PATTERNS IN SEQUENCE OF ANNUAL RIVER FLOW AND ANNUAL PRECIPITATION

by

Vujica M. Yevdjevich

EXTRACT OF PUBLICATION No. 63

Berkeley 1963

International Association of Schentific Hydrology

PATTERNS IN SEQUENCE OF ANNUAL RIVER FLOW AND ANNUAL PRECIPITATION (*)

Vujica M. YEVDJEVICH (**)

ABSTRACT

Nonrandomness in time series of annual values of flow and precipitation is investigated. Simplified mathematical models are derived for the relationship of annual flow, annual effective precipitation (defined as precipitation minus evaporation), annual precipitation at the ground, and annual precipitation at the cloud base.

Inferences about the nonrandomness in time series derived from these relationships are tested on data of annual values for four large samples of river gauging and precipitation gauging stations. Sampling is made on both a global scale (flow) and a continental scale (flow and precipitation). The effect of nonhomogeneity in data on the nonrandomness of time series is also tested.

Conclusions are that (1) the water carryover in river basins from year to year, (2) the evaporation from river basin, (3) the evaporation of rainfall in the air between the ground and the cloud base, and (4) the inconsistency and nonhomogeneity in data are four essential causes of nonrandomness in the series of annual flow. The last three factors are mainly responsible for nonrandomness in series of annual effective precipitation, and the last two factors for the nonrandomness in series of annual precipitation.

There is a small margin of nonrandomness left that could be explained by causes from the upper atmosphere, oceans, and/or solar and cosmic activities.

RÉSUMÉ

Les séries temporelles des débits et des précipitations annuels ne sont pas rigoureusement aléatoires. Des relations simplifiées sont déduites entre les valeurs annuelles du débit, des précipitations effectives (précipitation moins évaporation), des précipitations au sol, et de celles à la base des nuages.

Les conclusions relatives aux écarts des séries temporelles par rapport à des séquences purement aléatoires sont confrontées avec quatre échantillons d'observations de stations limnimétriques et pluviométriques. L'échantillonnage est fait sur une échelle à la fois globale (débits) et continentale (débits et précipitations).

L'effet d'inhomogénéité dans les observations sur les écarts des séries temporelles par rapport à des séquences purement aléatoires est également étudié.

Les conclusions sont que (1) les reports d'eau d'une année à l'autre dans les bassins versants, (2) l'évaporation dans ces bassins, (3) l'évaporation des pluies au cours de leur chute dans l'air et (4) l'inconsistance et l'inhomogénéité dans les observations sont les causes essentielles des écarts entre les séries temporelles des débits annuels et des séquences purement aléatoires. Ces écarts pour les précipitations annuelles effectives sont surtout dûs aux trois derniers facteurs et ceux pour les précipitations annuelles au sol aux deux derniers facteurs. Il n'existe qu'un faible espoir de trouver une relation significative entre ces écarts et des causes liées à la haute atmosphère, les océans et/ou des activités solaires et cosmiques.

(*) Research activities from which this paper results are sponsored by U.S. National Science Foundation, and partly by U.S. Office of Naval Research.
(**) Professor-in-Charge of Hydrology Program, Civil Engineering Department, Colorado State University, Fort Collins, Colorado, U.S.A.

1. Introduction

1.1. *Objectives and definitions*

The subject of the paper is the study of fluctuations of wet and dry years through analysis of runoff and precipitation. Dependence in their time series is studied.

A random series is defined as a series which has no link among its members. Series with any type of link among members are nonrandom in sequence.

1.2. *Approach to investigation*

Carryover of water stored in a river basin from year to year, evaporation and evapotranspiration from river basin area, and evaporation of rainfall in the air are considered here as some of the most essential physical causes of nonrandomness in the time series of annual flow and annual precipitation. Inconsistency and nonhomogeneity in data are also considered as important causes of nonrandomnes.

A random series of a normal standard variable is used here as a bench-mark variable and series. In the text it will be referred to as random normal variable. Properties of series of annual flow and annual precipitation are compared with the bench-mark variable and series, and departures in this comparison are explained by factors which are known to be the causes of nonrandomness.

Mathematical models of stochastic processes are derived in simplified forms, in order to show the potential causes of nonrandomness in series. Hypotheses underlying the models are based on physical relationships among hydrologic variables.

Tests of nonrandomness in sequence of annual flow and annual precipitation were made more reliable by using large samples of data. By sampling the stations in wide regions, on global and continental scales, the usual limitations in time length of observations were to some extent eliminated.

1.3. *Statistical methods of analysis*

Techniques used in the study of series of annual flow and annual precipitation as given here were serial correlation analysis and analysis by range. Statistical inference of the results has been made when the appropriate technique has been available.

1.4. *Samples used in research*

Four large samples of data were used as the research material.

The first sample consists of data of 140 river gauging stations with a total of 7667 annual flows, and average length of observation per station of 55 years, as sampling on the global scale (U.S.A. 72, Canada 13, Europe 37, Australia and New Zealand 11, and Asia and Africa 7). The second large sample consists of data of 446 river gauging stations with a total of 16509 annual flows, and average length of observation per station of 37 years, as sampling on the continental scale of Northwestern America (Western U.S.A. 431, Western Canada 14). The third large sample consists of 1140 precipitation stations considered as being with relatively homogeneous data, with a total of 61600 annual values and average length of observation per station of 54 years, as sampling on the continental scale of Northwestern America (Western U.S.A. 1059, Western Canada 81), covering the same region as the second sample. The fourth large sample consists of 472 precipitation stations considered as being of nonhomogeneous data, with a total of 27133 annual values, and an average length of observation per station of 57 years (all stations from Western U.S.A.). This last sample was used exclusively to test the effect of nonhomogeneity in data on the characteristics of time series. For detailed description of samples see reference (9).

2. Mathematical models

2.1. *Variables*

Variables used in the mathematical models, all expressed as annual values for water years and for a river basin, are : V, flow; P_e, effective precipitation (net water yield of atmosphere to a river basin, or precipitation minus evaporation for a year); P_i, precipitation at the ground level; E, evaporation and evapotranspiration from river basin surface; P_c, precipitation at the cloud base; E_a, evaporation of rainfall between the cloud base and the ground; W_e and W_b, water stored in river basin at the end and at the beginning of a water year, respectively; ΔW, change in the total volume of water stored in river basin for a year; e, random error in any variable; i, inconsistency in any variable (trends and jumps in time series caused by the systematic errors); h, nonhomogeneity in any variable (trends and jumps in time series caused by natural and manmade causes which alter the virgin values of flow and precipitation); g, errors in ΔW. For detailed description of mathematical models see reference (9).

2.2. *General mathematical model*

The general mathematical model used for investigation of dependence in series of annual flow, annual effective precipitation, and annual precipitation is

$$P_e = P_c - (E_a + E) = P_i - E = V + W_e - W_b = V + \Delta W \qquad (1)$$

with $P_i = P_c - E_a$, and $\Delta W = W_e - W_b$.

Seven variables P_c, E_a, P_i, E, P_e, ΔW and V are characterized by their probability distributions and sequential patterns. These properties are related, because $P_c = P_i + E_a$; $P_i = P_e + E$; and $P_e = V + \Delta W$. The properties of E_a, E, and ΔW determine the relationship among the characteristics of P_c, P_i, P_e and V.

Taking the errors and nonhomogeneity in V and ΔW into account, the true value P_{te} of annual effective precipitation is

$$P_{te} = V \pm e_v \pm i_v \pm h_v + \Delta W \pm g_w \qquad (2)$$

in which the random errors e_v decrease the nonrandomness of a series, while the inconsistency i_v and the nonhomogeneity h_v in V, and nonrandom errors g_w in ΔW, in the form of trends and jumps, increase on the average the nonrandomness in the P_e-series. In general, the effect of random errors in the computed annual flows is smaller than the effect of inconsistency and nonhomogeneity in data.

2.3. *Relationship of annual flow to annual effective precipitation*

The change ΔW in water carryover may be expressed as a simplified Markov linear model, figure 1, because W_b and W_e may be expressed as the linear functions of annual effective precipitation of previous years

$$\Delta W_n = P_n - \sum_{j=0}^{j=\infty} b_j P_{n-j} \qquad (3)$$

with ΔW_n, change in carryover for nth year; P_n, annual effective precipitation or P_e-value for the nth year; P_{n-j}, annual effective precipitation for the year which precedes the nth year by j years; and b_j, coefficients which represent the proportions of P_n flowing out in successive years.

Equation (3) is based on the properties of outflow of water stored in a river basin. Assuming an average recession curve of river flows at the end of water years, and the average distribution of rainfall and evaporation over river basin and within the year, then the b_j coefficients have these six properties : (1) their sum is unity; (2) they are all positive; (3) they decrease monotonically; (4) theoretically there is an infinite number of them, but practically they are insignificant after b_m, and m number of years, which depends on the available storage spaces, and inflow and outflow conditions for storage spaces; (5) they are assumed constant for given river basin and given distribution of rainfall and evaporation within the year and over the river basin, but usually the value b_0 changes from year to year; and (6) they do not depend on P_e-values, but each individual b_j depends on the time elapsed since the occurrence of a given annual effective precipitation.

From eqs. (1) and (3) comes

$$V_n = \sum_{j=0}^{j=\infty} b_j P_{n-j} \pm e \tag{4}$$

with e, variable which takes care of random errors, of the difference in distributions of precipitation and evaporation within the year and over the river basin, and the use of average b_j values in eq. (4). Equation (4) and properties of b_j coefficients show that any significant and changing water carryover from year to year makes the annual flow V_n nonrandom in sequence when the annual effective precipitation P_{n-j} is random in sequence, or the degree of nonrandomness in V-series is increased if P_e-series is also nonrandom.

2.4. *Relationship of annual effective precipitation to annual precipitation*

A simple linear equation for annual evaporation is assumed here as

$$E = aP_i + bW_b \pm f \tag{5}$$

with P_i, annual precipitation; W_b, annual storage of water and moisture inside the river basin at the beginning of a water year; a and b, coefficients which depend on river basin and climatic factors; and f, a variable embracing the random errors and the effects of neglected factors. W_b is a function of precipitation, evaporation and runoff conditions of previous years.

As $P_e = P_i - E$, and by using eq. (5) and a Markov linear model for W_b, as a function of annual effective precipitation of previous years, figure 1, and applying a recurrence procedure to express P_e as function of P_i, then

$$P_n = \sum_{j=0}^{j=\infty} k_j P_{p-j} \pm d \tag{6}$$

with P_n the P_e-value for the nth year; P_{p-j}, annual precipitation at the ground for the jth year previous to the nth year; k_j, coefficients (their sum is not unity, they are not all positive); and d, a variable taking care of the simplified assumptions and neglected factors and errors.

Equation (6) with the Markov linear model shows that the dependence of annual evaporation from a river basin on the moisture history of previous years introduces nonrandomness in the sequence of annual effective precipitation, when the annual precipitation is random in sequence. However, the effect of evaporation is not simple as the effect of changes in water carryover, because a and b coefficients in eq. (5) are functions of many variables, and are related.

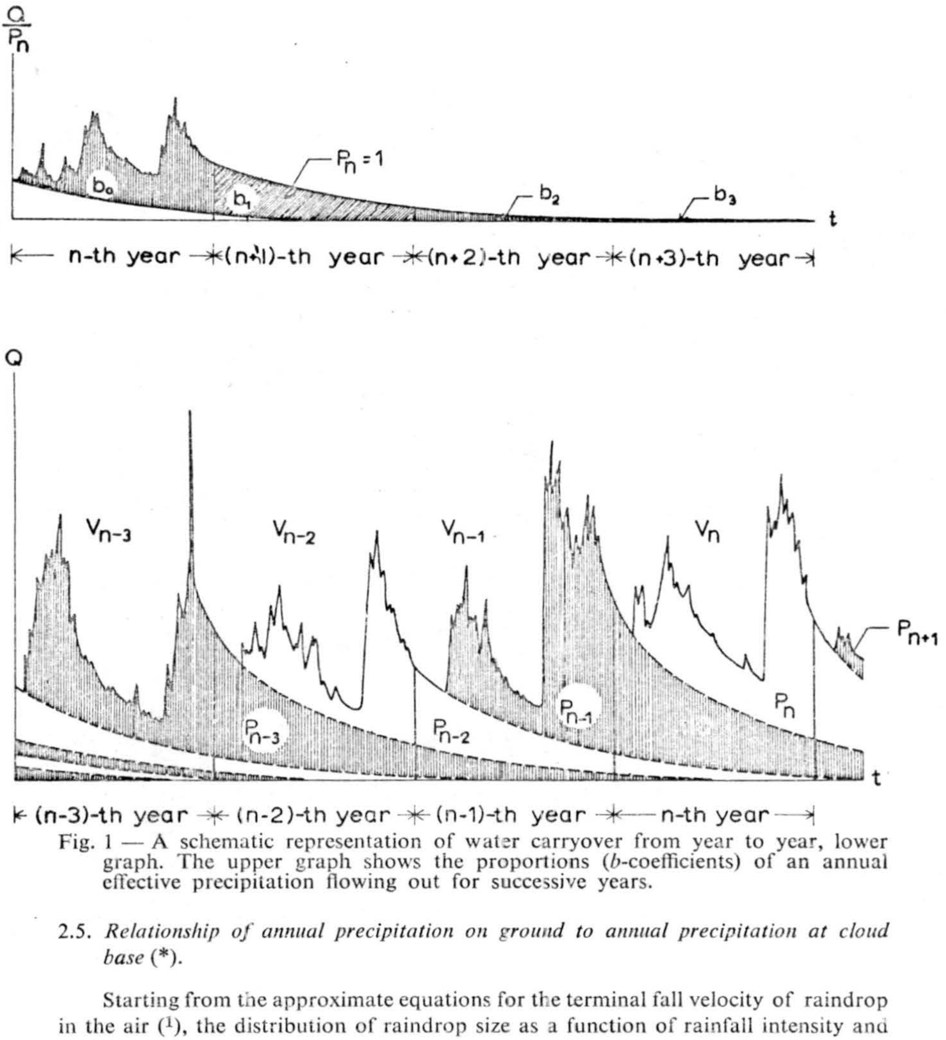

Fig. 1 — A schematic representation of water carryover from year to year, lower graph. The upper graph shows the proportions (*b*-coefficients) of an annual effective precipitation flowing out for successive years.

2.5. *Relationship of annual precipitation on ground to annual precipitation at cloud base* (*).

Starting from the approximate equations for the terminal fall velocity of raindrop in the air [1], the distribution of raindrop size as a function of rainfall intensity and elevation [2], the standard atmosphere, the equation for total transfer of heat from the air to an evaporating and ventilated spherical drop, and the rate at which the total mass of water vapor and the latent heat are transferred from a raindrop to the ventilated air [3], the evaporation of rainfall in a column of air of unit cross section area between the cloud base (elevation Z_b) and the ground (elevation Z_g) for the duration of rainfall t_0 is given approximately by the equation

$$E_a = a_0 \alpha g \int_{t=o}^{t_o} \int_{Z=Zg}^{Z_b} \int_{r=o}^{\infty} \frac{e(T - T_s) r \rho(r)(1 + 0.246\sqrt{R_e})}{e - b_o} dt\, dZ\, dr \qquad (7)$$

(*) This study of the evaporation of rainfall in the air and of its effects on the nonrandomness of annual precipitation at the ground was initiated by the author while working as the guest with the U.S. National Center for Atmospheric Research, Boulder, Colorado, in July 1962.

where $a_0 = 4\pi k/L = 1.21 \times 10^{-6}$ (k, thermal conductivity of air $= 5.66 \times 10^{-5}$, cal/cm sec °C; L, latent heat of evaporization $= 594.9 - 0.51T$, cal/gm, T in °C); α, factor greater than unity which takes care of nonsphericity of raindrops; g, gravitational constant; r, raindrop radius; e, water vapor pressure at the altitude Z; T, dry-bulb temperature and T_s, wet-bulb temperature in °C; $b_0 = kR^2K^3/\varepsilon^2L^2D^2J = 7200$ dynes/cm² (R, universal gas constant; K, Kelvin temperature, $K = 273 + T$, T in °C); ε, specific gravity of water vapor with respect to dry air; D, coefficient of diffusion of water vapor in air; J, mechanical equivalent of heat); $p(r)$, distribution function of raindrop sizes [2]; R_e, Reynolds number ($R_e = 2rV_z/\nu$; with V_z, falling velocity of the drop relative to the dry air; $\nu = 0.13$ cm²/sec, kinematic viscosity).

The values e, T, T_s, $p(r)$, r, and R_e change with time and altitude. As the evaporation E_a depends on e, T, and Z_b, and as these three variables depend to some extent on water evaporation from the ground in a region, and on ground conditions, the evaporation of rainfall in the air depends also on water stored in the river basin in different forms and places. However, the climatic factors are of a much more significant influence than the evaporation from the ground.

As $P_i = P_c - E_a$, the annual precipitation on the ground depends also on the climatic factors and in a small measure also on the moisture conditions in a river basin and around it of previous years, insomuch as the evaporation E_a depends on the moisture stored in a river basin or around it. This factor may be of a significant effect in some regions and may introduce the nonrandomness in the series of annual precipitation, when the annual precipitation at the cloud base is random in sequence.

2.6. *Test of conclusions from the above mathematical models*

From the simple mathematical models, eq. (1), (4), (6), and (7), it results that the carryover of water from year to year, in any form, the dependence of evaporation from the ground on the moisture in a river basin, stored from previous years, the evaporation of rainfall in the air, affect the nonrandomness of series of annual precipitation, annual effective precipitation, and annual flow.

The series of V, P_e, and P_i are used to test the effect of E and ΔW. Lack of sufficient data on series P_c makes a direct test of the effect of variable E_a unfeasible.

2.7. *Determination of variables V, P_e, and P_i*

The variable V is obtained directly from the river flow records. The P_e variable is obtained from the equation $P_e = V + \Delta W$. Values of ΔW are obtained by computing the water volumes W_e and W_b using the average recession curve and the flow at the end of water year as an index discharge [4] and [9]. The variable P_i is obtained from precipitation records.

3. Analysis by serial correlation

3. 1. *Serial correlation coefficient and its confidence limits*

The serial correlation coefficient r_k of lag k was computed by

$$r_k = \frac{\sum_{i=1}^{N-k} X_i X_{i+k} - \frac{1}{N-k} \sum_{i=1}^{N-k} X_i \sum_{i=1}^{N-k} X_{i+k}}{\left[(N-k-1)s_k^2\right]^{1/2} \left[\sum_{i=1}^{N-k} X_{i+k}^2 - \frac{1}{N-k}\left(\sum_{i=1}^{N-k} X_{i+k}\right)^2\right]^{1/2}} \tag{8}$$

with the variance

$$s_k^2 = \frac{1}{N-k-1}\left[\sum_{i=1}^{N-k} X_i^2 - \frac{1}{N-k}\left(\sum_{i=1}^{N-k} X_i\right)^2\right] \tag{9}$$

with N, sample size, and X_i and X_{i+k}, members of the series.

The confidence limits on 95% level for r_k are given by R.L. Anderson ([5]) for random and circular time series approximately as

$$L(95\%) = \frac{-1 \pm 1.64\sqrt{N-k-2}}{N-k-1} \tag{10}$$

For r_1 of a random time series, but taken as circular (the last member of time series is supposed to be followed by the first member), R.L. Anderson ([5]) gives the expected value

$$\bar{r}_1 = -\frac{1}{N-1} \tag{11}$$

and the standard error of r_1

$$s(r_1) = \frac{\sqrt{N-2}}{N-1} \tag{12}$$

Equations (8) through (12) were used in the analysis of serial correlation coefficients for the variables V, P_e, and P_i.

Time series in this paper are used as the open series, so it may be assumed that the expected value of r_1 for random normal variable is zero. Both expected values, that of eq. (11) and zero, will be used here.

3.2. *Serial correlation analysis of the first large sample*

The first serial correlation coefficient r_1 is computed for the data of 140 stations for both V-series and P_e-series. The average values are : $\bar{r}_1(V) = 0.176$ and $\bar{r}_1(P_e) = 0.130$. The $\bar{r}_1$ for random circular series with $N = 55$ is -0.018, and $\bar{r}_1 = 0$ for open random series.

Water carryover is the factor which accounts for 35.4% of the positive value of r_1 of V-series as compared with P_e-series. Figure 2 shows the probability distributions of r_1 for 140 stations both of V-series and P_e-series, and the random series with $N = 55$.

In the range of 20%-95%, the r_1-distributions for V- and P_e-series follow closely the slope of the r_1-distribution of random series for $N = 55$. The departures of the $r_1(V)$ and $r_1(P_e)$ distributions on the extremes from the normal distribution may be partly explained by the use of the average sample size $N = 55$ for 140 series, with a range of N from 40 to 150.

This large sample of flows for stations from several parts of the world includes many river basins with unusually great storage reservoirs (St. Lawrence, Göta, Neva, Nile, Lake Victoria, Lake Albert, etc.). Therefore, the differences in $\bar{r}_1$ and in r_1-distributions for V- and P_e-series would be smaller than given by the above results for other samples of river basins with less water carryover from year to year.

Figure 3 gives the correlograms (r_k versus the lag k) for both V-series (upper graph) and P_e-series (lower graph) for four river gauging stations from the first large sample, with 120 or more years of flow observations for each station. They are : Göta River at Sjötorp-Vänersburg, Sweden, with 150 years (1808-1957); Rhine River at

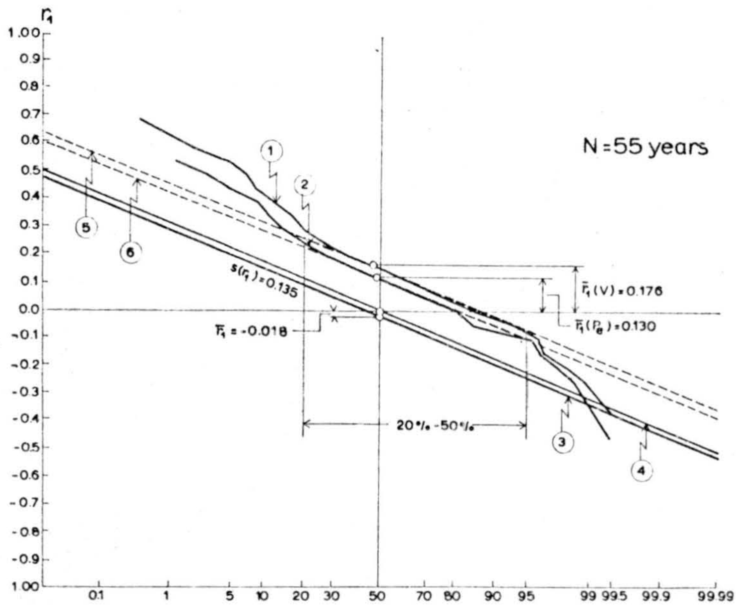

Fig. 2 — Cumulative frequency distributions of first serial correlation coefficient (r_1) for 140 river stations, using Cartesian-probability scales : (1) V-series, computed; (2) P_e-series, computed; (3) random circular series for $N = 55$; (4) random open series ($\bar{r}_1 = 0$); (5) V-series, fitting a straight line for range 20%-95%; (6) P_e-series, fitting a straight line for 20%-95% range of cumulative frequency.

Basle, Switzerland, with 150 years (1808-1957); Nemunas at Smalininkai, Lithuania, U.S.S.R., with 132 years (1811-1943); and Danube at Orshava, Romania, with 120 years (1838-1957). The $r_1(V)$-values for these four stations are 0.463, 0.076, 0.185, and 0.096, respectively. The $r_1(P_e)$-values for these four stations are 0.009, 0.015, 0.118, and −0.001, respectively. The confidence limits on 95% level are computed by using eq. (10). The r_k-values are determined up to $k = N/4$, with N the length of series.

The comparison of $\bar{r}_1$ values of V-series to those of P_e-series for these stations with long records shows that in all four cases the water carryover accounts for a large portion of the positive first serial correlation.

The correlograms lead to the conclusions, that the r_k-values, except for r_1-values, may be considered as not significant from zero on 95% level. In this test 5% of all r_k-values should be outside the confidence limits, what occurs approximately for these four correlograms.

3.3. *Serial correlation analysis of the second and third large samples*

There are two cases in these two samples, (a) the length of each of the three series (V, P_e, and P_i) is taken as 30 years (1931-1960) for all stations, and (b) the actual length of record at each station is used, with the average size of series $N_1 = 37$ years for V and P_e variables, and $N_2 = 54$ years for P_i variable.

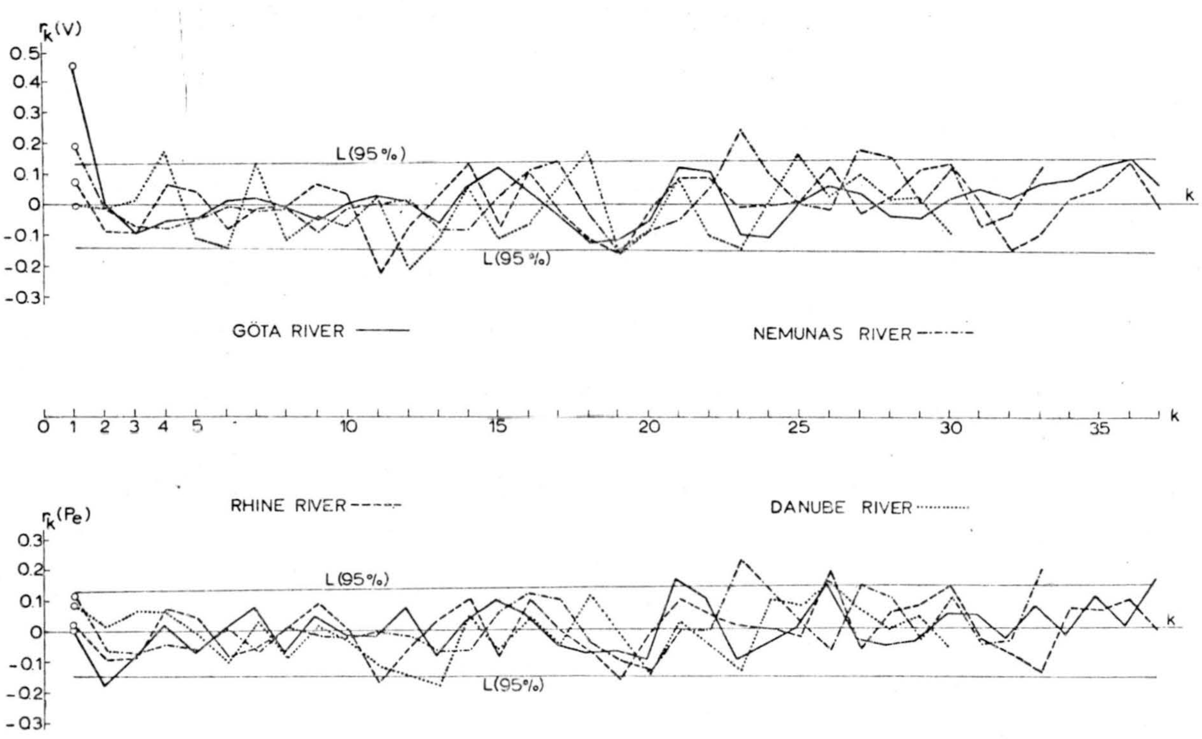

Fig. 3 — Correlograms of series of annual flows (upper graph) and annual effective precipitation (lower graph) for the rivers with longest observations : Göta (150), Rhine (150), Nemunas (132), and Danube (120).

The mean values for r_1 are:

(a) Length of series 30 years:

V-series, mean of 446 stations, $\bar{r}_1(V) = 0.163$.
P_e-series, mean of 446 stations $\bar{r}_1(P_e) = 0.146$.
P_i-series, mean of 1140 stations $\bar{r}_1(P_i) = 0.028$.

(b) Average length of V and P_e series is 37 years, and of P_i series 54 years:

V-series, mean of 446 stations $\bar{r}_1(V) = 0.197$.
P_e-series, mean of 446 stations $\bar{r}_1(P_e) = 0.181$.
P_i-series, mean of 1140 stations $\bar{r}_1(P_i) = 0.055$.

The $\bar{r}_1$-values for a random and circular series of length 30, 37, and 54, eq. (11), are -0.0345, -0.0278, and -0.019, respectively, and zero for random open series. The standard errors of r_1 for these three cases, eq. (12), are 0.182, 0.164, and 0.136, respectively.

This comparison of the mean values of r_1 for the three series shows that the change ΔW in water carryover accounts for about 10.5% or 8.2% of the positive correlation of r_1 in V-series, while the evaporation E accounts for about 80.9% or 69.7% of the positive correlation of r_1 in P_e-series, respectively for the periods of 30 years and for the longest available length of observations.

The $\bar{r}_1$ values for three series V, P_e, and P_i increase with an increase of the average length of observations. The mean value is increased from 0.163 to 0.197 (or 20.5%) for V-series by an increase of average N from 30 to 37; from 0.146 to 0.181 (or 24%) for P_e-series for the same change of average N; and from 0.028 to 0.055 (or 100%) for P_i-series for an increase of average N from 30 to 54. This change may be explained by the fact that the period 1931-1960 was simultaneous for all stations, and as such may have a smaller average value of r_1 than the average for 30-year periods.The increase of $\bar{r}_1$ with an increase of average N may be partly explained also by the greater occurrence and influence of inconsistency and nonhomogeneity in data in a longer period of observation, than in a shorter one.

The example of correlograms of four stations with longest records, figure 3, does not support a hypothesis for $\bar{r}_1$ to increase constantly with an increase of the average length of series.

Figure 4 gives the three probability distributions of r_1 for V, P_e and P_i, for 30 years' length of series. Figure 5 gives the same information for the longest available length of observations. The probability distributions of random series (both circular and open) for each length of series are also given for comparative purposes.

Figure 4 shows that the probability distributions of $r_1(V)$ and $r_1(P_e)$ follow closely the slope of the probability distribution of random normal variable with $N = 30$ and $s(r_1) = 0.182$ in the range of 5%-95% of probability. This conclusion is also applicable for the distribution of $r_1(P_i)$, which is normally distributed with $\bar{r}_1(P_i) = 0.028$, and standard error $s = 0.168$ for all the range of probability, except that the slope with $s = 0.168$ does not correspond closely to the slope with $s(r_1) = 0.182$ of the random normal variable with $N = 30$. The difference is not substantial.

Figure 5 shows that the probability distributions of $r_1(V)$ and $r_1(P_e)$ follow closely the slope of the distribution of random normal variable with $N = 37$ and $s(r_1) = 0.164$ in the range of 2.5%-97.5%. The same conclusions are valid for the distribution of $r_1(P_i)$, with $s(r_1) = 0.136$ for $N = 54$.

This comparison of observed series with the random normal variable leads to the conclusion that a portion of the nonrandomness in the series of annual flow is explained by water carryover from year to year in river basins. However, the major portion is explained by the evaporation from the river basin surface. Though the

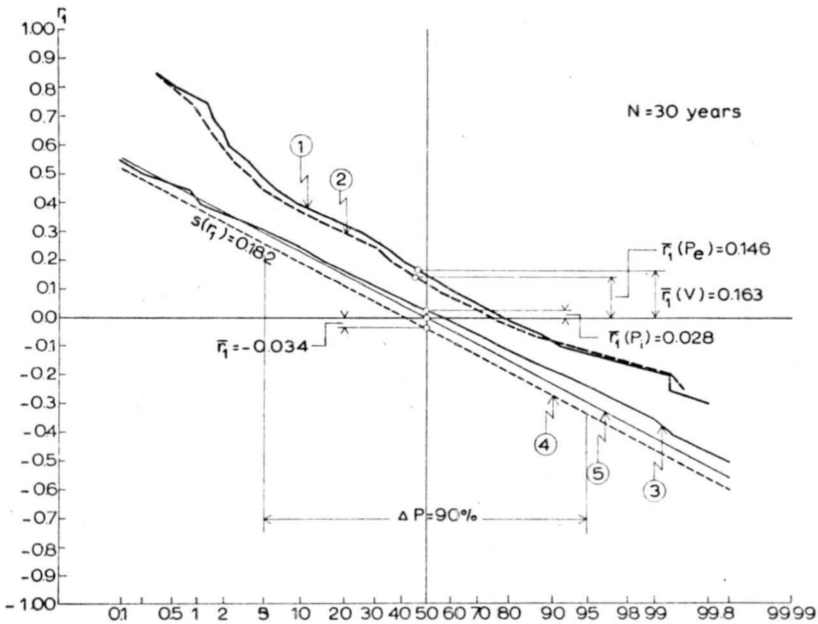

Fig. 4 — Cumulative frequency distributions of first serial correlation coefficient (r_1). for series of 446 river stations (V- and P_e-series) and for series of 1140 precipitation stations (Pi-series), with the length of these three series of 30 years (1931-1960) in Carthesian-probability scales: (1) V-series (annual flow); (2) P_e-series (annual effective precipitation); (3) Pi-series (annual precipitation); (4) random circular series for $N = 30$; and (5) random open series ($\bar{r}_1 = 0$) for $N = 30$.

precipitation stations do not coincide with the river basins, the fact that there are 1140 precipitation stations in the same area as 446 river gauging stations makes the inference of the essential effect of evaporation reliable.

It is expected that one part of the nonrandomness measured by the first serial correlation coefficient of precipitation on the ground may be explained by the evaporation of raindrops in the air, and by the inconsistency and nonhomogeneity in data. The data is not available for a direct test of this first hypothesis, but it will be tested indirectly. The second hypothesis will also be briefly tested.

3.4. *Example of correlograms for river gauging station with large carryover*

As an example of the effect of water carryover, the river basin with the largest ΔW-values, the correlograms of the St. Lawrence River at Ogdensburg, N.Y., both for V-series and P_e-series are given in figure 6, together with the confidence limits on 95% level. They show clearly that the carryover is mainly responsible for the nonrandomness of series of annual flow. This example is an upper limit case, but it proves that the simplified mathematical model of eq. (4) in the form of Markov linear equation is an attractive approximation for the patterns in sequence of V-series as it concerns carryover. For river basins with small carryovers, the simple relationship of eq. (4) is not as evident as in the case of the St. Lawrence River.

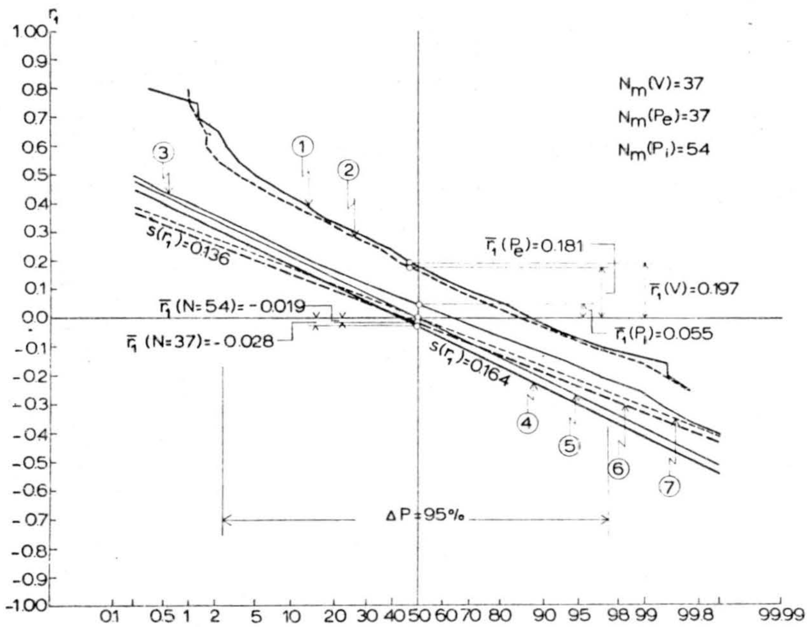

Fig. 5 — Cumulative frequency distributions of first serial correlation coefficient (r_1) for series of 446 stations (V- and P_e-series), with the average length of series $N = 37$, and for series of 1140 precipitation stations ($P\bar{i}$-series), with the average length of series $N = 54$, in Carthesian-probability scales : (1) V-series; (2) P_e-series; (3) $P\bar{i}$-series; (4) random circular series for $N = 37$; (5) random open series ($\bar{r}_1 = 0$) for $N = 37$; (6) random circular series for $N = 54$; (7) random open series ($\bar{r}_1 = 0$) for $N = 54$.

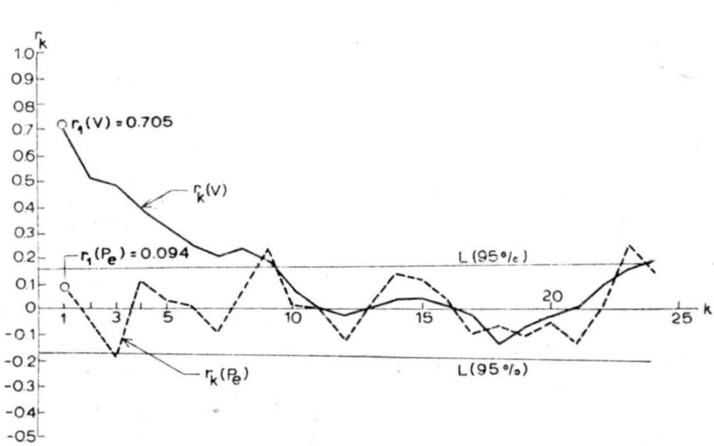

Fig. 6 — Correlograms for series of annual flow (V-series) and annual effective precipitation (P_e-series) for St. Lawrence River at Ogdensburg, N.Y., U.S.A., for 97 years of observation.

Most carryovers of moisture in a river basin for both the first and the second sample are either insignificant or are significant only for one year ($m = 1$). This indicates that a model $V_n = b_0 P_n + b_1 P_{n-1}$ may be the simplest fit for the nonrandomness model of V-series as related to the carryover.

3.5. *Indirect test of the effect of evaporation in the air on nonrandomness of annual precipitation at the ground*

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

The third large sample of 1140 precipitation stations with relatively homogeneous data is divided in three subsamples : (a) stations with the average annual precipitation greater than 500 mm, (b) stations with the average precipitation between 250 and 500 mm; and (c) stations with the average precipitation below 250 mm. The three subsamples have 582, 463, and 95 stations, respectively.

The values $\bar{r}_1(P_i)$ for the three subsamples and for $N = 30$ are 0.040, 0.024, and −0.027, respectively. The values $\bar{r}_1(P_1)$ for the three subsamples and maximum available length of records are 0.063, 0.051 and 0.024, respectively.

The differences show that the arid and semi-arid regions, with small precipitation at the ground, have on the average a smaller first serial correlation of annual precipitation at the ground than the semi-humid or humid regions. They indicate indirectly that there is an influence of evaporation of rainfall in the air on the nonrandomness of annual precipitation at the ground.

3.6. *Effect of inconsistency and nonhomogeneity in data*

Figures 7 and 8 give the probability distributions of the r_1 coefficient for the series of annual precipitation of 1140 precipitation stations with homogeneous data, and for 472 precipitation stations with nonhomogeneous data, for both the period of 30 years and the longest observation periods.

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

The values $\bar{r}_1$ for homogeneous and nonhomogeneous data of series 30 years long are 0.028 and 0.053, respectively, and for data of series of maximum available length $\bar{r}_1$ are 0.055 and 0.071, respectively. These values and the r_1-distributions show that the nonhomogeneity in data is not a negligible factor in creating the nonrandomness in time series of rainfall and runoff.

Though there may be a disagreement about the classification of precipitation data into homogeneous and nonhomogeneous, the large number of stations in the two classes tends to minimize error due to this cause, and validate the conclusion that nonhomogeneity in data increases on the average the nonrandomness of time series.

These two samples of annual precipitation may be supported by theoretical analysis. Whenever a trend or a jump, or the combination of the two are introduced into a random time series, the average result is that the series becomes nonrandom in sequence.

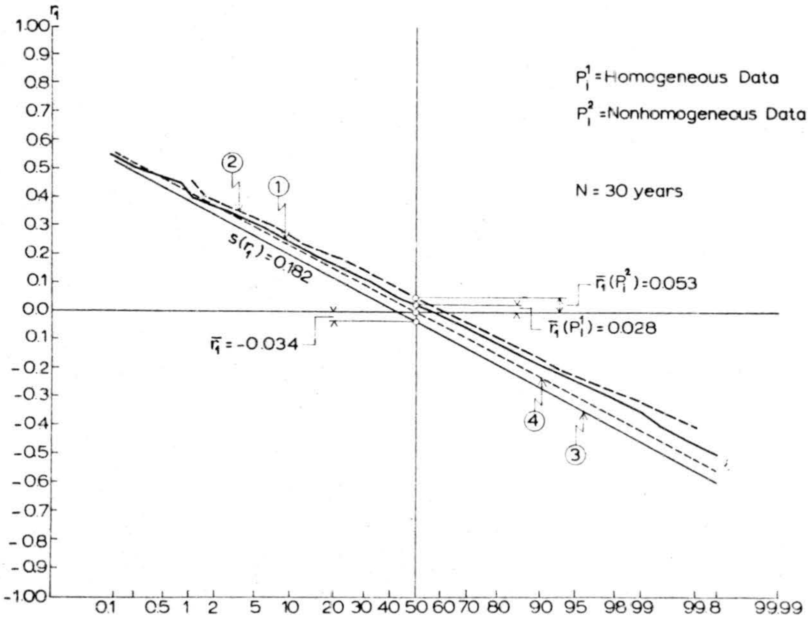

Fig. 7 — Cumulative frequency distributions of first serial correlation coefficient (r_1) for series of 1140 precipitation stations with homogeneous data (P^1i-series), and for series of 472 precipitation stations with nonhomogeneous data (P^2i-series) with the average length of both groups of series $N = 30$ years (1931-1960), in Carthesian-probability scales: (1) P^1i-series; (2) P^2i-series; (3) random circular series for $N = 30$; and (4) random open series ($\bar{r}_1 = 0$) for $N = 30$.

4. Analysis by range

4.1. *Definition of range*

The maximum range for a discrete time series and for n time-unit intervals between its successive members is defined here as the maximum difference $R_n = S^+ - S^-$ of the accumulated sum of departures from the mean value, with R_n the maximum range, S^+ the maximum positive and S^- maximum negative value of the accumulated sum of departures for these n time-unit intervals. According to H. E. Hurst (6) the maximum range can be conceived as the maximum accumulated storage, when there is never a deficit in outflow (which is equal here 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.

4.2. *Distribution of maximum range for a random time series*

The asymptotic values as for expected mean of the maximum range of random normal variable for a large value of n is given by W. Feller (7) as

$$R_n = 1.6\sqrt{n} \tag{13}$$

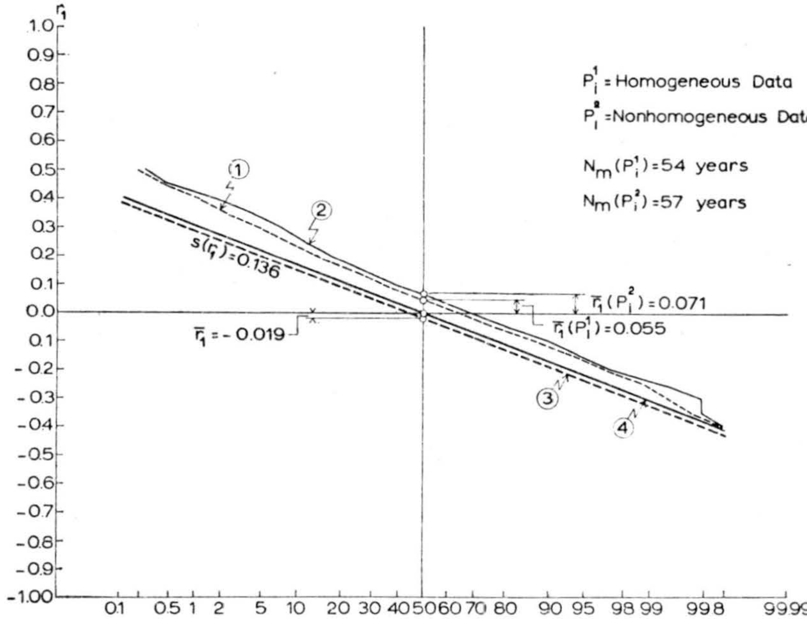

Fig. 8 — Cumulative frequency distributions of first serial correlation coefficient (r_1) for series of 1140 precipitation station of homogeneous data (P^1i-series), with average length $N = 54$, and for series of 472 precipitation stations of nonhomogeneous data (P^2i-series), with average length $N = 57$, in Carthesian-probability scales: (1) P^1i-series; (2) P^2i-series; (3) random circular series for $N = 54$; (4) random open series ($\bar{r}_1 = 0$) for $N = 54$.

The exact distribution of range $\bar{R}_n$ is difficult to obtain even for the small values of $n = 4$ or 5 [7].

The expression for the expected values R_n are given by A. A. Anis and E. H. Lloyd [8] as

$$\bar{R}_n = \sqrt{\frac{2}{\pi}} \sum_{i=1}^{n} i^{-1/2} \tag{14}$$

4.3. *Comparison of mean range of annual effective precipitation with that of random series*

The comparison is made here for a subsample containing 85 river gauging stations (U. S. A. 72, Canada 13) from the first large sample of 140 stations. The comparison is made particularly for the means of all maximum ranges of standardized variables of annual effective precipitation (P_e-series) for 20 values of n(1-20).

If a time series had the length $N = 60$, then there were 60 values of range for $n = 1$, 30 values for $n = 2$, 20 values for $n = 3$, and so on, until there are only 3 values for $n = 20$. The P_e variable is first standardized as $x_i = (X_i - \bar{X})/s$, and the sum S_i of x_i are computed for all values of i from 0 to N. For a period of the length n (n, number of successive years), without overlapping with the previous or the next period of length n, the extreme values of the sum are determined, with $R_n(P_e) =$

$S^+ - S^-$ for the series of each station, and for $n = 1 - 20$. All R_n-values of all stations for a given n are averaged. In this manner, for 85 series there were 4321 values of R_1, 2137 for R_2, 1417 for R_3, and so decreasing, with only 183 values for R_{20}. The reliability of the computed $\bar{R}_n$ decreases as n increases.

Figure 9 shows three lines, the asymptotic means for very large n of maximum range computed by eq. (13), the expected values, computed by eq. (14), and the average values $\bar{R}_n(P_e)$ for 85 stations, as functions of n. The number of values R_n used in computing $\bar{R}_n(P_e)$ is also given as function of n in figure 9.

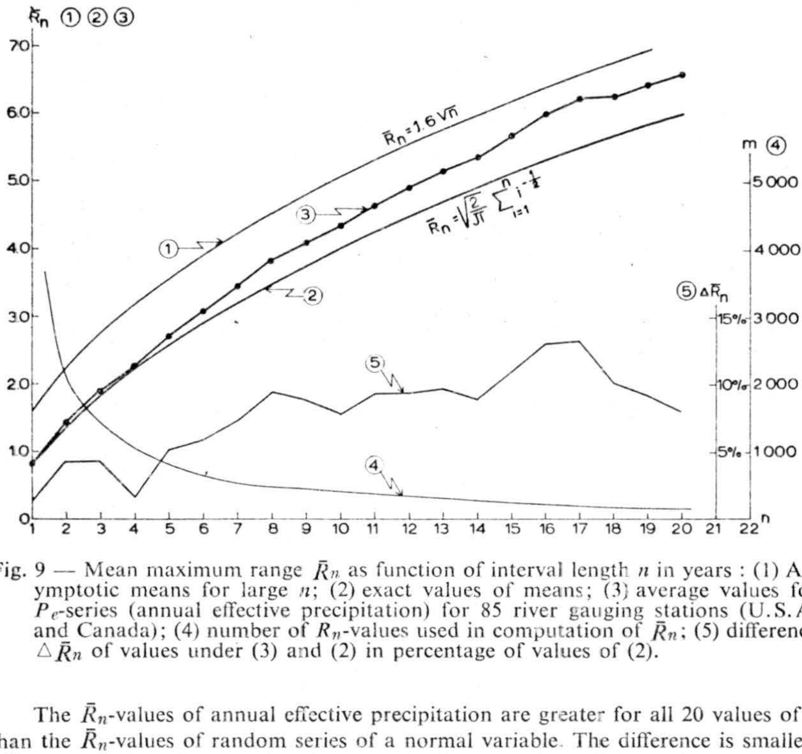

Fig. 9 — Mean maximum range $\bar{R}_n$ as function of interval length n in years : (1) Asymptotic means for large n; (2) exact values of means; (3) average values for P_e-series (annual effective precipitation) for 85 river gauging stations (U.S.A. and Canada); (4) number of R_n-values used in computation of $\bar{R}_n$; (5) difference $\Delta\bar{R}_n$ of values under (3) and (2) in percentage of values of (2).

The $\bar{R}_n$-values of annual effective precipitation are greater for all 20 values of n than the $\bar{R}_n$-values of random series of a normal variable. The difference is smallest for $n = 1$, 2, 3, and 4 and then increases for large n-values. This result corresponds to the standardized variable with a small positive nonrandomness in the series.

The departures of computed $\bar{R}_n(P_e)$ from the exact values of a random normal variable may be explained by the following three factors : (a) Small nonrandomness in the sequence of annual effective precipitation (r_1 value is significantly different from zero); (b) Average P_e-distribution is skewed (average coefficient of skewness C_s for 85 stations is different from zero); and (c) Inconsistency and nonhomogeneity in data (which on the average increase the nonrandomness).

The average value of r_1 of P_e-series for the subsample of 85 stations is $\bar{r}_1(P_e) = 0.143$.

The average value of C_s for the P_e-series of 85 stations is $\bar{C}_s(P_e) = 0.422$. This is significantly different from zero.

Inconsistency and nonhomogeneity in the data must exist, because there were many changes in the past both through river basin developments, and in the methods of sensing, recording and processing the data.

Regardless of the effect of these factors, which may have opposite influences on the maximum range, the difference of the computed and exact mean ranges is small for n small and large for n large. The relative difference of $\bar{R}_n(P_e)$ from the expected value of R_n, expressed as a percentage of $\bar{R}_n$, $d = 100[R_n(P_e) - \bar{R}_n]/\bar{R}_n$ is given for each n value in figure 9. These differences in mean range of P_e-series show that the annual effective precipitation is not random in sequence, but the degree of nonrandomness is small as measured by the first serial correlation coefficient.

4.4. *Distribution of ranges*

The exact distributions of R_1, R_2, and R_3 for random normal variable can be obtained analytically, and numerically integrated. The greater n the more complex becomes the equation of exact distribution of R_n, and the more difficult becomes the numerical integration of this equation.

The distribution of $R_n(P_e)$ for n from 1 to 3, and the numerically integrated exact distributions of R_n for random series of normal variable are compared in fig. 10.

The comparison of the $R_n(P_e)$ and R_n-distributions also shows that there is a small difference between the two series. In other words, for many practical purposes the effective annual precipitation may be considered as random in sequence for small values of n, but this conclusion is not valid for large values of n.

4.5. *Example of the change of mean range with n for large water carryover in river basins*

Figure 11 shows the mean values of range, $\bar{R}_n(V)$ and $\bar{R}_n(P_e)$, as compared with the asymptotic mean values for large n and the exact values, computed by eq. (13) and (14), respectively, for St. Lawrence River at Ogdensburg, N.Y. It shows that the line of $R_n(P_e)$ is much closer to the values R_n of random series than the line of $\bar{R}_n(V)$. In this case P_e and V refer to the standardized variables. The V-series is essentially nonrandom in sequence because its average values of R_n are much greater than the values $\bar{R}_n$ of the random normal variable. This difference increases with an in-increase of n. However, P_e-series is very close to a random series because there is no essential difference between their average values of R_n even for large n. This difference between V- and P_e-series is not so evident for most of the river basins with relatively small water carryover.

5. Conclusions

From the above mathematical models, samples, tests and comparisons, the following conclusions may be drawn :

(1) Nonrandomness in sequence of time series of annual values of flow, effective precipitation, and precipitation at the ground decreases from flows to effective precipitation, and from effective precipitation to precipitation.

(2) Nonrandomness in series of annual flows is produced largely by the four factors : water carryover from year to year, evaporation and evapotranspiration from river basin surface, evaporation of rainfall in the air, and inconsistency and nonhomogeneity in data.

(3) Nonrandomness left in series of annual values after the above factors are accounted for is relatively small.

(4) Before attempts are made to correlate with and attribute the nonrandomness in series of annual flow, annual effective precipitation, and annual precipitation to the factors of the upper atmosphere, oceans, and solar and cosmic activities, the known causes of nonrandomness as water carryover, evaporation, inconsistency and nonhomogeneity in data, should be first accounted for.

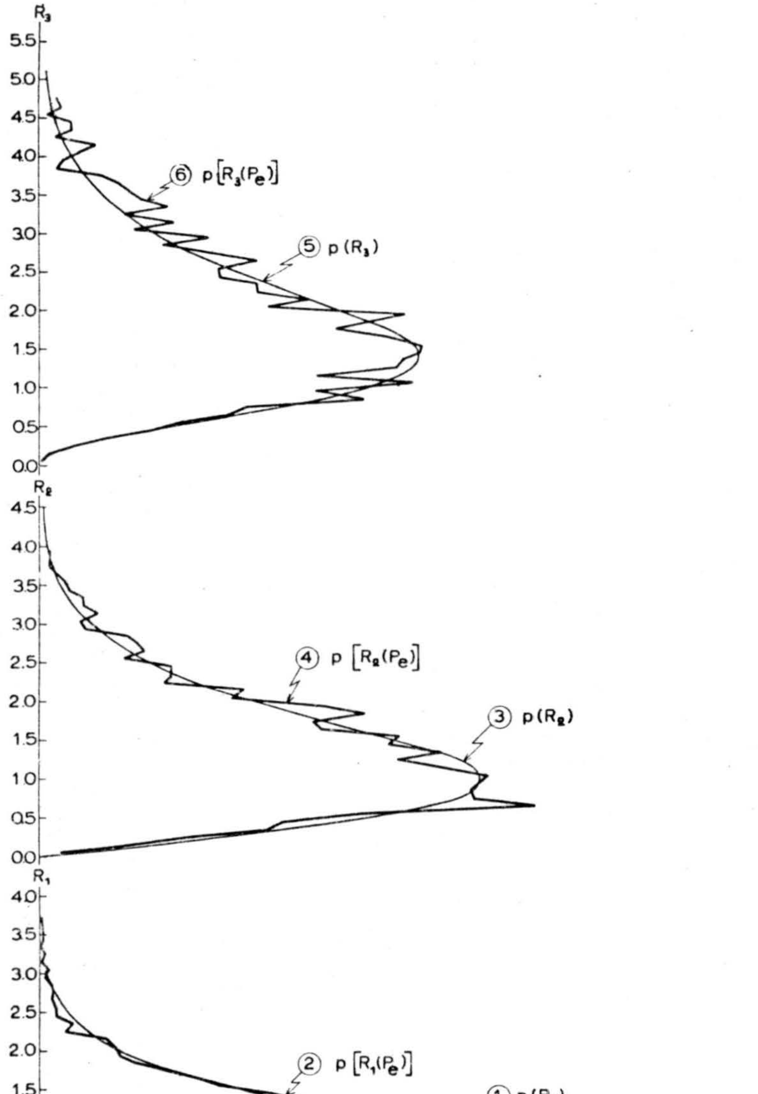

Fig. 10 — Comparison of probability density distributions for random series of the maximum range R_n for R_1, R_2, and R_3, respectively, and the maximum range R_n for the series of annual effective precipitation for 85 river gauging stations (U.S.A. and Canada), (2), (4), and (6) for $R_1(P_e)$, $R_2(P_e)$ and $R_3(P_e)$, respectively.

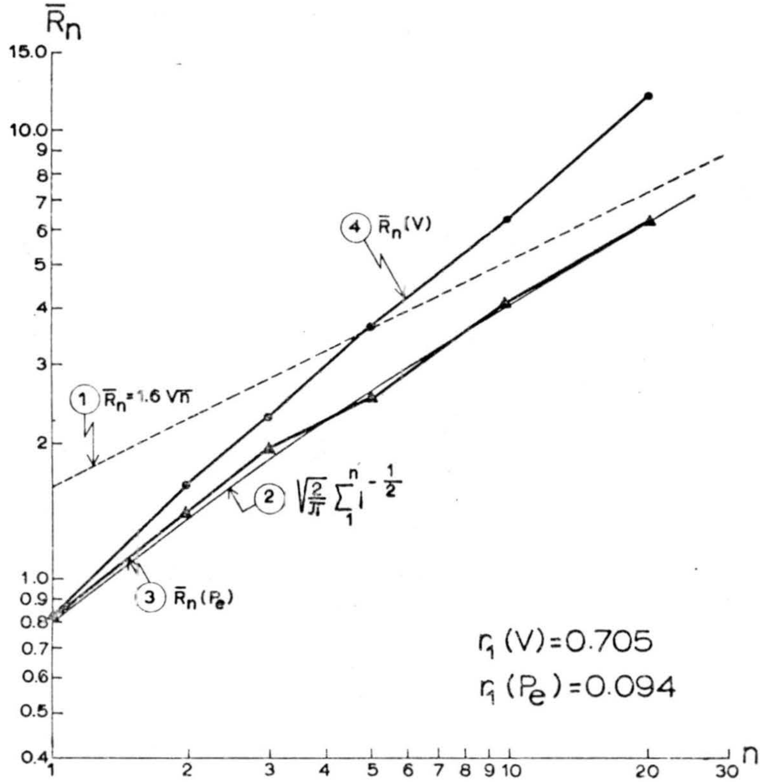

Fig. 11 — Mean maximum ranges $\bar{R}_n$ as function of interval length n (in years) for St. Lawrence River at Ogdensburg, N.Y., U.S.A., for 97 years of observation, as compared with the theoretical values. (1) Asymptotic mean range R_n for large n; (2) Exact values $\bar{R}_n$ for random series of random normal variable; (3) $\bar{R}_n(P_e)$-values for series of effective precipitation of St. Lawrence River; and (4) $\bar{R}_n(V)$-values for series of annual flow of St. Lawrence River.

(5) Simple mathematical models describing the relationship of variables of annual values of flow, carryover, evaporation and precipitation may explain the differences in the nonrandomness of time series studied. These models are, however, complex in a detailed analysis because of the changes in distributions of precipitation and evaporation both inside the river basins and within the year from one year to another.

(6) Samples of data of stations scattered on a large area (global or continental sampling), with the average intrastation correlation small, enable a reliable inference about the amount and the causes of nonrandomness in the sequence of annual flow and annual precipitation.

(7) The nonrandomness in annual flow and annual effective precipitation is relatively small on the average, and it changes from station to station, and/or from region to region.

(8) A net of river gauging and precipitation gauging stations, with homogeneous data and consistent records, well distributed on global and continental scales, and a

thorough study of the effects of water carryover, evaporation on the ground and in the air, as well as of nonhomogeneity and inconsistency in data may throw more light in the future on the patterns in sequence of wet and dry years, than is now available.

REFERENCES

(1) BEST, A.C., 1950, Empirical formulae for the terminal velocity of water drops falling through atmosphere : *Quart. Jour. of Royal Meteorology Soc.* Vol. 76, No. 329, p. 302-311.

(2) BEST, A.C., 1950, The size distribution of raindrops : *Quart. Jour. of Royal Meteorology Soc.*, Vol. 76, No. 327, p. 16-36.

(3) SQUIRES, P., 1952, The growth of cloud drops by condensation, I, General characteristics : *Australian Jour. of Sci. Research*, series A-Physical Sciences, Vol. 5, No. 1, P. 59-86.

(4) YEVDJEVICH, V.M., 1961, Some general aspects of fluctuations of annual runoff in the Upper Colorado River Basin : *Colorado State University, Engineering Research Report*, Oct. 1961.

(5) ANDERSON, R.L., 1941, Distribution of the serial correlation coefficient : *Math. Statistics Annals*, Vol. 8, No. 1, Mar. 1941, p. 1-13.

(6) HURST, H.E., 1951, Long-term storage capacity of reservoirs : *Amer. Soc. Civil Engineers Trans.*, Vol. 116, Paper No. 2447, p. 776.

(7) FELLER, W., 1951, The asymptotic distribution of the range of series of independent random variables : *Math. Statistics Annals*, Vol. 22, Sept., p. 427.

(8) ANIS, A.A., and LLOYD, E.H., 1953, On the range of partial sums of a finite number of independent normal variaties : *Biometrica*, No. 40, p. 35-47.

(9) YEVDJEVICH, V. M., 1963, Fluctuations of wet and dry years, Part I, Research data assembly and mathematical models : Hydrology Papers, Colorado State University, No 1, July 1963, Fort Collins, Colorado, U.S.A.