0.12
19r
961
no.54-62

SOME GENERAL ASPECTS OF FLUCTUATIONS OF ANNUAL RUNOFF IN THE UPPER COLORADO RIVER BASIN

by

Vujica M. Yevdjevich

Engineering Research
Colorado State University

CIVIL ENGINEERING SECTION
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO

October 1961

CER61VMY54

PAST AND PROBABLE FUTURE VARIATIONS IN STREAM FLOW IN THE UPPER COLORADO RIVER

I. Summary and Conclusions

Morris E. Garnsey,
Project Director

II. A Study of the Statistical Predictability of Stream Runoff in the Upper Colorado River Basin

Paul R. Julian,
Research Staff,
High Altitude Observatory,
University of Colorado

III. Some General Aspects of Fluctuations of Annual Runoff in the Upper Colorado River Basin

Vujica M. Yevdjevich,
Engineering Research,
Colorado State University

IV. Probability Analysis Applied to the Development of a Synthetic Hydrology for the Colorado River

Margaret R. Brittan,
Assistant Professor of Statistics,
University of Denver

V. Analysis of Precipitation Data in the Upper Colorado River Basin

Richard A. Schleusener,
Engineering Research,
Colorado State University
and
Loren W. Crow,
Consulting Meteorologist
Denver, Colorado

SPONSORS

This project was sponsored by the State of Colorado, supported by it and by the States of New Mexico, Utah and Wyoming, acting through the Upper Colorado River Commission

SOME GENERAL ASPECTS OF FLUCTUATIONS OF ANNUAL RUNOFF IN THE UPPER COLORADO RIVER BASIN

by

Vujica M. Yevdjevich
Engineering Research
Colorado State University

Civil Engineering Section
Colorado State University
Fort Collins, Colorado

October 1961

CER61VMY54

U18401 3297852

TABLE OF CONTENTS

ABSTRACT

The fluctuations of annual river flows and derived effective annual precipitations (defined as precipitation minus evapotranspiration on an area) of the Upper Colorado River Basin are subjects of this study. The data for this study are composed of annual flows for 14 river stations in the Upper Colorado River Basin and around it, and of the annual flows of the key station of the Upper Colorado River, at Lee Ferry, Arizona.

Three statistical techniques were used: 1) distribution of the first serial correlation coefficient; 2) correlogram analysis; and 3) distribution of range. The series of annual flows and series of derived effective annual precipitations were compared with the same statistical characteristics of random time series. The comparison of these statistical parameters shows that the fluctuations of effective annual precipitations on river basins are very close to the fluctuations of random series.

Probability characteristics of random variables may be applied to annual river flows of the Upper Colorado River, provided due corrections are made for the effect of water carryover from year to year, and for non-homogeneity in data.

A discussion of hydrological characteristics of annual runoff in the Upper Colorado River Basin is given, with general conclusions of this study and some recommendations for further studies.

ACKNOWLEDGEMENTS

The author wishes to thank the sponsors of this study, the State of Colorado and the other states of the Upper Colorado River Basin, for the financial support which has made this study possible.

A special vote of thanks is due to the U. S. Bureau of Reclamation, the U. S. Geological Survey, Water Resources Division, and the Colorado Water Conservation Board for their assistance in supplying hydrologic data and other information.

I. INTRODUCTION

A. SUBJECT AND OBJECTIVE OF THE STUDY

The subject of this study is the fluctuations of annual runoff and derived effective annual precipitation* (adjusted annual runoff, or net yield of atmosphere) of the Upper Colorado River Basin. The question to be answered by this study could be summarized as follows: Are there any regular patterns** in the sequence of annual flows and effective annual precipitation in the Upper Colorado River Basin, and if there are, how can they be explained by physical factors or some other known causes?

The reasons why the annual river flows of the Upper Colorado River Basin were studied in this report, and not the monthly flows or even the daily flows, comes from the main objective of the study, namely to inquire whether there are any regularities or patterns in the fluctuation of annual runoff. If the patterns were clearly and cyclically regular, the predictability of annual runoff by extrapolation in the future of these regular patterns determined from the past observations for the Upper Colorado River Basin at its key station at Lee Ferry, Arizona would be feasible. If random (or chance) processes in the sequence of annual flows predominate, the possibility or outlook of making long-range forecasts of expected runoff in acre-feet (for calendar periods say of five or ten years) by extrapolation is small or negligible. In this last case, however, probability methods should be used for the estimate of expected flows and their probability instead of extrapolating regular fluctuations from the past in the future. In other words, if randomness predominates in the sequence of annual flows on the border between the Upper and the Lower Colorado River Basins, then the study of the flow availability and the flow regulation by large reservoirs as Lake Mead, or the future reservoir at Glenn Canyon, should be carried out by using probability methods. The same should be applied for the other large water storage and other large water resources projects in the Upper Colorado River Basin for which the availability of successive annual flows is important information for project planning, design and operation.

* The effective annual precipitation is defined as difference for any water year between the total precipitation and the total evapotranspiration on a given river basin. It can be conceived as the net water yield of the atmosphere for a river basin.

** Under regular patterns any dependence in consecutive values of a time series is understood here, or any regular difference between the observed or derived time series and the random time series.

B. MAIN APPROACH

The main approach in this study is not to look for regular or combined different hidden periodicities in the runoff of the Upper Colorado River Basin, but the opposite, to compare the fluctuations of the available annual runoff samples with random series* or with time series created by stochastic** processes. Differences were determined between the observed fluctuations of runoff and fluctuations of effective precipitation, and they were compared with the fluctuations of random time series or series derived from random time series by some known processes.

If there would be sufficient and known physical causes for explaining one part of differences between the random time series and the observed time series of annual flows or derived time series of annual effective precipitations, then the question arises (after taking into account the effect of those physical factors) what room is left for the other causes which could influence the dependence in annual flow fluctuations. It is of a particular interest to see what room is left for causes outside the river basin itself (besides the effect of over-year storage of water in river basin, man's effect through the changes in river basin) and especially for causes from atmosphere (meteorological effects on dependence of consecutive values of effective annual precipitations), and for causes beyond the atmosphere (solar activities, cosmic effects).

C. DATA USED

The basic data to carry out the investigation, and to derive conclusions from it, are:

1. In a recent study by the author of this paper (ref. 1)*** an analysis was made of annual runoff data from 72 selected stations in the United States, and 13 stations in Canada, for the purpose of studying the fluctuations of annual river flows on a world-wide basis. These stations were the part of a total of 140 river gaging stations, gathered from many

* The term random series is defined here in the classical way, that there is no systematic link between successive values of an infinite time series, or that the consecutive values of the time series are independent.

** The term stochastic processes is applied here in the sense that non-random series are derived from random time series by some processess as moving average procedure or similar.

*** References are given at the end of this paper.

parts of the world, with 40-150 years of observations (ref. 1). The general results of the earlier analysis of data for these 140 stations will be used in this study also, to check and supplement the results derived from the stations in the Upper Colorado River Basin and around it.

2. An analysis was made of the data from 14 river gaging stations (see table 1) in the Upper Colorado River Basin (4 stations), and around it (10 stations). These 14 stations are the part of the total of 72 stations in the United States used for the general study of annual flow fluctuations.

3. The data available on annual flows from 1896 to 1959 of the Upper Colorado River at Lee Ferry Station, Arizona.

The river gaging stations, four stations inside the Upper Colorado River Basin, and ten stations around it, were not especially selected for this study, but rather were selected as a part of 140 stations for study of reference 1 , because they were considered as the most homogeneous stations as possible to find in the region, within the framework of 72 selected stations from the United States. At the time of selection of these stations, the potential specific study of the Upper Colorado River Basin was not considered. If a study of this type should be done on a more extensive basis for the Upper Colorado River Basin, it would be feasible to include several other stations, and also of the stations with shorter records than forty years. It might be that analysis of data of about 30 to 40 stations from the Upper Colorado River Basin and around it could give somewhat different results. Although the 14 stations constitute a small sample, it is believed that the analysis gives the general characteristics of fluctuations of annual river flows inside and around the Upper Colorado River Basin. On the other hand, the analysis of 20 to 30 additional river gaging stations inside the same region would probably not contribute much new additional information, due to high correlation of annual flows between adjacent stations.

D. PROCEDURES

The procedure followed in studying the patterns in fluctuations of annual flows in the Upper Colorado River Basin was based on the following principles and programs:

1. The discrete time series of annual river flows, and of derived effective annual precipitations on the corresponding river basins were used as the two basic time series for the determination of flow patterns. The effective annual precipitation is defined as the difference between the total annual precipitation and the total annual evapotranspiration on a river basin for each water year. The water year starting on the first of October (and carrying the name of the year for which nine months are included in the water year) was used throughout the study.

2. The effective annual precipitation, as will be shown later, is obtained by adding to the annual measured flow the difference of the total water stored in the river basin at the end and at the beginning of a water year. This difference can be positive (wet years) or negative (dry years).

3. The fluctuation of random variables in time was used as bench-mark time series. A random time series was thus used as a yardstick for comparing it with the time series of observed annual flows and derived effective annual precipitations, where the effective annual precipitations were determined from the annual flows and water carryover from one year to the other.

4. The following statistical parameters were used in the study:

a. Serial correlation coefficients (correlation coefficients of successive values of a time series with different lags between successive values) and correlograms (a graph relating the serial correlation coefficients to the lag between successive values of time series) which give a simple method of studying the dependence in successive terms of a time series; they are convenient when the absolute values of serial correlation coefficients are small, which is the case usually with time series of annual river flows and effective annual precipitations;

b. Range, defined as the difference between previous maximum and previous minimum on the cumulative curve of departures of annual flows from the average annual flow (or departures between effective annual precipitations and the average effective annual precipitation).

The use of harmonic analysis, based on Fourier series either on classical patterns of harmonic analysis or by power spectrum analysis, is considered less feasible than, or of the same value as, serial correlation due to two facts: the serial correlation of time series studied is generally of a small absolute value, and the fluctuation of a random variable is the yardstick in this study.

5. Physical factors and other known causes which affect the patterns of flow fluctuations, such as inconsistency and non-homogeneity in data, were studied in order to explain some of the differences between the observed time series and the random time series.

6. Analysis of flow characteristics and fluctuation patterns of the Upper Colorado River Basin at the gaging station Lee Ferry, Arizona, is carried out by using a depletion model in order to derive the expected actual and future annual flows (mean and fluctuations around the mean).

II. DETERMINATION OF EFFECTIVE ANNUAL PRECIPITATION

A. FIRST APPROXIMATION DETERMINATION OF EFFECTIVE ANNUAL PRECIPITATION

Time series of annual volumes of river flows and of effective precipitation were used for the analysis of flow patterns.

The effective annual precipitation is defined as:

$$P_e = P_i - E_1 = V_i + W_e - W_b = V_i - \Delta W_i \quad (1)$$

where

P_e = effective annual precipitation for a given water year and a river basin (net available water in each water year for a river basin, or net input of water into a basin);

P_i = total annual precipitation on that river basin and for a given water year;

E_i = annual evapotranspiration (total water losses from the basin, from surface and underground waters, into the atmosphere);

V_i = annual volume of flow of a river for a given water year;

W_e = total stored water in a river basin at the end of a water year;

W_b = total stored water in a river basin at the beginning of the corresponding water year;

ΔW_i = difference of stored water for the end and beginning of a water year (positive in wet water years, negative in dry water years).

As there are usually great errors in determination of P_i and E_i for a river basin, the effective annual precipitation was obtained for this study exclusively by using W_e and W_b, or their difference ΔW_i. The value of ΔW_i was determined for the approximate values W_b and W_e.

The mean flow recession curve of average daily or monthly flows was determined for each station, and for the season around the end or beginning of water years, and they were approximated by exponential functions, either of the type $Q = Q_o e^{-ct}$, or of the type $Q = Q_o e^{-ct^n}$, with Q_o the initial discharge, Q (any discharge of recession curve) and t (time) as variables, and c, or c and n, as the parameters which characterize the mean flow recession curves. The integration of the above functions of mean recession curves, from given Q_o for $t = O$, to t = infinity gives the volume of water W, or the stored water in a river basin. In this case, W_b or W_e are functions (for given c, or for given c and n) of the discharge Q_o at the end of a water year, if this end happens to be during a recession curve period (for the United States and Canada the end was exclusively on the 30 September, and the beginning of the next water year on the 1 October). In case a flood wave (rising level especially) occurs at the end of a water year, a special procedure was followed to obtain the stored water volume W.

The illustration of this method of determining W is given in fig. 1 (Ashley Creek, Utah). The recession curve of monthly flows is a straight line in semi-log paper, with the average c-value equal 0.25×10^{-6} for the simple exponential function. In this case the integration of the volume from $t = O$ and Q_o to t = infinity gives $W = Q_o/c$, or in this example $W = 4.0 \times 10^6\ Q_o$, with Q_o = discharge on the 1 October. By using eq. (1) for each water year, the effective annual precipitation was computed for all river gaging stations.

The effective annual precipitation represents the net input of water in a river basin for a water year, or this is the water volume after the storage effects of the surface and underground reservoirs and other storage capacities in a river basin were excluded. As they represent annual flows without a carryover of water from year to year, the interdependence among successive values of effective annual precipitations should be smaller than for the annual flows, as will be shown later.

B. DETERMINATION OF TRUE VALUES OF EFFECTIVE ANNUAL PRECIPITATION

The true value of the effective annual precipitation is, however,

$$P_t = V_i - \Delta W_i \pm (E_i + I_i + H_i + G_i) \quad (2)$$

where

E_i = random error in annual flows (usually small for annual river flows);

I_i = inconsistency or systematic errors, produced by measurements and computation techniques, in general given as trends or jumps, or their combination;

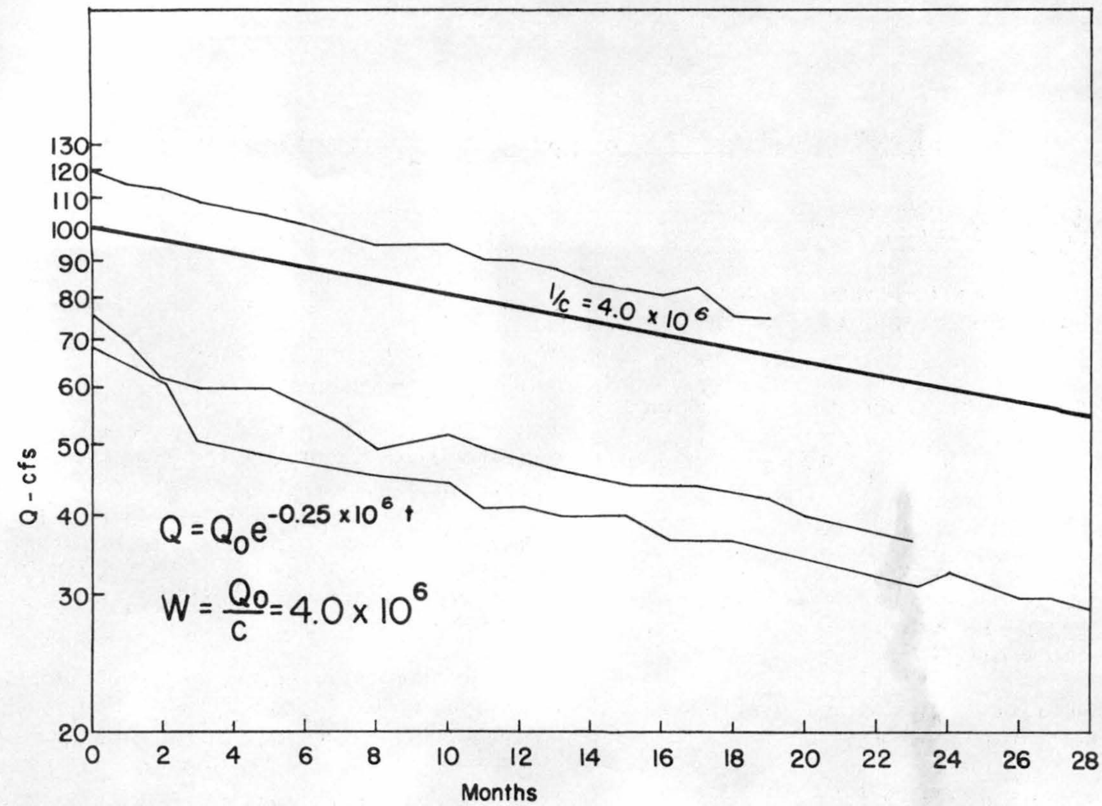


Fig. 1. Determination of the mean recession curve of Ashley Creek, Utah, by using monthly flows, for the purpose of compiling the water carryovers from one water year to another.

H_i = non-homogeneity or change in the virgin flow (virgin flow is defined as the flow uninfluenced by man's activities), which is produced by different man-made influences in the river basin (in general this is man-made depletion or increase of flow, of different type and patterns in time);

G_i = error in determining ΔW_i approximately by using in this study the average recession curves, instead of a recession curve for each individual water year and gaging station.

The experience shows that the influence of I_i, H_i, and G_i is relatively large, and that these factors and errors cannot be neglected in treating the fluctuations of annual flow and of effective annual precipitation.

In some cases the patterns in fluctuations of flows determined as dependence of successive values of time series could be explained by the inconsistency and non-homogeneity existing in data. The sufficient accuracy of computed members of the two time series V_i and ΔW_i should not be assumed without a study of it. If the effect of errors, inconsistency and non-homogeneity were currently neglected, without proving that it is justified, the derived dependence in time series may be highly biased.

III. PATTERNS IN FLUCTUATIONS, MEASURED BY FIRST SERIAL CORRELATION COEFFICIENT

A. FREQUENCY DISTRIBUTION OF FIRST CORRELATION COEFFICIENT

1. Procedure

To determine more clearly the general characteristics of annual flow patterns in the Upper Colorado River Basin, the characteristics of some river gaging stations in the Upper Colorado and around it were compared with the results of general characteristics of a much greater sample, namely with the results of 140 river gaging stations (ref. 1), from which 85 stations are in the United States and Canada (included are the 14 stations of the Upper Colorado, and around it).

The first serial correlation coefficient, defined as the correlation coefficient of successive pairs of annual flows or effective annual precipitations, were used as a measure of dependence of successive values in the two time series (flow and precipitation), or as an index of the possible flow fluctuation patterns. The unbiased first serial correlation coefficient is given in classical statistical books as:

$$r_1 = \frac{\sum_1^{N-1} U_i U_{i+1} - \frac{1}{N-1} \sum_1^{N-1} U_1 \sum_1^{N-1} U_{i+1}}{(N-2)\, s_i s_{i+1}} \qquad (3)$$

where U_i denotes any member of annual flow time series (Y_i symbol is used here for the effective annual precipitations), U_{i+1} is the next member to U_i, so that (U_i, U_{i+1}) represents the successive pairs of members of the time series, N = total number of members in a time series, $N-1$ is total number of correlated pairs, s_i is the standard deviation of $(N-1)$ - first members, and s_{i+1} is the standard deviation of $(N-1)$ - last members of the time series, with r_1 = first serial correlation coefficient.

The unbiased standard deviation is given in standard books of statistics as:

$$s_i = \left[\frac{1}{N-2} \sum_1^{N-1} U_i^2 - \frac{1}{(N-2)(N-1)} \left(\sum_1^{N-1} U_i\right)^2\right]^{1/2} \qquad (4)$$

For s_{i+1}, U_i is replaced by U_{i+1}, and as the difference between s_i, or s_{i+1} and (standard deviation for all N members) is very small, it is justified to write the unbiased standard deviation as:

$$s_i \approx s_{i+1} \approx s = \left[\frac{1}{N-1} \sum_1^{N} U_i^2 - \frac{1}{N(N-1)} \left(\sum_1^{N} U_i\right)^2\right]^{1/2} \qquad (5)$$

However, as the digital computer was used to determine the first serial correlation coefficients for all 140 stations (ref. 1), the Upper Colorado stations included, s_i and s_{i+1} through eqs. (3) and (4) were used in this computation. The r_1-values were determined for both series: U (annual river flows), and Y (effective annual precipitations on the corresponding river basins), with U- and Y- series generally expressed as the modular coefficients or relative values (absolute values divided by the mean), so that all members of series oscillate around the unity, as dimensionless numbers.

For a pure random time series (fluctuation of a random variable), considered as an open series (or that N-1 pairs are used for the computation of the first serial correlation coefficient) in contrast with a circular time series for which the last term is supposed to be succeeded by the first term of the series (in which case there are N pairs), R. L. Anderson (ref. 2) gives the expected value with the symbol $E(r_1)$ (the mean value of r_1-distribution) of the first serial correlation coefficient for circular time series as:

$$E(r_1) = -\frac{1}{N-1} \qquad (6)$$

and of the variance (square of standard deviation) of first serial correlation coefficient distribution.

$$\text{var } r_1 = \frac{N-2}{(N-1)^2} \qquad (7)$$

which both converge toward zero by an increase of N. For $N \geq 40$ the open time series gives r_1-values close to those of circular series. As the minimum of years is 40 ($N_{min} = 40$), the maximum value of $E(r_1)$ is - 0.0256, which is practically close to zero. The standard deviation s_r of r_1 is maximum for N minimum, and in this case of $N_{min} = 40$ it is $s_r(\max) = 0.158$, which is a relatively high value. For the maximum value of N = 150 years in the study of 140 stations (ref. 1), $s_r(\min) = 0.082$, also a rather large value. Therefore, it is to be expected that the values of r_1 for many stations, under the assumption of random fluctuations, would cover a relatively large range, both positive and negative values, with the mean close to zero.

The 14 river gaging stations were selected in the Upper Colorado River Basins (4 stations), and very close to it in adjacent river basins (10 stations) for the study of patterns in annual flow fluctuations. The basic hydrologic characteristics and the statistical parameters for these 14 stations, for both annual flows (U-series), and effective annual precipitations (Y-series), are presented in table 1. The position of stations is shown in fig. 3. The data for all these stations (and other stations in the United States of America) were obtained from the Basic Data Section, Surface Water Branch, Water Resources Division, U. S. Geological Survey, Washington, D. C.

Table 1 shows many characteristics of these 14 stations: the basin areas vary from 1.0 sq. mi. to 6160 sq. mi.; the mean discharge varies from 0.563 cfs to 2987 cfs; the specific yields are from 0.126 to 1.51 cfs/sq. mi.; the length of period of observations vary from 42 to 59 years, with the mean length $N = 47$ approximately; the coefficients of variation (standard deviation divided by the mean) vary from 0.213 to 0.709 for U-series, and from 0.220 to 0.712 for Y-series; the C_v-values are always greater for Y-series (effective annual precipitation) than for U-series (annual flows), because the storage in river basins and water carryover from year to year act as the attenuators of extremes for the same mean annual values, or they decrease with a decrease of standard deviation.

The effective annual precipitations and the annual flows are connected in a type of moving average model (Markov chains) of the type

$$U_i = b_o Y_i + b_1 Y_{i-1} + b_2 Y_{i-2} + \dots + b_{m-1} Y_{i-m+1} \qquad (8)$$

with b_o, b_1, ... b_{m-1} being monotonically decreasing and positive coefficients, with $\sum_{o}^{m-1} b_i = 1$, where m theoretically should be infinite (recession exponential curve is the reason for that), but practically, m does not pass 10 (the case of annual flows of Saint-Lawrence River at Ogdensburg, New York). The properties of eq. (8) show that the standard deviation of U-series is smaller than that of Y-series under given conditions for b-coefficients, because the moving average attenuates the extremes of Y-series in producing U-series.

The C_v-values for U- and Y-series, and their relationship could be used for measuring the long term water carryover factor in a river basin, if there would be no other factors causing the dependence, but serial correlation coefficients are a more convenient tool for this purpose.

The skew coefficients C_s are very changeable for given 14 stations, because N-values are relatively small to produce a dependable third statistical moment (table 1).

The index of variability, I_v, defined as the standard deviation of logarithms of modular coefficients (slope of the cumulative distribution curve in log-probability scales, for a distribution which could be fitted by log-normal function), shows that in all cases I_v for Y-series is greater than I_v for U-series. This result is also derived from eq. (8), if Y_i-values are assumed to be log-normally distributed. The relation of I_v-values for U-series and those for Y-series could also be used as a measure for water carryover from year to year, if the carryover is the main cause for the dependence in a time series.

2. Results

The first serial correlation coefficients (table 1) show that their average value for Y-series (0.18 approximately) is smaller than the average value for U-series (0.22 approximately), which must be so, when the statistical moving average model of eq. (8) and the conditions for b-values are applied to Y-series to derive U-series.

The first serial correlation coefficients of $\log U_i$ (or $\log Y_i$) show the same patterns, as the first serial correlation coefficients of U_i (or Y_i), table 1. The differences between r_1-values for $\log U_i$-series and U_i-series (or for $\log Y_i$-series and Y_i-series) are relatively small, and either positive or negative. There is no clear pattern to be distinguished among two sets of r_1-values (U_i versus $\log U_i$, or Y_i versus $\log Y_i$).

Table 1 shows also that $\Delta r_1 = r_1(U) - r_1(Y)$, the difference between the first serial correlation coefficients of U-series and the first serial correlation coeficients of Y-series, increases with an increase of parameter W/A, given in feet, a relative value of basin storage, as the ratio of mean annual carryover to the area of river basin, for 14 stations in the Upper Colorado River Basin and around it. Generally the greater W/A the greater is the difference Δr_1.

A general trend can be also derived, when the specific yield of river basins, q in cfs/sq. mi., is related to the decrease Δr_1 of the first serial correlation coefficients from U-series to the corresponding Y-series. The smaller the specific yield, the greater is Δr_1, on the average. This can be easily explained by the fact that for given topographical and geological conditions in a river basin, for its given area, the available space for underground and surface storage of water in relation to average annual runoff is greater in dry climates (small specific yields) than in humid climates (great specific yields). The relation $\Delta r_1 = f(q)$ is not simple, due to the fact that topography, geology, vegetative cover, climatic conditions vary very much from one river basin to another.

The above conclusion on the relationship of statistical parameters reflects the same patterns as in the analysis of 140 stations, previously mentioned.

Figure 2 gives the cumulative frequency distribution for first serial correlation coefficient, plotted on arithmetic-probability paper for the following series:

a. First serial correlation coefficient r_1 (U) of U-series for N = 140 stations, selected from different parts of the world (ref. 1), as line (1);

b. First serial correlation coefficient r_1 (Y) of Y-series for N = 140 stations (ref. 1), as line (2);

c. First serial correlation coefficient r_1 for random time series, with the length of series being equal to $N_m = 55$, the mean length of observations for 140 stations (ref. 1), and using eqs. (6) and (7), with the mean first serial correlation coefficient $\bar{r}_1 = E(r_1) = -0.018$, and standard deviation $s_r = 0.135$, as line (3);

d. The line (3) shifted parallely to pass through the median value of r_1 (U), equal 0.16, of the distribution given under line (1), as line (4);

e. The line (3) shifted parallely to pass through the median value of r_1 (Y) equal 0.115, of the distribution given under line (2), as line (5);

f. First serial correlation coefficient for U-series of the 14 Upper Colorado stations (and around it), with the approximate mean value $\bar{r}_1$ (U) = 0.220, as line (6);

g. First serial correlation coefficient for Y-series of 14 Upper Colorado stations, with the approximate value $\bar{r}_1$ (Y) = 0.18, as line (7);

h. First serial correlation coefficient r_1 for random time series with the length of series being N = 47, equal to the mean length of observation of 14 stations of the Upper Colorado River (and around it), using eqs. (6) and (7), with $\bar{r}_1 = -0.0217$, and $s_r = 0.146$, as line (8);

i. The line (8) shifted parallely to pass through the approximate median value, equal 0.220, of U-series of the distribution given under line (6), as line (9), and

j. The line (8) shifted parallely to pass through the approximate median value, equal 0.18, of Y-series of the distribution given under line (7), as line (10).

3. Discussion of results and conclusions

From the comparison of the distributions represented on fig. 2 the following general discussion of results and the conclusions may be derived:

a. The cumulative frequency distributions of U-series and Y-series either for 140 or 14 stations (Upper Colorado) are very close to the straight lines (extremes excluded) on arithmetic-probability paper, so that a normal function can fit these distributions, or their first serial correlation coefficients are normally distributed.

b. The slopes of the lines, (1), (2), (6) and (7), which represent the standard deviation of first serial correlation coefficient distributions are also very close (at least for probabilities in the range 10-90 percent) to the slopes of cumulative distributions of first serial correlation coefficients, lines (3) and (8), in the case these random time series have the same lengths as the mean length for U-series and Y-series. In other words, the frequency distributions of first serial correlation coefficients for U- and Y-series have approximately the same standard deviations as the corresponding random time series.

c. The difference between the frequency distribution of first serial correlation coefficients for U- and Y-series, and the frequency distribution of random time series is only in the mean values of coefficients. The U- and Y-series, lines (1), (2), (6) and (7), have positive mean values of r_1, while the means of random series are negative, lines (3) and (8), but practically very close to zero. The lines (1), (2), (6) and (7), approximated by straight lines (4), (5), (9) and (10), seem as those of random time series but shifted for the difference of mean of first serial correlation coefficients for U- and Y-series, and the mean of those for corresponding random time series. These differences for 140 stations selected from many parts of the world, as $\Delta\bar{r}_1 = \bar{r}_1 - E(r_1)$, are

U-series: $\Delta\bar{r}_1$ (U) = 0.160 + 0.018 = 0.178

Y-series: $\Delta\bar{r}_1$ (Y) = 0.115 + 0.018 = 0.133

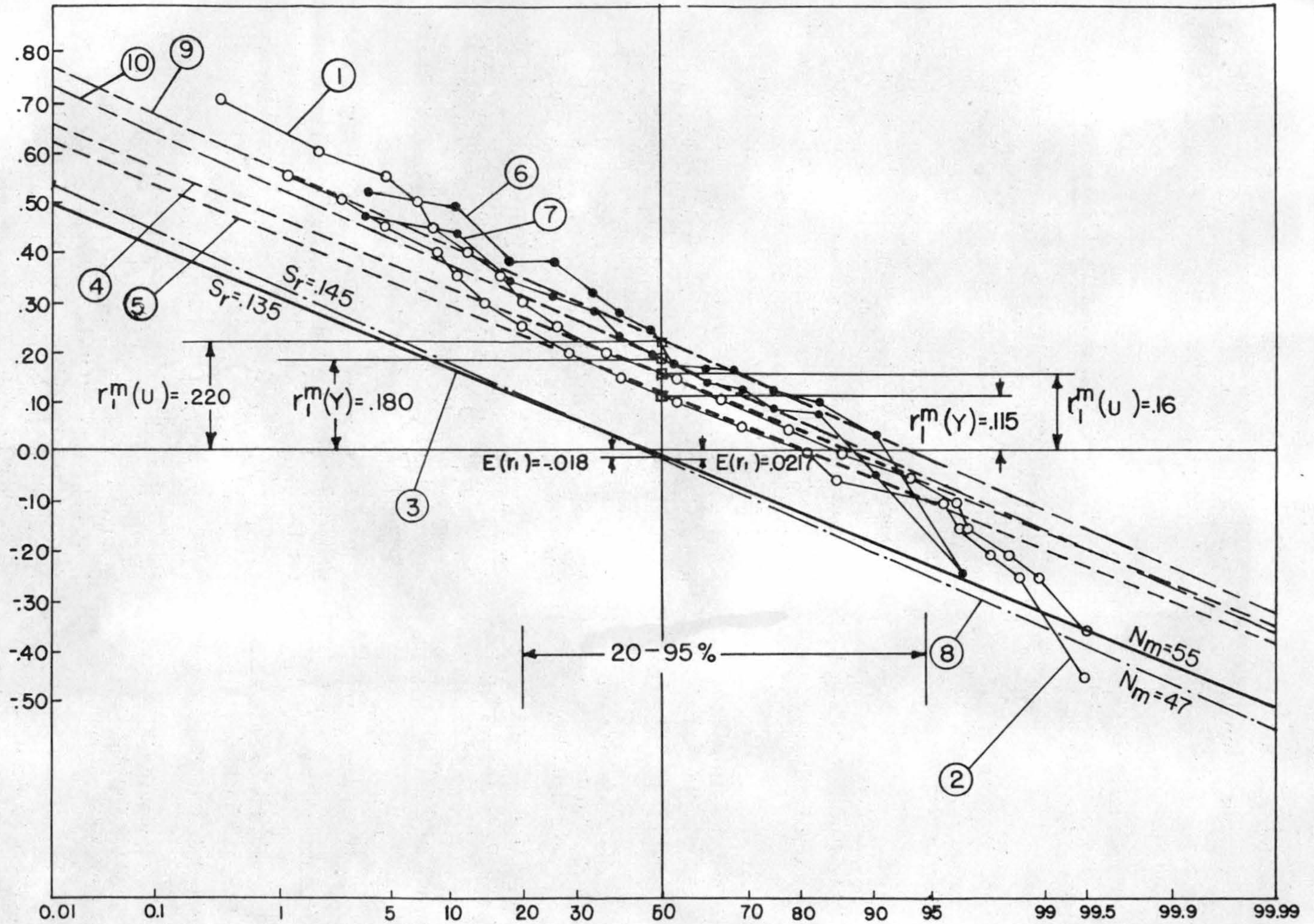


Fig. 2 Cummulative frequency distributions of first serial correlation coefficent: (1) For series of 140 stations and annual flows; (2) Same as under (1) but for effective annual precipitation; (3) For random series, $N_m=55$; (4) Line parallel to line (3) through $r=0.16$; (5) Same as (4) through $r=0.115$; (6) For series of 14 stations of Upper Colorado and annual flows; (7) Same as under (6) but for effective annual precipitation; (8) For random series, $N_m=47$; (9) Line parallel to line (8) through $r=0.22$; (10) Same as (9) through $r=0.18$.

and for 14 stations from the Upper Colorado River Basin, and around it, they are:

U-series: $\Delta\bar{r}_1$ (U) = 0.220 + 0.022 = 0.242

Y-series: $\Delta\bar{r}_1$ (Y) = 0.180 + 0.022 = 0.202

The main problem in detecting the patterns in the fluctuations of annual flows and of effective annual precipitations by using the frequency distributions of first serial correlation coefficients as a statistical tool is the interpretation and the analysis of these differences in the means or medians of first serial correlation coefficients, first between U- and Y-series, on one side, and the random time series, on the other side, and then of differences among U- and Y-series themselves.

d. The differences $\Delta\bar{r}_1$ for the 14 stations of the Upper Colorado River Basin and of the areas adjacent to it, and for the 140 stations of a larger sample are:

$$\Delta\left[\Delta\bar{r}_1\,(U)\right] = 0.064$$

$$\Delta\left[\Delta\bar{r}_1\,(Y)\right] = 0.069$$

The difference is approximately the same for both U- and Y-series. It seems that the semi-arid and arid regions have the tendency to have greater differences $\Delta\bar{r}_1$, than the humid regions, as will be shown later by a discussion of regional distributions of the first serial correlation coefficient. It seems logical to conclude, that the same factors which cause the mean first serial correlation coefficients to be greater than zero in humid regions, are more pronounced and emphasized in the semi-arid and arid regions, so that the means of first serial correlation coefficients are somewhat greater than in the humid regions.

e. There are departures at the extremes of distributions of first serial correlation coefficients of lines (1), (2), (6), and (7) from the straight line passed through the median values with the slopes of corresponding random time series. There could be many reasons for these departures (i.e., rivers with glaciers and snow carryover from year to year) apart from sampling departures, but the use of mean length N of observed annual flow series in the computation of parameters of first serial correlation coefficient distributions for random time series may be partly responsible for these departures.

f. The first serial correlation coefficient of annual flows is greater than the first serial correlation of the effective annual precipitations. In the case of 140 stations the quantity of water carried over from year to year in river basins increases the first serial correlation coefficient from its mean value 0.133 for effective annual precipitations to the mean value 0.178 for annual flows. In other words, the mean first serial correlation coefficient is increased on the average nearly 50 percent by the carryover. Or, by excluding the carryover, the mean first serial correlation coefficient is decreased by one-third. For the 14 stations in the Upper Colorado River Basin and the stations around it, the carryover is smaller, so that it increases the mean first serial correlation coefficient from 0.202 for the effective annual precipitations to 0.242 for the annual flows, or for 21 percent. In other words, by excluding the carryover from year to year, the first serial correlation coefficient of annual flows is decreased by about one-sixth. This is about one-half as much as for 140 stations from many parts of the world. Supposing that the figures for U- and Y-series are true values, it can be concluded that the first serial correlation coefficients are decreased by excluding the carryover of water from year to year. The carryover is, therefore, principally responsible for one part of the greater values of median or mean first serial correlation coefficients of U-series than those of random time series.

g. It is important to analyze which could be the factors influencing the positive mean first serial correlation coefficient for the effective annual precipitations, or for both the U- or the Y-series. The statistical model given by eq. (2) points out, that the difference between the given values of effective annual precipitation and the true values of effective annual precipitations could be caused also by four types of errors: random errors, inconsistency, non-homogeneity, and the errors in determining the carryover from year to year, apart from the carryover effect (ΔW). Suppose that a random time series is changed by introducing all four types of errors. The random errors could only increase the standard deviation of the random time series. It is easy to prove that random errors in annual flows are small and could be practically neglected. Any inconsistency (errors in one side which can change from place to place in time series or any inconsistency in the form of trends or jumps), and any non-homogeneity of data, which comes from man's activities in the river basin, and any error in computing the carryover in a river basin, increase on the average the first serial correlation coefficients. If the random time series of length N, with normal distribution of first serial correlation coefficients and their mean close to zero, has any inconsistency, non-homogeneity, and the errors similar to error in the corrections for the carryover, this biased series will have on the average the mean value of first serial correlation coefficients greater than the random time series.

4. Example of effect of carryover

As an example of a significant impact of water carryover from one water year to another on the first serial correlation coefficient the case of St. Lawrence River at Ogdensburg, New York, is given here. For the period of observations of 97 years, from 1860 to 1957, for St. Lawrence, the first serial correlation coefficient of actual annual flows is 0.705. When the effect of stored water in the Great Lakes was taken into consideration by correcting the time series and computing effective annual precipitation, the first serial correlation coefficient of the effective annual precipitation was dropped to 0.094. If the stored water in the St. Lawrence River Basin outside the Great Lakes, which means in the small lakes, rivers and in the underground, would be taken into consideration also, then the first serial correlation coefficient of 0.094 would probably be decreased still further. It must be pointed out, that the good data on levels and stored water in Great Lakes gave a good accuracy in computation of the overyear carrying quantity of water. It can be concluded, therefore, that in many river basins the largest factor in creating dependence between the successive values of annual flows is the water storage in river basin.

B. REGIONAL DISTRIBUTION OF FIRST SERIAL CORRELATION COEFFICIENT

1. Procedure and results

Figure 3 shows the position of 72 river gaging stations in the United States, used for a general study of patterns in annual flow fluctuations. The solid lines divide 14 hydrological regions as designated by the U.S. Geological Survey. Figure 4 shows the position of 13 river gaging stations in Canada, used as the part of 140 stations from many parts of the world. Each station has an identification number. At the same time there are two figures for each station, the upper figures giving the first serial correlation coefficients for the annual flows, and the lower figure giving the first serial correlation coefficients for the effective annual precipitations. Taking the effective annual precipitations as a measure, all stations having the first serial correlation coefficient greater than +0.10, on one side, and all the stations having the first serial correlation coefficients lower than +0.10, on another side, are specially marked. The negative first serial correlation coefficient for the effective annual precipitations is given also an additional sign.

It could be distinguished clearly that 48 stations have the first serial correlation coefficient of Y-series above and 37 stations below +0.10. There are 14 stations with negative first serial correlation coefficients. About half the stations have an $r_1 > 0.15$ for the Y-series.

2. Conclusions and discussion of results

The general conclusions from the results represented in figs. 3 and 4 are as follows:

1. The humid regions of the East of the United States, and the humid regions of the West of the United States (fig. 3) more frequently have the first serial correlation coefficients for the effective annual precipitation below +0.10 than above +0.10.

2. The dry regions in the Middle West and in the Rocky Mountains (fig. 3) more frequently have the first serial correlation coefficients of the effective annual precipitation above +0.10 than below +0.10. The regions around the Gulf of Mexico would be considered as approximately having the same number of stations with first serial correlation coefficients above or below +0.10.

3. The same patterns seem to be valid for Canada (fig. 4).

4. The station inside the Upper Colorado River Basin, given as numbers 41, 42, and 44 have r_1 above 0.10, and r_1 for station 43 is below 0.10. The stations to the West of the Upper Colorado River, in Utah, nearly all have first serial coefficients greater than 0.10. The stations to the East of the Colorado River Basin have the coefficients both above and below +0.10.

It can be concluded from the approximate analysis, that stations in arid regions are more likely to have greater first serial correlation coefficients for the effective annual precipitation, than stations in humid regions. The main question which arises is, what are the reasons for this difference? Could one part of this difference be explained by greater errors which come from the inconsistency in data, or by non-homogeneity with high relative depletion in the river basins due to man's activities, or by the errors in determining the carryover in arid regions (the fact is that usually the recession curves of rivers in arid regions fluctuate more around the mean recession curve, than it is the case in humid regions)?

It seems a quite attractive conclusion, before any climatic reason is studied for explaining this difference in first serial correlation coefficients between arid and humid regions, or before any climatic reason is advanced for the positive mean first serial correlation coefficients in effective annual precipitation, that the effects of inconsistency, non-homogeneity, and errors in determining the volumes of carryover should first be analyzed.

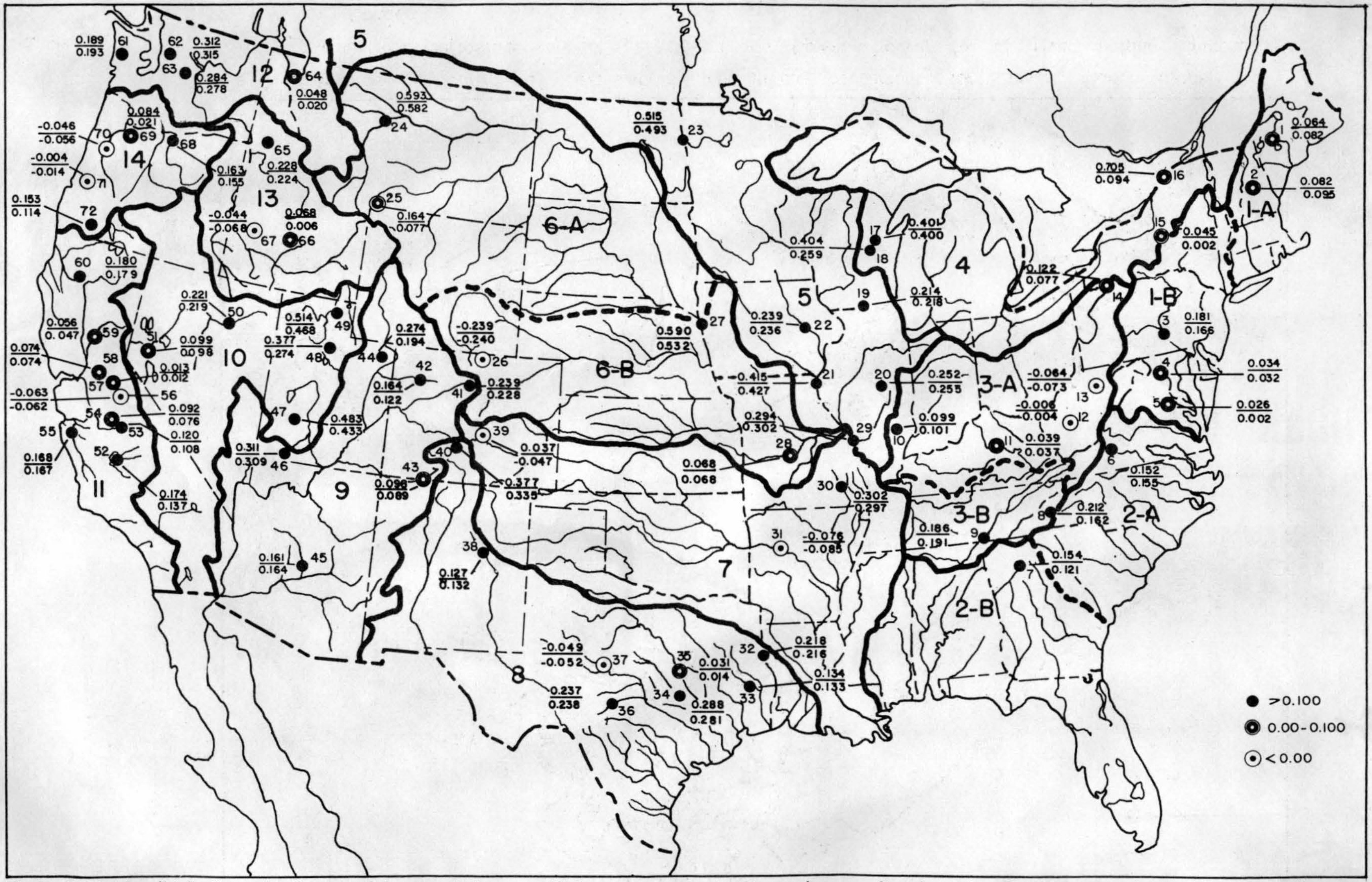


Fig. 3 Regional distribution of first serial correlation coefficient for annual flows (upper figure) and effective annual precipitations (lower figure) for 72 river gaging stations in the United States.

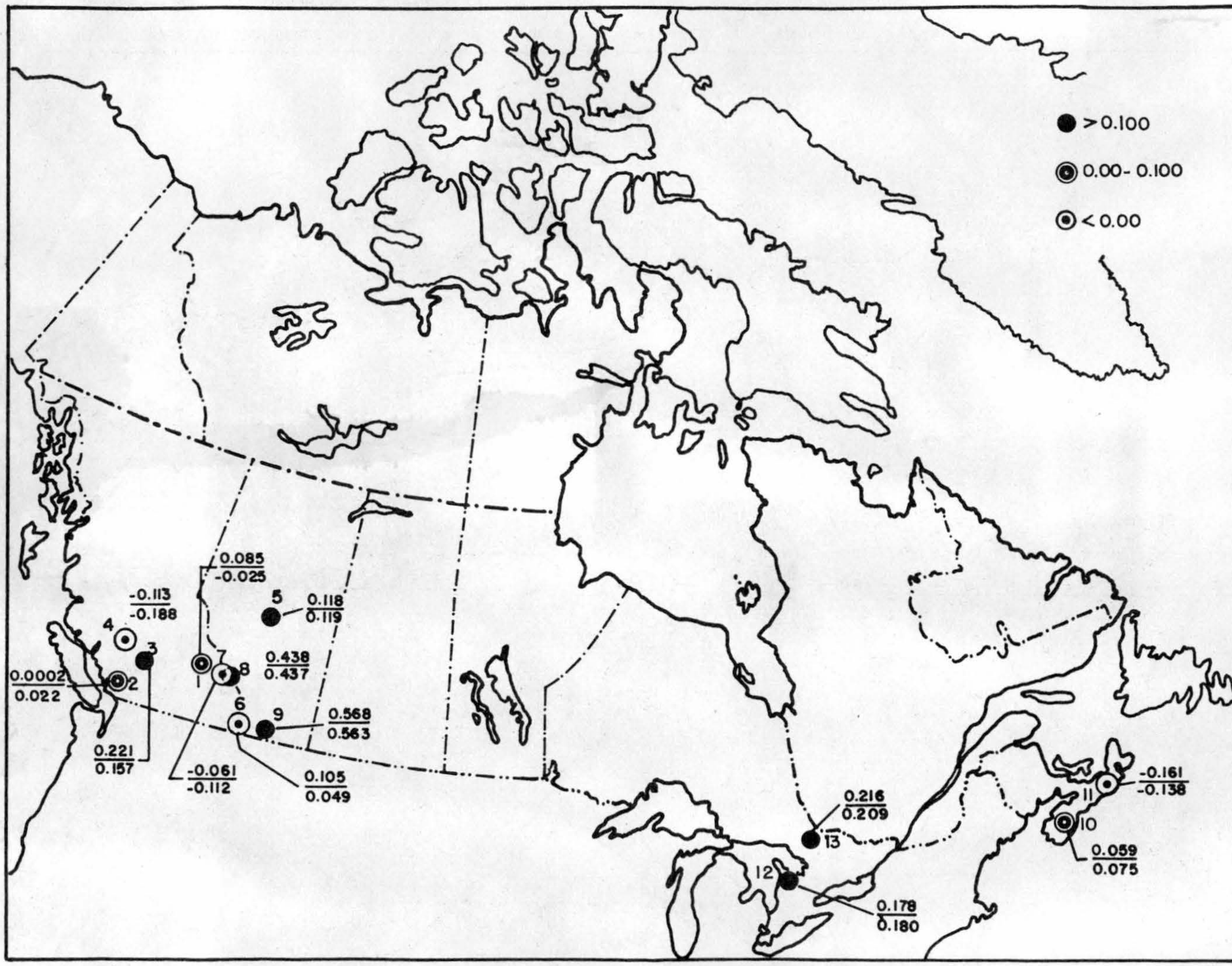


Fig. 4 Regional distribution of first serial correlation coefficient for annual flows (upper figure) and effective annual precipitations (lower figure) for 13 river gaging stations in Canada.

Table 1. Hydrologic data for fourteen river gaging stations in Upper Colorado River Basin and around it.

Station identification number *	River	Station	State	Basin area A	Mean discharge Q	Mean yield q	Number of years	Period of record from-to	Coeff. of variation C_v		Skew coeff. C_s		First serial corr. coeff. r_1		Index of variability (coeff. of variation of logarithms) I_v		First serial coeff. of logarithms		Relative basin storage W/A**	Δr_1
				sq. mi.	cfs	cfs/sq.mi.			U	Y	U	Y	U	Y	U	Y	U	Y	10^6 cf/sq.mi.	
25	Yellowstone	Corwin Springs	Mont.	2630	2987	1.14	45	1910-1955	.218	.227	.130	.137	.164	.077	.225	.235	.156	.063	3.54	0.087
26	Middle Boulder Creek	Nederland	Colo.	35.5	53.6	1.51	50	1907-1957	.240	.242	.205	.236	-.239	-.240	.248	.249	-.209	-.213	0.845	+0.001
38	Pecos	near Anton Chico	New Mex.	1050	144.5	0.138	46	1911-1957	.709	.712	1.49	1.48	.127	.132	.749	.776	.040	.033	0.0738	-0.005
39	Little Beaver Creek	near Pikes Peak	Colo.	1.0	0.563	0.563	48	1909-1957	.481	.514	.909	.864	.037	-.047	.509	.551	.082	.020	2.23	0.084
40	Arkansas	Salida	Colo.	1218	588	0.483	47	1909-1956	.213	.220	-.407	-.359	.377	.335	.234	.241	.362	.328	1.28	0.042
41	Fraser	near Winter Park	Colo.	27.6	40.6	1.47	47	1910-1957	.247	.249	.288	.218	.239	.228	.253	.258	.275	.258	5.18	0.011
42	White River	near Meeker	Colo.	762	636	0.836	48	1909-1957	.220	.233	.529	.611	.164	.122	.220	.231	.149	.109	2.62	0.042
43	San Juan	Rosa	New Mex.	1990	1235	0.621	47	1910-1957	.431	.433	.296	.325	.098	.089	.468	.469	.125	.121	0.264	0.009
44	Ashley Creek	near Vernal	Utah	101	106.2	1.05	43	1914-1957	.311	.324	.488	.483	.274	.194	.319	.334	.264	.185	2.59	0.080
45	Verde	below Bartlett Dam	Ariz.	6160	778	0.126	50	1888-1938	.639	.639	1.30	1.28	.161	.164	.625	.628	.198	.198	0.0635	-0.003
46	Virgin	Virgin	Utah	934	217.3	0.233	42	1909-1951	.408	.409	1.13	1.13	.311	.309	.376	.377	.340	.337	0.0524	0.002
47	Beaver	near Beaver	Utah	82	51.9	0.633	43	1914-1957	.348	.359	-.360	-.376	.483	.433	.431	.455	.593	.480	1.98	0.050
48	City Creek	nr. Salt Lake City	Utah	19.2	16.4	0.854	59	1898-1957	.274	.302	.629	.438	.377	.274	.277	.322	.363	.241	8.28	0.103
49	Blacksmith Fork	above U.P. Co. near Hyrum	Utah	260	126.7	0.499	44	1913-1957	.343	.355	.247	.245	.514	.468	.367	.382	.515	.470	1.43	0.046
					Mean value		47						0.22	0.18						

* See figure 2.

** Storage in a river basin at the end of a water year divided by the basin area.

The analysis of data given in figs. 3 and 4 shows a regional grouping of the first serial correlation coefficient of the same order of magnitude. This areal persistence is the consequence of simultaneity of occurrence of dry and wet years over small regions. This occcurrence seems to follow, however, a random areal distribution from one such region to another. If this random distribution is assumed, it means that a small region which has had for many stations and on the average a large positive value of first serial correlation coefficient for effective annual precipitations (say around 0.25) in the past period of observation may have a much smaller positive value than that, or even a negative value in the future period of observation. This would correspond to the concept of random sampling from one small region to another. This may be proved by dividing the past observation period on stations in two parts, by computing the first serial correlation coefficients for both parts, and by comparing their regional distributions.

If the centers of regions are sufficiently distant so as to be considered to having within them independent effective annual precipitations, the random regional distribution of first serial correlation coefficient may be accepted. However, if all stations between the centers of two such regions would be studied, their first serial correlation coefficients would change gradually from the coefficient of one center to that of another center, if both regions are sufficiently hydrologically homogenous.

The regional distribution of first serial correlation coefficient of effective annual precipitation may be used as one of criteria for dividing an area into homogeneous hydrologic regions.

The Upper Colorado River Basin may be according to fig. 3 considered from the point of view of first serial correlation coefficient distribution for effective annual precipitation as fairly homogeneous hydrologic region.

C. EFFECTS OF INCONSISTENCY, NON-HOMOGENEITY, AND ERRORS IN DETERMINING THE VOLUMES OF CARRYOVER ON THE FIRST SERIAL CORRELATION COEFFICIENT

1. Inconsistency

Any inconsistency in measuring or computing the annual flows caused by human errors of a systematic type, generally as jumps or trends introduced in data (which jumps and trend do not exist in true data), increase on the average the mean first serial correlation coefficient.

The analyses presented above are not valid for the fourteen river gaging stations in the Upper Colorado, if there is any significant inconsistency in the data of annual flows. It is assumed that there are no such errors. But it is very probable, due to changing techniques and ways of computing the annual flows, that small inconsistencies might exist in the records of some stations used for this study.

If an inconsistency is introduced into a random time series in the form of a jump at the position between N_1 and N_1+1 in time series ($N_1 + N_2 = N$, with N length of series), then the second part N_2 of series is $Y_i = (1+i)\, Y_{it}$, where Y_{it} is true value, Y_i is inconsistent value, and i is the relative value of inconsistency (percentage of change due to the inconsistency). Designating the ratio N_2/N as q, and neglecting small order terms, the expected value of first serial correlation coefficient is now approximately, as developed by the author of the paper (ref. 1),

$$E(r_1) = \frac{1}{C_v^2}\, \frac{q(1-q)i^2}{(1+qi)^2} \tag{9}$$

which is positive. C_v is the coefficient of variation of the time series.

For example, if the inconsistency is introduced in the middle of a time series, $q = 1/2$, then $E(r_1) = i^2/C_v^2(2+i)^2$. For $C_v = 0.25$ and $i = -0.10$ (10% of change in the sense of decreased values in relation to true values), the expected value of r_1 is 0.044. Therefore, any jump (created by inconsistency and non-homogeneity) increases on the average the serial correlation coefficients. The same can be proved for inconsistency trends.

The general model of inconsistency in data of river gaging stations is a combination of jumps and trends between the jumps, and it can be proved that this model applied to a random time series (superimposed on this series) increases on the average the serial correlation coefficients in comparison with those of random time series.

The example of the Nile at Aswan Dam shows how inconsistency can introduce a kind of fictitious dependence in time series of annual flows. To explain the meaning and effect of inconsistency, the Nile example will be here described.

The Nile River at Aswan Dam has observations from 1869 to 1955, used in the study referred previously as ref. 1. In the period 1869-1903 there were only stage measurements downstream from the place where Aswan Dam was built in 1903. After the Aswan Dam was built in 1903, the measurements of the flow were done by sluice gate openings. The rating curve (discharge versus water level) is derived as the discharge through the sluice gates of the Dam versus the downstream level at the previous gaging station. This rating curve was used to determine the annual flows for the period 1869 to 1903. This period has 25 percent more water than the period of 52 years from 1903 to 1955. The study of the other 139 stations, used in the study of ref. 1, showed that there is no station with so large a difference in the mean discharge between the periods of about 34 and 52 years. The occurrence of this difference has very small probability.

Assuming, however, that this difference is correct, then the first serial correlation coefficient for U-series and for period of 86 years from 1869 to 1955 is 0.553, while the first serial correlation coefficient of Y-series is 0.438. They are rather large values, which points out that the effective annual precipitations were highly serially correlated. The period 1869 to 1903 taken separately has the first serial correlation coefficient of U-series 0.381, and that of Y-series 0.194. The period 1903 to 1955, with reliable measurements, gave the first serial correlation coefficient of U-series 0.163, and that of Y-series 0.099. If only the period 1903 to 1955 would be taken into consideration, then the low first serial correlation coefficient of Y-series would point out clearly, that the dependence in the series is low. It is quite probable that by the closing of Aswan Dam, all sediment being retained in the reservoir, a degradation took place, namely a clearing of all sediment deposits on or among rocks in the bed on the downstream part of the river occurred, and so decreased the river stage levels for the same discharge. In the same time, the increased evaporation from the lake, and upstream irrigation development introduced the non-homogeneity in data. It is highly probable that both the inconsistency in data for the period of 1869 to 1903, and also the non-homogeneity caused by increased evaporation losses from the reservoir and from upstream irrigated land did create the high dependence, or very large first serial correlation coefficient of the Y-series for the period 1869 to 1955.

2. Non-homogeneity in data

There is a depletion of the total annual runoff in many rivers in the world, in the United States, and particularly for the case of this study, in the Upper Colorado River Basin and around it. This region has been influenced for the last eighty years by settlements, by development of agriculture, irrigation, and industry, with building of dams and reservoirs, water diversions and so on.

It is clear that any irrigation, evaporation from the new reservoirs, or any water diversion out of the river basin mean a depletion of the total yield of a river basin. It is not, however, so clear how much the other factors, as the increase of population, changes in agricultural practices, building of small dams and pools, general changes of biological cover, etc., influence the depletion of water.

All these man-made measures in the Upper Colorado Basin have created a depletion trend, which results in non-homogeneity of data. This causes a dependence of successive values in a time series of annual flow or of effective annual precipitations, so that the first serial correlation coefficients must be on the average more positive than negative.

As the non-homogeneity of data may be described by a general model of jumps and trends between the jumps, similar to the general model of inconsistency, the same conclusions given for inconsistency are valid also for non-homogeneity in data.

As the example of the effect of non-homogeneity in data on the serial correlation coefficients, the example of Lee Ferry Station of the Upper Colorado will be used here and in subsequent analyses of correlograms.

3. Non-homogeneity in data of Upper Colorado River at Lee Ferry Station

The non-homogeneity in data of the Upper Colorado River at the Lee Ferry Station, Arizona, will be analyzed here as an important characteristic of Upper Colorado flows. Figure 5 shows the virgin and measured (historical) annual flows at this station. The data, until 1947, is from the House Document No. 364, Washington, D. C., 1954 (ref. 3): Colorado River Storage Project, and for years 1948-1959 the data is obtained from the U. S. Department of Interior, Bureau of Reclamation, Regional Office, Region 4, Salt Lake City, Utah.

It might be that some inconsistency exists in data prior to 1914. In the ref. 3, page 141, it is stated: "Although inaccuracies are risked with the extension of records prior to 1914, the Bureau of Reclamation made extentions to include the 1896-1947 period at Lee Ferry ...".

For the determination of depletion of the water yield, in the same reference on page 143 it is stated: "Stream depletions from upper basin development, therefore, have been estimated only at sites of use, and aggregate depletions so determined are considered representative of the depletion at Lee Ferry," and on the same page: "This includes depletions from all causes, such as irrigation and uses incident to irrigation, water exports to areas outside of the drainage basins, domestic and industrial uses, and evaporation from storage reservoirs. The estimate allows credits for water imporations and channel salvage."

Figure 6 gives for the Lee Ferry Station the relationship of three variables: 1) Annual depletion, D in 10^6 acre-feet; 2) Annual virgin flow, V, in 10^6 acre-feet; and 3) Time (as parameter). It shows clearly, that the depletion has increased fast

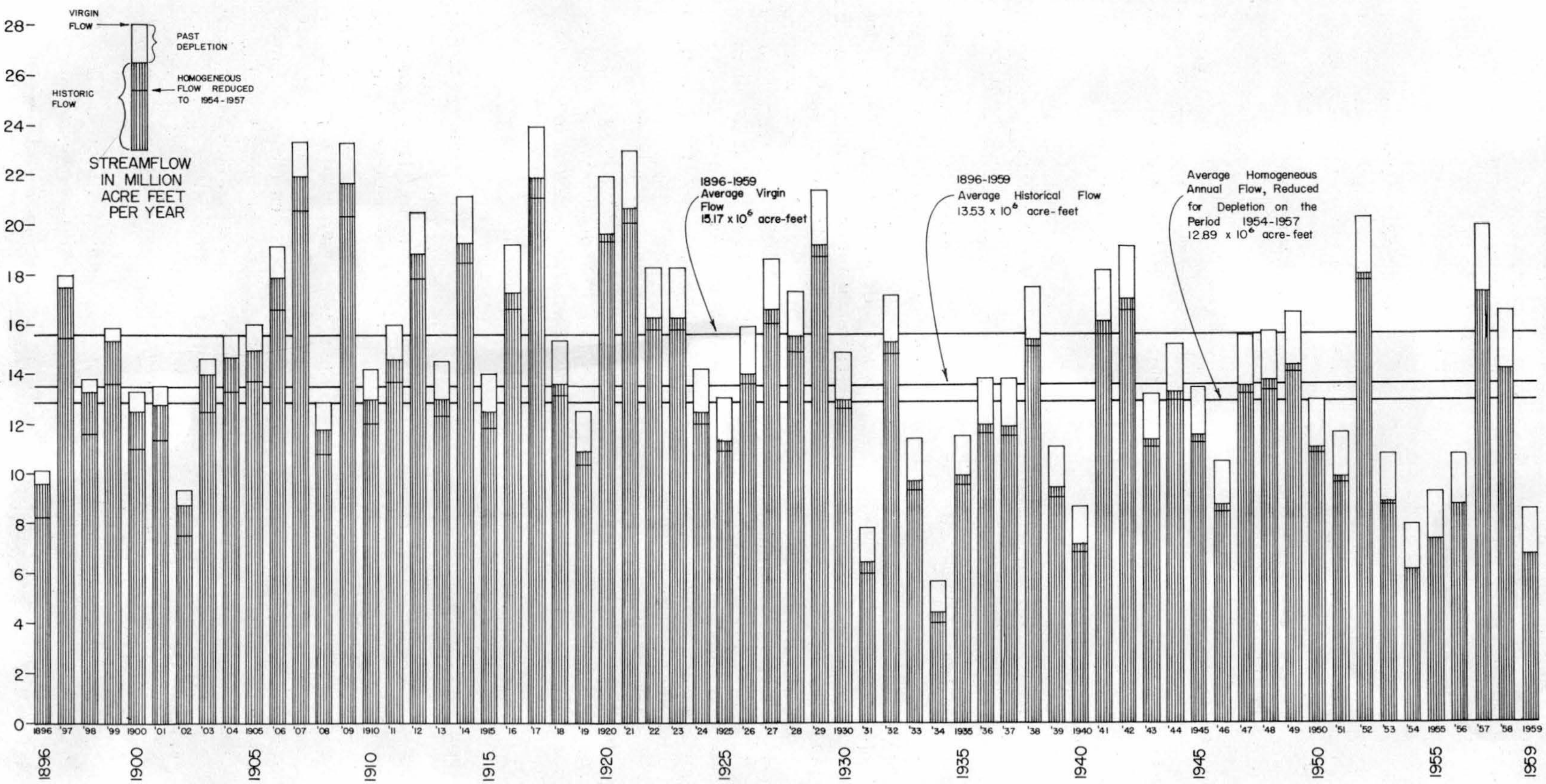


FIG. 5 COLORADO RIVER ANNUAL FLOWS AT LEE FERRY, ARIZONA

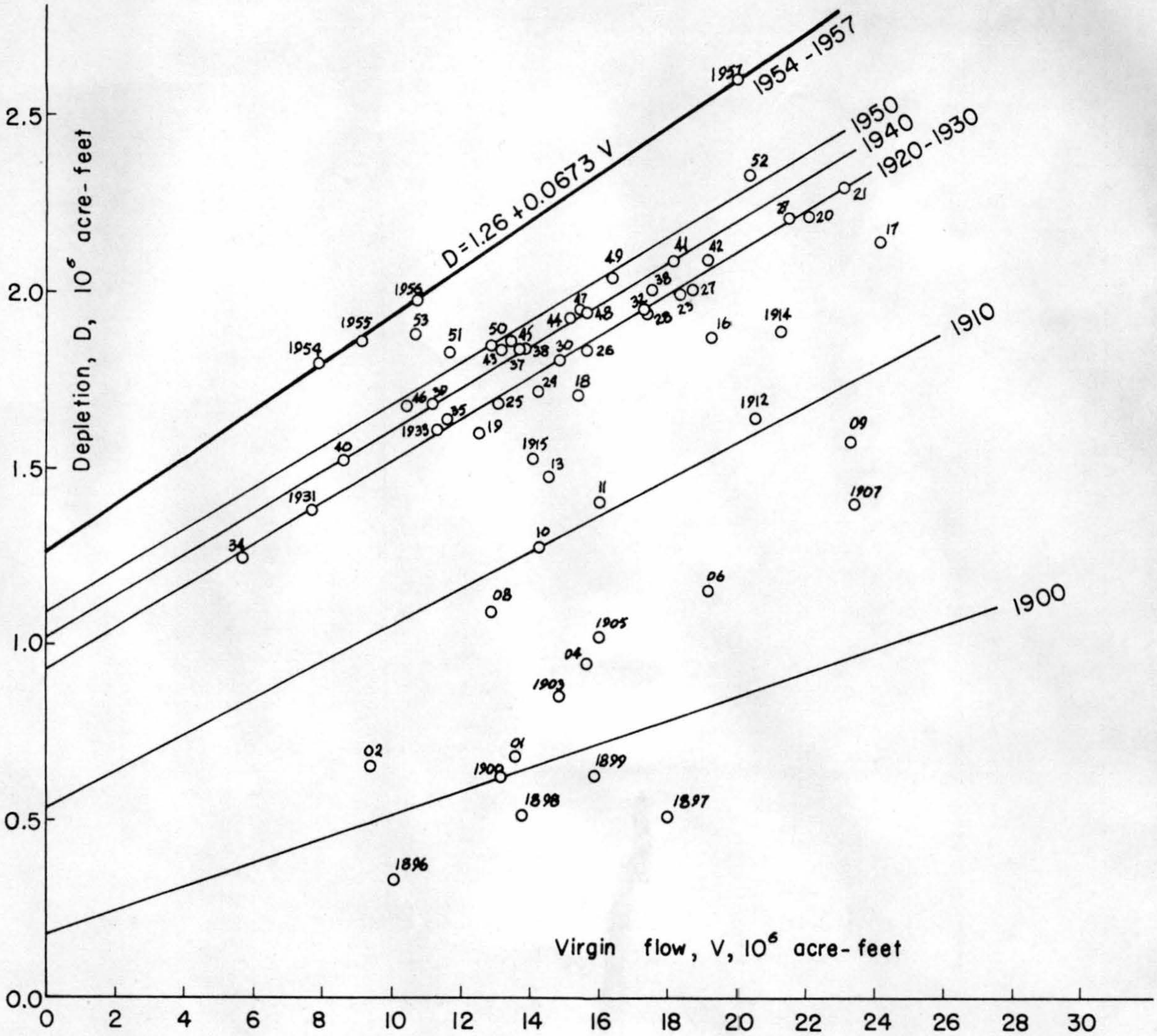


Fig. 6. Relation of man-made annual flow depletion in the Upper Colorado River at Lee Ferry, Arizona, to its virgin annual flows (historical flow plus depletion) for given time intervals or years, for the period 1896-1957.

from the turn of century until the end of First World War, then stayed approximately constant for the period 1920-1930, slowly increasing from 1930-1950, then increasing faster from 1950-1954, and approximately constant from 1954-1957. In this case, the historical annual flow at Lee Ferry Station is an evolutive time series (and not a stationary time series). To make the series homogeneous (to compensate for depletions), the virgin annual flows approximately determined give an insight what would be the flow, if the hydrologic factors of Upper Colorado River Basin would stay unchanged by man's activities. Though the depletions determined have errors (because they depend on many factors, on some rough and approximate evaluations of net consumptive water uses, and on net evaporation from the new water surfaces), and though the computed virgin flows are less accurate than in the case they would coincide with historical (measured) flows, both these depletions and virgin flows derived give a sufficient measure of man-made non-homogeneity in hydrologic records of this river basin.

The depletion model is defined here as the relation of annual depletion to annual virgin flow for given interval of years of for a year. They are approximated by straight lines (fig. 6), because

Table 2. Annual flows and depletion of the Upper Colorado River at Lee Ferry Station, Arizona

Water year	Annual historical flows 10^6 ac-ft	Annual historical depletions 10^6 ac-ft	Annual virgin flows 10^6 ac-ft	Homogeneous * annual flows reduced to period 1954-1957 10^6 ac-ft	Water year	Annual historical flows 10^6 ac-ft	Annual historical depletions 10^6 ac-ft	Annual virgin flows 10^6 ac-ft	Homogeneous * annual flows reduced to period 1954-1957 10^6 ac-ft
1896	9.8	0.3	10.1	8.16	1928	15.3	2.0	17.3	14.88
1897	17.5	0.5	18.0	15.53	1929	19.2	2.2	21.4	18.70
1898	13.3	0.5	13.8	11.61	1930	13.1	1.8	14.9	12.64
1899	15.2	0.7	15.9	13.57	1931	6.4	1.4	7.8	6.02
1900	12.6	0.6	13.2	11.05	1932	15.3	1.9	17.2	14.78
1901	12.9	0.7	13.6	11.42	1933	9.7	1.7	11.4	9.37
1902	8.7	0.7	9.4	7.51	1934	4.4	1.2	5.6	3.96
1903	14.0	0.8	14.8	12.54	1935	9.9	1.6	11.5	9.47
1904	14.7	0.9	15.6	13.29	1936	12.0	1.8	13.8	11.61
1905	15.0	1.0	16.0	13.66	1937	11.9	1.8	13.7	11.51
1906	18.0	1.1	19.1	16.55	1938	15.4	2.1	17.5	15.06
1907	22.0	1.4	23.4	20.56	1939	9.4	1.7	11.1	9.09
1908	11.8	1.1	12.9	10.77	1940	7.1	1.5	8.6	6.76
1909	21.7	1.6	23.3	20.47	1941	16.0	2.1	18.1	15.62
1910	13.0	1.2	14.2	11.98	1942	17.0	2.1	19.1	16.55
1911	14.6	1.4	16.0	13.66	1943	11.3	1.8	13.1	10.96
1912	18.9	1.6	20.5	17.86	1944	13.2	2.0	15.2	12.92
1913	13.0	1.5	14.5	12.26	1945	11.5	1.9	13.4	11.24
1914	19.3	1.9	21.2	18.51	1946	8.7	1.7	10.4	8.44
1915	12.5	1.5	14.0	11.80	1947	13.5	2.0	15.5	13.20
1916	17.3	1.9	19.2	16.65	1948	13.7	1.9	15.6	13.29
1917	21.9	2.1	24.0	21.12	1949	14.4	2.0	16.4	14.04
1918	13.6	1.8	15.4	13.10	1950	11.1	1.8	12.9	10.77
1919	10.9	1.6	12.5	10.40	1951	9.8	1.8	11.6	9.56
1920	19.7	2.3	22.0	19.26	1952	18.0	2.3	20.3	17.67
1921	20.7	2.3	23.0	20.19	1953	8.0	2.7	10.7	8.72
1922	16.3	2.0	18.3	15.81	1954	6.1	1.8	7.9	6.11
1923	16.3	2.0	18.3	15.81	1955	7.3	1.9	9.2	7.32
1924	12.5	1.7	14.2	11.98	1956	8.8	1.9	10.7	8.72
1925	11.3	1.7	13.0	10.86	1957	17.3	2.7	20.0	17.39
1926	14.0	1.9	15.9	13.57	1958	14.2	2.3	16.5	14.13
1927	16.6	2.0	18.6	16.09	1959	6.7	1.8	8.5	6.67
					Mean	13.53	1.64	15.17	12.89

* Homogeneous annual flows reduced to period 1954-1957 is a sample of annual flows which would be experienced if the depletion causes in the period 1954-1957 would be prevailing all along the period 1896-1959.

the complex models would not be justified in view of errors which are inherent in determination of depletions. For the period 1954-1957, the depletion model is

$$D = 1.26 + 0.0673\, V \qquad (9)$$

Figure 6 shows that both coefficients A and B in equation $D = A + B\,V$ increase with time. This increase of A and B means a greater average depletion per year with time, and an increase of B means that the depletion fluctuates more in function of the absolete value of virgin flow with time than earlier depletions, as the mean depletion increases. The increase of B means also, that the number of factors which affect the depletion, but are proportional to virgin flow, increases with time (more water diverted to irrigations inside the basin or more water diverted out of basin in wet years than in dry years, more evaporation from small and large reservoirs because of larger free surface areas in wet years than in dry years, and similar factors).

The historical flows of Lee-Ferry Station are a non-homogeneous time series. In statistical words, the sample of 64 years of annual flows at Lee Ferry (table 2) is derived from a super-universe* (superpopulation), and not from a unique universe well defined of all possible annual flows.

If the causes affecting flows at a river cross section do not change with time, their effect in form of all possible annual flows (infinite number of annual flows) represent a unique universe (population), or the samples derived from this universe are homogeneous. When the causes change with time (changes in consumptive use, in evaporation, transpiration, diversion out of the basin or into it, etc.), each change produced in form of a jump creates a new composition of causes, or there is a new universe of annual flows. If the factors affecting runoff change gradually, a universe passes gradually to another one. Theoretically speaking, a superuniverse and samples drawn from it may be treated statistically or probabilistically, if the law of change in time from one universe to the next one in the superuniverse would be known. Due to the fact that most of the changes are introduced by man's activity, and that the laws of change with time for runoff are complex and mostly unpredictable with sufficient accuracy, the approach of treating the superuniverse and samples from it in annual runoff is not usually feasible. This is the reason why the techniques of correcting the non-homogeneous samples into homogeneous ones are introduced and practiced currently. It would be extremely difficult to project the depletion of annual runoff in amount and in time at Lee Ferry Station for the next 3-5 decades with sufficient accuracy (all future storage reservoirs, diversion projects, irrigation schemes predicted exactly in runoff amount and in time for the next 30-50 years). If that very approximate projection of depletion model would be acceptable for economic and engineering studies, it would be possible at least theoretically to treat future possible samples of runoff starting from the superuniverse.

The computation of virgin flows for Lee Ferry from 1896-1959 is a procedure for making this sample homogeneous. In other words, the computed virgin flow sample is drawn from the universe that existed prior to any depletion, under the combination of a very large number of natural causes affecting the runoff. Practical problems require, however, that computations in engineering and economics should be carried out with homogeneous samples or with the universe (inferred from these samples) which are valid for the momemt of computation, or for the time interval a measure would serve. For the period of 1954-1957, applying its depletion model, eq. (9), to the virgin flows of the sample 1896-1959, a new homogeneous sample valid for the period 1954-1957 only is obtained. Assuming that small changes in depletion have taken place from 1958-1960 in comparison with the period 1954-1957, the new homogeneous sample can be considered as if drawn from the universe of annual runoffs at Lee Ferry, valid for late fifties. If these new annual flows are designated by V_n, then $V_n = V - D_n$, or

$$V_n = 0.9327\, V - 1.26 \qquad (10)$$

in millions of acre-feet. Applying this equation, the new series V_n together with other samples (historic flow, virgin flow, and real historical depletion) is given in Table 2, and also is represented in fig. 5.

The average annual flows are, therefore, for the period 1896-1959 as follows:

			millions of acre-feet
1. Historical flow	$\bar{V}_h$	= 13.53	acre-feet
2. Depletion	$\bar{D}$	= 1.64	" "
3. Virgin flow	$\bar{V}$	= 15.17	" "
4. Homogeneous sample, reduced at the period 1954-1957	$\bar{V}_n$	= 12.89	" "
5. Mean depletion for the sample reduced at the period 1954-1957	$\bar{D}_n = \bar{V} - \bar{V}_n$	= 2.28	" "

Extrapolating the depletion model $D = A + BV$ in the future by computing A and B as functions of time, it would be possible to reduce the virgin

* Superuniverse (or superpopulation) defined as a set of all different universes (or populations) which existed in the past, and which passes one to another as some causes of runoff change with time.

flow sample 1896-1959 to any future date. The planning of the Upper Colorado River Basin development should be able to project the depletion model, if not in function of time, then at least in function of different projects to be implemented, and even in function of population growth. In this case, the new variate V_t, the annual flows at Lee Ferry, for a given date is

$$V_t = V - D_t = (1 - B_t)\ V - A_t \tag{11}$$

where A_t and B_t are parameters of depletion model $D_t = A_t + B_t V$, at the date t. Assuming that $A_t(t)$ and $B_t(t)$ are given, then $V_t(V, t)$ would also be given. With the probability distribution of V, as well as the characteristics of sequence patterns of V given in analytical form, both probability distribution and sequence model for V_t may be derived as function of t. This approach would enable the computations of average hydrologic characteristics during a depreciation time for a water resources development project.

The above analysis leads to a conclusion, namely that the computation of effects of man-made structures and measures in river basins has an important bearing on the reliability of hydrologic data to be used for further water resources developments and water project operations. The studies and computations aimed to make hydrologic samples homogeneous (and also consistent by removing the eventual inconsistency in data) through computation and analysis of depletion models (or systematic errors) is a new and quite important task of hydrologic activities.

The first serial correlation coefficient for the annual virgin flows (historical flow plus depletion) for 64 years at Lee Ferry is $r_1 = 0.184$, while it is for the historical flow $r_1 = 0.218$. The depletion as non-homogeneity in data has, therefore, caused an increase of the first serial correlation coefficient of historical data. The homogeneous sample reduced to the period 1954-1957 has also the value $r_1 = 0.184$. The non-homogeneity in data has increased the first serial correlation coefficient by 18.5%. The average increase of serial correlation coefficients of the correlogram is still greater than that of the first serial correlation coefficient, as it will be shown in the next chapter.

4. Errors in determining the volumes of carryover

The method of recession curves, and especially the mean recession curve of those for many dry seasons, was used to determine the stored water in the river basin at the end of water years. The use of mean recession curve introduces errors in correcting the annual flows to obtain the effective annual precipitations.

It is a known fact, that the recession curves fluctuate from year to year around the mean recession curve, which depends on many conditions in the dry season of the year. These departures of individual recession curves from the mean recession curve are probably much larger in dry regions than in humid regions. The errors in determination of carryovers is a factor which influences the dependence between the successive values of a time series, so that one part of the positive mean first serial correlation coefficient of effective annual precipitations computed in the above way can be attributed to these errors.

It would be very difficult to determine the order of magnitude of this influence without a thorough study of each station. In order to determine the effect of these errors, some stations should be studied by determining the carryover in two manners: (a) by using the mean recession curve, and (b) using for each individual year its recession curve. The difference would give out the order of magnitude of these errors. This study was not undertaken.

In some cases, the mean monthly flows instead of the mean daily flows were used to determine the total carryover. This probably introduced larger errors than in the case the carryovers are computed by using mean daily discharges. Most of carryovers of 14 stations in the Upper Colorado River Basin and around it were computed by using the monthly flow recession curves.

5. Carryover of Upper Colorado River at Lee Ferry Station

The method of determining the water carryover from one water year to another, based on the mean recession curve and on the index-discharge (mean daily or mean monthly) at the end of a water year, assumes a time series of virgin annual flows. In the case there is an artificial influence on carryover by storage reservoirs, diversions, etc., the computation of carryovers and from them determination of effective annual precipitations becomes much more complex than it is the case with virgin flows.

This work has not been carried out in this study for the Upper Colorado River at Lee Ferry, Arizona, since it was beyond the scope of work foreseen for this study.

The positive first serial correlation coefficients of virgin annual flows, and of homogeneous annual flows reduced by the depletion model of eq. (9) to 1954-1957 period are affected, therefore, by two carryovers: 1) ΔW_1, carryover resulting from the natural storage in river basin (underground storage, surface storage, snow storage when carried from one water year to another); and 2) ΔW_2, carryover resulting from the artificial storage and other factors which are subject to man's control. It is legitimate to assume here, that the accurate computation of both carryovers for a 64-year period of observations of the Upper Colorado River would reduce somewhat the first serial correlation coefficient of effective annual precipitations on river basin in comparison with r_1-value for the virgin annual flows.

Table 1 and fig. 3 show that for the four stations (No. 41-44) inside the Upper Colorado River Basin the first serial correlation coefficients are smaller for effective annual precipitation than those for annual flows and for these four stations respectively: 0.239 to 0.228 (5% difference); 0.164 to 0.122 (34%); 0.098 to 0.089 (10%); and 0.274 to 0.194 (41%). At the average the difference is about 20%. It is justified therefore, also, that the first serial correlation coefficient for the Lee Ferry Station and for effective annual precipitation should be smaller than that for annual flows.

It is necessary to stress that the effective annual precipitation (total precipitation on river basin minus total evapotranspiration) applies for conditions of virgin flows. The effective annual precipitations for the homogeneous sample reduced to the 1954-1957 period are to be increased by that part of depletion which is evapotranspiration (generally it is D-T, depletion minus trans-mountain diversion out of basin), so that the equation

$$P_n = P_e - (D_n - T_n) \tag{12}$$

with

P_e = effective annual precipitations for virgin flows;
P_n = effective annual precipitations for homogeneous sample reduced to the period 1954-1957;
D_n = annual depletions from depletion model for period 1954-1957;
T_n = annual net diversions out of basin (diversion out of basin minus diversion into the basin) for conditions in period 1954-1957.

Taking that

$P_e = V - \Delta W_1 - \Delta W_2$, then

$$P_n = V - \Delta W_1 - \Delta W_2 - D_n + T_n = V_n - \Delta W_1 - \Delta W_2 + T_n \tag{13}$$

is the net yield of atmosphere at the status of river basin development which corresponds to the period 1954-1957. Four variates (D_n, T_n, ΔW_1, and ΔW_2) should be determinated from the sample of river flows and from different depletion factors and storage operations in the river basin, as well as from transmountain diversions in order to compute the effective annual precipitations corresponding to homogeneous sample for period 1954-1957.

The determination of ΔW_1 and ΔW_2, together with D_n and T_n, cannot be carried out for Lee Ferry Station without a substantial error, so that the first serial correlation coefficient of effective annual precipitation P_n would be always affected by these errors.

D. ANALYSIS OF FIRST SERIAL CORRELATION COEFFICIENT ON ANNUAL FLOWS FOR UPPER COLORADO RIVER AT LEE FERRY STATION

1. Comparison with random time series

The virgin annual flows of the Upper Colorado River at Lee Ferry and the homogeneous sample of annual flows at Lee Ferry, reduced to period 1956-1957, have the first serial correlation coefficient $r_1 = 0.184$. The random time series has the expected value of r_1, eq. (6), for $N = 64$,

$$E(r_1) = -\frac{1}{N-1} = -0.016$$

so that the difference is $\Delta r_1 = 0.200$.

Assuming that the taking into account of carryover would decrease the first serial correlation coefficient of effective annual precipitations by about 30%, the difference in r_1 of series of effective annual precipitation and random series would be approximately 0.140. This estimate is rather arbitrary.

2. Probable reasons for positive coefficients

The positive coefficient $r_1 = 0.184$ for annual flows of virgin and homogeneous series is probably produced by: a) regional sampling effect, because the majority of nearby stations show the same trend; b) effect of carryover from one water year to another; and c) by systematic errors in measured flow and errors in determination of depletions.

Figure 3 shows that the most of the differences in r_1 between series of annual flows or effective annual precipitations and random series might be attributed to regional sampling departures of r_1 values from the expected value r_1 of random time series. The Upper Colorado River Basin is close to the Upper Missouri River Basin, and other adjacent dry climate regions and river basins (Lower Colorado, Western Texas, Utah), which have r_1 values generally much above the expected values for random time series. This especially refers to effective annual precipitations, after the carryover influence is excluded. This hypothesis of regional sampling with positive deviations for Upper Colorado River from the expected values of r_1 is assumed in this study, with high probability that the next 50-60 years would produce regional distribution of r_1 different from that given in figs. 3 and 4.

The actual lack of accurate determination of water carryover from one water year to another for the Upper Colorado River at Lee Ferry makes it difficult to judge which part of positive first serial correlation coefficient is a result of regional sampling, and which part is caused by water carryover.

IV. PATTERNS IN FLUCTUATIONS MEASURED BY CORRELOGRAMS

A. DEFINITION, GENERAL REMARKS AND PROCEDURE

1. Definition

The correlogram is defined as a graph of discrete points relating the serial correlation coefficients and the lag between successive correlated pairs of members of a time series. It is expressed as $r_k(k)$, where r_k is k-th serial correlation coefficient and k is the lag between correlated values. The polygon connecting the successive points (r_k, k) should be considered as a manner of representing the discrete points of the correlogram. For the U- and Y-series the lag k represents the number of years between the correlated values.

2. General remarks

The correlogram is a measure and an indicator of the type of dependence or independence among the members of a time series. If a cosine function fits the fluctuations of a correlogram, the correlogram indicates that the fluctuations of the time series may be either a sine or a cosine function. The moving average procedure applied to random time series (in practical problems either caused by physical factors as the storage in river basin, or by the moving average procedure of n successive values for smoothing the time series) creates a correlogram decreasing from $r_o = 1$ (r_o is unity by definition) slowly to $r_n = 0$. A random time series, if it is sufficiently long, has a random sequence of coefficients in correlogram, but the serial correlation coefficients are confined within the confidence limits of given probability, with the exceedence of these limits also with the given probability (sum of both is unity). There are many possible combinations of these three and other basic types of correlograms (see Kendall, ref. 4).

The confidence limits for random time series are, according to R. L. Anderson (ref. 2) for 95% level approximately.

$$R_{95\%} = \frac{-1 \pm 1.64\sqrt{N-k-2}}{N-k-1} \tag{14}$$

with the meaning that 5% of the points of correlogram should be on the average outside the confidence limits.

3. Procedure

The serial correlation coefficients for the annual flows and effective annual precipitation for 14 river gaging stations in the Upper Colorado River Basin and around it (shown in table 1) were compiled by using the following equations similar to eqs. (3) and (4), in which r_1 is replaced by r_k and unity by k (according to classical statistical formulae)

$$r_k = \frac{\sum_{1}^{N-k} U_i U_{i+k} - \frac{1}{N-k}\sum_{1}^{N-k} U_i \sum_{1}^{N-k} U_{i+k}}{(N-k-1)\, s_i s_{i+k}} \tag{15}$$

with

$$s_i^{\,2} = \frac{1}{N-k-1}\left[\sum_{1}^{N-k} U_i^{\,2} - \frac{1}{N-k}\left(\sum_{1}^{N-k} U_i\right)^2\right] \tag{16}$$

and

$$s_{i+k}^{\,2} = \frac{1}{N-k-1}\left[\sum_{1}^{N-k} U_{i+k}^{\,2} - \frac{1}{N-k}\left(\sum_{1}^{N-k} U_{i+k}\right)^2\right] \tag{17}$$

The computation of these unbiased serial correlation coefficients was carried out up to $m < N/4$, where N = length of time series, for all 14 stations, by using a digital computer.

The confidence limits on 95% level for random time series were computed by using eq. (14).

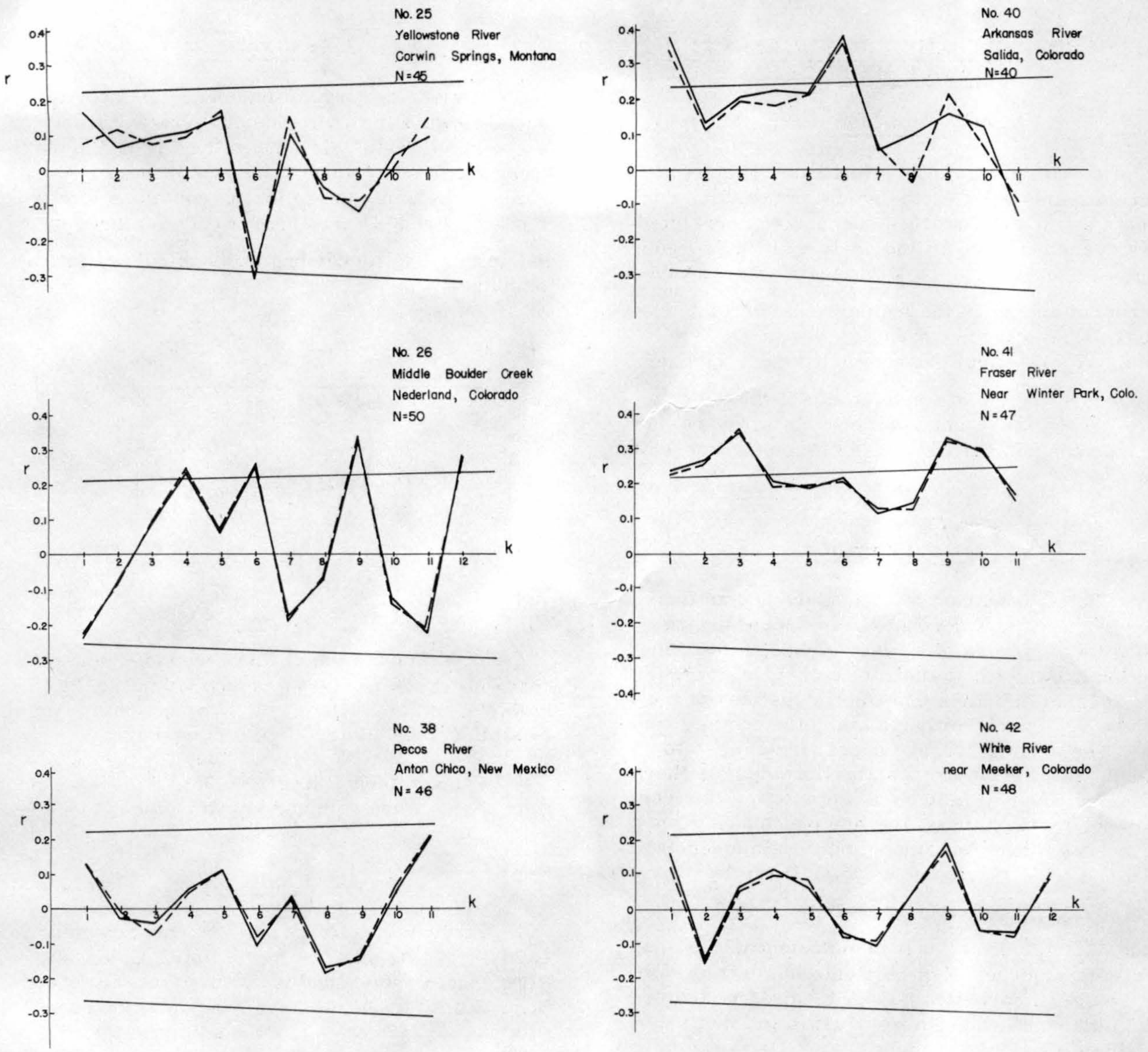
B. RESULTS

Figures 7 and 8 give the 14 correlograms for 14 river gaging stations in the Upper Colorado River Basin and around it together with their confidence limits. As N-k-1 in eq. (14) decreases with an increase of k, the confidence interval between the confidence limits increases with an increase of k, figs. 7 and 8.

Correlograms for both the U-series and the Y-series are represented on these figures.

C. ANALYSIS OF RESULTS AND CONCLUSIONS

Due to the relative small carryovers from year to year for the most of above 14 river basins, the difference between U-correlograms and Y-correlograms are rather small, except for r_1 -values for some stations. This means that the carryover affects mostly the link of successive values of lag



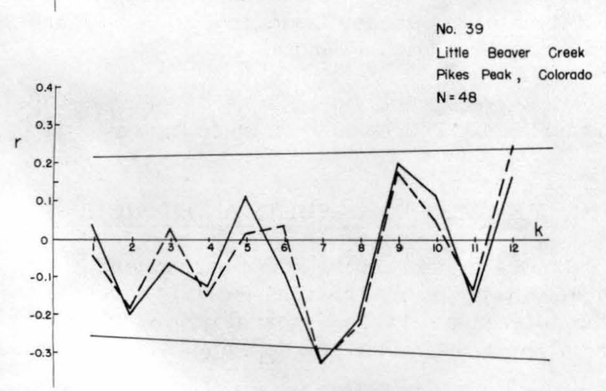
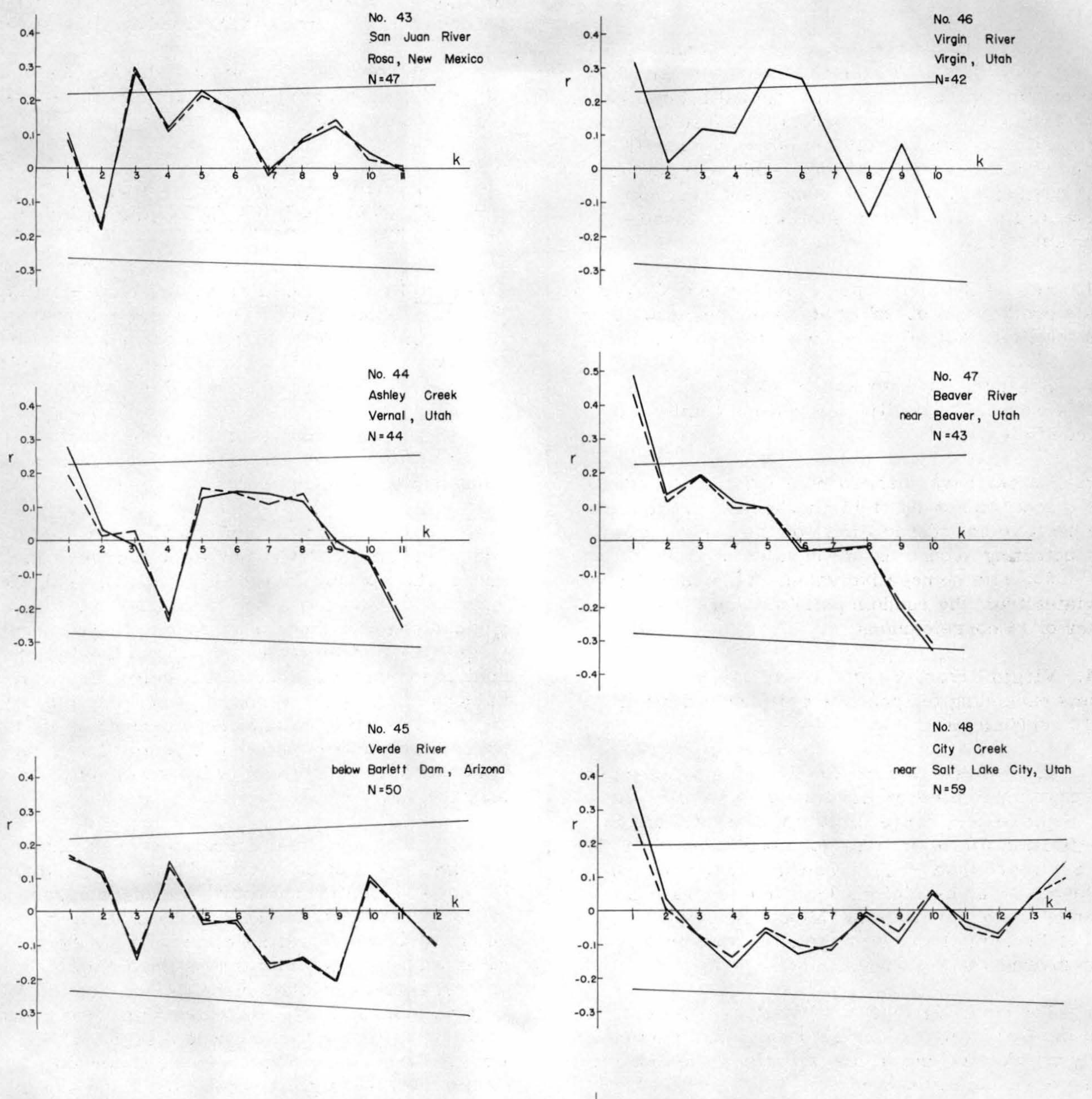


Fig. 7 Correlograms for 7 stations of and around the Upper Colorado River Basin for series of annual flows (full, heavy lines) and effective annual precipitations (dashed, heavy lines), with confidence limits (light lines) on 95% level.



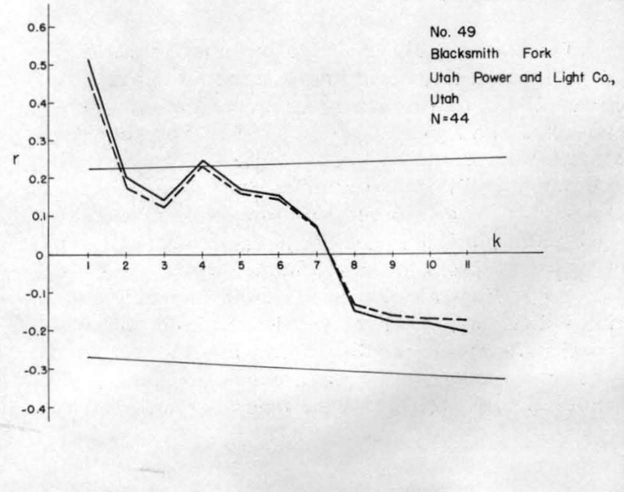


Fig. 8 Correlograms for 7 stations of and around the Upper Colorado River Basin for series of annual flows (full, heavy lines), effective annual precipitations (dashed, heavy lines), with confidence limits (light lines) on 95% level.

unity. This means also that the carryover from a given water year is felt significantly only in the next water year, and much less in subsequent water years.

The 14 double correlograms when analyzed jointly show that the majority of annual river flows and the correpsonding effective annual precipitations are close to random time series, because the majority of correlograms (10 out of 14) are confined in the confidence limits of 95% level, with no more than one serial correlation coefficient exceeding these limits. The 4 other stations have these characteristics:

1. Middle Boulder Creek at Nederland, Colorado, has 4 coefficients out of 12 above the positive confidence limit, but all 4 are very close to this limit;

2. Arkansas River at Salida, Colorado, has 2 points outside the positive confidence limit out of 11 coefficients;

3. Fraser River near Winter Park, Colorado, has 5 coefficients out of 11 coefficients greater than the positive confidence limit, or the correlogram is fluctuating around the mean value of about $r = 0.22$, with no negative value. This correlogram deviates from the random patterns more than any other of 14 correlograms.

4. Virgin River, Virgin, Utah, with 3 coefficients exceeding the positive confidence limit out of 10 coefficients.

There are a total of 20 coefficients exceeding the positive confidence limit (none is smaller than the negative confidence limit) out of a total of 159 coefficients for 14 stations, or approximately 13%. This is more than 5%, what should be the percentage expected in the case of random time series. Excluding 2 river stations, of Middle Boulder Creek and of Fraser, the percentage of exceedences is only around 7%.

The correlograms of 14 stations point out that the fluctuations of annual runoff and particularly of effective annual precipitations are close to those of random time series.

The mean serial correlation coefficient of each of 14 correlograms has a value which is positive and greater than the expected mean coefficient of a random time series. Several known factors which might be responsible for this positive departure should be examined before this might be attempted to be attributed to some general climatic or even solar influence. These factors are: 1) regional sampling deviations toward somewhat greater coefficients than in the case of random time series; 2) carryovers determined on an approximate basis (monthly flows are mostly used for the determination of mean recession curve, and as index discharges); 3) non-homogeneity in data (similar as in the case of the Upper Colorado River at Lee Ferry Station); and 4) some inconsistencies in data. All these factors may be sufficient to explain, at least in a major part, the difference between the Y-series and the random time series.

The room left for general climatic and solar causes in order to explain the above differences is rather small.

D. CORRELOGRAMS OF THE UPPER COLORADO RIVER AT LEE FERRY STATION

Figure 9 gives 2 correlograms for the Upper Colorado River at Lee Ferry, Arizona, using the historical annual flows, and the virgin annual flows. Only the first 10 serial correlation coefficients were computed and plotted. The confidence limits were determined by using eq. (14).

Two major conclusions may be derived from a study of these two correlograms and from a comparison between them:

1. The correction of historical annual flows by adding the annual depletions to obtain the virgin annual flows shows that all the serial correlation coefficients are decreased in virgin flows in comparison with those of historical flows. The virgin flow correlogram is more confined to the confidence interval than the historical flow correlogram. In other words, by removing the depletion in the form of a trend created by it, the correlogram of time series of virgin flows shows less dependence than that of the series of historical annual flows.

2. Though all serial correlation coefficients are positive, and 3 of them exceed by a small amount the positive confidence limit, the correlogram of virgin annual flows is rather very close to a correlogram of random time series. It has been stressed previously that the regional distribution of first serial correlation coefficients for the Upper Colorado Basin indicates somewhat greater coefficients than those of random series. The Fraser River, near Winter Park, Colorado, (Station No. 41, fig. 7), shows the same aspect of correlogram as the Lee Ferry Station. Its correlogram is closer to the positive confidence limit than to the value $r = 0$. This is similar also for the Arkansas River at Salida, Colorado, and the San Juan River at Rosa, New Mexico.

3. The conclusion that the correlogram of virgin annual flows of Lee Ferry Station is close to the correlogram of random time series may be reached here after taking into account the regional sampling factor, the fact that the determination of depletion is subject to errors, and that the correction for the carryover of water from one water year to another would also shift somewhat the correlogram of virgin annual effective precipitation still closer to that of a random time series.

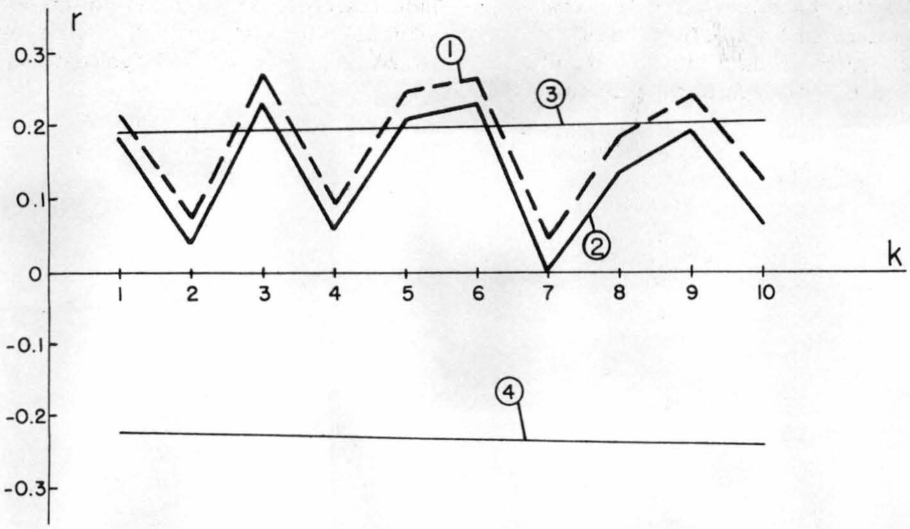


Fig. 9. Correlograms of annual flows of Upper Colorado River at Lee Ferry, Arizona, for period 1896-1959:

(1) Correlogram of historical (measured) annual flows;
(2) Correlogram of virgin annual flows;
(3) Positive confidence limit on 95% level;
(4) Negative confidence limit on 95% level.

4. for practical purposes, therefore, the fluctuation of effective annual precipitations of the Upper Colorado River at Lee Ferry as well as for other stations in the river basin itself may be considered as nearly random. It is important, however, to compile the effective annual precipitations (for virgin flows) on the Upper Colorado River Basin, taking into account the natural and the artificial carryover of water in the river basin from one water year to another. In the same time it would be useful to determine the statistical model of relation between virgin annual flows and their corresponding effective annual precipitations in a simplified form of eq. (8). Examination of these two problems is beyond the objectives of this study.

5. The mean serial correlation coefficient of virgin annual flows for 10 first coefficients $\bar{r}_v = \frac{1}{10} \sum_{1}^{10} r_k$ is 0.135, while that for the historical flows is 0.177. Depletion as factor of non-homogeneity in the data has thus produced an increase in the mean coefficient of 10 serial correlation coefficients of historical flows by 30% in comparison with that of virgin flows. This comparison shows that any inconsistency and non-homogeneity in data creates an increase in the average of the serial correlation coefficients.

The fluctuation of the serial correlation coefficients of historical annual flows of the Upper Colorado River at Lee Ferry around a value of 0.180 (or close to the positive confidence limit on 95% level), fig. 9, may be thus explained in great part by the following three basic factors:

1. Regional sampling factor, which for the arid region in the Upper Colorado Basin or around it shows a general trend toward a greater first and other serial correlation coefficients of effective annual precipitations (and also of annual flows), than it is the case with the river stations in some other regions, generally more humid than the Upper Colorado River region.

2. Depletion of annual flows, which clearly shows that a part of dependence in annual flows and effective annual precipitations is created by the constantly increasing depletion with time.

3. Carryover of water from one water year to another by both natural and artificial water storage in the river basin. This effect is proved by the previous analysis, and some correlograms in figs. 7 and 8 show this fact (comparison of correlograms for U-series and Y-series). The lack of a systematic study in computing the carryovers for the Upper Colorado River at Lee Ferry does not permit the evaluation of how this third factor influences the dependence of successive values of historical annual flows, but this effect is certainly present.

The above analysis shows also that the room left for any significant persistence or regularity of annual flow fluctuations is relatively very small. This is shown at least by the study of correlograms, after the three above factors are taken into consideration, and also after the effect of some systematic errors in data is allowed for. There is not very much room left in the amount of dependence of time series, which can justify any conclusion of the existence of hidden periodicities, at least with a statistical significance worthy to be considered in practical applications.

V. PATTERNS IN FLUCTUATIONS MEASURED BY RANGE

A. DEFINITION OF RANGE

The maximum range for a time series of length N and for the period N is defined as the difference on the cumulative curve of departures of the maximum and minimum values.

Figure 10 gives the curve of accumulated departures for the relative values U_i (modular coeffieicnets) of virgin annual flows of the Upper Colorado River at Lee Ferry. The maximum range for this time series of $N = 64$ years is defined as $R_{max} = S_{max} - S_{min}$, where S represents the values of the curve of cumulative departures, $\Sigma \Delta U_i = \Sigma(U_i - 1)$, with unity the mean value of modular coefficients U_i. In this case, according to H. E. Hurst (ref. 5), the maximum range can be conceived as the maximum accumulated storage when there is never a deficit in outflow (outflow from reservoir equal to the mean discharge), or as the maximum deficit, where there is never any storage, or as the sum of accumulated storage and deficit, when both storage and deficit exist.

The basic characteristics for the above definition of range is the use of departures from the mean value of flows for N-years also for the determination of range for shorter periods than N.

In a broader sense, any constant value U_o different from unity may be used to determine departures and the corresponding ranges on the cumulative curve of departures so defined. For the purpose of this study the values U_o may be conceived as a constant draw of flows from a reservoir, which is generally smaller then $\bar{U} = 1$, or smaller than the mean for the period of N-years.

There is also a definition of adjusted maximum range, which is defined by W. Feller (ref. 6) as the difference of maximum and minimum of the cumulative curve of departures, but with the changing mean. If the period has a length of N-years, the mean of that period is used for computing the departures and the adjusted range. An example of this maximum adjusted range is given in fig. 10 for two periods of $n = 32$ years (first half and second half of the total period of 64 years). The means $\bar{U}_1$ and $\bar{U}_2$ are plotted, and by using lines parallel to them S_{max} and S_{min} are obtained for both half periods, and then the adjusted ranges R_a^1 and R_a^2 are determined. The use of the mean $\bar{U} = 1$ for determining the ranges for both 32-year periods gives the range values R_1 and R_2, also shown in fig. 10.

All these definitions of ranges, based on $\bar{U} = 1$, or any U_o, or as the adjusted range can be used, depending on the type of problem at hand. In using the range as a statistical tool for comparing the observed time series of annual flows or the derived time series of effective annual precipitation with the random time series the range defined on the basis of the mean for the period of observation of N years will be used here exclusively.

Using the modular coefficients U_i to compute the range, the range is expressed in a dimensionless form as relative value R_{re}, and to determine the absolute value of the range R_{ab}, the relative value must be multiplied by the mean annual flow $\bar{V}$, so that $R_{ab} = R_{re} \bar{V}$.

B. DISTRIBUTION OF RANGE OF DIFFERENT PERIODS FOR A RANDOM TIME SERIES

1. General distribution function

Let a period of length of n-years be fixed for study (i.e., five years, ten years, 25 years, etc.). Let also the total period of N-years be divided in m smaller n-year periods, so that $m \cdot n + n_1 = N$, where $n_1 < n$ is a residual. If the range was determined for each of m periods on the basis of the mean for the total period N, there would be m values of the maximum range, one for each of n-year periods. The distribution of these maximum ranges (which are in fact the sum of storage and deficit of flow, if it is assumed that always the mean discharge should be available) is an important statistical tool for designing and operating large storage reservoirs.

Figure 11 shows for the homogeneous annual river flows (flows reduced on depletion conditions of 1954-1957 and expressed in modular coefficients) of the Upper Colorado River at Lee Ferry the eight values of R_8 for $n = 8$ years. There is a distribution of R_8-values, in this case ranging from 0.82 to 1.65. This corresponds to a *fixed* value 8 of n, and for a mean discharge as basic reference for computing the departures. Assuming n as a variable, the reference discharge in modular coefficient as U_o also as a variable, there is a general four variables function

$$F(R, p, n, U_o) = 0 \qquad (15)$$

with p = probability of the range R for given n and given U_o.

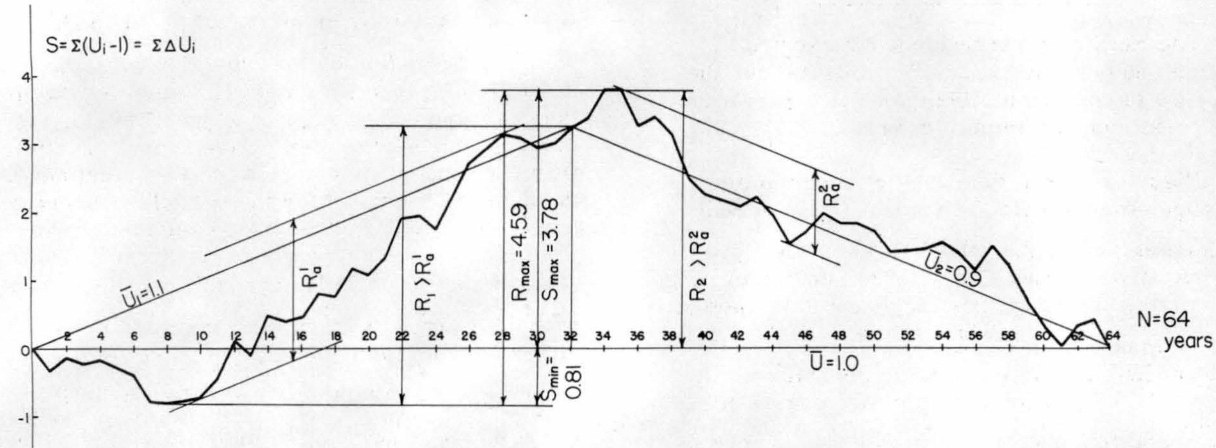


Fig. 10 Cumulative curve of departures of virgin annual flows (expressed as modular coefficients U_i) of Upper Colorado River at Lee Ferry, Arizona, from the mean $\bar{U} = 1$, for 64 years, with maximum range R_{max}, two adjusted ranges R_a^1 and R_a^2, and maximum ranges R_1 and R_2 for two periods of $n = 32$ years.

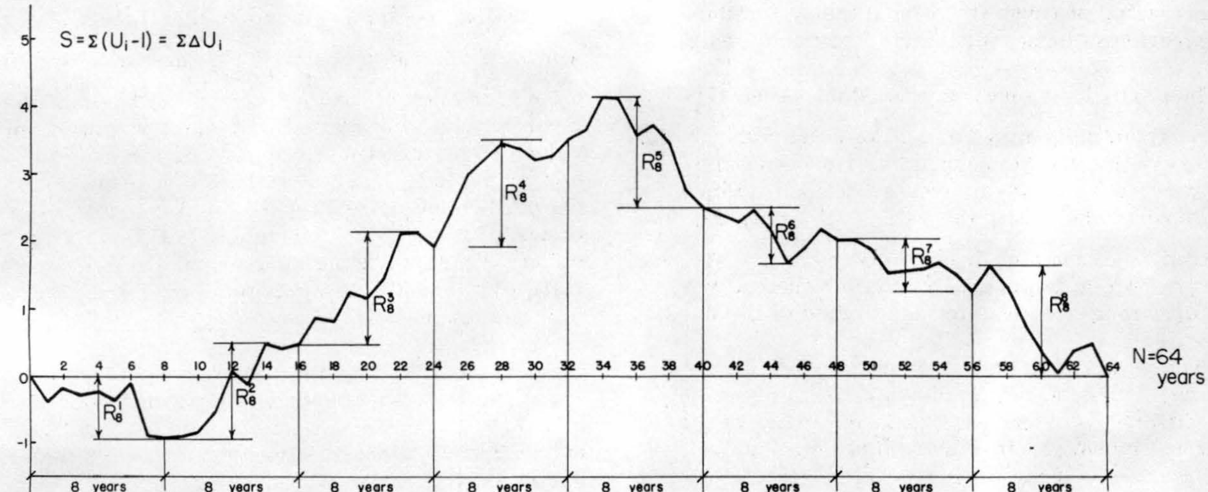


Fig. 11 Determination of ranges for 8 periods each, 8 years long for the homogeneous sample of annual river flows (reduced to depletion conditions in 1954 - 1957) of Upper Colorado River at Lee Ferry, Arizona.

The availability of a reliable function of the type of eq. (15) for a large storage project would help substantially to solve the problems of planning, design and operation of large reservoirs. This function can be determined either analytically (by using some approximations), or by numerical procedures in the case of a random time series, for a variable normally distributed, with given mean and standard deviation. It is also possible to develop that function for some other time series, derived from random time series by some processes, and also when the random variable is skew distributed (but with a known distribution function). Probably only the numerical integration of some basic probability distributions could be used in this last case.

2. Distribution of ranges of random series as yardstick distribution

The range distribution theoretically developed for random time series was used here, as in a recent study (ref. 1), as a yardstick to compare the range distribution of time series of annual flows and of effective annual precipitation with this yardstick.

The range distribution for both random and observed or derived time series was defined here by three statistical parameters: mean, coefficient of variation and skew coefficient. If the three types of parameters for given n and $\bar{U} = 1$ (mean of the time series) do not differ for the compared random and observed (or derived) time series in a significant manner, then it could be assumed that observed or derived time series are close either to random time series, or to some time series which are derived by simple processes from random time series. As an example of this derived time series eq. (8) may be used. If Y_i-series is a random time series, and b-values are given coefficients, then U_i-series is that derived time series.

3. Distribution of range for a random time series for large n

The asymptotic values for expected mean and for variance of the range of random series is given by W. Feller (ref. 6)

$$\bar{R}_n = 1.5958\ldots s\sqrt{n} \backsimeq 1.6\ s\sqrt{n} \tag{16}$$

and

$$S_n^2 = \text{var}(R_n) = 4s^2n\left(\ln 2 - \frac{2}{\pi}\right) = 0.2182\ s^2n \tag{17}$$

where s = standard deviation of time series of length N; n = length of period for which the mean range and variance of range are determined, $\bar{R}_n$ is the expected mean of range for period of length n, and S_n^2 is the variance of the range distribution.

It follows from eqs. (16) and (17) that the asymptotic value of the coefficient of variation of range distributions is a constant equal to 0.292.

The condition for the application of eqs. (16) and (17) is the large value of n (the period for which the maximum range is considered). As the regulation periods are generally small (some years), the practical application of eqs. (16) and (17) is restricted in the case of small n. The asymptotic value of $\bar{R}_n$ and S_n^2 mean that by an increase of n to infinity, the mean and variance of the range would converge to the values of eqs. (16) and (17). The use of eqs. (16) and (17) for small values of n shows, however, an upper limit for these statistical parameters, and they will be used here in this aspect.

The analysis of ranges in hydrology can be carried out from two points of view: a) as a statistical tool of comparing the observed or from it derived time series with random time series, or with series derived from random series; b) as a tool for studying and computing the storage reservoir capacities in the case of an overyear regulation. Both these views should be appled for the Upper Colorado River, the first to compare the series among them, and the second to study the potential use of large storage reservoirs (Lake Mead, Glen Canon Reservoir, etc.) on the border between the Upper and Lower Colorado River.

For the adjusted range, the asymptotic expected mean, according to H. E. Hurst (ref. 5) is

$$\bar{R}_n = s\sqrt{\frac{\pi n}{2}} = 1.25\ s\sqrt{n} \tag{18}$$

and the variance is according to W. Feller (ref. 6)

$$S_n^2 = \left(\frac{\pi^2}{6} - \frac{\pi}{2}\right) s^2 n = 0.0741\ s^2 n \tag{19}$$

4. Distribution of range of a random time series for small n

In case the random variate is normally distributed, then the mean range for n = 1 is

$$\frac{\bar{R}_1}{s} = \int_0^{\infty} \sqrt{\frac{2}{\pi}}\, R_1 e^{-R_1^2/2}\, dR_1 = \sqrt{\frac{2}{\pi}} \backsimeq 0.80 \tag{20}$$

and variance of R_1

$$\frac{S_1^2}{s^2} = \int_0^{\infty} (R_1 - \bar{R}_1)^2\, p_1(R_1)\, dR_1 =$$

$$= \left(1 - \frac{2}{\pi}\right) = 0.363 \tag{21}$$

where $\bar{R}_1/s$ and S_1^2/s^2 represents the mean and variance of a standardized variate, with variance unity. When eqs. (20) and (21) are applied to modular coefficients, with mean unity and variance s^2, then

$$\overline{R}_1 = 0.80\ s; \text{ and } S_1^2 = 0.363\ s^2$$

The range is a truncated distribution of half the normal distribution, with a skew coefficient

$$C_s = \left(2 - \frac{\pi}{2}\right)\left(\frac{2}{\pi - 2}\right)^{2/3} \backsimeq 0.995$$

The distribution of R_1 is given in fig. 12, and its statistical parameters in table 3.

For $n = 2$ the probability of range R_2 is according to author's study (ref. 1)

$$p_2(R_2) = \frac{2}{\sqrt{\pi}}\ e^{-R_2^2/4}\left[\int_0^{R_2/\sqrt{2}} \frac{e^{-t^2/2}dt}{\sqrt{2\pi}} + \sqrt{2}\, e^{-R_2^2/4} \int_0^{R_2} \frac{e^{-t^2/2}dt}{\sqrt{2\pi}}\right] \qquad (22)$$

where t = a standardized variate; R_2 = any range for $n = 2$; $p_2(R_2)$ = probability of a given range R_2.

The distribution or R_2 and its statistical parameters are computed by numerical integration of eq. (22), the distribution is shown in fig. 12, and the statistical parameters in table 3.

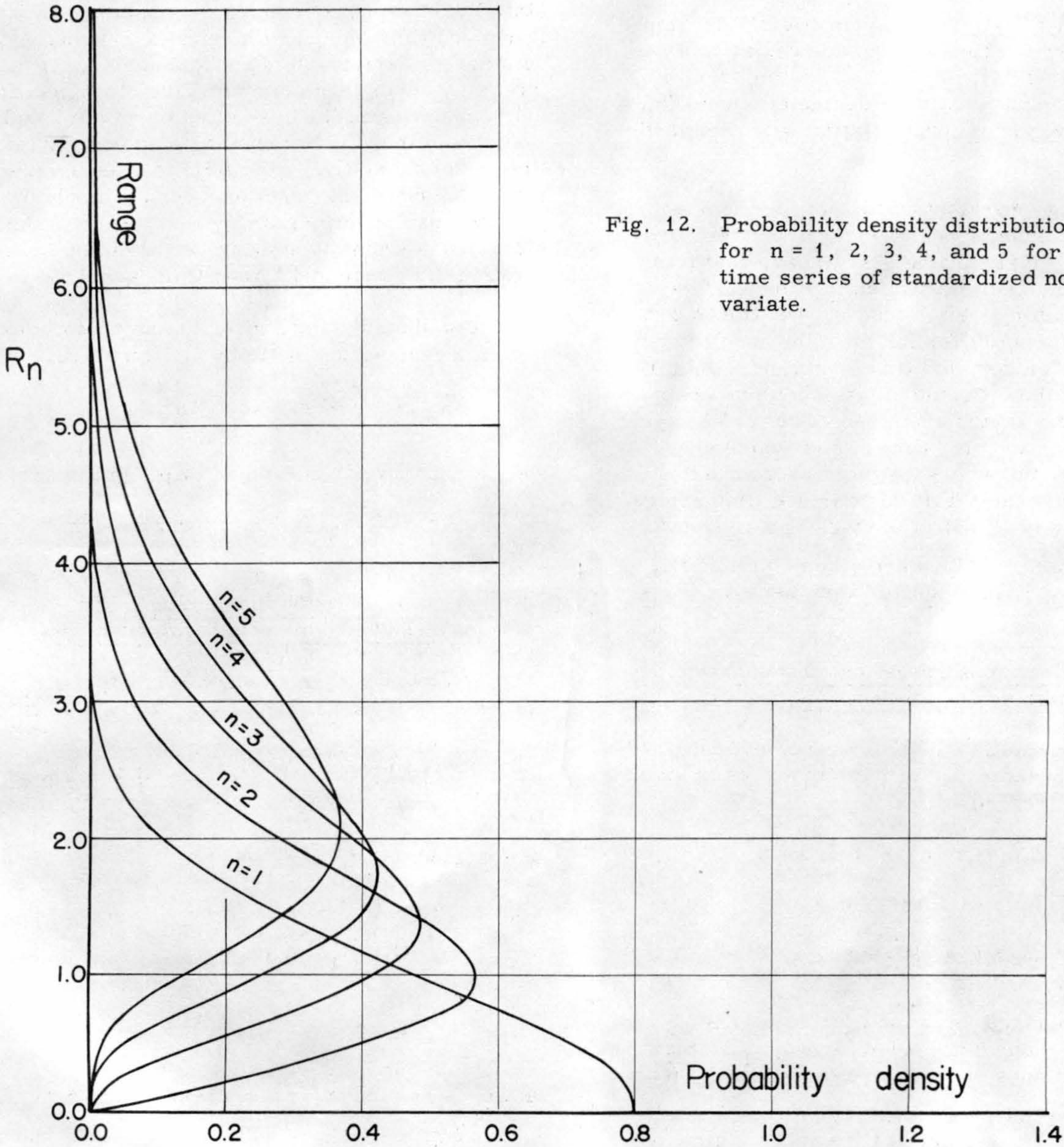


Fig. 12. Probability density distribution of ranges for $n = 1, 2, 3, 4$, and 5 for random time series of standardized normal variate.

Table 3. Statistical parameters of distribution of ranges for small n

Parameters	n = 1	n = 2	n = 3	n = 4	n = 5
	R_1	R_2	R_3	R_4	R_5
$\bar{R}_n$	0.79789	1.36257	1.84085	2.28061	2.69751
Median range	0.6725	1.2367	1.6962	2.2173	2.6916
Mode range	0.00	0.99	1.39	1.73	2.10
S_n^2	0.36439	0.59992	0.84694	1.11765	1.41536
S_n	0.60365	0.77455	0.92027	1.05719	1.18969
${}_nC_v$	0.75722	0.56844	0.49991	0.46356	0.44103
${}_nC_s$	0.98793	0.91143	0.88916	0.85106	0.79232
$\bar{R}'_n$	0.79789	1.36208	1.82272	2.22166	2.57846
$\Delta\bar{R}_n = \bar{R}_n - \bar{R}'_n$	-0.00067	0.00049	0.01813	0.05895	0.11905
$\frac{100\Delta\bar{R}_n}{\bar{R}_n}$	-0.084	0.036	0.985	2.580	4.420

The distributions and statistical parameters for R_3, R_4, and R_5 are also expressed in similar forms as eq. (22), then numerically integrated or computed, and the distributions are shown in fig. 12, and parameters in table 3.

The statistical parameters of ranges for standardized variate: mean, coefficient of variation and skew coefficient for n = 1, 2, 3, 4, and 5 are given in fig. 13, and median, mode, variance, and standard deviation for n = 1, 2, 3, 4, and 5 are given in fig. 14, taken from author's study (ref. 1).

A. A. Anis and E. H. Lloyd (ref. 7) have developed the expression for the mean value of the range for finite small n as

$$\bar{R}_n = \sqrt{\frac{2}{\pi}} \sum_{i=1}^{n} i^{-1/2} \qquad (23)$$

with i integers from 1 to n .*

* The authors of ref. 7 give the coefficient of eq. (23) both as $\sqrt{2/\pi}$, and $1/\sqrt{2\pi}$, and A. A. Anis in two successive papers, Biometrica vol. 42, 1955, pages 96-101, and Biometrica vol. 43, 1956, pages 79-84, gives always the value of coefficient as $1/\sqrt{2\pi}$. The author of this paper has found out, that to his approach the value of $\sqrt{2/\pi}$ of eq. (23) was correct.

For n = 4, $\bar{R}_4 = \sqrt{\frac{2}{\pi}}\left(\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{4}}\right) = 2.22166$

The values $\bar{R}_1 \ldots \bar{R}_5$ computed by eq. (23) are given in table 3 as $\bar{R}'_n$, and represented in fig. 13, as curve (3).

The comparison of the mean values of the range distribution for different small n is given for three types of values: a) asymptotic values according to eq. (16); b) values obtained by numerically integration of exact distributions, eqs. (20), (21), (22), and similar; and c) values obtained from analytical expression of eq. (23). Though there are some departures among the curves (b) and (c), it could be assumed that eq. (23) approximates closely the exact values given by curve (b), but for very small n. The asymptotic values, curves (a), depart greatly from the exact values for very small n , and should not be used in practical computations. For n = 1 the asymptotic value is double the exact value for the mean range. Table 3 gives the difference of means for (b) and (c) and also the departure in percentage of the $\bar{R}'_n$ from $\bar{R}_n$ The formula of eq (23) should be used with caution for middle values of n.

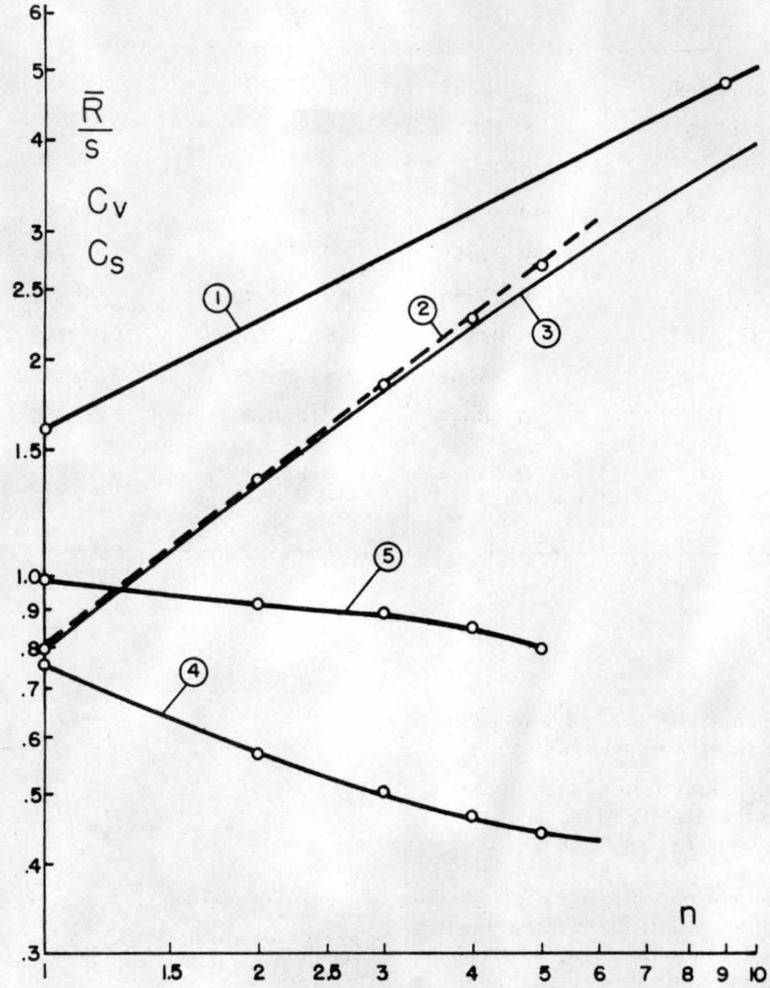


Fig. 13. The statistical parameters of range distribution of random time series of normal variable for small n (1-10).

(1) Mean range ($\overline{R}/s$), asymptotic values
(2) Exact values of mean range, obtained by numerical integration
(3) Mean range obtained by formula, given by Anis and Lloyd
(4) Exact values of coefficient of variation
(5) Exact values of skew coefficient.

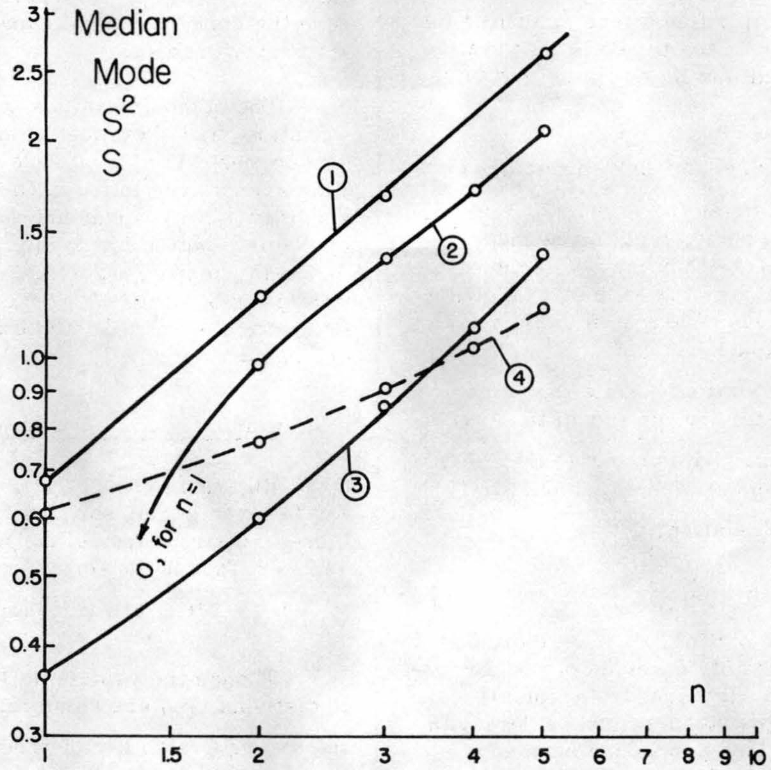


Fig. 14. The statistical parameters of range distribution (median, mode, variance, standard deviation) for random time series of normal variable.

(1) Medium range
(2) Mode range
(3) Variance of range
(4) Standard deviation of range.

C. DISTRIBUTION OF MAXIMUM RANGE OF EFFECTIVE ANNUAL PRECIPITATION FOR 14 RIVER GAGING STATIONS OF THE UPPER COLORADO RIVER BASIN AND AROUND IT

1. Procedure

The sums of departures $\Delta U_i = U_i - 1$ for modular coefficients of annual flows, and sums of departures $\Delta Y_i = Y_i - 1$ for modular coefficients of effective annual precipitations were computed for 14 river gaging stations in the Upper Colorado River Basin and around it (stations listed in table 1) on a digital computer.

The cumulative curves of ΔY_i-departures were plotted, and the ranges for n = 1, 2, 3, 4, 5, 7, 10, 15, 20, 25 were determined for as many periods m of length n as they may be included in the total length N of series for each station, without overlapping of periods of length n . Thus series of the corresponding values $R_1, R_2, \ldots . R_{25}$ were obtained for each station. While the number m of R-values were greater for R_1 (equal to N-1) the number m decreased with an increase of n , so that for R_{25} this number was either one ($N < 50$) or two ($N \geq 50$) for each station.

After the values $R_1, R_2, \ldots . R_{25}$ had been computed for each station, they were divided by the standard deviation of Y-series of the corresponding station, so that all values R/s for all stations refer to a unique standardized variate with the variance $s^2 = 1$. This procedure enabled the pulling together of values R for given n of all 14 stations in one large sample. In this way, the total number of R-values was:

n = 1	for R_1	$m_1 = 646$
n = 2	" R_2	$m_2 = 327$
n = 3	" R_3	$m_3 = 221$
n = 4	" R_4	$m_4 = 164$
n = 5	" R_5	$m_5 = 132$
n = 7	" R_7	$m_7 = 94$
n = 10	" R_{10}	$m_{10} = 65$
n = 15	" R_{15}	$m_{15} = 43$
n = 20	" R_{20}	$m_{20} = 29$
n = 25	" R_{25}	$m_{25} = 28$

For each of these 10 distributions of the mean, coefficient of variation, and skew coefficient were computed.

2. Results

The distributions of ranges for 14 stations, with all corresponding values pulled together, are given in fig. 15, and for the first 5 values of n (1, 2, 3, 4, 5) the distributions of ranges of random time series (fig. 12) are plotted also in order to compare the distributions of range of Y-series for 14 stations with them.

The computed values of mean, coefficient of variation, and skew coefficient are plotted in fig. 16. The corresponding statistical parameters for random time series are plotted in this figure also, and specifically: a) mean and coefficient of variation of R-distribution for asymptotic range values; b) mean, coefficient of variation and skew coefficient for 5 values of n , computed by numerical integration of range distribution; and c) the mean, computed by eq. (23).

3. Comparison of Y-series and random series

Figure 15 shows that the distributions of ranges for 14 river gaging stations for small n are very close to distributions of ranges of random time series. This is especially valid for values R_2 , and R_3 , but a little less valid for R_1 , R_4 and R_5 .

Though the values m (sample sizes for R-distributions) are relatively large for R_1 to R_5 , the sampling stability of range is, however, relatively small. This is mainly due to the fact that the concurrent values of Y-series (values for the same water year) of 14 stations in the Upper Colorado River Basin and around it have been pulled together in order to increase the sample sizes. The concurrent annual flows and the derived concurrent effective annual precipitations of 14 stations are not independent. There is a simultaneity of occurrence of wet and dry years in a large region, which creates the dependence among the concurrent annual flows and derived effective annual precipitations of 14 stations. It means that m-values given above are fictitiously large, or that the effective size of samples (under assumption that the reduced size takes care of that dependence) are much smaller than given above. This fact may partly explain why there is a relatively large variation of frequencies for different range intervals in range distributions of Y-series.

It can be also seen from the comparison of distributions in fig. 15, that both the mean and the standard deviation of range distributions of Y-series increase constantly with the increase of n , as it is the case with a random time series.

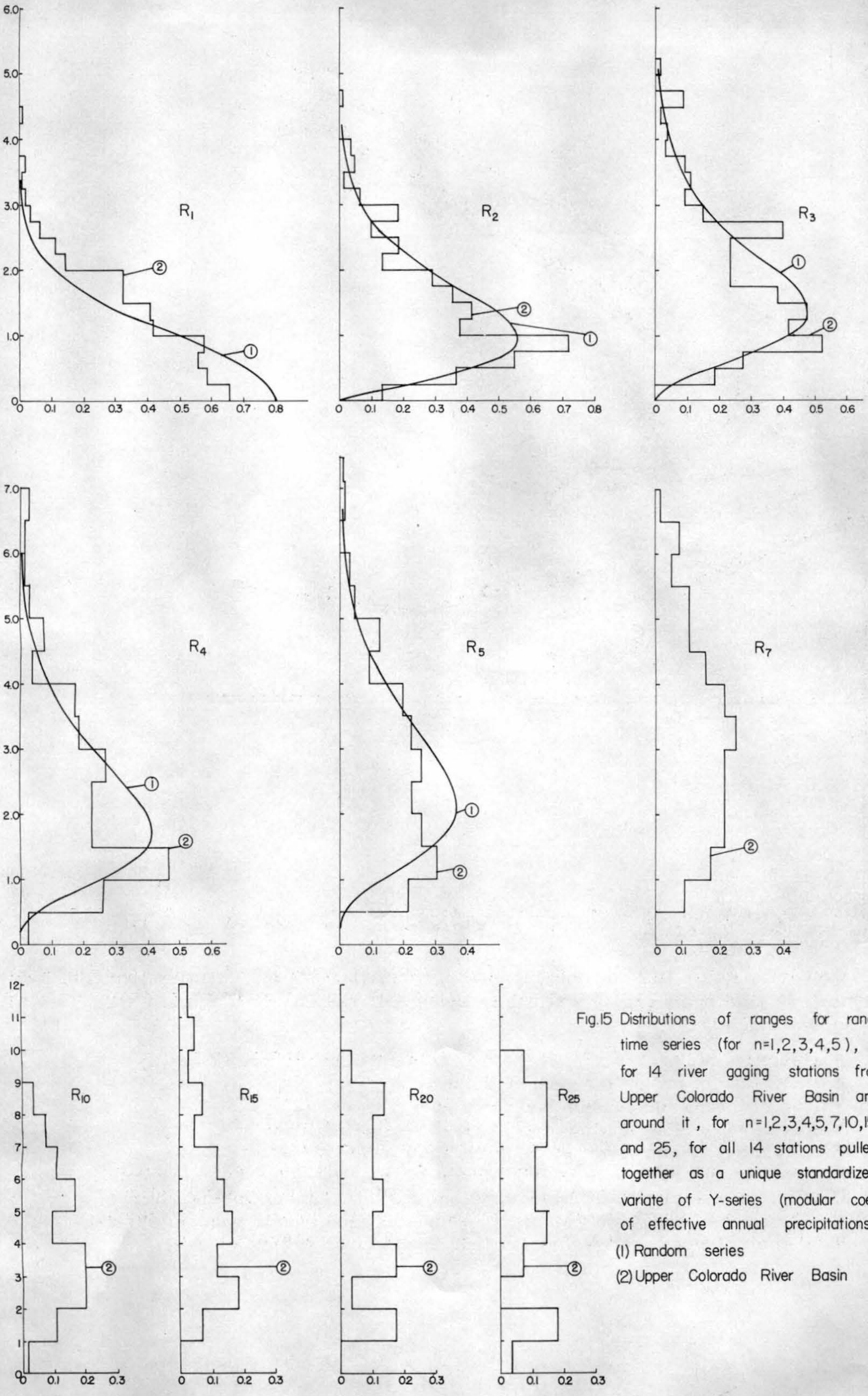


Fig.15 Distributions of ranges for random time series (for n=1,2,3,4,5), and for 14 river gaging stations from Upper Colorado River Basin and around it, for n=1,2,3,4,5,7,10,15,20 and 25, for all 14 stations pulled together as a unique standardized variate of Y-series (modular coefficients) of effective annual precipitations.
(1) Random series
(2) Upper Colorado River Basin

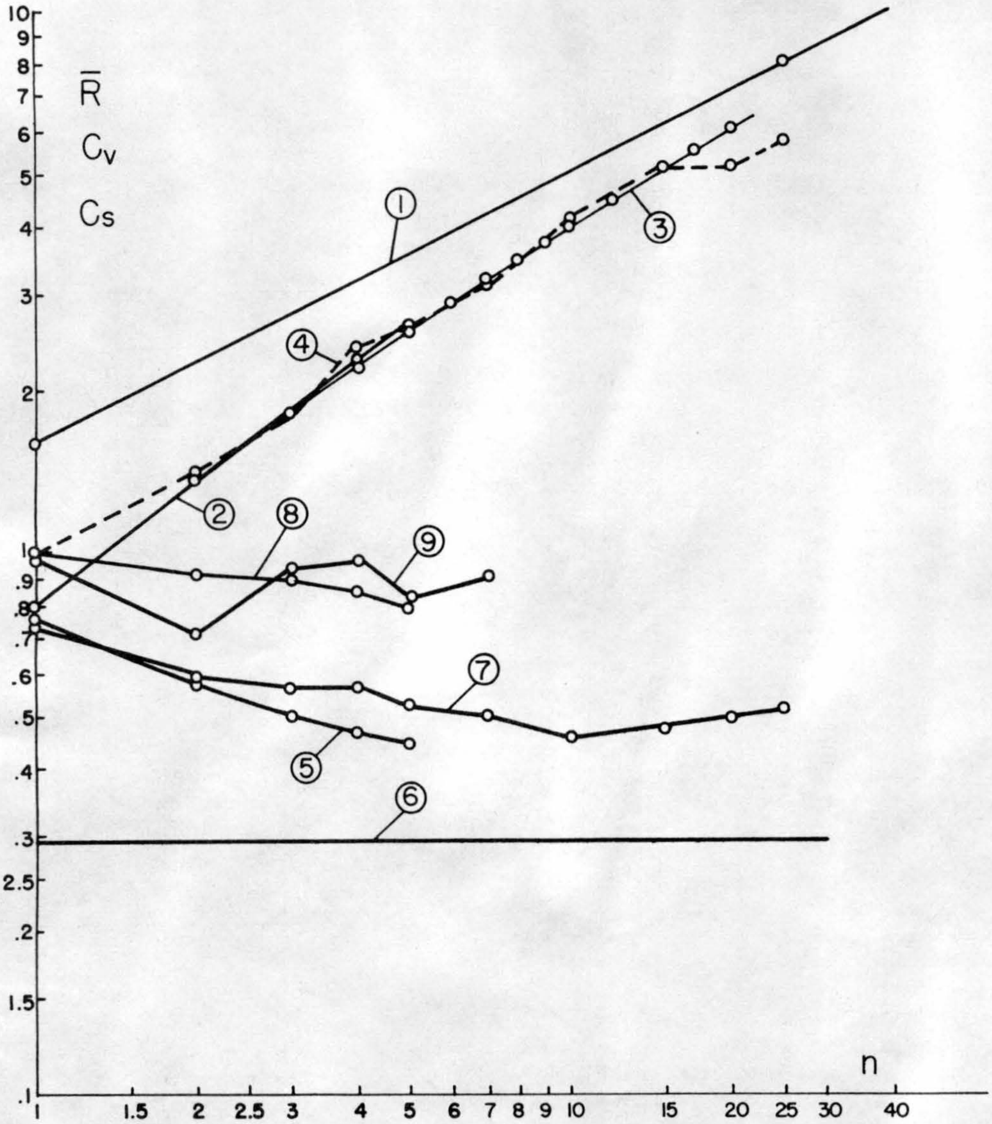


Fig. 16. Comparison of statistical parameters of range distributions for effective annual precipitations (Y-series) of 14 river gaging stations in Upper Colorado River Basin and around it, with those of random time series:

(1) Asymptotic values ($1 \cdot 6 \sqrt{n}$) of the expected mean of random series;
(2) Exact values of means for random series;
(3) Values of means for random series computed by eq. (23);
(4) Means of range distributions for 14 river gaging stations;
(5) Exact values of coefficient of variation for random series;
(6) Asymptotic constant value for coefficient of variation for random series;
(7) Coefficient of variation of range distributions for 14 river gaging stations;
(8) Exact values of skew coefficients for range distributions of random series;
(9) Skew coefficients of range distributions for 14 river gaging stations.

Taking into consideration the following factors: a) regional sampling; b) errors in computing the carryovers from one water year to another (which are determined by an approximate method); and especially c) non-homogeneity of data (created by man-made changes in river basins), it can be assumed here that the distributions of range of the effective annual precipitations in the Upper Colorado River Basin are very close to those of random time series.

The comparison of statistical parameters of distributions of ranges of Y-series for 14 rating gaging stations in the Upper Colorado River Basin and around it, with the statistical parameters of range distributions of random series shows that the mean values of ranges of Y-series are nearly identical with mean values of range of random time series, because the curves (2) and (4) of fig. 16 are very close, at least for values $n = 2 - 15$.

The mean range for R_1-distribution of Y-series is 1.00, while that for random time series is 0.80 (fig. 16). The R_1-distribution of Y-series (fig. 15) has smaller values of maximum probability than that of random series. The reason for this difference may be mainly the approximate determination of carryovers, but the presence of non-homogeneity in data may be responsible partly for differences.

The asymptotic values of mean range, curve (1) in fig. 16, are much larger than the values of the mean ranges of Y-series. Therefore, they should not be used in practical applications for small n. It is also evident that the mean ranges computed by eq. (23) approximate well the computed mean ranges of Y-series.

The comparison of the coefficients of variation of distributions of ranges of Y-series with those of random time series shows that the departures between them are not too great, curves (5) and (7), fig. 16, while the asymptotic constant value of the coefficient of variation, curve (6), is much smaller than the observed values, and therefore, should not be used for practical purposes in the case of a small n.

The comparison of skew coefficients of distributions of ranges of Y-series with those of random time series shows, in the limits of the sampling instability of the third statistical moment of distributions of Y-series, that the closeness of two curves, (7) and (8), fig. 16, is sufficiently good to derive the conclusion that even the skew coefficients are nearly identical for two range distributions.

The general comparison of curves in fig. 16 points out that the differences between range distributions of Y-series and those of random series are small and could be neglected for practical purposes.

D. DISTRIBUTION OF MAXIMUM RANGE OF ANNUAL VIRGIN FLOWS OF UPPER COLORADO RIVER AT LEE FERRY

1. Procedure

The values of maximum ranges for U-series (given as modular coefficients) of the virgin annual flows of the Upper Colorado River at Lee Ferry station, Arizona, were computed from the cumulative curve of departures $\Delta U_i = U_i - 1$, according to definition of the range, and for $n = 1-10$. For each group of R-values the mean, and coefficient of variation were also computed.

2. Results

The distribution of ranges for $n = 1-5$ is given in fig. 17, and the distributions of random series are plotted also for comparative purposes. For this comparison of distributions the values of ranges of U-series were divided by $s = 0.278$, the standard deviation of that series, and frequency density for each range-interval is computed, in order to arrive at a comparable distribution with the probability distribution of range of random series, given in fig. 12.

The values of the mean and of the coefficient of variation for ranges of U-series for $n = 1-10$ are plotted in fig. 18, and the same values for asymptotic range, numerically determined distributions of range, and also for the mean values obtained by eq. (23) are plotted in fig. 18 for comparative purposes.

3. Comparison of ranges of U-series of Lee Ferry Station with those of random series

Regardless that the U-series is compared with random time series in figs. 17 and 18, instead of comparing the Y-series (which was not available for Lee Ferry Station) with random time series, the differences between compared ranges are relatively small.

The distributions of ranges of U-series for $n = 1-5$, fig. 17, show a very clear trend of their approaching the theoretical probability distributions of random time series, of a variable normally distributed. The sampling stability of the range distributions of U-series is small, mainly because the sample sizes are small (R_1 with $m = 63$; R_2 with $m = 32$; R_3 with $m = 21$; R_4 with $m = 16$; and R_5 with $m = 13$).

The skew coefficient of virgin annual flows at Lee Ferry is very small, $C_s = 0.092$, so that it can be assumed that these flows are nearly normally distributed, and that the skewness does not play a significant role in the differences of range distributions.

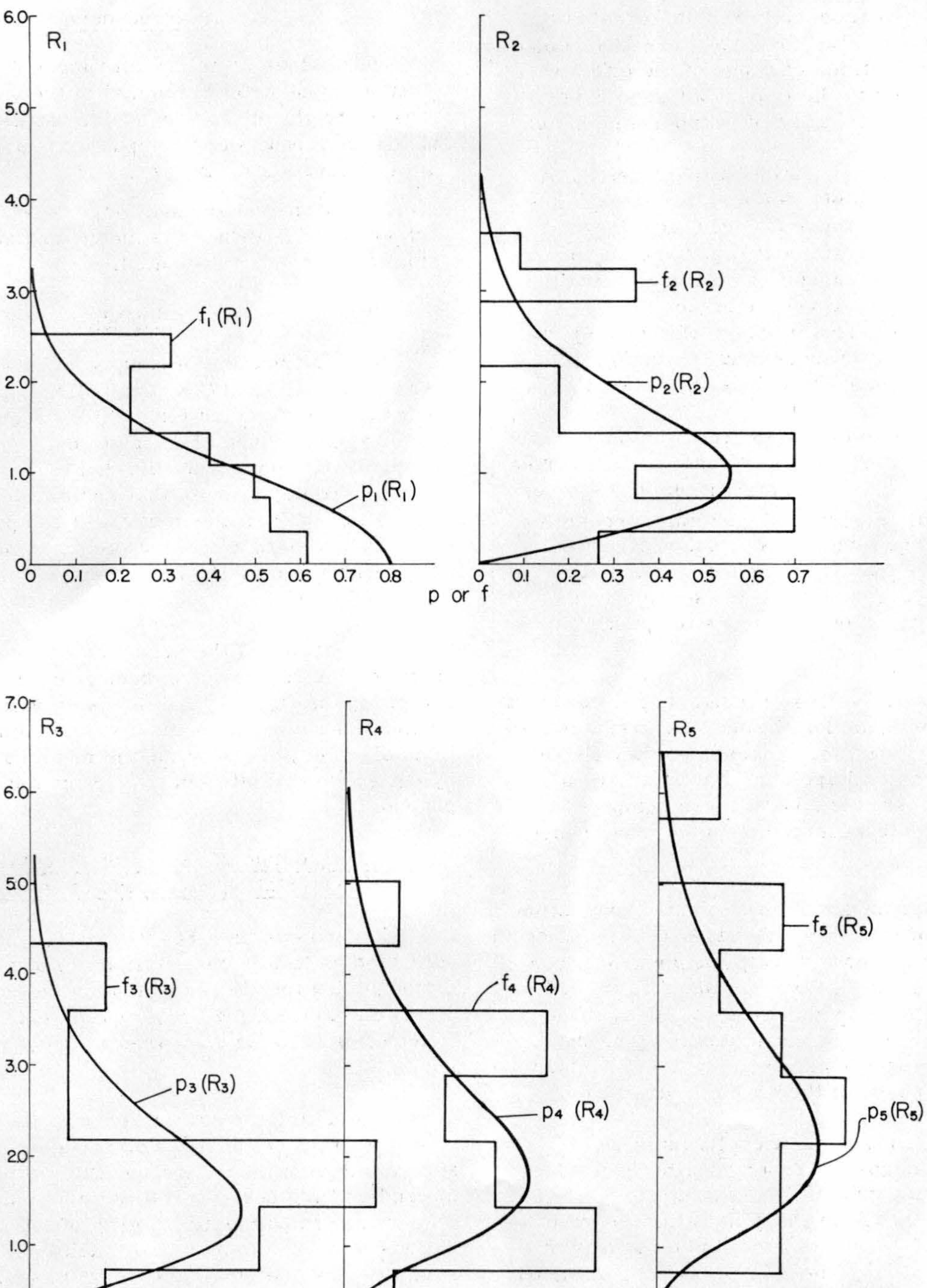


Fig. 17 Frequency distributions of ranges, f (R), for n = 1, 2, 3, 4 and 5 for virgin annual flows of Upper Colorado River at Lee Ferry, Arizona (histograms) and probability distributions, p (R), of ranges of random time series (continuous lines) given as densities. R-values are in demensionless form for standardized variate (R_n/s).

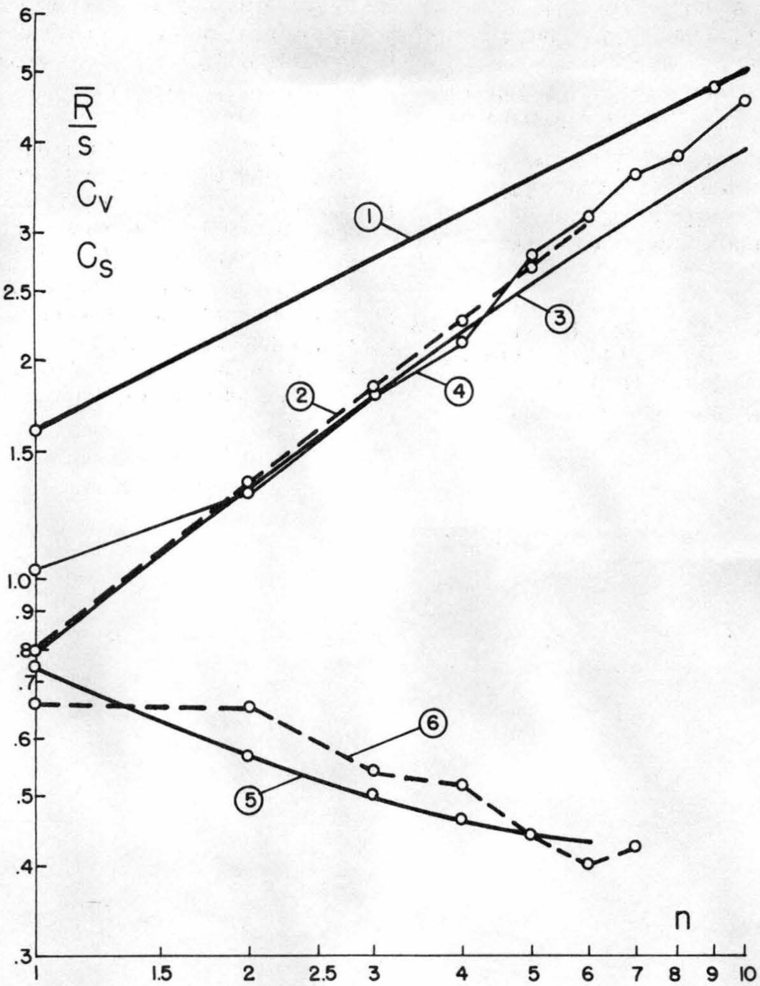


Fig. 18. Comparison of statistical parameters of range distributions for different n of virgin annual flows of Upper Colorado River at Lee Ferry, with the corresponding values of random time series.

(1) Mean range ($\bar{R}/s$), asymptotic values;
(2) Exact values of mean range;
(3) Mean range obtained by formula given by eq. (23);
(4) Mean range of virgin annual flows of Upper Colorado River;
(5) Coefficient of variation, exact values;
(6) Coefficient of variation of virgin annual flows of Upper Colorado River.

The comparison of the basic statistical parameters of distributions of ranges of U-series for Lee Ferry with those of random series, fig. 18, shows that the means and coefficients of variation of range of U-series are very close to those of exact distributions of range of random series, except for $\overline{R}_1$ and ${}_1C_v$, for which $(n = 1)$ the mean and the coefficient of variation of range of U-series are different than those of random series. The reason for this departure may be the water carryover from one water year to the next, apart from possible sampling deviations.

Both figures, therefore, suggest the conclusion that the virgin annual flows for the Upper Colorado River at Lee Ferry Station follow fluctuations which are very close to fluctuations of random time series. The departures between the distributions of range, though small, may be first explained by the water carryover, errors in computing the depletions, and by some probable systematic errors in data, before any other cause (from the atmosphere or beyond it) should be considered.

For the practical purposes, the fluctuation of virgin annual flows, or of the homogeneous annual flows (i.e., the historical sample of Lee Ferry Station, 1896-1959, reduced to depletion conditions of 1954-1957) of the Upper Colorado River at Lee Ferry, may be considered as nearly random. It can be also concluded that the fluctuations of effective virgin annual precipitations at Lee Ferry are still closer to fluctuations of random time series, than is the case with fluctuations of annual flows.

VI. GENERAL HYDROLOGICAL CHARACTERISTICS OF THE UPPER COLORADO RIVER BASIN AT LEE FERRY STATION, ARIZONA

The previous analysis of some general aspects of fluctuations of annual runoff in the Upper Colorado River Basin permits one to point out hydrological characteristics of the Upper Colorado River Basin at the Lee Ferry Station. These characteristics will be discussed as frequency distributions, patterns in sequence of annual flows, and some other characteristics of flow.

A. FREQUENCY DISTRIBUTION OF ANNUAL FLOWS

Four different annual flows may be distinguished at Lee Ferry Station for practical applications: 1) historical or measured annual flows (data supplied by the U.S. Geological Survey); 2) virgin annual flows, assuming that these flows would occur in the past, if there were no man-made changes in the river basin, or these would be the flows with no depletion in the runoff; 3) actual homogeneous sample of annual flows, defined as the reduced virgin annual flows to the conditions of depletion, which are valid for the actual river basin development and man-introduced changes; and 4) expected homogeneous samples in the future, obtained by reducing the virgin flows from the past records to the conditions of depletion which will prevail at different stages of the future river basin development.

The classical hydrological analysis of data of historical flows is not valid, when the samples are non-homogeneous as is the case of the Lee Ferry Station. The historical flow data should be considered in that case only as the basic material for computation of virgin flow or of homogeneous flow samples.

As the computation of depletion involves the study of all losses of water produced by river basin development and mean's activity; and as many of these losses are difficult to determine accurately without a great effort, the precision of data in the computed homogeneous samples of annual flows depends highly on the accuracy of computed depletion.

It would be very useful to develop standard procedures for the computation of annual depletions in the Upper Colorado River Basin on rigorous hydrological principles and methods.

Figure 19 shows the distribution of annual flows for the Upper Colorado River at Lee Ferry Station, and for three types of annual flows: historical flows, virgin flows, and homogeneous sample reduced to the conditions of depletion in the period 1954-1957. The coefficients of variations are respectively 0.297, 0.278, and 0.305, and corresponding skew coefficients are 0.018, 0.092, and 0.100. The skew coefficients are low, and this fact justifies the application of normal function for the frequency or probability distributions of annual flows at this station.

B. FREQUENCY DISTRIBUTION OF 10-YEAR MEAN ANNUAL FLOWS

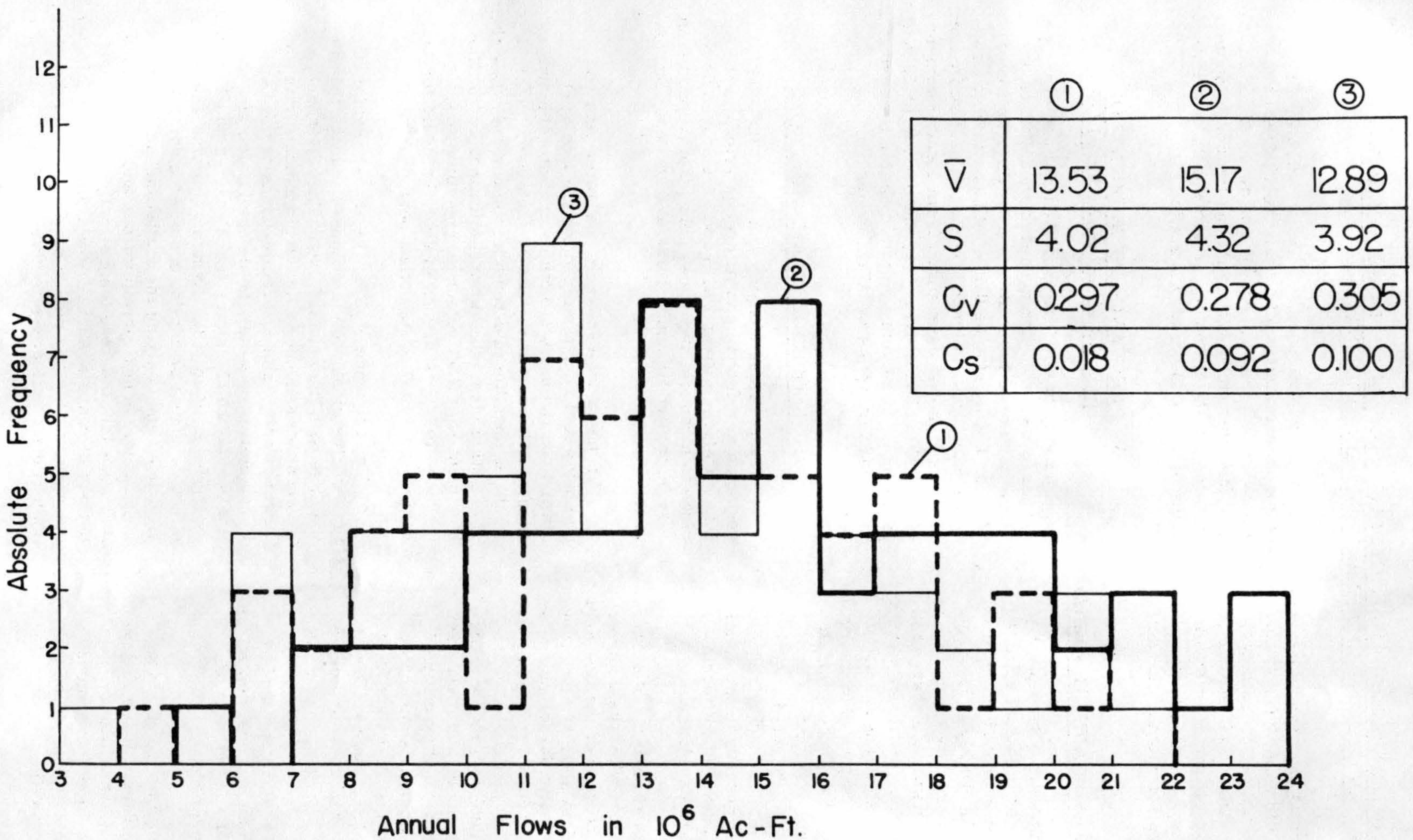
Having in mind that the distributions of virgin annual flows, or of homogeneous annual flows are close to a normal distribution, the distribution of 10-year means may be derived theoretically, as it was shown by L. Leopold (ref. 8). The distribution of 10-year means determined from a small sample of available observations (say six periods of 10 years in a sample of 64 years) is not reliable because of a very small sample size of six values.

C. SEQUENCE OF ANNUAL FLOWS

The previous analysis has shown that the sequence of effective annual precipitations in the Upper Colorado River Basin and around it is very close to a random sequence. This general conclusion may be applied to annual flows at Lee Ferry Station also.

If the carryovers from one water year to another would be determined for the Lee Ferry Station, on the basis of both natural storage and artificial storage at the end of each water year, it would be possible to determine the time series either of effective virgin annual precipitations, or of the homogeneous effective annual precipitations (reduced to the conditions of depletion valid for the period 1954-1957 or any other period).

The sequence of effective annual precipitation thus will be closer to the random sequence than is the case with the annual flows. The difference between two sequences, of virgin annual flows and virgin effective annual precipitations at the Lee Ferry Station, enables the determination of a statistical model which relates the U-series and Y-series, on the basis of eq. (8). Assuming further that the effective annual precipitations are random in sequence, and applying the statistical model relating effective annual precipitations and annual flows, it is possible to derive theoretically both the distributions of range of annual flows, and the annual flow distributions for any period of n-years, starting from normal distribution of effective annual precipitations.



	①	②	③
$\bar{V}$	13.53	15.17	12.89
S	4.02	4.32	3.92
C_V	0.297	0.278	0.305
C_S	0.018	0.092	0.100

Fig. 19 Distribution of annual flows of Upper Colorado River at Lee Ferry, Arizona:

(1) Historical annual flows;

(2) Virgin annual flows;

(3) Actual homogeneous sample of annual flows (virgin flows reduced to depletion conditions in 1954-1957), with respective means (in 10^6 acre-feet), standard deviations (in 10^6 acre-feet), coefficients of variation, and skew coefficients.

This probability method enables the analysis of flow sequence at Lee Ferry Station without using either the small sample method (historical method) or the synthetic hydrology method (in generating artificially the sequences of annual flows).

D. HYDROLOGICAL CHARACTERISTICS FOR OPERATION OF LARGE RESERVOIRS BETWEEN UPPER AND LOWER COLORADO RIVER

The previous analysis shows that the probability methods based on the properties of random time series may be and should be used for predicting the expected future flow characteristics and the expected future needs for storage capacities of the Upper Colorado River Basin. These probability methods may be used for carrying out the operation of large reservoirs, especially of Lake Mead and Glen Canyon Reservoir.

If the 10-year period is selected as the unit for considering the future expected flows, the probability distribution of 10-year mean flows, and the distribution of range for the period of 10 years are two basic probability distributions to be used in probability methods. The simultaneous occurrence of a given 10-year mean flow and of a given 10-year range needs, however, the computation of a joint bivariate probability distribution: simultaneous occurrence of both a given 10-year mean and a given 10-year range. This joint distribution may be computed by using random series of normal distributed variates of annual flows, under condition that the following information was available: 1) Mean of the effective annual precipitation of homogeneous sample; 2) Standard deviation of distribution of these effective annual precipitations; and 3) The statistical model which relates the effective annual precipitation and annual river flows.

VII. GENERAL CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSION

The analysis of hydrological characteristics in fluctuation of annual river flows in the Upper Colorado River Basin which was made in this study leads to the following conclusions:

1. Distribution of first serial correlation coefficient, correlogram analysis, and distribution of maximum range have shown that the sequence of effective annual precipitation of the Upper Colorado River Basin is very close to random sequence. This is shown by comparison of fluctuations of effective annual precipitation on river basins (conceived as precipitation minus evapo-transpiration) with the fluctuation of random time series. The sequence of effective annual precipitations is, therefore, governed by a chance process, with no significant regular fluctuation patterns.

2. Most of the dependence between the successive values of annual flows in the Upper Colorado River Basin can be explained: a) by changing of water carryover from one water year to another in the form of different water storage in a river basin; b) by non-homogeneity in data; c) by some systematic errors in compilation of annual flows; and d) by regional sampling. After these factors are taken care of, the room left for the causes of this linkage, which come either from the atmosphere or solar activities, remains small.

3. There is no statistical evidence that the fluctuations of annual flows may be composed of hidden periodicities, or of some regular patterns in the fluctuations, which can be extrapolated in the future with a reasonable expectancy that they would occur and would be verified by future flow records.

4. Reliable forecasts of future annual flows (of order 2-15 years, or more) by methods which are based on extrapolation of regular patterns in annual flow fluctuations (for instance, of hidden periodicities, or sun-spot cycles) do not seem possible.

5. For periods longer than one year, probability methods are the best techniques for computing the chances of occurrence in the near future of given flows. Due to the fact that the randomness dominates the sequence of annual flows in the Upper Colorado River, the use of properties of random series, or of series derived from random series by some known process, is feasible for the hydrologic analysis of annual flows of the Upper Colorado River Basin for purposes of planning, design, and operation of water resources projects.

6. Long-term forecasts of annual flows of the Upper Colorado River Basin (for periods longer than one year) will be possible only if based on a relationship of physical factors, linked as causes and effects, but where there is a time lag between the occurrence of causes and the occurrence of their effect. Eventual linkage between effective annual precipitations on river basins with some variables of atmospheric circulations, or with some factors of ocean, or even solar activity, may be related in such a manner that the observation of these variables or factors precedes the occurrence of the effective annual precipitations for a sufficient time lag. Forecasts then may be based on observations of some quantities prior to the occurrence of river flows.

7. Non-homogeneity in data of annual flows for some gaging stations in the Upper Colorado River Basin, especially at the Lee Ferry Station, is an important hydrological characteristic of this river basin. This non-homogeneity, defined as the man-made changes in the flows in the form of runoff depletion, affects substantially the characteristics of available records of river flows. This non-homogeneity should be determined through depletion studies with the best available hydrologic methods and with sufficient accuracy, and removed from data.

8. Carryover of water from one water year to the next in the Upper Colorado River Basin is also an important hydrological characteristic which greatly affects the linkage between successive values of annual flows. If the carryover was to be determined for the Lee Ferry gaging station of the Upper Colorado River, the effect of carryover on the linkage could be evaluated, and a statistical model for the relation of annual flows and annual effective precipitations could be derived. This would enable the use of probability properties of random variables for deriving the probability characteristics of annual river flows.

9. Average flow of historical annual flows at Lee Ferry Station for the period 1896-1959 is 13.53 million acre-feet, while the average value for the virgin annual flows is 15.17 million acre-feet. The average depletion in that period of 64 years was 1.64 million acre-feet, or 10.8 per cent of the virgin flow of that period (12.1 per cent of historical flows). If the depletion conditions which are valid today were to be applied throughout the period of 64 years, the average annual flow in the past for 64 years would be 12.89 million acre-feet, or depletion would be 2.28 million acre-feet (15 per cent of the virgin flow). The expected mean of actual annual flows in the

Upper Colorado River at Lee Ferry Station is, therefore, 12.89 million acre-feet, or 15 per cent less than the average virgin annual flow.

10. A depletion model for Lee Ferry, defined as the annual depletion versus the virgin annual flow, may be approximated by a straight line which changes with time. The extrapolation of the depletion model in the future will permit the study of the Upper Colorado River projects on the basis of probability of occurrence of future annual flows.

B. RECOMMENDATIONS

To understand better the hydrological characteristics of Colorado River flows, and for forecasting the future runoff at different places, continued hydrological and hydrometeorological studies are necessary and useful. Some recommended studies are:

1. Development of methods sufficiently accurate for current use in determining depletion of river flows by man-made changes in the river basin. This would improve the accuracy of derived virgin annual flows, or of any other homogeneous sample of annual flows.

2. Determination of carryovers of water from one water year to another at the important river gaging stations in the Upper Colorado River Basin by computing natural and artificial water storage at the beginning of each water year. This would permit the computation of effective annual precipitations, which is closer to random fluctuations than annual flows. This would enable the design of statistical models for linkage between annual precipitations and annual flows. This would furnish the basic material for rational application of probability methods in the analysis of Colorado River flows.

3. Selection, improvement, or development of probability methods to be used in hydrological studies of the Upper Colorado River Basin. This study would bring a replacement of current historical hydrologic methods, or of the synthetic hydrology method by the more reliable probability methods.

4. Forecasts of future flows by analysis of relationships between physical factors. Future forecasting studies should be directed toward searching for relationships between such factors as ocean temperatures, variables connected with activities in the lower and upper atmosphere, and with solar activity.

5. Selection and maintenance of some river gaging stations in the Upper Colorado River Basin as virgin flow stations (benchmark stations). This would permit the study of changes in hydrological conditions in the Upper Colorado River Basin with time.

REFERENCES

1. Yevdjevich, V. M., Fluctuations of Annual River Flows, manuscript of a study to be published, which was carried out by a sponsorship of the U. S. National Bureau of Standards (1958-1959) and the U.S. Geological Survey (1959-1960), Washington, D. C.

2. Anderson, R. L., Distribution of the Serial Correlation Coefficient, Mathematical Statistics Annals, v. 8, no. 1, Mar. 1941, p. 1-13.

3. 83rd Congress, 2nd Session, House Document No. 364, Colorado River Storage Project, U. S. Gov. Printing Office, Washington, D. C., 1954, Chapter IV, p. 141-155.

4. Kendall, U. G., The Advanced Theory of Statistics, v. 2, 1951, p. 404.

5. Hurst, H. E., Long-term Storage Capacity of Reservoirs: Am. Soc. Civil Engineers Trans., 1951, v. 116, no. 2447, p. 776.

6. Feller, W., The Asymptotic Distribution of the Range of Series of Independent Random Variables: Math. Statistical Annals, v. 22, Sept. 1951, p. 427.

7. Anis, A. A., and Lloyd, E. H., On The Range of Partial Sums of a Finite Number of Independent Normal Variates: Biometrica, no. 40, 1953, p. 35-47.

8. Luna, Leopold B., Probability Analysis Applied to a Water-Supply Problem, U. S. Geological Survey Circular 410, Washington, D. C., 1959.