kurtosis on the estimated X_i distribution, indicating that a large α_1 causes the skewness to decrease faster than the kurtosis increases.

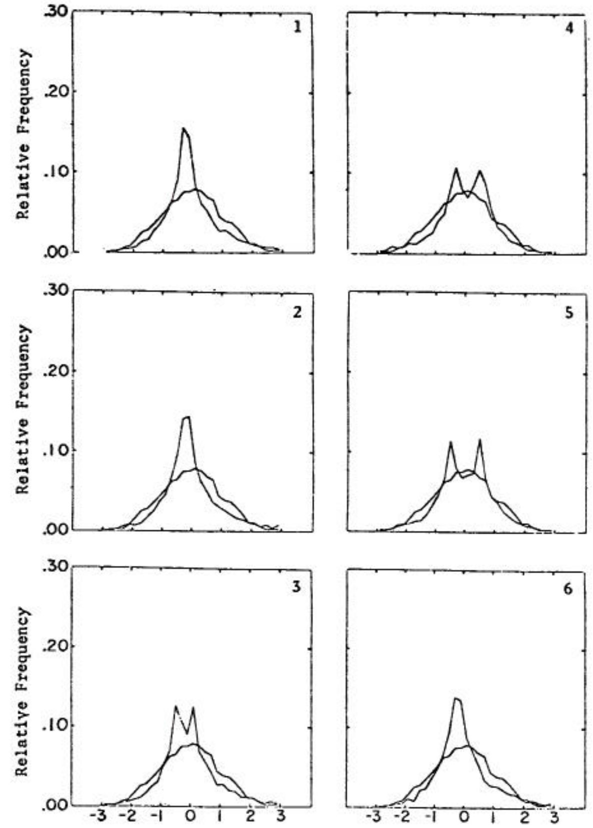

Fig. 5-5. Distribution of the X_i Series When the Phase Angle Between the Harmonics in Mean and Standard Deviation are Allowed to Vary for: (1) $\Delta\theta$ = 0°, (2) $\Delta\theta$ = 30°, (3) $\Delta\theta$ = 90°, (4) $\Delta\theta$ = 150°, (5) $\Delta\theta$ = 180°, (6) $\Delta\theta$ = 330°.

The final set of samples was used with an autoregressive scheme of the order two in the ε_i series, and the presence of unremoved harmonics in μ_τ and σ_τ of Eq. (4-4). The fifth sample of ξ_i was used. Again the percent of the explained variance by unremoved harmonics in μ_τ and σ_τ was held constant at 40 percent, the fundamental harmonics in both parameters were in phase, while autoregressive coefficients were varied according to the following relationship

$$2\,\alpha_1 + \alpha_2 = 1 \qquad (5\text{-}4)$$

$$0.1 \le \alpha_1 \le 1.0 \qquad (5\text{-}5)$$

and

$$-1.0 \le \alpha_2 \le 1.0 \qquad (5\text{-}6)$$

Table 5-5. Distribution Statistics of the X_i Series When the Phase Angle Between the Harmonics in the Mean and the Standard Deviation Vary. No Autoregressive Term Present.

(1) Sample	(2) θ_s Degrees	(3) θ_m Degrees	(4) $\Delta\theta$ Degrees	(5) Modes Right	(6) Modes Left	(7) Skewness	(8) Kurtosis	(9) Chi-Square N(0,1) Fit
1	45	45	0	---	-0.3	0.915	5.36	376.74
2	75	45	30	---	-0.2	0.886	5.33	264.54
3	105	45	60	0.1	-0.5	0.799	5.24	326.23
4	135	45	90	0.1	-0.1	0.210	4.78	123.86
5	165	45	120	0.5	-0.3	-0.468	4.64	131.03
6	195	45	150	0.5	-0.5	-0.289	4.39	66.61
7	225	45	180	0.5	-0.5	-0.736	4.29	71.08
8	255	45	210	0.5	-0.5	0.164	4.37	103.50
9	285	45	240	0.3	-0.5	0.401	4.61	112.20
10	315	45	270	0.3	-0.5	0.619	4.91	173.56
11	345	45	300	0.1	-0.5	0.776	5.16	359.38
12	375	45	330	---	-0.2	0.879	5.32	394.38
13	405	45	360	---	-0.3	0.915	5.36	376.74

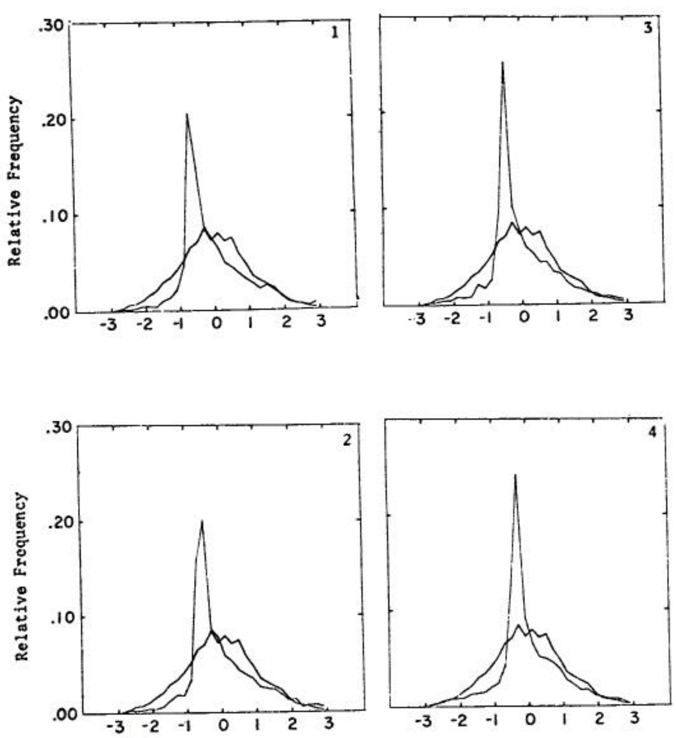

Fig. 5-6. Distribution of the X_i Series with Equal Amplitudes and Phases of the Harmonics in the Mean and Standard Deviation, and an Autoregressive Order-One Series, with α_1 Allowed to Vary: (1) α_1 = 0.2, (2) α_1 = 0.4, (3) α_1 = 0.6, (4) α_1 = 0.8.

Table 5-6. Distributions Statistics for the X_i Series When the Amplitudes and Phases of Harmonics in the Mean and Standard Deviation are Held Constant, with an Autoregressive Series of Order-One, with α, Allowed to Vary.

(1) Sample	(2) α_1	(3) Mode	(4) Skewness	(5) Kurtosis	(6) Chi-Square N(0,1) Fit	(7) $\hat{\alpha}_1$
1	0.1	-0.7	1.028	4.634	250.42	0.29
2	0.2	-0.7	1.021	4.674	279.66	0.36
3	0.3	-0.6	1.007	4.737	250.42	0.43
4	0.4	-0.5	0.988	4.819	279.46	0.50
5	0.5	-0.5	0.958	4.923	289.49	0.56
6	0.6	-0.5	0.910	5.063	312.36	0.63
7	0.7	-0.5	0.829	5.264	280.38	0.71
8	0.8	-0.3	0.663	5.555	173.97	0.80
9	0.9	-0.1	0.209	5.975	145.44	0.90

This allowed the autoregressive process to remain stable, and allowed several different spectra for autoregressive processes to be used [10]. Figure 5-7 shows the estimated distribution of X_i for four cases.

Table 5-7 gives the distribution statistics for all ten cases of this category. The noticeable results for this set of samples show that when the α_1 coefficnet increases the α_2 coefficient decreases, the negative distribution mode decreases, the positive skewness increases while the kurtosis decreases. It is again not clear what are the worst effects in this case. All of the estimated distributions of X_i

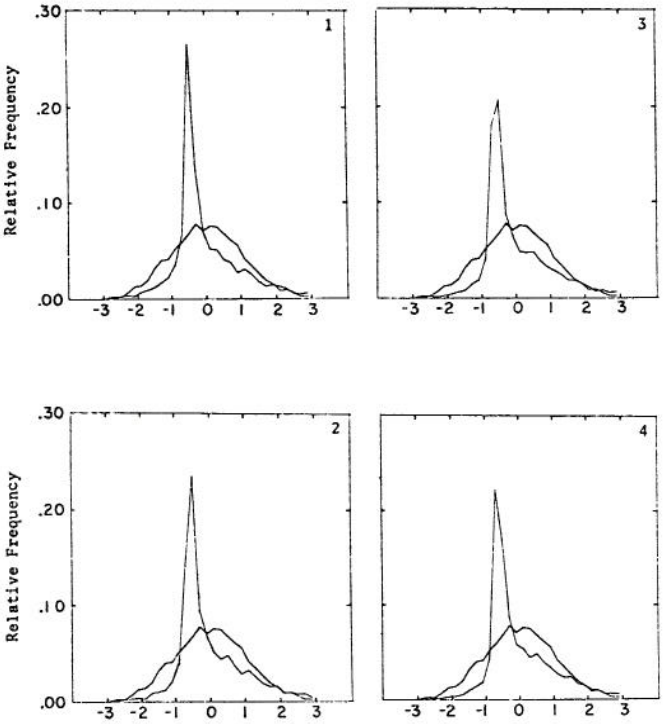

Fig. 5-7. Distribution of the X_i Series with Equal Amplitude and Phase in Unremoved Harmonics of the Mean and Standard Deviation, and an Autoregressive Order-Two Dependence Which is Allowed α_1 and α_2 to Vary: (1) α_1 = 0.2, α_2 = 0.6, (2) α_1 = 0.4, α_2 = 0.2, (3) α_1 = 0.5, α_2 = 0.0, (4) α_1 = 0.6, α_2 = -0.2.

Table 5-7. Distribution Statistics of the X_i Series When Amplitudes and Phases in the Unremoved Standard Deviation Series are Equal, with an Autoregressive Series of Order-Two Which Allowed α_1 and α_2 to Vary.

(1) Sample	(2) α_1	(3) α_2	(4) Mode	(5) Skewness	(6) Kurtosis	(7) Chi-Square for N(0,1) Fit
1	0.1	0.8	-0.3	1.086	5.77	377.35
2	0.2	0.6	-0.4	1.108	5.26	310.73
3	0.3	0.4	-0.5	1.114	4.95	326.81
4	0.4	0.2	-0.5	1.117	4.78	343.76
5	0.5	0.0	-0.6	1.119	4.66	328.28
6	0.6	-0.2	-0.7	1.121	4.55	414.74
7	0.7	-0.4	-0.7	1.123	4.47	443.25
8	0.8	-0.6	-0.7	1.124	4.43	378.99
9	0.9	-0.8	-0.7	1.127	4.42	331.54

represent a complex transformation of the initial generated sample from a standard normal distribution, ξ_i.

Table 5-8 shows the assumed true population autoregressive coefficients α_1 and α_2, the estimated values of the first and second serial correlation coefficients r_1 and r_2, and the estimated $\hat{\alpha}_1$ and $\hat{\alpha}_2$ by using the estimates r_1 and r_2 and Eqs. (2-22) and (2-23). The wide discrepancy between the population values α_1 and α_2 and the estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ could mainly be attributed to the presence of unremoved harmonic terms in μ_τ and σ_τ. Again, an investigator can make erroneous conclusions regarding the true state of the autoregressive process ε_i in the observed series of hydrologic variable.

Table 5-8. Effect of Unremoved Harmonics in the Mean and Standard Deviation on Autoregressive Order-Two Estimates

(1) Sample	(2) α_1	(3) α_2	(4) r_1	(5) r_2	(6) $\hat{\alpha}_1$	(7) $\hat{\alpha}_2$
1	0.1	0.8	0.907	0.824	0.900	0.008
2	0.2	0.6	0.824	0.683	0.813	0.013
3	0.3	0.4	0.747	0.568	0.730	0.023
4	0.4	0.2	0.677	0.477	0.654	0.034
5	0.5	0.0	0.610	0.406	0.577	0.054
6	0.6	-0.2	0.545	0.351	0.503	0.077
7	0.7	-0.4	0.482	0.310	0.433	0.101
8	0.8	-0.6	0.417	0.282	0.362	0.131
9	0.9	-0.8	0.351	0.264	0.295	0.161

5-4 Discussion of Experimental Results

It was demonstrated that the distribution of the estimated independent stochastic component X_i series by Eq. (4-4) from a generated periodic-stochastic process would exhibit a sharp-peaked distribution in comparison with the generated sample of an independent normal process. This is true with and without the presence of the autoregressive terms in the $\hat{\varepsilon}_i$ series; however, it does require a periodic standard deviation σ_τ. Therefore, the unremoved or overremoved harmonics in the standard deviation of the estimated independent stochastic component of periodic-stochastic process probably account for the sharply peaked distributions of the estimated independent stochastic component. This distribution problem is illustrated in Fig. 5-8. The periodic standard deviation creates problems in estimating the distribution of the independent stochastic component, $\hat{\xi}_i$. The series ξ_i may not be identically distributed, nor necessarily independent, if the mean is periodic and the $\hat{\varepsilon}_i$ series is an autoregressive process. A strictly analytical solution of this problem seems to be difficult but the validity of any solution could be checked by simulation on computers.

It is also apparent that the presence of a periodic mean, μ_τ in Eq. (4-4), causes the mean and the mode of the peaked distribution of estimated independent component, $\hat{\xi}_i$, to be somewhat different, and in fact causes the mode to be negative, for the unremoved or overremoved harmonics in μ_τ and σ_τ while they are in phase. The presence of the unremoved or overremoved autogressive dependence in $\hat{\varepsilon}_i$ series may also have a slight effect on the kurtosis.

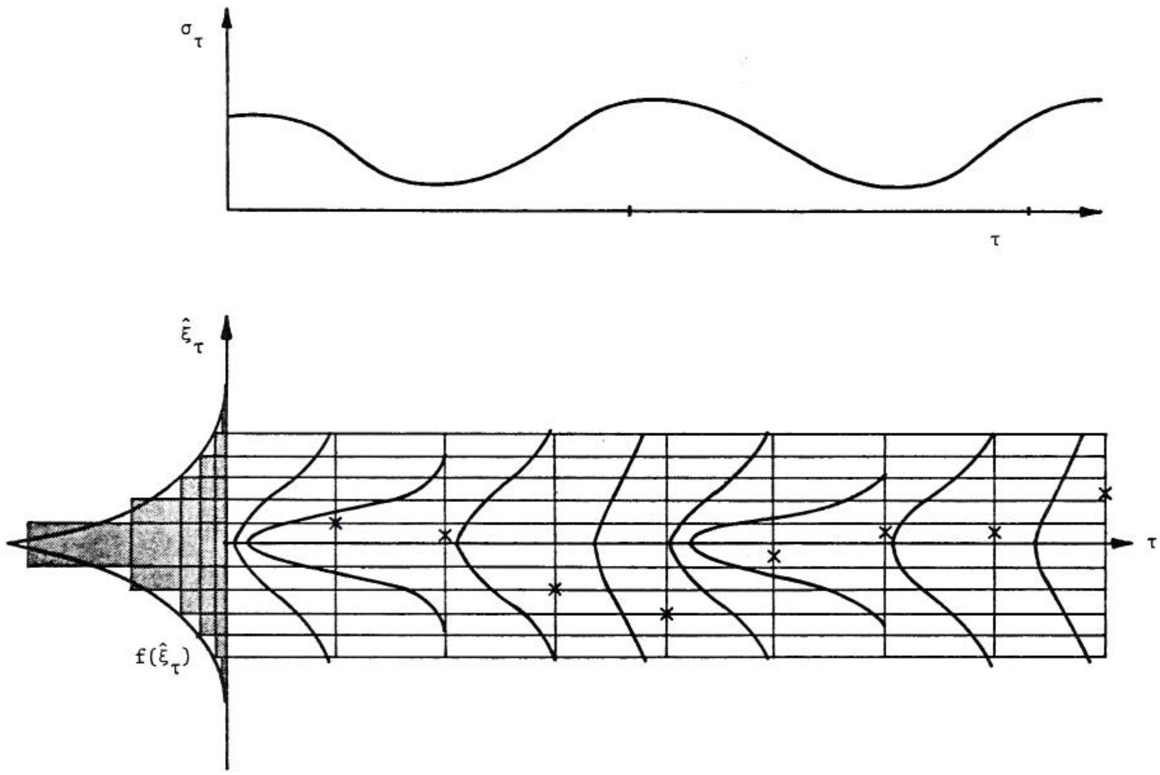

Figure 5-8. Illustration of the Effect of a Periodic Standard Deviation on the Distribution of the Estimated ξ_i Component.

Table 6-1. Simulated X_i Series Using Standard Normal Noise with Harmonics in the Mean and Standard Deviation and Non-Periodic Autoregressive Parameters Estimated from Historical Daily Data. Skewness and Kurtosis of Estimated Independent Residuals, ξ_i, after Estimating and Removing Significant Harmonics and Non-Periodic Autoregressive Dependence.

		SIMULATED										ESTIMATED											
	Coefs. from	Periodicity in Mean					Periodicity in Std. Dev.					Periodicity in Mean					Periodicity in Std. Dev.						
Series	Stn.	NHM	j	A_j	B_j	ΔP_j	NHS	j	A_j	B_j	ΔP_j	NHM	j	A_j	B_j	ΔP_j	NHS	j	A_j	B_j	ΔP_j	Skewness	Kurtosis
1	2	5	1	-200.3	-112.4	0.3784	0					5	1	-218.7	-144.4	.4337	0					0.033	3.015
			2	145.4	185.0	0.3971							2	138.1	189.2	.3465							
			3	-85.5	-79.9	0.0982							3	-18.2	44.4	.0145							
			4	58.0	65.6	0.0550							4	-1.2	17.6	.0020							
			5	-39.7	-72.5	0.0490							5	-14.7	-11.8	.0022							
2				Experiment repeated with same simulated data								5	1	-234.1	-88.6	.4115	0					0.009	2.958
													2	134.6	173.3	.3163							
													3	22.4	22.8	.0067							
													4	14.8	-11.9	.0024							
													5	-11.8	5.6	.0011							
3				Experiment repeated with same simulated data								5	1	-180.7	-129.5	.2880	0					-0.005	3.003
													2	189.9	206.6	.4586							
													3	-17.7	3.7	.0019							
													4	-51.4	14.6	.0166							
													5	-33.7	31.9	.0126							
4	12	3	1	-563.8	-6.91	.7026	1	1	-382.9	202.7	.6556	6	1	-573.7	-32.7	.7135	1	1	-457.1	257.5	.9492	0.032	3.018
			2	198.8	193.8	.1703							2	216.9	189.7	.1794							
			3	-100.0	-158.4	.0779							3	-3.9	28.9	.0018							
													4	2.3	3.3	.0000							
													5	-22.4	-21.1	.0020							
													6	-84.4	-146.9	.0621							
5				Experiment repeated with same simulated data								6	1	-598.9	25.5	.7506	1	1	-496.5	241.0	0.9610	0.033	3.182
													2	188.8	161.1	.1287							
													3	20.6	18.9	.0016							
													4	2.8	-15.7	.0005							
													5	-2.4	19.1	.0008							
													6	-117.1	-164.0	.0848							
6	11	5	1	-888.4	509.9	0.8546	1	1	-798.3	648.3	0.6009	6	1	-858.1	383.0	.6699	1	1	-1435.6	1158.1	.9471	0.002	2.999
			2	-20.2	-311.0	0.0791							2	163.1	-217.5	.0561							
			3	-17.7	-1.5	0.0010							3	-42.5	64.3	.0045							
			4	-6.2	11.0	0.0005							4	-88.9	49.7	.0079							
			5	13.6	-5.3	0.006							5	-24.7	86.0	.0061							
													6	-168.9	-78.2	.0263							
7				Experiment repeated with same simulated data								6	1	-935.9	502.9	.7352	1	1	-1454.6	1237.5	.9530	0.016	2.969
													2	159.6	-319.8	.0832							
													3	-92.4	-22.6	.0059							
													4	83.6	26.3	.0050							
													5	19.6	99.0	.0066							
													6	-11.5	-88.0	.0051							
8	2	5	1	-200.3	-112.4	0.3784	1	1	-123.3	-85.5	0.2706	5	1	-215.7	-144.9	.4364	1	1	-197.7	-132.1	.8984	0.040	2.973
			2	145.4	185.0	0.3971							2	134.6	186.6	.3423							
			3	-85.5	-79.9	0.0982							3	-18.3	40.6	.0128							
			4	58.0	65.6	0.0550							4	14.1	11.2	.0021							
			5	-39.7	-72.5	0.0490							5	-16.5	-18.2	.0039							
8	12	5	1	-563.8	-6.91	.7026	2	1	-382.9	202.7	.6556	6	1	-561.1	-32.6	.6393	1	1	-440.5	229.3	.9104	0.009	3.582
			2	198.8	193.8	.1703		2	11.4	91.7	.0298		2	242.1	218.8	.2155							
			3	-100.8	-158.4	.0779							3	-5.2	12.1	.0004							
													4	-18.5	13.1	.0010							
													5	-2.3	12.1	.003							
													6	-136.9	-183.6	.1061							
10				Experiment repeated with same simulated data								6	1	-579.2	-5.2	.6819	1	1	-449.1	242.9	.9192	0.008	3.657
													2	243.1	189.8	.1934							
													3	-27.1	-8.9	.0017							
													4	15.1	10.1	.0007							
													5	4.3	24.6	.0013							
													6	-99.8	-180.3	.0863							

Table 6-1. (Continued)

		SIMULATED										ESTIMATED											
	Coefs.	Periodicity in Mean					Periodicity in Std. Dev.					Periodicity in Mean					Periodicity in Std. Dev.						
Series	from Stn.	NHM	j	A_j	B_j	ΔP_j	NHS	j	A_j	B_j	ΔP_j	NHM	j	A_j	B_j	ΔP_j	NHS	j	A_j	B_j	ΔP_j	Skewness	Kurtosis
11	11	5	1	-888.4	509.9	0.8546	2	1	-798.3	648.3	0.6009	6	1	-781.3	291.6	.6919	2	1	-1462.0	1174.9	.8526	0.020	2.972
			2	-20.2	-311.0	0.0791		2	-235.9	-267.6	0.0723		2	135.0	-156.4	.0425		2	-461.7	-498.8	.1120		
			3	-17.7	-1.5	0.0010							3	-152.1	-37.3	.0244							
			4	-6.2	11.0	0.0005							4	-12.9	2.0	.0002							
			5	13.6	-5.3	0.0006							5	19.9	-70.6	.0054							
													6	100.9	75.3	.0158							
12				Experiment repeated with same simulated data								5	1	-920.7	558.6	.6875	2	1	-1448.7	1160.3	.8456	-0.003	3.084
													2	-10.3	-350.2	.0728		2	-441.4	-472.0	.1025		
													3	-74.1	-3.1	.0033							
													4	69.6	86.3	.0073							
													5	3.8	-137.5	.0112							
13	2	5	1	-200.3	-112.4	0.3784	2	1	-123.3	-85.5	0.2706	5	1	-222.4	-126.2	.4261	2	1	-195.0	-136.8	.3933	0.059	3.118
			2	145.4	185.0	0.3971		2	141.6	105.7	0.3750		2	135.1	189.4	.3526		2	233.1	163.5	.5621		
			3	-85.5	-79.9	0.0982							3	-22.8	29.1	.0089							
			4	58.0	65.6	0.0550							4	17.2	13.9	.0032							
			5	-39.7	-72.5	0.0490							5	-25.8	-11.9	.0053							
14	12	3	1	-563.8	-6.91	0.7026	3	1	-382.9	202.7	.6556	2	1	-546.4	-47.4	.6767	1	1	-432.4	210.0	.8656	0.081	3.794
			2	198.8	193.8	0.1703		2	11.4	91.7	.0298		2	227.4	224.7	.2300							
			3	-100.8	-158.4	0.0779		3	-50.2	-108.9	.0502												
15				Experiment repeated with same simulated data								6	1	-569.9	3.8	.7138	1	1	-430.2	207.6	.8652	-0.026	3.939
													2	205.7	182.9	.1665							
													3	-28.1	-4.5	.0018							
													4	25.4	20.9	.0024							
													5	-7.5	-35.4	.0029							
													6	-89.8	-145.9	.0645							
16	2	5	1	-200.3	-112.4	0.3784	3	1	-123.3	-85.5	0.2706	5	1	-232.1	-102.7	.4116	2	1	-239.7	-151.0	.3589	0.378	8.382
			2	145.4	185.0	0.3971		2	141.6	105.7	0.3750		2	147.0	171.6	.3262		2	280.1	183.3	.5012		
			3	-85.5	-79.9	0.0982		3	-66.4	-46.2	0.0786		3	-0.2	36.5	.0085							
			4	58.0	65.6	0.0550							4	11.8	-28.4	.0060							
			5	-39.7	-72.5	0.490							5	-5.1	18.5	.0024							
17				Experiment repeated with same simulated data NHS in estimates increased to three								5	1	-221.3	-125.9	.4228	(3)	1	-197.5	-138.6	.3540	0.191	4.756
													2	130.0	192.3	.3510		2	228.9	164.5	.4830		
													3	-24.4	36.0	.0123		3	1.3	-6.0	.0002		
													4	20.7	-14.1	.0041							
													5	-20.6	-14.2	.0041							
18				Experiment repeated with same simulated data NHS in estimates increased to six								5	1	-232.1	-102.8	.4116	(6)	1	-239.7	-151.0	.3589	-0.189	4.546
													2	147.0	171.6	.3262		2	280.1	183.3	.5012		
													3	-0.2	36.5	.0085		3	2.0	-3.3	.0001		
													4	11.8	-28.4	.0060		4	5.5	-3.5	.0002		
													5	-5.1	18.5	.0024		5	15.7	25.1	.0039		
19				Experiment repeated with same simulated data NHS in estimates increased to twelve								5	1	-198.1	-106.3	.2956	(12)	1	-192.6	-148.3	.3368	-0.089	3.294
													2	174.3	220.7	.4626		2	239.9	186.3	.5261		
													3	-8.9	-5.4	.0006		3	-7.4	-16.2	.0018		
													4	-50.2	18.1	.0167		4	11.4	20.4	.0031		
													5	-8.3	22.2	.0033		5	-23.7	-43.2	.0138		
20				Experiment repeated with same simulated data, non-parametric method for estimates										Non-Parametric								0.019	3.074
21	2	5	1	-200.3	-112.4	.3784	5	1	-123.3	-85.5	0.2706	5	1	-196.0	-105.6	.2878	5	1	-189.6	-144.1	.2918	0.760	9.584
			2	145.4	185.0	.3971		2	141.6	105.7	0.3750		2	176.1	221.1	.4638		2	233.8	183.4	.4545		
			3	-85.5	-79.9	.0982		3	-66.4	-46.2	0.0786		3	-8.3	-3.4	.0005		3	-5.5	-18.8	.0020		
			4	58.0	65.6	.0550		4	75.7	31.7	0.0809		4	-52.0	18.4	.0177		4	7.7	22.3	.0029		
			5	-39.7	-72.5	.0490		5	-47.2	-43.2	0.0492		5	-7.6	19.9	.0026		5	-24.2	-43.8	.0129		
22				Experiment repeated with same simulated data								5	1	-220.1	-125.7	.4166	5	1	-198.7	-133.1	.3032	0.732	7.541
													2	130.4	193.7	.3540		2	224.7	100.8	.4048		
													3	-27.5	34.8	.0128							
													4	21.0	15.3	.0044							
													5	-21.6	-12.4	.0040							

Chapter VI
APPLICATIONS OF EXPERIMENTAL METHOD IN USING HARMONIC COEFFICIENTS FROM HISTORICAL DATA, NORMAL AND GAMMA NOISE AND EFFECT OF PERIODIC AUTOCORRELATION

6-1 Methodology

In the previous chapter frequency distributions of the constituted X_i series were studied when the input was the normal noise with periodicities left in the mean and, with the exclusion of results given by Tables 5-1 and 5-2, in the standard deviation. Also, in four cases (Tables 5-4, 5-6, 5-7 and 5-8) the effects of an added autoregressive scheme were investigated.

In this chapter estimates of parameters from recorded historical data, used earlier by Quimpo [16] and others, are applied to produce the input data which has a variable distribution in its random component and the distribution of the independent residuals, ξ_i, is re-examined.

The stations pertaining to this study are as follows: (1) Tioga, Far Twins, New York; (2) Oconto, Gillette, Wisconsin; (11) Powell, Arthur, Tennessee; and (12) St. Maries, Lotus, Idaho.

A larger sample size of 40 years of daily data is used throughout and generated data are of the form given by

$$X_i = \mu_\tau + \sigma_\tau \varepsilon_i \qquad (6\text{-}1)$$

where $\tau = i$ (module ω) and

$$\varepsilon_i = \sum_{j=1}^{m} \alpha_j \varepsilon_{i-j} + \xi_i \quad . \qquad (6\text{-}2)$$

The standardized random series, ξ_i, have either normal or gamma distributions. The gamma variates are obtained by Eq. (4-3). This algorithm produces numbers which are serially uncorrelated but when a skewness, λ, greater than 2 is applied the skewness of the variates differs significantly from λ (the difference increases when λ increases) and also require standardization to zero mean and unit variance. In addition, the distribution is not strictly gamma. On the other hand, with the one-to-one transformation the algorithm is more economical than others and is therefore chosen for this study.

The order, m, of the autoregressive process and the coefficients, α_j, are those estimated from historical data. Harmonic coefficients from the original data are used but the numbers of significant harmonics are varied. Periodicities in the X_i series are then removed by inferring the numbers of harmonics and estimating the coefficients. The skewness and kurtosis of the independent residuals, ξ_i, were evaluated after removing the serial dependence through a fitted non-periodic autoregressive scheme.

For the second phase, a periodic autoregressive scheme as discussed in Chapter II is incorporated, with the means and standard deviations, μ_τ and σ_τ, constant for all values of τ. As before, a non-periodic scheme is used to remove serial dependence and the distribution statistics are evaluated for the cases of normal and gamma noise. For the gamma input, the skewness and kurtosis given in tables are those evaluated after the algorithm is applied.

6-2 Periodicities in the Mean and Standard Deviation

The results for a normal input with a variable number of harmonics in the mean and standard deviation, denoted by NHM and NHS respectively, are given in Table 6-1. If the input does not contain any periodicity in the standard deviation, as in the first three series, the ξ_i of the output are found to be normally distributed. The addition of one harmonic to the standard deviation of the input followed by the removal of periodicities in the mean and standard deviation also resulted in normally distributed independent residuals as seen in series 4 through 8 of Table 6-1. The results from the next five series, 9-13, show that, with the input subjected to two harmonics in the standard deviation, the kurtosis can exceed three. The skewness and kurtosis increase when NHS is increased further and this indicates inferential errors regarding the number of significant harmonics in the periodic standard deviation.

Although it appears that from the results of the 16th series only two harmonics in the standard deviation were significant, the experiments were repeated with NHS in the output increased arbitrarily as given by numbers within brackets in the next three series, 14-16. This resulted in a decrease in the skewness and kurtosis. Results by using the non-parametric method given by the 20th series show the possibility of obtaining a normal output and emphasize the uncertainties in the parametric method especially in the case of the two subsequent series for the input of which NHS = 5.

The experiment was repeated using gamma noise and the harmonics from station 2 data with NHM = 5 and NHS varying from 0 to 4 in the input. The results in Table 6-2 show, as in the case when normal noise was used, that differences between the input and output distributions tend to increase when NHS increases. However in this case there is a decrease in the skewness and kurtosis compared to the results in Table 6-1.

6-3 Periodicities in the Autoregressive Coefficients

The effects of periodicities in the autoregressive structure on the probability distribution of independent residuals formed by using a non-periodic autoregressive scheme are shown in Table 6-3 (normal noise) and Table 6-4 (gamma noise). The 3rd order autoregressive coefficients estimated from station 1 data in which Tao [20] found periodicities in the serial correlation coefficients were used to produce series 1 of Table 6-3. It appears that the estimated coefficients do not affect distributions. The other series were produced using the arbitrary harmonic coefficients representing periodicities in the serial correlation coefficients. When the amplitudes increase, the skewness and kurtosis tend to increase also. However, increasing the order of the process seems to have a normalizing effect.

To produce the results of Table 6-4 skewness coefficients of 1 and approximately 4 were applied to the ξ_i series. For low values of skewness and low amplitudes in the periodic autocorrelation coefficients, the distribution statistics are

Table 6-2. Simulated X_i Series Using Standardized Gamma Noise with Harmonics in the Mean and Standard Deviation and Non-Periodic Autoregressive Parameters Estimated from Station 2 Data. Skewness and Kurtosis of Estimated Independent Residuals, ξ_i, after Estimating and Removing Significant Harmonics and Non-Periodic Autoregressive Dependence.

	Simulated				Estimated	
	Number of Harmonics, ℓ					
Series	Mean	Std. Dev.	Skewness	Kurtosis	Skewness	Kurtosis
1	5	0	1.062	4.81	0.919	4.47
2	5	0	2.158	10.21	1.851	8.72
3	5	0	4.30	31.69	3.61	24.61
4	5	1	0.993	4.40	0.784	4.13
5	5	1	1.988	8.83	1.575	7.44
6	5	1	3.885	27.06	2.976	19.21
7	5	2	0.982	4.39	0.635	4.02
8	5	2	1.985	8.80	1.314	6.50
9	5	2	3.872	26.16	2.617	16.53
10	5	3	1.025	4.67	0.632	4.14
11	5	3	2.096	9.95	1.287	6.75
12	5	3	4.197	32.78	2.542	17.28
13	5	4	1.026	4.66	0.954	6.23
14	5	4	2.097	9.85	1.612	9.28
15	5	4	4.256	31.65	2.796	19.66

Table 6-3. Simulated X_i Series Using Standard Normal Noise and Periodic Autoregressive Scheme of Order m, m = 1,2,3. Means of Lag 1, 2 and 3 Periodic Serial Correlation Coefficients from 365 Values are 0.370, 0.513 and 0.404 Respectively (Station 1 Data). Series 1 is Based on Harmonics in Serial Correlation Coefficients of Station 1 Data. Skewness and Kurtosis of Estimated Independent Residuals, ξ_i, after Estimating and Removing Non-Periodic Autoregressive Dependence.

	Simulated								Estimated	
			Harmonics in Ser. Cor. Coefficients. Periodic Autoregressive Model, order m						Ind. Res. NM-periodic Autogressive Model	
			1st Coef.		2nd Coef.		3rd Coef.			
Series	m	j	A_j	B_j	A_j	B_j	A_j	B_j	Skewness	Kurtosis
1	3	1	.0307	-.0278	.0297	-.0381	.0318	-.0400	.028	3.00
		2	-.0226	.0174	.0043	.0230	-.0129	.0219		
		3	-.0086	-.0188	-.0136	-.0232	-.0136	-.0249		
		4	-.0031	-.0082	-.0102	.0026	-.0236	-.0019		
		5	.0017	-.0622	.0021	-.0190	.00284	-.0273		
		6	.0164	-.0293	.0338	.0076	.0339	.0053		
2	1	1	-.1	.1					0.289	3.75
3	2	1	-.1	.1	-.1	.1			0.077	3.26
4	3	1	-.1	.1	-.1	.1	-.1	.1	0.053	3.43
5	1	1	-.2	.2					0.607	4.92
6	2	1	-.2	.2	-.2	.2			0.319	4.00
7	3	1	-.2	.2	-.2	.2	-.2	.2	0.336	4.33
8	1	1	-.3	.3					0.725	5.84
9	2	1	-.3	.3	-.3	.3			0.610	4.96
10	3	1	-.3	.3	-.3	.3	-.3	.3	0.049	3.09

Table 6-4. Simulated X_i Series Using Standardized Gamma Noise and Periodic First Order Autoregressive Scheme. Means of Lag 1,2, and 3 Periodic Serial Correlation Coefficients from 365 Values are 0.370, 0.513 and 0.404 Respectively (Station 1 Data). Skewness and Kurtosis of Estimated Independent Residuals, ξ_i, after Estimating and Removing Non-Periodic Autoregressive Dependence.

	Simulated										Estimated	
				Harmonics in Ser. Cor. Coefficients. Periodic Autoregressive Model, order m							Ind. Res. Non-periodic Autoregressive Model	
				1st Coef.		2nd Coef.		3rd Coef.				
Series	m	Skewness	Kurtosis	A_1	B_1	A_1	B_1	A_1	B_1		Skewness	Kurtosis
1	1	1.06	4.82	-.1	.1						0.60	4.25
2	1	1.03	4.66	-.1	.1	-.1	.1				0.44	3.63
3	2	1.00	4.56	-.1	.1	-.1	.1	-.1	.1		0.79	4.67
4	2	4.31	31.70	-.1	.1						4.99	45.27
5	3	4.26	31.66	-.1	.1	-.1	.1				2.34	16.58
6	3	4.17	32.84	-.1	.1	-.1	.1	-.1	.1		3.10	26.07
7	1	0.99	4.40	-.2	.2						0.85	5.18
8	1	1.04	4.76	-.2	.2	-.2	.2				0.66	4.51
9	2	1.03	4.67	-.2	.2	-.2	.2	-.2	.2		1.29	6.82
10	2	3.89	27.06	-.2	.2						4.90	42.95
11	3	4.25	31.78	-.2	.2	-.2	.2				2.74	22.51
12	3	4.24	33.49	-.2	.2	-.2	.2	-.2	.2		4.20	37.56
13	1	0.98	4.39	-.3	.3						1.08	6.18
14	1	1.06	4.85	-.3	.3	-.3	.3				0.99	6.19
15	2	1.01	4.53	-.3	.3	-.3	.3	-.3	.3		1.56	8.61
16	2	3.87	26.10	-.3	.3						5.15	45.12
17	3	4.44	35.31	-.3	.3	-.3	.3				3.68	32.62
18	3	4.12	31.52	-.3	.3	-.3	.3	-.3	.3		4.74	44.48

generally lower, regardless of the order of the process. An increase in amplitudes usually results in an increase in the skewness and kurtosis of the ξ_i series as in the case of normal residuals, but the interactions between the various combinations of stastics seem to be highly complicated.

The conclusions of this part of the study are firstly that the existing methods of inferring periodicities in the standard deviation are inadequate as noted from the results in Chapter V. Further research is required on this aspect especially if the objective is to determine a more accurate distribution of the independent residuals. The alternative method of using the non-parametric method is not attractive because of the multiplicity of parameters that need to be estimated and the likelihood of incorporating the effect of large sampling errors into the mathematical models. Secondly, regarding the periodicities in the serial correlation coefficients, these do not seem to cause serious problems with respect to the estimation of probability distributions, provided that periodicities are not highly significant. If periodicities are found, higher-order autoregressive processes than indicated by the usual methods of inference seem to be the remedy, and periodicities need to be removed prior to the evaluation of the distribution statistics.

Chapter VII
CONCLUSIONS

The main conclusions of this study are as follows:

(1) The use of the periodic-stochastic model for hydrologic series to describe daily time series, with the annual periodicity in parameters, results often in a sharp-peaked curve of the estimated frequency distribution of the independent stochastic component in comparison with the population distribution which is normal (0,1). The frequency curve is also distorted when the population is non-normal. These deviations are caused by the modeling process and the sampling errors in estimating periodicities.

(2) The distorted distribution of the estimated independent residuals of stochastic components of a periodic-stochastic process indicates the limitations in current methods of inferring the numbers of significant harmonics and in the least-square estimation of Fourier coefficients of inferred harmonics.

(3) Inferential errors pertaining to the periodicity in the mean do not seem to have a serious effect on the distribution of independent residuals. Larger errors in the estimated distributions are caused by errors in the estimates of harmonics in the standard deviation.

(4) Periodicity which remains in the dependent stochastic component of the periodic-stochastic process leads to biased estimates of autoregressive coefficients.

(5) The bias is reduced by taking account of periodicities in the autocorrelation structure. However, these may not be highly significant and the greatest uncertainty lies in estimating periodicity in the standard deviation.

(6) A reduction in the number of significant harmonics and an improved fit is obtained through a logarithmic transformation. The weighted Fourier analysis also leads to some improvement but there may be undesirable effects particularly if the approach is used with the logarithmic transformation.

REFERENCES

1. Adamowski, K., 1971, Special Density of River Flow Time Series, Journal of Hydrology, Vol. 14, October 1971, pp. 43-50.

2. Adamowski, K., and Smith, F.K., 1972, Stochastic Generation of Rainfall, Journal of Hydraulics Division, American Society of Civil Engineers, Vol. 98, No. H411, November 1972, pp. 1935-1945.

3. Anderson, R.L., 1942, Distribution of the Serial Correlation Coefficient, Annals of Mathematical Statistics, Vol. 13, 1942, pp. 1-13.

4. Box, E.P.G., and Jenkins, G.M., 1971, Time Series Analysis Forecasting and Control. Holden-Day Co., San Francisco, 553 p.

5. Chin, William Q., 1973, A Deterministic-Stochastic Model of Long-Term Climatic Changes. Ph.D. Dissertation, Colorado State University, Ft. Collins, Colorado, pp. 157-169.

6. Dickenson, R., 1971, On the Relationship of Variance Spectra of Temperature to Large Scale Atmospheric Circulation. Journal of Applied Meteorology, Vol. 10, No. 2, Feb. 1971, pp. 157-169.

7. Dwass, Meyer, 1970, Probability Theory and Applications. W.A. Benjamin Inc., New York, pp. 245-249.

8. Fiering, M.B., and Jackson, B.B., 1971, Synthetic Streamflow. American Geophysical Union, Water Resources Monograph No. 1, McGregor and Weiner Inc., Washington, D. C., 210 p.

9. Fisher, R.A. 1929, Tests of Significance in Harmonic Analysis, Proceedings of the Royal Society, Series A, Vol. 125, pp. 55-59.

10. Hastings, C., 1955, Approximations for Digital Computers. Princeton University Press. Princeton, New Jersey, 195 p.

11. Jenkins, G.M., and Watts, D.G., 1969, Spectral Analysis and Its Applications. Holden-Day, San Francisco, 525 p., pp. 149-152.

12. Jones, H.J., and Brelsford, W.M., 1967, Time Series with Periodic Structure. Biometrica, Vol. 54,1967, pp. 403-410.

13. Kendall, M.G., and Stuart, A., 1966, The Advanced Theory of Statistics, Vol. 3. Haffnner, New York, pp. 419-420.

14. Landsberg, H.E., J.M. Mitchell, and H.L. Crutcher, 1959, Power Spectrum Analysis of Climatological Data for Woodstock College, Maryland. Monthly Weather Review, Vol. 87, No. 8, 1959, pp. 283-298.

15. Quenoille, M.H., 1949, A Large Sample Test for the Goodness of Fit in Autoregressive Schemes. Journal of the Royal Statistical Society. Vol. 10, pp. 123-129.

16. Quimpo, R.G., 1967, Stochastic Model of Daily River Flow Sequences. Hydrology Paper 18, Colorado State University, Ft. Collins, Colorado, pp. 11-13.

17. Roesner, L.A., and V.M. Yevjevich, 1966, Mathematical Models for Time Series of Monthly Precipitation and Monthly Runoff. Hydrology Paper 15, Colorado State University, Ft. Collins, Colorado, 34 p.

18. Salaz-la Cruz, J.D., and V. Yevjevich, 1972, Stochastic Structure of Water Use Time Series, Hydrology Paper 52, Colorado State University, Ft. Collins, Colorado.

19. Schuster, A., 1898, On the Investigation of Hidden Periodicities with Application to a Supposed 26-Day Period of Meteorological Phenomenon, Terr. Mag., V. 3, No. 1, pp. 13-41, Journal of Geophysical Research, Washington, D. C.

20. Tao, Pen-Chih, 1973, Distribution of Hydrologic Independent Stochastic Components. Ph.D. Dissertation, Colorado State University, Ft. Collins, Colorado.

21. Walker, G.T., 1925, On Periodicity, Royal Meteorological Society Journal, V. 51, pp. 337-346.

22. Yevjevich, V.M., 1964, Fluctuations of Wet and Dry Years. Part II, Analysis by Serial Correlogram. Hydrology Paper 4, Colorado State University, Ft. Collins, Colorado.

23. Yevjevich, V.M., 1972, Structural Analysis of Hydrologic Time Series. Hydrology Paper 56, Colorado State University, Ft. Collins, Colorado, pp. 18-55.

24. Yevjevich, V.M., 1972, Stochastic Processes in Hydrology. Water Resources Publications, Ft. Collins, Colorado, 260 p., pp. 35-43.

Key Words: Hydrologic time series, harmonics of periodic parameters, misestimation of harmonics, residual stochastic series, structural analysis of time series.

Abstract: The periodic-stochastic model, used to describe the structure of historic hydrologic time series, is examined in the light of finding that it distorts the distribution of residuals, or of the time-independent stochastic component under certain circumstances. In general this distortion caused the distribution of an independent stochastic component to have a sharper peak, and sometimes it appears to follow a double branch exponential distribution function. The apparent cause

Key Words: Hydrologic time series, harmonics of periodic parameters, misestimation of harmonics, residual stochastic series, structural analysis of time series.

Abstract: The periodic-stochastic model, used to describe the structure of historic hydrologic time series, is examined in the light of finding that it distorts the distribution of residuals, or of the time-independent stochastic component under certain circumstances. In general this distortion caused the distribution of an independent stochastic component to have a sharper peak, and sometimes it appears to follow a double branch exponential distribution function. The apparent cause

Key Words: Hydrologic time series, harmonics of periodic parameters, misestimation of harmonics, residual stochastic series, structural analysis of time series.

Abstract: The periodic-stochastic model, used to describe the structure of historic hydrologic time series, is examined in the light of finding that it distorts the distribution of residuals, or of the time-independent stochastic component under certain circumstances. In general this distortion caused the distribution of an independent stochastic component to have a sharper peak, and sometimes it appears to follow a double branch exponential distribution function. The apparent cause

Key Words: Hydrologic time series, harmonics of periodic parameters, misestimation of harmonics, residual stochastic series, structural analysis of time series.

Abstract: The periodic-stochastic model, used to describe the structure of historic hydrologic time series, is examined in the light of finding that it distorts the distribution of residuals, or of the time-independent stochastic component under certain circumstances. In general this distortion caused the distribution of an independent stochastic component to have a sharper peak, and sometimes it appears to follow a double branch exponential distribution function. The apparent cause

for distortions is the failure of the least-squares method to accurately estimate amplitudes and phases of harmonics in periodic parameters of historic data. The unremoved and/or misestimated dependence of the autoregressive type in stochastic component was found to also affect the inferred distribution of residuals. The investigation method used is by generating new samples of given sizes and properties by the classical Monte Carlo method.

<u>Reference</u>: Bullard, K. L., V. Yevjevich and N. Kottegoda, <u>Colorado</u> State University, Hydrology Paper No. 88 (November 1976), Effect of Misestimating Harmonics in Periodic Hydrologic Parameters.

for distortions is the failure of the least-squares method to accurately estimate amplitudes and phases of harmonics in periodic parameters of historic data. The unremoved and/or misestimated dependence of the autoregressive type in stochastic component was found to also affect the inferred distribution of residuals. The investigation method used is by generating new samples of given sizes and properties by the classical Monte Carlo method.

<u>Reference</u>: Bullard, K. L., V. Yevjevich and N. Kottegoda, <u>Colorado</u> State University, Hydrology Paper No. 88 (November 1976), Effect of Misestimating Harmonics in Periodic Hydrologic Parameters.

for distortions is the failure of the least-squares method to accurately estimate amplitudes and phases of harmonics in periodic parameters of historic data. The unremoved and/or misestimated dependence of the autoregressive type in stochastic component was found to also affect the inferred distribution of residuals. The investigation method used is by generating new samples of given sizes and properties by the classical Monte Carlo method.

<u>Reference</u>: Bullard, K. L., V. Yevjevich and N. Kottegoda, <u>Colorado</u> State University, Hydrology Paper No. 88 (November 1976), Effect of Misestimating Harmonics in Periodic Hydrologic Parameters.

for distortions is the failure of the least-squares method to accurately estimate amplitudes and phases of harmonics in periodic parameters of historic data. The unremoved and/or misestimated dependence of the autoregressive type in stochastic component was found to also affect the inferred distribution of residuals. The investigation method used is by generating new samples of given sizes and properties by the classical Monte Carlo method.

<u>Reference</u>: Bullard, K. L., V. Yevjevich and N. Kottegoda, <u>Colorado</u> State University, Hydrology Paper No. 88 (November 1976), Effect of Misestimating Harmonics in Periodic Hydrologic Parameters.

LIST OF PREVIOUS 25 PAPERS

No. 63 Mechanics of Soil Erosion From Overland Flow Generated by Simulated Rainfall, by Mustafa Kilinc and Everett V. Richardson, September 1973.

No. 64 Area-Time Structure of the Monthly Precipitation Process, by V. Yevjevich and Alan K. Karplus, August 1973.

No. 65 Almost-Periodic, Stochastic Process of Long-Term Climatic Changes, by William Q. Chin and Vujica Yevjevich, March 1974.

No. 66 Hydrologic Effects of Patch Cutting of Lodgepole Pine, by Thomas L. Dietrich and James R. Meiman, April 1974.

No. 67 Economic Value of Sediment Discharge Data, by Sven Jacobi and Everett V. Richardson, April 1974.

No. 68 Stochastic Analysis of Groundwater Level Time Series in Western United States, by Albert G. Law, May 1974.

No. 69 Efficient Sequential Optimization in Water Resources, by Thomas E. Croley II, September 1974.

No. 70 Regional Water Exchange for Drought Alleviation, by Kuniyoshi Takeuchi, November 1974.

No. 71 Determination of Urban Watershed Response Time, by E. F. Schulz, December 1974.

No. 72 Generation of Hydrologic Samples, Case Study of the Treat Lakes, by V. Yevjevich, May 1975.

No. 73 Extraction of Information on Inorganic Water Quality, by William L. Lane, August 1975.

No. 74 Numerical Model of Flow in Stream-Aquifer System, by Catherine E. Kraeger Rovey, August 1975.

No. 75 Dispersion of Mass in Open-Channel Flow, by William W. Sayre, August 1975.

No. 76 Analysis and Synthesis of Flood Control Measures, by Kon Chin Tai, September 1975.

No. 77 Methodology for the Selection and Timing of Water Resources Projects to Promote National Economic Development, by Wendim-Agegnehu Lemma, August 1975.

No. 78 Two-Dimensional Mass Dispersion in Rivers, by Forrest M. Holly, Jr., September 1975.

No. 79 Range and Deficit Analysis Using Markov Chains, by Francisco Gomide, October 1975.

No. 80 Analysis of Drought Characteristics by the Theory of Run, by Pedro Guerrero-Salazar and Vujica Yevjevich, October 1975.

No. 81 Influence of Simplifications in Watershed Geometry in Simulation of Surface Runoff, by L. J. Lane, D. A. Woolhiser and V. Yevjevich, January 1976.

No. 82 Distributions of Hydrologic Independent Stochastic Components, by Pen-chih Tao, V. Yevjevich and N. Kottegoda, January 1976.

No. 83 Optimal Operation of Physically Coupled Surface and Underground Storage Capacities, by Dragoslav Isailovic, January 1976.

No. 84 A Salinity Management Strategy for Stream-Aquifer Systems, by Otto J. Helweg, and J. Labadie, February 1976.

No. 85 Urban Drainage and Flood Control Projects Economic, Legal and Financial Aspects, by Neil S. Grigg, Leslie H. Botham, Leonard Rice, W. J. Shoemaker, and L. Scott Tucker, February 1976.

No. 86 Reservoir Capacity for Periodic-Stochastic Input and Periodic Output, by Kedar Nath Mutreja, September 1976.

No. 87 Area-Deficit-Intensity Characteristics of Droughts, by Norio Tase, October 1976.