FLUCTUATIONS OF WET AND DRY YEARS

PART II

ANALYSIS BY SERIAL CORRELATION

By

Vujica M. Yevdjevich

June 1964

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

4

Several departments at Colorado State University have substantial research and graduate programs oriented to hydrology. These Hydrology Papers are intended to communicate in a fast way the current results of this research to the specialists interested in these activities. The papers will supply most of the background research data and results. Shorter versions will usually be published in the appropriate scientific and professional journals, or presented at national or international scientific and professional meetings and published in the proceedings of these meetings.

This research is sponsored by the U. S. National Science Foundation. Part of the research material is from the project "Water Cycle" sponsored previously by the U. S. Office of Naval Research. The research on this topic was initiated and carried out parly while the author was associated with the U. S. National Bureau of Standards and with the U. S. Geological Survey. The research was conducted at Colorado State University Civil Engineering Section, Fort Collins, Colorado.

EDITORAL BOARD

Dr. Arthur T. Corey, Professor, Agricultural Engineering Department

Dr. Robert E. Dils, Professor, College of Forestry and Range Management

Dr. Vujica M. Yevdjevich, Professor, Civil Engineering Department

Colorado State University

FLUCTUATIONS OF WET AND DRY YEARS

PART II

ANALYSIS BY SERIAL CORRELATION

by

Vujica M. Yevdjevich

HYDROLOGY PAPERS

COLORADO STATE UNIVERSITY

FORT COLLINS, COLORADO

June, 1964

No. 4

PREFACE

This second part of the paper "Fluctuations of Wet and Dry Years" refers only to the analysis of data by serial correlation in the study of patterns in sequence of annual river flow, annual effective precipitation, and annual precipitation. The other parts will contain: the analysis of patterns in sequence by ranges, by runs, and by variance spectrum (power spectrum); the effect of inconsistency and nonhomogeneity in data on these patterns; the effect of selected beginning of year; and various relationships among statistical characteristics.

TABLE OF CONTENTS

	Page
Preface	iii
Abstract	ix
A. Introduction	1
1. Summary of Part I	1
2. Definitions	1
3. Conditions of stationarity.	1
4. Expressions for serial correlation coefficients	2
5. Objectives of this part of study	3
6. Methods used in the analysis	3
B. Serial Correlation of Variables, Independent or Dependent in Sequence	4
1. Distribution, expected value, and variance of serial correlation coefficients of normal independent variables	4
2. Significance test for serial correlation coefficients of normal independent variables	5
3. Non-normal variables	6
4. Distribution, expected value, and variance of serial correlation coefficients of dependent variables	7
5. Effective number of stations with interstation correlation	9
C. Analysis of First Large Sample of River Flow Records	12
1. Method of data analysis	12
2. Frequency distribution of first serial correlation coefficient	12
3. Frequency distributions of other serial correlation coefficients.	16
4. Average values of serial correlation coefficients	18
5. Correlograms of individual rivers	19
D. Analysis of the Second Large Sample of River Flow and the Large Sample of Precipitation	25
1. Simultaneous analysis of flow and precipitation	25
2. Frequency distribution of the first serial correlation coefficients	25
3. Frequency distribution of other serial correlation coefficients	27
4. Average values of the serial correlation coefficients	30
5. Individual correlograms	33
6. Effect of length of time series	39
7. Regional distribution of first serial correlation coefficient	39
8. The case of the Missouri River	39
E. Effect of Inconsistency and Nonhomogeneity of Data	42
1. Errors and nonhomogeneity	42
2. Comparison of homogeneous or consistent data with samples of nonhomogeneous or inconsistent data	42
3. Hypothesis of quasi-stationarity	42
F. Effects of Climatic Conditions on Serial Correlation	45
1. Climatic conditions	45
2. Comparison of distributions of first serial coefficient for various groups of water yield	45
G. Conclusions	49
References	50

LIST OF FIGURES AND TABLES

Figures

Page

1. Distribution of the time lengths for the series of the first large sample of river gaging stations 12

2. Cumulative distributions and frequency histograms of the first serial correlation coefficient for the series of the first large sample of river gaging stations 13

3. Cumulative distributions of the first serial correlation coefficient for the series of the first large sample of river gaging stations 14

4. Cumulative distributions of Fisher's z-transform of the first serial correlation coefficient for the series of the first large sample of river gaging stations 14

5. Cumulative distributions of the serial correlation coefficients of r_2, r_3, r_4 and r_5 for the first large sample of river gaging stations 16

6. Average values of the serial correlation coefficients ($\bar{r}_1$ through $\bar{r}_{11}$) for the series of the first large sample of river gaging stations versus lag k 18

7. Correlograms of the four rivers with longest records: the Göta River, the Nemunas River, the Rhine River, and the Danube River 19

8. Correlograms of 28 individual series from the first large sample of river gaging stations 20

9. Correlograms of the St. Lawrence River at Ogdensburg, New York 23

10. Correlograms and their differences for the St. Lawrence River at Ogdensburg, New York 24

11. Cumulative distributions of the first serial correlation coefficient for the series of the second large sample of river gaging stations and for the series of the large sample of precipitation gaging stations, both in Western North America 25

12. Cumulative distributions of the first serial correlation coefficient for the series of the second large sample of river gaging stations and for the series of the large sample of precipitation stations, both in Western North America 27

13. Cumulative distributions of serial correlation coefficients r_2 through r_{11} for the series of the second large sample of river gaging stations and the large sample of precipitation gaging stations from Western North America 28

14. Cumulative distributions of serial correlation coefficients r_2 through r_{11} for the series of the second large sample of river gaging sations and the large sample of precipitation gaging stations from Western North America 29

15. Average values of $\bar{r}_k$, standard deviation s_r, skewness coefficients C_{sr}, and kurtosis k_r of the serial correlation coefficients r_1 through r_{25} for the second large sample of river gaging stations and the large sample of precipitation gaging stations from Western North America 31

16. Average value of $\bar{r}_k$, standard deviation s_r, skewness coefficient C_{sr}, and kurtosis k_r of the serial correlation coefficients r_1 through r_{15} for the second large sample of river gaging stations and the large sample of precipitation gaging stations from Western North America for the simultaneous period of observation (1931-1961) 32

17. Correlograms of 62 individual series in groups of 5 or 6 from V-, P_e- and P_i-series . . 34

18. Correlograms of 62 individual series in groups of 5 or 6 from V-, P_e- and P_i-series as a continuation of Fig. 17 36

19. Regional distribution of the first serial correlation coefficient for the large sample of precipitation gaging stations from Western North America 38

20. The Missouri River Basin with main tributaries and main storage reservoirs as developed by 1963 40

21. Time series of the annual flow of the Missouri River near Sioux City, Iowa 40

22. Comparison of the distributions of the first serial correlation coefficient of homogeneous and nonhomogeneous time series of annual precipitation in Western North America 43

23. Comparison of the distributions of the first serial correlation coefficient of homogeneous and nonhomogeneous time series of annual precipitation in Western North America for the simultaneous period 1931-1960 with N = 30 43

24. Distributions of the first serial correlation coefficient for V-series in Western North America with N_m = 37 for the three ranges of specific water yield 45

25. Distributions of the first serial correlation coefficient for V-series in Western North America for the simultaneous observations of 1931-1960 with N = 30 for the three ranges of specific water yield 45

26. Distributions of the first serial correlation coefficient for P_e-series in Western North America with N_m = 37 for the three ranges of specific water yield 46

27. Distributions of the first serial correlation coefficient for P_e-series in Western North America for the simultaneous observations of 1931-1960 with N = 30 for the three ranges of water yield 46

28. Distributions of the first serial correlation coefficient for P_i^1-series in Western North America from all available data with N_m = 54 for the three ranges of average annual precipitation 46

29. Distributions of the first serial correlation coefficient for P_i^1-series in Western North America for the simultaneous observations of 1931-1960 with N = 30 for the three ranges of average annual precipitation 46

30. Distributions of the first serial correlation coefficient for P_i^2-series in Western North America from all available data with N_m = 57 for the three ranges of average annual precipitation 47

31. Distributions of the first serial correlation coefficient for P_i^2-series in Western North America from the simultaneous observations of 1931-1960 with N = 30 for the three ranges of average annual precipitation 47

Tables

1. Statistics of the frequency distribution of the first serial correlation coefficient for 140 stations of annual flow (V-series) and of annual effective precipitation (P_e-series) as well as for the normal independent variable 13

2. Statistics of the frequency distributions of the serial correlation coefficients, r_2, r_3, r_4, and r_5 for 140 stations of annual flow (V-series) and annual effective precipitation (P_e-series) as well as simple estimates of mean for r_6, r_7, r_8, r_9, r_{10}, and r_{11} 17

3. Serial correlation coefficient r_k for the annual flows of the Göta River 22

4. Statistical parameters or r_1-distributions of V-, P_e-, P_i^1-, and P_i^2-series, for the three ranges of specific water yield or average annual precipitation for each series, and for the longest or simultaneous period of observation, as given in Figs. 24 through 31 47

ABSTRACT

Serial correlation analysis is applied to investigate the patterns in the sequence of series of annual river flow, annual effective precipitation and annual precipitation at the ground. Four large samples of series are used as research material: (1) first large sample of river gaging stations (140) from many parts of the world; (2) second large sample of river gaging stations (446) from Western North America; (3) large sample of precipitation gaging stations (1141) with consistent and/or homogeneous data from Western North America; and (4) large sample of precipitation gaging stations (473) with inconsistent and/or nonhomogeneous data from Western North America.

Statistical techniques and final expressions for the serial correlation analysis are given in summary form in Chapter B. Analysis of the first sample of river gaging stations is the subject of Chapter C. Analysis of the second large sample of river gaging stations and the large sample of precipitation gaging stations with homogeneous data is the subject of Chapter D . Effects of climatic conditions on serial correlation are discussed in Chapter F. A brief study of the effect of inconsistency and/or nonhomogeneity in precipitation data is given in Chapter E.

The carryover of water from year to year in river basins, which is disposed of in successive years either by river runoff or by evaporation and evapotranspiration, is the main factor of time dependence in series of annual river flow, annual effective precipitation, and annual precipitation at the ground. A factor worthwhile of further study is the nonhomogeneity and/or inconsistency in hydrologic data of river flow and precipitation.

There is no statistical evidence that cycles exist in river flow or precipitation time series beyond the astronomic cycle of the year. Moving average schemes in general and the first and second order autoregressive schemes (Markov linear mathematical models) in particular fit sufficiently well the patterns in the sequence of annual river flows of river basins with large water carryover.

FLUCTUATIONS OF WET AND DRY YEARS

PART II

ANALYSIS BY SERIAL CORRELATION

By: Vujica M. Yevdjevich*

A. INTRODUCTION

1. Summary of Part I. Part I deals with the assembly of research data and with the mathematical models of this study of fluctuations of wet and dry years.

Two large samples of river gaging stations, one on the global and the other on the continental sample scale have been described as the research material for the analysis of patterns in sequence of annual river flow and annual effective precipitation. Two large samples of precipitation gaging stations, one with the homogeneous and the other with the nonhomogeneous data, both on a continental sampling scale, have been described as research material for the analysis of patterns in sequence of annual precipitation. More information in a summary form about these samples is given in the analysis of serial correlation of each sample. The reader is referred to Part I, Colorado State University Hydrology Paper No. 1 [1]** for all detailed information about the samples which are used in Part II.

Mathematical models as derived in Part I are the basis for this investigation of patterns in sequence of annual flow, annual effective precipitation and annual precipitation by using a serial correlation analysis.

2. Definitions. The serial correlation coefficient r_k of the order k, or of the lag k, is defined as the product-moment correlation coefficient between the members of a discrete time series that are k items apart from one another. The serial correlation coefficient sometimes is abbreviated here as s. c. c. for the sake of shorter text.

For the series of annual flow, annual effective precipitation and annual precipitation, the coefficients r_k are computed for the stations of this study among all members x_i and x_{i+k}, where $i = 1, 2, 3, \ldots (N-k)$, with N the size of time series, k the lag of s. c. c., and x_i (and x_{i+k}) the standardized values of the time series X. The number of correlated pairs (x_i, x_{i+k}) is $N-k$. The greater k for a given N, the smaller is the number $N-k$ of correlated pairs. These s. c. c. refer, therefore, to an open time series as distinguished from a circular time series (the last member of time series x_N is followed by the first member x_1, and again $x_2, x_3, \ldots$).

The discrete graph of s. c. c. r_k against the lag k is the correlogram.

The term "autocorrelation" refers here to the correlogram and correlation coefficients of a continuous time series. Assuming a continuous hydrograph of river flow of finite length, its correlogram is an autocorrelation correlogram with the equation $r_t = f(t)$, where r_t is the autocorrelation coefficient (a. c. c.), and t the time lag between the correlated values x_s and x_{s+t}. Theoretically, there is always in this case of continuous time series an infinite number of ordinates, regardless of the finite size T of a time series.

The symbols r_k and r_t are used here for finite time series of lengths N and T, respectively for a discrete and a continuous time series. These values are considered as the best estimates of population values. When a reference is made to the time series of infinite length (referred to the total population of all possible observations), the serial correlation coefficients and autocorrelation coefficients are designated as ρ_k and ρ_t respectively for a discrete and a continuous time series.

The autocorrelation function is defined here as the mathematical expression which describes analytically the continuum of autocorrelation coefficients r_t or ρ_t. The correlogram of a discrete time series is a discrete series. The expression "autocorrelation function" will also be used, if a continuous mathematical function is fitted to the discrete correlogram of s. c. c.

3. Conditions of stationarity. It is assumed here that the time series are stationary, except when the contrary is stated. All hydrologic time series have a small degree of nonstationarity. It comes either from man-made changes in river basins and around precipitation gaging stations or from the inconsistency in data. Inconsistency is conceived as systematic errors in data in form of trends or jumps, or the systematic difference between the values in nature and data on the desk.

By the selection of stations in this study, an attempt was made to eliminate the records with known or supposed high nonstationarity or to incorporate the data which evidently is not stationary in a special sample of nonhomogeneous data (second large sample of precipitation data in this study). However, the small degree of nonstationarity which is left inside the selected samples of "stationary series,"

*Professor-in-charge of Hydrology Program, Civil Engineering Department, Colorado State University.

**References are designated by these brackets and given at the end of the paper.

and called here "pseudostationarity," will not seriously impair basic conclusions of the study about time dependence in series of annual flow, annual effective precipitation, and annual precipitation. All natural phenomena, similar to hydrologic phenomena, observed or measured with similar procedures and methods over a long time and with man-made influence on these phenomena, are subject to this pseudostationarity. It is considered here that the samples of time series as selected for this study are stationary in the practical sense, or that the nonstationarity or pseudostationarity when present will not mask the effects of primarily physical factors of time dependence.

Conditions of stationarity which are considered here as approximately satisfied are:

(1) The expected value of any X_i in a time series is equal to the population mean which is a constant, or

$$E(X_i) = \mu = \int_{-\infty}^{\infty} X dP(X) \qquad 1.1$$

with $P(X)$ the probability distribution of X, and μ the constant population mean.

(2) The expected value of the second order covariance for any position in the time series of X_i and X_{i+k} depends only upon k and not on the position i; or that

$$E[(X_i - \mu)(X_{i+k} - \mu)] = \sigma^2 \rho_k \qquad 1.2$$

with σ^2 the population variance, and ρ_k the population s. c. c. of the lag k. These two conditions, if satisfied, imply that the time series have second order stationarity. The next condition is that of ergodicity (which should be satisfied also); i. e., the time averages converge in probability to theoretical averages.

4. Expressions for serial correlation coefficients. In practice the serial correlation coefficients are computed either for an open time series or for a circular time series. Regardless of whether an open or a circular time series approach is used, the expression for ρ_k is

$$\rho_k = \frac{\text{cov}(x_i, x_{i+k})}{\text{var } x_i} \qquad 1.3$$

with cov (x_i, x_{i+k}) the population covariance, and var $x_i = \sigma^2$ the population variance. The value of ρ_k is usually estimated from the available sample. Two approaches may be used in this case:

(a) Values of the mean and the variance of population are assumed as known, so that the estimate of ρ_k from the sample is

$$r_k = \frac{\sum_{i=1}^{N-k} (x_i - \mu)(x_{i+k} - \mu)}{(N-k)\sigma^2} \qquad 1.4$$

for an open time series, or with N-k replaced by N in eq. 1.4 for a circular time series; and

(b) Values of the mean and the variance of population are estimated from the sample, or

$$r_k = \frac{\sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{(N-k)\, s^2} \qquad 1.5$$

with $\bar{x}$ and s^2 the estimates of μ and σ^2 from the available sample, respectively, for an open time series, and with N-k replaced by N in eq. 1.5 for a circular time series.

For an independent time series and for an underlying normally distributed population of x, the most efficient estimate of μ is

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \,, \qquad 1.6$$

and the most efficient estimate of σ^2 is

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \qquad 1.7$$

which is an unbiased estimate of σ^2. Under the assumption of stationarity of second order, however, this estimate of σ^2 is no more unbiased.

An asymptotically unbiased and consistent estimate of cov (x_i, x_{i+k}) is

$$C_k = \frac{1}{N-k} \sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x}) \qquad 1.8$$

for an open time series or with N-k replaced by N for a circular time series. In this case $E(C_k) = \sigma^2 \rho_k + \theta(1/N)$.

A consistent estimate of ρ_k is, therefore, $r_k = C_k / s^2$, which is obtained by using eq. 1.5 with $\bar{x}$ given by eq. 1.6 and s^2 by eq. 1.7. This estimate of ρ_k is not unbiased. Its expected value is

$$E(r_k) = \rho_k + \theta\left(\frac{1}{N}\right) \qquad 1.9$$

where $\theta(\frac{1}{N})$ may be neglected for N, in this study, as large as 30 or more.

In practice sometimes, on the analogy of the ordinary correlation coefficient, the estimate of ρ_k is calculated by the following equation

$$r_k^* = \frac{\sum_{i=1}^{N-k} (x_i - \bar{x}_i)(x_{i+k} - \bar{x}_{i+k})}{(N-k)\, s_i s_{i+k}} \qquad 1.10$$

for an open time series. In eq. 1.10 the values $\bar{x}_i$, $\bar{x}_{i+k}$, s_i, and s_{i+k} are given as

$$\bar{x}_i = \frac{1}{N-k} \sum_{i=1}^{N-k} x_i \,; \quad \bar{x}_{i+k} = \frac{1}{N-k} \sum_{i=1}^{N-k} x_{i+k} \,; \ldots \qquad 1.11$$

$$s_i^2 = \frac{1}{N-k} \sum_{i=1}^{N-k} (x_i - \bar{x}_i)^2 \,; \text{ and}$$

$$s_{i+k}^2 = \frac{1}{N-k} \sum_{i=1}^{N-k} (x_{i+k} - \bar{x}_{i+k})^2 \qquad 1.12$$

The above four values differ slightly from $\bar{x}$ and s^2 computed by eqs. 1.6 and 1.7.

For small values of k the estimates of ρ_k by using r_k^* of eq. 1.10 give negligible differences as compared with r_k computed by eq. 1.5, or even eq. 1.9. As it will be shown later, the first one or two ρ_k values are the most important in the analysis of patterns in sequence of annual flow and annual precipitation.

The computation of serial correlation coefficients in samples in this study was done by using eqs. 1.10 through 1.12. A greater amount of bias is, however, introduced when eq. 1.10 is used instead of eq. 1.5 for the computation of r_k with large k . It is assumed here that a replacement of N by N-k in equations for estimating the mean and the variance of distributions of r_k of normal independent variables will account for this bias in the case k is large.

As it is difficult to calculate the distribution of r_k^* of eq. 1.10, its distribution will be replaced or substituted by the distribution of r_k computed by eq. 1.5, either for an open or for a circular time series, and either for μ and σ^2 assumed as known, or for μ and σ^2 estimated from the sample by eqs. 1.6 and 1.7. The values of k are 0, 1, 2, . . ., m, where m is the highest value of k selected to be computed, but with m usually much smaller than N. The negative value of k gives $r_{-k} = r_k$. The value of r_o is unity, because in this case each x_i is correlated with itself. The correlogram starts, therefore, from the coordinate [k= 0, r_o = 1] and decreases both to the positive and negative values of k . The first serial correlation coefficient (f. s. c. c.) r_1 is often the most useful in the current analysis of hydrologic time series. It is computed by eq. 1.10 with k = 1 .

As the sequence of annual flows cannot be taken as repeating itself with the same patterns, the closed (circular) series approach has not been used in this study for the computation of r_k . However, if the lag k is very small, and N is high (i. e., k = 1 , N_{min} = 30), the equations developed for the distribution of statistical parameters of circular time series may be used as approximations for distributions of statistical parameters of open time series.

5. Objectives of this part of study. This analysis of annual flow, annual effective precipitation, and annual precipitation by using the serial correlation technique has the following objectives:

(a) to show the degree of time dependence of each type of time series analyzed;

(b) to obtain an insight into how this dependence in time series differs from one series to other, i. e., to compare changes in dependence between measured river flow and effective precipitation in the river basin and also between effective precipitation and measured precipitation;

(c) to infer from the results the most appropriate mathematical models for this dependence as outlined in Part I;

(d) to show the regional distribution of first serial correlation coefficient, and to get an insight into how r_1 changes from one region to another, and what may be the variation of future values of r_1 for those regions; and

(e) to investigate the effect of general climate (i. e., humid or arid regions) on time dependence.

6. Methods used in the analysis. Properties of normal independent variables and of normal dependent variables of a known stochastic process are discussed first.

Distributions of serial correlation coefficients, and the statistical parmeters which describe these distributions from observed time series of annual flow, annual effective precipitation and annual precipitation at the ground are compared with distributions of normal independent variables and their parameters, respectively. Differences are analyzed and an attempt is made to explain and relate these differences to the physical and nonphysical factors.

Properties of the first serial correlation coefficient in particular are investigated, because this parameter is usually sufficient to describe the time dependence in hydrologic series of annual values.

Other items concerned with herein are: analysis of correlograms of individual series, as well as average values of serial correlation coefficients for entire samples; statistical inference, whenever appropriate techniques are available; discussion of regional patterns in the distribution of first serial correlation coefficient; and investigation of mathematical models for dependent time series of annual values of flow are investigated for their limitations.

B. SERIAL CORRELATION OF VARIABLES, INDEPENDENT OR DEPENDENT IN SEQUENCE

1. Distribution, expected value, and variance of serial correlation coefficients of normal independent variables. The following is stated on page 207 in Appendix 2 by P. Whittle of the book, Stationary Time Series, by Herman Wold [2]: "There are very few statistics in time series whose distributions may be evaluated exactly, and approximations are the rule rather than the exception. One of the stumbling blocks in the way of exact analysis is the 'end effect' of a finite series, which must usually be neglected, the justification being that it is sensible for a short series." Expressions given here for probability density functions as well as for expected values and variances of s. c. c. must be understood, therefore, as being only approximations.

Distributions of serial correlation coefficients, which are available in the literature, refer either to circular time series, (which was a historical development at the beginning of analyses) or to open time series, and two cases are currently treated: (a) mean and variance of the variable x in the expression for the estimate of ρ_k are considered as known; and (b) mean and / or variance of the variable x are estimated from the sample like the covariance itself.

The approximate probability density function for r_1, the first serial correlation coefficient (f. s. c. c.), of a sequence of normal independent variables estimated by eq. 1.4, or with mean μ and variance σ^2 known (in a circular time series) is given by Dixon [3, page 125, eq. 322], which is based on derivations by Koopmans [4] and Anderson [5], as

$$f(r_1) = \frac{\Gamma(\frac{N}{2}+1)}{\Gamma(\frac{N}{2}+\frac{1}{2})\sqrt{\pi}} (1 - r_1^2)^{(N-1)/2} \qquad 2.1$$

According to Whittle [2] the result of eq. 2.1 is valid for any of the first few s. c. c. of normal independent variables, or for any r_k with k relatively small.

Moments of the probability density function of eq. 2.1 are [3]:

$$M_1 = 0;\ M_2 = \frac{1}{N+2};\ M_3 = 0;\ M_4 = \frac{3}{(N+2)(N+4)} \qquad 2.2$$

This function is symmetrical and the kurtosis is $M_4/M_2^2 = 3(N+2)/(N+4)$, which is approximately 3 for sufficiently large N (say N = 30). Standardizing the variable r_1 by setting $x = r_1/\sqrt{M_2} = r_1\sqrt{N+2}$, the density function of x becomes

$$f(x) = \frac{\Gamma(\frac{N}{2}+1)\sqrt{N+2}}{\Gamma(\frac{N+1}{2})\sqrt{\pi}} (1 - \frac{x^2}{N+2})^{(N-1)/2}$$

which converges to $e^{-x^2/2}/\sqrt{2\pi}$ as N increases to infinity, for every fixed x. Therefore, the central portion of the distribution of eq. 2.1 may be approximated by a normal function with zero mean and standard deviation $1/\sqrt{N+2}$, if N is sufficiently large (say 30 or more), except near the extremes of the interval $[-1, +1]$. Dixon has shown also [3, page 126, eq. 3.21] that in case the mean and variance of x are estimated from a sample by eqs. 1.6 and 1.7 and r_k is computed by eq. 1.5 for a circular series, the moments of probability density function of r_1 (or for other r_k with k relatively small) are:

$$M_1 = \frac{-1}{N-1};\ M_2 = \frac{1}{N+1};\ M_3 = \frac{-3}{(N-1)(N+3)};$$

$$M_4 = \frac{3}{(N+1)(N+3)} \qquad 2.3$$

These results are similar to Anderson's [5]. Anderson gives the expected value of r_1 of a circular time series with r_1 estimated by eq. 1.5 in the case $k = 1$, and the mean and variance of x estimated by eqs. 1.6 and 1.7, as

$$E(r_1) = \frac{-1}{N-1} = \frac{1}{1-N} \qquad 2.4$$

but the variance of r_1 in the asymptotic form as

$$\text{var } r_1 = \frac{N-2}{(N-1)^2} \qquad 2.5$$

Siddiqui [unpublished study, 1957] used the following estimate of ρ_k for large samples and small k:

$$r_k = \frac{\sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \qquad 2.6$$

or an open time series approach. For normal independent variables, he obtained

$$E(r_1) = -\frac{1}{N} \qquad 2.7$$

and

$$E(r_1^2) = \frac{N^3 - 2N^2 + 3}{N^2(N^2-1)} \qquad 2.8$$

with the distribution of r_k,

$$f(r_k) = \frac{(1-r_k)^{(N-1)/2}(1+r_k)^{(N-3)/2}}{2^{N-1} B(\frac{N+1}{2}, \frac{N-1}{2})} + \phi(\frac{1}{N^2}) \qquad 2.9$$

For sufficiently large N (say N = 30) the last terms of eq. 2.9 in the form $\phi(1/N^2)$ may be neglected.

2. Significance test for serial correlation coefficients of normal independent variables. The single-tail significance points are given by Anderson [5] for a circular time series of size N for two different levels, 5% and 1%. For the case in which ρ_1 is estimated by eq. 1.5 and the mean and variance of x are estimated by eqs. 1.6 and 1.7, the following two equations are obtained:

$$r_1(5\%) = \frac{-1 \pm 1.645\sqrt{N-2}}{N-1} \qquad 2.10$$

and

$$r_1(1\%) = \frac{-1 \pm 2.326\sqrt{N-2}}{N-1} . \qquad 2.11$$

Anderson gives a comparison between the exact and these approximate values of significance points for positive and negative tails for 5% and 1% level and for N = 45 and N = 75. The differences are shown to be relatively small. He recommends the use of exact values for $N < 75$, and normal distribution values for $N > 75$. Dixon has shown that the normal approximation may be used for N somewhat less than 75, because differences between the exact and approximate values in Anderson's table are greater since he used an asymptotic second moment, as given by eq. 2.5, and not that given by eq. 2.3. Dixon, using the Pearson Type I approximation and the first two moments of eq. 2.3 obtained as the distribution of r_1 the fitted curve

$$f(r_1) = \frac{(1+r_1)^{p-1}\,(1-r_1)^{q-1}}{B(p,q)\;\;2^{p+q-1}} \qquad 2.12$$

in which

$$p = \frac{(N-1)\,(N-2)}{2(N-3)}\;;\quad q = \frac{N(N-1)}{2(N-3)} \qquad 2.13$$

and $B(p, q)$ is the Beta-function with p and q parameters.

The comparison between exact values, values obtained by Pearson Type I approximation, and values obtained by normal function approximation for 5% point and 1% point is given below for both positive and negative tail. This comparison is based on expressions given by Dixon [3, page 127], for r_1 estimated by eq. 1.5 with $k = 1$, and the mean and variance of x estimated by eqs. 1.6 and 1.7 for a circular time series. The sample size used is N = 30, which is the minimum value used in this study for series of annual flow and annual precipitation. This comparison is

Exact	Type I	Normal
	Positive tail, 5% point	
0.257	0.257	0.255
	Positive tail, 1% point	
0.370	0.371	0.375
	Negative tail, 5% point	
0.324	0.324	0.324
	Negative tail, 1% point	
0.433	0.433	0.444

Differences decrease with an increase of N, so that for N = 75 there is practically no difference between the three values. As tests in hydrology do not go to a point below 1%, but are stopped usually at the point of 2.5% or 5%, it can be assumed that the normal distribution is a sufficient approximation for all necessary tests in the case of normal independent variables (null hypothesis for r_k) even when N = 30. If eq. 1.4 is assumed to have been used for the estimate of ρ_k (sufficiently long time series), the normal function approximation to eq. 2.1 and the first two moments of eq. 2.2 may be used for the test of significance of the computed r_k if tail point is not below 2.5%. Equation 2.9 may also be approximated by a normal function, if N is sufficiently large (say N = 30), and the tail point test is not below 2.5%. If the point is smaller than 2.5% (say 1%), Fisher's z-transformation of r_k may be used. In this case, the normal function for $\bar{z}$ and $s(z)$ may be applied, where $s(z)$ is the standard deviation of the z-transform.

Assuming $\rho = 0$ in a simple linear correlation of two variables, then according to Fisher [6]

$$p(r) = \frac{(1-r^2)^{(N-4)/2}}{B[\frac{1}{2}\,,\,\frac{N-2}{2}]} \qquad 2.14$$

This equation is similar to eq. 2.1 except that the power is $(N-1)/2$ in eq. 2.1 with a different constant part. Equation 2.1 becomes equivalent to the above equation if N in eq. 2.1 is replaced by N-3. Therefore, Fisher's z-transformation (usually applied to simple correlation coefficients) may be applied also to serial correlation coefficients, if N is replaced by N+3 in eq. 2.14. The currently used Fisher's z-transform is

$$z = \frac{1}{2}\log_e\frac{1+r}{1-r} \qquad 2.15$$

with the expected value

$$E(z) \approx \frac{1}{2}\log_e\frac{1+\rho}{1-\rho} . \qquad 2.16$$

The variance of z is given by

$$\text{var } z = \frac{1}{N-3} . \qquad 2.17$$

For serial correlation coefficient with small k of an open time series, N in eq. 2.14 should be replaced by N+3. In this case

$$\text{var } z = \frac{1}{N} \qquad 2.18$$

gives the estimate of the variance of z-transform of the serial correlation coefficient r_k for the first few k values.

For a large number of series and for normal independent variables, the values r_k of k relatively small in comparison to N are normally distributed in the range 2.5% and 97.5% of probability for N at least 30. To test whether or not the s.c.c.'s, r_k, are significantly different for an observed series from those of normal independent variables, the 2.5% and 97.5% tail levels are usually determined from the normal function in this study. The test of significance

will be designated as that at 95% level.

The number of stations of flow or precipitation in this study is n. Individual stations have records of different length N_j. There is, therefore, a need for computation of the weighted mean and the weighted variance of r_k for both the observed series and the normal independent variables. For observed series the mean $\bar{r}_k$ is computed in two ways:

(a) Simply as

$$\bar{r}_k = \frac{1}{n} \sum_{j=1}^{n} r_{kj} \qquad 2.19$$

where r_{kj} represents the r_k value for the j-th station with the sample size N_j;

(b) As the weighted mean by using different sample sizes for time series as

$$\bar{r}_k^* = \sum_{j=1}^{n} N_j r_{kj} \Big/ \sum_{j=1}^{n} N_j \quad , \qquad 2.20$$

thus giving a larger weight to r_k values for series with long records.

The variance of r_k for a sample of n stations, and N_j which is changing from station to station, is also computed in two ways:

(a) Simply as

$$\text{var } r_k = \frac{1}{n} \sum_{j=1}^{n} (r_{kj} - \bar{r}_k)^2 \qquad 2.21$$

by using either n values in ungrouped r_k value approach, or using the frequency of r_k intervals in the grouped r_k value approach; and

(b) As weighted variance,

$$\text{var } r_k = \sum_{j=1}^{n} N_j (r_{kj} - \bar{r}_k^*)^2 \Big/ \sum_{j=1}^{n} N_j \qquad 2.22$$

For a sample of stations the value $\sum_{j=1}^{n} N_j$ is a constant.

For normal independent variables (or benchmark variables) the mean $\bar{r}_k = 0$, if eq. 1.4 is assumed to be used for the computation of r_k. If eq. 1.5 or its approximation eq. 1.10 is used for the estimates of ρ_k, the mean $\bar{r}_k$ is computed in two ways:

(a) By using the average length N_m of series for a given sample of stations and eq. 2.4 as

$$\bar{r}_k = \frac{1}{1 - N_m} \qquad 2.23$$

and

(b) By using the weight N_j and eq. 2.4 as

$$\bar{r}_k^* = \sum_{j=1}^{n} \frac{N_j}{1-N_j} \Big/ \sum_{j=1}^{n} N_j \qquad 2.24$$

When eq. 2.7 is used, there also are two cases:

(a) By using the average length N_m of series with $\bar{r}_k = -1/N_m$, and

(b) By using the weight N_j as

$$\bar{r}_k^* = -\frac{1}{n} \sum_{j=1}^{n} \frac{1}{N_j} \qquad 2.25$$

The variance of r_k for normal independent variables with given n, and the weight N_j, is computed also in two ways:

(a) By using the mean length N_m and equation 2.2

$$\text{var } r_k = \frac{1}{N_m + 2} \qquad 2.26$$

or N_m and eq. 2.3

$$\text{var } r_k = \frac{1}{N_m + 1} \qquad 2.27$$

(b) By using the weight N_j and eq. 2.2

$$\text{var } r_k = \sum_{j=1}^{n} \frac{N_j}{N_j + 2} \Big/ \sum_{j=1}^{n} N_j \qquad 2.28$$

and the weight N_j and eq. 2.3

$$\text{var } r_k = \sum_{j=1}^{n} \frac{N_j}{N_j + 1} \Big/ \sum_{j=1}^{n} N_j \qquad 2.29$$

For testing the significance whether the mean $\bar{r}_k$ or the mean $\bar{r}_k^*$ of an observed sample of series (n series) is or is not significantly different from the corresponding means $\bar{r}_k$ or $\bar{r}_k^*$ of normal independent variables; the normal distribution of means of normal independent variables is determined by using the variance

$$\text{var } \bar{r}_k = \frac{1}{n} \text{var } r_k \qquad 2.30$$

and similarly for $\bar{r}_k^*$ by replacing r_k in eq. 2.30 by r_k^*. In eq. 2.30 var r_k or var r_k^* is given by respective expressions of eqs. 2.26 through 2.29. The values $\bar{r}_k$ and $\bar{r}_k^*$ are assumed to be normally distributed about $\bar{\rho}_k$ or $\bar{\rho}_k^*$, respectively.

3. Non-normal variables. Bartlett [7] has shown that for large samples the variance and covariance of r_k are independent of the distribution of x under fairly wide conditions. According to Quenouille [8] "this means that their joint distribution function obtained for normal independent variable will often give a good approximation for nonnormal independent variables, and can be used as the basis for any test of the correlogram." In general, if the kurtosis of

the nonnormal distribution is around 3 (or excess around zero), the above statement is valid. Usually this condition is satisfied with variables like annual flow and annual precipitation. It can be assumed, therefore, that the distribution functions of annual flow or annual precipitation are nearly independent of the properties of s. c. c. of their series. Furthermore, the significance test of these properties may be carried out regardless of whether or not the distributions of flow or precipitation are skewed.

4. Distribution, expected value, and variance of serial correlation coefficients of dependent variables. It is assumed here that the properties of serial correlation coefficients of non-normal dependent variables are approximately the same as those of normal dependent variables, so that it is sufficient to investigate the latter case.

The general mathematical model which relates V- and P_e-series [1, page 13, eq. 10] as affected by water carryover in river basins (which storage is bound to flow out of river basins either by surface or underground flow) is

$$V_n = \sum_{j=o}^{j=\infty} b_j P_{n-j} + e_n \qquad 2.31$$

with V_n the annual flow of the n-th year; P_{n-j} the annual effective precipitation of (n-j)-th year, or j-years preceding the n-th year; b_j the coefficients; e_n a random component which takes care of the fact that b_j coefficients are only the average values. The b_j coefficients should satisfy six conditions [See Part I, 1]. The main ones are: sum of b_j coefficients is unity; all are positive; and they decrease monotonically.

The general moving average scheme of eq. 2.31 embraces simple mathematical models like autoregressive linear schemes of the first and second order (Markov first and second order linear models), and similar models.

Assuming that e_n in eq. 2.31 is zero, and that $\{P_i\}$ is a sequence of mutually independent variables, which replaces P_{n-j} variable in eq. 2.31, this equation may be transformed by a recurrence procedure in V_n as function of V_{n-j} values. Thus eq. 2.31 for $b_o P_n = \epsilon_n$ becomes

$$V_n = \epsilon_n + a_1 V_{n-1} + a_2 V_{n-2} + \ldots =$$

$$= \sum_{j=o}^{\infty} a_j V_{n-j} + \epsilon_n \qquad 2.32$$

in which

$$a_1 = \frac{b_1}{b_o} \quad ;$$

$$a_2 = \frac{b_2}{b_o} - \frac{b_1^2}{b_o^2} \quad ;$$

$$a_3 = \frac{b_3}{b_o} + \frac{b_1^3}{b_o^3} - \frac{2b_1 b_2}{b_o^2} \quad ;$$

$$a_4 = \frac{b_4}{b_o} - \frac{b_1^4}{b_o^4} - \frac{b_2^2}{b_o^2} - \frac{2b_1 b_3}{b_o^2} + \frac{3b_1^2 b_2}{b_o^3} \quad ; \text{ etc.}$$

These coefficients a_j do not decrease monotonically; they may be either positive or negative, and their sum is not unity. It should be stressed that the a_j coefficients have been derived from b_j coefficient under the assumption that P_{n-j} or $\{P_i\}$ in eq. 2.31 are mutually independent variables. It will be shown later that P_{n-j} or $\{P_i\}$ are not actually mutually independent variables.

Replacing $\{V_n\}$ in eq. 2.31 by a sequence of dependent standard normal variables $\{x_i\}$, and $\{P_i\}$ by a sequence of independent standard normal variables $\{y_i\}$, and putting $e_n = 0$, eq. 2.31 becomes

$$x_i = \beta_o y_i + \beta_1 y_{i-1} + \ldots = \sum_{k=0}^{\infty} \beta_k y_{i-k} \qquad 2.33$$

The relationship of β_k and b_k coefficients in this case is

$$\beta_k = \frac{b_k}{\left[\sum_{j=o}^{\infty} b_j^2\right]^{1/2}} \qquad 2.34$$

because the ratio of standard deviations of $\{P_i\}$ and $\{V_i\}$ variables in eq. 2.31 is

$$1 \Big/ \left[\sum_{j=o}^{\infty} b_j^2\right]^{1/2} .$$

Since $\sum_{j=o}^{\infty} b_j = 1$, $E(x_i) = E(y_i)$,

$E(x_i^2) = E(y_i^2)$, and $E(y_i y_{i+k}) = 0$ for $k \neq 0$ and

var x_i = var y_i = 1,

$$\sum_{k=o}^{\infty} \beta_k^2 = 1 . \qquad 2.35$$

It follows from eq. 2.33 also that

$$\beta_k = E(x_i y_{i-k}) . \qquad 2.36$$

Since $\rho_k = E(x_i x_{i+k})$, then

$$\rho_k = \sum_{j=o}^{\infty} \beta_j \beta_{j+k} = \frac{\sum_{j=o}^{\infty} b_j b_{j+k}}{\sum_{j=o}^{\infty} b_j^2} \qquad 2.37$$

The mathematical expression of eq. 2.31 requires that $\beta_k > 0$ for all k, with $\beta_o > 0$.

According to eq. 2.34 the condition

$$\sum_{j=o}^{\infty} \beta_k = \left[\sum_{j=o}^{\infty} b_j^2\right]^{-1/2} > 1 \qquad 2.38$$

must be also satisfied because $\sum_{j=o}^{\infty} b_j^2$ is smaller than unity.

The special case of eq. 2.32 is the first order linear autoregressive scheme (Markov first order linear model), for which eq. 2.32 is expressed with only one term under summation. Putting $a_1 = \rho$ and $j=1$ in eq. 2.32, this scheme becomes

$$x_i = \rho x_{i-1} + \epsilon_i \qquad 2.39$$

where $\rho = \rho_1$ is the first serial correlation coefficient of the variable x. Inserting for x_i and x_{i+1} the corresponding values from eq. 2.39 into eq. 2.33, which substitution implies that $\epsilon_i = \beta_o y_i$ for all i, then

$$\beta_k = \rho\beta_{k-1} = \rho^2\beta_{k-2} = \ldots = \rho^k\beta_o$$

For conditions of stationarity $|\rho|$ is smaller than unity. Since

$$\Sigma\beta_k^2 = 1\,, \quad \beta_o^2\Sigma\rho^{2k} = \beta_o^2/(1-\rho^2) = 1\,, \text{ or}$$

$$\beta_k = \rho^k(1-\rho^2)^{1/2} \qquad 2.40$$

Equation 2.37, with the values for β_j and β_{j+k} given by eq. 2.40, becomes

$$\rho_k = \rho^k \qquad 2.41$$

As $\beta_k > o$, then $1 > \rho > o$. In this case ρ is estimated by the first serial correlation coefficient r_1 of the sample as

$$r_1 = \frac{N}{N-1}\,\frac{\sum_{i=1}^{N-1} x_i x_{i-1}}{\sum_{i=o}^{N} x_i^2} \qquad 2.42$$

Since $\rho_k = \rho^k$, the correlogram is of an exponential type, with ρ_k decreasing asymptotically to zero as k increases to infinity.

According to Madow [9] and Leipnik [10] for ρ in eqs. 2.39 and 2.40 estimated by eq. 2.42, if $N \geq 20$, the distribution of r_1 ($\equiv r$) is

$$f(r) = \frac{\Gamma\left[\frac{N+2}{2}\right]}{\Gamma\left[\frac{N+1}{2}\right]\sqrt{\pi}}(1-2r\rho+\rho^2)^{-\frac{N}{2}}(1-r^2)^{\frac{N-1}{2}} \qquad 2.43$$

with

$$E(r) = \frac{N\rho}{N+2}\,; \text{ and } E(r^2) = \frac{1}{N+2} + \frac{N(N+1)\rho^2}{(N+2)(N+4)} \qquad 2.44$$

and

$$\text{var } r = \frac{1}{N+2} - \frac{N(N-2)\rho^2}{(N+2)^2(N+4)}\,. \qquad 2.45$$

Quenouille [11] states that the transform $z = \tanh^{-1} r$, with $\xi = \tanh^{-1}\rho$ is approximately normally distributed with

$$E(z) = \xi - \frac{\rho}{N(1-\rho^2)} + \frac{\rho(1+\rho^2)}{N(1-\rho^2)} \qquad 2.46$$

and

$$\text{var } z = \frac{1}{N(1-\rho^2)} - \frac{2\rho^2}{N^2(1-\rho^2)^2} \qquad 2.47$$

which requires the knowledge of ρ.

Quenouille [12] and White [13, 14] use t-transform of r as

$$t = \frac{(r-\rho)\sqrt{N+1}}{\sqrt{1-r^2}} \qquad 2.48$$

which is distributed as

$$f(t) = K(t)\left[1 - \frac{\rho t}{[t^2 + (N+1)(1-\rho^2)]^{1/2}}\right] \qquad 2.49$$

where

$$K(t) = \frac{\Gamma\left[\frac{N}{2}+1\right]}{\Gamma\left[\frac{N+1}{2}\right]}\,\frac{\left[1+\frac{t^2}{N+1}\right]^{-(N+2)/2}}{\sqrt{\pi(N+1)}} \qquad 2.50$$

which is the Student t-distribution with N+1 degrees of freedom. To test the hypothesis that $\rho_1 = 0$, eq. 2.43 with $\rho = 0$ may be used and the Student t-distribution with N+1 degrees of freedom applies for a two-sided test. A knowledge of the mean is required.

To show the applicability of the model of eq. 2.39, it is assumed here that the flow recession curves at the end of water years are approximated by an exponential curve of the type $Q = Q_o e^{-ct}$ [1, p. 19]. The b_j coefficients are relative areas (area divided by total area $W = Q_o/c$) for t between 0-1 (b_o), 1-2(b_1), 2-3(b_2), etc. In this case the b_j coefficients are

$$1 - \frac{1}{e^c}\,;\ \frac{1}{e^c}\left[1 - \frac{1}{e^c}\right];\ \frac{1}{e^{2c}}\left[1 - \frac{1}{e^c}\right];\ \frac{1}{e^{3c}}\left[1 - \frac{1}{e^c}\right];\ \ldots$$

and they give $\sum_{j=o}^{\infty} b_j = 1$; $b_k = e^{-c}b_{k-1}$; and $\beta_k = e^{-c}\beta_{k-1}$. Substituting x of eq. 2.32 into eq. 2.33 and equating the coefficients for y_i, y_{i-1}, y_{i-2}, ... and using the relationship

$$\beta_k = e^{-ck}\beta_o \qquad 2.51$$

then $\epsilon_i = \beta_o y_i$; $\beta_1 = a_1\beta_o$; $\beta_2 = a_1\beta_1 + a_2\beta_2$; $\beta_3 = a_1\beta_2 + a_2\beta_3 + a_3\beta_4$; etc., which give $a_1 = e^{-c}$; $a_2 = o$; $a_3 = o$; $a_4 = o$; etc. The simple exponential curve fitted to recession curves of flow leads to the first order linear autoregressive scheme of eq. 2.39 with $\rho_1 = e^{-c}$, because $a_1 = \rho_1$; and $\rho_k = e^{-ck}$.

The next special case is the second order linear autoregressive scheme, in which model only two terms with a_i coefficients different from zero are used in eq. 2.32, or

$$x_i = a_1 x_{i-1} + a_2 x_{i-2} + \epsilon_i \qquad 2.52$$

Substituting the values of x of eq. 2.52 into eq. 2.33, and equating coefficient for y_i, y_{i-1}, . . . , then $\epsilon_i = \beta_o y_i$; $\beta_1 = a_1\beta_o$; $\beta_k - a_1\beta_{k-1} - a_2\beta_{k-2} = 0$, for $k = 2, 3, \ldots$ To solve the last equation by finite differences, $\beta_k = \beta^k$ is assumed. It gives $\beta^2 - a_1\beta - a_2 = o$, or $2\beta = a_1 \pm \sqrt{a_1^2 + 4a_2}$. For stationarity the absolute value of β must be less than one, i.e. $\beta = r\cos\theta + i\, r\sin\theta$, with $|r| < 1$. So $a_1^2 + 4a_2 \leq o$, or $4a_2 \leq -a_1^2$, or a_2 is negative. From this $\cos\theta = a_1 / 2\sqrt{-a_2}$; and $2\sin\theta = \sqrt{4 + a_1^2/a_2}$, so that $\beta = r\, e^{\pm i\theta}$, and $\beta_k = r^k(Ae^{ik\theta} + Be^{-ik\theta})$. For $k = o$, and $k = 1$, $A + B = \beta_o$; $a_1\beta_o = r(Ae^{-i\theta} + Be^{i\theta})$. As $r(e^{-i\theta} + e^{+i\theta}) = a_1$, then

$$\beta_k = \frac{\beta_o r^k}{\sin\theta}\sin(k+1)\theta \qquad 2.53$$

Equation 2.53 implies that β_k can be also negative when $\sin(k+1)\theta/\sin\theta$ becomes negative. This model is used regardless of the fact that β_k becomes negative when k is sufficiently large. However, only positive values of β_k are used. It is assumed that $\beta_k = 0$ as soon as it becomes negative. The smaller the angle θ , the larger is the lag k before β_k becomes negative. Since $(k+1)\theta$ should be smaller than π for β_k to be positive, then $k \leq (\frac{\pi}{\theta} - 1)$. In this case if θ is 60^0 then $k \leq 2$, or $\beta_2 = o$, and only β_o and β_1 are assumed as positive and different from zero. This gives

$$\beta_o^2 = \frac{(4a_2 + a_1^2)(1 + a_2)(1 + 2a_2\cos 2\theta + a_2^2)}{2a_2(1 - a_2)(1 - \cos 2\theta)} \qquad 2.54$$

The values a_1 and a_2 for the determination of β_o are obtained from the expression

$$\rho_k = a_1\rho_{k-1} + \rho_{k-2} \cdot \qquad 2.55$$

By using r_1 and r_2 as estimates of ρ_1 and ρ_2 respectively, by putting $\rho_{-k} = \rho_k$ and $\rho_o = 1$, and using values of $k = 1$ and $k = 2$, a_1 and a_2 are found to be

$$a_1 = \frac{r_1 - r_1 r_2}{1 - r_1^2} \quad ; \quad a_2 = \frac{r_2 - r_1^2}{1 - r_1^2} \qquad 2.56$$

with the condition that $a_2 \leq -a_1^2/4$ must be satisfied if the model is to fit. The multiple correlation coefficient R is estimated by

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

and $\sqrt{1 - R^2}$ is also an estimate of β_o .

The question arises also what should be the shape and mathematical function fitted to flow recession curves in order that the second order linear autoregressive scheme would be an appropriate mathematical model for time dependence of annual flows. In general, an expression of the type

$$Q = Q_o e^{-ct^n} \qquad 2.58$$

can be fitted to flow recession curves with Q_o initial flow (flow at the end of the year on the recession curve), and c and n the two main parameters, with n any real number, usually greater than unity. Assuming that the area of this curve from $t = o$ to $t = \infty$ ($Q = Q_o$ to $Q = o$) is W, then the b_k coefficient may be assumed to be

$$b_k = \frac{Q_o}{W}\int_k^{k+1} e^{-ct^n} dt \quad . \qquad 2.59$$

Substituting $ct^n = x$, then

$$b_k = \frac{Q_o}{W}\frac{1}{nc^{1/n}}\int_{ck^n}^{c(k+1)^n} e^{-x} x^{\frac{1}{n} - 1} dx \qquad 2.60$$

which is an incomplete gamma function of $1/n$ for given limits ck^n and $c(k+1)^n$, with $k = o, 1, 2, \ldots$ Using the tables of incomplete gamma function for given limits and the coefficient $Q/(Wnc^{1/n})$, the values b_o , b_1 , . . . may be computed. From b_j coefficients thus computed, the a_j coefficients of eq. 2.32 may be determined. If it happens that for given values of c and n the coefficient a_1 is positive but a_2 is negative, while a_3 and a_4 are very small, the second order linear autoregressive scheme of eq. 2.52 may be applied. The fact that eq. 2.58 is usually only an approximation to a group of recession curves plotted for a river gaging station should not be overlooked.

5. Effective number of stations with interstation correlation. When such statistics as the mean for many stations or the mean of the means, mean variance, mean first serial correlation, etc. are

studied for river flow or precipitation, the fact that these variables are correlated for stations taken pairwise (interstation correlation) cannot be disregarded. The effective number of stations n_e for the interstation mean $\bar{x}$ of a statistic x is defined here from the relation $\text{var}\,\bar{x} = \text{var}\,x/n_e$. The effective number of stations n_e means that n stations which are correlated have the same variance of $\bar{x}$ as n_e stations which are uncorrelated (independent). Correlated or uncorrelated stations in this case imply that the values of a statistic x are or are not correlated, respectively.

The effective number of stations n_e depends upon the statistical parameter under investigation. One n_e value is obtained for the regional mean, another for the regional mean variance, a third for the regional mean first serial correlation coefficient, etc., if the same basic variable with the same number of stations is investigated.

It is assumed here that ρ_1 values of the f. s. c. c. for time series of annual flow and annual precipitation of different stations are not independent among themselves. Unfortunately, the length of time series for most stations is too small to permit a division of the series into several parts for each of the two stations, or to allow the computation of corresponding r_1 values of the f. s. c. c. for each part, and then to determine the simple correlation coefficient for the concurrent values of r_1. Therefore, the interstation correlation coefficients between values of basic variable x must be used. *

The expression for effective number of stations, n_e, of f. s. c. c. is determined here by assuming that the same variable is investigated for two stations (say annual flow, or annual precipitation). This variable is designated here by x for the first station and by y for the second station (any pair of stations). They are assumed to be standardized (mean 0, variance 1). Stations are assumed to have series of the same length N. The covariances of successive values of two series are

$$C_1(x) = \frac{\sum_{i}^{N-1} x_i x_{i+1}}{N-1} \quad ; \quad C_1(y) = \frac{\sum_{j}^{N-1} y_j y_{j+1}}{N-1} \qquad 2.61$$

with

$$EC_1(x) = \rho_1(x) \quad \text{and} \quad EC_1(y) = \rho_1(y)$$

which are the first serial correlation coefficients for x and y, respectively. The expected value of the product of these two covariances is

$$E\,C_1(x)\,C_1(y) = \frac{1}{(N-1)^2} \sum_{j=1}^{N-1} \sum_{j=1}^{N-1} E(x_i x_{i+1} y_j y_{j+1}) \ .$$

If the x and y variables are approximately normally distributed (Gaussian distribution), then

$$E\,C_1(x)\,C_1(y) = \frac{1}{(N-1)^2} \sum_{j=1}^{N-1} \sum_{i=1}^{N-1} [\,E(x_i x_{i+1})\,E(y_j y_{j+1}) +$$

$$+ E(x_i y_j)\,E(x_{i+1} y_{j+1}) + E(x_i y_{j+1})\,E(x_{i+1} y_j)\,] .$$

It is assumed here that the simultaneous values x_k, y_k are correlated, while the time lag correlation of x_k, y_{k+s}, for $s \neq o$, is negligible, or that $E(x_i y_j) = o$, if $i \neq j$; and $E(x_i y_j) = \rho$, if $i=j$. Also the stationarity condition gives that $E(x_i x_{i+1}) = \rho_1(x)$, and $E(y_i y_{j+1}) = \rho_1(y)$, because $\sigma_x^2 = 1$ and $\sigma_y^2 = 1$.

The value of ρ is the simple correlation coefficient among simultaneous values of x and y. Then $E(x_i x_{i+1} y_j y_{j+1}) = \rho_1(x)\,\rho_1(y)$, if $i \neq j$ and $E(x_i x_{i+1} y_j y_{j+1}) = \rho_1(x)\rho_1(y) + \rho^2$, if $i = j$. The expected value of the product of covariances is

$$EC_1(x)\,C_1(y) = \frac{1}{(N-1)^2} \Big\{ (N-1)\,[\,\rho_1(x)\rho_1(y) + \rho^2\,] +$$

$$+ (N-2)(N-1)\rho_1(x)\rho_1(y) \Big\} = \rho_1(x)\rho_1(y) + \frac{\rho^2}{N-1} \ . \qquad 2.62$$

This gives

$$\text{Cov}\,[\,C_1(x)\ ,\ C_1(y)\,] = \frac{\rho^2}{N-1} \ . \qquad 2.63$$

The sample estimates of $\rho_1(x)$ and $\rho_1(y)$ are $r_1(x)$ and $r_1(y)$, respectively. The expected value of their product is

$$Er_1(x)\,r_1(y) = E\,\frac{C_1(x)\,C_1(y)}{C_0(x)\,C_0(y)} =$$

$$= \frac{EC_1(x)\,C_1(y)}{EC_0(x)\,C_0(y)}\,[\,1 + \theta(1/N)\,],$$

where $C_0(x)$ and $C_0(y)$ are sample variances of x and y, respectively, and $\theta(1/N)$ represents terms which are of the order $1/N$ as compared to unity. The latter mentioned terms will be considered here as negligible. Now

$$EC_0(x)\,C_0(y) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} E(x_i^2 y_j^2)$$

$$= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} [\,Ex_i^2 Ey_j^2 + 2(Ex_i y_j)^2\,]$$

$$= 1 + \frac{2\rho^2}{N} \ . \qquad 2.64$$

Thus,

$$Er_1(x)r_1(y) = [\,\rho_1(x)\,\rho_1(y) + \frac{\rho^2}{N-1}\,]\,[\,1 - \frac{\rho^2}{N}\,] + \theta(\frac{1}{N^2})$$

$$= \rho_1(x)\,\rho_1(y) + \frac{\rho^2}{N-1} - \frac{2\rho^2\rho_1(x)\rho_1(y)}{N} + \theta(\frac{1}{N^2}) \ . \qquad 2.65$$

*The following equations 2.61-2.73 have been derived in consultation with M. M. Siddiqui, Colorado State University.

Finally

$$\text{Cov}\,[r_1(x),\ r_1(y)] = \frac{\rho^2}{N-1} - \frac{2\rho^2\rho_1(x)\rho_1(y)}{N} + \theta\left(\frac{1}{N^2}\right) \qquad 2.66$$

Again, if either $\rho_1(x)$ or $\rho_1(y)$ or both are small, the second term will be negligible compared to the first. Hence, under these circumstances,

$$\text{Cov}\,[r_1(x),\ r_1(y)] = \frac{\rho^2}{N-1} \qquad 2.67$$

approximately.

Variances of $r_1(x)$ and $r_1(y)$ depend upon the mathematical model of serial correlation of x, and of y. It is assumed here that the simple model may be used in the form of the first order linear auto-regressive scheme. Then the variances of $r_1(x)$ and $r_1(y)$ are given by eq. 2.45 in which ρ^2 is replaced by $\rho_1^2(x)$ and $\rho_1^2(y)$, respectively for the two series. The last term in eq. 2.45 is not negligible in comparison with the preceding term which is $1/(N+2)$. For $N = 30$, and $r_1(x)$ or $r_1(y)$ being assumed equal to 0.5 (which are large values for f. s. c. c. of annual flow and annual precipitation), the two terms in eq. 2.45 are: $1/(N+2) = 0.0313$, and the last term is 0.00603. The last term is 19.2% of the preceding term. For small values $r_1(x)$ or $r_1(y)$, or around 0.10 - 0.20, the last term in eq. 2.45 may be neglected. However, due to different values of $r_1(x)$ or $r_1(y)$, the last term in eq. 2.45 is retained here and approximated by:

$$[\text{var}\, r_1(x)\ \text{var}\, r_1(y)]^{1/2} =$$

$$= \frac{1}{N+2}\left\{1 - \frac{N(N-2)\,[\rho_1^2(x) + \rho_1^2(y)]}{(N+2)(N+4)}\right\}^{1/2} \qquad 2.68$$

with the term $[\rho_1^2(x)\rho_1^2(y)]$ neglected because of the fourth power of the product $\rho_1(x)$ and $\rho_1(y)$.

For ρ_{ij} designating $\rho[\rho_1(x),\ \rho_1(y)]$ or the correlation coefficient of the f. s. c. c. between two stations, it becomes

$$\rho_{ij} = \frac{\text{Cov}\,[r_1(x)\, r_1(y)]}{[\text{var}\, r_1(x)\ \text{var}\, r_1(y)]^{1/2}}\ , \text{ or}$$

$$\rho_{ij} = \frac{N+2}{N-1}\rho^2\left\{1 + \frac{N(N-2)\,[\rho_1^2(x) + \rho_1^2(y)]}{2(N+2)(N+4)}\right\} \qquad 2.69$$

To estimate ρ_{ij} the value ρ in eq. 2.69 is estimated by r, product-moment correlation coefficient of x and y; $\rho_1(x)$ by $r_1(x)$, the first serial correlation coefficient of x; $\rho_1(y)$ by $r_1(y)$ the first serial correlation coefficient of y; and N by N_{ij} the number of simultaneous values of x and y in their correlation.

The statistic $\bar{r}_1$, or the regional mean of the first serial correlation coefficient for a sample of series, of sample size n, is tested for significance by using the following expressions. The statistic $\bar{r}$ is

$$\bar{r}_1 = \frac{\sum_{i=1}^{n} r_1}{n}\ . \qquad 2.70$$

Its variance is

$$\text{var}\,\bar{r}_1 = \frac{\text{var}\, r_1}{n}[1 + \Sigma(n-1)\bar{\rho}]\ , \qquad 2.71$$

where

$$\bar{\rho} = \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\rho_{ij}\ , \qquad 2.72$$

and ρ_{ij} are defined by eq. 2.69. As seen from eq. 2.69, to estimate ρ_{ij} one needs the estimates of ρ, $\rho_1(x)$ and $\rho_1(y)$. These are easily obtained from the samples of x and y series. To be more specific, let $\hat{\rho}_{ij}$ designate the estimate of ρ_{ij}, r_{ij} the simple correlation coefficient between x and y series based on N_{ij} observations, and $r_1(x)$ and $r_1(y)$ the f. s. c. c. based on the same number of observations, then, from eq. 2.69,

$$\hat{\rho}_{ij} = \frac{N_{ij}+2}{N_{ij}-1}\, r_{ij}^2\left\{1 + \frac{N_{ij}(N_{ij}-2)[r_1^2(x) + r_1^2(y)]}{2(N_{ij}+2)(N_{ij}+4)}\right\} \qquad 2.73$$

Finally, the effective sample size n_e for $\bar{r}_1$ is estimated to be

$$n_e = \frac{n}{1 + \bar{r}\,(n-1)}\ , \qquad 2.74$$

where $\bar{r}$ is the estimate of $\bar{\rho}$ obtained by using eq. 2.72 with ρ_{ij} replaced by $\hat{\rho}_{ij}$ from eq. 2.73.

As $r_1(x)$ and $r_1(y)$ are computed for the total length of series x and y, respectively, and not only for simultaneous observations in the two series, N_{ij}, the value $r_1(x)$ and $r_1(y)$ for the total series x and y will be used in eq. 2.73 as better estimates of $\rho_1(x)$ and $\rho_1(y)$, respectively, than $r_1(x)$ and $r_1(y)$ values for sample size N_{ij}.

The value $\hat{\rho}_{ij}$ of eq. 2.73 is a biased value, because it is always positive (like the multiple correlation coefficient R). Even if the series of x and y are serially uncorrelated and mutually independent, the sampling deviations of $r_1(x)$; $r_1(y)$; and r_{ij} about their true values will produce positive values of $r_1^2(x)$; $r_1^2(y)$, and r_{ij}^2, and therefore, a positive $\hat{\rho}_{ij}$. However, if r_{ij}, $r_1(x)$ and $r_1(y)$ are small, their squares are very small also, and they will produce a relatively negligible positive value of $\hat{\rho}_{ij}$.

C. ANALYSIS OF FIRST LARGE SAMPLE OF RIVER FLOW RECORDS

1. Method of data analysis. A dilemma existed when the method for data analysis had to be selected: (1) to treat initially and separately the characteristics of series of annual flow, then of annual effective precipitation, and finally of precipitation, and afterwards to study the relationships of their individual patterns in sequence; or (2) to study series of all these variables--either flow and effective precipitation; or flow, effective precipitation and precipitation at the ground--simultaneously for a sampling region, both for their patterns in sequence and for their interrelationship. This second method has been adopted here for the simple reason that the two aspects: patterns in sequence of individual variables and the relationship between the three variables may be best investigated simultaneously in their complex interdependence.

2. Frequency distribution of first serial correlation coefficient. The number of river gaging stations of the first large sample of river flow records is 140, so there are 140 f. s. c. c. for the study of its distribution. The mean length of these 140 time series of annual flow (V-series) and annual effective precipitation (P_e-series) is $N_m = 55$, and the extreme lengths are $N_{min} = 37$ and $N_{max} = 150$.

Two cases are analyzed in comparing these two series with the normal independent variables: (a) the average length $N_m = 55$ of time series in this sample is used as if all series were of that same length; and (b) the length N_j of a particular series is used as a weight of r_1 for that series. For this second approach and for the computation of statistics of normal independent variables the distribution of time length N_j of n = 140 stations is given in fig. 1 for a grouped data computational approach.

Average r_1-values, variances of r_1, and the variance of $\bar{r}_1$ are computed by equations 2.19 through 2.30 by setting k = 1. These were computed for each of the three sample series: V, P_e, and normal independent variables.

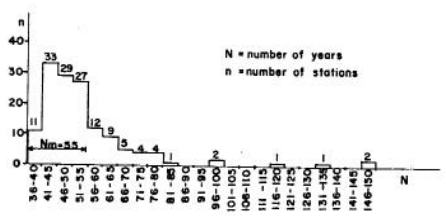

Fig. 1 Distribution of the time lengths, N_j, for the series of the first large sample of river gaging stations (n = 140).

Table 1 gives the statistics computed for the first serial correlation coefficient for the three samples of time series. When N_j of series is taken into account for the computation of weighted statistics of V-series and P_e-series, the same N_j distribution was assumed for the sample of series of normal independent variables. Table 1 gives also the equation which has been used for the estimation of a particular parameter.

Values in Table 1 show that on the average the weighted length N_j of series in this first sample does not give substantially different estimates of statistics of r_1 when compared with the values estimated by using the average length N_m.

Cumulative distributions of first serial correlation coefficient, as well as frequency histograms of the f. s. c. c. are given in Fig. 2 in cartesian scales for both annual flow and annual effective precipitation for 140 stations.

Figure 3 gives frequency distributions of the first serial correlation coefficient r_1 in cartesian-probability scales for both annual flow (V) and annual effective precipitation (P_e). The means given are simple averages estimated by eq. 2.19. Distributions of r_1 for normal independent variables with series length $N = N_m = 55$ are given as straight lines in these scales for two cases: (a) Mean and variance of r_1 estimated by moments of eq. 2.2; and (b) Mean and variance of r_1 estimated from moments of eq. 2.3 by using eqs. 2.23 and 2.27.

It is necessary to stress that many values of r_1 are negative: 16 (or 11.4%) in the V-series and 26 (or 18.6%) in the P_e-series, as it is shown in Fig.2. Approximate corrections for the water carryover from year to year, made in this study, have thus increased the number of negative r_1 values for annual effective precipitation in comparison with that of annual flow.

Shapes of histograms and curves, (1) through (4) in Fig. 2, and curves (1) and (2) in Fig. 3 show that frequency distributions in their middle part are close to the normal distribution, but tails at both extremes have an appreciable departure from the normal distribution.

Table 1 shows that the correction of V-series to obtain P_e-series by approximate values of the change in carryover ΔW_i [1, page 18] decreased

TABLE 1

Statistics of frequency distribution of first serial correlation coefficient for 140 stations of annual flow (V - series) and of annual effective precipitation (P_e - series), as well as for normal independent variable.

Statistics	V - Series		P_e - Series		Normal Independent Variable			
	Equation Used	Value	Equation Used	Value	Equation Used	Value	Equation Used	Value
Simple estimate of mean	2.19	0.1748	2.19	0.1353	2.23	-0.0185	2.2	0
Weighted estimate of mean	2.20	0.1844	2.20	0.1336	2.24	-0.0186	2.2	0
Median		0.1600		0.1150				
Simple estimate of variance	2.21	0.0354	2.21	0.0303	2.27	0.0178	2.26	0.0176
Weighted estimate of variance	2.22	0.0370	2.22	0.0292	2.29	0.0179	2.28	0.0176
Simple estimate of standard deviation	2.21	0.1875	2.21	0.1738	2.27	0.1335	2.26	0.1325
Weighted estimate of standard deviation	2.22	0.1920	2.22	0.1708	2.29	0.1335	2.28	0.1325

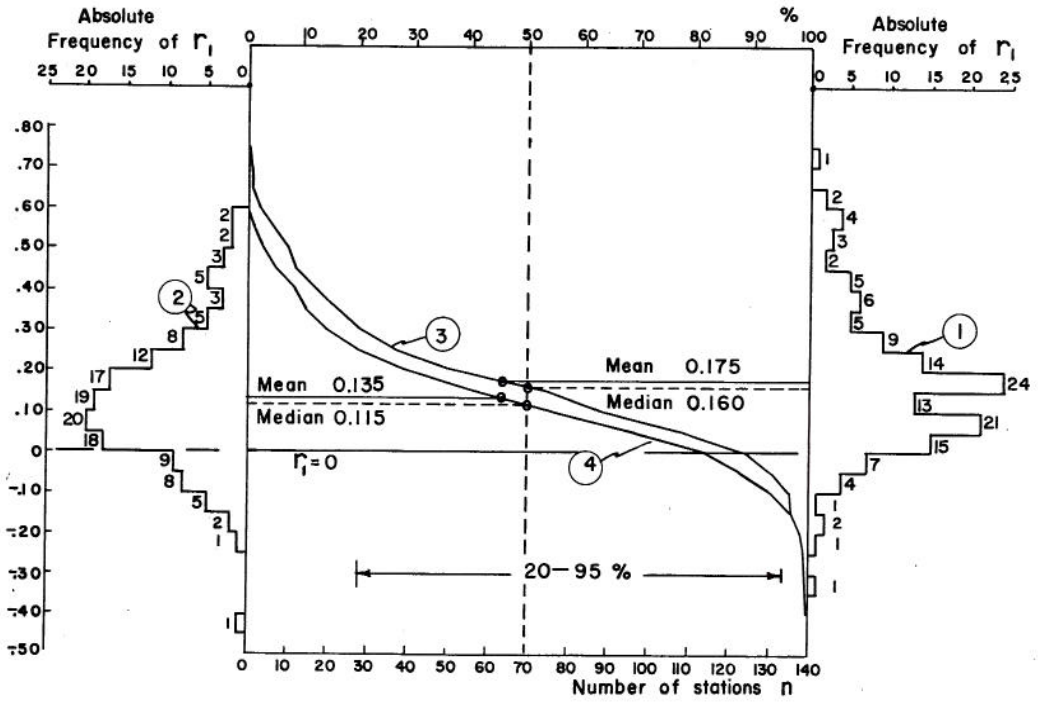

Fig. 2 Cumulative distributions and frequency histograms of the first serial correlation coefficient for the series of the first large sample of river gaging stations (n = 140): (1) frequency histogram of the V-series; (2) frequency histogram of the P_e-series; (3) cumulative distribution of the V-series; and (4) cumulative distribution of the P_e-series.

the mean value $\bar{\rho}_1$, estimated by the simple average $\bar{r}_1$. As $V = P_e + \Delta W$, and since P_e and ΔW are not mutually independent, it is not simple to determine how much serial correlation in V-series is affected by each of these two variables. Differences in r_1 are given here for V- and P_e-series:

$$\bar{r}_1(V) - \bar{r}_1(P_e) = 0.1748 - 0.1354 = 0.0395 \approx 0.040$$

The weighted means $\bar{r}_1{}^*(V)$ and $\bar{r}_1{}^*(P_e)$ are a little different from the simple average values; in this case

$$\bar{r}_1{}^*(V) - \bar{r}_1{}^*(P_e) = 0.1844 - 0.1336 = 0.0508 \approx 0.051 .$$

It can be concluded that a substantial part of the positive correlation in V-series may be attributed to the changes ΔW in water carryover from year to year in river basins in the form of stored water which is bound to flow out through the river in subsequent years.

Distributions in Figs. 2 and 3 and comparisons of their parameters show:

(1) that the means of r_1 for V- and P_e-series depart from those of normal independent variables;

(2) that distributions of r_1 in cartesian-probability scales for the range 20% - 95% for both V- and P_e-series may be approximated by straight lines of the same slope as the distributions of r_1 for the corresponding normal independent variable of equivalent sample lengths;

(3) that distributions of r_1 for V- and P_e-series are skewed; this should be expected for r_1-distributions if the population values of ρ_1 were different from zero in the case of normal dependent variables; and

(4) that some general conclusions about patterns in sequence of annual flow and annual effective precipitation may be derived from this large sample of river flow records.

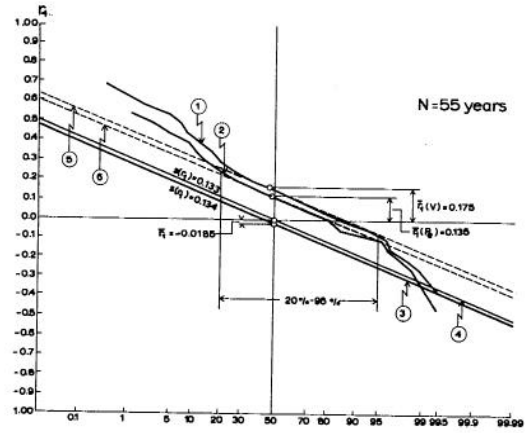

Fig. 3 Cumulative distributions of the first serial correlation coefficient for the series of the first large sample of river gaging stations (n = 140) in cartesian-probability scales: (1) V-series; (2) P_e-series; (3) normal independent variables, with the mean $\bar{\rho}_1$ estimated by eq. 2.3 and 2.23; and the variance estimated by eq. 2.27; (4) normal independent variables, with the mean $\bar{\rho}_1$ estimated by eq. 2.2, and the variance estimated by eq. 2.27; (5) fitted straight line which is parallel to lines (3) and (4) and passes through the mean $\bar{r}_1(V)$; (6) the same as under (5) but for P_e-series and $r_1(P_e)$.

Figure 4 shows r_1 (V) and r_1 (P_e) as Fisher's z-transforms plotted in cartesian-probability scales. It still shows like Fig. 3, that the extremes of r_1-distributions for both V- and P_e-series depart from normal independent variables with $\bar{z}$ estimated by eq. 2.15, ρ estimated by $\bar{r}_1$, and var z by eq. 2.18. The slope of the straight lines in Fig. 4 is $s_r = 1/\sqrt{N_m} = 0.135$ for $N_m = 55$.

When the first serial correlation coefficient is estimated by eq. 1.5 or approximately by eq. 1.10, the value of $\bar{\rho}_1$ for a normal independent variable may be estimated by eq. 2.23 as $\bar{r}_1 = -0.0185$ for the mean $N_m = 55$, or $\bar{r}_1{}^* = -0.0194$ for the weighted N_j. The mean of the f.s.c.c. $\bar{r}_1(V)$ or $\bar{r}_1{}^*(V)$ for annual flow (V-series) departs from the corres-

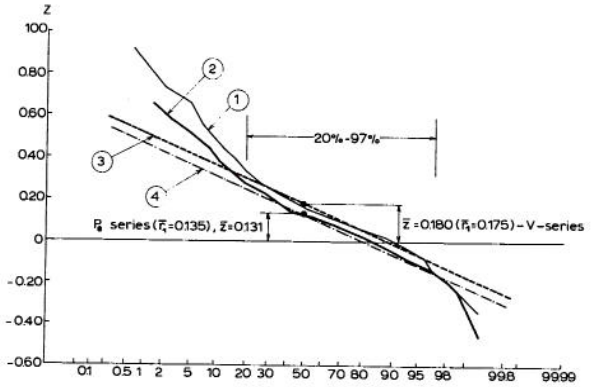

Fig. 4 Cumulative distributions of Fisher's z-transform of the first serial correlation coefficient for the series of the first large sample of river gaging stations (n = 140) in cartesian-probability scales: (1) z-transform of r_1 for V-series; (2) z-transform of r_1 for P_e-series; (3) straight line fit to z-transform of V-series; and (4) straight line fit to z-transform of P_e-series.

ponding expected values of the normal independent variables. The difference is $\Delta\bar{r}_1(V) = \bar{r}_1(V) - \bar{r}_1 = 0.1748 + 0.0185 = 0.1933 \approx 0.20$ for simple means $\bar{r}_1(V)$ and $\bar{r}_1$. It is $\Delta\bar{r}_1{}^* = \bar{r}_1(V) - \bar{r}_1{}^* = +0.1844 + 0.0194 = 0.2038 \approx 0.20$ for weighted means $\bar{r}_1{}^*(V)$ and $\bar{r}_1{}^*$. Similarly, the two differences for P_e-series are $\bar{\Delta}r_1(P_e) = \bar{r}_1(P_e) - \bar{r}_1 = 0.1353 + 0.0185 = 0.1538 \approx 0.154$ and $\Delta\bar{r}_1{}^*(P_e) = \bar{r}_1{}^*(P_e) - \bar{r}_1{}^* = 0.1336 + 0.0194 = 0.153$, with $\Delta\bar{r}_1{}^*$ being the difference of $\bar{r}_1{}^*$ values computed by using the weighted N_j.

The first large sample of river flow records of 140 stations from several parts of the world shows that on the average the annual river flows are serially correlated with the mean value of the f. s. c. c. approximately given by $\rho_1 = 0.20$. The annual effective precipitation (precipitation minus evapotranspiration on a river basin in a year or the net annual water yield of the atmosphere to the river basin surface) has an approximate mean value given by $\rho_1 = 0.15$. The water carryover in river basins in the form of surface and underground storage which will flow out of the river basin as surface runoff in subsequent years is responsible for the first serial correlation coefficient of annual flow being greater than the f. s. c. c. of annual effective precipitation. Since the change in water carryover from year to year is determined in this study in an approximate way [1], the above average ρ_1-values should be considered also as approximate values.

Before statistical inference is used to test whether or not the means $\bar{r}_1(V)$ and $\bar{r}_1(P_e)$ or $\bar{r}_1{}^*(V)$ and $\bar{r}_1{}^*(P_e)$ are significantly different from the corresponding values of normal independent variables, the interstation correlation is discussed here in general terms.

The positive interstation correlation of concurrent values of V- or P_e-series makes the effective size, n_e, of the first large sample of river flow records smaller than will be the case if these variables were free of interstation correlation. The effective sample size, n_e, represents that number of stations, a smaller number than 140, which would contain the same information as the 140 stations which have significant interstation correlation. The effective sample, n_e, might be said to be independent of the interstation correlation. The effect of interstation correlation is studied quantitatively in the analysis of the second large sample of river flow records for stations in Western North America and the first large sample of precipitation stations (same continental sampling area).

Assuming that the mean interstation correlation coefficient of r_1 is not significantly different from zero, or that $n_e = 140$, the statistical inference can be carried out for the hypothesis that the mean values $\bar{r}_1(V)$ and $\bar{r}_1(P_e)$, or $\bar{r}_1{}^*(V)$ and $\bar{r}_1{}^*(P_e)$ are significantly different either from zero or from $\bar{r}_1 = -0.020$ of normal independent variables; in other words, that the differences $\bar{\Delta}r_1(V) = 0.20$ and $\Delta\bar{r}_1(P_e) = 0.15$ are significantly different from -0.020 at the 95% level of significance.

The expected value of $\bar{\rho}_1$ for normal independent variables is $\bar{r}_1 = -0.0182$ for $N_m = 55$. The variance of r_1 is $\operatorname{var} r_1 = 0.0178$. For the 95% significance level each tail having the total probability $\epsilon = 0.025$, the confidence limits about $\bar{r}_1$ are

$$\bar{r}_1 \pm t\sqrt{\frac{\operatorname{var} r_1}{n-1}},$$

with $t = 1.96$ from tables of normal function for $\epsilon = 0.025$; and $n = 140$, number of stations. This gives two limits, 0.004 and -0.040. The values $\bar{r}_1(V) = 0.175$ and $\bar{r}_1(P_e) = 0.135$ are much greater than the positive confidence limit 0.004. For the expected value of $\bar{\rho}_1$ being zero, then the limits are $+0.022$ and -0.022, and still $\bar{r}_1(V)$ and $\bar{r}_1(P_e)$ are much greater than the positive limit 0.022.

An opposite approach is also investigated here. It is assumed that $\bar{r}_1$ is barely significant. This means that $\bar{r}_1$ corresponds to the positive confidence limit at the 95% level of the normal independent variables. Since $\bar{r}_1 = 0$, $t = 1.96$, $\operatorname{var} r_1 = 0.0178$, the positive limits are 0.150 for P_e-series and +0.200 for V-series. Equation 2.74 gives the effective sample sizes: $n_e = 3.5$ for P_e-series, and $n = 2.9$ for V-series. The fact is that the river gaging stations of the first large sample are greatly dispersed around the world. It is unlikely that the average interstation correlation of r_1 values is large enough to produce the small effective sample sizes of 2.9 and 3.5.

By using eq. 2.74 the average interstation correlation coefficient of r_1 is $\bar{r} = (n-n_e)/n_e(n-1)$. For $n = 140$ and $n_e = 2.9$ and 3.5 respectively, for V- and P_e-series, the corresponding $\bar{r}$ values are 0.34 and 0.28. It is shown later that for Western North America those two values are much smaller than 0.34 or 0.28. The effective sample sizes from $\bar{r}_1$ for V- and P_e-series must be greater than 2.9 or 3.5, respectively, because the sampling of river gaging stations on a global basis should produce a smaller interstation correlation than the sampling on a continental basis.

Therefore, the conclusion is that the average first serial correlation coefficients for V-series and P_e-series are significantly different from zero (or from -0.020) at the 95% level of significance test. However, the absolute measures of serial correlation with $\bar{\rho}_1(V) = 0.175$ and $\bar{\rho}_1(P_e) = 0.135$ may be considered as relatively small, though they are not negligible in many problems, especially in the determination of overyear reservoir storage capacities for regulating river flows.

It should be pointed out also, that the first large sample of river flow records includes many rivers with unusually large natural storage capacities (the St. Lawrence River with Great Lakes, USA; the Göta River in Sweden; the Neva River in U.S.S.R.; outflow of Victoria Lake in Africa; outflow of Albert Lake in Africa; etc.). This fact indicates that samples of stations from river basins with relatively small water storage capacities would tend to have smaller average first serial correlation coefficients.

Great departures in $r_1(V)$ and $r_1(P_e)$ distributions or z(V) and $z(P_e)$-transform distributions at their tails from the normal function, especially the high positive r_1-values or z-values, may be partly explained by a biased selection of river gaging stations and a stress on stations with large upstream surface storage (mostly large lakes). The rivers with regulated flow have always attracted economic development. Hydrologic services usually gave preference to flow measurements on such rivers in many countries at an earlier period of streamflow gaging. This bias in the first large sample will be more evident through the analysis of correlograms for V- and P_e-series for several rivers with long flow records.

3. Frequency distributions of other serial correlation coefficients. Distributions of only four other r_k coefficients, namely r_2, r_3, r_4, and r_5, are presented and discussed here for both V-series and P_e-series. They are computed for open time series by eq. 1.10. Figure 5 shows these four distributions in cartesian-probability scales.

Table 2 gives the estimated values of $\bar{\rho}_2$ through $\bar{\rho}_{11}$ by eq. 2.19, or by average values $\bar{r}_2$ through $\bar{r}_{11}$. The equations used for estimation of a particular statistic are also given in Table 2. For $\bar{r}_2$, $\bar{r}_3$, $\bar{r}_4$, and $\bar{r}_5$, the weighted estimates of means

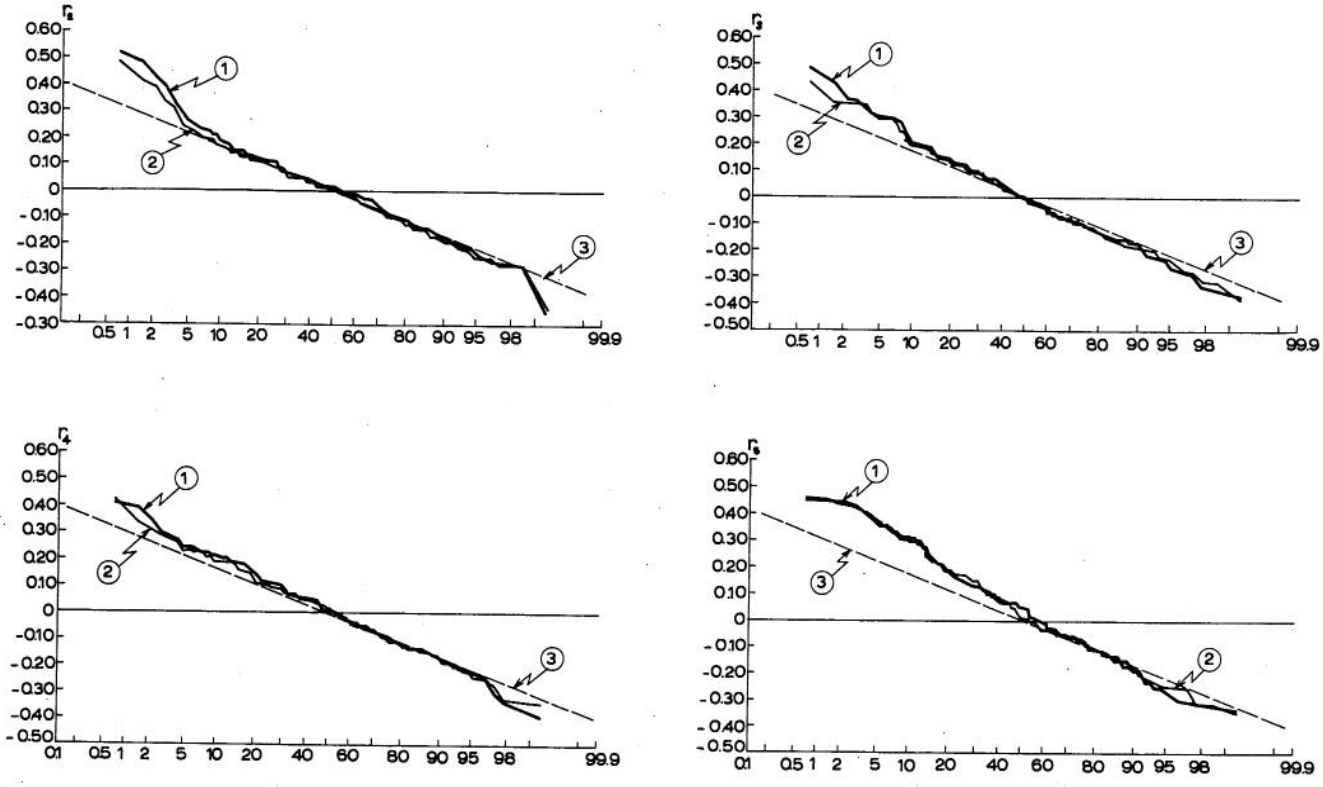

Fig. 5 Cumulative distributions of the serial correlation coefficients of r_2, r_3, r_4, and r_5 for the series of the first large sample of river gaging stations (n=140) in cartesian-probability scales: (1) V-series; (2) P_e-series; and (3) normal independent variables with $\bar{\rho}_1$ zero and the variance estimated by eq. 2.27.

are given, as determined by eq. 2.20. Both the simple and weighted estimates of variances for r_2, r_3, r_4, and r_5 are computed by eq. 2.21 and by eq. 2.22, respectively. For the normal independent variable and its r_2, r_3, r_4, and r_5 distributions as given in Fig. 5, the following statistics are used: $\bar{r}_k = 0$, and $\text{var } r_k = \frac{1}{N_m + 2} = 0.0175$, with $N_m = 55$, and with the standard deviation $s(r_k) = 0.132$.

Figure 5 and Table 2 show that the departures of distributions between V- and P_e-series on one side, and normal independent variables, on the other side, for r_2 through r_{11}, are relatively small. Taking the expected value of normal independent variables, $\bar{r}_k = -0.0185$, then the differences $\bar{\Delta r}_k$ of estimated $\bar{r}_k$- values for V-series and P_e-series, and $\bar{r}_k$ for the normal independent variables rounded to three decimal places are:

	V-Series	P_e-Series
$\bar{\Delta r}_2$	0.033	0.022
$\bar{\Delta r}_3$	0.028	0.026
$\bar{\Delta r}_4$	0.033	0.030
$\bar{\Delta r}_5$	0.064	0.064
$\bar{\Delta r}_6$	0.022	0.023
$\bar{\Delta r}_7$	-0.007	-0.008
$\bar{\Delta r}_8$	-0.002	-0.002
$\bar{\Delta r}_9$	-0.022	-0.018
$\bar{\Delta r}_{10}$	-0.006	-0.006
$\bar{\Delta r}_{11}$	0.016	0.017

The greatest difference is for $\bar{r}_5$, and it is 0.064. Distributions of r_2, r_3, r_4, and r_5 are close to those of normal independent variables, except that there is a tendency for the standard deviation of distributions

TABLE 2

Statistics of frequency distributions of serial correlation coefficients, r_2, r_3, r_4, and r_5 for 140 stations of annual flow (V-series) and annual effective precipitation (P_e - series), as well as simple estimates of mean for r_6, r_7, r_8, r_9, r_{10}, and r_{11}.

r_k	Statistics estimated	V - Series				P_e - Series			
		Equation Used	Value	Equation Used	Value	Equation Used	Value	Equation Used	Value
r_2	Mean	2.19	0.0146	2.20	0.0178	2.19	0.0036	2.20	0.0004
	Variance	2.21	0.0228	2.22	0.0231	2.21	0.0210	2.22	0.0204
r_3	Mean	2.19	0.0097	2.20	0.0101	2.19	0.0070	2.20	0.0029
	Variance	2.21	0.0253	2.22	0.0259	2.21	0.0248	2.22	0.0233
r_4	Mean	2.19	0.0150	2.20	0.0186	2.19	0.0110	2.20	0.0164
	Variance	2.21	0.0224	2.22	0.0214	2.21	0.0204	2.22	0.0189
r_5	Mean	2.19	0.0460	2.20	0.0406	2.19	0.0450	2.20	0.0357
	Variance	2.21	0.0323	2.22	0.0306	2.21	0.0313	2.22	0.0295
r_6	Mean	2.19	0.0039			2.19	0.0041		
r_7	Mean	2.19	-0.0251			2.19	-0.0269		
r_8	Mean	2.19	-0.0207			2.19	-0.0203		
r_9	Mean	2.19	-0.0407			2.19	-0.0370		
r_{10}	Mean	2.19	-0.0246			2.19	-0.0259		
r_{11}	Mean	2.19	-0.0027			2.19	-0.0014		

of V- and P_e-series to increase with an increase of k, while that of normal independent variables is considered to be constant for a given N and to be independent of k. This departure of standard deviations is greatest for r_5, similar to $\bar{r}_5$, and it may be considered as the sampling fluctuation of an independent time series with several r_k values.

There is no indication that $\bar{r}_{11}$ for k = 11, or the approximate time lag of average sun-spot peaks is significantly different from zero.

4. Average values of serial correlation coefficients. Figure 6 gives the average values of serial correlation coefficients or mean values of r_k as they change with k, from k = 1 to k = 11, for both V-series and P_e-series. It is quite clear that $\bar{r}_1$ values are significantly different from all other values. The other conclusion is that the difference, $\bar{r}_k(V) - \bar{r}_k(P_e)$, decreases with an increase of k, and this is evident especially from k = 1 to k = 5. After k = 5, the difference between the two series is small.

The estimate of $\bar{\rho}_k$ for N_m = 55 is also plotted in Fig. 6, with $E\bar{\rho}_k = \bar{r}_k = -0.0185$. The variance of r_k is 0.0178. By using eq. 2.30 and n = 140, the variance of $\bar{r}_k$ is 0.000125. This gives the standard deviation of $\bar{r}_k$ as $s(\bar{r}_k) = 0.0112$. The confidence limits at 95% level for $\bar{r}_k$ of normal independent variables are $\bar{r}_k = 0.0034$ and $\bar{r}_k = -0.0404$. Assuming that r_k values are not intercorrelated for stations taken pairwise, or $n = n_e = 140$ for this sample, all first five values, $\bar{r}_1$ through $\bar{r}_5$, for V-series are significantly different from $\bar{r}_k$ of normal independent variables at the 95% level of significance, although $\bar{r}_2$, $\bar{r}_3$, $\bar{r}_4$ are not very far from the positive confidence limit. For P_e-series $\bar{r}_2$ and $\bar{r}_3$ may be considered as not significantly different at the 95% level from the $\bar{r}_k$-value of normal independent variables. Values of $\bar{r}_1(V)$ and $\bar{r}_1(P_e)$ are evidently significantly different from the corresponding value of -0.0185 for normal independent variables. Both $\bar{r}_5(V)$ and $\bar{r}_5(P_e)$ are above the positive confidence limit at the 95% level, but much below $\bar{r}_1(V)$ and $\bar{r}_1(P_e)$. The values $\bar{r}_5(V)$ and $\bar{r}_5(P_e)$ may be considered as sampling deviation because 5% of $\bar{r}_k$-values should be on the average outside the confidence limits.

Assuming that there is a positive mean interstation correlation between r_k-values, or as an

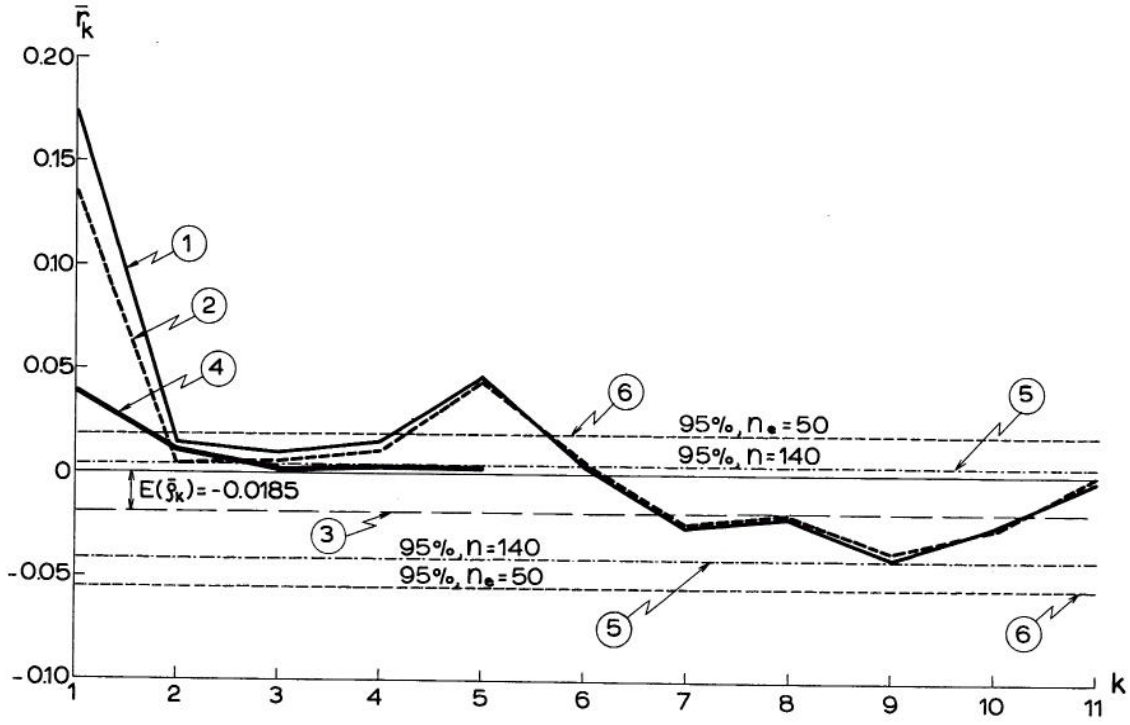

Fig. 6 Average values of the serial correlation coefficients ($\bar{r}_1$ through $\bar{r}_{11}$) for the series of the first large sample of river gaging stations versus the lag k: (1) V-series; (2) P_e-series; (3) expected value of $\bar{\rho}_k$ estimated by eqs. 2.3 and 2.23; (4) difference $\Delta\bar{r}_k = \bar{r}_k(V) - \bar{r}_k(P_e)$; (5) confidence limits at the 95 percent level for normal independent variables; and (6) confidence limits for normal independent variables with n_e = 50 (assumed effective sample size).

approximation, a positive correlation of annual flow effective precipitation among pairs of 140 stations, and assuming that the effective sample size $n_e = 50$, the confidence limits at the 95% level of $\bar{r}_k$ would be $\bar{r}_k = 0.0181$; and $\bar{r}_k = -0.551$. In this case only $\bar{r}_1(V)$, $\bar{r}_1(P_e)$, $\bar{r}_5(V)$, and $\bar{r}_5(P_e)$ are outside the confidence limits. It is concluded here, as previously, that $\bar{r}_1(V)$ and $\bar{r}_1(P_e)$ are significantly different from the corresponding statistics of normal independent variables.

The water carryover for some river basins is greater than one year. Therefore, the difference, $\bar{r}_2(V) - \bar{r}_2(P_e)$, may also be considered as significantly different from zero. The large values of $\bar{r}_5(V)$ and $\bar{r}_5(P_e)$ may be considered as sampling departures from the expected value of normal independent variables. This may be so even though the probability of their exceeding a positive confidence limit is 10% (one in ten values, $\bar{r}_2$ through $\bar{r}_{11}$) instead 5% as the confidence level implies.

The sequence of $\bar{r}_k$ in Fig. 6, and especially the differnce, $\Delta\bar{r}_k = \bar{r}_k(V) - \bar{r}_k(P_e)$, which is also plotted in Fig. 6, leads to the conclusion that on the average the linear autoregressive schemes may be applied to the relationship of V- and P_e- variables. It is shown later that the first order or the second order linear autoregressive schemes for the patterns in sequence of V-series may well fit some typical cases.

5. Correlograms of individual rivers. Figures 7, 8, and 9 give correlograms of individual river stations of both the annual flow and the annual effective precipitation. Usually the upper graph (Fig. 7) or the left graph (Fig. 8) gives the correlogram for the V-series, and the lower graph (Fig. 7) or the right graph (Fig. 8) gives the correlogram for the P_e-series. The confidence limits for normal independent variables at the 95% level of significance, estimated by eq. 2.10, are also given in both graphs. As each correlogram refers to a series with N usually different from N_m used in the graph, the confidence limits for a given N_m should be considered here only as indications of correlogram fluctuations for normal independent variables.

In using eq. 2.10 the value of N in that equation has been replaced by N-k or by the number of correlated pairs in the computation of r_k-values. Thus the confidence limits expand slowly as the lag k increases. Confidence limits at the 95% level for normal independent variables are given for two values of sample size (in Fig. 7): N = 150 (the Göta, the Rhine), and N = 120 (the Danube). The expected means of r_k for normal independent variables as given by eq. 2.7 are not plotted in Fig. 7.

Figure 7 gives correlograms of V-series and P_e-series for four rivers with longest records: the Rhine River at Basle, Switzerland (150 years); the Göta River at Sjörtop-Vänersburg, Sweden (150 years); the Nemunas River at Smalininkai, Lithuania, USSR, (132 years); and the Danube River at Orshava, Romania (120 years). The number of computed r_k-values

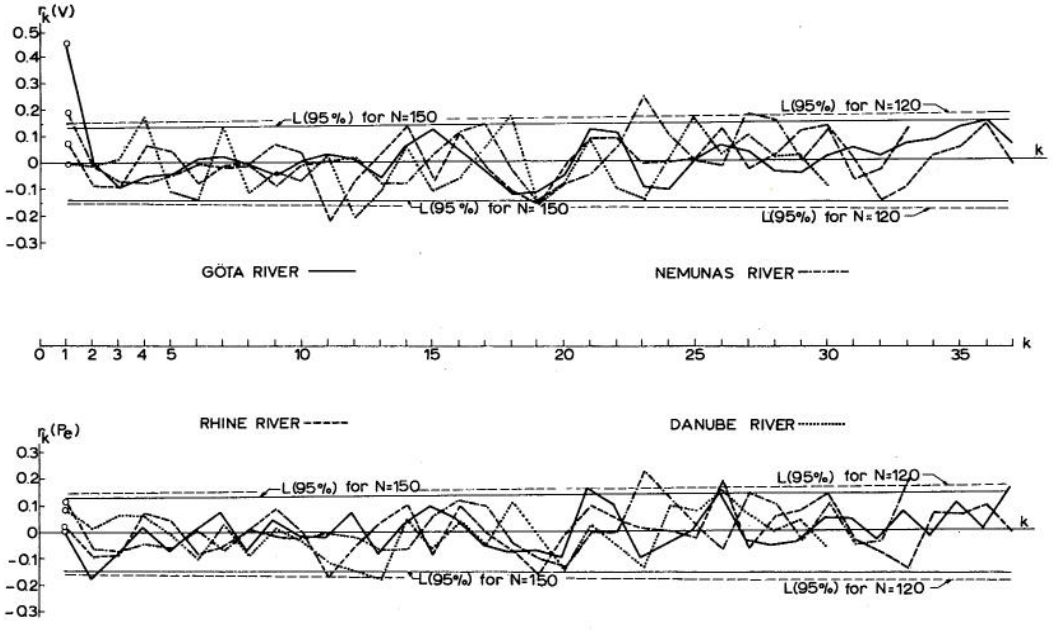

Fig. 7 Correlograms of the four rivers with longest records: the Göta River (N = 150), the Nemunas River (N = 132), the Rhine River (N = 150), and the Danube River (N = 120). The upper graph refers to the V-series, and the lower graph to the P_e-series. Two confidence limits at the 95 percent level are given for normal independent variables for two lengths: N = 150 (max) and N = 120 (min).

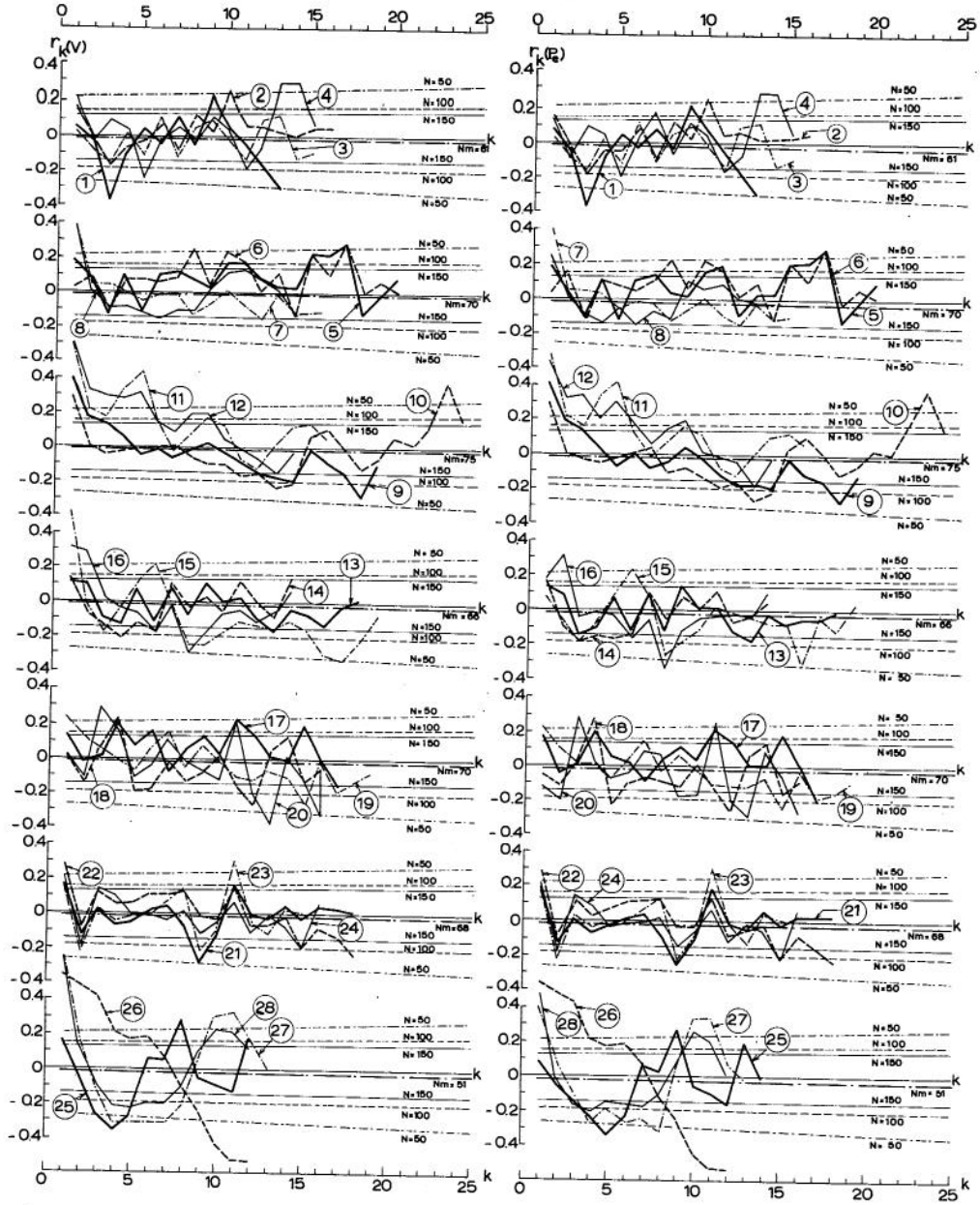

Fig. 8 Correlograms of 28 individual series from the first large sample of river gaging stations (n = 140) in groups of four series are shown. The left graphs refer to the V-series, and the right graphs refer to the P_e-series. For each of the 7 groups, three confidence limits at the 95 percent level for normal independent variables are given for the following lengths of series; N = 150, N = 100, and N = 50. The expected value of $\overline{r}_k$ is given for the average length N_m of the series for each group of four correlograms, as computed by eqs. 2.3 and 2.23. The groups are from these regions: (I) USA; (II) USA; (III) USA; (IV) Europe; (V) Europe; (VI) Australia; and (VIII) Africa. Following are listed the 28 stations, their period of observation, and the length of their series.

(1) River Piscataquis near Dover-Foxcroft, Me., 1902-1956 (54); (2) River Susquehanna, Harrisburg, Pa., 1890-1957 (67); (3) Potomac River, near Point of Rocks, Md., 1896-1957 (61); (4) French Broad River near Asheville, N. C., 1895-1957 (62) (5) Tennessee River near Chattanooga, Tenn., 1874-1956 (82); (6) Kanawha River, near Kanawha Fall, W. Va., 1877-1957 (80); (7) Wolf River near New London, Wisc., 1896-1957 (61); (8) Fox River, near Berlin, Wisc., 1898-1957 (59); (9) Mississippi River, near Keokuk, Iowa, 1878-1957 (79); (10) Mississippi River, near St. Louis, Mo., 1861-1957 (96); (11) Missouri River, near Fort Benton, Mont., 1890-1955 (65); (12) Missouri River, near Sioux City, Iowa, 1897-1955 (58); (13) Dnieper River, at Dnieperpetrovsk, USSR, 1881-1955 (74); (14) Volga River at Gorkii, USSR, 1877-1935 (58); (15) Neva River, at Petrokrepost, USSR, 1859-1935 (76); (16) Kama River, at Berczniki, USSR, 1881-1938 (57); (17) Thames River, near Teddington, England, 1883-1954 (71); (18) Danube River, at Vienna-Mussdorf, Austria, 1893-1957 (64); (19) Indals River, near Ostersund, Sweden, 1893-1957 (64); (20) Riul Mures River, at Arad, Romania, 1876-1955 (79); (21) Goulburn River, near Murchison, Victoria, Australia, 1881-1954 (73); (22) Kiewa River, at Kiewa, Victoria, Australia, 1885-1957 (72); (23) Owens River at Wangaratta, Victoria, Australia, 1886-1948 (62); (24) River Murray at Jingellic, Victoria, Australia, 1890-1957 (67); (25) Nile (Second Half), at Aswan Dam, Egypt, Africa, 1903-1955 (52); (26) Niger River at Koulicoro, Africa, 1906-1957 (51); (27) Outflow from Lake Victoria, Nile Basin, Africa, 1898-1952 (54); (28) Lake Albert at Mongalla, Africa, 1904-1952 (48).

was N/4 so that for these four rivers there are altogether 137 r_k-values.

By the definition of 95% confidence interval, seven r_k-values (or 5%) for the correlograms of four rivers should be outside the confidence limits, if the four series would be serially uncorrelated. In the case of the P_e-series there are nine values of r_k which are evidently outside the confidence interval for N = 150, and five values for N = 120. The average is seven, or 5.2%, which approximately satisfies this condition of 5%. Neglecting the r_1-values for the Göta River and the Nemunas River, for V-series there are also 9 values of r_k which are evidently outside the confidence limits at the 95% level for N = 150, and five values for N = 120, or the average of seven values. This gives 5.2% of values outside the confidence limits, which also satisfies the condition of 5%.

The r_1-value of the Göta River for V-series may well be explained by the large lake storage of that river resulting in water carryover from year to year. The significant difference between $r_1(V)$ and $r_1(P_e)$ can logically be attributed to water carryover. The r_1-value of the Nemunas River for the V-series eventually may be explained by some inconsistency in the data as well as by water carryover. The fact is that the flow rating curve obtained in the second half of the nineteenth century has been applied to stage observations of previous years to derive the river flows. It might be that the natural fluctuations of river bed elevation produced some inconsistency in data, when the above method of obtaining the river flow was applied. The difference between r_1 values of V- and P_e-series for the Nemunas River is also significant, so that water carryover may also be responsible partly for the significantly positive r_1 value of the V-series.

It may be concluded from the results given in Fig. 7 that the four rivers for both V- and P_e-series have sequences which do not depart significantly from independent time series except for r_1 values in two cases. This fact supports the hypothesis that a linear moving average scheme may well be fitted to describe mathematically the dependence of annual flow to annual effective precipitation for these rivers with long records. The first order autoregressive scheme for the time dependence of V-series with $r_k = r_1^k$ seems less appropriate for the Göta and Nemunas River in describing the relationship of V- and P_e-variables.

For the Göta River and its V-series the actual values of s.c.c. are: $r_1 = 0.461$; $r_2 = -0.005$; and $r_3 = -0.095$. Using the above relationship $r_k = r_1^k$ and taking $r_1 = 0.461$, the other two values should be $r_2 = 0.213$; and $r_3 = 0.098$. The difference between these two values and the actual values are $\Delta r_2 = 0.213 + 0.005 = 0.218$; and $\Delta r_3 = 0.098 + 0.095 = 0.193$. The average is about 0.205, and they are not likely to be explained only by sampling fluctuations in r_2 and r_3 values. The second order linear autoregressive scheme is fitted to the patterns in sequence of annual flow of the Göta River, because the actual r_1 is positive, while the actual value r_2 is negative. Using eq. 2.56 and the computed values r_1 and r_2, the estimates of a_1 and a_2 for the Göta River are $a_1 = 0.586$; and $a_2 = -0.277$. The condition that $a_2 \leq -a_1^2/4$ is satisfied because $a_2 = -0.277 < -a_1^2/4 = -0.086$.

The angle θ in eq. 2.53 is given by the expression $2 \sin \theta = \sqrt{4 + a_1^2/a_2}$, and in this case $\theta = 56^0$, so that the positive values of β_k are only for $k \leq (\frac{\pi}{\theta} - 1)$, or $k \leq 2.22$. For this river there are only three positive values of β_k: namely, β_0, β_1, and β_2. Using eq. 2.53, $\beta_1 = 0.518\beta_0$; and $\beta_2 = 0.0535\beta_0$. Therefore, the main carryover is in the first year (k = 1). In the second year (k = 2) the carryover is very small. In other words, the b_j-coefficients are positive and significantly different from zero only for j = 0, 1, and 2.

The time dependence model for this river, therefore, is

$$x_i = 0.586\, x_{i-1} - 0.277\, x_{i-2} + \epsilon_i \qquad 3.1$$

As $\epsilon_i = \beta_o y_i$, and as $Ey_i = 0$, and $Ey_i^2 = 1$, the estimate of β_o^2 by eq. 2.57 is $\beta_o^2 = 0.728$. In the above mathematical model ϵ_i is a normal independent variable with mean zero and variance 0.728.

As $x_i = \frac{K_i - 1}{s}$, with $\bar{K} = 1$ and standard deviation of annual flows given in modular coefficients being equal to C_v, then for $s = C_v = 0.182$ [1, Appendix 1], the following model in modular coefficients is applicable to the Göta River:

$$K_i = 0.586\, K_{i-1} - 0.277\, K_{i-2} + 0.182\, \epsilon_i + 0.691 \qquad 3.2$$

Replacing $0.182\, \epsilon_i + 0.691$ by η_i - variable, the expected mean of η_i is 0.691, and the variance is 0.1325, or

$$K_i = 0.586\, K_{i-1} - 0.277\, K_{i-2} + \eta_i . \qquad 3.3$$

In order to obtain absolute annual flow the K_i-value should be multiplied by the mean $\bar{V}$ in m^3/sec. or

in cfs. If a normal standard variable t is used,

$$K_i = 0.586\, K_{i-1} - 0.277\, K_{i-2} + 0.1325\, t + 0.691 \qquad 3.4$$

A probability statement can be made about the annual flow of the next year, if two limits for t have been selected. By using t_1 and t_2 from tables of normal function for a given probability that a value t falls between t_1 and t_2, the limits K_1 and K_2 may be computed from eq. 3.4 and the corresponding values t_1 and t_2.

Using eq. 2.55 and $a_1 = 0.586$, $a_2 = -0.277$, $r_1 = 0.461$ and $r_2 = -0.005$ for the Göta River, the values r_1 through r_9 are computed and given in Table 3. Also the computed values of r_1 through r_9 for the observed series of 150 years are given, as well as their difference, Δr_k. Table 3 shows that the differences for r_3 through r_9 are relatively small, and that r_3, r_4, and r_5 have the same negative sign in both the computed r_k values by eq. 2.55 and the computed r_k values from the actual series. Therefore, the second order linear autoregressive scheme gives a good fit for the patterns in sequence of annual flow of the Göta River.

TABLE 3

Serial correlation coefficient r_k for the annual flows of the Göta River, computed by the second order linear autoregressive scheme of eq. 3.1 and by using eq. 2.55, the computed r_k-values for the actual series, and their differences.

k	Computed r_k by eqs. 2.55 and 3.1.	Computed r_k from actual series	Difference
1	0.461	0.461	0.000
2	-0.005	-0.005	0.000
3	-0.133	-0.095	-0.038
4	-0.077	-0.057	-0.020
5	-0.009	-0.048	0.039
6	0.016	-0.010	0.026
7	0.012	-0.018	0.030
8	0.002	-0.016	0.018
9	-0.002	-0.055	0.053

For the Nemunas River and its V-series the computed values of s.c.c. are: $r_1 = 0.185$; $r_2 = -0.015$ and $r_3 = -0.077$. Using eq. 2.50 and $r_1 = 0.185$, the other two values are $r_2 = 0.034$ and $r_3 = 0.006$. The differences are $\Delta r_2 = 0.034 + 0.015 = 0.049$ and $\Delta r_3 = 0.006 + 0.077 = 0.083$. They are not significant either as large or small values, so that it would be difficult to conclude which of the two models, eq. 2.39 or eq. 2.52 would better fit the patterns in sequence of annual flows. The fact that the computed values or r_2 and r_3 are negative, while eq. 2.39 should produce positive values, gives some advantage to the second order linear autoregressive scheme.

The values of r_1, r_2, and r_3 for the V-series of the Rhine River and the Danube River are small. No conclusion can be made about the best mathematical model which would fit the patterns in sequence, if the series have a relatively small time dependence.

Figure 8 shows correlograms for 28 river gaging stations: 12 stations (No. 1-12) in first three groups of four stations for U.S.A., 8 stations (No. 13-20) in fourth and fifth group of four stations for Europe, 4 stations (No. 21-24) in Australia (Victoria State) in sixth group of four stations, and 4 stations (No. 25-28) for Africa in seventh group of four stations. The left graphs give correlograms of V-series and the right graphs those of P_e-series. The stations for U.S.A. are mostly in the eastern part of the country. The analysis of the second large sample of river gaging stations will deal with correlograms of selected stations in the western part of the country.

Figure 8 has three confidence intervals at the 95% level for normal independent variables and for three sample sizes: N = 50, N = 100, and N = 150. The expected value $\overline{r}_k$ of normal independent variables is given for each group of stations for the average sample size N_m of groups of four stations. Equations 2.10 and 2.23 are used for the computation of confidence limits and expected $\overline{r}_k$, respectively, with N replaced by N-k and N_m replaced by N_m-k.

The correlograms of the 12 stations in the U.S.A. (1 through 12 in Fig. 8) usually show a greater fluctuation of r_k-values around the expected values at their end. This is due to increased sampling variance with a decrease of N-k as k increases. In general, most correlograms are confined well inside the confidence interval at the 95% level for the corresponding sample size and for normal independent variables, except for r_1-values.

The correlograms of the 8 stations in Europe (13 through 20 in Fig. 8) also are well inside

the corresponding confidence intervals at the 95% level for normal independent variables. The correlograms of the 4 stations in Australia (21 through 24 in Fig. 8) are well within the corresponding confidence intervals with the same characteristics as the previous groups. However, correlograms of the 4 stations in Africa (25 through 28 in Fig. 8) show some departures from the expected correlograms of normal independent variables. Lake Victoria and Lake Albert stations show fluctuations of the correlograms close to the confidence limits while the Nile River at Aswan Dam has a relatively important fluctuation also very close to confidence limits. The characteristics of these correlograms indicate that a second order linear autoregressive scheme may be fitted to the patterns in sequence of annual flow and annual effective precipitation for these three gaging stations.

The Niger River has the most unexpected correlogram in comparison with that of normal independent variables. Its correlogram decreases continuously for the V-series from a high value $r_1 =$ $= 0.555$ to a low value $r_1 = -0.533$. The same holds for the P_e-series. A logical question is whether this type of correlogram is a product of regional sampling or not. Has it been produced by pure chance in the past and is not expected to be produced in the future? Or is it related to some specific climatic conditions in Central Africa and will be produced again in similar patterns in the future?

This problem of the Niger River is important, but only a careful study of all available data (and also of data quality) in the river basin and around it would produce a reliable answer to the above questions. The long range precipitation stations in the region may offer the clue to whether the cyclic movement of the correlogram is a pattern or a chance product.

Although it seems that the first three stations (Lake Victoria outlet, Lake Albert outlet, and the Nile River) have a first damping oscillation of 11, 20, and 8 years respectively, and the Niger River at Koulicoro has half oscillation of 12 years (or full first oscillation of 24 years), this may not have a significant relationship with average time lag between sunspot peaks. The series length of $N = 51$ for the Niger River is too short to draw any reliable conclusion about a potential cycle of 24 years. In the case of correlograms of the first three stations the damping is rather fast. These types of correlograms may be produced by a special type of moving average scheme, especially, since the large lakes upstream from those three stations and the evaporation from them may create a particular type of moving average scheme.

The St. Lawrence River at Ogdensburg, (N. Y., U.S.A.) is treated as a particular case of water carryover because of the large effect of the Great Lakes. Observations from 1860 to 1957 (97 years) give correlograms for V- and P_e-series which are plotted in Fig. 9. While the correlogram of P_e-series is well within the confidence interval at the 95% level for normal independent variables, the correlogram of V-series has positive values for the first eleven r_k, and most of them (r_1 through r_9) are above the positive confidence limit. The shape of this correlogram points toward the conclusion that either the first order linear autoregressive scheme, eq. 2.39, or a moving average scheme of special type may be the most appropriate mathematical model for describing the patterns in sequence of annual flow. Figure 9 gives the first order linear autoregressive scheme $\rho_k = \rho_1^k$, with $\rho_1(V)$ estimated by $r_1(V) = 0.705$.

In order to see how well the first order linear autoregressive scheme $\rho_k = \rho^k$ fits the series of annual flow of St. Lawrence River, a least squares fit of this model is carried out. Minimizing the sum $\sum_1^k (\rho_1^k - r_k)^2$, for $k = 1, 2, \ldots, 17$ gives $\rho_1 = 0.785$. By giving a larger weight to the first values of r_k, a least squares fit to the log r_k is used, in which case the sum $\sum_1^k (k \log \rho_1 - \log r_k)^2$ is minimized with $\rho_1 = 0.767$.

Figure 10, upper graph, gives the observed correlogram, and the correlograms of $\rho_k = \rho_1^k$, where ρ_1 is estimated in three ways: (a) as observed value of $\rho_1 = r_1(V) = 0.705$; (b) by least square fit to r_k, and $\rho_1 = 0.785$; and (c) by least square fit to log r_k, and $\rho_1 = 0.767$. The lower graph of Fig. 10 gives the differences Δr_k between the correlogram of observed series and the first order linear autoregressive scheme in three cases: (a) $\rho_1 = 0.705$,

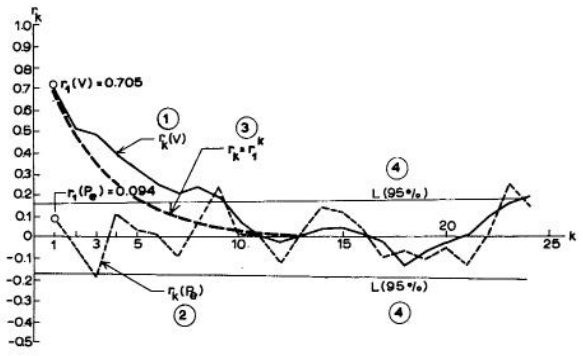

Fig. 9 Correlograms of the St. Lawrence River at Ogdensburg, New York: (1) V-series; (2) P_e-series; (3) first order linear autoregressive scheme $\rho_k = \rho_1^k$, with ρ_1 estimated by the actual value $r_1 = 0.705$; and (4) confidence limits at the 95 percent level for normal independent variables with N = 97.

(b) $\rho_1 = 0.785$, and (c) $\rho_1 = 0.767$. This last graph also has confidence limits at the 95% level about the expected mean of normal independent variables with $N = 97$, or $\bar{r}_1 = -0.0104$, and limits computed by eq. 2.10 of +0.156 and -0.177. Considering Δr_k, after the correlogram of the first order scheme has been deducted, as the correlogram of an independent series, the test shows that for $\rho_1 = 0.767$ and $\rho_1 = 0.785$ these differences Δr_k may be considered as being those of normal independent variables, because they are well within the confidence limits at the 95% level for all values of r_1 through r_{17}.

The model of the first order linear autoregressive scheme is, therefore, applicable to the series of annual flow of the St. Lawrence River. Assuming t to be a normal standard variable (0, 1), ϵ_1 in eq. 2.39 is $\epsilon_1 = t\sqrt{1 - \rho_1^2}$, so that the model in modular coefficients becomes $K_i = \rho_1 K_{i-1} + t\, C_v \sqrt{1 - \rho_1^2} + 1 - \rho_1$, or for $C_v = 0.087$ [1, Appendix 1], and $\rho_1 = 0.767$

$$K_i = 0.767\, K_{i-1} + 0.0557\, t + 0.233 \qquad 3.5$$

with $K_i = V_i / \bar{V}$. For given values of t_1 and t_2, the probability of occurrence of K_i within these limits may be obtained from tables of the normal function, and K_1 and K_2 may be obtained from the above equation.

Figures 9 and 10 show that the water carry-over from year to year in the Great Lakes and in the other parts of the St. Lawrence River basin is evidently responsible for most of the positive serial correlation. This is clear from the large positive initial ten r_k values, and the large length k of positive r_k values.

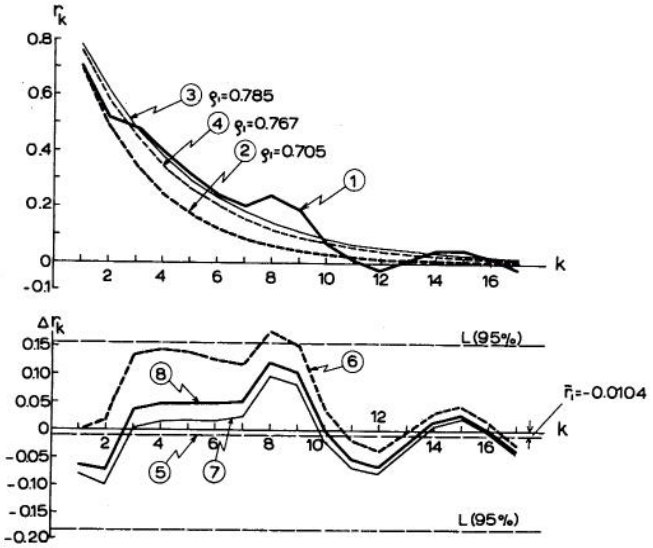

Fig. 10 Correlograms and their differences for the St. Lawrence River at Ogdensburg, New York. Upper graph: (1) V-series; (2) first order linear autoregressive scheme, with ρ_1 estimated in $\rho_k = \rho_1^k$ by the actual value $r_1 = 0.705$; (3) first order linear autoregressive scheme with ρ_1 estimated by least square fit to the r_k values of $\rho_k = \rho_1^k$; and (4) the same as under (3) but with the fit of r_k to lag r_k values of $\rho_k = \rho_1^k$. Lower graph: (5) expected value of $\overline{\rho_k}$, estimated by eqs. 2.3 and 2.23 with $N = 97$; (6) difference Δr_k of correlograms (1) and (2); (7) difference Δr_k of correlograms (1) and (3); (8) difference Δr_k of correlograms (1) and (4).

D. ANALYSIS OF THE SECOND LARGE SAMPLE OF RIVER FLOW AND THE LARGE SAMPLE OF PRECIPITATION

1. Simultaneous analysis of flow and precipitation. The first large sample of 140 stations of annual flow records has been analyzed simultaneously for two series, annual flow and derived annual effective precipitation. The second large sample of annual flow for 446 river station records (Western North America) is also analyzed simultaneously for annual flow and derived annual effective precipitation. In addition, the large sample of annual precipitation for 1141 stations, which covers the same area (Western North America), is analyzed simultaneously with annual flow and derived annual effective precipitation. For a detailed description of these two samples see Part I [1, pages 8-9 and 18-21].

The second large sample of river flow records has been treated in such a way that when two or more stations are situated on the same stream, all flow that has been previously measured upstream of a station has been subtracted from that station. In this way the large interdependence of annual flow and annual effective precipitation between the upstream and the downstream stations has been substantially reduced.

The study of patterns in simultaneous sequences of annual flow, annual effective precipitation and annual precipitation at the ground in a large region will thus give an insight into different river basin factors which are responsible for the dependence obtained in the particular type of time series.

Two parallel cases of investigation for V-series, P_e-series, and P_i-series (annual precipitation at the ground) are investigated: (a) all annual values available are used whether the record is continuous or not; in this case N_j, the record length, changes from station to station; and (b) the longest continuous period common to all stations is used which includes the years 1931-1960, a record of N = 30. The objective in investigating these two parallel cases is to study the influence of the length of time series on conclusions derived. The first case will give more reliable results than the second case, since the average sample size is greater in the first case, namely N_m = 37 years for the second large sample of river flow records for 446 stations, and N_m = 54 years for the large sample of homogeneous data of annual precipitation for 1141 stations.

2. Frequency distribution of the first serial correlation coefficients. The distributions of the f. s. c. c., computed from all data available at a station, are plotted in Fig. 11 on normal probability

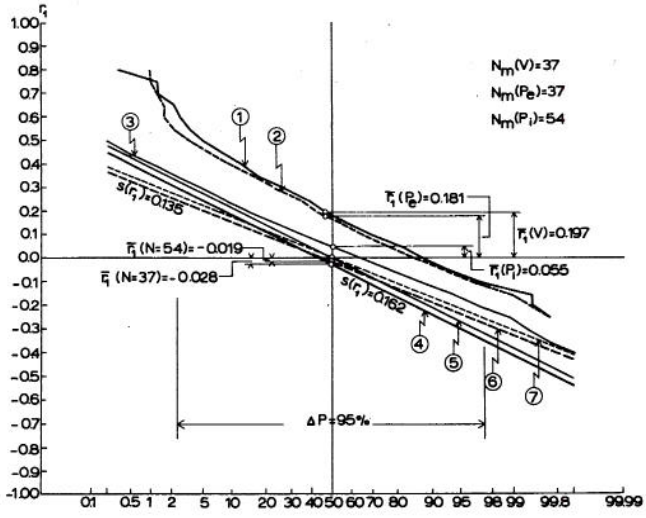

Fig. 11 Cumulative distributions of the first serial correlation coefficient for the series of the second large sample of river gaging stations (n = 446) and for the series of the large sample of precipitation gaging stations (n = 1141), both in Western North America on cartesian-probability scales: (1) V-series (annual flows); (2) P_e-series (annual effective precipitation; (3) P_i-series (annual precipitation at the ground); (4) normal independent variables, with the mean $\bar{\rho}_1$ estimated by eq. 2.3 and 2.23, and the variance estimated by eq. 2.27, both with N_m = 37 which corresponds to V- and P_e-series; (5) normal independent variables, the same as under (4) but with the mean $\bar{\rho}_1$ estimated by eq. 2.2., and the variance by eq. 2.27; (6) normal independent variables, the same as under (4) but with N_m = 54 which corresponds to P_i-series; and (7) normal independent variables, the same as under (5) but with N_m = 54 which corresponds to P_i-series.

paper. The distributions plotted are for V-series, P_e-series, and P_i-series, as well as for the normal independent variables with $N_j = 37$ (which corresponds to V- and P_e-series) and $N_j = 54$ (which corresponds to P_i-series). In both of these two cases, $\bar{r}_1$ and $s(r_1)$ for the normal independent variables are computed by moments from eq. 2.2 and eq. 2.3. The average values of r_1 are also given for all cases.

In the range of probability from 2.5% to 97.5%, a straight line may fit the r_1-distribution of V-series, P_e-series, and P_i-series. These straight lines are approximately parallel to the straight lines which represent the r_1-distributions of normal independent variables.

Assuming that the moments of eq. 2.3 for normal independent variables approximate well the case when r_1 is computed by the approximate expression of eq. 1.10, then the differences between various $\bar{r}_1$-values, $\bar{r}_1$ denoting the case of normal independent variables, are

$$\Delta\bar{r}_1(V) = \bar{r}_1(V) - \bar{r}_1 = 0.197 + 0.028 = 0.225$$

$$\Delta\bar{r}_1(P_e) = \bar{r}_1(P_e) - \bar{r}_1 = 0.181 + 0.028 = 0.209$$

$$\Delta\bar{r}_1(P_i) = \bar{r}_1(P_i) - \bar{r}_1 = 0.055 + 0.019 = 0.076$$

The above $\bar{r}_1$-values have been computed by simple average procedure or by using eq. 2.19 for V-, P_e-, and P_i-series, and by eq. 2.23 for the normal independent variables, because it was shown on the example of the first large sample of flow records that differences between the simple and weighted means are not substantial when sample sizes, N_j, are different.

For Western North America, the difference of $\bar{r}_1$ for V- and P_e-series is

$$\bar{r}_1(V) - \bar{r}_1(P_e) = 0.225 - 0.209 = 0.016.$$

This indicates that a part of the positive correlation in V-series is produced by water carryover in surface and underground storage in the river basins which is bound to flow out in following years. Similarly for P_e- and P_i-series the difference is

$$\bar{r}_1(P_e) - \bar{r}_1(P_i) = 0.209 - 0.076 = 0.133.$$

A substantial portion of the positive serial correlation in P_e-series is produced by evaporation (and evapotranspiration) from the water carryover of previous years which is evaporated (lost out of the basin as evaporation) in successive years. Because $V = P_e \pm \Delta W$, with ΔW being the difference in water carryover for each year, and $P_e = P_i - E_i$, with E_i being the annual evaporation, and because ΔW and E_i depend on the water carryover from previous years, these two magnitudes ΔW and E_i are the most important factors in producing the time dependence: both of them for V-series, and E_i only for P_e-series.

While the first large sample of river flow records from many parts of the world has shown that the carryover in the form of surface and underground storage (which flows out in successive years) accounted for a substantial portion of the positive serial correlation in V-series, the second large sample shows a relatively smaller impact of this carryover on the dependence of the V-series. The annual evaporation and the annual evapotranspiration (E_i) result as the major causal factors for time dependence in both the V- and P_e-series. Since Western North America encompasses a large arid and semi-arid area where the evaporation and evapotranspiration represent a large portion in the water balance of river basins, this result of a great effect of evaporation and evapotranspiration on time dependence of annual flow and annual effective precipitation should be expected.

It should be pointed out that the water carryover of previous years can be disposed of in next years in two ways: (a) by flowing out of river basin through surface runoff and through underground outflow; and (b) by flowing out into the atmosphere through evaporation and evapotranspiration. From the point of view of explaining the time dependence in a series of annual flow these two means of carryover depletion make no substantial difference except that different mathematical models, based on the different physical processes, may fit approximately these two manners of water outflow from river basins.

The fact stressed above that dependence in time series decreases substantially from annual flow to annual effective precipitation and from annual effective precipitation to annual precipitation at the ground is the main and the most significant general result of this study.

Similarly as in Fig. 11, Fig. 12 gives the distributions of the f. s. c. c. (r_1) on normal probability paper for V-, P_e-, and P_i-series for simultaneous records of 30 years (1931 - 1960) at all stations as well as for normal independent variables with $\bar{r}_1$ computed by the first moment of either eq. 2.2 or of eq. 2.3, or with $\bar{r}_1 = 0$ and $\bar{r}_1 = -0.034$, respectively. In this last case for $\bar{r}_1$

$$\Delta\bar{r}_1(V) = \bar{r}_1(V) - \bar{r}_1 = 0.163 + 0.034 = 0.197$$

$$\Delta\bar{r}_1(P_e) = \bar{r}_1(P_e) - \bar{r}_1 = 0.146 + 0.034 = 0.180$$

$$\Delta\bar{r}_1(P_i) = \bar{r}_1(P_i) - \bar{r}_1 = 0.028 + 0.034 = 0.062 .$$

The difference between $\bar{r}_1(\cdot)$ of V- and P_e-series are

$$\bar{r}_1(V) - \bar{r}_1(P_e) = 0.017$$

and

$$\bar{r}_1(P_e) - \bar{r}_1(P_i) = 0.118 .$$

The comparison between the two alternatives, longest records and simultaneous and continuous but shorter records, leads to the following conclusions:

(a) The absolute values of $\bar{r}_1(V)$, $\bar{r}_1(P_e)$, and $\bar{r}_1(P_i)$ are somewhat smaller for 30-year record than for longest records;

(b) The effect of water carryover and

evaporation on time dependence is shown to be approximately the same, regardless of the average length of time series and simultaneity of flow observations;

(c) The fitted straight lines to r_1-distributions for V-, P_e-, and P_i-series are similar in the case of 30-year record, except that the slope of r_1-distribution of P_i-series is somewhat smaller than that of r_1-distribution for normal independent variables of the same time series size of N = 30.

3. Frequency distribution of other serial correlation coefficients. Apart from r_1 the distributions of serial correlation coefficients, r_2 through r_{11}, are given in Fig. 13 for all years of observations and for the following variables: (1) V, (2) P_e, (3) P_i, (4) normal independent variables with N_m = 37, and (5) normal independent variables with N_m = 54. Case (4) serves for the comparison with V and P_e variables, and case (5) with the P_i variable. Similarly, Fig. 14 gives distributions for r_k-values with k = 2, 3, . . . , 11 for 30 years of observation (1931-1960) and for the following variables: (1) V, (2) P_e, (3) P_i, and (4) normal independent variables with N = 30.

Both $\bar{r}_k$ and var r_k for normal independent variables have been computed by using the average length of time series. However, in computing the covariance of x_i and x_{i+k} only (N-k)-values have been used.

The objectives of presenting Figs. 13 and 14 in this investigation are: (a) to show that the distributions of r_k for V, P_e, and P_i become closer and closer to those of normal independent variables as k increases; (b) to demonstrate that the distribution of r_k for the P_i variable is closer to the distribution of normal independent variables than the distributions of r_k for V and P_e variables; and (c) to demonstrate that the use of N instead of N-k in formulas for the expected value of r_k and of var r_k introduces a departure in slope between the distributions of r_k for V, P_e, and P_i, and those of normal independent variables, and that this departure increases with an increase of k. These objectives will be pursued further in the discussion of the change in the statistical parameters of the r_k-distributions as k increases.

The general conclusions from Figs. 13 and 14 are:

(1) Average values of r_k for V, P_e, and P_i-series are closer to the average values r_k of normal independent variables for large k-values (k = 7, 8, 9, 10, 11) than for small k-values (k = 2, 3, 4, 5, 6).

(2) Straight line fits to the distributions of r_k for V-, P_e-, and P_i-variables parallel better the straight line distributions for r_k of the corresponding

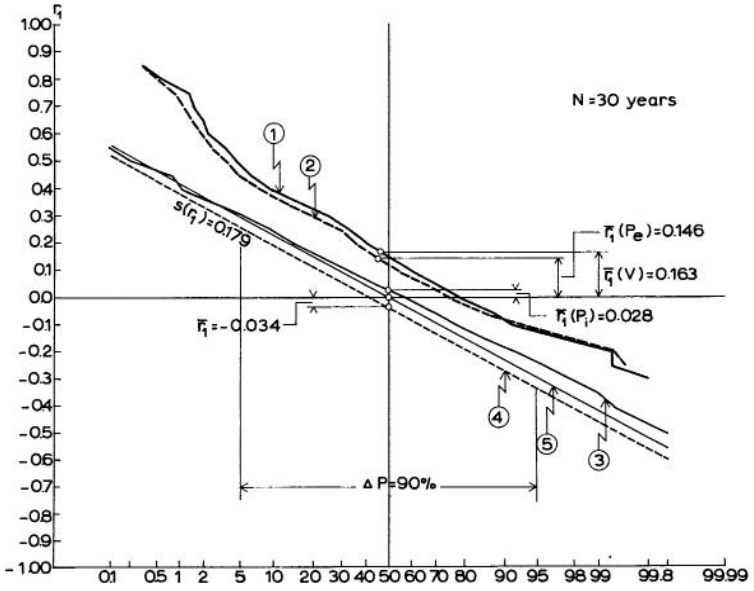

Fig. 12 Cumulative distributions of the first serial correlation coefficient for the series of the second large sample of river gaging stations (n = 446) and for the series of the large sample of precipitation stations (n = 1141), both in Western North America, for the simultaneous period of observation 1931-1960, with N = 30, on cartesian-probability scales: (1) V-series (annual flow); (2) P_e-series (annual effective precipitation); (3) P_i-series (annual precipitation at the ground); (4) normal independent variables, with the mean $\bar{\rho}_1$ estimated by eq. 2.3 and 2.23, and the variance estimated by eq. 2.27; and (5) normal independent variables, with the mean $\bar{\rho}_1$ estimated by eq. 2.2, and the variance estimated by eq. 2.27.

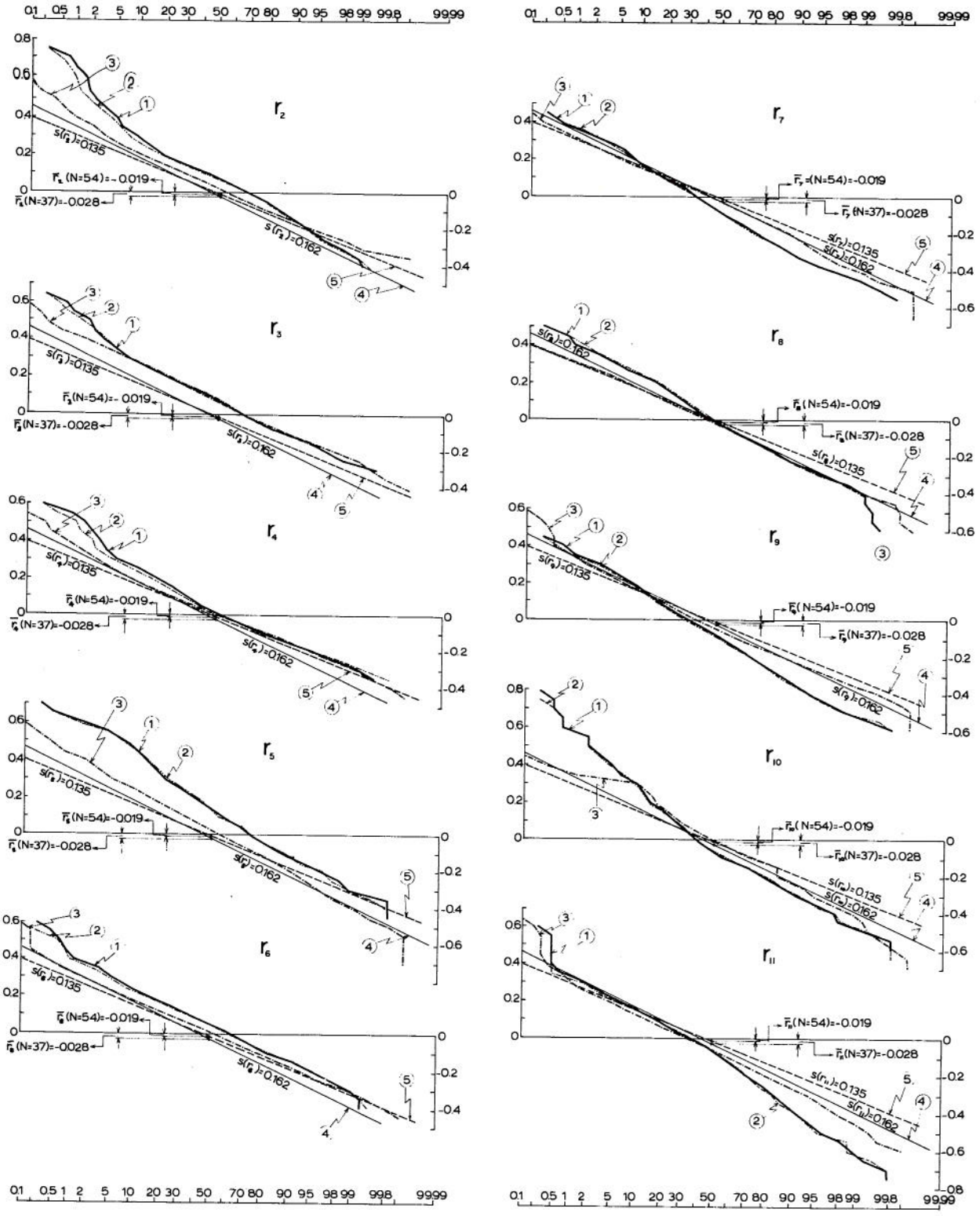

Fig. 13 Cumulative distributions of serial correlation coefficients r_2 through r_{11} for the series of the second large sample of river gaging stations (n = 446) and the large sample of precipitation gaging stations (n = 1141) from Western North America on cartesian-probability scales: (1) V-series (N_m = 37); (2) P_e-series (N_m = 37); (3) P_i-series (N_m = 54); (4) normal independent variables, with the mean $\bar{\rho}_1$ estimated by eqs. 2.3 and 2.23, and the variance estimated by eq. 2.27 with N_m = 37; and (5) normal independent variables, the same as under (4) except that N_m = 54.

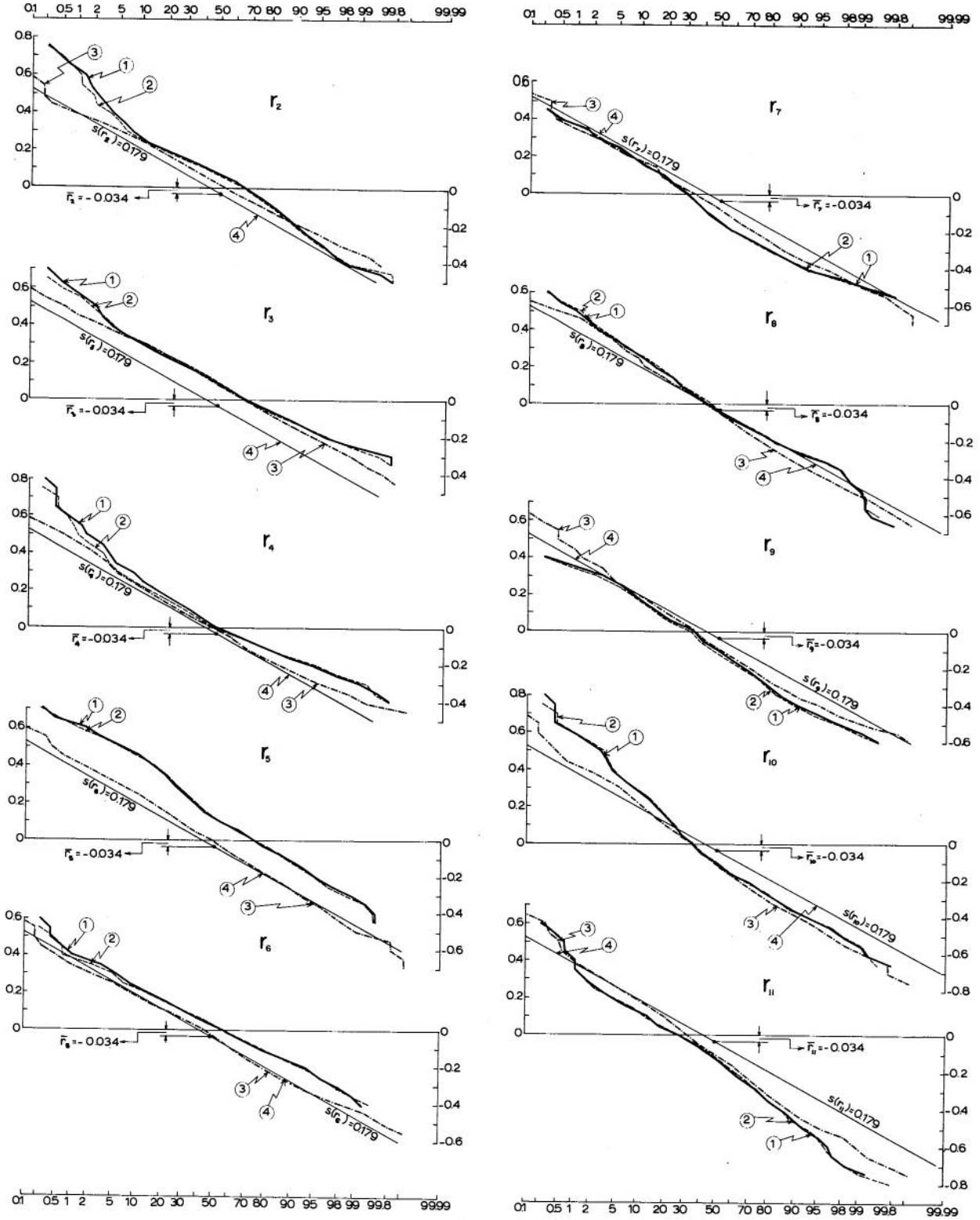

Fig. 14 Cumulative distributions of serial correlation coefficients r_2 through r_{11} for the series of the second large sample of river gaging stations (n = 446) and the large sample of precipitation gaging stations (n = 1141) from Western North America for the simultaneous period of observation 1931-1960, with N = 30, on cartesian-probability scales: (1) V-series; (2) P_e-series; (3) P_i-series; and (4) normal independent variables, with the mean $\bar{\rho}_1$ estimated by eqs. 2.3 and 2.23, and the variance estimated by eq. 2.27 with N = 30.

normal independent variables (same N) for small k (2, 3, 4, 5, 6) than for large k (7, 8, 9, 10, 11). Because the series are relatively short for V and P_e (N_m = 37), the value $N_m - k$ is changing from 35 to 26 for k changing from 2 to 11, and the use of N_m instead of $N_m - k$ in computing $\bar{r}_k$ and var r_k may be the explanation for this conclusion. For P_i with N_m = 54, the effect of k is smaller for this variable than for V and P_e because N_m-k is much larger for P_i-series than for V- and P_e-series for the same k .

(3) A replacement of N by N-k in the expressions for the mean and variance of r_k may better fit the distributions of r_k of V, P_e, and P_i for large values of k, than if N were used for all r_k-values. This conclusion may be partly the consequence of using eq. 1.10 in the computation of r_k instead of using eq. 1.5.

4. Average values of the serial correlation coefficients. Figure 15 is based upon computations from the longest available records of V-, P_e-, and P_i-series. It gives the following statistics of r_k-distributions as they change with the lag k: mean $\bar{r}_k$, standard deviation s_r, skewness coefficient C_{sr}, and kurtosis k_r. The mean and other statistics are computed by using the average length of the time series in the sample. The expected mean $\bar{r}_k$ and the standard deviation s_r of normal independent variables are given for the corresponding average length of time series (with N_m = 37, which corresponds to V- and P_e-series, and with N_m = 54 which corresponds to P_i-series). In this case $\bar{\rho}_k$ for normal independent variables is estimated by eq. 2.7 for all k values by $\bar{r}_k = -1/N_m$ as well as by the equation

$$E\bar{\rho}_k = \frac{1}{N_m - k} \quad . \qquad 4.1$$

Similarly, the standard deviation of r_k of normal independent variables is computed by eq. 2.3 as $s_r = \sqrt{(N_m + 1)}$ as well as by the equation

$$s_r = \frac{1}{\sqrt{N_m - k + 2}} \qquad 4.2$$

Figure 16 (as in Fig. 15) gives the same statistics for V-, P_e-, P_i-series, and normal independent variables for the simultaneous 30 years of records (1931-1960).

Both figures show clearly that the expressions given by eqs. 4.1 and 4.2 for the mean and standard deviation of r_k fit the observed values for large k better than the values obtained by eq. 2.7 or 2.3. Therefore, eqs. 2.2, 2.3, 2.4, 2.5, 2.7, and 2.8 are good approximations only for small k or better for ratios N/(N-k) close to unity. Note that s_r computed for normal independent variables by the use of eq. 4.2 with the appropriate series length approximates very well the s_r's of V-, P_e, and P_i-series. This is true for both cases where all available record and the 30-year period are used. If computed by eq. 2.3, s_r of normal independent variable is constant and departs significantly from s_r of the other variables for large k values.

The skewness coefficient C_{sr} of r_k-distributions fluctuates highly but, on the average, not very far above the value $C_{sr} = 0$. The kurtosis k_r of r_k-distributions fluctuates also about k = 3, or about the value for normal distribution of r_k .

It may be concluded from the values s_r, C_{sr}, and k_r that r_k-distributions may be considered normal with mean $\bar{r}_k$ and standard deviation s_r except near the ends of the range [+1, -1]. For large k the values $\bar{r}_k$ of V-, P_e-, and P_i-series fluctuate about $\bar{r}_k$-values computed by eq. 4.1.

As s_r's of V-, P_e-, and P_i-series are close to s_r's of normal independent variables, the test whether their difference is or is not significant from zero is not carried out here. The only test carried out here for the significance of differences is for $\bar{r}_k(V) - \bar{r}_k$; $\bar{r}_k(P_e) - \bar{r}_k$; and $\bar{r}_k(P_i) - \bar{r}_k$, with $\bar{r}_k$ the value for the corresponding normal independent variables. In this test of differences, the confidence limits at the 95% level are used about $\bar{r}_k$, assuming that r_k-values of normal independent variables are normally distributed about $\bar{r}_k$. The values $\bar{r}_k$ of normal independent variables are estimated by eq. 4.1 and are given in Figs. 15 and 16, upper graph. The simple average $\bar{r}_k$ for V-, P_e-, and P_i-series is used instead of a weighted average, with the expectation that the large values of n (n=446 for V- and P_e-series, and n = 1141 for P_i-series) will produce approximately the same values of $\bar{r}_k$ as the weighted means.

The effective number of stations for $\bar{r}_k$ is n_e, and it is estimated by eqs. 2.74 and 2.73, with var $\bar{r}_k$ = var $r_k/n_e = s_r^2/n_e$, and the standard deviation of $\bar{r}_k$ is $s(\bar{r}_k) = s_r/\sqrt{n_e}$. The confidence limits on a given level are, therefore

$$\bar{r}_k = \bar{\rho}_k \pm \frac{t\sigma_r}{\sqrt{n_e}} \approx \bar{r}_k \pm \frac{t s_r}{\sqrt{n_e}} \qquad 4.3$$

Using eq. 4.1 for the estimate of $\bar{\rho}_k$, and eq. 4.2 for the estimate of σ_r then

$$\bar{r}_k = \frac{-1}{N_m - k} \pm \frac{t}{\sqrt{n_e(N_m - k + 2)}} \qquad 4.4$$

From eq. 2.74 it follows that

$$\bar{r}_k = -\frac{1}{N_m - k} \pm t\sqrt{\frac{1 + \bar{r}\,(n-1)}{n(N_m - k + 2)}} \qquad 4.5$$

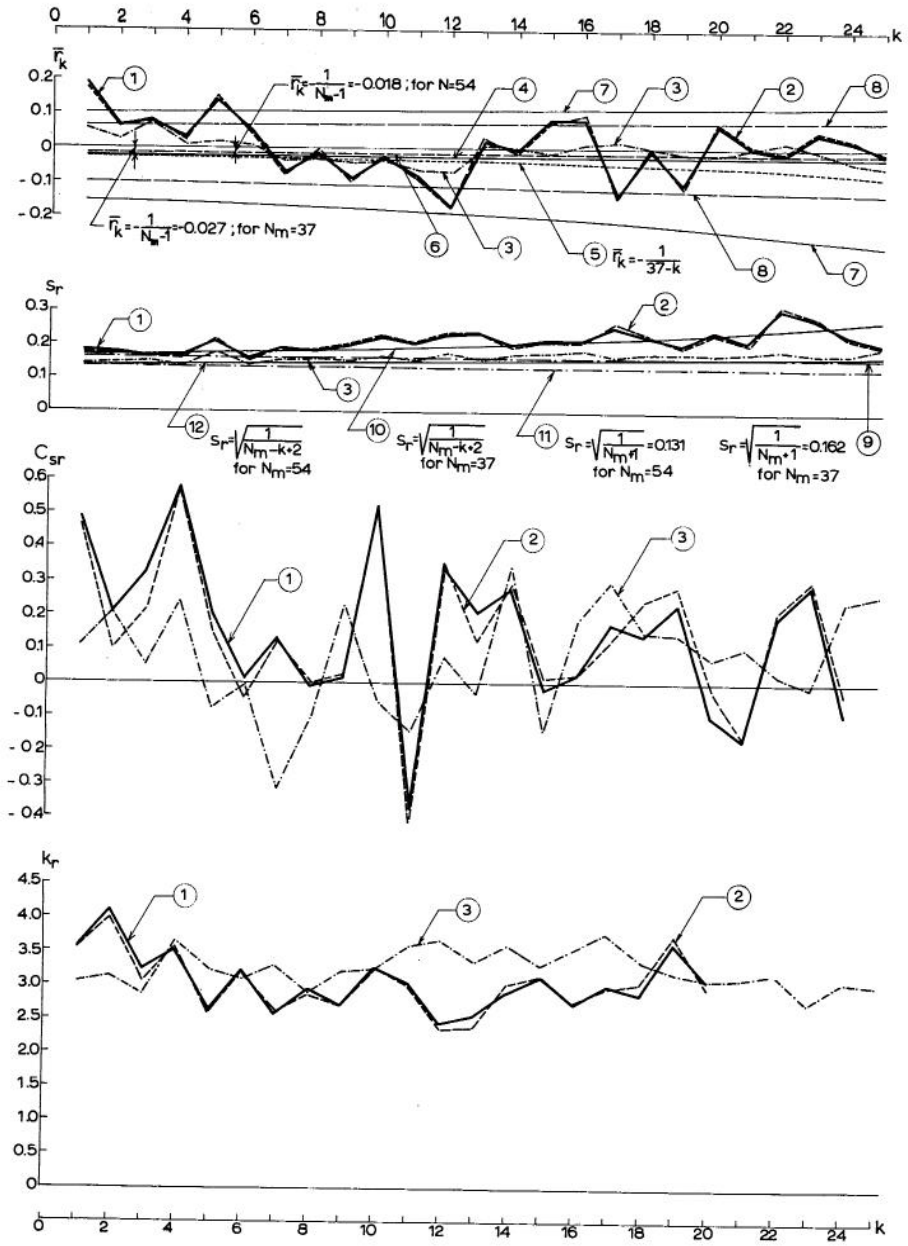

Fig. 15 Average values of $\bar{r}_k$, standard deviation s_r, skewness coefficients C_{sr}, and kurtosis k_r of the serial correlation coefficients r_1 through r_{25} for the secondlarge sample of river gaging stations (n = 446) and the large sample of precipitation gaging stations (n = 1141) from Western North America: (1) V-series (N_m = 37); (2) P_e-series (N_m = 37); (3) P_i-series (N_m = 54); (4) normal independent variables, the mean $\bar{\rho}_k$ estimated by eq. 2.3 with N_m = 37; (5) normal independent variables, the mean $\bar{\rho}_k$ estimated by eq. 4.1 with N_m = 37; (6) normal independent variables, the mean $\bar{\rho}_k$ estimated by eq. 4.1 with N_m = 54; (7) normal independent variables, the confidence limits at the 95 percent level for the effective number of stations n_e = 6.30, instead of the actual number n = 446; (8) normal independent variables, the confidence limits at the 95 percent level with the effective number of stations n_e = 9.65 instead of the actual number n = 1141; (9) normal independent variables, the standard deviation estimated by the second moment from eq. 2.3 with N_m = 37; (10) normal independent variables, the standard deviation estimated by eq. 4.2 with N_m = 54; (11) normal independent variables, the same as under (9) but with N_m = 54; (12) normal independent variables, the same as under (10) but with N_m = 54. Curves 1, 2, and 3 on all plots refer to V-, P_e- and P_i-series respectively.

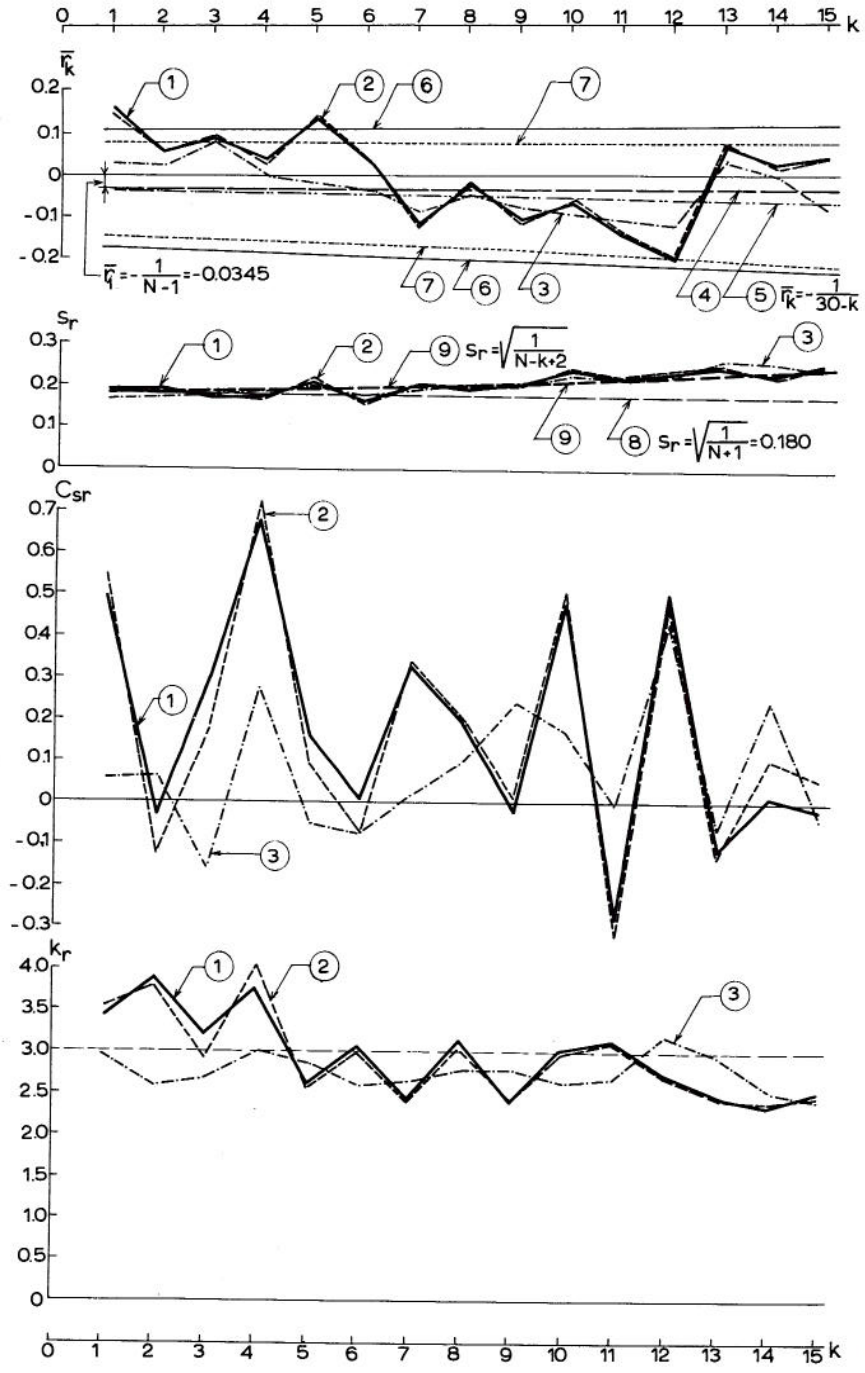

Fig. 16 Average value of $\bar{r}_k$, standard deviation s_r, skewness coefficient C_{sr}, and kurtosis k_r of the serial correlation coefficients r_1 through r_{15} for the secondlarge sample of river gaging stations (n = 446) and the large sample of precipitation gaging stations (n = 1141) from Western North America for the simultaneous period of observation 1931-1961, with N = 30: (1) V-series; (2) P_e-series; (3) P_i-series; (4) normal independent variables with the mean $\bar{\rho}_k$ estimated by eq. 2.3 with N = 30; (5) normal independent variables with the mean $\bar{\rho}_k$ estimated by eq. 4.1 with N = 30; (6) normal independent variables, the confidence limits at the 95 percent level for the effective number of stations n_e = 6.30 instead of the actual number n = 446; (7) normal independent variables, the confidence limits at the 95 percent level for the effective number of stations n_e = 9.65 instead of actual number n = 1141; (8) normal independent variables, the standard deviation estimated by the second moment from eq. 2.3 with N = 30; (9) normal independent variables, the standard deviation estimated by eq. 4.2 with N = 30. Curves 1, 2, and 3 on all plots refer to V-, P_e-, and P_i-series respectively.

with $\bar{r}$ given by eq. 2.72. In section B, equations 2.69 through 2.74 were developed to compute the average interstation correlation coefficient of the first serial correlation coefficients among the station series. The matrix of the simple interstation correlations between station series and eq. 2.73 are used for the digital computation of the pairwise estimates of the interstation correlation coefficient of their first serial correlation coefficients. The average interstation correlation coefficient of f. s. c. c. is then computed by eq. 2.72. The average interstation correlation coefficient, $\bar{r}$, is 0.157 for V-series and 0.159 for P_e-series. Equation 2.74 gives the effective number of stations, $n_e = 6.26$ and $n_e = 6.29$ for V- and P_e- respectively. Even though the interstation correlation of the f. s. c. c. is relatively small, the original sample size of 446 is reduced to an effective random sample size of approximately 6.30 for both series.

Equation 4.4 gives the confidence limits at the 95% level (t = 1.96) for the mean of first serial correlation coefficient, with k = 1, $n_e = 6.30$, and $N_m = 37$, as $\bar{r}_1 = +0.099$ and $\bar{r}_1 = -0.155$. Assuming that the effective sample size n_e for any r_k is the same as for r_1, or $n_e = 6.30$, then for V- and P_e-series

$$\bar{r}_k = -\frac{1}{37-k} \pm \frac{1.96}{\sqrt{6.30(39-k)}} \qquad 4.6$$

The confidence limits computed by eq. 4.6 are plotted in Fig. 15.

Though n_e for V- and P_e-series has been computed for the longest records of each station, the same value $n_e = 6.30$ is also used for the V- and P_e-series of the 30-year period of record (1931-1960). Using eq. 4.4 and N=30, the values of the confidence limits of $\bar{r}_k$ at the 95% level are computed and plotted in Fig. 16, upper graph.

Figures 15 and 16 show that only two average values of r_k, namely $\bar{r}_1$ and $\bar{r}_5$ for both V- and P_e-series are outside confidence limits. It is concluded here that $\bar{r}_1(V)$ and $\bar{r}_1(P_e)$ are significantly different from the expected value of $\bar{\rho}_1$ of normal independent variables at the 95% level. However, the fact that $\bar{r}_5(V)$ and $\bar{r}_5(P_e)$ are greater than the positive confidence limit at the 95% level may be explained by sampling fluctuations. There are 25 values of $\bar{r}_k$ for each of the two series and 5% of them or about 1.25 should be outside the confidence limits on the average.

The P_i-series (n = 1141) has the average interstation correlation coefficient between the first serial correlation coefficients of $\bar{r} = 0.095$. The effective number of stations, n_e, equals 9.65. Eq. 4.4 gives the confidence limits at the 95% level (t = 1.96) for the mean of the f. s. c. c. with k = 1, $n_e = 9.65$, and $N_m = 54$, as

$$\bar{r}_1 = +0.066 \quad \text{and} \quad \bar{r}_1 = -0.103 .$$

Assuming the effective sample size n_e for any r_k is the same as for r_1, or $n_e = 9.65$, then for the P_i-series

$$\bar{r}_k = -\frac{1}{54-k} \pm \frac{1.96}{\sqrt{9.65\,(56-k)}} \qquad 4.7$$

The confidence limits computed by eq. 4.7 are plotted in Fig. 15. Using the same effective number of stations $n_e = 9.65$ for the P_i-series with N=30 (period 1931-1960) and eq. 4.4, the confidence limits are plotted in Fig. 16 similarly as in Fig. 15.

Figure 15 shows that the only value of $\bar{r}_k$ for the P_i-series which is outside of confidence limits at the 95% level is $\bar{r}_3$. Although $\bar{r}_1$ is inside the confidence limits, Fig. 15 shows that $\bar{r}_1$ through $\bar{r}_6$ of P_i-, V-, and P_e- series are all positive. Also, the negative values $\bar{r}_7$ through $\bar{r}_{12}$ of V- and P_e-series are parallelled by negative values of P_i-series. This indicates that P_i-series has a small regression effect of moisture carryover similar to that of P_e-series. Figure 16 shows that all $\bar{r}_k$ values of the P_i-series are inside the confidence limits at the 95% level except $\bar{r}_3$ which touches the positive limit.

It may be concluded that the annual precipitation at the ground has no significant dependence in sequence and that the series of annual precipitation statistically cannot be distinguished in its sequential patterns from independent variables.

5. Individual correlograms. Figures 17 and 18 give correlograms of annual flows (V-series) and annual effective precipitation (P_e-series) for 40 river gaging stations and correlograms of annual precipitation at the ground (P_i-series) for 44 rainfall stations all taken from the large sample of stations from Western North America.

The expected values of $\bar{r}_k$ for normal independent variables are computed for an average length N_m of the time series for all 24 graphs (each containing 4-6 correlograms) for these three series, and they are plotted in each graph as shown in Figs. 17 and 18. Also given in Figs. 17 and 18 are the confidence limits at the 95% level as computed by eq. 2.10. The time series length used varies from one grouping to another depending on the mean length of the time series presented in that particular grouping. These mean lengths have been rounded to the nearest multiple of ten, i. e., N = 40, 50, 60, 70, or 80.

This massive presentation of correlograms is used here intentionally to show an overall confinement of correlograms of V-, P_e-, and P_i-series inside the confidence limits for an average series length of normal independent variables. Figures 17 and 18 show clearly that most of the correlograms are well inside the approximate confidence limits, especially if one takes into consideration the fact that about 5% of r_k-values should be on the average outside the confidence limits for normal independent variables plus the fact that some of the serial correlation coefficients (r_1, r_2, r_3 . . .), particularly those of V- and P_e-series, should be significantly different from those of normal independent variables.

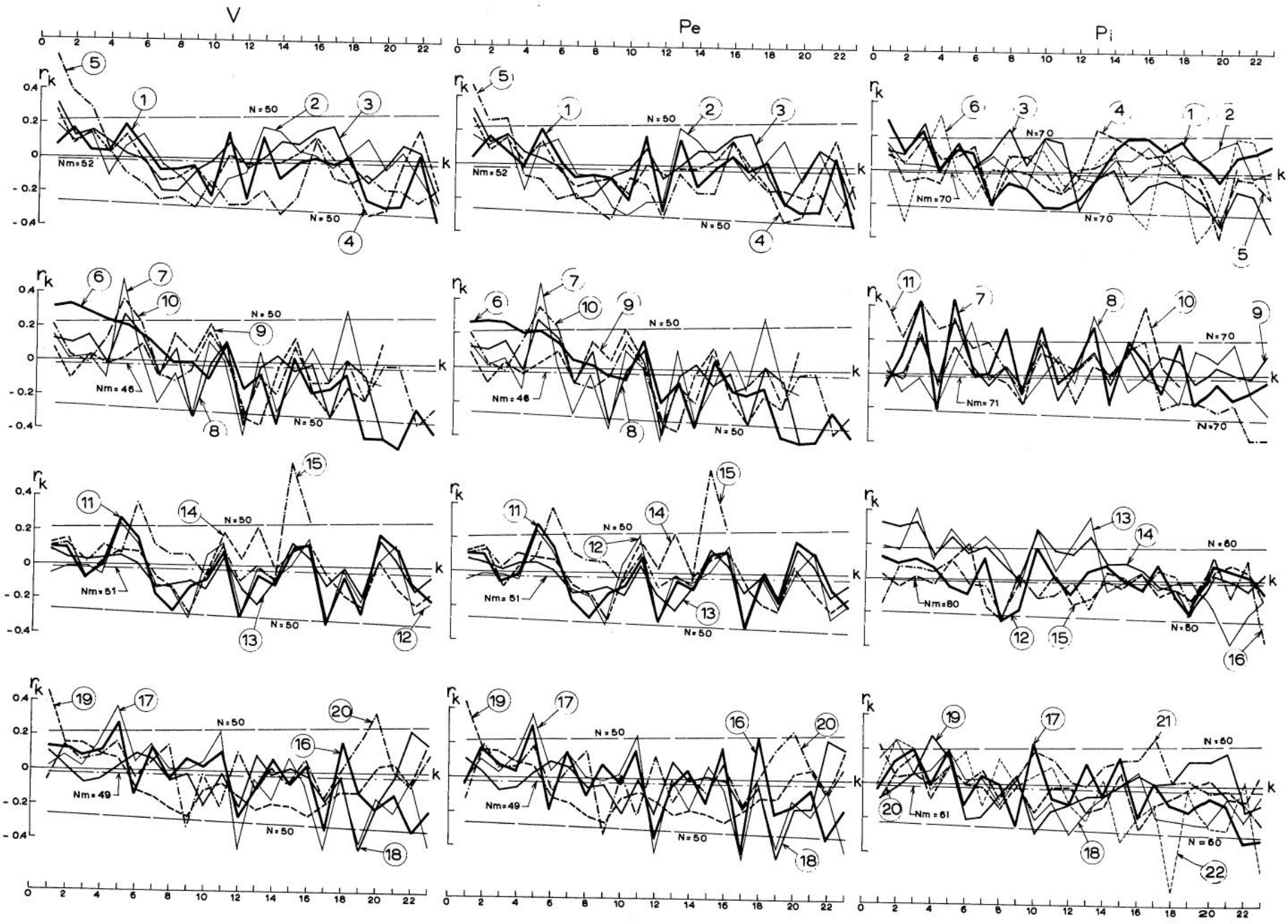
V
Pe
Pi
r_k
k
N = 50
Nm = 52
N = 50
N = 70
Nm = 70
N = 70
N = 50
Nm = 46
N = 50
N = 70
Nm = 71
N = 70
N = 50
Nm = 51
N = 50
N = 60
Nm = 80
N = 60
N = 50
Nm = 49
N = 50
N = 60
Nm = 61
N = 60

Fig. 17 Correlograms of 62 individual series in groups of 5 or 6 from V-, P_e-, and P_i-series. The left column shows correlograms for the first 20 series from the V-series and the middle column shows correlograms for the first 20 series from the P_e-series, all taken from the second large sample of river gaging stations (n = 446). The right column shows correlograms for the 22 series from the P_i-series from the large sample of precipitation gaging stations (n = 1141). For V-series and P_e-series the confidence limits at the 95 percent level of significance for the normal independent variables are given as estimated by eq. 2.10 with N = 50, and for P_i-series with three values of N = 60, 70, and 80. The mean $\overline{\rho}_k$ as estimated by eqs. 2.3 and 2.33 are also plotted on each graph. with N_m, the average length of series in each group.

The river gaging stations are:

(1) Klickitat River near Glenwood, Washington, 1910-1960 (51)
(2) Quinault River at Quinault Lake, Washington, 1912-1960 (49)
(3) Cedar River near Landsburg, Washington, 1896-1960 (65)
(4) Wenatchee River at Plain, Washington, 1911-1960 (50)
(5) Thompson River at Spences Bridge, CANADA, 1917-1960 (44)
(6) Oak Grove Fork above Power Plant Intake, Oregon, 1910-1960 (51)
(7) Siletz River at Siletz, Oregon, 1906-1911, 1926-1960, (41)
(8) South Fork Big Butte Creek near Butte Falls, 1918-1922, 1926-1960 (40)
(9) Grande Ronde River at LaGrande, Oregon, 1904-1915, 1919-1960 (52)
(10) Silvies River near Burns, Oregon, 1904-1905, 1910-1912, 1918-1960 (48)
(11) Kaweah River near Three Rivers, California, 1911-1960 (50)
(12) Cherry Creek near Hetch Hetchy, California, 1911-1960 (50)
(13) Arroyo Seco near Soledad, California, 1902-1960 (59)
(14) Trinity River at Lewiston, California, 1912-1960 (49)
(15) West Fork Mohave River near Hesperia, California, 1905-1922, 1930-1960 (49)
(16) Snake River at Moran, Wyoming, 1904-1960 (57)
(17) Boise River near Twin Springs, Idaho, 1912-1960 (49)
(18) St. Joe River at Calder, Idaho, 1912, 1921-1960 (41)
(19) Milk River at Milk River, Alberta, CANADA, 1912-1960 (49)
(20) Kootenay River at Wardner, CANADA, 1928-1960 (33)

The precipitation gaging stations are:

(1) Anacortes, Washington, 1893-1960 (68)
(2) Walla Walla WB City, Washington, 1857-1859, 1860-1861, 1864, 1874-1960 (93)
(3) Chelan, Washington, 1892-1960 (69)
(4) Centralia, Washington, 1892, 1894, 1896-1897, 1902-1922, 1925-1960 (61)
(5) Seattle WB AP, Washington, 1892-1960, (69)
(6) Kelowna, British Columbia, CANADA, 1900, 1903-1960 (59)
(7) Rosenburg, WB AP, Oregon, 1878-1960 (83)
(8) LaGrande, Oregon, 1887, 1890-1891, 1893-1895, 1898-1960 (69)
(9) Albany, Oregon, 1879-1960 (82)
(10) Prineville 4NW, Oregon, 1897-1902, 1904-1909, 1911, 1914-1919, 1922-1926, 1928-1960 (57)
(11) Lakeview, Oregon, 1885-1887, 1891-1892, 1895-1898, 1901-1907, 1913-1960 (64)
(12) Chico Experiment Station, California, 1871-1960 (90)
(13) Eureca WB City, California 1887-1960 (74)
(14) Fort Ross, California, 1875-1960 (86)
(15) San Jacinto, California, 1887, 1893-1960 (69)
(16) Visalia, California, 1878-1885, 1888-1960 (81)
(17) Caldwell, Idaho, 1905-1960 (56)
(18) Oakley, Idaho, 1894-1960 (67)
(19) Grace, Idaho, 1907-1960 (54)
(20) Ashton IS, Idaho, 1899, 1902-1913, 1915-1960 (59)
(21) Helena WB AP, Montana, 1881-1882, 1884-1960 (79)
(22) Lyndon, Alberta, CANADA, 1911-1960 (50)

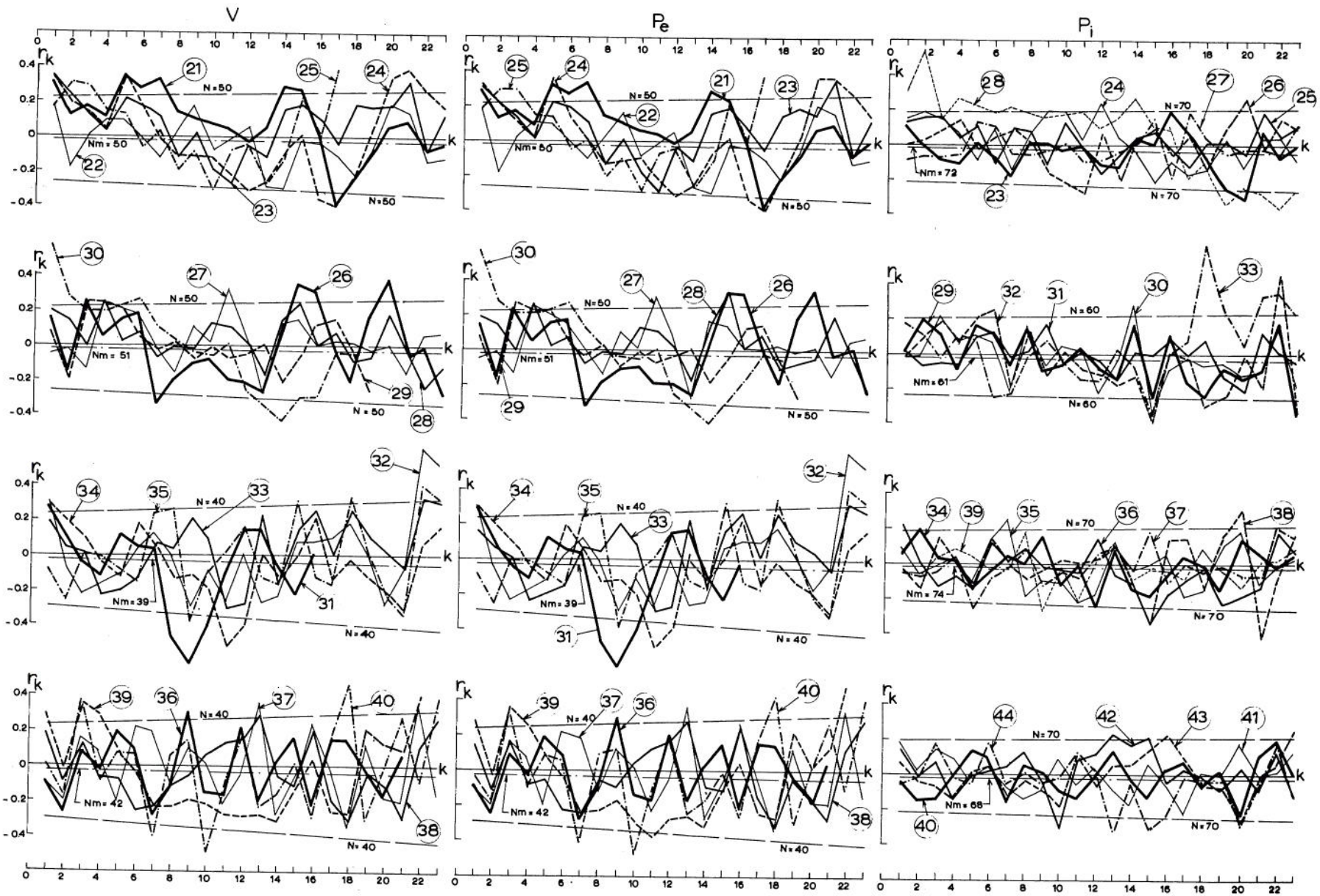
V
Pe
Pi
rk
k
N=50
Nm=50
N=50
N=50
Nm=72
N=70
N=70
Nm=51
N=50
N=50
Nm=51
N=50
Nm=61
N=60
N=40
Nm=39
N=40
N=40
Nm=39
N=40
N=70
Nm=74
N=70
N=40
Nm=42
N=40
N=40
Nm=42
N=40
N=70
Nm=68
N=70

Fig. 18 Correlograms of 62 individual series in groups of 5 or 6 from V-, P_e-, and P_i-series is a continuation of Fig. 17. The left column shows correlograms for 20 additional series from the V-series, numbers 21-40 and the middle column shows correlograms for 20 additional series from the P_e-series, numbers 21-40, taken from the second large sample of river gaging stations (n = 446). The right column shows correlograms for 22 additional series from the P_i-series, numbers 23-44, from the large sample of precipitation gaging stations (n , 1141). For V-series and P_e-series the confidence limits at the 95 percent level of significance for the normal independent variables are given as estimated by eq. 2.10 from two values N = 50 and N = 40 and for the P_i-series for the N = 70 and N = 60. The mean ρ_k as estimated by eqs. 2.3 and 2.23 are also plotted on each graph with N_m, the average length of series in each group.

The river gaging stations are:

(21) Weber River near Oakley, Utah, 1905-1960 (56)
(22) White River near Meeker, Colorado, 1902-1906, 1910-1960 (56)
(23) Lion Creek near Halfway, Colorado, 1909-1960 (52)
(24) Bull Lake Creek near Leonore, Wyoming, 1919-1960 (42)
(25) Little Missouri River near Alzada, Montana, 1913, 1916-1925, 1929-1960 (43)
(26) Dolores River at Dolores, Colorado, 1896-1903, 1911-1912, 1922-1960 (49)
(27) San Pedro River at Charleston, Arizona, 1905, 1913-1960 (49)
(28) Verde River below Bartlett Dam, Arizona, 1889-1960 (30)
(29) Red River near Questa, New Mexico, 1913-1960 (47)
(30) Bluewater Creek near Bluewater, New Mexico, 1913-1915, 1917-1918, 1928-1960 (38)
(31) Middle Fork Forked Deer River near Almo, Tennessee, 1930-1960 (31)
(32) Eleven Point River near Ravenden Springs, Missouri, 1922-1960 (39)
(33) Tarkio River near Fairfax, Missouri, 1923-1960 (38)
(34) Marias Des Cygnes River near Ottawa, Kansas, 1903-1905, 1920-1960 (44)
(35) Petit Jean Creek at Danville, Arkansas, 1917-1960 (44)
(36) Yegua Creek near Somerville, Texas, 1925-1960 (36)
(37) Colorado River at Ballinger, Rexas, 1908-1960 (53)
(38) Llano River near Junction, Texas, 1916-1960 (45)
(39) Leon River near,Belton, Texas, 1924-1960 (37)
(40) Neches River at Evadale, Texas, 1923-1960 (38)

The precipitation gaging stations are:

(23) Fillmore, Utah, 1892-1960 (69)
(24) Moab, Utah, 1890-1960 (71)
(25) Salt Lake City WB AP, Utah, 1875-1960 (86)
(26) Lusk, Wyoming, 1890-1892, 1896-1899, 1902-1909, 1912-1918, 1921-1960 (62)
(27) Fort Collins, Colorado, 1873-1874, 1880, 1882-1884, 1887-1960 (80)
(28) Leadville, Colorado, 1889-1890, 1896-1904, 1908-1960 (64)
(29) Durango, Colorado, 1889, 1895-1960 (67)
(30) Jerome, Arizona, 1898-1899, 1901-1915, 1917, 1919-1960 (60)
(31) Tuscon, U of A, Arizona, 1876-1960 (85)
(32) Wolf Canyon, New Mexico, 1912-1960 (49)
(33) Ione, New Mexico, 1916-1960 (45)
(34) Ashland DDC8, Kansas, 1889-1960 (72)
(35) Manhattan No. 2, Kansas, 1858-1960 (103)
(36) New Madrid, Missouri, 1894-1960 (67)
(37) Stephenville, IE, Missouri, 1893-1960 (68)
(38) Marlow 1 WSW, Oklahoma, 1901-1960 (60)
(39 Mena, Arkansas, 1888-1889, 1891-1909, 1911-1960 (71)
(40) Coleman, Texas, 1878-1880, 1882, 1895-1960 (70)
(41) Galveston WB City, Texas, 1871-1960 (90)
(42) Greenville 2SW, Texas, 1900-1960 (61)
(43) New Braunfels, Texas, 1889-1906, 1908-1960 (71)
(44) Raymondville, Texas, 1911-1960 (50)

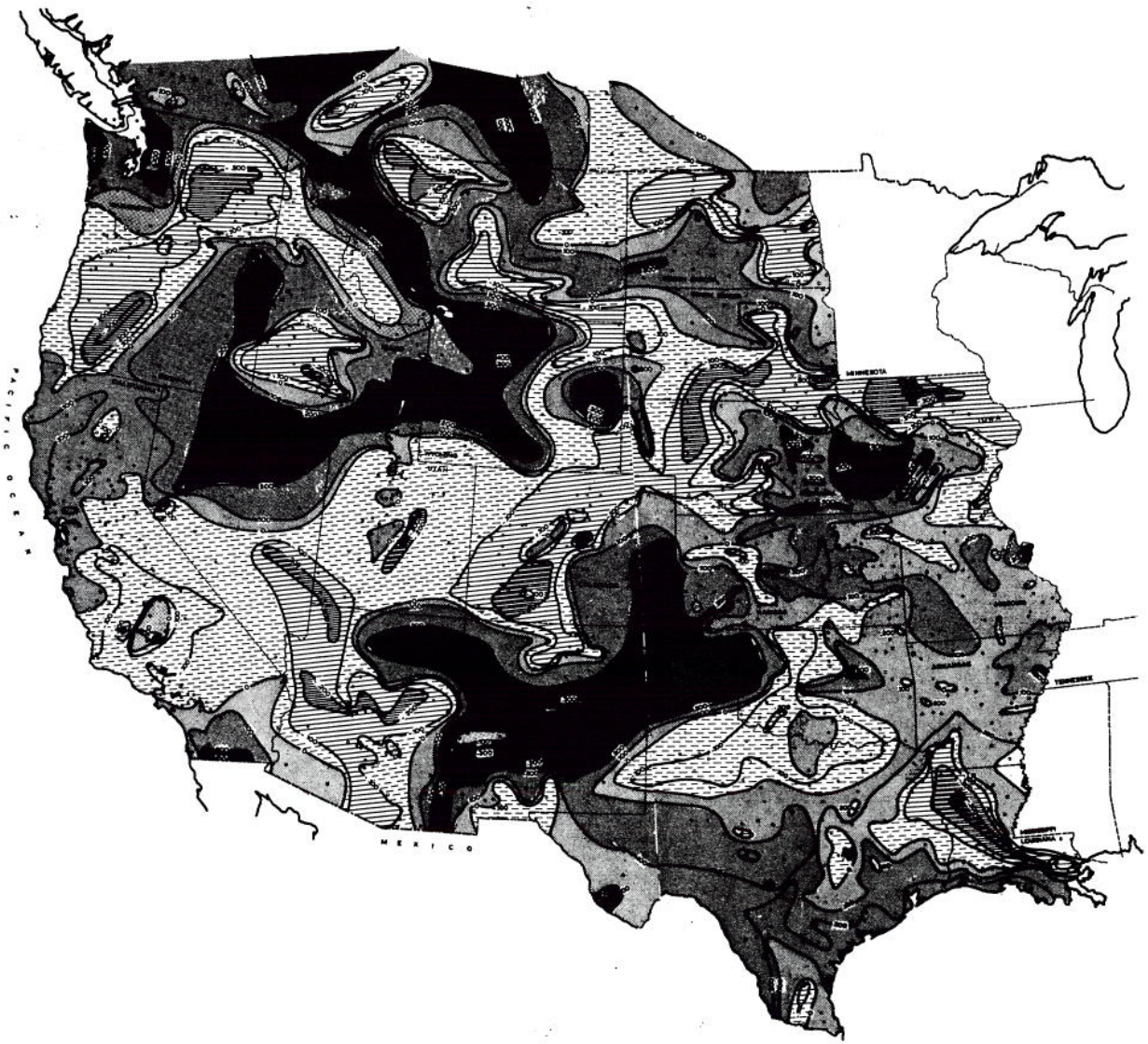

Fig. 19 Regional distribution of the first serial correlation coefficient for the large sample of precipitation gaging stations (n = 1141) from Western North America. The small circles are locations of precipitation gaging stations. Isolines of equal values of r_1(f. s. c. c.) are drawn showing no systematic patterns in the areal distribution of positive or negative values, especially of the extreme values.

Correlograms of the series of annual river flow, annual effective precipitation, and annual precipitation in North Western America, according to the above analysis, do not point to any patterns in sequence of cyclical character or to a significant trend. The fact that several correlograms start with high average values of r_1 through r_3 and then decrease slowly for a greater lag k, only points to the fact that some moving average scheme may be present. These schemes are likely to be only the product of water carryover in river basins from year to year and its outflow in subsequent years either through surface and ground water flow or through evaporation and evapotranspiration into the atmosphere.

6. Effect of length of time series. As has been shown above, the dependence increases with an increase of the time series length. This conclusion should be accepted very cautiously, because the comparison of a 30-year period (1931-1960) with a longer average length (37 and 54 years, respectively for V- and P_i-series) may be misleading. By pure chance or sampling fluctuations, the period of 30 years may show either a greater or a smaller time dependence than a somewhat longer average length which includes this period of 30 years.

It is likely that a longer average period than 30 years will have a larger amount of nonhomogeneity in the data (a larger quasi-stationarity), so that this factor may produce a somewhat greater dependence in longer series than in shorter series.

7. Regional distribution of first serial correlation coefficient. Figure 19 shows the isolines of the first serial correlation coefficients of annual precipitation. A total of 1141 precipitation stations have been used with 1141 values of r_1. The small circles represent the positions of these stations.

The isolines of r_1 show islands of high values of r_1 (0.3 to 0.4) as well as of low values r_1 (-0.2 to -0.3). There is no systematic pattern in the areal distribution of r_1. Islands of high and low values cover equally very humid regions as well as very arid regions.

It is quite unlikely that the patterns in areal distribution of r_1 as shown in Fig. 19 would repeat themselves in a future sample which would be independent of this sample used for Fig. 19. Therefore, the main hypothesis is advanced here that the regional distribution of r_1 as shown in Fig. 19 is mostly the result of sampling fluctuations. In other words, by an increase of the period of observation those islands of high and low values will decrease in absolute values of r_1 and converge to the expected mean, which is somewhere around zero. It is logical to expect that a prevalence of more positive values than negative values is partly produced by some small effect of nonhomogeneity in data (quasi-stationarity). The other factor for this prevalence of positive values may be the carryover of moisture from year to year and its impact on the evaporation of precipitation between the cloud base and the ground [See Part I, 1, p. 15-17].

8. The case of the Missouri River. The sampling fluctuation of the first s.c.c. is discussed here with the Missouri River as an example. Figure 20 shows the Missouri River Basin with its main tributaries and main storage reservoirs as developed by 1963. Figure 21 shows the fluctuation of wet and dry years in millions of acre feet for the Missouri River gaging station near Sioux City, Iowa, for the calendar years of the period 1898-1962. The characteristics of this station are given in Part I [1, Appendix 1, pages 28 and 29, and Appendix 2, page 36] for the water years of the period 1897-1955. The period after 1955 of this study in Part I has been excluded because of the large impact of storage reservoirs upon the fluctuation of wet and dry years. The river basin at this station covers 314,000 square miles. The first serial correlation coefficient for annual flow at this station is $r_1 = 0.590$ and for annual effective precipitation it is $r_1 = 0.532$ [1, Appendix 1, page 29].

The upstream station of the Missouri River at Fort Benton, with an area of 24,000 square miles, shows similar patterns in sequence as measured by r_1(f.s.c.c.) as the downstream station near Sioux City. For the 65-year period of observation (1890-1955) at Fort Benton the first serial correlation coefficient of annual flows is $r_1 = 0.593$ and of annual effective precipitation it is $r_1 = 0.582$ [1, Appendix 1, pages 28 and 29]. Therefore, the main stream of the Missouri River has annual flows which are highly correlated with $r_1 = 0.60$ approximately. However, the Mississippi River below the confluence of the Mississippi River at St. Louis shows a smaller first serial correlation for both annual flow and annual effective precipitation. They are 0.294 and 0.302, respectively or $r_1 = 0.30$ approximately. This means that the Missouri River Basin has a particularly high value of r_1 in comparison to the Mississippi after their confluence.

The main question that arises is whether the high correlation is a permanent pattern for the Missouri River or whether it is mostly a product of sampling fluctuations of wet and dry years. In the case the sampling fluctuation predominates the population series of annual flow is supposed to be highly correlated. In this case most of the high positive correlation comes from the sampling fluctuation of r_1 about a much lower population value of ρ_1. It can be concluded on the other side that a part of the positive serial correlation (as measured by r_1) of the Missouri River is produced by water carryover from year to year which is released to the atmosphere in subsequent years by evaporation and evapotranspiration. In support of this conclusion is the fact that there is a significant difference between the first serial correlation coefficient of annual precipitation in the Missouri River Basin as shown by isolines of r_1 in Fig. 19 and the first correlation coefficient of annual flow as given above for the two gaging stations. It can be assumed also that a part of the high positive first serial correlation coefficient in annual flows was produced in the historical data by nonhomogeneity and/or inconsistency in the data. The period 1890-1960 in the Missouri River Basin was one of a constant increase in population and water depletion by man-made structures and other measures. There must be a small trend in average annual flow (a decrease) which may account for an increase in r_1 (f.s.c.c.).

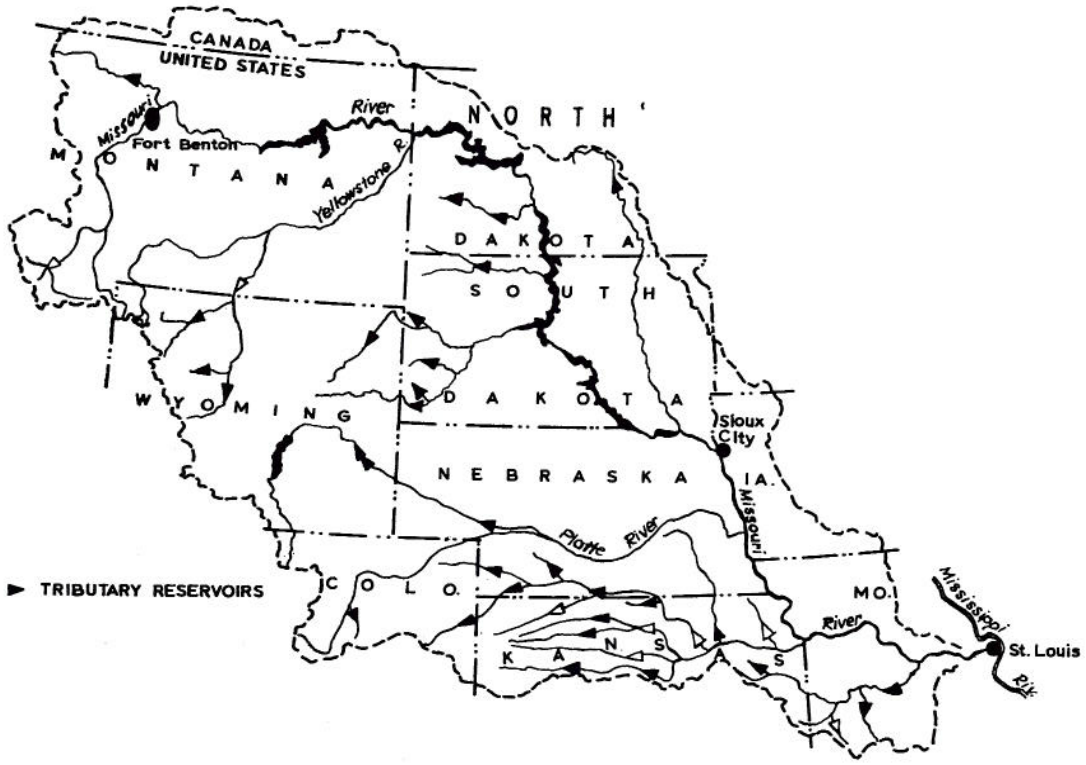

Fig. 20 The Missouri River Basin with main tributaries and main storage reservoirs as developed by 1963. The three main river gaging stations used are: Fort Benton and Sioux City on the Missouri River and St. Louis on the Mississippi River.

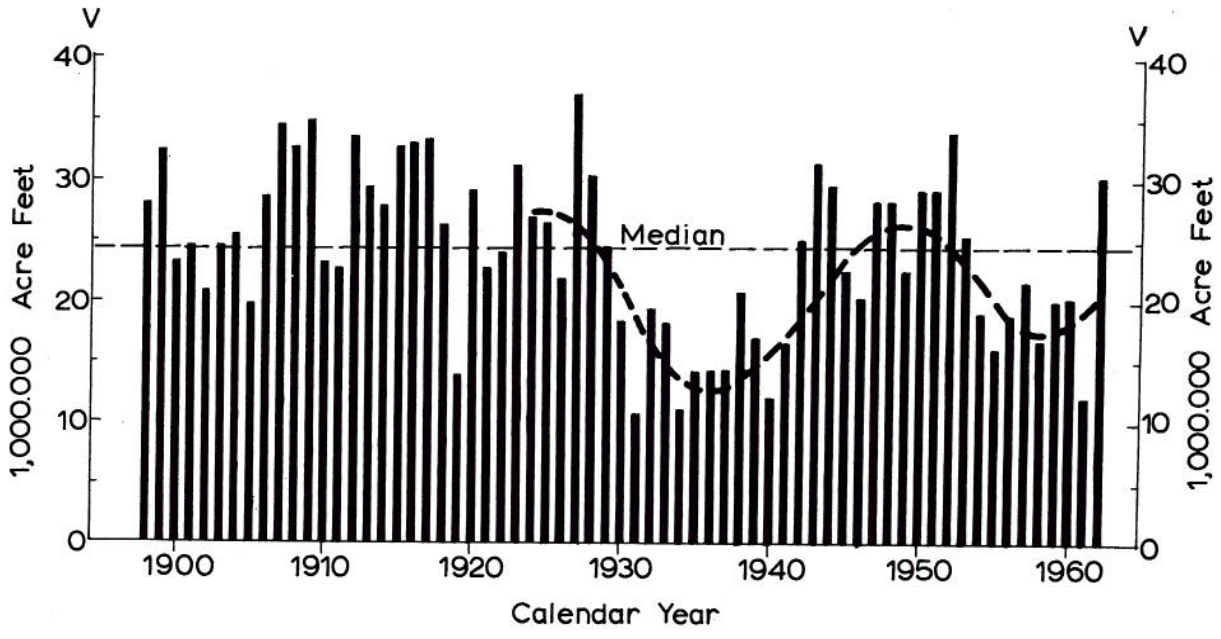

Fig. 21 Time series of the annual flow of the Missouri River near Sioux City, Iowa, for calendar years 1898-1962. The main pseudo-cyclicity is emphasized for the period 1925-1962 by a dashed heavy line.

Figure 21 shows a particular cluster of wet and dry years from 1929 to 1962. On the average there was a dry period in the thirties, a wet period in the forties and early fifties, and a dry period again in the late fifties and early sixties. This type of cyclic movement in the last 30 years is not evident systematically throughout the United States, or even in the Middle West or West. It is particularly marked, however, in the Missouri River Basin and some surrounding areas [1, Appendix 1, pages 28 and 29]. This particular movement or cluster of wet and dry years is mostly responsible for a large first serial correlation coefficient in the annual flow of the Missouri River.

A hypothesis advanced here is that the particular cluster and the resulting positive serial correlation of wet and dry years along the Missouri River is largely a product of sampling fluctuation, but the other factors also affect this positive correlation. The high first serial correlation coefficient is assumed here to be produced by these factors: (1) sampling cluster of wet and dry years in a pseudo-cyclicity for the last 30 - 35 years; (2) water carryover in the river basin from year to year which is released to the atmosphere by evaporation in subsequent years (difference between r_1 for V- and P_e-series is small, so that the carryover from year to year which is released through river flow in subsequent years is small); and (3) nonhomogeneity in data, which is a consequence of the constant flow depletion. The hypothesis of sampling fluctuation as the dominant factor is worthwhile for further investigation because of its implication to design criteria and to the operation of large storage reservoirs in the Missouri River Basin for overyear flow regulations.

A question arises as to the probability that the next period of 30 - 35 years will have a similar sequence of wet and dry years as the past period of 30 - 35 years. The next question is are there any particular physical causal factors which would produce the pseudo-cyclicity of the type shown in Fig. 21 in the Missouri River Basin which do not act in other regions of the United States or elsewhere. The comparison of the period 1898-1928 with the period of 1929-1962 for the Missouri River near Sioux City as shown in Fig. 21 (it should be noted that Fig. 21 gives annual flows for calendar years while Part I, Appendix 2, page 36 gives the annual flows for water years) points out that the probability of a repetition of the cluster of wet and dry years of the period 1929-1962 is at least smaller than 50 percent. Any comparison with other regions in the United States will decrease this probability. The study of the atmospheric processes which lead to precipitation and create evaporation does not support any assumption that there may be some particular factors which are proper only for the Missouri River Basin and which favor or even consistently produce the above described patterns in the sequence of wet and dry years of the period 1929-1962.

A systematic study of clusters of wet and dry years in the Missouri River Basin is worthwhile. This may lead to a separation of the effects of the three main causal factors of high values of r_1 (f. s. c. c.): the sampling effect, the carryover effect, and the nonhomogeneity in data (depletion) effect. The practical importance of this analysis would lie in better design and operation of large reservoir storage capacities in the river basin. The cluster of wet and dry years as experienced in the period 1929-1962 requires special operation criteria and larger storage capacities for a given degree of flow regulation than the cluster of wet and dry years of the period 1898-1928. If the cluster of the period 1929-1962 has a small probability of repeating itself in the near future, which is the hypothesis advanced here, then the actual storage capacities are either overdesigned when based on that cluster or the effect and benefit of storage reservoirs will be greater in the future than those evaluated or computed by using this cluster.

E. EFFECT OF INCONSISTENCY AND NONHOMOGENEITY OF DATA

1. Errors and nonhomogeneity. Errors and nonhomogeneity in data have been discussed previously in Part I (1, Chapter E, pages 22-25). Random errors and systematic errors, being defined here as inconsistency in data, have been analyzed briefly in Part I with the conclusion that the data of many river gaging stations and many precipitation gaging stations have a higher or lower degree of inconsistency. It can be proved by statistical analysis that an independent time series becomes dependent when the series has been subjected to modifications by inconsistencies in the form of jumps or trends. The nonhomogeneity in data has already been defined as changes in time series produced by significant accidents in nature and by man-made structures and other regulatory measures in river basins.

The study of the effects of inconsistency and nonhomogeneity in data will be the subject of further investigation and these results will be presented in a special paper. A comparison of distributions of the first serial correlation coefficient for two large samples of series of annual precipitation (a sample of homogeneous and a sample of nonhomogeneous data) is given here for the simple purpose of showing that the positive serial correlation in the series of annual flow, annual effective precipitation and annual precipitation is partly produced by inconsistency and/or nonhomogeneity in data.

2. Comparison of homogeneous or consistent data with samples of nonhomogeneous or inconsistent data. Figures 22 and 23 give probability distributions of the first serial correlation coefficient (r_1) for the series of annual precipitation of the large sample of 1141 precipitation stations with the data considered as homogeneous and/or consistent (P_i^1 - -series) and of the large sample of 473 precipitation stations with the data found or considered as nonhomogeneous and/or inconsistent (P_i^2 - series). Figure 20 refers to the longest period of observation available for each series of both samples of stations, and Fig. 21 refers to the simultaneous period of observation 1931-1960 for all series of both samples. The distributions of the first serial correlation coefficient of normal independent variables are also given in Fig. 22 and 23 for comparative purposes. The mean $\bar{\rho}_1$ and the variance of the r_1 - distributions of the normal independent variables are estimated by moments of either eq. 2.2 or eq. 2.3 .

The series with nonhomogeneous and/or inconsistent data are those which were found as such either by a consistency or a homogeneity test by the river flow forecasting service of the U.S. Weather Bureau or by the author because of a substantial change in the station position (horizontal or vertical change of gage position during the observation period).

The average values $\bar{r}_1$ for homogeneous (and/or consistent) and nonhomogeneous (and/or inconsistent) data of series 30 years long are 0.028 and 0.053, respectively. The average $\bar{r}_1$ for homogeneous (and/or consistent) and nonhomogeneous (and/or inconsistent) data of series of maximum available length of observation are 0.055 and 0.071, respectively. The r_1 - distribution of the P_i^2 - -series is above the r_1 - distribution of the P_i^1 - -series in both figures. These distributions are also above the r_1 - distribution of the normal independent variables for the lengths of series which correspond to P^1_i - and P^2_i - series, respectively. The average values $\bar{r}_1$ (P^1_i) and $\bar{r}_1$ (P^2_i) as well as the distributions of r_1 (P^1_i) and r_1 (P^2_i) show that the nonhomogeneity and/or inconsistency in data is not a negligible factor in producing the dependence in the time series of annual precipitation. Nonhomogeneity and inconsistency are very often present in annual values of river flow and derived effective precipitation as shown to exist in the data of annual precipitation.

Although there may be a disagreement about the classification of precipitation data into homogeneous or consistent and nonhomogeneous or inconsistent samples, the large number of stations in the two classes (P^1_i - and P^2_i - series) tends to minimize error due to this cause and to validate the conclusion that nonhomogeneity and inconsistency in data increase, on the average, the dependence in time series.

The conclusion derived from these two samples of annual precipitation about an increase of dependence in time series by an increase of nonhomogeneity may be supported by theoretical analysis. Whenever a trend of a jump or the combination of the two is introduced by any process into an independent time series, the average result is that the series becomes dependent in sequence. The reservation here is made by stating that this occurs on the average. Sampling in time from a population of independent time series produces series with small dependence (in the range of sampling fluctuation). It may happen that the process of introducing the nonhomogeneity or inconsistency into data of a series in the form of jumps or trends or both may increase or decrease this sampling dependence. However, for many series subjected to this analysis the average result will be that their dependence will increase.

3. Hypothesis of quasi-stationarity. The previous, as well as the above analysis, leads to a hypothesis of the existence of nonhomogeneity and inconsistency in all data of river flows and precipi-

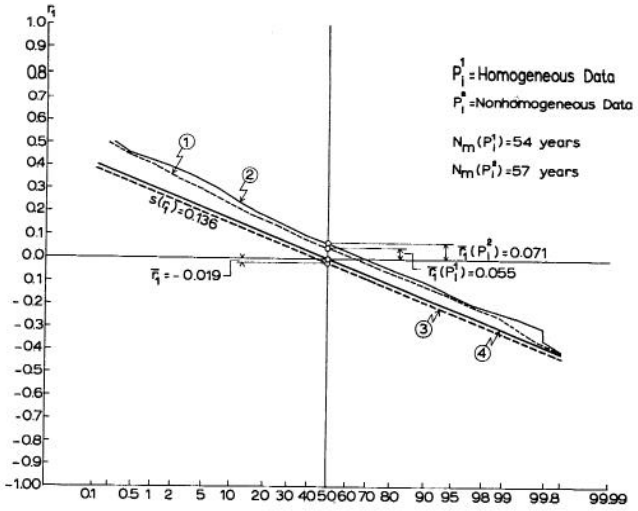

Fig. 22 Comparison of the distributions of the first serial correlation coefficient of homogeneous (and/or consistent) and nonhomogeneous (and/or inconsistent) time series of annual precipitation in Western North America: (1) distribution of r_1 (f. s. c. c.) from the large sample of 1141 precipitation gaging stations of the series of annual precipitation considered to be homogeneous and/or consistent (P_i^1-series) with an average length of series N_m = 54; (2) distribution of r_1 (f. s. c. c.) from the large sample of 473 precipitation gaging stations of the series of annual precipitation found or considered to be nonhomogeneous and/or inconsistent (P_i^2-series) with an average length of series N_m = 57 years; (3) distribution of r_1 (f. s. c. c.) from normal independent variables with the mean $\bar{\rho}_1$ estimated by eqs. 2. 3 and 2. 23 and the variance estimated by eq. 2. 7 with N_m = 55. 5 (average of 54 and 57); and (4) distribution of r_1 (f. s. c. c.) from normal independent variables with the mean $\bar{\rho}_1$ and the variance estimated by moments of eq. 2. 2 with N_m = 55. 5.

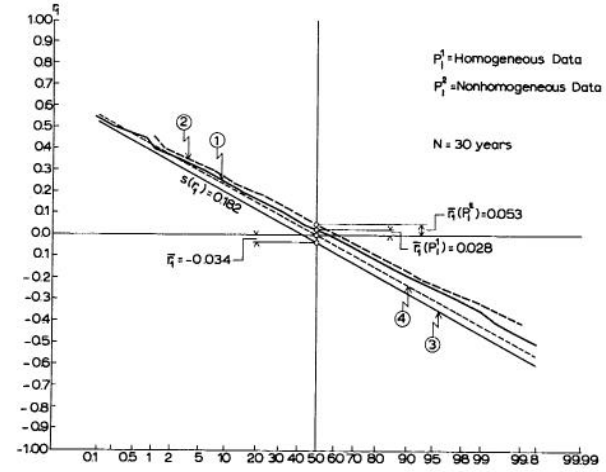

Fig. 23 Comparison of the distributions of the first serial correlation coefficient of homogeneous and/or consistent and nonhomogeneous and/or inconsistent time series of annual precipitation in Western North America for the simultaneous period 1931-1960 with N = 30: (1) distribution of r_1 (f. s. c. c.) from the large sample of 1141 precipitation gaging stations of the series of annual precipitation considered to be homogeneous and/or consistent (P_i^1-series); (2) distribution of r_1 (f. s. c. c.) from the large sample of 473 precipitation gaging stations of the series of annual precipitation found or considered to be nonhomogeneous and/or inconsistent (P_i^2-series); (3) distribution of r_1 (f. s. c. c.) from normal independent variables with the mean $\bar{\rho}_1$ estimated by eqs. 2. 3 and 2. 23 and the variance estimated by eq. 2. 7 with N = 30; and (4) distribution of r_1 (f. s. c. c.) from normal independent variables with the mean $\bar{\rho}_1$ and the variance estimated by moments of eq. 2. 2 with N = 30.

tation. With a degree of nonhomogeneity and inconsistency which varies greatly from series to series and from variable to variable (flow, effective precipitation, precipitation at the ground, evaporation, change in water carryover, etc.), even the detection and removal of jumps or trends by appropriate corrections will still leave nonhomogeneity and inconsistency in the data, but to a small degree. This small degree of nonhomogeneity and inconsistency in data which is unremovable in the practical sense, is defined here as the quasi-stationarity of the data. The impact of quasi-stationarity on the dependence of time series may be small, even sometimes undetectable by current statistical techniques or tests, but it is always present.

This hypothesis of quasi-stationarity in hydrologic data warrants a systematic investigation, both by theoretical and by practical analysis and tests.

F. EFFECTS OF CLIMATIC CONDITIONS ON SERIAL CORRELATION

1. Climatic conditions. Climatic conditions are understood here as being measured by the water yield either from the atmosphere to the ground as the average annual precipitation in inches or as the specific water yield of river basins at the stream flow gaging stations expressed as the average flow rate in cubic feet per second per square mile. The effect of climatic conditions on serial correlation may be studied either by regions or through a relationship between the serial correlation coefficients and the measures of water yield. The regional distribution of the first serial correlation coefficient for time series of annual precipitation is shown in Fig. 19. The relationship of the first serial correlation coefficient of the various series (V-, P_e-, P_i^1-, and P_i^2-series) to the water yield is analyzed here. The total range of variation of the average annual precipitation is divided into three groups: small, medium, and large. For each group the distribution of r_1 (f. s. c. c.) is given, and the three r_1-distributions are compared. Similarly, the average specific water yield of the river basins is used to divide the time series of annual river flow into three groups: small, medium and large average specific yield. For each group the distribution of r_1 (f. s. c. c.) is also given, and the three r_1-distributions are compared.

2. Comparison of distributions of first serial coefficient for various groups of water yield. Figure 24 gives r_1-distributions for V-series (annual flow) of the second large sample of river gaging stations (n = 446) in Western North America for the three different ranges of the specific water yield in cfs/sq. mi. of river basin area and for the maximum length of observation of each series with the average length N_m = 37: (1) q = 1.4-10.6; (2) q = 0.5-1.4; and (3) q = 0.0-0.5 cfs/sq. mi. Figure 25 gives the same r_1-distributions but for the simultaneous period of observation 1931-1960 with N = 30.

Figures 26 and 27 represent the same r_1-distributions and ranges of specific water yields as the corresponding Figs. 24 and 25 except that Figs. 26 and 27 refer to P_e-series (annual effective precipitation) of the second large sample of river gaging stations (n = 446) in Western North America.

Figures 28 and 29 represent the same r_1-distributions as the corresponding Figs. 24 and 25 except that Figs. 28 and 29 refer to the P_i^1-series (annual precipitation of homogeneous and/or consistent data) of the first large sample of precipitation gaging stations (n = 1141) in Western North America for three ranges of average annual precipitation: (1) P = 27-180; (2) P = 16.5-27; and (3) P = 0.0-16.5 inches. Figure 28 refers to all available observations with N_m = 54 while Fig. 29 refers to the simultaneous observation 1931-1960 with N=30.

Figures 30 and 31 represent the same r_1-distributions corresponding to Figs. 28 and 29 except that Figs. 30 and 31 refer to P_i^2-series (annual precipitation with nonhomogeneous and/or inconsistent data) of the second large sample of precipitation gaging stations (n = 473) in Western North America for the ranges of average annual precipitation:(1) P = 32.5-130; (2) P = 17.0-32.5; and (3) P = 0.0-17.0 inches. Figure 30 refers to all available observations with N_m = 57 while Fig. 31 refers to the simultaneous observation 1931-1960 with N = 30.

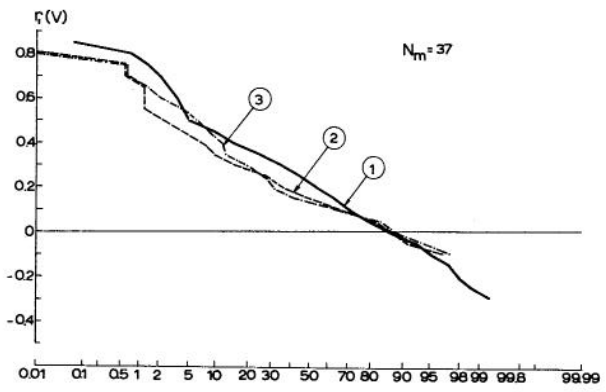

Fig. 24 Distributions of the first serial correlation coefficient for V-series (annual flow) of the second large sample of river gaging stations (n = 446) in Western North America on cartesian-probability scales with N_m = 37 for the three ranges of specific water yield: (1) q = 1.4-10.6 cfs/sq. mi.; (2) q = 0.5-1.4 cfs/sq. mi.; and (3) q = 0.0-0.5 cfs/sq. mi.

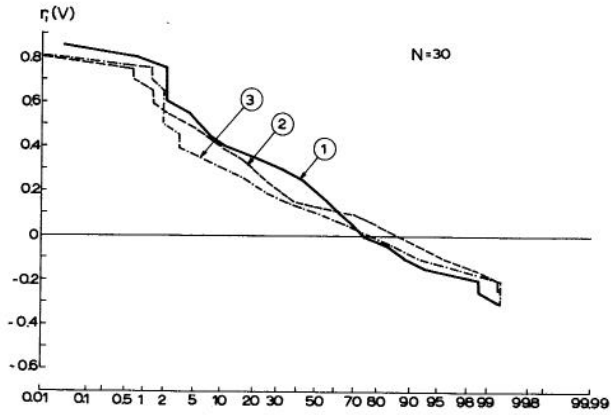

Fig. 25 Distributions of the first serial correlation coefficient for V-series (annual flow) of the second large sample of river gaging stations (n = 446) in Western North America on cartesian-probability scales for the simultaneous observations of 1931-1960 with N = 30 for the three ranges of specific water yield: (1) q = 1.4-10.4 cfs/sq. mi.; (2) q = 0.5-1.4 cfs/sq. mi.; and (3) q = 0.0-0.5 cfs/sq. mi.

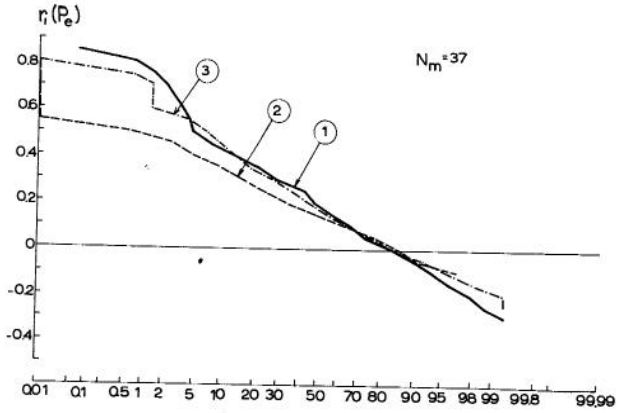

Fig. 26 Distributions of the first serial correlation coefficient for P_e-series (annual effective precipitation) of the second large sample of river gaging stations (n = 446) in Western North America on cartesian-probability scales with N_m = 37 for the three ranges of specifid water yield: (1) q = 1.4-10.4 cfs/sq. mi.; (2) q = 0.5-1.4 cfs/sq. mi.; and (3) q = 0.0-0.5 cfs/sq. mi.

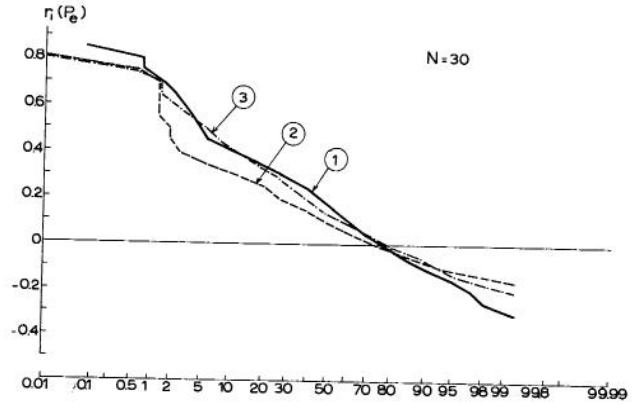

Fig. 27 Distributions of the first serial correlation coefficient for P_e-series (annual effective precipitation) of the second large sample of river gaging stations (n = 446) in Western North America on cartesian-probability scales for the simultaneous observations of 1931-1960 with N = 30 for the three ranges of water yield: (1) q = 1.4-10.4 cfs/sq. mi.; (2) q = 0.5-1.4 cfs/sq. mi.; and (3) q = 0.0-0.5 cfs/sq. mi.

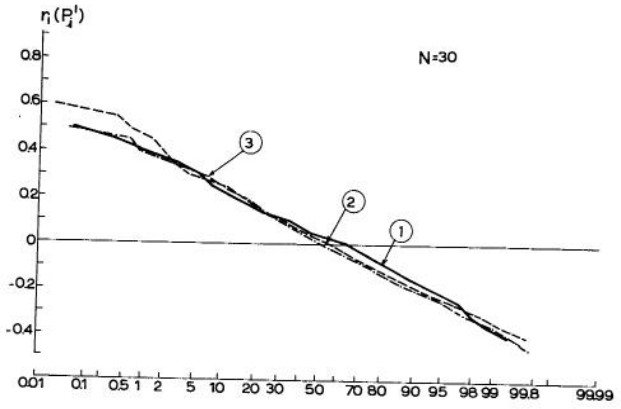

Fig. 28 Distributions of the first serial correlation coefficient for P_i^1-series (annual precipitation with homogeneous and/or consistent data) of the first large sample of precipitation gaging stations (n = 1141) in Western North America from all available data with N_m = 54 on cartesian-probability scales for the three ranges of average annual precipitation: (1) P = 27-180 inches; (2) P = 16.5-27 inches; and (3) P = 0.0-16.5 inches.

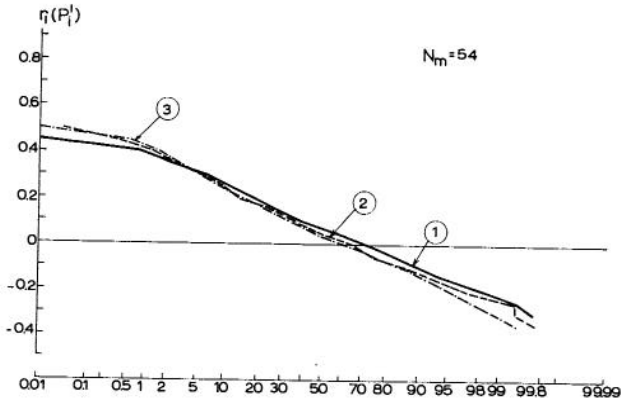

Fig. 29 Distributions of the first serial correlation coefficient for P_i^1-series (annual precipitation with homogeneous and/or consistent data) of the first large sample of precipitation gaging stations (n = 1141) in Western North America for the simultaneous observations of 1931-1960 with N = 30 on cartesian-probability scales for the three ranges of average annual precipitation: (1) P = 27-180 inches; (2) P = 16.5-27 inches; and (3) P = 0.0-16.5 inches.

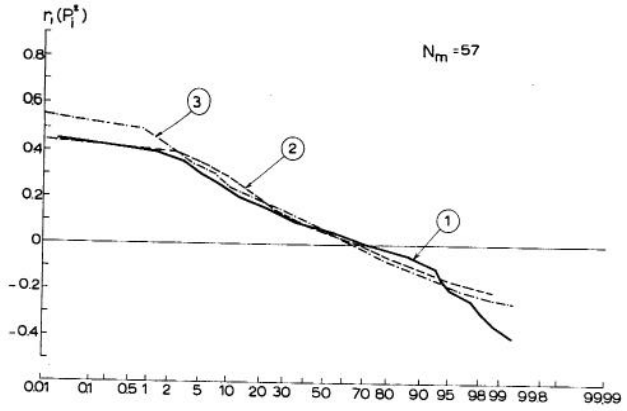

Fig. 30 Distributions of the first serial correlation coefficient for P_i^2-series (annual precipitation with nonhomogeneous and/or inconsistent data) of the second large sample of precipitation gaging stations (n = 473) in Western North America from all available data with N_m=57 on cartesian-probability scales for the three ranges of average annual precipitation: (1) P = 32.5-130 inches; (2) P = 17.0-32.5 inches; and (3) P = 0.0-17.0 inches.

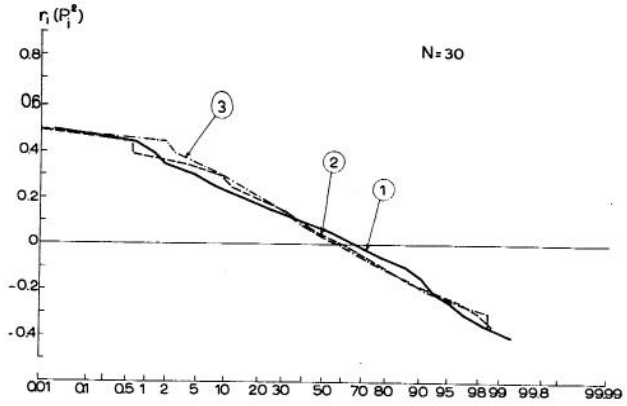

Fig. 31 Distributions of the first serial correlation coefficient for P_i^2-series (annual precipitation with nonhomogeneous and/or inconsistent data) of the second large sample of precipitation stations (n = 473) in Western North America from the simultaneous observations of 1931-1960 with N = 30 on cartesian-probability scales for the three ranges of average annual precipitation: (1) P = 32.5 - 130 inches; (2) P = 17.0-32.5 inches; (3) P = 0.0-17.0 inches.

TABLE 4

Statistical parameters or r_1-distributions of V-, P_e-, P_i^1-, and P_i^2-series, for three ranges of specific water yield or average annual precipitation for each series, and for longest or simultaneous period of observation, as given in Figs. 24 through 31.

Series	Number of Stations	Average size of Series	Range q in cfs/sq. mi. or P in inches	Number of stations per range	Average $\bar{r}_1$	Standard deviation $s(r_1)$	Skewness Coefficient C_{sr}	Kurtosis k_r
V	446	37	1.4 - 10.6	146	0.217	0.198	0.134	3.4
			0.5 - 1.4	155	0.165	0.154	0.691	4.4
			0.0 - 0.5	145	0.172	0.168	1.000	4.2
		30 (1931-1960)	1.4 - 10.6	146	0.178	0.218	0.178	3.6
			0.5 - 1.4	155	0.123	0.166	0.963	5.5
			0.0 - 0.5	145	0.165	0.168	0.699	4.0
P_e	446	37	1.4 - 10.6	146	0.193	0.202	0.353	3.4
			0.5 - 1.4	155	0.160	0.156	0.826	4.9
			0.0 - 0.5	145	0.187	0.189	0.457	2.9
		30 (1931-1960)	1.4 - 10.6	146	0.165	0.206	0.225	3.0
			0.5 - 1.4	155	0.114	0.162	1.019	4.9
			0.0 - 0.5	145	0.163	0.193	0.509	3.2
P_i^1	1141	54	27.0 -180.0	380	0.070	0.136	0.243	2.8
			16.5 - 27.0	381	0.050	0.140	0.225	3.0
			0.0 - 16.5	380	0.047	0.147	0.136	3.2
		30 (1931-1960)	27.0 -180.0	380	0.043	0.154	-0.158	3.3
			16.5 - 27.0	381	0.029	0.167	0.202	3.0
			0.0 - 16.5	380	0.016	0.171	0.116	2.7
P_i^2	473	57	32.5 -130.0	156	0.062	0.138	-0.265	4.5
			17.0 - 32.5	160	0.069	0.147	0.429	2.8
			0.0 - 17.0	157	0.064	0.154	0.302	2.9
		30 (1931-1960)	32.5 -130.0	156	0.051	0.156	-0.426	3.7
			17.0 - 32.5	160	0.046	0.173	-0.017	2.5
			0.0 - 17.0	157	0.047	0.180	0.141	2.6

The results of a comparison of the r_1-distributions in the eight figures are summarized in Table 4. Each range contains approximately one third of the sample. Figures 24 through 31 show clearly that the r_1-values are greater on the average for a greater range of either specific water yield or the average annual precipitation. Statistical parameters of r_1-distributions of Fig. 24-31 show that for nearly all eight figures except for P_i^2-series and $N_m = 57$) the greatest $\bar{r}_1$ values are for the greatest specific water yields (q) or the greatest average annual precipitation (P). The $\bar{r}_1$-values decrease with a decrease of either q or P. The average value of the first serial correlation coefficient for a sample is greater for a more humid region. Comparisons of the average values of $\bar{r}_1$ among the high ranges of V-, P_e-, P_i^1-, and P_i^2-series, and the medium and the low ranges of the same series show clearly that $\bar{r}_1$-values decrease from V-series to P_e-series, and especially from P_e-series to P_i^1-series. The P_i^2-series usually has a greater $\bar{r}_1$-value for most of the ranges than the P_i^1-series.

The above conclusion may serve as an indirect test of the effect of evaporation in the air on the dependence in the series of annual precipitation at the ground. In Part I [1, pages 15-16] mathematical expressions have been developed for the evaporation of precipitation in the air between the cloud base and the ground. Based on that derivation, the hypothesis was advanced in Part I that the annual evaporation of precipitation in the air is a physical factor which depends on the humidity of the air, and this in turn depends on the annual evaporation from the ground and water surfaces in a river basin and from the areas around it. Since the annual evaporation from the ground is affected by water carryover from previous years, the annual evaporation of precipitation in the air is also dependent on the carryover, but less than either the annual evaporation from the ground or the annual runoff from a river basin.

The direct test of the effect of evaporation of raindrops in the air on the dependence in series of annual precipitation at the ground cannot be carried out because of the lack of data on the precipitation at the cloud base for large number of stations. However, the above indirect test may show that there is an effect by this evaporation in the air upon the dependence in the series of annual precipitation at the ground.

The above analysis of the first serial correlation coefficient of the ranges of the specific water yield and average annual precipitation is, in fact, the analysis of the differences in r_1-distributions between humid, moderately humid, and arid regions. The fact is that humid regions have an annual precipitation at the ground which has a higher value of $\bar{r}_1$ than arid regions. This points out that the greater dependence must be a result of a physical factor which has a greater influence on the time series of precipitation in humid regions than in arid regions. The most attractive physical factor seems to be the evaporation of precipitation in the air between the cloud base and the ground. The water carryover from year to year is usually greater in absolute value and per unit area in humid regions than in arid regions. The impact of evaporation from the ground upon the evaporation of precipitation in the air should be somewhat greater in humid regions than in arid regions because of the water carryover from year to year. Thus the hypothesis advanced in Part I may have some support from the above analysis and discussion. However, this hypothesis must wait for sufficient data from precipitation measurements at the cloud base in order to be proved or disproved.

Table 4 shows that the standard deviation of r_1-distributions is also the greatest for the high range (or humid regions) of specific water yield for V- and P_e-series. For annual precipitation or P_i^1- and P_i^2-series the greatest values of standard deviation are for the most arid regions. However, the differences in standard deviations are so small between ranges in both cases that neither explanation nor hypothesis is advanced here about these differences.

The skewness coefficient of r_1-distributions shows a clear pattern between the ranges for V- and P_e-series. The humid regions have the smallest C_{sr} values for all cases of V- and P_e-series. For annual precipitation or P_i^1- and P_i^2-series there is no clear pattern in C_{sr} coefficient although one third of them are negative and all others are relatively small in comparison with the V- and P_e-series. Similar conclusions may be derived for the kurtosis k_r of the r_1-distributions. For V- and P_e-series the average values of k_r for 12 cases are greater than three, while for 12 cases of P_i^1- and P_i^2-series the average value of k_r is close to three. It can be concluded that the r_1-distributions for the three ranges of P_i^1- and P_i^2-series are much closer to the normal distribution than are the r_1-distributions for the three ranges of V- and P_e-series.

G. CONCLUSIONS

From the preceding analysis by serial correlation of the four large samples of series of annual river flow, annual effective precipitation, and annual precipitation at the ground some basic conclusions may be derived.

The water carryover from year to year, and especially the constant change in this carryover from year to year is the basic physical factor in time dependence of the annual river flow, annual effective precipitation and annual precipitation.

This change in carryover affects the dependence in three basic ways:

(1) The part of the carryover which flows out of river basins by surface or underground runoff is the basic casual factor for the annual river flow having on the average a greater time dependence than the annual effective precipitation;

(2) The part of the carryover which goes into the atmosphere through evaporation and evapotranspiration from the surface is the basic casual factor for the annual effective precipitation having on the average a much greater time dependence than the annual precipitation at the ground; and

(3) The part of the carryover which goes into the atmosphere through evaporation and evapotranspiration from the surface of a river basin and adjacent basins is likely to be the basic casual factor for the annual precipitation at the ground having on the average a somewhat greater time dependence than the annual precipitation at the cloud base. This last statement should be understood to be more a hypothesis than a final conclusion.

Serial correlation analysis, with appropriate statistical tests, has been used with success to detect some patterns in the sequence of annual river flow, annual effective precipitation and annual precipitation at the ground. There is no statistical or other evidence that any cyclic (or deterministic) movement exists in the sequences of these variables. There is no evidence, at least by the statistical tests used, that the average sun-spot cycle affects significantly the fluctuation of wet and dry years of runoff and rainfall. The sequence of wet and dry years may be considered as a pure stochastic process. The most likely general stochastic mathematical model fitting the time dependence of the above series is the moving average scheme. Among the various mathematical models for the moving average scheme the first and the second order linear autoregressive models, or Markov linear models, have been shown to fit well the stochastic process of the time series of annual flow of river basins with substantial water carryover from year to year.

Apart from the physical factor of water carryover, the inconsistency and nonhomogeneity of the data are shown to be factors which increase on the average the time dependence. The jumps and trends created by nonhomogeneity and/or inconsistency in data increase the time dependence.

The serial correlation analysis further supports the conclusion made in Part I (1, page 26) that the causal factors of time dependence for annual flow and annual precipitation should be analyzed and accounted for before attempts are made to search for causal factors of this time dependence in the upper atmosphere, in oceans, and in solar and cosmic activities.

REFERENCES

1. Yevdjevich, V. M., Fluctuations of wet and dry years, Part I, Research data assembly and mathematical models: Colorado State University Hydrology Paper No. 1, July, 1963.

2. Wold, Herman, A study in the analysis of stationary time series, second edition, with an appendix by Peter Whittle: Almquest and Wiksell, Stockholm, 1954.

3. Dixon, Wilfrid, Further contributions to the problem of serial correlation: Annals of Math. Statistics, v. 14, no. 2, June, 1944, pp. 119-144.

4. Koopmans, Tjalling, Serial correlation and quadratic forms in normal variables: Annals of Math. Statistics, v. 13, 1942, pp. 14-33.

5. Anderson, R. L., Distribution of the serial correlation coefficients: Annals of Math. Statistics, v. 8, no. 1, Mar. 1941, pp. 1-13.

6. Fisher, R. A., Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population: Biometrica, v. 10, 1915, pp. 507-521.

7. Bartlett, M. S., On the theoretical specification of sampling properties of auto-correlated time series: Royal Stat. Soc. Suppl., v. 8, 1946, pp. 27-41.

8. Quenouille, M. H., The joint distribution of serial correlation coefficients: Annals of Math. Statistics, v. 20, no. 4, December, 1949, pp. 561-571.

9. Madow, G. W., Note on the distribution of the serial correlation coefficient: Annals of Math. Statistics, v. 16, no. 3, September, 1945, pp. 308-310.

10. Leipnik, R., Distribution of the serial correlation coefficient in circularly correlated universe: Annals of Math. Statistics, v. 18, 1947, pp. 80-87.

11. Quenouille, M. H., Some results in the testing of serial correlation coefficients: Biometrica, v. 35, 1948, pp. 261-267.

12. Quenouille, M. H., Approximate tests of correlation in time series 3: Proc. Cambridge Phil. Soc., v. 45, 1949, part 3, pp. 433-484.

13. White, J. S., Approximate moments for the serial correlation coefficient: Annals of Math. Statistics, v. 28, 1957, pp. 798-802.

14. White, J. S., A t-test for the serial correlation coefficient: Annals of Math. Statistics, v. 28, 1957, pp. 1046-1048.

Key Words: Hydrology, Time Series, Serial Correlation Analysis, Patterns in Sequence, Flow Sequence, Precipitation Sequence, Wet and Dry Years, Annual Flows, Annual Precipitation.

Abstract: Four large samples of annual river flow, annual effective precipitation and annual precipitation are investigated by serial correlation analysis. Statistical techniques and final expressions are given for the serial correlation analysis in a summary form. The carryover of water from year to year in river basins, which is disposed of in successive years either by river runoff or by evaporation and evapotranspiration is the main factor of time dependence in series of annual river flow, annual effective precipitation, and annual precipitation at the ground. A factor worthwhile for further study is the nonhomogeneity and/or inconsistency in hydrologic data of river flow and precipitation.

Reference: Yevdjevich, V. M., Colorado State University, Hydrology Papers No. 4 (June 1964) "Fluctuations of Wet and Dry Years, Part II, Analysis by Serial Correlation"

Key Words: Hydrology, Time Series, Serial Correlation Analysis, Patterns in Sequence, Flow Sequence, Precipitation Sequence, Wet and Dry Years, Annual Flows, Annual Precipitation.

Abstract: Four large samples of annual river flow, annual effective precipitation and annual precipitation are investigated by serial correlation analysis. Statistical techniques and final expressions are given for the serial correlation analysis in a summary form. The carryover of water from year to year in river basins, which is disposed of in successive years either by river runoff or by evaporation and evapotranspiration is the main factor of time dependence in series of annual river flow, annual effective precipitation, and annual precipitation at the ground. A factor worthwhile for further study is the nonhomogeneity and/or inconsistency in hydrologic data of river flow and precipitation.

Reference: Yevdjevich, V. M., Colorado State University, Hydrology Papers No. 4 (June 1964) "Fluctuations of Wet and Dry Years, Part II, Analysis by Serial Correlation"

Key Words: Hydrology, Time Series, Serial Correlation Analysis, Patterns in Sequence, Flow Sequence, Precipitation Sequence, Wet and Dry Years, Annual Flows, Annual Precipitation.

Abstract: Four large samples of annual river flow, annual effective precipitation and annual precipitation are investigated by serial correlation analysis. Statistical techniques and final expressions are given for the serial correlation analysis in a summary form. The carryover of water from year to year in river basins, which is disposed of in successive years either by river runoff or by evaporation and evapotranspiration is the main factor of time dependence in series of annual river flow, annual effective precipitation, and annual precipitation at the ground. A factor worthwhile for further study is the nonhomogeneity and/or inconsistency in hydrologic data of river flow and precipitation.

Reference: Yevdjevich, V. M., Colorado State University, Hydrology Papers No. 4 (June 1964) "Fluctuations of Wet and Dry Years, Part II, Analysis by Serial Correlation"

Key Words: Hydrology, Time Series, Serial Correlation Analysis, Patterns in Sequence, Flow Sequence, Precipitation Sequence, Wet and Dry Years, Annual Flows, Annual Precipitation.

Abstract: Four large samples of annual river flow, annual effective precipitation and annual precipitation are investigated by serial correlation analysis. Statistical techniques and final expressions are given for the serial correlation analysis in a summary form. The carryover of water from year to year in river basins, which is disposed of in successive years either by river runoff or by evaporation and evapotranspiration is the main factor of time dependence in series of annual river flow, annual effective precipitation, and annual precipitation at the ground. A factor worthwhile for further study is the nonhomogeneity and/or inconsistency in hydrologic data of river flow and precipitation.

Reference: Yevdjevich, V. M., Colorado State University, Hydrology Papers No. 4 (June 1964) "Fluctuations of Wet and Dry Years, Part II, Analysis by Serial Correlation"

Key Words: Hydrology, Time Series, Serial Correlation Analysis, Patterns in Sequence, Flow Sequence, Precipitation Sequence, Wet and Dry Years, Annual Flows, Annual Precipitation.

Abstract: Four large samples of annual river flow, annual effective precipitation and annual precipitation are investigated by serial correlation analysis. Statistical techniques and final expressions are given for the serial correlation analysis in a summary form. The carryover of water from year to year in river basins, which is disposed of in successive years either by river runoff or by evaporation and evapotranspiration is the main factor of time dependence in series of annual river flow, annual effective precipitation, and annual precipitation at the ground. A factor worthwhile for further study is the nonhomogeneity and/or inconsistency in hydrologic data of river flow and precipitation.

Reference: Yevdjevich, V. M., Colorado State University, Hydrology Papers No. 4 (June 1964) "Fluctuations of Wet and Dry Years, Part II, Analysis by Serial Correlation"

There is no statistical evidence that cycles exist in river flow or precipitation time series beyond the astronomic cycle of the year. Moving average schemes in general and the first and second order autoregressive schemes (Markov linear mathematical models) in particular fit sufficiently well the patterns in the sequence of annual river flows of river basins with large water carryover.

There is no statistical evidence that cycles exist in river flow or precipitation time series beyond the astronomic cycle of the year. Moving average schemes in general and the first and second order autoregressive schemes (Markov linear mathematical models) in particular fit sufficiently well the patterns in the sequence of annual river flows of river basins with large water carryover.

There is no statistical evidence that cycles exist in river flow or precipitation time series beyond the astronomic cycle of the year. Moving average schemes in general and the first and second order autoregressive schemes (Markov linear mathematical models) in particular fit sufficiently well the patterns in the sequence of annual river flows of river basins with large water carryover.

There is no statistical evidence that cycles exist in river flow or precipitation time series beyond the astronomic cycle of the year. Moving average schemes in general and the first and second order autoregressive schemes (Markov linear mathematical models) in particular fit sufficiently well the patterns in the sequence of annual river flows of river basins with large water carryover.

PREVIOUSLY PUBLISHED PAPERS

Colorado State University Hydrology Papers

No. 1. "Fluctuations of Wet and Dry Years, Part I, Research Data Assembly and Mathematical Models," by Vujica M. Yevdjevich, July 1963.

No. 2. "Evaluation of Solar Beam Irradiation as a Climatic Parameter of Mountain Watersheds," by Richard Lee, August 1963.

No. 3. "Hydraulic Properties of Porous Media" by R. H. Brooks and A. T. Corey, March 1964.

Colorado State University Fluid Mechanics Papers

No. 1. "A Resistance Thermometer for Transient Temperature Measurements," by J. L. Chao and V. A. Sandborn, March 1964.

No. 2. "Measurement of Turbulence in Water by Electrokinetic Transducers," by J. E. Cermak and L. V. Baldwin, April 1964.