Journal of Hydrology, 35 (1977) 311—324

RUN-SUMS OF ANNUAL FLOW SERIES

ZEKÂI ŞEN

Civil Engineering Faculty, Department of Hydraulics and Water Power, Istanbul Technical University, Istanbul (Turkey)

(Received August 19, 1976; revised and accepted December 10, 1976)

ABSTRACT

Şen, Z., 1977. Run-sums of annual flow series. J. Hydrol., 35: 311—324.

A general methodology which is very useful for determining various run-sum properties of a given hydrologic process, has been developed on the basis of the random sum of random variables. Application of the method to some independent and dependent processes has been given.

INTRODUCTION

The objective of this paper is to provide simple analytical solutions to various problems concerning the amount of water over a dry or wet period in water-resources system design. One of the main problems in water-resources system design is to predict the total amount of water available over an operational period of the project considered. For instance, in reservoir design, if the total amount of water is in excess of the demand, the excess water will be stored, whereas during a deficit period the extraction of water is necessary, either from the previously stored water in the reservoir or/and if possible, from alternative water resources. Thus the accumulated water excess plays an important role in the reservoir design. The total amount of water from all of the water resources considered together for a given drainage basin represents the water potential of that basin. This water potential varies from one region to another, the water potential of humid areas being greater than that of arid areas. If in a period of time, water supply is greater than the demand this period is considered a wet period; otherwise, it is a dry period.

In the design of hydroelectric plants the maximum length of a dry period as well as the total amount of water over such a period is the most frequently required quantity both in dimensioning the hydraulic structures and in fixing the capacity of turbines. In addition, some other important factors which require the prediction of total amount of water are water pollution, sedimentation, problems of erosion and sizing of pipes.

DEFINITIONS

The main purpose of a water-resources system planner is to extract the maximum relevant information out of a given observation sequence and then to incorporate this information in an objective way in the prediction of future possible occurrences of the same event. Apart from the basic statistical information such as mean, standard deviation, correlation coefficients, empirical probability distribution function (PDF), etc., there is information concerning run properties, which seem to be directly related to some important quantities necessary in a water-resources system design.

An objective definition of runs will be made on the basis of a given sequence of observations $x_1, x_2, x_3, \ldots, x_n$ and a preselected level of reference x_0, see Fig.1. The preselected reference level can correspond to various hydrological quantities. In the reservoir design it corresponds to the level of demand; when a flood control scheme is considered, it is a function of the height or the capacity of a spillway. During one period of water excess the deviation $x - x_0 > 0$ is referred to as surplus and the deviation $x - x_0 \leqslant 0$ is referred to as deficit. For a given basic time period, say, one year, there exist two simple events $[x - x_0 > 0]$ or $[x - x_0 \leqslant 0]$ which are mutually exclusive but collectively exhaustive with probabilities:

$$p = P(x - x_0 > 0) \qquad \text{and} \qquad q = 1 - p = P(x - x_0 \leqslant 0)$$

respectively, or instead with

$$p = P(x > x_0) \qquad \text{and} \qquad q = P(x \leqslant x_0)$$

Generally, runs are compound events made up from these two basic events. For instance, a group of successive surpluses preceded and succeeded by at least one deficit is referred to as positive run. Moreover, the period of positive run is called positive run-length, n_{p}, whereas the sum of surpluses over the positive run-length is referred to as positive run-sum, S, see Fig.1. Thus a compound event such as:

$$[x_{i+1} > x_0, x_{i+2} > x_0, \ldots, x_{i+n-1} > x_0]$$

describes positive run-length of length at least equal to n. In the case of stationary stochastic processes, the magnitude of any x_i does not depend on a particular time instant, i.e. the statistical behaviour of x_i is independent of time instant, and thus by ignoring i the compound event becomes:

$$[x_1 > x_0, x_2 > x_0, \ldots, x_{n-1} > x_0]$$

As a result, the probability of positive run-length to be equal to or greater than n is:

$$P(n_p \geqslant n) = P(x_1 > x_0, x_2 > x_0 \ldots x_{n-1} > x_0) \tag{1}$$

In the case of independent observations, eq.1 reduces to:

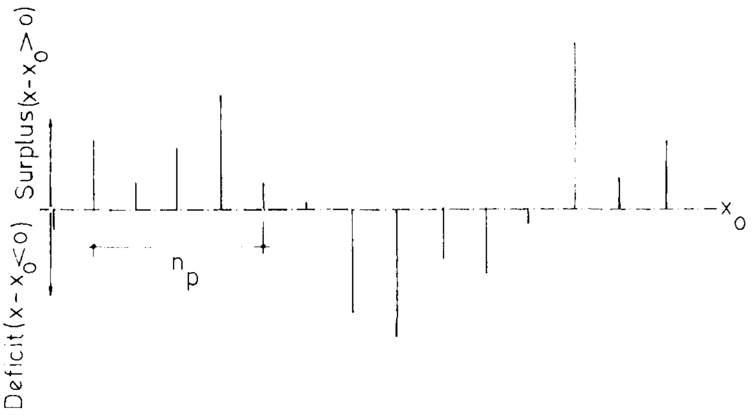

Fig. 1. Definition of surplus, deficit and runs.

$$P(n_p \geqslant n) = P(x_1 > x_0) \cdot P(x_2 > x_0) \cdot P(x_3 > x_0) \ldots P(x_{n-1} > x_0) \tag{2}$$

Or, if the observations are independent and identically distributed (IID) eq.2 becomes:

$$P(n_p \geqslant n) = p^{n-1} \tag{3}$$

On the other hand, Feller (1957) has proven that:

$$P(n_p = n) = P(n_p \geqslant n) - P(n_p \geqslant n + 1) \tag{4}$$

Thus, the probability of positive run-length of length n can be obtained by substituting eq.3 into eq.4, which leads to:

$$P(n_p = n) = q \cdot p^{n-1} \tag{5}$$

Eq.5 has been previously obtained by Feller (1957). If the underlying generating mechanism of the observations is of the first-order Markov-process-type, then eq.1 becomes:

$$P(n_p \geqslant n) = P(x_2 > x_0 | x_1 > x_0)\ P(x_3 > x_0 | x_2 > x_0) \ldots P(x_{n-2} > x_0 | x_{n-1} > x_0) \tag{6}$$

When the observations are identically distributed, all of the conditional probabilities become the same, and eq.6 reduces to:

$$P(n_p \geqslant n) = P^{n-1}(x_2 > x_0 | x_1 > x_0) \tag{7}$$

For the conditional probabilities the following expressions have been derived by Şen (1976):

$$P(x_i > x_0 | x_{i-1} > x_0) = r \tag{8}$$

$$P(x_i > x_0 | x_{i-1} \leqslant x_0) = \frac{p}{q}(1 - r) \tag{9}$$

$$P(x_i \leqslant x_0 | x_{i-1} \leqslant x_0) = 1 - \frac{p}{q}(1 - r) \tag{10}$$

$$P(x_i \leqslant x_0 | x_{i-1} > x_0) = 1 - r \tag{11}$$

where $P(x_i \leqslant x_0 | x_{i-1} \leqslant x_0)$ has been given in an integral form by Cramer and Leadbetter (1967) as:

$$P(x_i \leqslant x_0 | x_{i-1} \leqslant x_0) = q + \frac{1}{2\pi q} \int_0^{\rho_n} e^{-x_0^2/2(1+z)} \cdot (1 - z^2)^{-\frac{1}{2}} \, dz \tag{12}$$

where ρ is the first-order autocorrelation coefficient of the Markov process which can be given in its simplest form as:

$$x_i = \rho x_{i-1} + \epsilon_i \tag{13}$$

where ϵ_i is a normal independent variable. For $\rho = 0$, eq.8 yields $r = p$. The numerical solutions of eq.12 for various ρ and x_0 values have been given in tables 1—6 of a paper by Şen (1976). In the light of the above calculations, the same author has also shown that for the first-order Markov process, eq.7 leads to:

$$P(n_p \geqslant n) = r^{n-1} \tag{14}$$

and further, by making use of eq.4:

$$P(n_p = n) = (1 - r)r^{n-1} \tag{15}$$

On the other hand, the compound event corresponding to positive run-sum over a positive run-length is given as:

$$\{(x_1 - x_0) + (x_2 - x_0) + \ldots + (x_{n-1} - x_0) > 0\}$$

The partial sum of surpluses over a positive run-length can be calculated as:

$$S = \sum_{i=1}^{n} (x_i - x_0) \tag{16}$$

or, by defining a new random variable (RV), y_i, which represents the surplus at time instant i:

$$y_i = x_i - x_0 \tag{17}$$

Hence, eq.16 can be rewritten as:

$$S = \sum_{i=1}^{n} y_i \tag{18}$$

The PDF of y_i can be obtained by truncating the underlying PDF of x_i at x_0 level.

RUN-SUM PROBABILITY

An inspection of the right-hand side of eq.18 shows that the RV, S, has, in fact, two random terms in its structure. The first one is n which denotes the RV characterizing positive run-length, and the other RV is the summand, y_i. In other words, eq.18 is the random sum of RV's. In general, theoretically, there exists a joint PDF of S and n; therefore, one can write:

$$P(S,n) = P(S|n)P(n) \tag{19}$$

where $P(S|n)$ is the conditional probability of positive run-sum given that the positive run-length is equal to n and $P(n)$ is equivalent to $P(n_p = n)$ given by eq.5 or eq.15 for independent and dependent processes, respectively. It is important to note at this level that the RV y_i is continuous whereas the RV n is discrete and changes from zero to infinity. Consequently, the marginal PDF of run-sum can be written as:

$$P(S) = \sum_{n=0}^{\infty} P(S|n)P(n) \tag{20}$$

Although it is possible to solve eq.20 numerically on a digital computer to obtain the PDF of S, it has not been undertaken in this study. In the derivation of eq.20 no assumption as regards to the form of the PDF of observations has been made, which means that eq.20 can be applied even to skewed distributions. It is also applicable to dependent processes.

STATISTICAL MOMENTS OF RUN-SUM

The first objective study in relation to run-sum has been performed by Downer et al. (1967) who have succeeded in deriving analytical expressions for statistical moments of run-sum by making use of the moment and cumulative generating functions on the basis of normal independent processes. Their results will be rederived in this study by an entirely different method based on the random sums of RV's which seem to be rather attractive and simple to understand. Later, the application of the same methodology to the first-order Markov processes will yield analytical results for run-sums similar to the independent process results. Furthermore, it is possible to apply the methodology developed herein to any kind of stationary stochastic process provided that their run-length probabilities can be expressed analytically. The expected value of positive run-sum can be found as:

$$E(S) = \int_0^{\infty} SP(S)\mathrm{d}S \tag{21}$$

The substitution of eq.20 into eq.21 yields:

$$E(S) = \sum_{n=0}^{\infty} P(n) \int_{0}^{\infty} SP(S|n)\mathrm{d}S \tag{22}$$

The integration term is, in fact, the conditional expectation of S given n which can be denoted by:

$$E(S|n) = \int_{0}^{\infty} SP(S|n)\mathrm{d}S$$

Thus, eq.22 becomes:

$$E(S) = \sum_{n=0}^{\infty} E(S|n)P(n) \tag{23}$$

This is the general expression of the expectation, i.e., the first moment of positive run-sums. In a similar manner, higher-order moments, say m-th order, can be evaluated which leads to:

$$E(S^m) = \sum_{n=0}^{\infty} E(S^m|n)P(n) \tag{24}$$

The conditional expectation in eq.23 can be calculated from eq.18 by considering n as a given value: taking the expectation of both sides leads to:

$$E(S|n) = n \cdot E(y_i) \tag{25}$$

Substitution of eq.25 into eq.23 yields:

$$E(S) = E(y_i) \sum_{n=0}^{\infty} nP(n) \tag{26}$$

The infinite sum in eq.26 is the expected value of positive run-length, thus:

$$E(S) = E(y_i)E(n_p) \tag{27}$$

The conditional expectation of the second-order moment can again be calculated from eq.18 as:

$$E(S^2|n) = \sum_{i=1}^{n} E(y_i^2) + \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} E(y_i y_j)$$

or:

$$E(S^2|n) = \sum_{i=1}^{n} [V(y_i) + E^2(y_i)] + \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} [\mathrm{Cov}(y_i,y_j) + E^2(y_i)]$$

Furthermore:

$$E(S^2|n) = nV(y_i) + n^2E^2(y_i) + \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} \mathrm{Cov}(y_i,y_j) \qquad (28)$$

or, in terms of the autocorrelation coefficients:

$$\mathrm{Cov}(y_i,y_j) = V(y_i)\rho_{i,j} \qquad (29)$$

where $\rho_{i,j}$ is the autocorrelation coefficient between surpluses at time instances i and j. If the original observations are generated by a stationary stochastic process, then their surpluses above x_0 will also be stationary which implies that $\rho_{i,j} = \rho_{|i-j|}$, i.e. the autocorrelation coefficient of surpluses is a function of time difference between the two surpluses considered but not a function of absolute time. Therefore, eq.29 can be rewritten as:

$$\mathrm{Cov}(y_i,y_j) = V(y_i)\cdot\rho_{|i-j|} \qquad (30)$$

the substitution of which into eq.28 after some algebra yields:

$$E(S^2|n) = nV(y_i) + n^2E^2(y_i) + 2\,V(y_i)\cdot \sum_{i=1}^{n-1} (n-i)\rho_i \qquad (31)$$

The general expression of the second moment of positive run-sum can be obtained from eq.24 with $m = 2$ as:

$$E(S^2) = V(y_i)E(n_n) + E^2(y_i)E(n_p^2) + 2\,V(y_i) \sum_{n=0}^{\infty} \sum_{i=1}^{n-1} (n-i)\rho_i\cdot P(n) \qquad (32)$$

By considering eq.27 together with eq.32 the variance of positive run-sum becomes:

$$V(S) = E(S^2) - E^2(S) = V(y_i)E(n_p) + E^2(y_i)V(n_p) + $$

$$+ 2\,V(y_i) \sum_{n=0}^{\infty} \sum_{i=1}^{n-1} (n-i)\rho_i\cdot P(n) \qquad (33)$$

Any order of moments of S can be found by a similar calculation as performed above. In the case of independent observations, $\rho_i = 0$ for $i \geqslant 1$ and eq.33 reduces to:

$$V(S) = V(y_i)E(n_p) + E^2(y_i)V(n_p) \qquad (34)$$

or, more explicitly:

$$V(S) = [qV(y_i) + pE^2(y_i)]/q^2 \qquad (35)$$

The correlation between positive run-sum and corresponding positive run-length can in general be expressed for a given n value as:

$$\mathrm{Cov}(S,n) = E(Sn) - E(S)E(n) \tag{36}$$

where $E(Sn)$ denotes the cross-product moment of RV's S and n; it can be calculated by making use of eq.19 as:

$$E(Sn) = \int_0^\infty \sum_{n=0}^\infty nSP(S|n)P(n)\mathrm{d}S$$

or:

$$E(Sn) = \sum_{n=0}^\infty nP(n) \int_0^\infty SP(S|n)\mathrm{d}S$$

where the integration term is the conditional expectation of S given. Therefore:

$$E(Sn) = \sum_{n=0}^\infty nP(n)E(S|n) \tag{37}$$

Substituting eq.25 into eq.37 gives:

$$E(Sn) = E(y_i) \sum_{n=0}^\infty n^2P(n) \tag{38}$$

where the infinite summation is in fact the second order moment of positive run-length, $E(n_p^2)$; hence, finally eq.38 becomes:

$$E(Sn) = E(y_i)E(n_p^2) \tag{39}$$

or, from eq.36:

$$\mathrm{Cov}(S,n) = E(y_i)E(n_p^2) - E(y_i)E^2(n_p) = E(y_i)\,[E(n_p^2) - E^2(n_p)]$$

by taking into account that $V(n_p) = E(n_p^2) - E^2(n_p)$

$$\mathrm{Cov}(S,n) = E(y_i)V(n_p) \tag{40}$$

The autocorrelation coefficient, R, between positive run-sum and positive run-length can be found to be:

$$R = \mathrm{Cov}(S,n)/[V(S)V(n_p)]^{\frac{1}{2}} = E(y_i)\cdot[V(n_p)/V(S)]^{\frac{1}{2}} \tag{41}$$

As a conclusion, $E(S)$, $V(S)$ and R have been expressed in terms of the expected values $E(y_i)$ and $E(n_p)$ and the variances $V(y_i)$ and $V(n_p)$ of the surpluses and positive run-lengths, respectively. For an independent process, Feller (1957) has given that:

$$E(n_p) = 1/q \tag{42}$$

and:

$$V(n_p) = p/q^2 \tag{43}$$

In the case of first-order Markov process the corresponding expressions have been given by Şen (1976) as:

$$E(n_p) = 1/(1 - r) \tag{44}$$

and:

$$V(n_p) = r/(1 - r)^2 \tag{45}$$

APPLICATION TO INDEPENDENT PROCESSES

Although various moments of positive run-sum in the case of normal independent process have been analytically given by Downer et al. (1967), the same results will be derived on the basis of the methodology developed in the previous section. Downer et al. (1967) have employed the data generation method in order to derive the run properties of the skewed variables.

In general, the surpluses are distributed according to the truncated distribution, $f_t(x)$, of the original PDF, $f(x)$, with a truncation level at x_0. The truncated distribution can be obtained from $f(x)$ as:

$$f_t(x) = \frac{1}{N} \cdot f(x) \tag{46}$$

where N is a normalizing factor which makes the area under $f_t(x)$ equal to unity. Thus:

$$N = \int_{x_0}^{\infty} f(x)\mathrm{d}x \tag{47}$$

or:

$$N = 1 - q = p \tag{48}$$

Hence, eq.46 can be written as:

$$f_t(x) = (1/p) \cdot f(x) \tag{49}$$

Normal distribution

If the original RV's are normally distributed with mean μ and standard deviation σ then the surpluses above x_0 are distributed according to a truncated normal distribution which has its general form as:

$$f_t(x) = \frac{1}{p(2\pi)^{\frac{1}{2}}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2 \right], \qquad x_0 \leqslant x < \infty \tag{50}$$

or in the standardized form:

$$f_t(x) = \frac{1}{p(2\pi)^{\frac{1}{2}}} \exp(-x^2/2), \qquad x_0 \leqslant x < \infty \tag{51}$$

where now,

$$x_0 = (x_0 - \mu)/\sigma \tag{52}$$

The first two moments about x_0 of eq.51 can be found after some algebra as:

$$E(x) = \frac{1}{p(2\pi)^{\frac{1}{2}}} \int_{x_0}^{\infty} x \ \exp(-x^2/2)\mathrm{d}x = \frac{\exp(-\frac{1}{2}x_0^2)}{p(2\pi)^{\frac{1}{2}}} \tag{53}$$

and:

$$E(x^2) = \frac{1}{p(2\pi)^{\frac{1}{2}}} \int_{x_0}^{\infty} x^2 \ \exp(-x^2/2)\mathrm{d}x = 1 + \frac{x_0 \exp(-\frac{1}{2}x_0^2)}{p(2\pi)^{\frac{1}{2}}} \tag{54}$$

or, the moments about the origin become:

$$E(x) = \frac{\exp(-\frac{1}{2}x_0^2)}{p(2\pi)^{\frac{1}{2}}} + x_0 \tag{55}$$

and:

$$V(x) = 1 + \frac{x_0 \exp(-\frac{1}{2}x_0^2)}{p(2\pi)^{\frac{1}{2}}} - \frac{\exp(-x_0^2)}{2\pi p^2} \tag{56}$$

Now, it is possible to find various positive run-sum properties by substituting eqs.42 and 55 into eq.27 which yields:

$$E(S) = \frac{\exp(-\frac{1}{2}x_0^2)}{pq(2\pi)^{\frac{1}{2}}} + \frac{x_0}{q} \tag{57}$$

The variance and autocorrelation coefficient between the positive run-sum and corresponding run-length can be calculated through eqs.35 and 41, respec tively.

Log-normal distribution

The log-normal probability has been applied in water resources system design for a long time. It was adopted early in the statistical studies of hydrological data by Hazen (1914). The main reason for its adoption is its skewness which seems to represent the observed data more realistically than a symmetric PDF. It has been applied by various researchers to model daily stream flows, flood peak discharges and annual, monthly and daily rainfall.

If any hydrologic variable, y, were log-normally distributed then $x = \ln y$ is normally distributed. The general form of a log-normal distribution is given by Benjamin and Cornell (1970) as:

$$f(y) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma_{\ln Y}} \exp\left[-\frac{1}{2}\left\{\frac{1}{\sigma_{\ln Y}} \ln(y/\check{m}_Y)\right\}^2\right], \qquad 0 < y < \infty \tag{58}$$

where $\check{m}_Y$ and $\sigma_{\ln Y}$ are the two parameters of the distribution, namely, the median and standard deviation, respectively. On the other hand, the relationships between parameters and the moments are given as:

$$\check{m}_Y = m_Y \exp(-\tfrac{1}{2}\sigma_{\ln Y}) \tag{59}$$

$$\sigma^2_{\ln Y} = \ln(V_Y^2 + 1) \tag{60}$$

and the coefficient of skewness, γ, is:

$$\gamma = 3\, V_Y + V_Y^3 \tag{61}$$

where m_Y and V_Y are the mean and coefficient of variation of the log-normally distributed variable. From eq.49 one can obtain the truncated log-normal distribution as:

$$f(y) = \frac{1}{p(2\pi)^{\frac{1}{2}}\sigma_{\ln Y}} \exp\left[-\frac{1}{2}\left\{\frac{1}{\sigma_{\ln Y}} \ln(y/\check{m}_Y)\right\}^2\right], \qquad y_0 \leqslant y < \infty \tag{62}$$

where y_0 is the level of truncation of log-normally distribution variate. The n-th order moment can be easily obtained from eq.62 as:

$$E(y^n) = \frac{\check{m}_Y^n \cdot \exp(-\frac{1}{2}n^2\sigma^2_{\ln Y})}{p}\left[1 - F\left\{\frac{1}{\sigma_{\ln Y}} \ln(y_0/\check{m}_Y) - n \cdot \sigma_{\ln Y}\right\}\right] \tag{63}$$

where $F[\cdot]$ is the area under the standardized normal distribution form $-\infty$ to the value in the brackets. Thus, the expected value about the origin becomes:

$$E(y) = \check{m}_Y \cdot \exp(-\tfrac{1}{2}n^2\sigma^2_{\ln Y})\, \frac{1 - F\left[\frac{1}{\sigma_{\ln Y}}\ln(y_0/\check{m}_Y) - \sigma_{\ln Y}\right]}{1 - F\left[\frac{1}{\sigma_{\ln Y}}\ln(y_0/\check{m}_Y)\right]} \tag{64}$$

It is possible to obtain the variance, $V(y)$ of surpluses in the truncated log-normal distribution case by making use of the second-order moment for $n = 2$ in eq.63. The calculations of $E(S)$, $V(S)$ and R in this case have been provided in Table I. It is clear from this Table that as the coefficient of skewness increases $E(S)$ and $V(S)$ decrease.

APPLICATION TO DEPENDENT PROCESSES

The dependence between successive observations effects the magnitude of the design value such as the reservoir size, flood peaks, etc. It has been observed that an increase in the first-order autocorrelation coefficient, ρ, causes

TABLE I

The various run-sum properties of the log-normal independent process

		$m_Y = 1.00$, $\sigma_Y = 0.597$, $V_Y = 0.597$, $\gamma = 2.000$, $\sigma_{\ln Y} = 0.5521$, $\breve{m}_Y = 0.8586$				$m_Y = 1.000$, $\sigma_Y = 0.325$, $V_Y = 0.325$, $\gamma = 1.00$, $\sigma_{\ln Y} = 0.3169$, $\breve{m}_Y = 0.951$				$m_Y = 100.00$, $\sigma_Y = 0.030$, $V_Y = 0.01$, $\gamma = 0.030$, $\sigma_{\ln Y} = 0.1732$, $\breve{m}_Y = 0.9851$			
q	p	y	$E(S)$	$V(S)$	R	y	$E(S)$	$V(S)$	R	y	$E(S)$	$V(S)$	R
0.10	0.90	0.4235	14.966	205.276	0.9909	0.6339	16.839	100.143	0.9954	0.7892	18.192	95.810	0.9982
0.20	0.80	0.5340	8.426	28.142	0.9704	0.7287	9.120	24.421	0.9912	0.8517	9.539	22.41	0.9977
0.40	0.60	0.7479	5.156	7.314	0.9412	0.8785	5.175	5.458	0.9872	0.9433	5.125	4.639	0.9949
0.50	0.50	0.8586	4.552	4.694	0.9252	0.9510	4.402	3.257	0.9795	0.9851	4.240	2.619	0.9916
0.60	0.40	0.9856	4.226	3.205	0.9126	1.0290	3.911	2.029	0.9631	1.0280	3.674	1.156	0.9902
0.80	0.20	1.3650	4.124	1.601	0.8546	1.2410	3.432	0.769	0.9588	1.1390	3.002	0.506	0.9024
0.90	0.10	1.7406	4.075	1.107	0.7794	1.4260	1.771	0.4063	0.9260	1.2290	2.559	0.245	0.872

TABLE II

The various run-sum properties of the normal lag-one Markov process

x_0	$\rho = 0.2$			$\rho = 0.4$			$\rho = 0.6$		
	r	$E(S)$	$V(S)$	r	$E(S)$	$V(S)$	r	$E(S)$	$V(S)$
−1.00	0.8563	8.960	75.070	0.8747	10.276	103.52	0.8973	12.537	164.728
−0.50	0.7288	3.720	12.630	0.7687	4.362	19.076	0.8138	5.418	34.210
−0.40	0.6972	3.170	9.175	0.7424	3.734	14.275	0.7930	4.646	25.980
−0.30	0.6657	2.740	6.844	0.7162	3.231	10.467	0.7724	4.028	20.246
−0.20	0.6323	2.380	5.154	0.6888	2.810	8.332	0.7504	3.509	15.933
−0.10	0.5987	2.080	3.950	0.6602	2.457	6.351	0.7381	3.070	12.780
0.00	0.5640	1.830	3.068	0.6309	2.162	5.158	0.7047	2.702	10.367
0.10	0.5290	1.618	2.415	0.6011	1.910	4.137	0.6808	2.387	8.523
0.20	0.4943	1.441	1.934	0.5714	1.700	3.374	0.6319	1.980	6.443
0.30	0.4592	1.290	1.566	0.4992	1.397	2.398	0.6070	1.776	5.494
0.40	0.4252	1.162	1.287	0.5109	1.367	2.322	0.5818	1.597	4.728
0.50	0.3908	1.064	1.069	0.4803	1.233	1.960	0.5566	1.445	4.115
1.00	0.2400	0.507	0.3375	0.3375	0.792	1.142	0.4567	0.966	2.411

the expected value of the positive run-lengths to increase, Şen (1976). Since the expected value of positive run-length is explicitly related to the expectation of positive run-sum as is obvious from eq.27, it is possible to conclude that ρ influences $E(S)$. The autocorrelation coefficient does not affect the probability distribution function of observations. Thus, in the case of a normal distribution $E(S)$ can be obtained by substituting eq.44 and eq.55 into eq.27 which leads to:

$$E(S) = \frac{x_0}{(1-r)} + \frac{\exp(-\frac{1}{2}x_0^2)}{p(1-r)(2\pi)^{\frac{1}{2}}} \tag{65}$$

If the generating mechanism of observations is assumed to be a lag-one Markov process given in eq.13 then after the substitution of eqs.15, 44 and 45 into eq.33 and some tedious algebra the variance, $V(S)$, can be derived as:

$$V(S) = \frac{1}{(1-r)}\left[V(y) + \frac{rE^2(y)}{(1-r)}\right] + \frac{2\,V(y)\rho}{(1-\rho)}\left[\frac{r}{(1-r)} + \frac{(1-r)}{(1-\rho)(1-\rho r)}\right] \tag{66}$$

which in the case of normal independent process ($r = \rho$) reduces to eq.35. The solutions to eq.65 and eq.66 have been presented in Table II for various ρ values. This table reveals that in the case of Markov processes, $E(S)$ and $V(S)$ increase but R decreases with an increase in ρ value.

CONCLUSIONS

A general methodology for determining various run-sum properties of hydrological processes has been presented and its applications to log-normally distributed variables as well as to the normal lag-one Markov process have been performed. On the basis of the study in this paper, the following important conclusions can be drawn:

(1) The presented method can be used for dependent and independent processes with normal or non-normal distributions.

(2) The correlation between successive observations increases the mean and variance of run-sums.

REFERENCES

Benjamin, J.R. and Cornell, C.A., 1970. Probability Statistics, and Decision for Civil Engineers. McGraw-Hill, New York, N.Y., 684 pp.

Cramer, H. and Leadbetter, M.R., 1967. Stationary and Related Stochastic Processes. Wiley, New York, N.Y., 384 pp.

Downer, R.N., Siddigui, M.M. and Yevjevich, V., 1967. Applications of runs to hydrologic droughts. Proc. Int. Hydrol. Symp., Fort Collins, Colo., pp. 496—505.

Feller, W., 1957. An Introduction to the Probability Theory and its Application, 1. Wiley, New York, N.Y., 509 pp.

Hazen, A., 1914. Discussion on flood flows. ASCE, Trans., 77.

Saldarriaga, J. and Yevjevich, V., 1970. Application of run-lengths to hydrologic series. Colo. State Univ. Hydrol. Pap., 40: 56 pp.

Şen, Z., 1976. Wet and dry periods of annual flow series. J. Hydraul. Div., ASCE, Proc. Pap. 12457, 102(HY10): 1503—1514.