GENERATION OF HYDROLOGIC SAMPLES CASE STUDY OF THE GREAT LAKES

by
Vujica Yevjevich

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

72

GENERATION OF HYDROLOGIC SAMPLES CASE STUDY OF THE GREAT LAKES

by
Vujica Yevjevich*

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO

May, 1975 No. 72

*Professor-in-Charge of Hydrology and Water Resources Program, Department of Civil Engineering, Colorado State University, Fort Collins, Colorado 80523

TABLE OF CONTENTS

ACKNOWLEDGEMENT

This paper is an application of various methods and techniques developed and/or perfected under the US National Science Foundation-Grant GK-31521X at Colorado State University, "Stochastic Processes in Water Resources," with the writer of this paper the principal investigator. This continuous support for research and publishing of results by NSF is gratefully acknowledged.

The case study of generating the new samples of hydrologic processes of the Great Lakes is the result of the agreement No. DAC-23-72-C-0042 between the North Central Division of the U.S. Army Corps of Engineers, Chicago, Illinois, and the writer of this paper. In carrying out this study, the writer has cooperated with the following individuals involved in the International Great Lakes Levels Board (IGLLB) study: Dr. Lewis H. Blakey, Chief, Engineering Division, and Mr. Joseph Raoul, Jr., Hydraulic Engineer, Water Control Center, of the North Central Division, U. S. Army Corps of Engineers, Chicago, Illinois, and Dr. Geoff Kite, at the time of these studies, Department of Environment of Canada, Ottawa, Canada. Their help in initiating the study and in reviewing the reports is acknowledged.

Dr. Jose Salas-La Cruz, at present with the Ministry of Agriculture, Lima, Peru, helped the writer during the investigations, while he was the postdoctoral fellow at Colorado State University. Mr. W. W. Burt, Computer Consultant, helped in detailed computer programming. Their help is duly acknowledged.

ABSTRACT

The subject matter of this paper is the generation of 20 samples, each 50 years long, of three variables related to water inputs into and retardation of flows in connecting channels of the Great Lakes: (1) Mean monthly net basin water supplies of five lakes; (2) Mean quarter-monthly net basin water supplies of two smaller lakes (Ontario and Erie); and (3) Flow retardations in connecting channels because of freezing and weed effects. The methods of obtaining the net basin water supplies and channel flow retardations are described as developed by the Great Lakes various committees. For each of series of the above three variables, first the tests of homogeneity (trends) in data have been performed, basically by using the t-statistic and Student t-distribution. The Lake St. Clair mean monthly net basin supplies have been found to have a trend. Also the two connecting channels (the St. Mary River, the St. Clair-Lake St. Clair-Detroit Rivers Systems) had decreasing trends in flow retardation series. All series are studied further with the trends in parameters removed.

Periodic parameters in all three variables are found to be the mean and the standard deviation. The autocorrelation coefficients and the skewness coefficient are found not to be periodic. The stochastic components, after the periodic mean and standard deviation are identified and removed, are found to be greatly autocorrelated. The simple first- and second-order autoregressive linear models are found sufficient to describe these dependences. For the resulting white noise (independent, identically distributed stochastic components) of all series, the three-parameter lognormal distributions have been found as good approximations.

The principal component analysis has been used in generating the new samples of the mean monthly net basin supplies. The approach of generating first the monthly values, and then superimposing the generated four differences of mean quarter-monthly values was shown to be difficult to apply, because both the conservation of mass (sum of four differences to be zero) and the autoregressive model could not be satisfied simultaneously. For small number of series of the same variable, the sample correlation between the independent stochastic components was used in generating new samples.

PART I
INTRODUCTION

This part relates to two aspects of generation of new samples of a set of station series, namely (1) The reduction of periodic-stochastic time processes to normal, independent, identically distributed random variables, space dependent, as an n-dimensional process, transformed to n principal components; and (2) The description of the Great Lakes data used in simulating the various series of net basin supplies of these lakes and flow retardations in connecting channels.

Chapter 1

GENERAL APPROACH TO GENERATION OF SAMPLES OF A SET OF STATION SERIES

1.1 Basic Character of Hydrologic Area-Time Processes. In sampling area-time hydrologic processes by various data collection services, the most current practice is in approximating the areal variation by a set of points, and in observing the variation of particular variables in time either as continuous recordings, discrete time observations, or the cumulative values over the specified intervals. The general area-time process $x\{X,Y;t\}$, as a three-dimensional process, is then separated in n time processes, such as: $x_1\{X_1,Y_1,t\}$; $x_2(X_2,Y_2;t)$; ...; $x_n\{X_n,Y_n;t)$, where (X_i,Y_i) are the coordinates of n points, $i = 1,2,\ldots,n$, and t are either discrete intervals or continuous time.

The basic character of nearly all hydrologic processes, for small Δt (say Δt smaller than a year), is that they are periodic-stochastic processes dependent over the area. Mostly the station random variables $x_1,x_2,\ldots,x_n$ are of mixed distributions, because the parameters (mean, variance, autocovariances, skewness, etc.) vary periodically over the year (or in case of Δt smaller than a day, also often over the day). These mixed variables most often have skewed distributions with a kurtosis coefficient different from three of the normal distribution.

1.2 Generation of New Samples. To simulate m samples each of the size N for all n stations, by preserving the time structure and areal dependence, if n is not too small (say five or more), the best presently available approach is to use the stationary multivariate normal distributions of n identically distributed components, time independent and areally dependent, and the principal component analysis. By removing the periodicity in parameters of time series, the stationary dependent stochastic components are obtained. By inferring the proper dependence model for the stochastic component, the independent, identically distributed process in time (the noise) can be singled out. If this noise is not normally distributed, a further transformation (logarithmic, cubic root, and similar) may be used to arrive as close as practically feasible to normal independent identically distributed stochastic components along the time intervals, as the white noise. The n mutually dependent components then represent the multivariate normal distribution as the starting information for the use of principal components analysis.

By using the n x n covariance matrix, or for standardized normal variables the correlation matrix, the well known procedures of transformation of n mutually dependent normal variables to n mutually independent normal variables, or the principal components, can be used. Computer oriented procedures and programs are available currently for this analysis.

For principal components as the normal, identically distributed random variables, independent both in time and in area, it is then simple to generate by the Monte Carlo or experimental method the m samples, each of the size N (or several sizes N_i if this is needed), for each of n components. The inverse process then transforms the generated samples of principal components into the multivariate normal components, dependent among themselves but independent in time. This means a preservation of areal dependence specified by the correlation matrix. Applying to each component the corresponding time dependence model, inferred in the structural analysis of time series, the n time dependent stationary stochastic components are produced. The superposition of inferred functions for periodic parameters then produces the n components, which preserve both the time structure and areal dependence.

The method described has been used to generate the net basin supplies and channel flow retardations for the system of Great Lakes.

1.3 Procedure Followed. The three parts, II, III and IV, which relate to generation of samples for monthly net basin supplies to five of Great Lakes, quarter-monthly supplies to two lakes, and the flow retardations in channels connecting the lakes, because of the winter ice cover, explain the procedures used in details.

Chapter 2

DESCRIPTION OF GREAT LAKES AND VARIABLES TO GENERATE

2.1 The System of Great Lakes. The data on Great Lakes in this paper and various descriptions are taken from official reports [1 through 6] and published papers. The total area of Great Lakes is about 95,000 square miles, with the drainage area approximately 203,000 square miles. The basin map is given in Fig. 2-1. Principal hydrologic data for five of the Great Lakes are shown in Table 2-1.

The immense storage capacity of the lakes represents a large natural regulating water system. The ratios of the maximum to minimum flows at the lake outlets are only two to three. Lake Superior, completely regulated, is the uppermost and largest of the Great Lakes with the outflow through the St. Mary River into Lake Huron. Lakes Michigan and Huron are at the same level since they are connected by the broad and deep Straits of Mackinac. They are hydrologically treated as one lake. The outflow from these lakes is through the St. Clair River, then Lake St. Clair and the Detroit River into Lake Erie. The outlet from Lake Erie is through the Niagara River into Lake Ontario Lake Ontario, the lowest of the Great Lakes, is the smallest. Following the construction of the St. Lawrence Seaway and Power projects, the outflows from Lake Ontario are also regulated.

2.2 Regulation of Great Lakes. At the request of the Governments of Canada and the United States, the

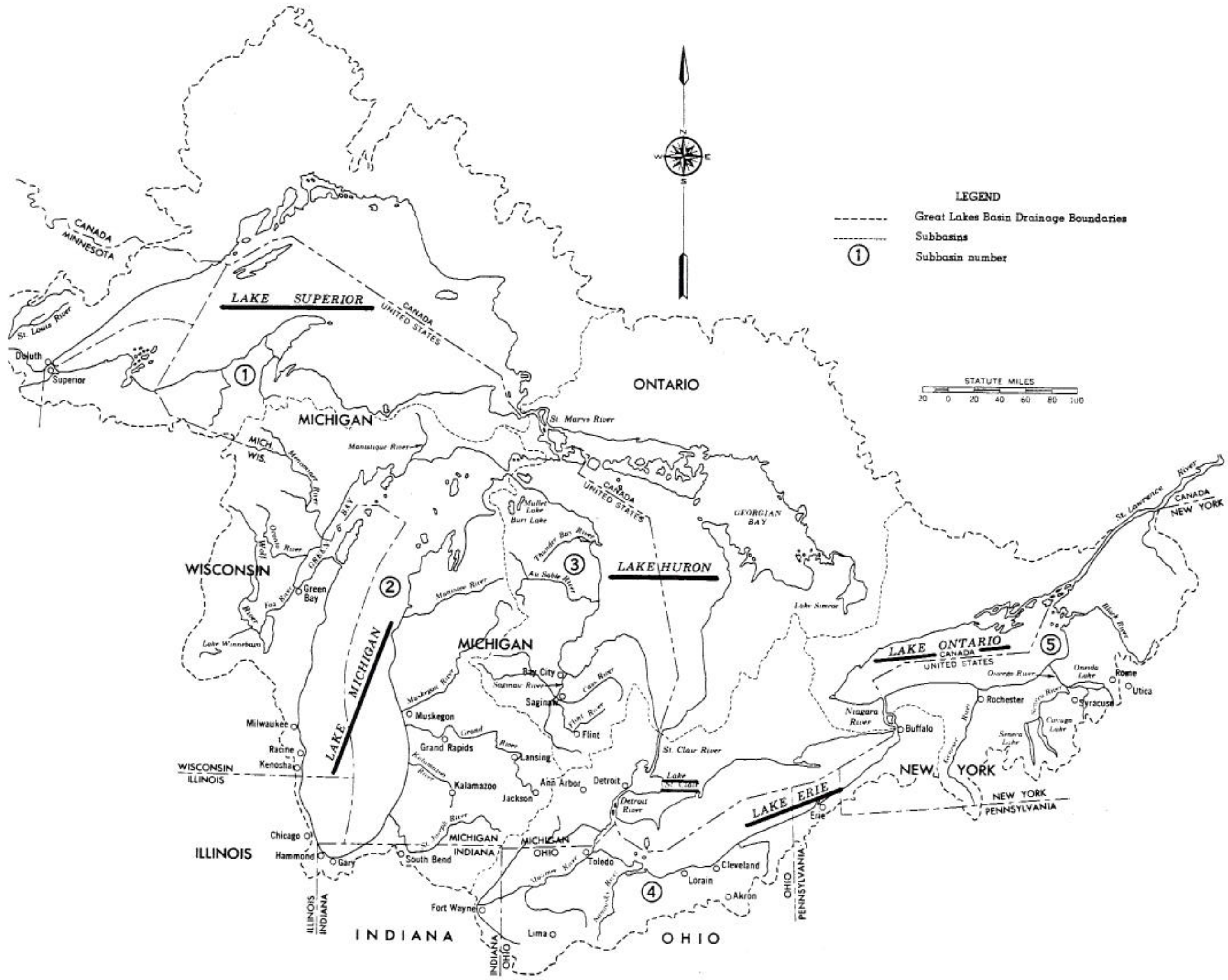

Fig. 2-1 The Great Lakes basin, with the subbasins (The map of the Great Lakes Basin Commission).

TABLE 2-1
DATA ON THE GREAT LAKES (1860-1972), ACCORDING TO IGLLB STUDY

Lake	Drainage Area		Storage Capacity Per Foot of Stage of Lake (cfs for one month)	Average Elevation (above m.s.l.) (IGLD,1955)*	Range of Monthly Mean Stage (ft.)	Outlet River	Mean Outflow	
	Land Area (Square Miles)	Water Surface Area (Square Miles)					(cfs)	Depth on Total Drainage Basin (inches)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Superior	49,300	31,700	337,000	600.4	3.8	St. Marys	75,400	12.6
Michigan-Huron	45,600 51,800	22,300 23,000	481,000	578.7	6.6	St. Clair Lake St. Clair	187,900	11.6
St. Clair	6,100	400	5,000	573.1	5.8	Detroit	188,900	
Erie	23,600	9,900	105,000	570.4	5.3	Niagara	202,300	10.6
Ontario	27,200	7,600	80,000	244.6	6.6	St. Lawrence	239,700	11.1
TOTALS	201,500	94,600	1,003,000					

*International Great Lakes datum, 1955 - elevation in feet above Father-Point, Quebec.

International Joint Commission has undertaken an investigation to determine whether it would be feasible and in the public interest to regulate further the levels of the Great Lakes so as to reduce the extremes of stage which have been experienced in the past and bring about a more beneficial range of stage for the various water users. The Commission established the International Great Lakes Levels Board on December 2, 1964 to undertake, through appropriate agencies in Canada and the United States, the necessary investigations and studies and to advise the Commission on all matters which it must consider in reporting on this matter.

In 1965 the Working Committee, which was established by the Levels Board, appointed a Regulation Subcommittee composed, in part, of members of the U. S. Lake Survey, Corps of Engineers and the Inland Waters Branch, Canada Department of Energy, Mines and Resources. The Regulation Subcommittee, and in particular, these two agencies, were assigned the task of computing and coordinating the water supply data required for the development of regulation plans and to derive the basis of comparison for these plans. Preliminary water supplies and comparison data, covering the period January 1900 through December 1964, were issued in April 1967. Subsequently, final water supplies and the basis of comparison data, covering the period January 1900 through December 1967, have been developed and coordinated.

2.3 Methods Employed to Obtain the Final Data. The historic or recorded data from which the required water supplies for the regulation plans have been developed, were themselves developed, in part, by another international committee known as the Coordinating Committee on Great Lakes Basic Hydraulic and Hydrologic Data. The Committee developed data on water levels, river flows and physical characteristics of the Great Lakes system.

Because of their large areas, the levels of Lakes Superior and Michigan-Huron respond to changes in outflow much more slowly than do the levels of Lakes Erie and Ontario. The lake regulation studies are conducted on a monthly basis for Lakes Superior and Michigan-Huron and on a quarter-monthly basis for Lakes Erie and Ontario. Monthly data are also used for Lake St. Clair.

The quarter-monthly periods consist of seven or eight days sub-divided as:

Quarter Month	Months of 28 or 29 days	Months of 30 or 31 days
First	1-7	1-8
Second	8-14	9-15
Third	15-21	16-23
Fourth	22-28 or 29	24-30 or 31

The data used for obtaining net basin water supplies and winter and weed retardations are: beginning-of-period lake levels, recorded outflows, and recorded diversions.

2.4 Net Basin Supplies. The net basin supply is a term used to describe the water which a lake receives from precipitation on both its surface and its drainage basin less the net evaporation and condensation on the lake surface. Some of these factors cannot be determined accurately. The net basin supplies were computed by employing reliable lake level and flow records for the required monthly and quarter-monthly periods. The relationship used is

$$\Delta S = P + R + U - E + I - O \pm D \qquad (2\text{-}1)$$

where ΔS = the change in amount of water stored in the lake (positive if supplies exceed outflows; negative if outflows exceed supplies), P = the precipitation on the lake's surface, R = the runoff from the lake's land drainage area, U = the ground-water contribution (considered positive in the aggregate), E = the evaporation from the lake's surface (net of evaporation and condensation), I = the inflow from the lake above, O = the outflow from the lake through its natural outlet, and D = the diversion (positive if into lake; negative if out of lake).

The changes in storage, inflow, outflow and diversion are determinable directly from reliable lake level and flow records, while the precipitation, runoff, ground-water contribution and evaporation cannot be determined accurately with presently available data and techniques. The first four terms of the right-hand side of the equation are combined in a single term which is called the net basin supply to the lake, Q, and the equation written as:

$$Q = \Delta S + O - I \mp D , \qquad (2\text{-}2)$$

with all terms expressed in units of cubic feet per second for the period. The sum of ΔS and O (including any outflow diversion) represents the net total supply to the lake.

The distribution of average water supply to lake in percent of average outflow is given in Table 2-2. The values for changes in storage, outflows and inflows are determined directly from the recorded values of lake levels, river flows and diversions.

TABLE 2-2
DISTRIBUTION OF AVERAGE WATER SUPPLY TO LAKE
IN PERCENT OF AVERAGE OUTFLOW

	Lake Superior	Lake Michigan-Huron	Lake Erie	Lake Ontario
Inflow from upstream lake (I)	0	46	86	86
Precipitation on lake surface (P)	88	59	12	8
Evaporation from lake surface (E)	-55	-57	-13	-7
Net (P-E)	+33	+2	-1	+1
Runoff from land basin	62	49	12	13
Percent of total outflow accounted for	95	97	97	100

2.5 Winter and Weed Retardations. The freezing over and weed retardations for connecting channels are determined as follows. The level of Lake Superior and the outflow through the St. Marys River are regulated by the International Lake Superior Board of Control by a dam at the head of the St. Mary Rapids. Under present regulation conditions, the winter effect on the discharges is virtually zero for all months.

Lake St. Clair normally freezes over in the early winter. Subsequently, heavy runs of Lake Huron ice in the St. Clair River jam in the channels through the St. Clair Flats and upstream, to the extent that the river flow is reduced. Ice jams seldom occur in the Detroit River, although the river is frequently frozen over in its lower reaches. Winter retardation in the St. Clair River was estimated by subtracting the coordinated recorded flow from the corresponding discharge based on the two-gauge open-water stage-discharge relationship for the Harbour Beach-Grosse Pte. reach.

In order to avoid ice problems in the Niagara River, an ice boom has been placed across the head of the Niagara River by the Power Entities for each winter season commencing in 1964. The presence of this ice boom has reduced the retardation of Niagara River flow by ice to a very small amount which can be considered insignificant. However, since the outlet conditions of 1953 were adopted as the basis to be used for comparing regulation plans, average winter retardation was assumed for the Niagara River over the period of record.

Lake Ontario is at present regulated by the International St. Lawrence River Board of Control. Therefore, direct estimates of winter retardation could be made only for the period prior to commencement of the St. Lawrence project in 1955. For the period 1900-1955 retardation values were calculated as the difference between the outflows resulting from the open-water Oswego stage-discharge relationship, and the recorded outflows.

Reduction in the winter flow at the outlet of Lake St. Louis was calculated directly as the difference between the discharge derived from the appropriate open-water stage-discharge curves and the recorded discharge.

The determination of winter flow retardation values are based on the following relationship:

$$I = Q_A - Q_C \qquad (2\text{-}3)$$

where I = the winter flow retardation in cfs-months, Q_A = the adopted flow through connecting channel in cfs-months and Q_C = the flow computed from the open water stage-discharge relationship in cfs-months. In the aforementioned relationship Q_A represents the mutually accepted values among the agencies of both U. S. and Canadian Governments and are tabulated in the report on coordinated basic data.

2.6. Generation of Samples of Net Basin Supplies and Flow Retardations. The U. S. Army Corps of Engineers, North Central Division, Chicago, Illinois, and the writer of this paper concluded an agreement for the writer to analyze data and to generate 20 samples each 50 years long of:

(1) Monthly net basin supplies of five lakes;

(2) Quarter-monthly net basin supplies of two lakes; and

(3) Winter flow retardations of the outflows at four connecting channels.

These three items are covered in Parts II, III and IV of this paper, respectively. The two reports, submitted to North Central Division of U. S. Army Corps of Engineers, Chicago, Illinois, namely in August 1972 and September 1972, respectively, have served as the basic material in shaping this paper, with modifications.

The following data, in the form of punched cards have been supplied by the Corps of Engineers for the generation of the new samples:

1. Lake Superior monthly mean net basin supply
2. Lake Michigan/Huron monthly mean net basin supply
3. Lake St. Clair monthly mean net basin supply
4. Lake Erie quarter-monthly mean net basin supply
5. Lake Erie monthly mean net basin supply
6. Lake Ontario quarter-monthly mean net basin supply
7. Lake Ontario monthly mean net basin supply
8. Winter retardation of the outflow from Lake Michigan/Huron
9. Winter retardation of the outflow from Lake St. Clair
10. Winter retardation of the outflow from Lake Ontario
11. Winter retardation of the outflow from Lake St. Louis

Dr. Jose Salas-La Cruz assisted the writer in all computations of sample generations; this help is appreciated and acknowledged.

PART II
GENERATION OF SAMPLES OF MEAN MONTHLY NET BASIN SUPPLIES OF GREAT LAKES

This part refers to the structural analysis of historic data, made available by the U. S. Army Corps of Engineers, North Central Division, Chicago, of the five mean monthly net basin supplies--in further text abbreviated as NBS--of Lake Ontario, Lake Erie, Lakes Michigan/Huron, Lake Superior and Lake St. Clair, and the generation by the experimental (Monte Carlo) method of 20 samples each 50 years long of each of these five mean monthly net basin supplies. It contains the following: tests of homogeneity of the above five NBS time series, structural analysis of these series found or made homogeneous with their mathematical description, generation of samples for four series (all except NBS of Lake St. Clair), analysis of generated samples, development of a multiple linear regression for Lake St. Clair NBS series to the other four series, generation of samples for NBS series of Lake St. Clair, and analysis of these generated samples.

Chapter 3

TESTS OF HOMOGENEITY OF NBS SERIES

3.1 Selection of Test Statistics. Tests of homogeneity are carried out by the split-sample approach in ascertaining whether differences between the means of the two unequal subsamples (36 and 33 years in the case of four series) are or are not significantly different from zero on the 95 percent probability level of significance. Only if the probability is less than 5 percent that a difference is greater than the critical value of these differences are the two subsample means considered not to be from the same population, or the series considered to be nonhomegeneous.

The t-statistic is used for testing whether the difference of the two means, $\overline{x}_1$ and $\overline{x}_2$, is significant with

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}} , \qquad (2\text{-}1)$$

and

$$s = \sqrt{\frac{\sum_{i=1}^{n_1} (x_i - \overline{x}_1)^2 + \sum_{j=1}^{n_2} (x_j - \overline{x}_2)^2}{n_1 + n_2 - 2}} , \qquad (2\text{-}2)$$

where n_1 and n_2 are subsample sizes, x_i are values of the series in the n_1 subsample and x_j in the n_2 subsample. This t has the Student t-distribution. The critical value, t_c, for the significance probability level of 95 percent was then taken from the Student t-distribution tables.

Similar equations to Eqs. 2-1 and 2-2 are also used for testing whether the differences of standard deviations s_1 and s_2 of two subsamples are significantly different, with $\overline{x}_1$ and $\overline{x}_2$ in Eq. 2-1 replaced by $\overline{s}_1$ and $\overline{s}_2$, and s of Eq. 2-1 is the value of s of Eq. 2-2. These $\overline{s}_1$ and $\overline{s}_2$ are the averages of s of each year for the two subsamples.

3.2 Results of Homogeneity tests. Table 2-1 gives the results of tests of homogeneity in the subsample means, with only the series of monthly mean NBS of Lake St. Clair found to be nonhomogeneous. The monthly mean NBS series for Lake St. Clair is then corrected with $\overline{x}$ = 5.24 of the last 26 years of records also being the mean for the first 43 years. In fact all values of the first subsample are increased by the difference $\overline{x}_2 - \overline{x}_1$. The new series of NBS for Lake St. Clair is then further analyzed as a homogeneous series.

TABLE 2-1
ANALYSIS OF DIFFERENCE IN TWO SUBSAMPLE MEANS

	Subsample Sizes		Statistic t(95%)		
Lake	n_1	n_2	From t-Tables	Computed	Change in the Mean
Ontario	36	33	2.0	0.299	No
Erie	36	33	2.0	0.635	No
Superior	36	33	2.0	1.525	No
Michigan	36	33	2.0	0.866	No
St. Clair	43	26	2.0	4.477	Yes

The tests of homogeneity in the standard deviation of the NBS series are presented in Table 2-2 in a similar manner to the presentation of results of tests of the two subsample means of Table 2-1. All five series are found to be homogeneous in the standard deviation.

Usually and under the natural conditions, when a hydrologic series has no significant trend or slippage (positive or negative jump) in the mean and the standard deviation, the entire series may be safely inferred as being homogeneous. No tests were considered necessary to determine whether the difference of two subsample statistics of other parameters are or are not significantly different from zeros,or for such periodic parameters as monthly means m_τ, monthly standard deviations s_τ, monthly autocorrelation coefficients $r_{k,\tau}$ of the stochastic dependent component, or monthly skewness coefficient ${}_\tau C_s$ of the independent stochastic component.

TABLE 2-2
ANALYSIS OF DIFFERENCE IN TWO SUBSAMPLE STANDARD DEVIATIONS

	Subsample Sizes		Statistic t(95%)		
Lake	n_1	n_2	From t-Tables	Computed	Change in the St. Deviation
Ontario	36	33	2.0	0.598	No
Erie	36	33	2.0	0.992	No
Superior	36	33	2.0	1.748	No
Michigan	36	33	2.0	0.269	No
St. Clair	43	26	2.0	1.056	No

The four major historic series of mean monthly NBS (Ontario, Erie, Superior, Michigan) are used in the further investigations as made available by the Corps of Engineers, while the Lake St. Clair NBS series is used with a mean of the second subsample of the last 26 years of data, with the monthly average $\bar{x} = 5.24$.

Figures 2-1 through 2-5 give the series of the mean annual NBS at the lake outlet flow gauging stations (Fig. 2-1 - Ontario, Fig. 2-2 - Erie, Fig. 2-3 - Superior, Fig. 2-4 - Michigan, and Fig. 2-5 - St. Clair). These figures also show the two selected subsamples in testing whether the difference of their two means or the difference of their two standard deviations are or are not significantly different from zeros. A visual inspection also shows that the first four lakes, Figs. 2-1 through 2-4, either do not show a trend in mean annual NBS, or that the trend is so mild that it can be safely attributed to sampling fluctuations.

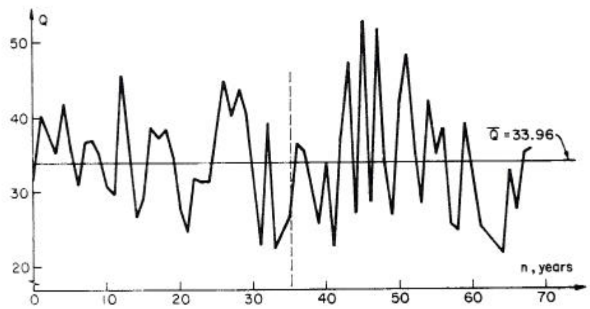

Fig. 2-1 Annual mean net basin supply for Lake Ontario.

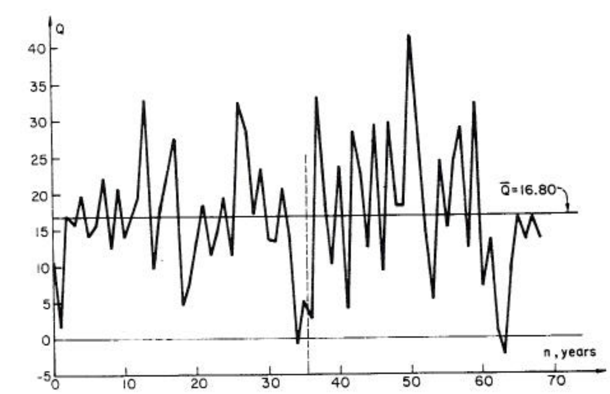

Fig. 2-2 Annual mean net basin supply for Lake Erie.

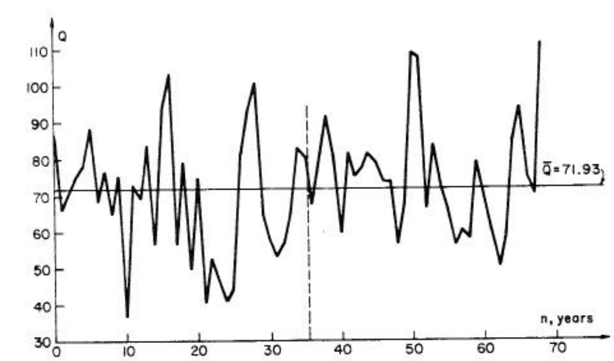

Fig. 2-3 Annual mean net basin supply for Lake Superior

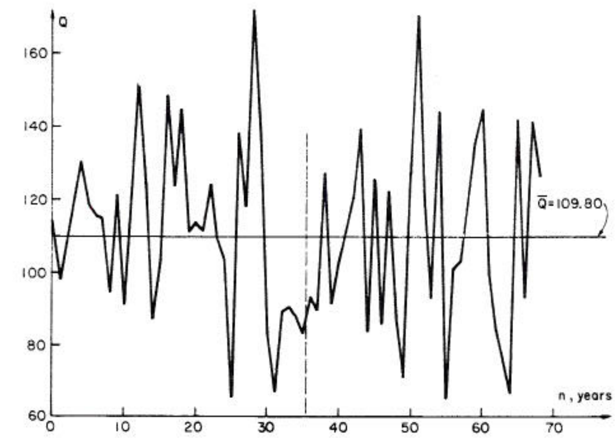

Fig. 2-4 Annual mean net basin supply for Lake Michigan.

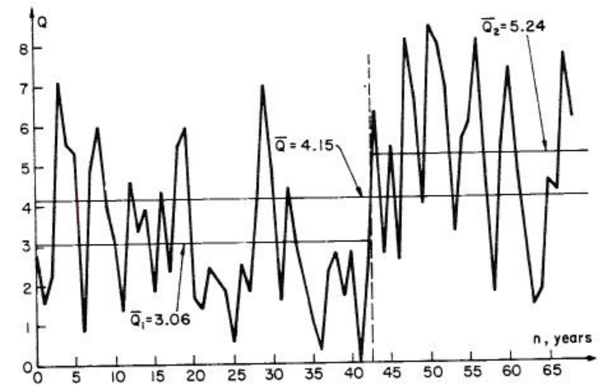

Fig. 2-5 Annual mean net basin supply for Lake St. Clair.

Chapter 4

STRUCTURAL ANALYSIS AND MATHEMATICAL DESCRIPTION OF MONTHLY MEAN NET BASIN SUPPLY

4.1 General Structural Analysis and Description. The structural analysis and mathematical description are based on the following four concepts:

(1) A relationship is established between the periodic parameters of the monthly means and monthly standard deviations, and the dependent stochastic component in the form

$$\varepsilon_{p,\tau} = \frac{x_{p,\tau} - \mu_\tau}{\sigma_\tau} \quad , \qquad (2\text{-}3)$$

in which $x_{p,\tau}$ is the observed mean monthly NBS; μ_τ is the fitted periodic function to the estimated 12 average monthly values, m_τ; σ_τ is the fitted periodic function to the 12 monthly standard deviation, s_τ; p is the year of the NBS series with $p = 1,2,\ldots,n$, and n the sample size in years; τ is the month of the year with $\tau = 1,2,\ldots,12$, and $\varepsilon_{p,\tau}$ is an approximately standardized but dependent stochastic component.

(2) The mathematical description of m_τ and s_τ by the periodic functions μ_τ and σ_τ is achieved by selecting a given number of harmonics out of maximum six harmonics of the monthly series, with

$$\mu_\tau = \bar{x} + \sum_{j=1}^{m} (A_j \cos 2\pi j f \tau + B_j \sin 2\pi j f \tau) , \qquad (2\text{-}4)$$

where $\bar{x}$ is the mean of m_τ, A_j and B_j are Fourier coefficients for the j-th harmonic, and $f = 1/12$ is the ordinary frequency of the 12-month harmonic. A similar equation is obtained for σ_τ, with $\bar{x}$ replaced by the mean of the twelve estimated monthly standard deviations s_τ, and the corresponding A_j and B_j are Fourier coefficients. The Fourier coefficients for μ_τ are estimated by

$$A_j = \frac{1}{6} \sum_{\tau=1}^{12} (m_\tau - \bar{x}) \cos \pi j \tau/6 \qquad (2\text{-}5)$$

and

$$B_j = \frac{1}{6} \sum_{\tau=1}^{12} (m_\tau - \bar{x}) \sin \pi j \tau/6 . \qquad (2\text{-}6)$$

For $j = 6$, the Fourier coefficients are $A_6 = A_j/2$, and $B_6 = 0$.

(3) The dependence model of the dependent stochastic component $\varepsilon_{p,\tau}$ is found in the analysis to be well approximated by the second-order linear autoregressive (Markov) model of the type

$$\varepsilon_{p,\tau} = \alpha_1 \varepsilon_{p,\tau-1} + \alpha_1 \varepsilon_{p,\tau-2} + \xi_{p,\tau} , \qquad (2\text{-}7)$$

where $\xi_{p,\tau}$ is the independent, assumed also to be the second-order stationary stochastic component, α_1 and α_2 are the population autoregressive coefficients (estimated by the sample values a_1 and a_2), related to the first two population autocorrelation coefficients ρ_1 and ρ_2 (estimated by the sample first two serial correlation coefficients, r_1 and r_2), or

$$\alpha_1 = \frac{\rho_1 - \rho_1\rho_2}{1 - \rho_1^2} , \text{ and } \alpha_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} . \qquad (2\text{-}8)$$

By replacing ρ_1 and ρ_2 in Eq. 2-8 by r_1 and r_2, and α_1 and α_2 by a_1 and a_2, then a_1 and a_2 are computed from Eq. 2-8.

(4) A probability distribution function is fitted to the empirical frequency distribution function of $\xi_{p,\tau}$, either separately for each $\xi_{p,\tau}$ of the NBS series, or when shown feasible for all series. In this later case a probability distribution function is fitted to the frequency distribution of $\xi_{p,\tau}$ of all NBS series put together, representing now a unique sample.

4.2 Periodic Components. Fourier coefficients are computed from the 12 values of m_τ, and the 12 values of s_τ, and for all six harmonics, $j = 1,2,\ldots,6$ (or 12-month, 6-month, 4-month, 3-month, 2.4-month, and 2-month harmonic). The amplitudes of these six harmonics are computed by

$$C_j = (A_j^2 + B_j^2)^{1/2} . \qquad (2\text{-}9)$$

The four harmonics with the highest C_j values are considered as having amplitudes significantly different from the amplitude values which series of m_τ and s_τ would have if they would not be periodic.

Table 2-3 presents Fourier coefficients of μ_τ and σ_τ for the five NBS series, and for the four harmonics for each series (given as j in Table 2-3) found significant by the approximate procedure of testing this significance. The harmonics j are sorted in the decreasing order of their amplitudes.

Figures 2-6 through 2-10 give graphs for each of the five NBS series and each graph has four lines: (1) the estimated monthly means, m_τ, (2) the fitted periodic function μ_τ, with four significant harmonics; (3) the estimated monthly standard deviations, s_τ, and (4) the fitted periodic function σ_τ, with four significant harmonics. The study of these five figures leads to the following conclusions:

(1) Periodic components μ_τ and σ_τ have similar patterns but do not seem proportional for these five series of the mean monthly NBS;

(2) Though a parallelism of σ_τ and μ_τ are expected for each NBS series, the sampling fluctuations (and maybe some other factors) produce this non-proportionality;

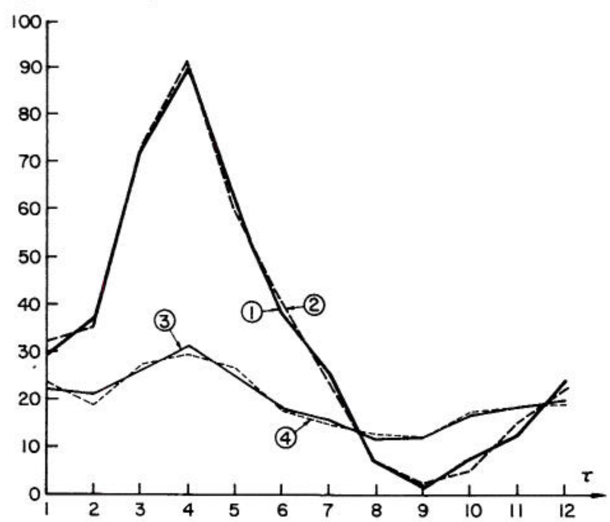

Fig. 2-6 Periodic mean fitted (1) and computed (2) and periodic standard deviation fitted (3) and computed (4) for monthly net basin supply of Lake Ontario.

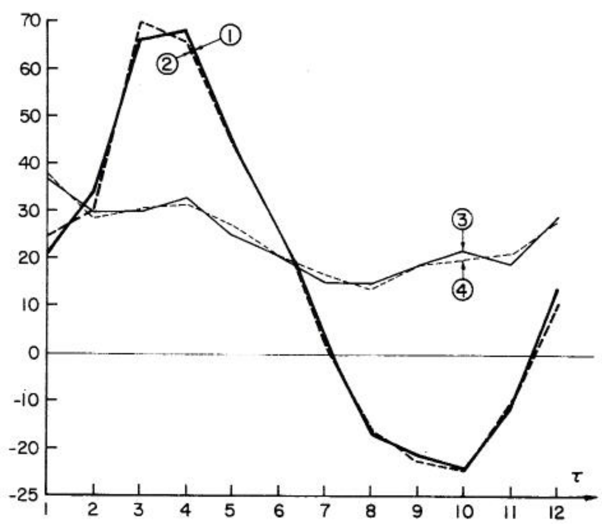

Fig. 2-7 Periodic mean fitted (1) and computed (2) and periodic standard deviation fitted (3) and computed (4) for monthly net basin supply of Lake Erie.

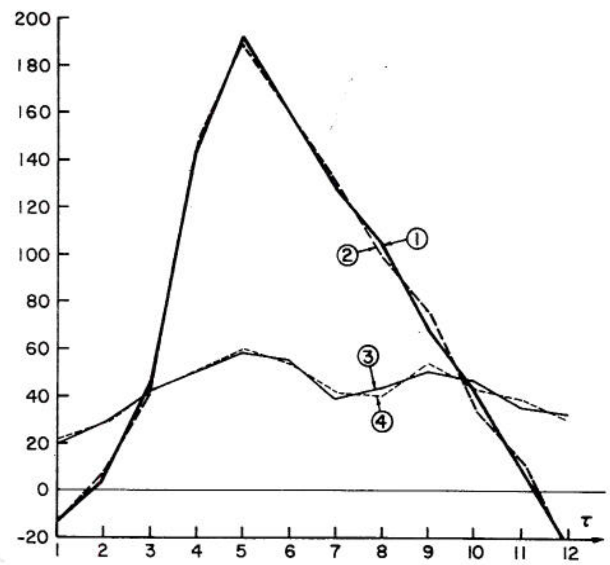

Fig. 2-8 Periodic mean fitted (1) and computed (2) and periodic standard deviation fitted (3) and computed (4) for monthly net basin supply of Lake Superior.

TABLE 2-3

FOURIER COEFFICIENTS OF FITTED HARMONICS TO MEANS AND STANDARD DEVIATIONS OF MONTHLY MEAN NET BASIN SUPPLIES

	Mean μ_τ			St. Deviation σ_τ		
Lake	harmonic j	A_j	B_j	harmonic j	A_j	B_j
Ontario	1	-13.162	32.586	1	-0.166	6.608
	2	-3.048	-10.522	2	0.024	-2.987
	3	6.398	-3.169	4	-0.312	1.238
	4	0.398	5.079	3	1.213	0.160
Erie	1	-11.062	41.503	1	3.732	7.761
	3	4.780	-2.611	3	0.721	2.681
	2	-1.371	-5.022	4	0.023	2.329
	4	4.250	2.751	5	-0.829	0.259
Superior	1	-94.327	-0.243	1	-11.270	-2.557
	2	5.407	-23.493	2	-2.006	-9.090
	3	2.930	10.799	4	2.404	-0.912
	4	-8.776	-1.776	3	-0.343	2.487
Michigan	1	-93.848	84.036	2	-2.364	-14.685
	2	8.191	-34.842	1	-6.062	3.574
	5	-8.567	-9.185	3	3.638	1.990
	3	9.792	-0.353	5	0.656	2.590
St. Clair	1	-0.415	3.148	1	0.408	2.361
	4	0.068	0.276	3	0.514	0.437
	3	0.251	0.043	4	-0.330	0.319
	2	-0.227	0.050	5	0.210	-0.056

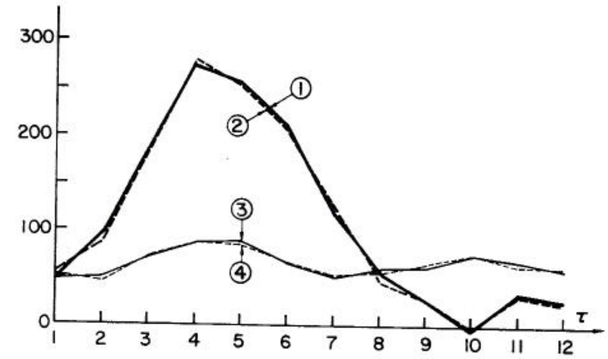

Fig. 2-9 Periodic mean fitted (1) and computed (2) and periodic standard deviation fitted (3) and computed (4) for monthly net basin supply of Lake Michigan.

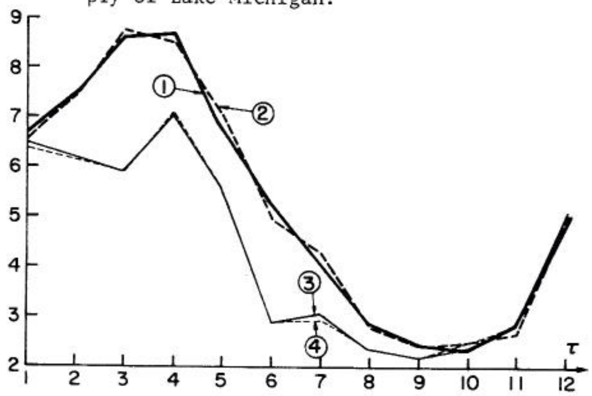

Fig. 2-10 Periodic mean fitted (1) and computed (2) and periodic standard deviation fitted (3) and computed (4) for monthly net basin supply of Lake St. Clair.

(3) Excellent fits of μ_τ to m_τ and of σ_τ to s_τ point out that for all practical purposes in this case there is no significant difference whether $\varepsilon_{p,\tau}$ of Eq. 2-3 has been computed by using m_τ and s_τ instead of the fitted μ_τ and σ_τ; and

(4) The use of μ_τ and σ_τ in Eq. 2-3 should remove the major periodicity in the monthly mean NBS, so that it is only necessary to check whether the autoregressive coefficients a_1 and a_2 are periodic or not by testing whether the serial correlation coefficients r_1 and r_2 are periodic or not, as well as whether the skewness coefficient of the $\xi_{p,\tau}$ series is periodic or not.

4.3 Dependence model for the Dependent Stochastic Components. The computed dependent stochastic components $\varepsilon_{p,\tau}$ by Eq. 2-3 are first tested for periodicities by determining whether their first three serial correlation coefficients $r_{1,\tau}$, $r_{2,\tau}$ and $r_{3,\tau}$ are or are not periodic. Figures 2-11 through 2-15 show the monthly values of these coefficients, with $\tau = 1,2,...,12$, for $\varepsilon_{p,\tau}$ of the five NBS series, together with the means of these 12 monthly values. They are *column serial correlation coefficients*, which means they are computed for each month by using the appropriate autocovariances for n years of data. The serial correlation coefficients of $\varepsilon_{p,\tau}$ for the Lake St. Clair NBS series are computed with the nonhomogeneity removed. Shapes of 15 graphs in Figs. 2-11 through 2-15 show that--for all practical purposes--no periodicity or systematic changes in the column serial correlation coefficients could be detected. There is neither a parallelism of $r_{1,\tau}$, $r_{2,\tau}$, and $r_{3,\tau}$ of Figs. 2-11 through 2-15 with the m_τ and s_τ of Figs. 2-6 through 2-10, respectively for each mean monthly NBS series, nor there is an opposite pattern, namely that $r_{1,\tau}$, $r_{2,\tau}$ and $r_{3,\tau}$ are small when m_τ and s_τ are large or vice versa. Taking into account a high sampling variation of $r_{1,\tau}$, $r_{2,\tau}$, and $r_{3,\tau}$, particularly when the $\varepsilon_{p,\tau}$ series is dependent, so that $r_{k,\tau}$'s are also dependent in sequence, it can be safely assumed that the column serial correlation coefficients are not periodic. Because the average $\bar{r}_{1,\tau}$ is greater than $\bar{r}_{2,\tau}$, and this latter is greater than $\bar{r}_{3,\tau}$ for all series, the autoregressive models (Markov models) of time dependence in $\varepsilon_{p,\tau}$ are indicated. Besides, $\bar{r}_{3,\tau}$ is very close to zero for all series, so the study of $r_{k,\tau}$'s, with $k > 3$, was not considered necessary.

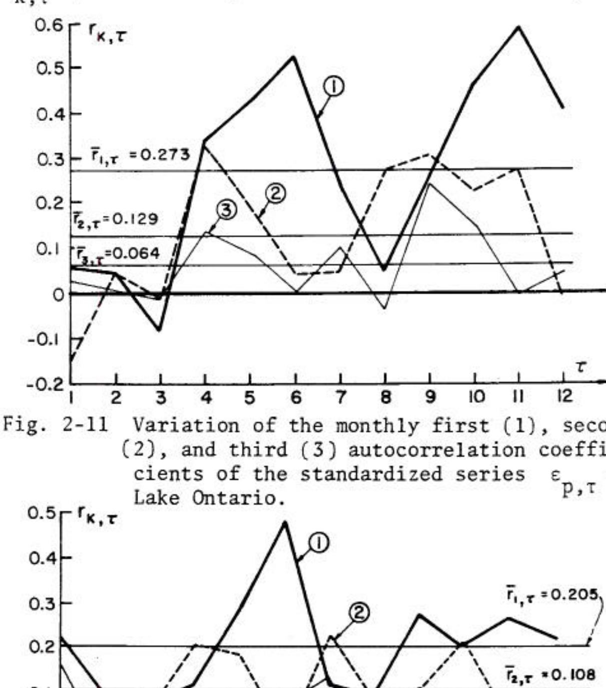

Fig. 2-11 Variation of the monthly first (1), second (2), and third (3) autocorrelation coefficients of the standardized series $\varepsilon_{p,\tau}$ for Lake Ontario.

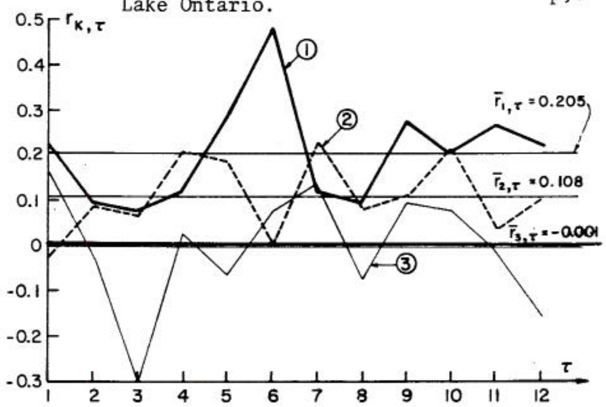

Fig. 2-12 Variation of the monthly first (1), second (2), and third (3) autocorrelation coefficients of the standardized series $\varepsilon_{p,\tau}$ for Lake Erie.

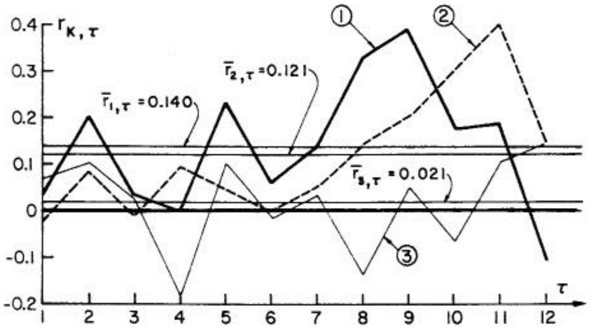

Fig. 2-13 Variation of the monthly first (1), second (2), and third (3) autocorrelation coefficients of the standardized series $\varepsilon_{p,\tau}$ for Lake Superior.

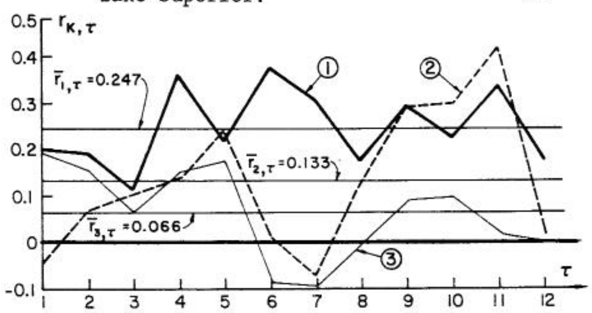

Fig. 2-14 Variation of the monthly first (1), second (2), and third (3) autocorrelation coefficients of the standardized series $\varepsilon_{p,\tau}$ for Lake Michigan

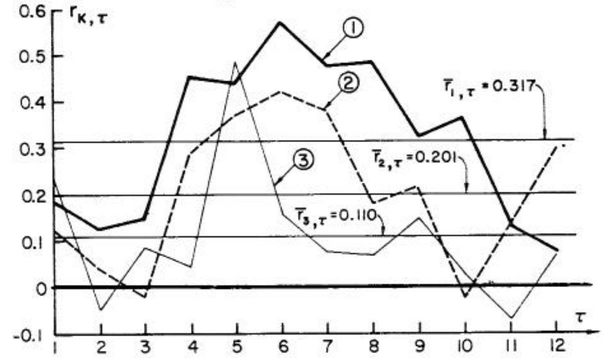

Fig. 2-15 Variation of the monthly first (1), second (2), and third (3) autocorrelation coefficients of the standardized series $\varepsilon_{p,\tau}$ for Lake St. Clair.

Table 2-4 presents a comparison of r_1, r_2, and r_3, or of the first three serial correlation coefficients of the entire $\varepsilon_{p,\tau}$ series, and $\bar{r}_{1,\tau}$, $\bar{r}_{2,\tau}$, and $\bar{r}_{3,\tau}$, or the averages of column serial correlation coefficients of Figs. 2-11 through 2-15. The corresponding differences between these coefficients are relatively small. Therefore, there is no practical difference whether the dependence model of the $\varepsilon_{p,\tau}$ series is based on $r_{1,\tau}$, $r_{2,\tau}$, and $r_{3,\tau}$ or on the coefficients r_1, r_2, and r_3 of the entire $\varepsilon_{p,\tau}$ series.

Because the $\varepsilon_{p,\tau}$ series is found to be the second-order stationary process (after μ_τ and σ_τ are removed, and after $r_{1,\tau}$, $r_{2,\tau}$, and $r_{3,\tau}$ are found not to be periodic), and because there are no significant differences between the mean column serial correlation coefficients and the serial correlation coefficients of the entire $\varepsilon_{p,\tau}$ series, these latter coefficients r_k, k = 1,2,3, are used in estimating the autoregressive coefficients of the inferred order of the autoregressive or Markov linear dependence models.

Figures 2-16 through 2-20, upper graphs, give correlograms of the $\varepsilon_{p,\tau}$ series for the five mean monthly NBS. They all show a dampening effect with the absolute value of r_k decreasing with an increase of the lag k . All these graphs for their first 5-6 values of r_k show the fitted first-order autoregressive linear model of the type

$$\varepsilon_{p,\tau} = \rho_1 \varepsilon_{p,\tau-1} + \xi_{p,\tau} \qquad (2\text{-}10)$$

with

$$\rho_k = \rho_1^k \quad , \qquad (2\text{-}11)$$

and the second-order autoregressive linear model, as given by Eq. 2-7, with the correlogram

$$\rho_k = \alpha_1\rho_{k-1} + \alpha_2\rho_{k-2} \quad , \qquad (2\text{-}12)$$

where α_1, α_2, ρ_1, ρ_{k-1}, ρ_{k-2}, and ρ_k are the population parameters estimated by the sample values a_1 ,

TABLE 2-4

COMPARISON OF FIRST THREE SERIAL CORRELATIONS COEFFICIENTS OF $\varepsilon_{p,\tau}$ SERIES FOR THE FIVE LAKE NET BASIN SUPPLIES: (1) MEANS $\bar{r}_{1,\tau}$, $\bar{r}_{2,\tau}$ AND $\bar{r}_{3,\tau}$ OF COLUMN VALUES, AND (2) r_1, r_2, AND r_3 OF ENTIRE $\varepsilon_{p,\tau}$ SERIES.

	Ontario	Erie	Superior	Michigan	St. Clair
$\bar{r}_{1,\tau}$	0.273	0.205	0.247	0.140	0.317
r_1	0.256	0.196	0.229	0.134	0.308
$\bar{r}_{2,\tau}$	0.129	0.108	0.133	0.121	0.201
r_2	0.134	0.111	0.135	0.130	0.199
$\bar{r}_{3,\tau}$	0.064	-0.001	0.066	0.021	0.110
r_3	0.065	-0.001	0.061	0.038	0.108

a_2, r_1, r_{k-1}, r_{k-2}, and r_k. The visual comparison of the fitted first-order and second-order linear models with the estimated correlograms show that the second-order model fits better the estimated correlograms. When the independent and second-order stationary component $\xi_{p,\tau}$ is computed either by Eq. 2-10 as

$$\xi_{p,\tau} = \varepsilon_{p,\tau} - r_1 \varepsilon_{p,\tau-1} \tag{2-13}$$

for the first-order model, or by Eq. 2-7 as

$$\xi_{p,\tau} = \varepsilon_{p,\tau} - a_1 \varepsilon_{p,\tau-1} - a_2 \varepsilon_{p,\tau-2} \tag{2-14}$$

for the second-order model, the correlograms of $\xi_{p,\tau}$ of Eqs. 2-13 and 2-14 are plotted in Figs. 2-16 through 2-20 as the lower graphs. The 95 percent tolerance limits for the correlograms of independent series show that about 95 percent of r_k values are confined within these tolerance limits for the second-order model, while this is less true for the first-order model. In general, the second-order model shows a better fit, so that it is selected for all five $\varepsilon_{p,\tau}$ series of the monthly mean NBS.

Table 2-5 gives the autoregressive coefficients a_1 and a_2, computed by Eq. 2-8 in using the first two serial correlation coefficients, r_1 and r_2, also given in Table 2-5.

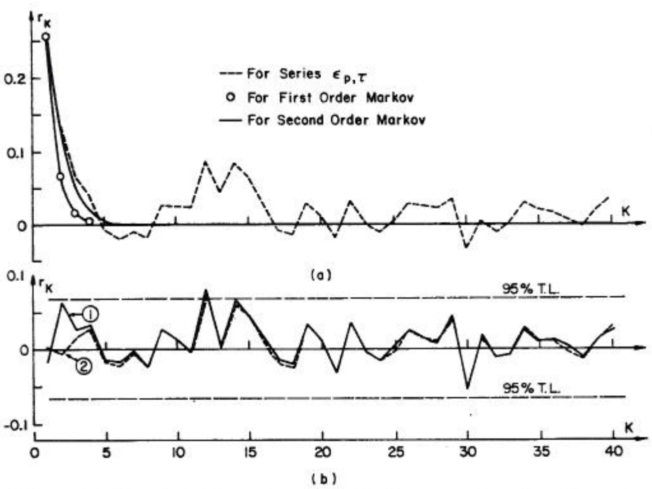

Fig. 2-16. (a) Correlogram of the standardized series $\varepsilon_{p,\tau}$ and expected correlograms for the 1st and 2nd order Markov models. (b). Correlograms of the independent series $\xi_{p,\tau}$ after fitting the 1st (1) and 2nd (2) order Markov models, Lake Ontario.

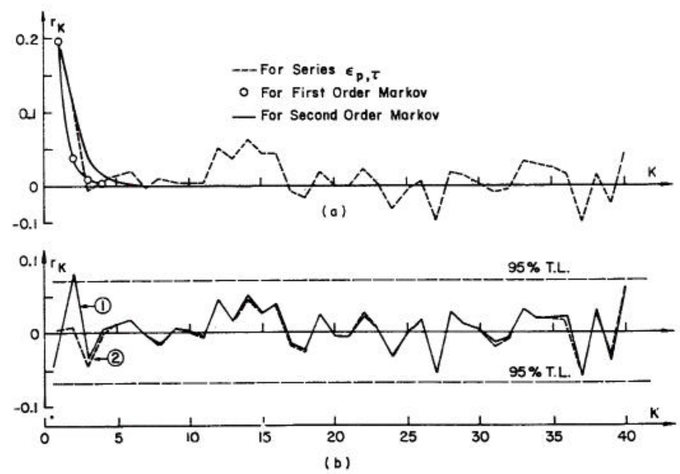

Fig. 2-17 (a) Correlogram of the standardized series $\varepsilon_{p,\tau}$ and expected correlograms for the 1st and 2nd order Markov models. (b). Correlograms of the independent series $\xi_{p,\tau}$ after fitting the 1st (1) and 2nd (2) order Markov models, Lake Erie.

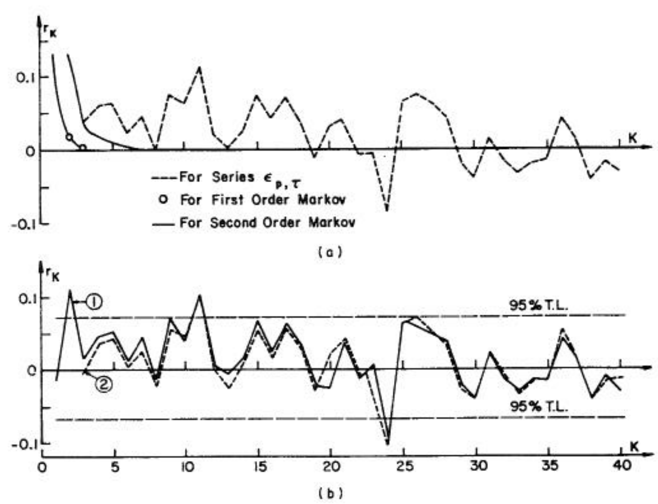

Fig. 2-18. (a) Correlogram of the standardized series $\varepsilon_{p,\tau}$ and expected correlograms for the 1st and 2nd order Markov models. (b). Correlograms of the independent series $\xi_{p,\tau}$ after fitting the 1st (1) and 2nd (2) order Markov models, Lake Superior

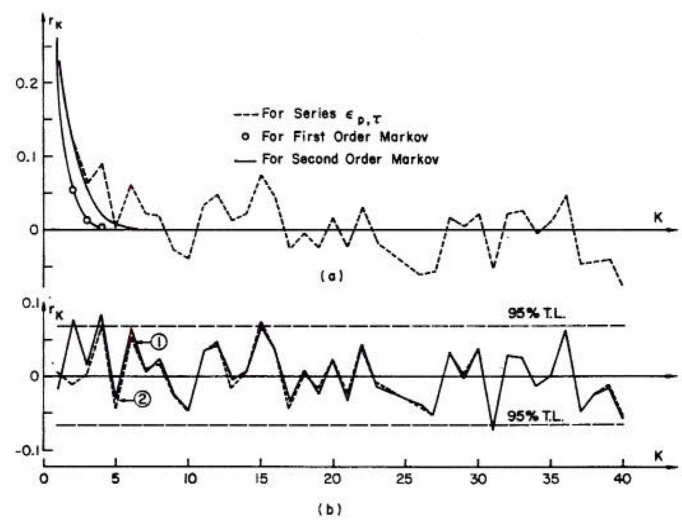

Fig. 2-19 (a) Correlogram of the standardized series $\varepsilon_{p,\tau}$ and expected correlograms for the 1st and 2nd order Markov models. (b). Correlograms of the independent series $\xi_{p,\tau}$ after fitting the 1st (1) and 2nd (2) order Markov models, Lake Michigan.

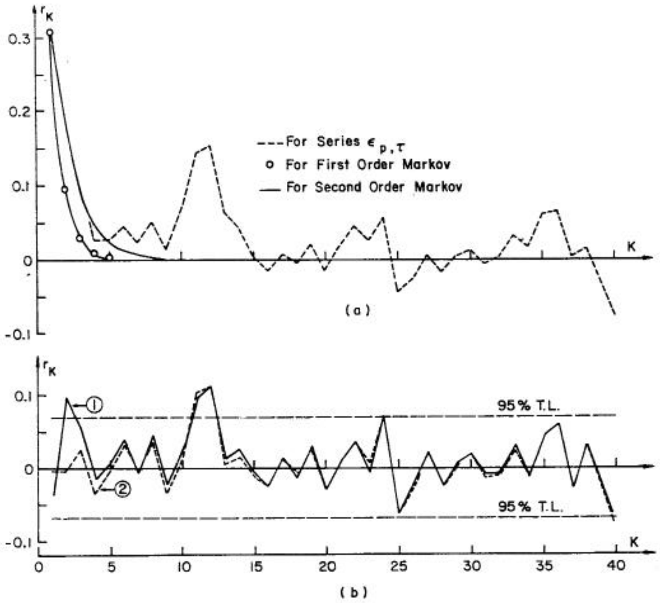

Fig. 2-20 (a) Correlogram of the standardized series $\varepsilon_{p,\tau}$ and expected correlograms for the 1st and 2nd order Markov models, (b) Correlograms of the independent series $\xi_{p,\tau}$ after fitting the 1st (1) and 2nd (2) order Markov models, Lake St. Clair.

TABLE 2-5
AUTOREGRESSIVE COEFFICIENTS OF THE SECOND-ORDER LINEAR MARKOV MODEL, FOR THE $\varepsilon_{p,\tau}$ DEPENDENT STOCHASTIC COMPONENT

Lake	Serial Correlation Coeff. r_1	Serial Correlation Coeff. r_2	Autoregressive Coeff. a_1	Autoregressive Coeff. a_2
Ontario	0.256	0.134	0.2375	0.0732
Erie	0.196	0.111	0.1809	0.0756
Superior	0.134	0.130	0.1183	0.1142
Michigan	0.229	0.135	0.2091	0.0870
St. Clair	0.308	0.199	0.2726	0.1150

4.4 Analysis of Skewness Coefficients of Independent Stochastic Components. Figure 2-21 represents the monthly skewness coefficients, ${}_{\tau}C_s$, of the independent second-order stationary stochastic components, $\xi_{p,\tau}$, with $\tau = 1,2,\ldots,12$, for the five series of the mean monthly NBS. The fluctuations of ${}_{\tau}C_s$ around their average values seem not to have any clear periodicity, though this coefficient should fluctuate in a relatively large range because of sampling variation. The large fluctuation of ${}_{\tau}C_s$ for Lake St. Clair may be explained by a relatively nonhomogeneous sample size of this series.

The entire $x_{p,\tau}$, $\varepsilon_{p,\tau}$ and $\xi_{p,\tau}$ series, neglecting in which month of the year they occur, have the skewness coefficients as given in Table 2-6. The averages of 12 values of monthly ${}_{\tau}C_s$ of the $\xi_{p,\tau}$ series are also given in Table 2-6. The differences of these averages and the C_s values of entire $\xi_{p,\tau}$ series are relatively small.

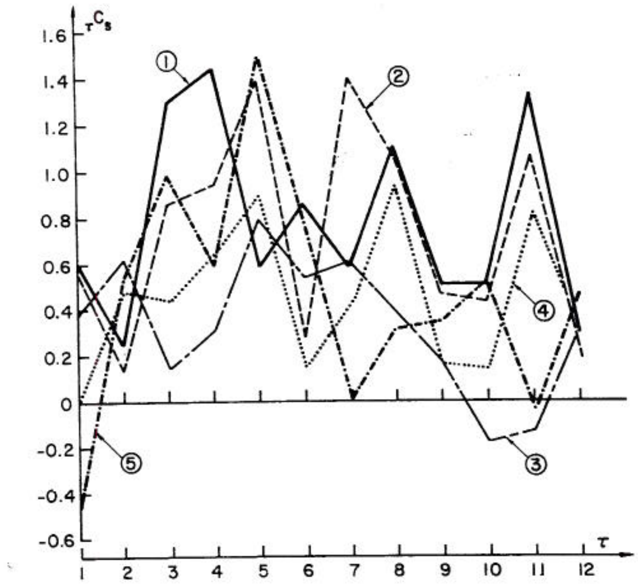

Fig. 2-21 Monthly skewness coefficients, ${}_{\tau}C_s$ of $\xi_{p,\tau}$ series of monthly mean net basin of Lake Ontario (1), Lake Erie (2), Lake Superior (3), Lake Michigan (4), and Lake St. Clair (5).

TABLE 2-6

SKEWNESS COEFFICIENTS FOR $x_{p,\tau}$, $\varepsilon_{p,\tau}$ and $\xi_{p,\tau}$ SERIES OF MEAN MONTHLY NET BASIN SUPPLIES FOR FIVE LAKES

Lake	C_s for Series $x_{p,\tau}$	C_s for Series $\varepsilon_{p,\tau}$	C_s for Series $\xi_{p,\tau}$	Average C_s of 12 ${}_{\tau}C_s$ for $\xi_{p,\tau}$	Average C_s for 20 simulated $x_{p,\tau}$
Ontario	1.005	0.776	0.765	0.760	0.882
Erie	0.648	0.785	0.728	0.730	0.563
Superior	0.489	0.341	0.312	0.310	0.496
Michigan	0.597	0.413	0.443	0.450	0.594
St. Clair	1.052	0.461	0.408	0.405	0.817

The average values of ${}_{\tau}C_s$ are computed for the purpose of analysis whether C_s is or is not a periodic parameter. Because C_s is concluded not to be periodic, it can be assumed for all practical purposes that the $\xi_{p,\tau}$ series are also approximately the third-order stationary independent stochastic variables.

The skewness coefficients of the entire $\xi_{p,\tau}$ series of five mean monthly NBS vary between 0.312 (Lake Superior) and 0.765 (Lake Ontario). It is difficult to conclude whether these five values of C_s: 0.765, 0.728, 0.312, 0.443 and 0.408 are only the sampling variations, or whether they represent the true differences because of different populations. Because the five $\xi_{p,\tau}$ series are mutually highly correlated, their sample skewness coefficients are also mutually dependent statistics. It is then not simple to test whether all five C_s values are or are not from the same population, though this assumption may be justified. Namely, $\xi_{p,\tau}$ represents mainly the *independent climatic noise* introduced into the NBS series. There is not a sufficient justification for the general climatic patterns to differ significantly over the five basins for which the monthly mean NBS are obtained, though the periodic patterns, the time dependent parameters, the general mean $(\bar{x})$ and the general standard deviation (s) change from one basin to another because of differences in basin factors.

Table 2-6 shows that C_s of $\xi_{p,\tau}$ series is smaller for large lake basins (Superior, Michigan) and larger for small lake basins (Ontario, Erie), while C_s value of Lake St. Clair can not be compared with the other four lakes because of built-in nonhomogeneity in data of the St. Clair mean monthly NBS.

A hypothesis is advanced here, namely that the method of computing NBS may be partly responsible for the differences in C_s of the $\xi_{p,\tau}$ series of the four major lakes. The greater a lake surface the less accurate is expected to be the mean monthly NBS, especially their extreme values, because of errors in determining the mean lake levels. Also, the ratio of the lake to land surfaces of each basin may affect the extreme values of NBS, because of evaporation part in the water balance producing each value of NBS. The lack of a small number of extreme high values of $\xi_{p,\tau}$ in NBS of Lakes Superior and Michigan would reduce significantly the values of C_s of their $\xi_{p,\tau}$ series, while the opposite is true for Lakes Ontario and Erie, for which a couple of extreme high values in NBS series would increase significantly the C_s values of their $\xi_{p,\tau}$ series. Therefore, it is difficult to definitely conclude whether C_s values of the $\xi_{p,\tau}$ series in Table 2-6 are siginficantly different or not among themselves.

4.5 Fitting of Lognormal Probability Density Function to Frequency Distributions of $\xi_{p,\tau}$ Because of an easy transformation of generated independent standard (0,1) normal random numbers into the independent random numbers which follow a lognormal distribution, the lognormal probability density function is used for all $\xi_{p,\tau}$ series of the five lake NBS. Because all $\xi_{p,\tau}$ series have negative values, a lognormal probability density function with three parameters is considered as the most feasible to use, or

$$f(\xi) = \frac{1}{(\xi-g)s_n\sqrt{2\pi}} e^{-[\ln(\xi-g)-m_n]^2/2s_n^2} \qquad (2\text{-}15)$$

where g is the lower boundary, m_n is the mean of $\ln(\xi-g)$ and s_n is the standard deviation of $\ln(\xi-g)$. The estimates of m_n, s_n, and g of Eq. 2-15 are given in Table 2-7. First, values of estimates of parameters of Eq. 2-15 are given for the $\xi_{p,\tau}$ series of the four major lakes. Figures 2-22 through 2-25 give the frequency density curves of $\xi_{p,\tau}$ and the fitted three-parameter lognormal probability density curves. Because frequency density curves look much less smooth than their corresponding cumulative frequency distribution curves, the fit looks excellent for all four $\xi_{p,\tau}$ variables. Only for Lake Superior is the critical chi-square value (35.20) approximately equal to computed chi-square values of a fitted lognormal function. The other three $\xi_{p,\tau}$ variables pass well the chi-square test of the goodness of fit.

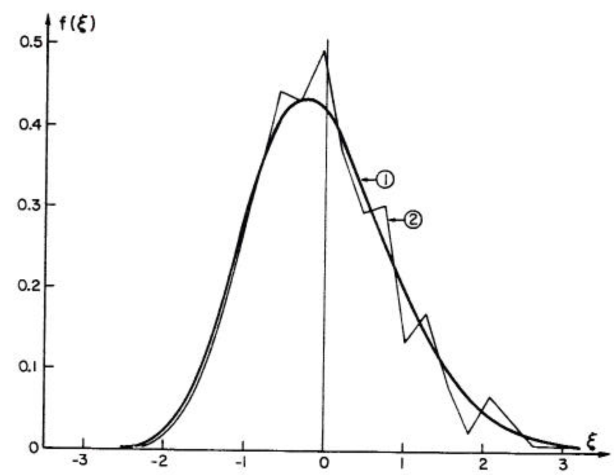

Fig. 2-22 The fit of three-parameters lognormal probability density function (smooth solid line) to the frequency density curve (broken line) of $\xi_{p,\tau}$ variable of monthly mean NBS of Lake Ontario.

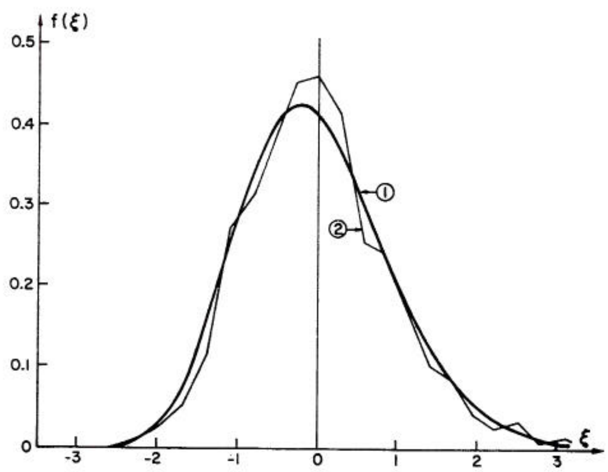

Fig. 2-23 The fit of three-parameters lognormal probability density function (smooth solid line) to the frequency density curve (broken line) of $\xi_{p,\tau}$ variable of monthly mean NBS of Lake Erie.

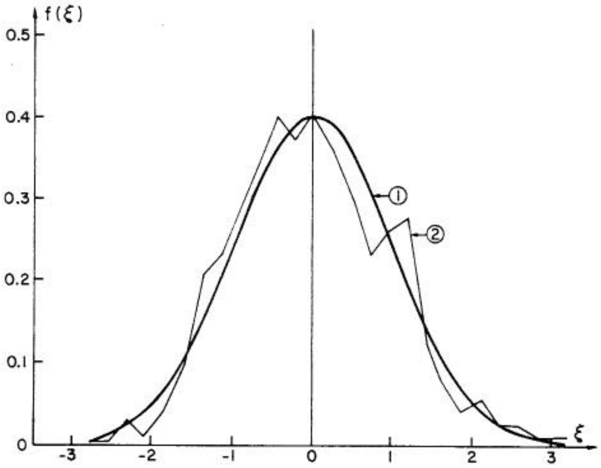

Fig. 2-24 The fit of three-parameters lognormal probability density function (smooth solid line) to the frequency density curve (broken line) of $\xi_{p,\tau}$ variable of monthly mean NBS of Lake Superior.

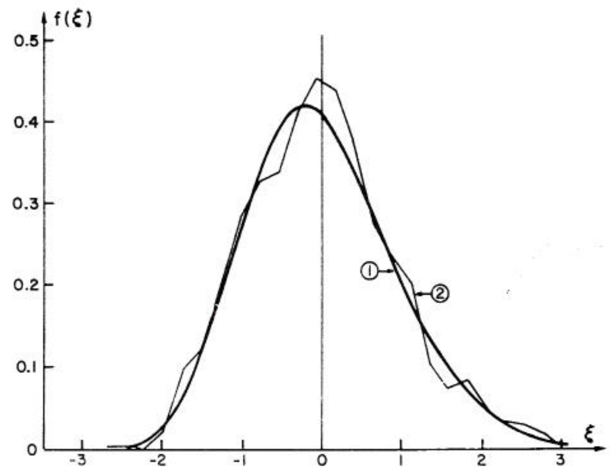

Fig. 2-25 The fit of three-parameters lognormal probability density function (smooth solid line) to the frequency density curve (broken line) of $\xi_{p,\tau}$ variable of monthly mean NBS of Lake Michigan.

TABLE 2-7

CHARACTERISTICS OF THE FITTED LOGNORMAL DISTRIBUTION FUNCTIONS TO FREQUENCY DISTRIBUTIONS OF THE INDEPENDENT STOCHASTIC COMPONENT $\xi_{p,\tau}$

		Chi-square (95%)		Estimates of Parameters		
Lake	Fitted Function	From Tables (critical)	Computed	m_n	s_n	g
Ontario	Lognormal-3	35.20	25.840	1.5969	0.1883	-5.0272
Erie	Lognormal-3	35.20	24.402	1.7454	0.1655	-5.8079
Superior	Lognormal-3	35.20	36.366	2.1801	0.1102	-8.9020
Michigan	Lognormal-3	35.20	24.859	1.7787	0.1608	-6.0000
Four Series Combined	Lognormal-3	35.20	39.195	1.8016	0.1572	-6.1354
St. Clair	Lognormal-3	35.20	22.960	1.7791	0.1584	-6.000

Because the parameters of Table 2-7 are relatively close, it was feasible to consider all four $\xi_{p,\tau}$ series as belonging to the same population. By putting together in one sample all four $\xi_{p,\tau}$ series, a new frequency density curve is obtained and plotted in Fig. 2-26 and a lognormal function is fitted. A visual inspection gives the conclusion of a very good fit. Estimated parameters are given in Table 2-7. Though the chi-square test shows the computed value (39.196) to be greater than the critical chi-square value (35.20), this test may be somewhat in question because all four $\xi_{p,\tau}$ series are highly mutually correlated. This fit, however, passes well the Smirnov-Kolmogorov test. The parameters of fitted function are very close to those of $\xi_{p,\tau}$ for the Lake Michigan NBS.

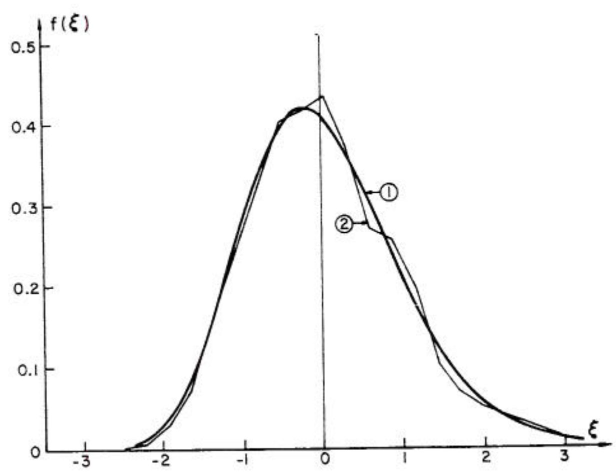

Fig. 2-26 The fit of three-parameters lognormal probability density function (smooth solid line) to the frequency density curve (broken solid line) of four $\xi_{p,\tau}$ series put together in one sample (of NBS series of Lakes Ontario, Erie, Superior and Michigan).

Figure 2-27 gives the frequency density curves for $\xi_{p,\tau}$ series of NBS of the four major lakes (the lines of Figs. 2-22 through 2-25) and the fitted lognormal-3 probability density function of Fig. 2-26 (fitted to all the $\xi_{p,\tau}$ variables put together). This visual test also shows a relatively good fit, and the fluctuations of four frequency density curves around the fitted probability density curve may be assigned mainly to sampling variations.

Figure 2-28 gives the fit of lognormal function to the frequency density curve of $\xi_{p,\tau}$ series of the Lake St. Clair NBS, while Table 2-7, last row, gives the estimated parameters. The chi-square test is satisfactory, while the three parameters show their estimates to be very close to the parameters of the overall fitted lognormal-function to the frequency curves of four $\xi_{p,\tau}$ series put together.

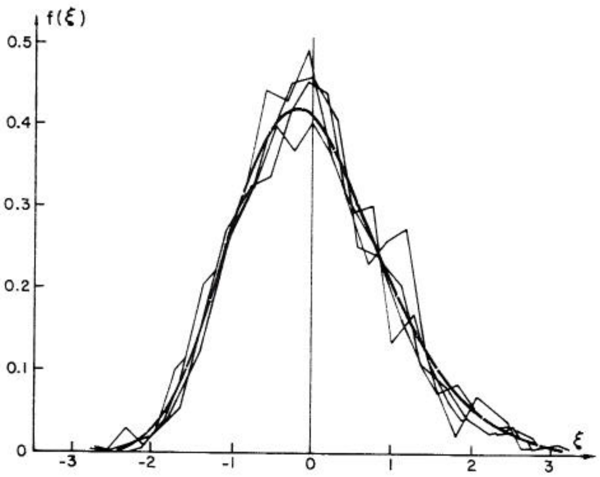

Fig. 2-27 The comparison of the fitted three-parameter lognormal probability density function of Fig. 2-26 with the four individual frequency density curves of Figs. 2-26 - 2-29 of $\xi_{p,\tau}$ series (NBS series of Lakes Ontario, Erie, Superior and Michigan).

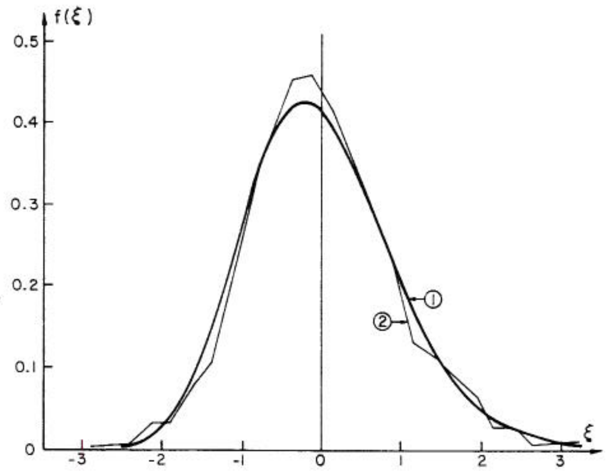

Fig. 2-28 The fit of three-parameters lognormal probability density function (smooth solid line) to the frequency density curve (broken line) of $\xi_{p,\tau}$ variable of monthly mean NSB of Lake St. Clair.

Chapter 5

GENERATION OF NEW SAMPLES OF MONTHLY MEAN NET BASIN SUPPLIES

5.1 Generation of Samples for Four Major Lakes. The method of multivariate normal distribution of four random variables is used in generating the new samples. Because the three-parameter lognormal probability density function has been shown to fit well both the four mutually dependent $\xi_{p,\tau}$ series, or their joint sample (all four series put together as a new sample), it is easy to transform a lognormal fourvariate probability distribution into a normal fourvariate distribution. For that purpose the parameters m_n, s_n, and g of lognormal probability function, fitted to the frequency density curve of the four combined $\xi_{p,\tau}$ series, are used. Furthermore, the transformation of each of the four $\xi_{p,\tau}$ series is made by

$$\zeta_{p,\tau} = \frac{\ln\,(\xi_{p,\tau} - g) - m_n}{s_n} \,, \qquad (2\text{-}16)$$

with m_n, s_n, and g given in Table 2-6 for the fit of four combined series, estimated for use in Eq. 2-15. The new four $\zeta_{p,\tau}$ series are approximately normal, uncorrelated in sequence; however, their means and standard deviations are not exactly zero and unity, respectively, though they are close to these values.

The standardization is then made by

$$y_{p,\tau} = \frac{\zeta_{p,\tau} - \bar{\zeta}}{s_\zeta} \,, \qquad (2\text{-}17)$$

with $\bar{\zeta}$ and s_ζ the mean and the standard deviation for each of the four $\zeta_{p,\tau}$ series. The four $y_{p,\tau}$ series are normal, standardized independent components but mutually dependent (correlated) random variables. The normalization and standardization of the three-parameter lognormal independent stochastic components $\xi_{p,\tau}$ into the four $y_{p,\tau}$ series enable the use of the classical normal multivariate method of principal components in generating the new samples. The generation of samples in this report follows the procedure which is described in the paper by G.K. Young and W.C. Pisano [8].

The standard normal independent $y_{p,\tau}$ components mutually correlated are further transformed into the new standard normal independent $\eta_{p,\tau}$ components which are mutually uncorrelated random variables. The $\eta_{p,\tau}$ series are then the four principal components of the $y_{p,\tau}$ normal fourvariate distribution. To obtain $\eta_{p,\tau}$

components, the transformation is made by

$$(y_{p,\tau})_i = B(\eta_{p,\tau})_i \qquad (2\text{-}18)$$

in which $i = 1,2,3,4$, $(y_{p,\tau})_i$ represent the four components of Eq. 2-17 as a 4 x 1 matrix, $(\eta_{p,\tau})_i$ are the four transformed normal standard components both serially and mutually uncorrelated, also as a 4 x 1 matrix, and B is a 4 x 4 matrix to be estimated from Eq. 2-18. The matrix B is obtained by post-multiplying [8] both sides of Eq. 2-18 by the transpose matrix of $(y_{p,\tau})_i$, given as $(y_{p,\tau})_i^T$, and by taking the expected values of both sides. This leads to

$$B\,B^T = E\{(y_{p,\tau})_i\ (y_{p,\tau})_i^T\} = M_o \quad , \qquad (2\text{-}19)$$

where B^T is the transpose of B and M_o is the lag-zero cross-correlation matrix of the four standard normal $y_{p,\tau}$ components, and given in Table 2-8. The values of the matrix in Table 2-8 are the pairwise correlation coefficients between the $y_{p,\tau}$ components. Then Eq. 2-19 and Table 2-8 permit the estimate of matrix B (the lower triangular matrix), which is given in Table 2-9.

The generation of new samples of mean monthly NBS for Lakes Ontario, Erie, Superior and Michigan then follows the procedure, which is the inverse operation of all above structural analysis and mathematical description:

(1) The standard, normal and independent random numbers are first generated for each of the four $\eta_{p,\tau}$ series. A total of 20 samples, each sample 50 years long (or 20 x 600 monthly values), of each component are generated. This represents a total of 4 x 20 x 600 = 48,000 standard normal random numbers divided into 20 samples each consisting of four series, and each series containing 600 random numbers.

TABLE 2-8

CROSS CORRELATION MATRIX (M_o) OF FOUR STANDARD NORMAL INDEPENDENT $y_{p,\tau}$ SERIES OF NBS SERIES OF FOUR LAKES

Lake	Ontario	Erie	Superior	Michigan
Ontario	1.000	0.608	0.246	0.541
Erie	0.608	1.000	0.191	0.487
Superior	0.246	0.191	1.000	0.444
Michigan	0.541	0.487	0.444	1.000

(2) Using the matrix B of Table 2-9 and Eq. 2-18, 20 samples each 600 long for each of four series $y_{p,\tau}$ are then obtained from the above $\eta_{p,\tau}$ series.

(3) Using Eq. 2-17, the $y_{p,\tau}$ series, and for each of the four $y_{p,\tau}$ series the corresponding $\bar{\zeta}$ and s_ζ, the 20 samples (each 600 long) of the four series of $\zeta_{p,\tau}$ are obtained by

$$\zeta_{p,\tau} = \bar{\zeta} + s_\zeta\, y_{p,\tau} \qquad (2\text{-}20)$$

TABLE 2-9

ESTIMATED B MATRIX (LOWER TRIANGULAR MATRIX) BY EQ. 2-19 IN USING M_o MATRIX OF TABLE 2-8

Lake	Ontario	Erie	Superior	Michigan
Ontario	1.000	0.000	0.000	0.000
Erie	0.608	0.794	0.000	0.000
Superior	0.246	0.051	0.968	0.000
Michigan	0.541	0.199	0.310	0.756

(4) The 20 samples of each of the four $\zeta_{p,\tau}$ series are then transformed by using Eq. 2-16 into the 20 samples of the four $\xi_{p,\tau}$ series (each 600 values long) by

$$\xi_{p,\tau} = g + e^{m_n + s_n \zeta_{p,\tau}} \quad , \qquad (2\text{-}21)$$

with m_n, s_n, and g the same values for all 20 samples of each of the four $\xi_{p,\tau}$ series.

(5) Using the second-order autoregressive linear model, with proper values of autoregressive coefficients a_1 and a_2 for each of the monthly mean NBS (Ontario, Erie, Superior and Michigan), the 20 samples (each 600 long) of each of the four dependent stochastic $\varepsilon_{p,\tau}$ series are produced.

(6) Adding the corresponding periodic components μ_τ and σ_τ for each of the four series by

$$x_{p,\tau} = \mu_\tau + \sigma_\tau\, \varepsilon_{p,\tau} \quad , \qquad (2\text{-}22)$$

the 20 samples of each of the four $\varepsilon_{p,\tau}$ series are transformed into the 20 samples of generated mean monthly net basin supplies of Lakes Ontario, Erie, Superior and Michigan, with each series 600 months long (or 50 years of data in monthly values).

These generated samples of NBS for each of the four lakes are then printed and reproduced in the form of punched cards for further use. The size of samples of 50 years for monthly values is selected by assuming that the life of any lake regulation project should be 50 years. Any 50-year regulation plan will then produce 20 results, one for each of the 20 generated samples, so that a frequency distribution of the 20 regulation values may be determined. The resulting 20-value frequency curves will then enable statements on probabilities of exceedences of lake levels during a given time period.

5.2 Analysis of Generated Samples for Four Major Lakes. The basic approach in this generation of new samples is the preservation, in the limits of sampling variation, of the basic parameters of historic samples. Some statistics, as sample properties, have very large variation from one sample to another. Attempting to reproduce all these properties as exactly the same values as in the historic samples would run against the basic objective, namely the obtaining of potential future samples of relatively sufficient probability to occur. The future will produce some extremes and new properties in samples, which have not been experienced in the observed data. The approach used here is based on the concept of producing the new samples which preserve the most reliable properties and parameters, namely: the general mean, the general standard deviation, the periodic monthly averages and standard deviations, the autoregressive coefficients of selected dependence model, and the basic parameters of selected probability distribution of independent stochastic components.

Tables 2-10 through 2-13 give the three parameters of 20 generated samples of the mean monthly NBS of the four lakes: the mean $\bar{x}$, the standard deviation s_x and the skewness coefficient ${}_xC_s$. At the top of each table these three parameters are given for the historic data. At the bottom of each table the means of the 20 values of each parameter are also given. The best preservation is in the mean $\bar{x}$, the second best is in the standard deviation s_x, and the skewness coefficient ${}_xC_s$ is least well reproduced. This should be expected.

5.3 Generation of New Samples of Monthly Mean NBS of Lake St. Clair. Because Lake St. Clair is a small lake, its monthly mean net basin supplies may be more in error than for other lakes. Already it is shown that its NBS series is nonhomogeneous, and this nonhomogeneity was corrected by adding to the first part of 43 years of series data ($x_{p,\tau}$ series) the difference between the means of the second part (26 years) and the first part (43 years).

It was concluded from the analysis that in the eventual use of Lake St. Clair data in a fivevariate normal distribution of the $y_{p,\tau}$ series, the effect of these data would decrease the reliability of the 20 generated samples of the mean monthly NBS series of the four major lakes. Therefore, a different procedure is used to generate the 20 new samples of NBS for Lake St. Clair.

A linear multiple regression equation is developed for the independent stochastic component $\xi_{p,\tau}$ of the NBS of Lake St. Clair to the four series $\xi_{p,\tau}$ of NBS of Lakes Ontario, Erie, Superior and Michigan, in the form:

$$\xi_{S.C} = a_o + a_1 \xi_{ONT} + a_2 \xi_{ER} + a_3 \xi_{SUP} + a_4 \xi_{MICH} + \nu_{p,\tau} \quad . \qquad (2\text{-}23)$$

In this equation $\nu_{p,\tau}$ is a new independent stochastic

TABLE 2-10
GENERAL STATISTICAL PROPERTIES OF ORIGINAL AND GENERATED MEAN MONTHLY NET BASIN SUPPLY FOR LAKE ONTARIO

	Mean, $\bar{x}$	St. Deviation, s_x	Skewness Coeff., ${}_xC_s$
Historic Sample	33.960	33.741	1.005
		Generated Samples	
1	33.986	33.563	1.005
2	34.048	34.055	0.817
3	34.306	35.568	1.041
4	34.239	34.571	0.957
5	33.488	32.043	0.874
6	34.016	32.819	0.642
7	33.744	32.563	0.820
8	34.101	33.639	0.752
9	34.065	33.879	0.940
10	34.046	33.349	0.867
11	33.784	33.095	0.897
12	34.225	35.584	1.012
13	33.900	34.077	1.024
14	34.140	34.359	0.776
15	33.967	32.719	0.896
16	33.633	32.548	0.912
17	33.504	32.441	0.962
18	33.616	32.262	0.767
19	33.960	33.201	0.789
20	33.940	33.911	0.895
Mean of 20 values	33.935	33.512	0.882

TABLE 2-11
GENERAL STATISTICAL PROPERTIES OF ORIGINAL AND GENERATED MEAN MONTHLY NET BASIN SUPPLY FOR LAKE ERIE

	Mean, $\bar{x}$	St. Deviation, s_x	Skewness Coeff., ${}_xC_s$
Historic Sample	16.797	40.122	0.648
		Generated Samples	
1	16.898	41.875	0.546
2	16.417	40.398	0.652
3	17.268	42.357	0.647
4	17.157	41.401	0.414
5	16.558	39.237	0.721
6	17.090	40.467	0.473
7	16.557	39.598	0.490
8	16.994	39.979	0.541
9	17.019	40.175	0.604
10	17.060	42.238	0.670
11	16.842	40.100	0.628
12	16.932	40.849	0.558
13	16.994	41.152	0.464
14	16.692	39.771	0.469
15	16.818	39.750	0.553
16	16.924	39.048	0.511
17	16.305	38.920	0.614
18	16.690	39.460	0.614
19	17.027	40.701	0.465
20	16.731	39.721	0.631
Mean of 20 values	16.849	40.360	0.563

TABLE 2-12
GENERAL STATISTICAL PROPERTIES OF ORIGINAL AND GENERATED MEAN MONTHLY NET BASIN SUPPLY FOR LAKE SUPERIOR

	Mean, $\bar{x}$	St. Deviation, s_x	Skewness, Coeff. $_xC_s$
Historic Sample	71.932	82.160	0.489
		Generated Samples	
1	71.861	82.534	0.523
2	71.429	80.368	0.465
3	71.912	83.005	0.482
4	72.828	84.521	0.602
5	72.181	82.933	0.466
6	71.546	79.440	0.483
7	72.481	82.042	0.416
8	71.800	81.416	0.496
9	71.511	80.457	0.419
10	72.357	82.478	0.436
11	71.560	81.160	0.451
12	71.922	83.157	0.484
13	72.029	83.011	0.544
14	72.406	83.493	0.403
15	72.519	85.262	0.649
16	72.343	84.889	0.599
17	71.711	78.496	0.431
18	71.973	83.051	0.589
19	72.212	82.151	0.482
20	71.243	80.263	0.507
Mean of 20 values	71.991	82.206	0.496

TABLE 2-13
GENERAL STATISTICAL PROPERTIES OF ORIGINAL AND GENERATED MEAN MONTHLY NET BASIN SUPPLY FOR LAKE MICHIGAN

	Mean, $\bar{x}$	St. Deviation, s_x	Skewness, Coeff. $_xC_s$
Historic Sample	109.787	114.599	0.597
		Generated Samples	
1	110.230	117.066	0.694
2	109.521	114.763	0.689
3	110.371	122.401	0.501
4	109.733	111.494	0.532
5	109.683	113.540	0.608
6	109.813	110.261	0.499
7	110.052	116.044	0.540
8	109.872	116.196	0.753
9	109.912	116.199	0.488
10	110.130	115.777	0.546
11	109.071	111.406	0.596
12	110.149	118.418	0.726
13	110.090	118.049	0.622
14	111.083	116.636	0.627
15	109.812	116.544	0.630
16	109.003	111.569	0.586
17	109.692	114.247	0.451
18	109.535	111.767	0.499
19	109.884	114.081	0.623
20	110.564	117.055	0.664
Mean of 20 values	109.900	115.176	0.594

component for Lake St. Clair, which is independent of the independent stochastic components of the other four lakes, while a_o, a_1, a_2, a_3 and a_4 are the multiple regression coefficients to be estimated from data. They are obtained from the five $\xi_{p,\tau}$ series, and they are: $a_o = -0.00065$, $a_1 = 0.12580$, $a_2 = 0.16721$, $a_3 = -0.04994$ and $a_4 = 0.19315$.

The historic sample variable $\nu_{p,\tau}$ is then computed from the five $\xi_{p,\tau}$ series and the above coefficients by using Eq. 2-23. The frequency density curve of the $\nu_{p,\tau}$ variable is plotted in Fig. 2-29 together with the fitted three-parameter lognormal probability density function of Eq. 2-15, with the following parameters: $m_n = 2.4821$, $s_n = 0.07197$, and $g = -12.00$. The critical value of chi-square, on the 95 percent probability level, for this fit is 35.20, while the computed chi-square is 31.04, showing the goodness of fit to be sufficient and to be accepted.

The generation of the 20 new samples was as follows:

(1) The 20 samples, each 600 values long, are generated for the $\nu_{p,\tau}$ series by first generating the standard, normal and independent random numbers $t_{p,\tau}$,

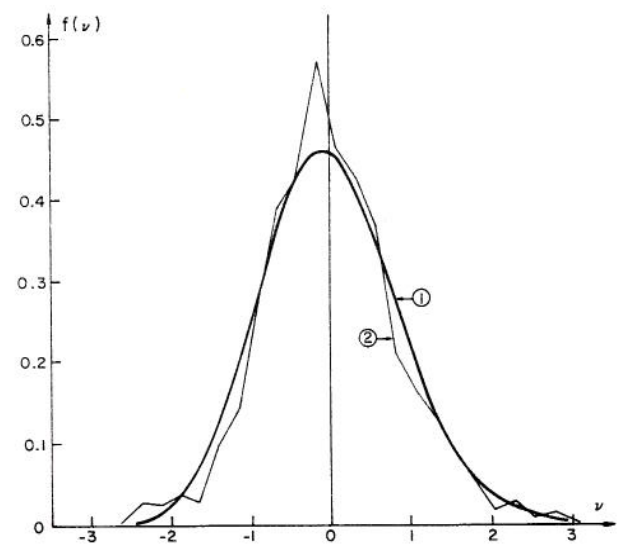

Fig. 2-29 Frequency density curve (broken line) of the independent stochastic variable $\nu_{p,\tau}$ of NBS series of Lake St. Clair, and the fitted three-parameter lognormal probability density curve (smooth line).

and by transforming them to $\nu_{p,\tau}$ by

$$\nu_{p,\tau} = -12.00 + e^{2.4821 + 0.07197\, t_{p,\tau}} . \quad (2\text{-}24)$$

(2) Using Eq. 2-23 and the four generated series $\xi_{p,\tau}$ (Ontario, Erie, Superior and Michigan) for each of the 20 samples, together with $\nu_{p,\tau}$ of each of the 20 samples, the new $\xi_{p,\tau}$ series of NBS for Lake St. Clair are generated.

(3) Following then the same procedure as it was done for the other four lakes, namely performing the transformation of the second-order autoregressive linear model to the $\varepsilon_{p,\tau}$ series, and adding periodicities to obtain the $x_{p,\tau}$ series, the 20 new samples of the mean monthly NBS of Lake St. Clair are produced.

Table 2-14 gives the comparison for Lake St. Clair of the three parameters ($\bar{x}$, s_x, and ${}_xC_s$) of historic sample and of the means of these parameters for the 20 generated samples. The same conclusion can be drawn for the generated mean monthly NBS of this lake as for the other four lakes.

A much better way of testing how well the estimated parameters of generated samples reproduce the parameters estimated from the historic series, is by using the sampling distributions of these estimated parameters. Then the chi-square test may be used to find how well the twenty estimates of generated samples conform with the expected sampling distributions. Difficulties arise because the determination of exact or good-approximation distributions of parameter estimates of periodic-stochastic processes is not simple. Reasons are that several parameters are periodic, while the stochastic component is only approximately stationary and is highly time dependent process.

TABLE 2-14
GENERAL STATISTICAL PROPERTIES OF ORIGINAL AND GENERATED MEAN MONTHLY NET BASIN SUPPLY FOR LAKE ST. CLAIR

	Mean, $\bar{x}$	St. Deviation, s_x	Skewness, Coeff. ${}_xC_s$
Historic Sample	5.245	5.190	1.052
		Generated Samples	
1	5.273	5.273	0.877
2	5.273	5.306	0.707
3	5.357	5.308	0.816
4	5.281	5.317	0.731
5	5.215	5.233	0.791
6	5.162	4.967	0.544
7	5.212	5.191	0.943
8	5.256	5.271	0.887
9	5.101	4.980	0.730
10	5.197	5.282	1.070
11	5.275	5.136	0.774
12	5.258	5.413	0.891
13	5.248	5.292	0.935
14	5.288	5.016	0.725
15	5.246	5.238	0.685
16	5.134	5.059	0.757
17	5.318	5.260	0.603
18	5.215	5.169	0.842
19	5.464	5.557	0.867
20	5.275	5.451	1.155
Mean of 20 values	5.252	5.236	0.817

PART III
GENERATION OF SAMPLES OF MEAN QUARTER-MONTHLY NET BASIN SUPPLIES OF LAKES ONTARIO AND ERIE

This part refers to the structural analysis of historic data, made available by the U.S. Corps of Engineers, Chicago Division, of the mean quarter-monthly net basin supplies--in further text abbreviated as NBS --of the Lakes Ontario and Erie, and also the generation by the experimental (Monte Carlo) method of 20 samples each 50 years long of each of the two mean quarter-monthly NBS. It contains the following: tests of homogeneity of the two discrete NBS time series, structural analysis and mathematical description of these two series found homogeneous, generation of 20 samples each 50 years long for the two series, and analysis of the generated samples.

Chapter 6

BASIC APPROACH IN GENERATING SAMPLES

6.1 Tests of Homogeneity of NBS Series. Tests of homogeneity of mean quarter-monthly NBS series of Lakes Ontario and Erie are made by using the split-sample approach and by ascertaining whether differences between the means of the two unequal subsamples (36 and 33 years) are or are not significantly different from zero at the 95 percent probability level of significance. The t statistics of Eq. 2-1 is also used in this test, namely now as Eq. 3-1,

$$t = \frac{x_1 - x_2}{s\sqrt{\dfrac{n_1 + n_2}{n_1 \cdot n_2}}} \quad . \tag{3-1}$$

As is true for the mean monthly NBS, it is also found for the mean quarter-monthly NBS series of Lakes Ontario and Erie, that these two series are homogeneous, so that the original data are used for the structural analysis and mathematical description of these two series.

6.2 Two Procedures for Generating Samples of Mean Quarter-Monthly Net Basin Supplies. Two procedures are considered in generating samples of mean quarter-monthly net basin supplies:

(1) Generating the samples of the monthly NBS series, and then superimposing the generated samples of four values for each month, which four values represent the differences between the mean quarter-monthly NBS values and the mean monthly NBS value, these latter means being the averages of the four mean quarter-monthly NBS values; and

(2) Generating the mean quarter-monthly NBS values, and then by averaging the four consecutive values produce the corresponding mean monthly NBS series.

The first approach permits the use of the previously generated mean monthly NBS series (20 series each 50 years long) of Lakes Ontario and Erie, and by superimposing the generated four differences for each month between the mean quarter-monthly NBS and the mean monthly NBS values for each month, the new samples of mean quarter-monthly NBS series are obtained. This approach assumes simple characteristics of differences to be superimposed on the mean monthly NBS values. However, it was found that the structure of the four differences, and for each month of the year, is not so simple as to promise their reliable generation. This difficulty is demonstrated later in this text, as support for the decision why this first procedure should not be used and should be replaced by the second procedure.

The second procedure of directly generating the mean quarter-monthly NBS samples is based on the following two factors:

(1) The Ontario and Erie Lakes are relatively small to re-regulate outflows from the upstream lakes; it can be assumed that they are assigned only to regulate their own mean quarter-monthly NBS, or the combined Ontario-Erie supplies. In other words, the outflow from the upstream lakes is passing through these two lakes without being re-regulated.

(2) The generation of the mean quarter-monthly NBS series and the computation of corresponding mean monthly NBS series are intended basically to determine whether the use of mean quarter-monthly series is more appropriate for the regulation of these two lakes than the use of the mean monthly NBS series. Therefore, it is not necessary to use the same monthly series for this objective as for the monthly series generated for regulating all five lakes jointly, in preserving the mutual correlation between the independent stochastic components of the mean monthly NBS of these five lakes.

The first procedure should be used if the mean quarter-monthly NBS series of Lakes Ontario and Erie would be used in conjunction with the mean monthly NBS series of the other three lakes (Superior, Michigan-Huron and St. Clair). When the study of selecting the quarter-monthly or the monthly time interval of NBS series for the regulations of Lakes Ontario and Erie is a primary objective, as postulated in this analysis, then the generation of samples by the second procedure is appropriate, as a more accurate procedure. Namely, in generating the mean quarter-monthly NBS series (20 samples each 50 years long) and from them by summing the four consecutive values, the corresponding mean monthly NBS series are obtained. Therefore, by selecting the second procedure, the problem of the Great Lakes regulations are divided into two:

(1) The study of the selection of the time interval of a discrete time series, which is mainly a problem of flow regulation by Lakes Ontario and Erie; and

(2) The general study of the Great Lakes regulations by using the mean monthly NBS series of all five lakes.

If the selection of the mean monthly NBS series for Lakes Ontario and Erie comes out to be a sufficiently accurate alternative, then the use of the mean monthly NBS series, as generated, may be used for all lakes. If, however, the comparison of regulations of Lakes Ontario and Erie with quarter-monthly and monthly series shows significant differences in the extreme levels of these lakes, then these results must be compared and taken into account. Namely, the five lakes can be studied in their water regulations by using the monthly series, and then the above differences should be added in a proper way for results of regulations of Lakes Ontario and Erie. The third alternative, namely to use in regulation studies the monthly series for the three lakes (Superior, Michigan-Huron, and St. Clair), and the quarter-monthly series for Lakes Ontario and Erie, though possible, would be less convenient.

First, the second procedure as used in the final analysis is described in Chapter 7. Here, the properties of the first procedure are discussed to show the reason for not using it.

6.3 Use of Differences between Mean Quarter-Monthly and Mean Monthly Net Basin Supplies to Generate New Samples of Mean Quarter-Monthly Series. A method, often suggested or sometimes used in generating new samples of discrete series with the time intervals less than a month, is as follows. First, the monthly series is analyzed and the new samples are generated. From historical data the differences are computed between the series for the interval or the intervals less than a month and the corresponding monthly series. These differences are then separately analyzed and then their new samples are generated. By superimposing the generated differences to generated new samples of monthly series, the new samples of discrete series with intervals less than a month are obtained.

This method of superimposing the differences, namely

$$\Delta x_{p,t} = (x_{p,\tau})_m - (x_{p,t})_d, \qquad (3\text{-}2)$$

with $t = 1,2,\ldots,12$ and $\tau = 1,2,\ldots,48$, and where $(x_{p,t})_m$ is the monthly series, and $(x_{p,t})_d$ the series with an interval less than a month, has difficulties which are briefly shown here. Only the mean quarter-monthly NBS and the mean monthly NBS of the Lake Ontario are used to demonstrate these difficulties. The NBS series of the Lake Erie gives the same results as obtained for the NBS series of the Lake Ontario.

Figure 3-1 gives the 48 means m_τ of the $\Delta x_{p,\tau}$ series of Eq. 3-2 with $(x_{p,t})_d$ being the mean quarter-monthly NBS values. This figure shows the following properties of m_τ:

(1) There is no evident periodicity in m_t of $\Delta x_{p,t}$, leading to the conclusion that the periodic component of the mean monthly NBS, described either by the computed m_τ or by the fitted periodic function μ_τ of mean monthly and mean quarter-monthly values, takes care also of the periodicity.

(2) The variance of $\Delta x_{p,t}$ changes with the position t, as is shown in Fig. 3-2.

(3) The constraint is imposed that the four values of $\Delta x_{p,t}$ of each month must sum to zero, or

$$\sum_{p,t+1}^{p,t+4} \Delta x_{p,t} = 0\ , \qquad (3\text{-}3)$$

with $(p,\tau+1)$ and $(p,\tau+4)$ designating the first and the fourth quarters of each month. If the first three values of each month follow a given structure of dependence, the constraint of Eq. 3-3 requires the fourth quarter to disrupt the dependence pattern. By using the fourth quarter value of the previous month, the first quarter value $\Delta x_{p,\tau+1}$ of the next month may be developed to follow the dependence structure. For a more complex structure depending on several previous values, the constraint of Eq. 3-3 introduces still further biases.

(4) On the average, every fourth value is either the peak or the trough in the sequence of m_τ, with an artificially built-in four -value (monthly) cycle, as

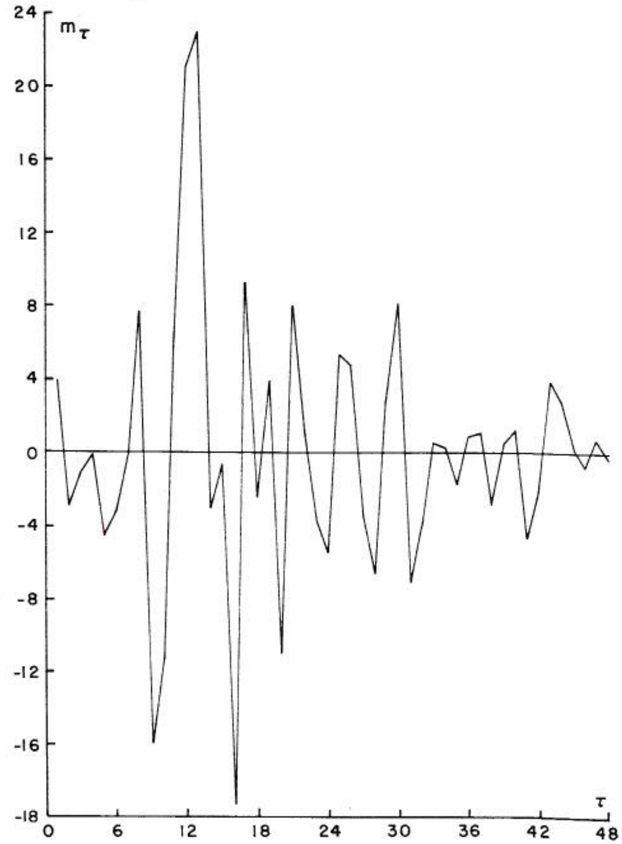

Fig. 3-1 The 48 means m_τ for the 48 quarter-monthly intervals of the differences between the mean quarter-monthly and the mean monthly net basin supplies of Lake Ontario.

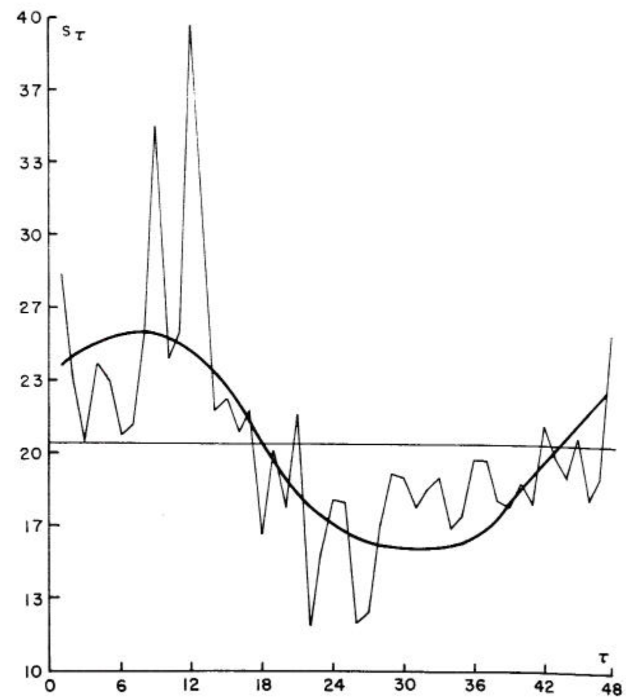

Fig. 3-2 The 48 standard deviations, s_τ (broken line) for the 48 quarter-monthly intervals of the differences between the mean quarter-monthly and the mean monthly net basin supplies of Lake Ontario, with an approximate fit of periodic function, σ_τ (smooth line).

shown by Fig. 3-1. This cycle does not exist in nature, and is basically due to the constraint of Eq. 3-3.

Figure 3-2 gives the 48 standard deviations s_τ of the $\Delta x_{p,\tau}$ series of Eq. 3-2. This figure shows the following properties of s_τ:

(1) There is a clear 12-month cycle inside the s_τ series, which leads to the conclusion that the monthly and quarter-monthly series did not take into account the complete periodicity in the standard deviation.

(2) The artificial cyclicity of one month (four values of the quarter-monthly series), created by the constraint of Eq. 3-3, is also present--on the average, but less clearly evident for s_τ than for m_τ--in the 48 standard deviations s_τ of the $\Delta x_{p,\tau}$ series of Fig. 3-2.

(3) For any generation of new samples of the mean quarter-monthly NBS series, a periodic function should be fitted to the 48 values of s_τ of the $\Delta x_{p,\tau}$ series.

Figure 3-3 gives the correlogram of the $\Delta x_{p,\tau}$ series of Eq. 3-3 and the first parts of correlograms of the residual $\xi_{p,\tau}$ series when the first-order and second-order linear autoregressive models are fitted to the $\Delta x_{p,\tau}$ series (the remaining parts of correlograms are nearly the same as the computed correlogram). These correlograms show the significant negative values for r_1 through r_5 in case of both the $\Delta x_{p,\tau}$ series and the two $\xi_{p,\tau}$ series of the two autoregressive models. It may be safe to postulate that the constraint of Eq. 3-3 has made a significant influence on the $\Delta x_{p,\tau}$ series. The derived $\xi_{p,\tau}$ series by the autoregressive models are not the independent stochastic series. The correlograms of $\varepsilon_{p,\tau}$, (1) in Fig. 3-3, and the correlogram of the $\xi_{p,\tau}$ series, (2) and (3) of Fig. 3-3--(not shown in Fig. 3-3), are relatively well confined within the tolerance limits of the 95 percent probability level of an independent series; however, only from the serial correlation coefficient r_6 on.

Figure 3-4 gives the skewness coefficients of the $\Delta x_{p,\tau}$ series for the 48 quarter-monthly intervals. This figure shows the following properties:

(1) There is a very large variation in the skewness coefficient along the 48 positions of τ; and

(2) There is no evidence of annual cyclicity in C_s of the $\Delta x_{p,\tau}$ series of Eq. 3-2.

For all the reasons discussed, the procedure of generating the new samples of the mean quarter-monthly NBS by first generating the new samples of mean monthly NBS and then by superimposing on them the generated differences $\Delta x_{p,\tau}$ of Eq. 3-2, should not be used until further investigations show how to avoid the built-in bias in using the constraint of Eq. 3-3.

Chapter 7

STRUCTURAL ANALYSIS AND MATHEMATICAL DESCRIPTION OF MEAN QUARTER-MONTHLY NET BASIN SUPPLIES OF LAKES ONTARIO AND ERIE

7.1 The Approach to Structural Analysis. The four concepts for structural analysis and mathematical description of monthly mean NBS as given in Part II are also used for the analysis of the mean quarter-monthly NBS series of the two lakes.

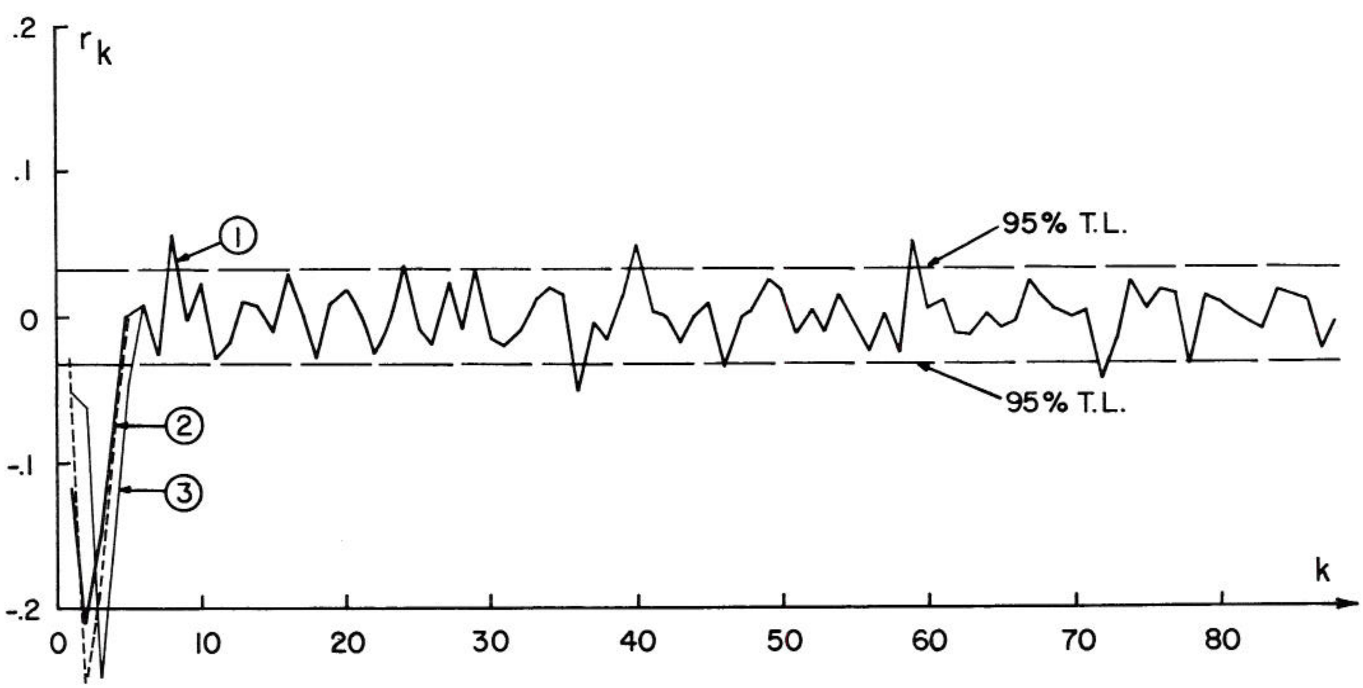

Fig. 3-3 The correlogram (1) of differences between mean quarter-monthly net basin supply and mean monthly $\xi_{p,\tau}$, net basin supply of Lake Ontario (solid line), with the correlograms of residuals, when the first two autoregressive linear models are applied to above differences: (2) first-order model: (3) second-order model.

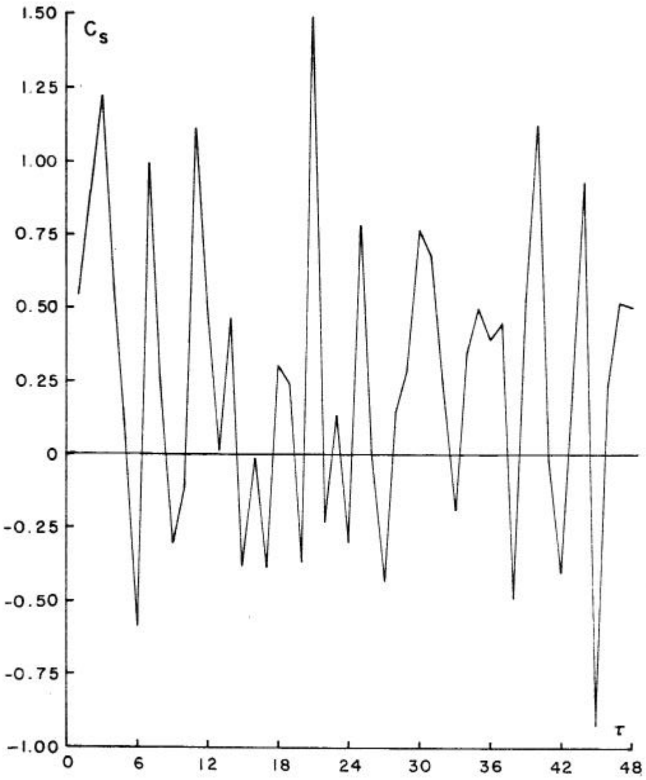

Fig. 3-4 The 48 skewness coefficients, C_s, for the 48 quarter-monthly intervals of the differences between the mean quarter-monthly and the mean monthly net basin supplies of Lake Ontario.

7.2 Structural Analysis and Mathematical Description. The mean values of the mean quarter-monthly NBS series are: Lake Ontario $\bar{x} = 33.93$, Lake Erie $\bar{x} = 16.89$, in units as given by the U.S. Corps of Engineers for the original historic series.

The Fourier coefficients for the four to six harmonics of the quarter-monthly means (48 values of m_τ), found significant, and of six harmonics of quarter-monthly standard deviations (48 values of s_τ), found significant, out of the possible 24 harmonics for the fit of periodic functions to $\omega = 48$ values of m_τ and s_τ, by using the functions of the type of Eq. 2-4, are presented in Table 3-1 for Lakes Ontario and Erie.

The means and variances of the fitted periodic functions μ_τ and σ_τ are:

For Lake Ontario: $\bar{\mu}_\tau = 33.93$, $s^2(\mu_\tau) = 784.13$

$\bar{\sigma}_\tau = 28.65$, $s^2(\sigma_\tau) = 52.60$

For Lake Erie: $\bar{\mu}_\tau = 16.89$, $s^2(\mu_\tau) = 1087.03$

$\bar{\sigma}_\tau = 40.02$, $s^2(\sigma_\tau) = 137.59$

Figures 3-5 and 3-6 give the 48 computed quarter-monthly means m_τ (1), the fitted periodic mean μ_τ (2), the computed quarter-monthly standard deviations s_τ (3), and the fitted periodic standard deviation σ_τ (4), with the first figure referring to the Lake Ontario, and the second figure to the Lake Erie. These figures lead to the following conclusions:

TABLE 3-1
FOURIER COEFFICIENTS OF FITTED HARMONICS TO MEANS AND STANDARD DEVIATIONS OF MEAN QUARTER-MONTHLY NET BASIN SUPPLIES OF LAKES ONTARIO AND ERIE

Lake	Mean μ_τ			St. Deviation σ_τ		
	Harmonic j	A_j	B_j	Harmonic j	A_j	B_j
Ontario	1	-6.72912	35.10671	1	2.80829	7.58073
	2	-7.28315	- 8.67878	2	-2.73167	-1.34601
	3	3.72747	- 7.46954	5	1.82896	2.38898
	4	5.29405	3.22931	3	1.20795	-1.43560
	5	-2.01569	3.28203	6	-1.59427	0.78324
	6	-2.81477	- 0.47560	4	-1.56599	-0.24596
Erie	1	-3.59848	43.55893	1	8.18817	9.88355
	3	3.65264	- 6.79597	4	4.44852	1.21878
	4	7.35812	- 1.03199	3	2.92948	-0.25768
	2	-3.92515	- 4.76208	5	0.97925	2.10916
				2	-2.20833	0.13688
				6	-1.27545	1.06196

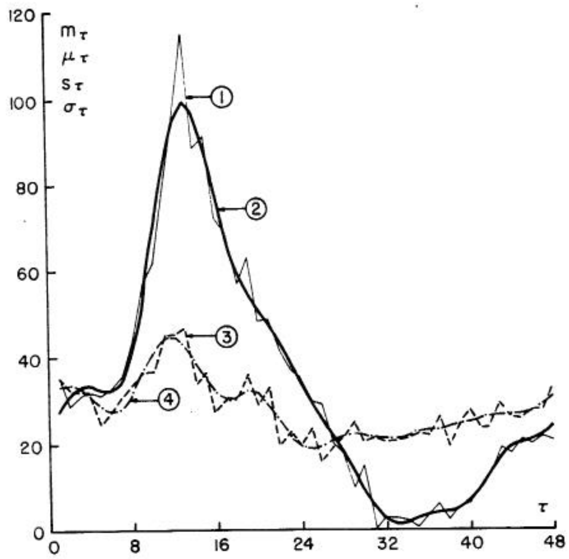

Fig. 3-5 Periodic mean, computed m_τ (1) and fitted μ_τ (2), and periodic standard deviation, computed s_τ (3) and fitted σ_τ (4), for the quarter-monthly net basin supplies of Lake Ontario.

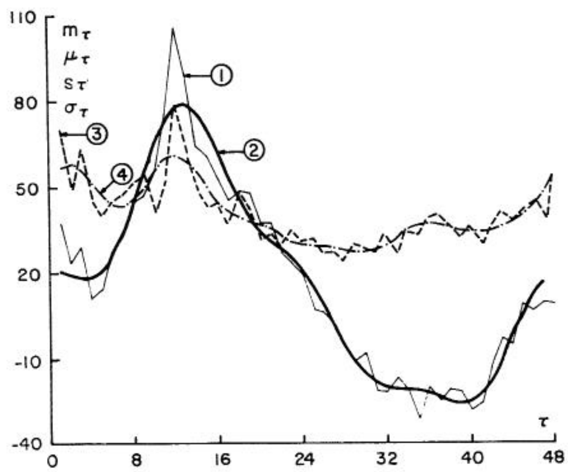

Fig. 3-6 Periodic mean, computed m_τ (1) and fitted μ_τ (2), and periodic standard deviation, computed s_τ (3) and fitted σ_τ (4), for the quarter-monthly net basin supplies of Lake Erie.

(1) Fitted periodic components μ_τ and σ_τ have a general parallelism but seem not to be proportional for these two quarter-monthly net basin supplies;

(2) Sampling fluctuations may be the reason for the nonparallelism between μ_τ and σ_τ, but some other factors may be also responsible (such as combined snow, rain, evaporation and evapotranspiration effects on the net basin supplies);

(3) Only 4-6 harmonics out of the 24 possible harmonics, in case the periodic functions would pass through every point of m_τ and s_τ series, are sufficient to fit well the periodic movements in the mean and the standard deviation; this is a large saving (parsimony) in the number of parameters in comparison with the use of all m_τ and s_τ values (96 values);

(4) The use of μ_τ and σ_τ of Figs. 3-5 and 3-6 is expected to remove the major periodicity in the mean quarter-monthly NBS series;

(5) Because it was found in Part II that the autoregressive coefficients α_1 and α_2 of $\varepsilon_{p,\tau}$ of the monthly series of five lakes are not periodic, it is implicitly concluded that α_1, α_2,...autoregressive coefficients of $\varepsilon_{p,\tau}$ of the mean quarter-monthly NBS series of Lakes Ontario and Erie also are not periodic; and

(6) Because the skewness coefficient of the independent stochastic series $\xi_{p,\tau}$ of the mean monthly NBS series for five lakes is found in Part II not to be periodic, it is assumed implicitly that the skewness coefficient of the $\xi_{p,\tau}$ series of the mean quarter-monthly NBS of Lakes Ontario and Erie also is not periodic.

The dependence model of the $\varepsilon_{p,\tau}$ component, obtained by Eq. 2-3 is found to be well approximated by the second-order linear autoregressive (Markov linear) model of Eq. 2-7. The $\varepsilon_{p,\tau}$ series is first standardized before Eq. 2-7 is used, as applied to $\varepsilon_{p,\tau}$ series in Part II.

The autoregressive coefficients α_1 and α_2 of Eq. 2-8 are estimated by

$$a_1 = \frac{r_1 - r_1 r_2}{1 - r_1^2} \quad \text{and} \quad a_2 = \frac{r_2 - r_1^2}{1 - r_1^2} , \qquad (3\text{-}4)$$

in using the historic data of the $\varepsilon_{p,\tau}$ component of the mean quarter-monthly NBS series of Lakes Ontario and Erie. Table 3-2 gives the estimates of the first three serial correlation coefficients, r_1, r_2, and r_3, with r_1 and r_2 used in Eq. 3-4 to compute a_1 and a_2 as the estimates of α_1 and α_2.

TABLE 3-2
ESTIMATES OF FIRST THREE SERIAL CORRELATION COEFFICIENTS OF THE $\varepsilon_{p,\tau}$ COMPONENT

Lake	r_1	r_2	r_3
Ontario	0.283	0.098	0.050
Erie	0.353	0.226	0.141

By using a_1 and a_2 in Eq. 2-8, the independent second-order stationary stochastic component $\xi_{p,\tau}$ is obtained. It was also shown that the first-order linear autoregressive model of dependence, Eq. 2-10, and the third-order linear autoregressive model, in the form

$$\varepsilon_{p,\tau} = \alpha_1 \varepsilon_{p,\tau-1} + \alpha_2 \varepsilon_{p,\tau-2} + \alpha_3 \varepsilon_{p,\tau-3} + \xi_{p,\tau} ; \qquad (3\text{-}5)$$

with α_1 estimated by r_1 of Table 3-2, and α_1, α_2, and α_3 estimated by using r_1, r_2, and r_3 of Table 3-2 in the appropriate equations connecting a_1, a_2, and a_3 to r_1, r_2, and r_3, did produce a less good fit for the dependence of $\varepsilon_{p,\tau}$ than the second-order linear autoregressive model.

Figures 3-7 and 3-8, respectively for the Lake Ontario and the Lake Erie, give the comparison of the results of these two models in fitting the time dependence of the $\varepsilon_{p,\tau}$ component of the mean quarter-monthly NBS series. Lines (1) in Figs. 3-7 and 3-8 are correlograms of $\varepsilon_{p,\tau}$, while the lines (2) and (3) give correlograms of $\xi_{p,\tau}$ of the two models of Eqs. 2-3 and 2-10. The tolerance limits for the variation of serial correlation coefficients r_k on the 95 percent tolerance level, are drawn in Figs. 3-7 and 3-8 for the case of independent time series, enabling the comparison of the two linear models to ascertain which of them produces the stochastic component $\xi_{p,\tau}$ which is best confined within the tolerance interval.

For all practical purposes, the dependent stochastic components $\varepsilon_{p,\tau}$ of the mean quarter-monthly NBS series of the Lakes Ontario and Erie follow well the second-order linear autoregressive model of dependence. This model was used in the form of Eq. 2-7 in computing the independent stochastic component, $\xi_{p,\tau}$.

The three-parameter lognormal probability density function of Eq. 2-15 is fitted to the frequency distributions of the $\xi_{p,\tau}$ series of mean quarter-monthly NBS of Lakes Ontario and Erie, as was done previously for the $\xi_{p,\tau}$ series of the mean monthly NBS series of five lakes. The main reasons for using the three-parameter lognormal probability function are:

(1) The $\xi_{p,\tau}$ variable, by its definition and by the way of its computation, has a negative lower boundare, so that only the three-parameter lognormal or the three-parameter gamma (Pearson Type III) probability functions, as simple functions with a minimum number of parameters to estimate, can meet this requirement.

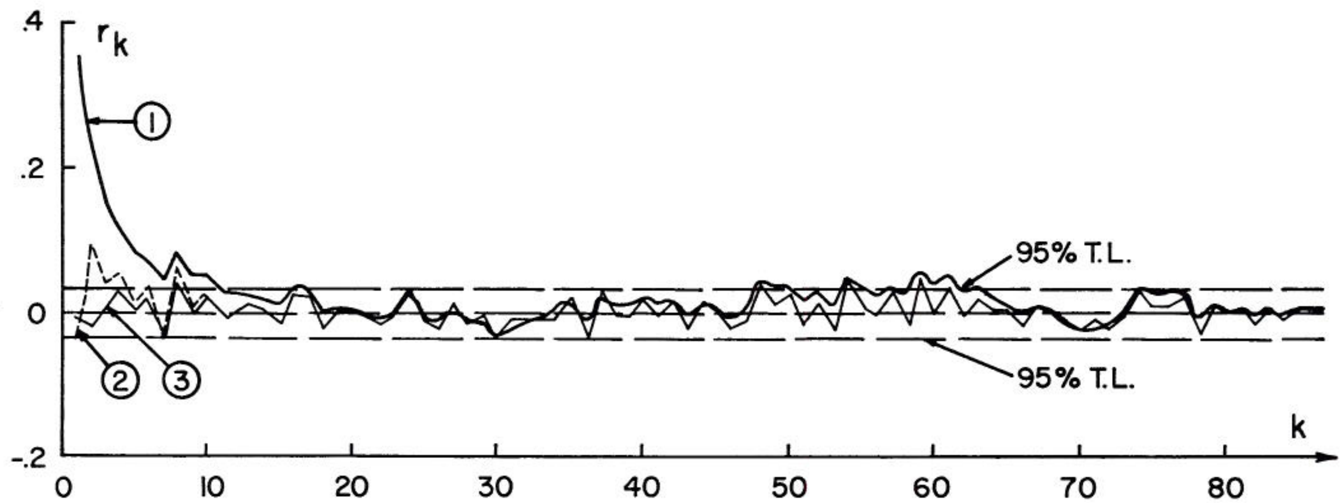

Fig. 3-7 Correlograms of stochastic components of mean quarter-monthly net basin supplies of Lake Ontario: (1) computed for $\varepsilon_{p,\tau}$ series of Eq. 2-10; (2) for the $\xi_{p,\tau}$ series of the fitted first-order autoregressive model; and (3) for the $\xi_{p,\tau}$ series of the fitted second-order autoregressive model.

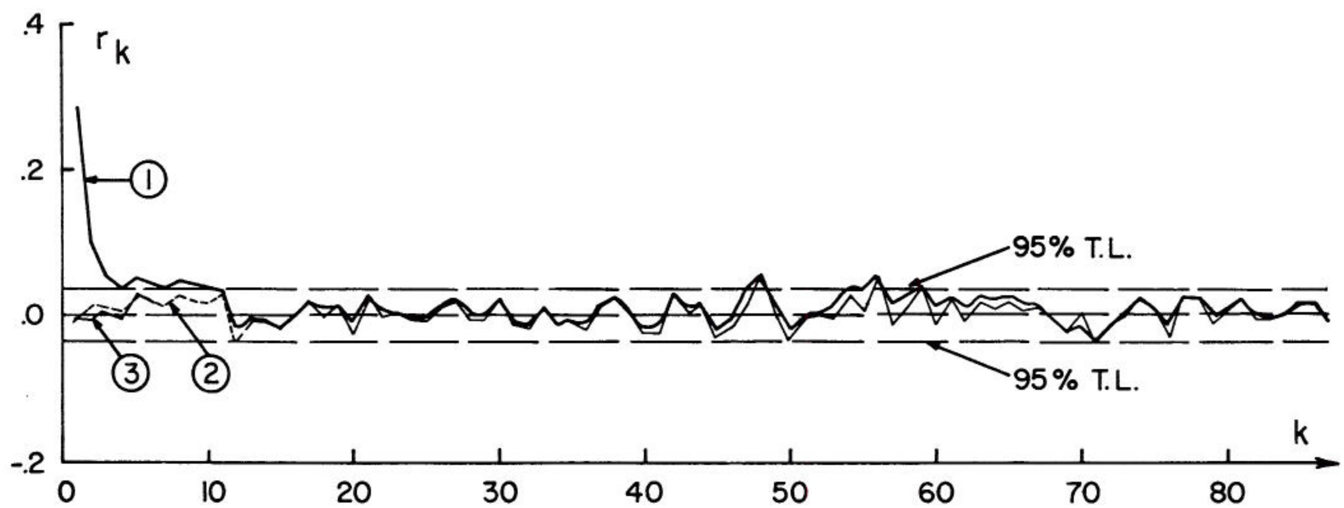

Fig. 3-8 Correlograms of stochastic components of mean quarter-monthly net basin supplies of Lake Erie: (1) computed for $\varepsilon_{p,\tau}$ series of Eq. 2-10; (2) for the $\xi_{p,\tau}$ series of the fitted first-order autoregressive model; and (3) for the $\xi_{p,\tau}$ series of the fitted second-order autoregressive model.

(2) To preserve the dependence among the $\xi_{p,\tau}$ series of these lakes, it is easy to transform the three-parameter lognormal distribution into a normal distribution in cases when the multivariate normal approach to generating new samples is used.

(3) The generated independent standard normal random numbers are easily transformed into the independent three-parameter lognormal random numbers, with given parameters.

The parameters of the three-parameter lognormal probability function, estimated from historic data on the $\xi_{p,\tau}$ series of the mean quarter-monthly NBS series of Lakes Ontario and Erie, are given in Table 3-3.

TABLE 3-3
PARAMETERS OF THREE-PARAMETER LOGNORMAL PROBABILITY DISTRIBUTION FUNCTION OF INDEPENDENT STOCHASTIC COMPONENTS OF MEAN QUARTER-MONTHLY NET BASIN SUPPLIES OF LAKES ONTARIO AND ERIE

Lake	Mean of Logarithms m_n	Standard Deviation of Logarithms s_n	Lower Boundary g
Ontario	1.7974	0.1529	-6.1046
Erie	1.6271	0.1766	-5.1695

Figure 3-9 for the Lake Ontario, and Fig. 3-10 for the Lake Erie, give the frequency density curve, the fitted three-parameter lognormal probability density function, the cumulative frequency distribution and the fitted lognormal probability distribution for the independent stochastic components $\xi_{p,\tau}$ of the mean quarter-monthly NBS series. The reasons for presenting the fits of probability functions not only as the density curves, similar to the method used in Part II, but also as the cumulative distribution curves are two:

(1) By a visual inspection the goodness of fit is usually assessed from the cumulative curves rather than from the density curves; and

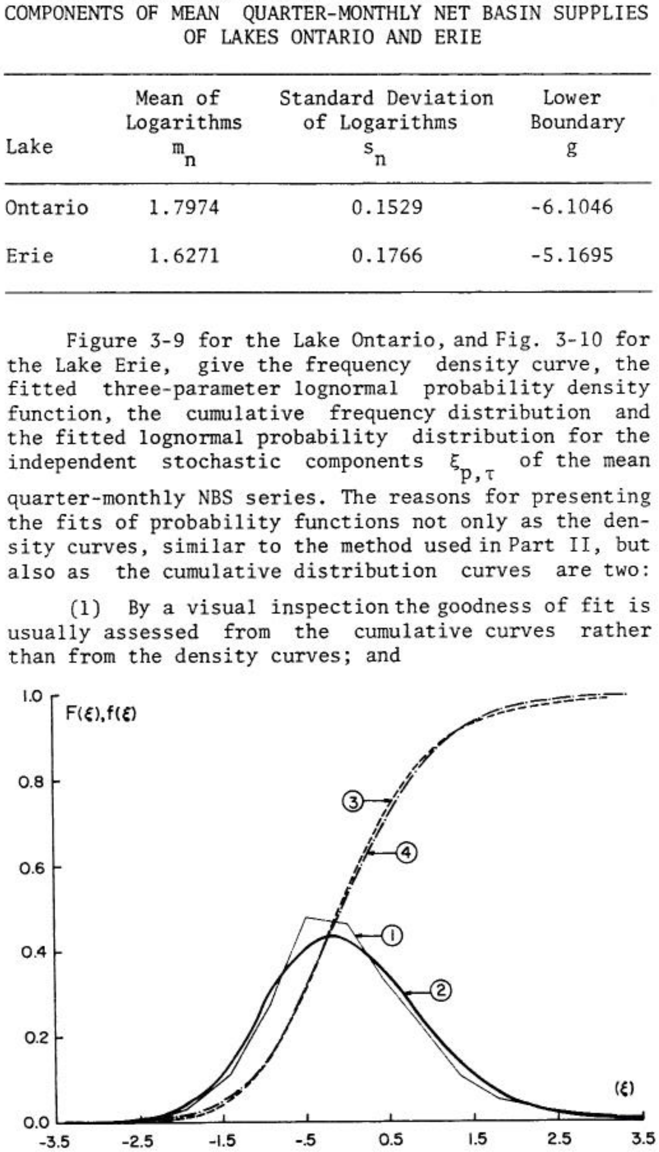

Fig. 3-9 Frequency density curve (1), the fitted three-parameter lognormal probability density function (2), the cumulative frequency distribution curve (3) and the fitted three parameter lognormal distribution function (4), for the independent stochastic component of the mean quarter-monthly net basin series of Lake Ontario.

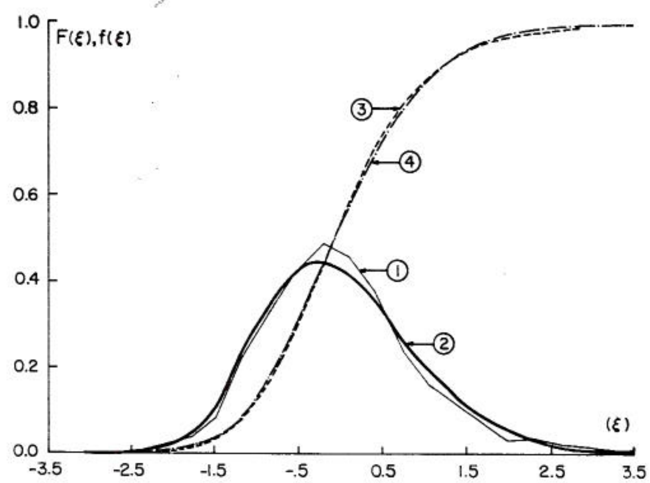

Fig. 3-10 Frequency density curve (1), the fitted three-parameter lognormal probability density function (2), the cumulative frequency distribution curve (3) and the fitted three parameter lognormal distribution function (4), for the independent stochastic component of the mean quarter-monthly net basin series of Lake Erie.

(2) The cumulative curves enable the use of the simple Smirnov-Kolmogorov statistic for the test of goodness of fit.

The Smirnov-Kolmogorov tests show that the fit is acceptable because the test statistics are as follows: 0.0238 from the table and 0.0175 from Fig. 3-9 for the $\xi_{p,\tau}$ series of the Lake Ontario, and 0.0243 from the table and 0.020 from the graph of Fig. 3-10 for the $\xi_{p,\tau}$ series of the Lake Erie. In both cases the critical values for the 95 percent probability level of the test statistic are larger than the maximum difference between the two cumulative distribution curves of Figs. 3-9 and 3-10. The fits are acceptable, although the test statistics are close to critical values.

By using the chi-square statistic the results of the tests are as follows: for the Lake Ontario, the chi-square from the tables and for the 95 percent probability level is 35.20, while the computed value is 40.26; and for the Lake Erie the tables give 35.20 while the computed value is 66.82. Regardless that the critical values are smaller, Figs. 3-9 and 3-10 show a relatively acceptable fit of the three-parameter lognormal probability function to the frequency distribution curves. It is likely that the three-parameter lognormal probability function does not give a very good fit at the extremes, thus making the chi-square statistic relatively large.

Table 3-4 gives the skewness coefficients C_s for the three variables, $x_{p,\tau}$, $\varepsilon_{p,\tau}$ and $\xi_{p,\tau}$ of the historic mean quarter-monthly NBS of Lakes Ontario and Erie. This table shows that the skewness coefficient decreases from $x_{p,\tau}$ to $\varepsilon_{p,\tau}$ and from $\varepsilon_{p,\tau}$ to $\xi_{p,\tau}$. When the periodic components μ_τ and σ_τ are removed from $x_{p,\tau}$, the dependent stochastic component $\varepsilon_{p,\tau}$ has a lower skewness coefficient than $x_{p,\tau}$. When the second-order linear autoregressive model is used to compute the $\xi_{p,\tau}$ series from the $\varepsilon_{p,\tau}$ series, the skewness coefficient of $\xi_{p,\tau}$ becomes smaller than the skewness coefficient of $\varepsilon_{p,\tau}$.

TABLE 3-4
SKEWNESS COEFFICIENTS OF THREE TYPES OF SERIES OF HISTORIC MEAN QUARTER-MONTHLY NET BASIN SUPPLIES OF THE LAKES ONTARIO AND ERIE

Lake	Skewness Coefficient		
	$x_{p,\tau}$	$\varepsilon_{p,\tau}$	$\xi_{p,\tau}$
Ontario	1.122	0.862	0.768
Erie	1.269	1.031	0.936

Chapter 8

GENERATION OF SAMPLES OF MEAN QUARTER-MONTHLY NET BASIN SUPPLIES

8.1 Regional Dependence of Two Series. In the generation of new samples of the mean quarter-monthly NBS of Lakes Ontario and Erie, the dependence among their two independent stochastic components $\xi_{p,\tau}$ must be preserved. When there are only two regional random variables for which new samples must be generated by preserving the dependence between their series, a simple regression analysis is sufficient. The new samples of the independent stochastic component of one variable may be first generated independently of the other variable. Then the residual term of the regression equation of the second variable $\xi_{p,\tau}$ against the first variable $\xi_{p,\tau}$ is generated for computing the stochastic independent component of the second variable. The use of the regression equation produces the new samples of the independent stochastic component of the second variable. The first variable is selected to be ξ_{ont} of the Lake Ontario, so that its new samples are independently generated. The second variable, ξ_{er} of the Lake Erie, is obtained by

$$(\xi_{er})_{p,\tau} = a + b(\xi_{ont})_{p,\tau} + \eta_{p,\tau} \ . \qquad (3\text{-}6)$$

The estimated regression coefficients are: $a = 0$ and $b = 0.4784$, while the correlation coefficient between ξ_{er} and ξ_{ont} is $r = 0.467$. With a and b given, the $\eta_{p,\tau}$ series of Eq. 3-6 is computed from the already available ξ_{er} and ξ_{ont} series.

Figure 3-11 gives the frequency density curve, and the fitted three-parameter lognormal probability distribution of $\eta_{p,\tau}$. The critical value of the Smirnov-Kolmogorov test statistic is 0.0243, as obtained from the table and for the 95 percent probability level, while it is 0.0180 in Fig. 3-11, showing an acceptable fit. However, the critical value of the chi-square statistic for the 95 percent probability level is 35.20, while the computed value from data is 54.19. Regardless of the fact that the computed chi-square statistic is greater than the critical value, Fig. 3-11 shows a relatively acceptable fit of the lognormal probability function, at least for the central part of the function (say from 1-99 percent of probability). The deviations at extremes are the main reason for the difference in the two values of the chi-square statistic.

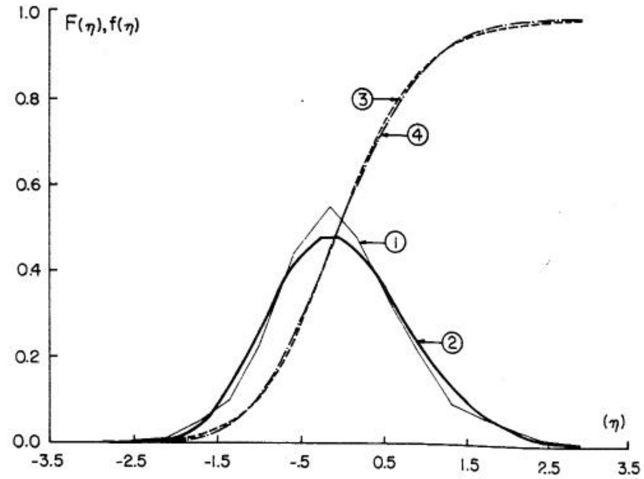

Fig. 3-11 The frequency density curve (1), the fitted three-parameter lognormal probability density function (2), the cumulative frequency distribution curve (3), and the fitted three-parameter lognormal probability distribution function (4), for the $\eta_{p,\tau}$ series of Lake Erie, computed by Eq. 3-6.

8.2 Generation of New Samples. The generation of the new 20 samples, each 50 years long (a total of 20 x 50 x 48 = 48,000 values) of the mean quarter-monthly NBS has followed the same procedure of generation as was described in Part II for the mean monthly NBS. First, two series, each 48,000 values long, of the normal standard, independent random numbers are generated. One series is transformed into the lognormal distribution with the three parameters (m_n, s_n, g) of the ξ_{ont} series of the Lake Ontario. The other series is transformed into the lognormal distribution with the three parameters of the $\eta_{p,\tau}$ series of Eq. 3-6 for the Lake Erie. By using Eq. 3-6 and the two generated series, ξ_{ont} and $\eta_{p,\tau}$, the 48,000 numbers of the ξ_{er} series are obtained. By further applying the corresponding autoregressive linear model of Eq. 2-7 to both $\xi_{p,\tau}$ series, and then by using Eq. 2-4, and taking into account the standardization of $\varepsilon_{p,\tau}$, the 48,000 numbers of $x_{p,\tau}$, the mean quarter-monthly NBS series are obtained for Lakes Ontario and Erie. Dividing the 48,000 numbers into 20 consecutive groups, each of 2400 values of $x_{p,\tau}$, the 20 new samples, each 50 years long, are obtained.

8.3 Analysis of Generated Samples. The basic parameters: the mean, the standard deviation and the skewness coefficient of each sample of the $x_{p,\tau}$ series for the Lake Ontario are given in Table 3-5, and for the Lake Erie in Table 3-6. The average of means of the 20 samples is 33.92 for the Lake Ontario, the same as the mean of the historic sample. Similarly, the average of means of the 20 samples is 16.94 for the Lake Erie, while the mean of the historic sample is 16.89. This shows a good preservation of the mean. Tables 3-5 and 3-6 give the average of standard deviations as 40.42 and 52.39, respectively. They cannot be compared with the mean standard deviations $\bar{\sigma}_\tau$ of 48 quarter-monthly because of different definitions of standard deviations, and built-in differences between these two parameters. The average skewness coefficients in

Tables 3-5 and 3-6 are 0.912 and 0.660 while the original samples of $x_{p,\tau}$ had the corresponding values 1.122 and 1.269. This is a decrease in the skewness coefficients of generated samples in comparison with the original data. The relatively smaller value of the average skewness coefficient for generated samples in case of the Lake Erie may be the result of the use of Eq. 3-6. Summing the two lognormal variables, ξ_{ont} and $\eta_{p,\tau}$, produces a variable, ξ_{er}, with a lesser skewness coefficient than if ξ_{ont} and ξ_{er} were first transformed to normal variables, and the two transformed normal variables then correlated. However, it is considered that the above differences in skewness coefficients are not crucial. The sampling errors in the skewness coefficients are very high. The second reason for the deviations between the skewness coefficients of the generated samples and the historic sample likely comes from the fact that the lognormal probability function may not be the best function to fit the extreme tails of frequency distribution curves. The differences in chi-square statistics, as shown above, also support this assertion.

8.4 Conclusion and Recommendation. The following conclusion and recommendation can be drawn for the use of the new generated samples of the mean quarter-monthly NBS of Lakes Ontario and Erie. If the lake regulations--carried out by using both the new generated samples of the mean quarter-monthly and the corresponding mean monthly NBS series, as analyzed in Part II and as generated--show substantial differences in extremes of lake levels, with the somewhat larger probabilities for a high lake level to be exceeded or for a low lake level not to be exceeded in the use of the mean quarter-monthly series than in the use of the mean monthly series, a refinement of generating technique then would be warranted. Namely, the fit of probability density functions to the $\xi_{p,\tau}$ series, or to the $\eta_{p,\tau}$ series of the Lake Erie, as defined above, should be refined. It may be shown by additional attempts, that some other functions may fit better the tails of frequency distributions of $\xi_{p,\tau}$ and $\eta_{p,\tau}$ than does the three-parameter lognormal function. Repeating the regulation procedures with a new set of 20 generated samples of the quarter-monthly and the monthly NBS series by this refined technique, the differences in properties of extreme levels of the two lakes, in using the quarter-monthly and the monthly NBS series, may then be more accurately determined.

TABLE 3-5

PARAMETERS OF 20 GENERATED SAMPLES, EACH 50 YEARS LONG, OF MEAN QUARTER-MONTHLY NET BASIN SUPPLIES OF LAKE ONTARIO

Sample	Mean	Standard Deviation	Skewness Coefficient
1	33.90588	40.43557	.89378
2	33.51754	38.36393	.94793
3	34.07477	41.12573	1.00206
4	33.71370	39.98800	.84853
5	34.20152	41.13454	.97354
6	33.96379	40.37421	.91199
7	33.72621	40.48558	.93618
8	33.78266	40.40658	.81592
9	34.21029	42.04338	.97776
10	33.34848	38.43994	.86102
11	33.94721	41.63354	.89228
12	34.03116	39.31787	.76881
13	33.94093	40.94372	1.02243
14	34.00472	40.35714	.85794
15	33.95356	40.76450	.94466
16	34.14399	40.96988	.94850
17	33.89417	39.72068	.80185
18	33.91233	40.31172	.84379
19	34.17697	41.05916	.97847
20	33.90994	40.63312	1.00667
Average	33.917491	40.425439	.911706

TABLE 3-6

PARAMETERS OF 20 GENERATED SAMPLES, EACH 50 YEARS LONG, OF MEAN QUARTER-MONTHLY NET BASIN SUPPLIES OF LAKE ERIE

Sample	Mean	Standard Deviation	Skewness Coefficient
1	16.91320	52.59721	.58602
2	16.82120	52.07000	.64685
3	17.08726	52.98467	.66651
4	16.58983	52.21053	.69666
5	17.31193	53.81653	.73299
6	16.95504	53.04667	.71606
7	17.14300	52.87276	.64686
8	16.57884	51.59427	.70001
9	16.95450	51.66996	.65689
10	16.85595	50.65117	.49309
11	16.99510	53.66775	.62276
12	16.85785	51.65828	.63080
13	17.03101	52.51445	.66985
14	16.85268	51.98974	.66893
15	17.08246	52.94295	.68316
16	17.28940	52.11988	.69848
17	16.75696	51.76269	.70377
18	16.62019	52.30916	.64904
19	17.12422	52.45803	.71294
20	17.06847	52.87705	.62630
Average	16.944455	52.390688	.660398

PART IV
GENERATION OF SAMPLES OF MONTHLY WINTER FLOW RETARDATIONS IN THE CONNECTING CHANNELS OF GREAT LAKES

This part refers to the structural analysis of historic data, made available by the U.S. Corps of Engineers, Chicago Division, of the monthly winter flow retardations in the connecting channels of the Great Lakes (the St. Mary's River, the St. Clair-Detroit River, the Niagara River, and the St. Lawrence River), and the generation by the experimental (Monte Carlo) method of 20 samples each 50 years long of each of the four series of water flow retardations. It contains the following: tests of homogeneity in the mean and standard deviation of the four series, removing trends in the mean and the standard deviation when they are found to be nonhomogeneous, structural analysis and mathematical description of the four series found or made homogeneous, generation of 20 samples each 50 years long for the four flow retardation series, and analysis of generated samples.

Chapter 9

CORRECTIONS FOR NONHOMOGENEITY

9.1 Tests of Homogeneity of Winter Flow Retardation Series. Tests of homogeneity are made in monthly series of winter flow retardation in the connecting channels of the Great Lakes. The names of channels in this text are used interchangeably as: the St. Mary's River or Michigan-Huron, the St. Clair-Detroit Rivers or St. Clair, the Niagara River or Ontario, and the St. Lawrence River or St. Louis. The tests are made by using the split-sample approach and by ascertaining whether differences between the means of the two subsamples are or are not significantly different from zero at the 95 percent probability level of significance. The t-statistic of Eq. 2-1 is used.

Figure 4-1 presents the series of the total annual retardations $\bar{x}$ (obtained by summing all monthly values of winter retardation in any given year) for the connecting channel of the St. Mary's River; it shows a clear downward trend. The split-sample test shows the difference between the means of two subsample to be significantly different from zero. Also, the test shows that the slope of the trend line is significantly different from zero on the 95 percent probability level. Similarly, Fig. 4-2 gives the change with time for every consecutive two-year period of the standard deviation of monthly winter retardation in the St. Mary's River. This series of s_2 values is obtained by finding the mean of nonzero values of monthly series for each two years in sequence, and then the standard deviation is computed. Two years are used instead of one year, because any year has only 4-5 nonzero values. The s_2 series is, therefore, a measure whether there is also a trend in the standard deviation, when it is found to be present in the mean of annual retardation series. Figure 4-2, and tests, particularly that the trend slope is significantly different from zero, confirm that the trend is also significant in the standard deviation for the connecting channel of the St. Mary's River.

Figures 4-3 and 4-4 show the same graphs as Figs. 4-1 and 4-2, only in this case the winter retardations are for the connecting channel of St. Clair-Detroit Rivers. The linear trends in both series, the annual retardation series and the series of consecutive two-year standard deviation of monthly winter retardations, have the slopes which are significantly different from zeros on the 95 percent probability level of significance.

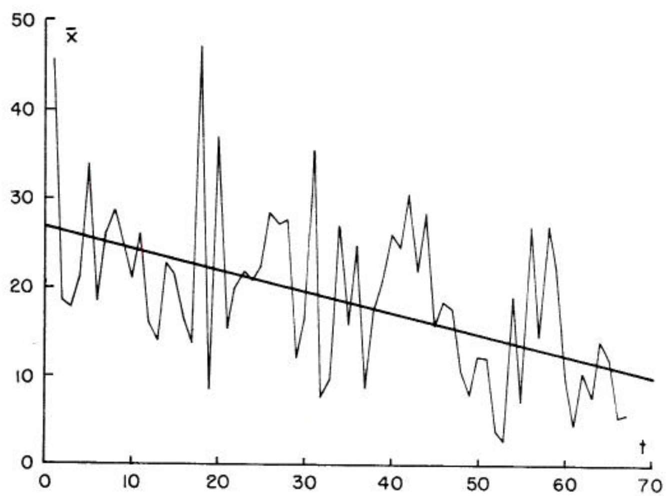

Fig. 4-1 Series of annual retardations in the connecting channel, St. Mary's River (Michigan-Huron), with fitted downward significant linear trend.

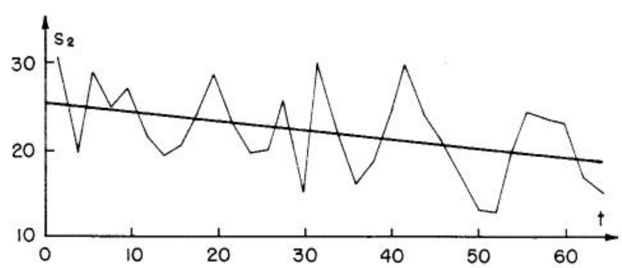

Fig. 4-2 Series of consecutive values of two-year standard deviation of monthly winter retardations in the lakes connecting channel of St. Mary's River (Michigan-Huron).

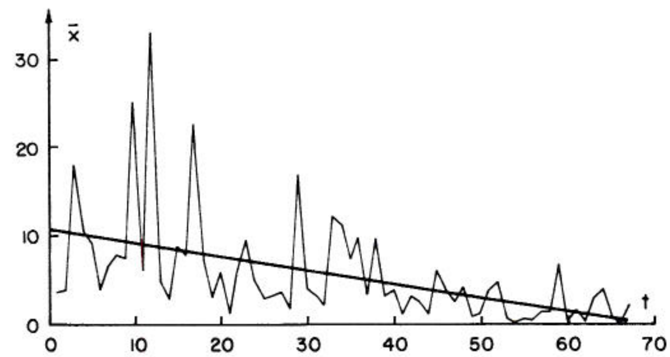

Fig. 4-3 Series of annual retardations R in the connecting channel, St. Clair-Detroit Rivers, with fitted downward significant linear trend.

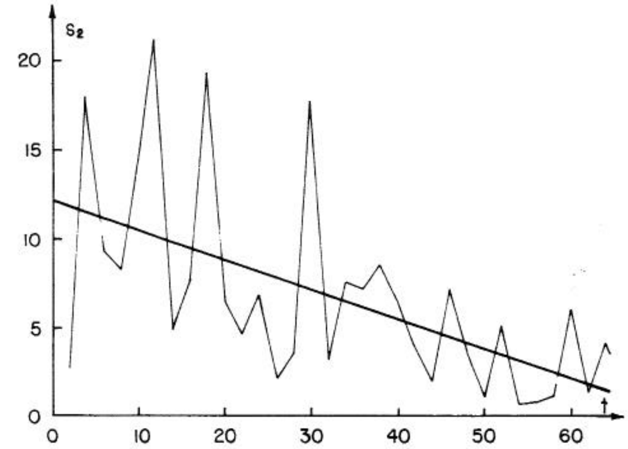

Fig. 4-4 Series of consecutive values of two-year standard deviation of monthly winter retardations in the lakes connecting channel of St. Clair-Detroit Rivers.

Figures 4-5 through 4-8 give the same series as Figs. 4-1 and 4-2 except that they refer, respectively, to the Niagara River (Ontario) and the St. Lawrence River (St. Louis). These four figures do not show the significant trends in the series analyzed.

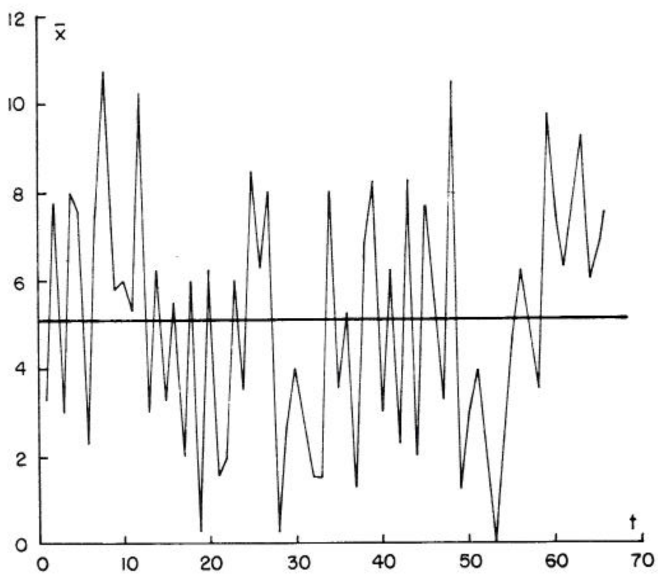

Fig. 4-5 Series of annual retardations in the connecting channel of the Niagara River (Ontario, in this paper), with no significant trend in the series.

Two cases out of four show the significant downward trends in monthly series of winter flow retardations in connecting channels of Great Lakes. The most attractive explanation--without a special study of this phenomenon--is the heat release by man-made uses, either into the air or into the water, or both. The analysis of factors which have produced a steady decrease in flow retardations may show a different cause (say the decrease of the average water levels during the winter by the way lakes were operated), or the combination of several factors. This analysis is outside the scope of this text.

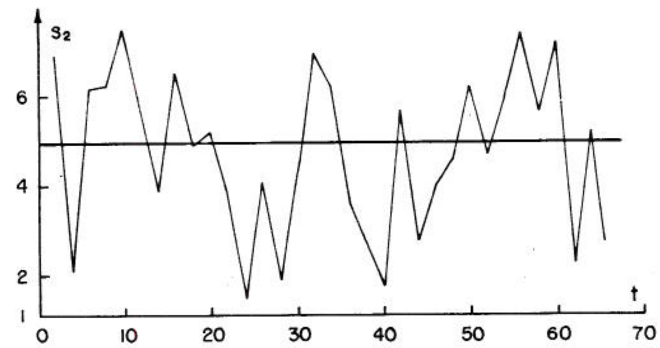

Fig. 4-6 Series of consecutive values of two-year standard deviation of monthly winter retardations in the Niagara River (Ontario, in this paper), with no significant trend.

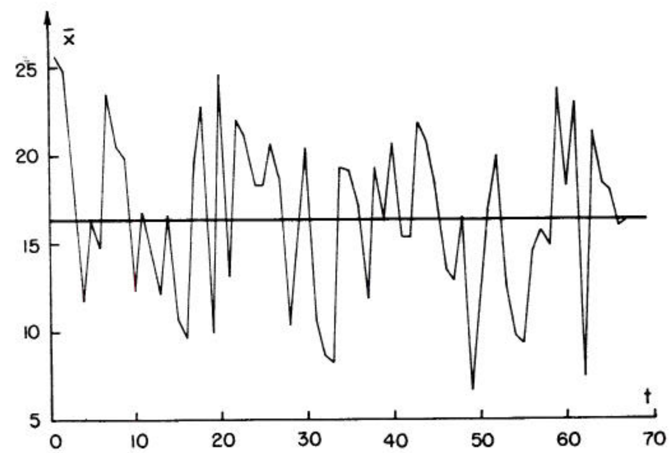

Fig. 4-7 Series of annual retardations in the connecting channel of the St. Lawrence River (St. Louis, in this paper), with no significant trend in the series.

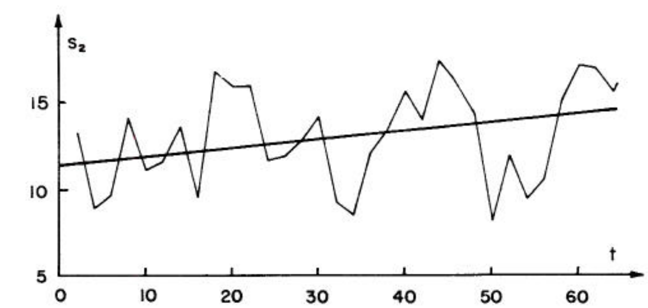

Fig. 4-8 Series of consecutive values of two-year standard deviation of monthly winter retardations in the St. Lawrence River (St. Louis, in this paper), with no significant trend.

9.2 Removing Trends in Mean and Standard Deviation of Winter Flow Retardation Series. The generation of new samples of monthly series of winter flow retardation of the St. Mary's and the St. Clair-Detroit Rivers by using the properties of historic data would over-estimate the future winter flow retardations. The trend in a hydrologic time series must be removed if it is not expected either to be repeated, or if it will not occur at all in the future. Three practical solutions seem attractive in this case:

(1) To extrapolate the trend into the future, and when it hits the zero value, to consider that there is not a winter flow retardation from that time on;

(2) To neglect completely the future winter flow retardations, particularly for the connecting channels of the St. Clair-Detroit Rivers; and

(3) To remove the trend from both series (from the mean of the annual series and from the series of the two-year standard deviation), and to use the general mean and the standard deviation of historic monthly series of winter flow retardation of the last 20 years or so, as the expected values of these two parameters in the future samples of 50 years.

This third alternative has been suggested to the U.S. Corps of Engineers, Chicago Division. It is used in the following procedure of removing trends and making the time series homogeneous.

Only the linear trends are used, because any non-linear trend though easy to fit may have small justification, and because the difference between the non-linear and linear trends may be partly or fully the result of sampling variations.

The trend in the monthly means of the St. Mary's River is

$$\bar{x} = 27.049 - 0.234\, t \,, \qquad (4\text{-}1)$$

with the correlation coefficient between $\bar{x}$ and t being $r = -0.474$.

Similarly the trend in the two-year standard deviation series is

$$s_2 = 25.878 - 0.308\, t \,, \qquad (4\text{-}2)$$

with the correlation coefficient between s_2 and t being $r = -0.555$. For this case, and as an example, the computed t-statistic of Eq. 2-1 is $t = 5.458$, while the t-critical value for the nonsignificant difference of the two standard deviations is $t_{cr} = 2.000$.

The general mean and the general standard deviation of the monthly series of winter flow retardations of the St. Mary's River for the historic period are: $\bar{x} = 19.108$ and $s = 21.156$. The last 20 years of data give these parameters as: $\bar{x} = 12.741$ and $s = 15.401$ in the units of original data, supplied by U.S. Corps of Engineers, Chicago Division. The removal of the two trends and reducing the historic series to the new mean of 11.840 and the standard deviation at 16.196 is by

$$x_t^* = \frac{x_t - 27.049 + 0.234\, t}{25.878 - 0.308\, t} \times 15.401 + 12.741 \,, \qquad (4\text{-}3)$$

in which x_t is the original historic series, t is the position of each month with winter flow retardation, and x_t^* is the new homogeneous historic series, reduced to the mean and standard deviation of the last 20 years of historic data.

Figures 4-9 and 4-10 give the annual series of winter flow retardation and the series of two-year standard deviation, of the new x_t^* monthly series of winter flow retardations of Eq. 4-3, after the two trends are removed and the mean and standard deviation of the last 20 years added.

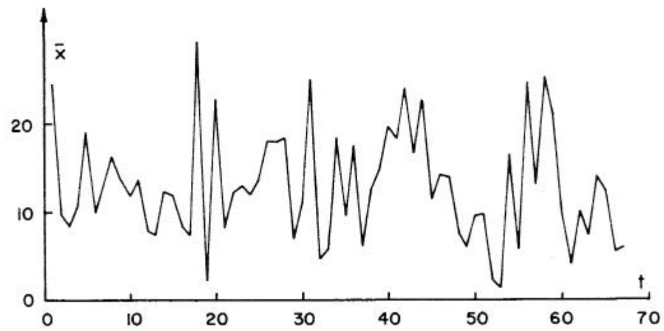

Fig. 4-9 Series of annual flow retardation for the new homogeneous historic monthly series of winter retardations of St. Mary's River (trends removed).

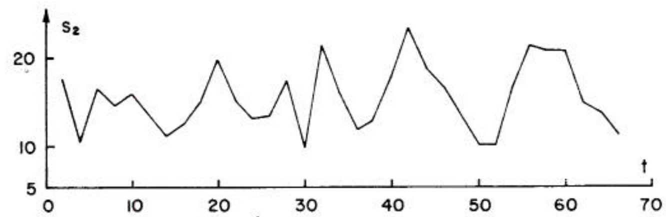

Fig. 4-10 Series of two-year standard deviation for the new homogeneous historic monthly values of winter retardations of St. Mary's River (trends removed).

Similarly as for the St. Mary's River, the trends in the monthly series of winter flow retardations of the St. Clair-Detroit Rivers are removed and the homogeneous series is adjusted to have the mean and standard deviation of the last 20 years of historic data. The trend in the mean is

$$\bar{x} = 10.849 - 0.154\, t \,, \qquad (4\text{-}4)$$

with $r = -0.493$, and the trend in the standard deviation, s_2, is

$$s_2 = 12.172 - 0.329\, t \,, \qquad (4\text{-}5)$$

with $r = -0.566$. The general mean and standard deviation for the historic series are: $\bar{x} = 5.600$ and $s = 9.451$, while these values for the last 20 years are: $\bar{x} = 2.001$ and $s = 3.281$. Then the new homogeneous series is

$$x_t^* = \frac{x_t - 10.849 + 0.154\, t}{12.172 - 0.329\, t} \times 3.281 + 2.001 \,, \qquad (4\text{-}6)$$

or it is homogeneous with the trends removed and with the basic parameters $(\bar{x}, s)$ of the last 20 years of historic data. Figures 4-11 and 4-12 give the same graphs for St. Clair-Detroit Rivers as Figs. 4-9 and 4-10 give for St. Mary's River.

The two monthly series, x_t^*, of winter flow retardation for the St. Mary's and St. Clair-Detroit Rivers, as homogeneous, are used with the two historic monthly series, x_t, of winter flow retardation of the Niagara River and the St. Lawrence River, in the basic structural analysis, mathematical description, and the generation of new samples of winter flow retardations.

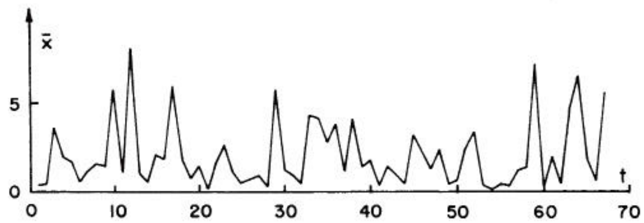

Fig. 4-11 Series of annual flow retardation for the new homogeneous historic monthly series of winter retardations of St. Clair-Detroit Rivers (trends removed).

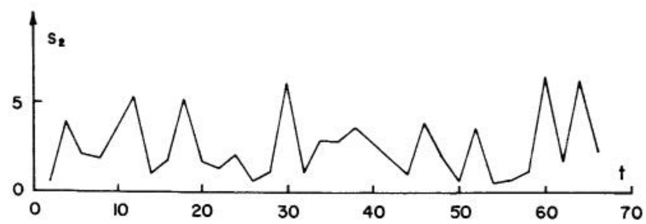

Fig. 4-12 Series of two-year standard deviation for the new homogeneous historic monthly values of winter retardations of St. Clair-Detroit Rivers (trends removed).

Chapter 10

STRUCTURAL ANALYSIS AND MATHEMATICAL DESCRIPTION OF MONTHLY SERIES OF WINTER FLOW RETARDATION FOR FOUR CONNECTING CHANNELS OF GREAT LAKES

10.1 Analysis and Description. The structural analysis and mathematical description of the monthly series of winter flow retardations follow approximately the same procedure as was used for the monthly and quarter-monthly NBS series of Great Lakes, as described in Parts II and III. The only major modification is that the monthly series of winter flow retardations have only either four or five winter months with the mean flow retardations greater than zero, the remaining eight or seven months being zero values. These series are intermittent processes, with a run of nonzero mean values for 4-5 winter months, and a run of zero values for the next 8-7 months. Only the Niagara River has four nonzero mean monthly values, though in some of the years there are zero flow retardations even during these four months. The other three monthly series (St. Mary's River, St. Clair-Detroit Rivers, St. Lawrence River) have five winter months with nonzero mean flow retardations.

Figures 4-13 and 4-14 give the mean monthly winter flow retardations, m_τ, and the monthly standard deviations, s_τ, for five months of nonzero mean monthly values for all the connecting channels except the Niagara River, for which only the four winter months have nonzero mean monthly flow retardations, and respectively for the Lakes Michigan-Huron, St. Clair, Ontario, and St. Louis.

10.2 Periodicity in Parameters. The m_τ and s_τ values of Figs. 4-13 and 4-14 may be used directly in a nonparametric method to obtain the corresponding $\varepsilon_{p,\tau}$ series by

$$\varepsilon_{p,\tau} = \frac{x_{p,\tau} - m_\tau}{s_\tau} . \qquad (4\text{-}7)$$

However, in order to be consistent with the previous work on the analysis of the mean monthly and mean quarter-monthly NBS series, the periodic functions μ_τ and σ_τ are fitted to m_τ and s_τ series. Because of only 4 or 5 values of m_τ and s_τ available for any one individual year, only the two harmonics are sufficient in fitting μ_τ and σ_τ to m_τ and s_τ. This is equivalent of the periodic functions passing through all points m_τ and s_τ. In other words, if Eq. 4-7 is used with the standard computer program in fitting μ_τ to m_τ and σ_τ to s_τ, this is equivalent to using Eq. 4-7 from the beginning.

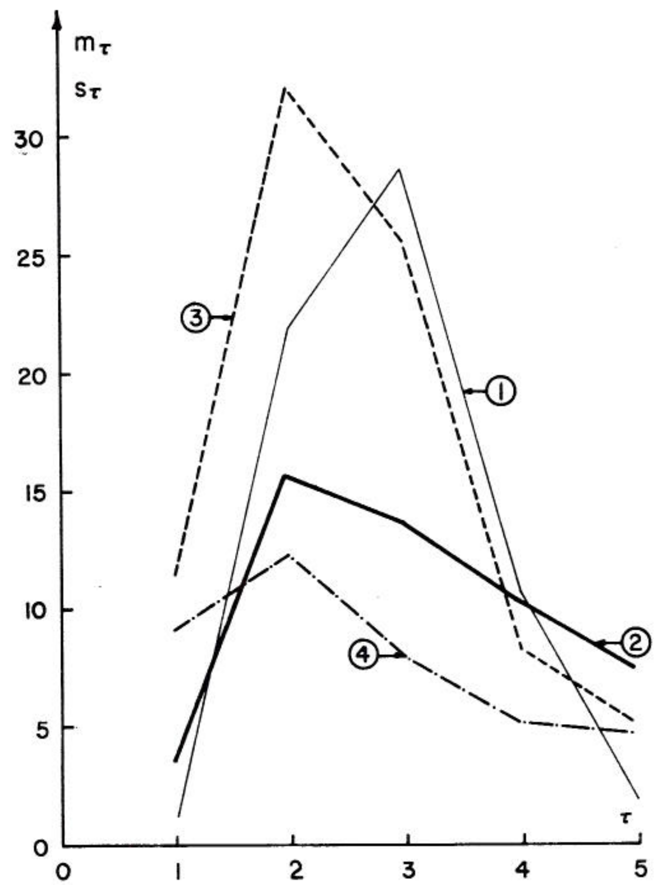

Fig. 4-13 Means, m_τ (1) and standard deviation s_τ (2) for the St. Mary's River (Michigan-Huron), and means, m_τ (3) and standard deviation s_τ (4) for the St. Louis River, of the monthly series of winter flow retardations for these two connecting channels.

10.3 Time Dependence of Stochastic Component. The serial correlation coefficients, r_1, r_2, and r_3 of the four $\varepsilon_{p,\tau}$ series, obtained by Eq. 4-7 are given in Table 4-1. Because of intermittent series, the computation of these three coefficients did use the pairs of the $\varepsilon_{p,\tau}$ values only inside each uninterrupted run of nonzero values of $\varepsilon_{p,\tau}$.

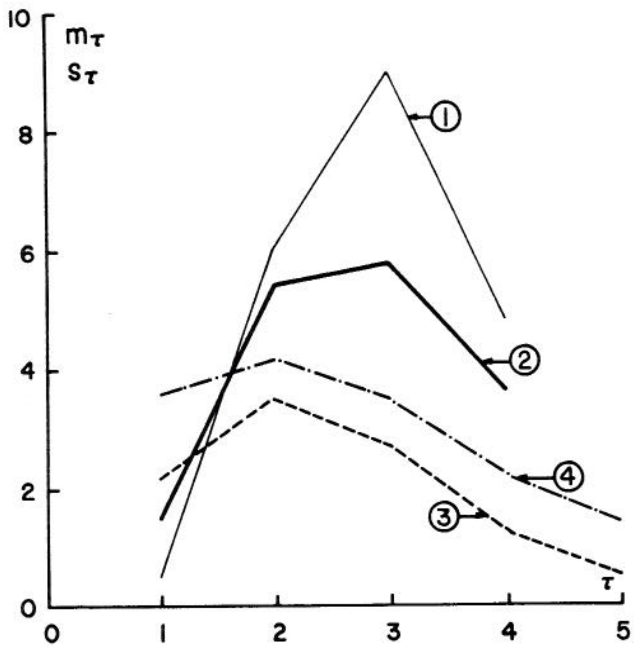

Fig. 4-14 Means, m_τ (1) and standard deviation s_τ (2) for Niagara River (Ontario) and means, m_τ (3) and standard deviation s_τ (4) for St. Clair-Detroit Rivers (St. Clair), of the monthly series of winter flow retardations for these two connecting channels.

TABLE 4-1
FIRST THREE SERIAL CORRELATION COEFFICIENTS OF THE $\varepsilon_{p,\tau}$ SERIES OF FOUR MONTHLY WINTER FLOW RETARDATION SERIES OF CONNECTING CHANNELS OF GREAT LAKES

Connecting Channel	Serial Correlation Coefficients r_1	r_2	r_3
St. Mary's River	0.286	0.085	0.030
St. Clair-Detroit Rivers	0.448	0.112	0.043
Niagara River	0.374	-0.009	-0.073
St. Lawrence River	0.229	0.137	0.064

The first-order and the second-order autoregressive linear models are tested for the $\varepsilon_{p,\tau}$ series of the monthly winter flow retardation, using either r_1 of Table 4-1 and Eq. 2-10 with ρ_1 estimated by r_1, or r_1 and r_2 of Table 4-1 and Eq. 2-7 with α_1 and α_2 of Eq. 2-8 estimated by replacing ρ_1 and ρ_2 by r_1 and r_2. Both the first-order and the second-order models are used in order to compute the $\xi_{p,\tau}$ series, for each of the four monthly winter flow retardation series. The test whether the first-order or the second-order model fits better cannot be performed by using a long correlogram, because the truncated series has the correlogram also truncated, and makes it unfeasible to study the correlogram with more than four r_k values, with $k = 1, 2, 3$, and 4. The correlograms of the $\xi_{p,\tau}$ series, computed both by using the Markov I (first-order autoregressive) model and the Markov II (second-order autoregressive) model, for the four monthly winter flow retardations show that only r_3 of the $\xi_{p,\tau}$ series of both Models I and II of the St. Mary's River is outside the tolerance limits. No other r_1, r_2, or r_3 value exceeds the tolerance limits on the 95 percent probability level of significance. The first-order linear autoregressive model is selected, and the four $\xi_{p,\tau}$ series are computed by using it, because no substantial improvement was shown in correlograms of the $\xi_{p,\tau}$ series when computed by using the second-order model in comparison with the computation by the first-order model.

10.4 Probability Distribution of Independent Stochastic Component. The next step in the analysis of monthly winter flow retardation series was the fit of probability density functions to the four independent stochastic components $\xi_{p,\tau}$ of these series, computed by the first-order autoregressive model from the $\varepsilon_{p,\tau}$ series, or by Eq. 2-13.

Table 4-2 gives the estimated parameters of the three probability functions: normal, three-parameter lognormal, and three-parameter gamma. The alpha (shape) parameter of the three-parameter gamma function of $\xi_{p,\tau}$, in the case of the first three connecting channels, is very high (α = 295, 326, and 302). Therefore, these values show that the distribution is normal. There was no point in testing the goodness of fit of the gamma function. The tests for the fit of the normal and the three-parameter lognormal functions are then carried out by using the chi-square statistic. For these two probability functions the fits to the $\xi_{p,\tau}$ series in the cases of the first three connecting channels are not especially good, because the computed chi-squares are much greater than the critical chi-square values on the 95 percent probability level of significance, with the computed chi-squares ranging from 142 through 248, while the critical chi-squares values are either 22.36 or 21.00 for these two functions. The fourth case, the St. Lawrence River, shows an acceptable fit, particularly for the three-parameter lognormal function.

Figure 4-15 shows the cumulative frequency distribution curves and the fitted normal probability distribution functions for the $\xi_{p,\tau}$ series of monthly winter flow retardations, with parameters given in Table 4-2. The fits are relatively acceptable by a visual inspection. The major reason for the large deviations of computed and critical chi-square values are not the skewness factors, but rather the high frequency densities in the center of distributions. As an example, Fig. 4-16 shows the absolute class frequencies for the 13 unequal class intervals, but with these intervals having the same probability of 1/13 of the normal probability distribution function, in case of the $\xi_{p,\tau}$ series of the Niagara River. For the normal function to be a very good fit, the 13 absolute class frequency of the $\xi_{p,\tau}$ series should also be very close, fluctuating about the average of the 13 values. The central class interval for the case of $\xi_{p,\tau}$ for the Niagara River shows a particular spike at $\xi = -0.40$ to $\xi = -0.50$, which is nearly four times as large as the

TABLE 4-2

PARAMETERS IN FITTING PROBABILITY FUNCTIONS TO FREQUENCY DISTRIBUTIONS OF THE $\xi_{p,\tau}$ SERIES OF MONTHLY WINTER FLOW RETARDATIONS IN THE CONNECTING CHANNELS OF GREAT LAKES

Function	Parameter	St. Mary's River	St. Clair-Detroit Rivers	Niagara River	St. Lawrence River
Normal	Mean	0.000	0.000	0.000	0.000
	Standard Deviation	0.965	0.914	6.944	0.977
	Chi-square critical	22.362	22.362	22.362	22.362
	Chi-square computed	206.851	248.785	232.179	23.352
Three-Parameter Lognormal	Lower Bound	-15.997	-15.997	-4.212	-6.393
	Mean of Logarithms	2.770	2.771	1.415	1.844
	St. Dev. of Logs.	0.057	0.055	0.211	0.152
	Chi-square critical	21.000	21.000	21.000	21.000
	Chi-square computed	176.379	141.704	213.313	13.227
Three-Parameter Gamma	Lower Bound	-15.988		-15.988	-4.529
	Alpha (shape)	294.895	325.500	301.581	21.494
	Beta (scale)	0.054	0.049	0.053	0.211

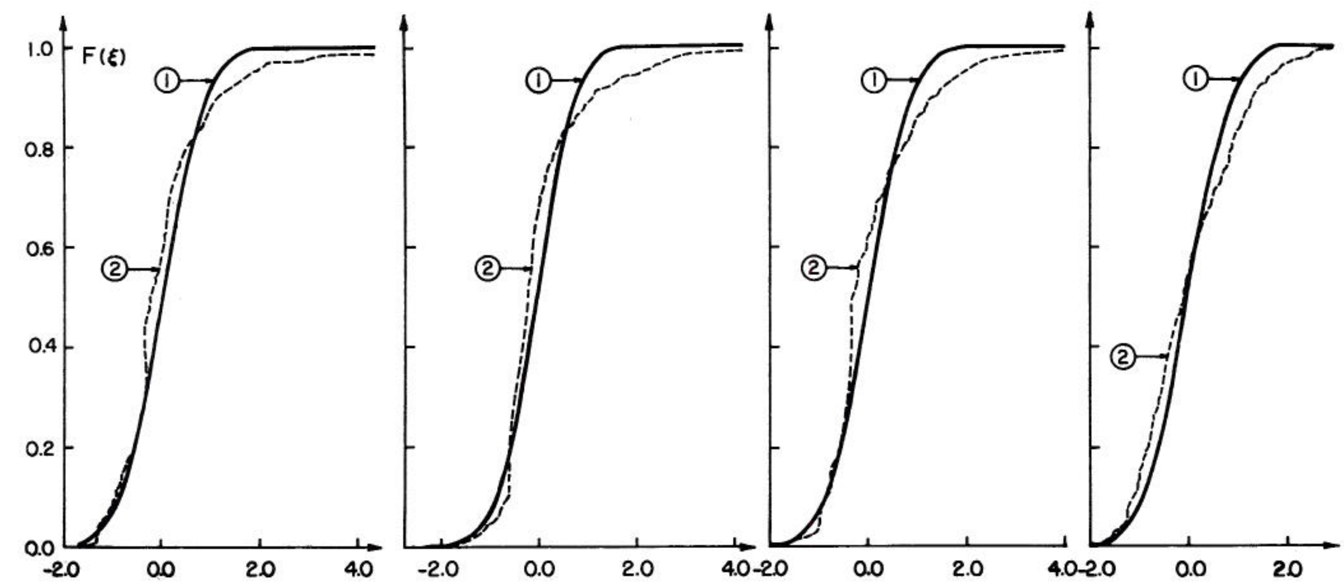

Fig. 4-15 Fitting the normal probability distribution function (1) to the cumulative frequency distribution (2) of the independent stochastic component, $\xi_{p,\tau}$ of the monthly winter flow retardation series of the four connecting channels: (I) St. Mary's River; (II) St. Clair-Detroit Rivers; (III) Niagara River, and (IV) St. Louis River.

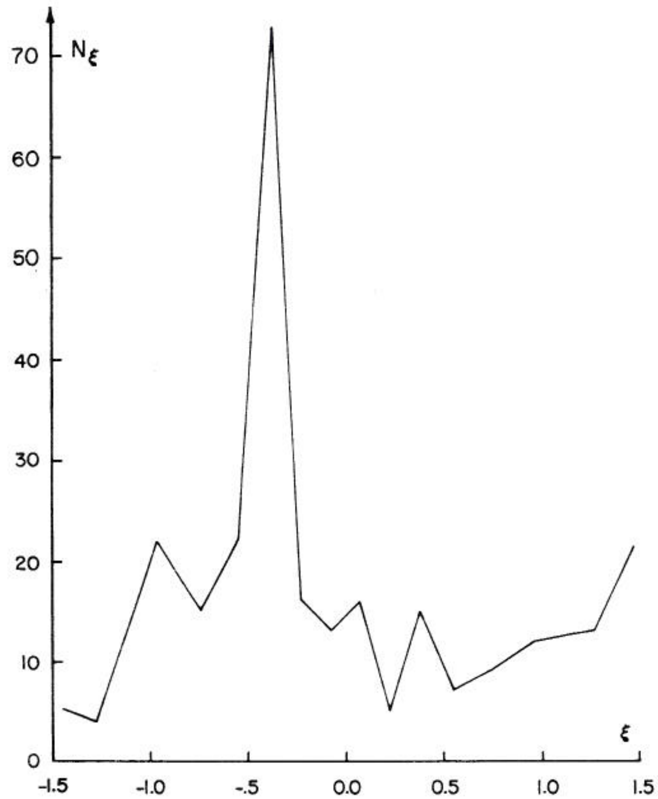

Fig. 4-16 Absolute frequencies, N_ξ, for the thirteen class intervals of equal probabilities following the normal distribution function, of the independent stochastic component, $\xi_{p,\tau}$, of monthly winter flow retardations of Niagara River (Ontario).

average of the other 12 values. Similarly, in other cases the absolute frequencies for the unequal class intervals of equal probabilities of the normal function, used in computing the chi-square statistic of $\xi_{p,\tau}$, show large variations. As for the normal function, the results of fits of the three-parameter lognormal function are shown in Fig. 4-17, demonstrating to be close to those of the normal function, though the lognormal function shows better fits than the normal function.

The dilemma is then either to continue to search for the new probability distribution functions, such as the normal function transformed by using the orthogonal polynomials, which would fit better the four $\xi_{p,\tau}$ series than the normal or the three-parameter lognormal function, or to accept the fits by this latter function. The reasons for accepting the fits by the lognormal function are:

(1) Significant errors must exist in the estimates of monthly winter flow retardations, so that a high level of goodness of fit cannot be justified.

(2) The goodness of fit is of the same order for many currently used hydrologic frequency distributions.

(3) Uncertainties in future trends of ice effects on flow retardations in the connecting channels of the Great Lakes--as demonstrated by the trends in the mean and the standard deviation of monthly winter flow retardations of the St. Mary's River and the St. Clair-Detroit Rivers--do not justify a close reproduction of the four historic $\xi_{p,\tau}$ series in the generation of new samples of monthly winter flow retardations.

In conclusion, the three-parameter lognormal distribution function is used for the four $\xi_{p,\tau}$ series, with parameters as estimated and given in Table 4-2, for monthly winter flow retardations in the connecting channels of the Great Lakes.

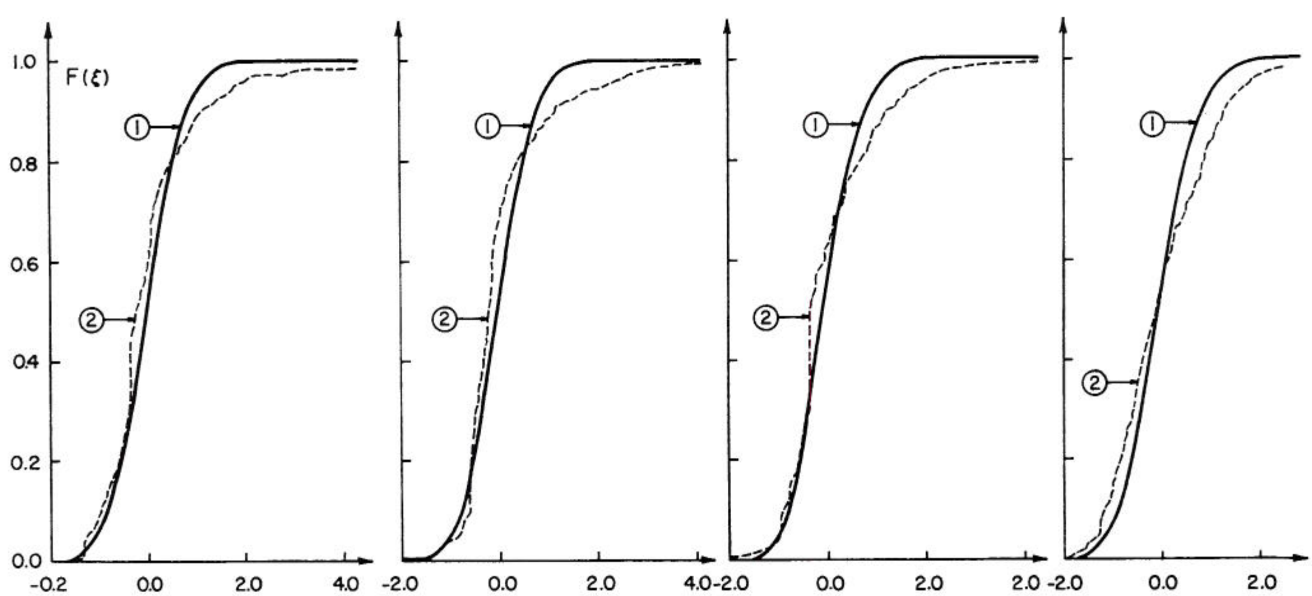

Fig. 4-17 Fitting the three-parameter lognormal probability distribution function (1) to the cumulative frequency distribution (2) of the independent stochastic component, $\xi_{p,\tau}$ of the monthly winter flow retardation series of the four connecting channels: (I) St. Mary's River; (II) St. Clair-Detroit Rivers; (III) Niagara River, and (IV) St. Louis River.

Chapter 11

GENERATION OF NEW SAMPLES

11.1 Basic Approach in Generating New Samples of Winter Flow Retardations. The basic concept in taking into account the monthly winter flow retardations in the connecting channels of the Great Lakes and in generating their new samples is to use the historic series of flow retardations in these channels. These series are then structurally analyzed and mathematically described to produce the four mutually dependent but sequentially independent stochastic components $\xi_{p,\tau}$ of the series of the four connecting channels. In this concept the basic hypotheses are as follows:

(1) That the sequentially independent stochastic components $\xi_{p,\tau}$ of winter flow retardation series of the four channels are mutually dependent random variables, similarly as they were sequentially independent stochastic components but mutually dependent random variables for the five mean monthly NBS series of the Great Lakes.

(2) That the sequentially independent stochastic components of monthly winter flow retardations are not dependent on the sequentially independent stochastic components of the mean monthly NBS series of adjacent lakes; and

(3) That the winter flow retardation series in generating new samples are not dependent on the water levels of the Great Lakes, in other words, on regulation patterns.

The approach used in this Part IV in generating the new 20 samples, each 50 years long, of the monthly series of winter flow retardations has followed the above three hypotheses. The first hypothesis is justified and is easy to prove, because the general freezing conditions usually occur all over the Great Lakes region, producing the mutually dependent random variables of winter flow retardations, in general, and of their independent stochastic components, in particular. The second hypothesis might not be fulfilled rigorously, for the simple reason that large winter freezing conditions which produce icing and flow retardations, may be associated with very low NBS values, with a negative correlation coefficient. However, it might come out that the dependence between the sequentially independent stochastic components of these two sets of random variables are relatively small to be of a practical importance. These eventual connections of negative correlation are not tested. The third hypothesis may be crucial in assessing how good the use of the generated new samples of winter flow retardation series may be. It is safe to assume that the winter flow retardations are a function also of the total discharge which would flow through channels without ice. Therefore, the flow retardation series in a connecting channel must depend on the levels of upstream and downstream lakes, if the upstream lakes are not controlled by regulating structures. This hypothesis was not tested.

If it is found that the independent stochastic components of winter flow retardation depend on: (1) the freezing conditions (climatic random variables), (2) the net basin supplies, and (3) the lake levels, then the generation of new samples of winter flow retardations should follow a different approach. Therefore, the results of Part IV in the generation of new samples of monthly winter flow retardations are based on the above three fundamental hypotheses, and cannot be better than the hypotheses.

11.2 Generation of New Samples. Because the independent stochastic components $\xi_{p,\tau}$ of winter flow retardations of the four connecting channels of the Great Lakes are mutually dependent random variables, the generation of new samples must preserve the regional dependence in $\xi_{p,\tau}$. Table 4-3 gives the correlation matrix in the form of correlation coefficients. This table shows that the cross correlation coefficients at the lag zero between the pairs of the $\xi_{p,\tau}$ series are relatively small, ranging from 0.072 through 0.220. The lag-one cross correlation coefficients are still smaller.

TABLE 4-3

CORRELATION MATRIX BETWEEN $\xi_{p,\tau}$ SERIES OF WINTER FLOW RETARDATIONS

Connecting Channel	St. Mary's River	St. Clair-Detroit Rivers	Niagara River	St. Lawrence River
St. Mary's River	1.000	0.072	0.220	0.220
St. Clair-Detroit Rivers	0.072	1.000	0.133	0.160
Niagara River	0.220	0.133	1.000	0.179
St. Lawrence River	0.220	0.160	0.179	1.000

Because the absolute values of skewness coefficients of $\xi_{p,\tau}$ are relatively small, especially for the first three connecting channels of Table 4-3, it is not necessary to transform the three-parameter lognormal functions of $\xi_{p,\tau}$ into normal, in order to use the simple multinormal distribution approach in generating new samples.

The same procedure and the same equations are used in the generation of new samples of winter flow retardations as they were used for generating the new samples of mean monthly and mean quarter-monthly net basin supply series. The differences, however, are as follows:

(1) Whenever a monthly value of winter flow retardation comes out to be negative, it is replaced by zero value, because by definition the flow retardation cannot be negative, in contrast to the NBS values which can be negative. The zero values of winter flow retardations in months for which the mean winter flow retardation is greater than zero are also a common occurrence in the historic series.

(2) Because of intermittency of winter flow retardations, the new generated samples do not carry the dependence between the values of one year to values of the next year.

11.3 Analysis of Generated Samples. The basic parameters of the historic series of winter flow retardations for the four connecting channels of the Great Lakes are given in Table 4-4. Table 4-5 gives the means and the standard deviations of the generated new samples of the independent stochastic components $\xi_{p,\tau}$ for the four

TABLE 4-4

BASIC PARAMETERS OF HISTORIC SAMPLES OF WINTER FLOW RETARDATIONS OF FOUR CONNECTING CHANNELS OF GREAT LAKES

Parameters	St. Mary's River	St. Clair-Detroit Rivers	Niagara River	St. Lawrence River
Mean of $\xi_{p,\tau}$ Series	0.000	0.000	0.000	0.000
Standard Deviation of $\xi_{p,\tau}$ Series	0.967	0.916	0.966	0.979
Periodicity in the Monthly Means of Winter Flow Retardation Series	1.174	2.126	0.493	11.403
	21.755	3.496	6.060	32.030
	28.645	2.645	9.030	25.419
	10.524	1.230	4.791	8.134
	1.606	0.514	--	4.940
Periodicity in the Monthly Standard Deviation of Winter Flow Retardation Series	3.499	3.545	1.386	9.081
	15.684	4.212	5.404	12.262
	13.647	3.468	5.737	7.913
	10.136	2.194	3.578	5.030
	7.310	1.393	--	4.546

TABLE 4-5

BASIC PARAMETERS OF $\xi_{p,\tau}$ COMPONENTS OF GENERATED NEW SAMPLES OF WINTER FLOW RETARDATIONS IN CONNECTING CHANNELS OF GREAT LAKES

Sample	St. Mary's River		St. Clair-Detroit Rivers		Niagara River		St. Lawrence River	
	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
1	-0.018	0.967	-0.008	1.015	0.058	0.947	-0.069	0.998
2	-0.043	0.977	-0.055	1.047	-0.082	0.973	-0.090	0.982
3	0.141	0.966	0.093	0.924	0.050	1.014	0.053	0.948
4	-0.023	1.007	-0.011	0.950	-0.077	1.024	-0.016	1.034
5	0.159	0.912	0.143	1.069	-0.113	0.942	0.053	0.981
6	-0.056	1.002	0.011	0.976	-0.001	0.952	-0.000	0.970
7	0.083	0.978	-0.044	1.021	0.057	0.956	-0.009	0.997
8	0.064	0.965	0.054	0.917	-0.076	1.022	0.008	0.964
9	-0.047	0.972	0.052	1.124	0.053	1.036	-0.009	1.025
10	0.028	1.007	0.044	0.972	0.026	0.936	-0.093	0.984
11	-0.064	0.953	-0.053	0.984	-0.133	1.004	-0.149	0.983
12	-0.052	1.020	-0.054	1.060	0.064	0.920	0.049	1.014
13	-0.124	1.038	0.156	1.054	0.124	1.060	0.099	0.933
14	-0.053	1.058	-0.029	1.002	-0.015	1.056	-0.024	0.984
15	-0.016	1.059	0.058	0.992	-0.029	0.990	0.015	0.951
16	-0.064	1.034	-0.020	1.061	-0.227	0.955	-0.065	1.025
17	0.051	0.972	-0.118	0.956	0.093	0.994	0.028	0.961
18	0.136	1.066	-0.060	0.998	0.170	1.027	0.069	1.020
19	-0.008	1.001	-0.054	1.065	-0.121	0.928	0.041	0.916
20	0.027	1.033	0.043	0.989	0.008	1.015	0.181	0.988
Average	0.006	1.003	0.007	1.009	0.000	0.988	0.003	0.983

winter flow retardation series. Their average values for 20 samples, each 600 monthly values long (50 years), also given, should be close to zero and unity, respectively. The differences are small. The generated $\xi_{p,\tau}$ series with mean zero and standard deviation unity are used here to demonstrate the good reproduction of the $\xi_{p,\tau}$ series. The standard deviations of historic series are smaller than unities, as Table 4-4 demonstrates. The multiplication of generated values of $\xi_{p,\tau}$ series by these values makes the generated $\xi_{p,\tau}$ series compatible with the historic series. Performing the transformation of $\xi_{p,\tau}$ by the autoregressive model and adding the periodic components does not change the goodness of reproduction of the $\xi_{p,\tau}$ series.

Table 4-6 gives the general mean and the general standard deviation for each generated sample of the winter flow retardation series for the four connecting channels of the Great Lakes, as well as their average values. The next row gives the mean and the standard deviation of the corresponding historic monthly series of winter flow retardations. In the last row are the differences between the averages of parameters of generated samples and of the historic sample. Because of replacing the negative values by zeros in the generated samples, the averages of means of generated samples for all four series are somewhat greater than the means of historic samples. The average standard deviations of generated samples, for the same reason, are somewhat smaller than the standard deviation of historic samples. Regardless of these differences, it may be assuming that the reproduction of parameters of historic samples is sufficiently good, especially in light of several factors, but particularly because of errors in data of historic samples, and because of trends in the winter flow retardation series.

TABLE 4-6
COMPARISON OF MEANS AND STANDARD DEVIATION OF GENERATED SAMPLES AND THE HISTORIC SAMPLE OF WINTER FLOW RETARDATION SERIES

Sample No.	St. Mary's River		St. Clair-Detroit Rivers		Niagara River		St. Lawrence River	
	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.	Mean	St.Dev.
1	13.469	14.222	2.317	2.512	4.301	5.193	16.477	12.976
2	13.339	15.315	2.550	2.812	3.974	4.841	16.499	13.418
3	13.459	13.980	2.537	2.983	4.173	4.940	16.659	13.298
4	13.816	14.872	2.549	2.753	4.276	5.326	16.529	12.869
5	13.998	15.304	2.580	2.745	4.217	4.944	16.491	13.083
6	13.343	13.823	2.454	2.729	4.131	4.911	16.613	13.172
7	13.198	13.694	2.502	2.822	4.150	4.900	16.348	12.701
8	14.222	15.170	2.481	2.983	4.200	4.954	16.256	12.685
9	14.087	15.946	2.466	2.830	4.269	5.200	16.607	13.654
10	13.881	14.563	2.558	2.779	4.278	5.102	16.642	13.974
11	13.379	13.880	2.496	3.055	4.126	5.063	16.710	13.592
12	13.190	13.839	2.482	2.821	4.083	4.848	16.305	13.152
13	13.922	14.362	2.376	2.833	4.008	4.696	16.369	12.949
14	12.875	13.228	2.433	2.747	4.208	4.964	16.439	12.948
15	13.383	13.447	2.463	2.844	4.101	5.051	16.267	12.266
16	13.601	13.904	2.416	2.785	4.206	4.910	16.520	12.885
17	13.362	13.137	2.383	2.719	4.219	4.944	16.376	13.031
18	13.704	14.788	2.507	2.805	4.303	4.957	16.380	13.348
19	12.872	12.756	2.442	2.710	4.229	5.045	16.365	13.188
20	13.946	15.222	2.434	2.671	4.245	5.082	16.574	13.527
Average	13.553	14.272	2.471	2.797	4.185	4.994	16.471	13.161
Historic Sample	12.741	15.401	2.001	3.281	3.780	4.985	16.331	13.280
Difference	0.812	- 1.120	0.470	-0.484	0.405	0.009	0.140	- 0.119

REFERENCES

1. Pentland, R. L., Runoff Characteristics in the Great Lakes Basin, 11th Conference on Great Lakes Research and 2nd meeting of the International Association for Great Lakes Research, Milwaukee, Wisconsin, April 18-20, 1968.

2. Coordinated Basic Data, Report to the International Great Lakes Levels Working Committee by the Regulation Subcommittee, May 1969.

3. Simulation of Great Lakes Basin Water Supplies and Winter Flow Retardation in the Connecting Channels, Interim Report to International Great Lakes Levels Working Committee by the Regulation Subcommittee, November 1969.

4. Great Lakes Basin Commission Maps, Lake Survey District, Corps of Engineers, Map and Chart Plant, Detroit, Michigan, 30 June 1970.

5. Witherspoon, D. F., General Hydrology of the Great Lakes and Reliability of Component Phases, Technical Bulletin No. 50, Inland Waters Branch, Department of the Environment, Ottawa, Canada, 1971.

6. Regulation of Lakes Superior and Ontario, Interim Report to the International Joint Commission by the International Great Lakes Levels Board, Second Draft, December 1971.

7. Morton, F. I., and H. B. Rosenberg, Hydrology of Lake Ontario, Paper No. 3188, Vol. 126, Part 1, pp. 705-822, Trans. ASCE, 1961.

8. Young, G. K., and W. C. Pisano, Operational Hydrology Using Residuals, ASCE Hydraulics Journal, Vol. 94, No. HY4, pp. 903-923, 1968.

KEY WORDS: Time series, Generation of time series, Application of Monte Carlo methods, Great Lakes hydrology, Simulation of lake hydrologic variables.

ABSTRACT: Generation of 20 samples, each 50 years long, of three variables related to water inputs into and retardation of flows in connecting channels of the Great Lakes is the subject matter of this paper. They are: (1) Mean monthly net basin water supplies of five lakes; (2) Mean quarter-monthly net basin water supplies of two smaller lakes (Ontario and Erie); and (3) Flow retardations in connecting channels due to freezing and weed effects. For each of series of the above three variables, first tests of trends in data have been performed. The St. Clair Lake mean monthly net basin supplies had a trend. Also the St. Mary's River and the St. Clair-Detroit Rivers connecting channels had decreasing trends in flow retardation. Series have been studied with trends in parameters removed.

Periodic parameters in all three variables were the mean and the standard deviation, while autocorrelation coefficients and skewness coefficient were found not to be periodic. The stochastic components are found greatly autocorrelated, according to the second-order autoregressive

KEY WORDS: Time series, Generation of time series, Application of Monte Carlo methods, Great Lakes hydrology, Simulation of lake hydrologic variables.

ABSTRACT: Generation of 20 samples, each 50 years long, of three variables related to water inputs into and retardation of flows in connecting channels of the Great Lakes is the subject matter of this paper. They are: (1) Mean monthly net basin water supplies of five lakes; (2) Mean quarter-monthly net basin water supplies of two smaller lakes (Ontario and Erie); and (3) Flow retardations in connecting channels due to freezing and weed effects. For each of series of the above three variables, first tests of trends in data have been performed. The St. Clair Lake mean monthly net basin supplies had a trend. Also the St. Mary's River and the St. Clair-Detroit Rivers connecting channels had decreasing trends in flow retardation. Series have been studied with trends in parameters removed.

Periodic parameters in all three variables were the mean and the standard deviation, while autocorrelation coefficients and skewness coefficient were found not to be periodic. The stochastic components are found greatly autocorrelated, according to the second-order autoregressive

KEY WORDS: Time series, Generation of time series, Application of Monte Carlo methods, Great Lakes hydrology, Simulation of lake hydrologic variables.

ABSTRACT: Generation of 20 samples, each 50 years long, of three variables related to water inputs into and retardation of flows in connecting channels of the Great Lakes is the subject matter of this paper. They are: (1) Mean monthly net basin water supplies of five lakes; (2) Mean quarter-monthly net basin water supplies of two smaller lakes (Ontario and Erie); and (3) Flow retardations in connecting channels due to freezing and weed effects. For each of series of the above three variables, first tests of trends in data have been performed. The St. Clair Lake mean monthly net basin supplies had a trend. Also the St. Mary's River and the St. Clair-Detroit Rivers connecting channels had decreasing trends in flow retardation. Series have been studied with trends in parameters removed.

Periodic parameters in all three variables were the mean and the standard deviation, while autocorrelation coefficients and skewness coefficient were found not to be periodic. The stochastic components are found greatly autocorrelated, according to the second-order autoregressive

KEY WORDS: Time series, Generation of time series, Application of Monte Carlo methods, Great Lakes hydrology, Simulation of lake hydrologic variables.

ABSTRACT: Generation of 20 samples, each 50 years long, of three variables related to water inputs into and retardation of flows in connecting channels of the Great Lakes is the subject matter of this paper. They are: (1) Mean monthly net basin water supplies of five lakes; (2) Mean quarter-monthly net basin water supplies of two smaller lakes (Ontario and Erie); and (3) Flow retardations in connecting channels due to freezing and weed effects. For each of series of the above three variables, first tests of trends in data have been performed. The St. Clair Lake mean monthly net basin supplies had a trend. Also the St. Mary's River and the St. Clair-Detroit Rivers connecting channels had decreasing trends in flow retardation. Series have been studied with trends in parameters removed.

Periodic parameters in all three variables were the mean and the standard deviation, while autocorrelation coefficients and skewness coefficient were found not to be periodic. The stochastic components are found greatly autocorrelated, according to the second-order autoregressive

models. For the resulting independent, identically distributed stochastic component of all series, the three-parameter lognormal distributions was a good approximation.

The principal component analysis has been used in generating the new samples of the mean monthly net basin supplies. The approach of generating first the monthly values, and then superimposing the generated four differences of mean quarter-monthly values was difficult to apply, because both the sum of four differences to be zero and the autoregressive model could not be satisfied simultaneously.

HYDROLOGY PAPER #72 - GENERATION OF HYDROLOGIC SAMPLES
CASE STUDY OF THE GREAT LAKES, by Vujica Yevjevich

models. For the resulting independent, identically distributed stochastic component of all series, the three-parameter lognormal distributions was a good approximation.

The principal component analysis has been used in generating the new samples of the mean monthly net basin supplies. The approach of generating first the monthly values, and then superimposing the generated four differences of mean quarter-monthly values was difficult to apply, because both the sum of four differences to be zero and the autoregressive model could not be satisfied simultaneously.

HYDROLOGY PAPER #72 - GENERATION OF HYDROLOGIC SAMPLES
CASE STUDY OF THE GREAT LAKES, by Vujica Yevjevich

models. For the resulting independent, identically distributed stochastic component of all series, the three-parameter lognormal distributions was a good approximation.

The principal component analysis has been used in generating the new samples of the mean monthly net basin supplies. The approach of generating first the monthly values, and then superimposing the generated four differences of mean quarter-monthly values was difficult to apply, because both the sum of four differences to be zero and the autoregressive model could not be satisfied simultaneously.

HYDROLOGY PAPER #72 - GENERATION OF HYDROLOGIC SAMPLES
CASE STUDY OF THE GREAT LAKES, by Vujica Yevjevich

models. For the resulting independent, identically distributed stochastic component of all series, the three-parameter lognormal distributions was a good approximation.

The principal component analysis has been used in generating the new samples of the mean monthly net basin supplies. The approach of generating first the monthly values, and then superimposing the generated four differences of mean quarter-monthly values was difficult to apply, because both the sum of four differences to be zero and the autoregressive model could not be satisfied simultaneously.

HYDROLOGY PAPER #72 - GENERATION OF HYDROLOGIC SAMPLES
CASE STUDY OF THE GREAT LAKES, by Vujica Yevjevich

LIST OF PREVIOUS 25 PAPERS

No. 47 Mathematical Simulation of Infiltrating Watersheds, by Roger E. Smith and David A. Woolhiser, January 1971.

No. 48 Models for Subsurface Drainage, by W. E. Hedstrom, A. T. Corey and H. R. Duke, February 1971.

No. 49 Infiltration Affected by Flow of Air, by David B. McWhorter, May 1971.

No. 50 Probabilities of Observed Droughts, by Jaime Millan and Vujica Yevjevich, June 1971.

No. 51 Amplification Criterion of Gradually Varied, Single Peaked Waves, by John Peter Jolly and Vujica Yevjevich, December 1971.

No. 52 Stochastic Structure of Water Use Time Series, by Jose D. Salas-La Cruz and Vujica Yevjevich, June 1972.

No. 53 Agricultural Response to Hydrologic Drought, by V. J. Bidwell, July 1972.

No. 54 Loss of Information by Discretizing Hydrologic Series, by Mogens Dyhr-Nielsen, October 1972.

No. 55 Drought Impact on Regional Economy, by Jaime Millan, October 1972.

No. 56 Structural Analysis of Hydrologic Time Series, by Vujica Yevjevich, November 1972.

No. 57 Range Analysis for Storage Problems of Periodic-Stochastic Processes, by Jose Salas-La Cruz, November 1972.

No. 58 Applicability of Canonical Correlation in Hydrology, by Padoong Torranin, December 1972.

No. 59 Transposition of Storms, by Vijay Kumar Gupta, December 1972.

No. 60 Response of Karst Aquifers to Recharge, by Walter G. Knisel, December 1972.

No. 61 Drainage Design Based Upon Aeration by Harold R. Duke, June 1973.

No. 62 Techniques for Modeling Reservoir Salinity by John Hendrick, August 1973.

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 the 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.