The numerical calculation of extreme wet and dry periods in hydrological time series

ZEKÂI SEN Civil Engineering Faculty, Department of Hydraulics and Water Power, Istanbul Technical University, Taskisla, Taksim, Istanbul, Turkey*

Received 14 May 1979; revised 31 August 1979

Abstract. A numerical method for evaluating the possibilities of the extreme wet and the extreme dry periods of a hydrological sequence is presented on the basis of stationary independent and Markovian processes that are currently employed in the planning and operation of water resources systems. The validity of the formula has been checked against the Monte Carlo simulation results obtained on a digital computer. Recurrence relationships for the probability distribution functions of the longest wet and dry periods have been derived by direct enumeration and the statistical properties of these extremes are presented.

Méthode numérique d'évaluer les périodes extrêmes de dessèchement et de sécheresse dans une série de temps hydrologique

Résumé. Une méthode numérique d'évaluer les potentiels des extrêmes durées de sécheresse et de dessèchement dans une série hydrologique est développée au base des processus Markovian qui sont généralement utilisés dans l'aménagement des ressources d'eau. La validité de des formules est contrôlée par les simulations obtenues en ordinateur digital. Une relation de récurrence est dérivée pour décrire la fonction de probabilité des durées extrêmes de sécheresse et de dessèchement par une application d'énnumeration. Par conséquence, les caractéristiques statistiques de ces variables sont présentées.

INTRODUCTION

Hydrological processes such as those concerned with rainfall, runoff, evaporation, groundwater etc. evolve continuously in nature. Water engineers aim at quantifying these processes in order to make objective assessments of water resources. For this purpose at various measurement stations scattered all over the world the behaviour of these processes is determined in time as well as in space. A time series of past observations is obtained and can be processed through statistical methods. Due to man's ignorance and lack of complete knowledge about the evolution of natural processes, the recorded time series are considered to be random, and future values cannot be known exactly. Therefore, any variable related to the time series, such as drought duration and intensity, reservoir capacity, design discharge etc., is also regarded as a random variable. The treatment of such a variable can only be

* Address until October 1980: King Abdulaziz University, Faculty of Earth Sciences, PO Box 1744, Jedidah, Saudi Arabia.
achieved through using probabilistic, statistical and mathematical tools. This paper concentrates only on the critical drought duration distribution function which provides a useful means of assigning risk statements to various drought durations. The probability distribution of the actual critical drought duration will be derived in this paper on the basis of stationary stochastic processes.

**CRITICAL DROUGHT DURATION**

An objective definition of drought properties of a given time series can be achieved by means of the statistical theory of runs. The first classical definition of runs was by Uspensky (1937), who regarded a success run of length $r$ in a series of independent trials, as an occurrence of success at least $r$ times in succession. A serious drawback to this definition can be detected when exactly $r$ successive successes are considered, in which case the occurrence of a success at the $(r+1)$th trial may invalidate the run completed at the $r$th trial. Feller (1968) gave a definition of runs based on recurrence theory and Bernoulli trials as follows. A sequence of $n$ events, S (success) and F (failure), contains as many S-runs of length $r$ as there are non-overlapping uninterrupted blocks containing exactly $r$ events S each. This definition is unsuitable from the practical point of view because it does not say anything about the start and finish of the run.

A definition of runs given by Mood (1940) seems to be most revealing for the analysis of various run properties since a run is defined as a succession of similar events preceded and succeeded by different events with the number of similar events in the run referred to as its length. In the light of this definition Yevjevich (1967) gave a definition of hydrological droughts by truncating a given time series $x_1, x_2, \ldots, x_n$ at a constant truncation level $x_0$ as in Fig. 1. In this figure a non-interrupted sequence of negative deviations preceded and succeeded by positive deviations is referred to as the negative run-length in statistics and drought duration in hydrology. These two labels are used interchangeably throughout this paper. The maximum negative run-length is adopted as the measure of critical drought duration. At any pre-selected truncation level there is one and only one critical drought duration. Given the probabilistic structure of the time series together with a truncation level, the probability distribution function of the critical drought duration in a finite sample can be found through the enumeration technique illustrated in the following example.

Although some studies concerning the longest negative run-length exist in the literature they are not particularly numerous. For instance, the expression of the probability distribution function of the longest negative run-length in finite but
independent series has been given as an approximation by Feller (1968). On the other hand Millan (1972) gave an exact formula for the longest negative run-length distribution for an irreducible Markov chain of two ergodic states on the basis of combinatorial analysis. These formulae are tedious to evaluate and an approximation to critical drought duration in stochastic processes is necessary.

**ENUMERATION OF CRITICAL DROUGHT DURATION**

**Independent Bernoulli trials**

In order to discuss the enumeration method explicitly, first independent processes are considered. Truncation of such a process at a constant level yields two elementary and mutually distinct events, namely, a surplus \( (x_i > x_0) \) and a deficit \( (x_i < x_0) \) with probabilities \( p \) and \( q = 1 - p \), respectively. Since the longest positive and negative run-lengths have similar derivations of probability, in the following example only the derivation of the longest positive run-length probability will be examined. If the probability of the longest run-length \( L \) in a sample size \( i \) equal to \( j \) is denoted by \( P_i(L = j) \), then for sample size \( i = 1 \) one can simply write

\[
P_1(L = 0) = q
\]
\[
P_1(L = 1) = p
\]

(1)

In the case of \( i = 2 \) there are \( 2^2 \) combinations of surplus and deficit events, out of which the two with sequential surplus-deficit and deficit-surplus events constitute the longest run-length occurrences of length one. On the other hand, the combination of surplus-surplus, deficit-deficit events gives rise to the longest run-length equal to two, whereas a deficit-deficit combination results in a length equal to zero. Since, at this stage the elementary events are assumed to be independent, the probabilities of these combinations are

\[
P_2(L = 0) = P_1(L = 0)q
\]
\[
P_2(L = 1) = P_1(L = 1)q + P_1(L = 0)p
\]
\[
P_2(L = 2) = P_1(L = 2)q + P_2(L = 1)p
\]

(2)

In the same way \( 2^3 \) combinations for \( i = 3 \) include combined events of deficit-deficit-deficit, deficit-deficit-surplus, deficit-surplus-surplus, deficit-super surplus-deficit, surplus-deficit-deficit, surplus-deficit-surplus, surplus-deficit-deficit, surplus-deficit-surplus, surplus-deficit-deficit, surplus-deficit-surplus. After a close inspection of these combinations, the probabilities of longest run-length can be found in terms of the previous probabilities:

\[
P_3(L = 0) = P_2(L = 0)q
\]
\[
P_3(L = 1) = P_2(L = 1)q + P_2(L = 0)p + P_1(L = 1)q p
\]
\[
P_3(L = 2) = P_2(L = 2)q + P_2(L = 1)q - P_2(L = 1)q p + P_2(L = 2)q p
\]
\[
P_3(L = 3) = P_2(L = 3)q + P_2(L = 2)q - P_2(L = 2)q p + P_2(L = 3)q p
\]

(3)

Similar types of recurrence relationships can be obtained for \( i = 4 \) as

\[
P_4(L = 0) = P_3(L = 0)q
\]
\[
P_4(L = 1) = P_3(L = 1)q + P_3(L = 0)p + P_3(L = 1)q p
\]
\[
P_4(L = 2) = P_3(L = 2)q + P_3(L = 1)q - P_3(L = 1)q p + P_3(L = 2)q p
\]
\[
P_4(L = 3) = P_3(L = 3)q + P_3(L = 2)q - P_3(L = 2)q p + P_3(L = 3)q p
\]
\[
P_4(L = 4) = P_3(L = 4)q
\]

(4)
If desired, it is possible to continue writing the probabilities of the longest run-length for values of $n$ higher than 4. However, a close inspection of Equations 1–4 reveals a general pattern of relationships. For instance, generally, the probability of a no run-length existing in a sample of size $n$ can be written as

$$P_n[L = 0] = P_{n-1}[L = 0]q$$  \hspace{0.2cm} (5a)

Furthermore, it is also straightforward to write a recurrence relationship for the longest run-length to be equal exactly to the sample size considered by the comparison of the last lines in Equations 1–4, leading to

$$P_n[L = n] = P_{n-1}[L = n-1]p$$  \hspace{0.2cm} (5b)

However, the general probability expression of the longest run-length of size $j$ where $0 < j < n$ is

$$P_n[L = j] = P_{n-1}[L = j]q + \left\{ P_{n-1}[L = j-1] - \sum_{i=1}^{k_1} P_{n-i-1}[L = j-1]qp^{i-1} \right\} p + \sum_{i=1}^{k_2} P_{n-i-1}[L = j]qp^j$$  \hspace{0.2cm} (5c)

where $k_1 = \min (n-j, j-1)$ and $k_2 = \min (n-j-1, j)$. Equations 5a, b, c represent completely the exact probability of the longest positive run-length. By interchanging $p$ and $q$ the probability of the longest negative run-length $N$, i.e. critical drought duration, can be obtained as

$$P_n[N = 0] = P_{n-1}[N = 0]p$$
$$P_n[N = j] = P_{n-1}[N = j]p + \left\{ P_{n-1}[N = j-1] - \sum_{i=1}^{k_1} P_{n-i-1}[N = j-1]pq^{i-1} \right\} q + \sum_{i=1}^{k_2} P_{n-i-1}[N = j]pq^j$$
$$P_n[N = n] = P_{n-1}[N = n]q$$  \hspace{0.2cm} (6)

Of course in the above formulation for any $n$ value

$$\sum_{i=0}^{n} P_n[L = i] = \sum_{i=0}^{n} P_n[N = i] = 1$$

**Dependent Bernoulli trials**

The exact distribution of the longest positive and negative run-length in $n$ trials for the case of dependent events has been given by Millan (1972) through the application of combinatorial analysis. In this case the result of each trial is dependent only on the outcome of the previous trial. For a run-length to develop two sources of information are required, namely, (i) information about the initial state, and (ii) information about transitions between successive states. The initial state is independent of any other state and therefore one can assume two values, either a surplus or a deficit. In terms of probability these two states will be denoted by $P(\cdot + \cdot)$ and $P(\cdot - \cdot)$, respectively.

There exist four possibilities of transitions between successive trials: (i) transition from a surplus to a surplus, (ii) transition from a surplus to a deficit, (iii) transition from a deficit to a surplus, and (iv) transition from a deficit to a deficit. These four
possibilities will be denoted in this paper by conditional probabilities as \( P(\cdot | \cdot) \). If necessary, one can also use joint probabilities provided that \( P(\cdot, \cdot) = P(\cdot | \cdot) P(\cdot) \), \( P(\cdot, \cdot) = P(\cdot | \cdot) P(\cdot) \), \( P(\cdot, \cdot) = P(\cdot | \cdot) P(\cdot) \), and \( P(\cdot, \cdot) = P(\cdot | \cdot) P(\cdot) \). Homogeneous Bernoulli trials are considered in this paper. Therefore, the transition probabilities as well as the state probabilities are independent of absolute time. Hence, the state probabilities can be written in terms of transition probabilities as

\[
\begin{align*}
P(+) &= P(\cdot | \\cdot) P(\cdot) + P(\cdot | \\cdot) P(\cdot) \\
P(-) &= P(\cdot | \\cdot) P(\cdot) + P(\cdot | \\cdot) P(\cdot)
\end{align*}
\]

where \( P(\cdot) + P(\cdot) = 1; P(\cdot, \cdot) + P(\cdot, \cdot) = 1 \) and \( P(\cdot, \cdot) + P(\cdot, \cdot) = 1 \). The transition probabilities can be put in matrix form as

\[
T = \begin{pmatrix}
P(\cdot, \cdot) & P(\cdot, \cdot) \\
P(\cdot, \cdot) & P(\cdot, \cdot)
\end{pmatrix}
\]

Following all these definitions the enumeration method can be applied similarly to the case of independent processes. However, a slight change of notation is necessary due to the dependence of a trial on the previous one. Hence, the probability of the longest positive run-length \( L \) being equal to an integer \( j \) value in a sample size of \( i \) with a surplus state at the final stage will be denoted by \( P^+\{L=j\} \). At the initial stage there exist no transition and one can simply write as in Equation 1:

\[
P^-\{L=0\} = P(\cdot) = q \\
P^+\{L=1\} = P(\cdot) = p
\]

For \( i = 2 \) the transitions come to play an important role, leading to

\[
\begin{align*}
P^-\{L=0\} &= P^-\{L=0\} P(\cdot) \\
P^+\{L=1\} &= P^+\{L=0\} P(\cdot) \\
P^-\{L=1\} &= P^-\{L=1\} P(\cdot) \\
P^+\{L=2\} &= P^+\{L=1\} P(\cdot)
\end{align*}
\]

By continuing the enumeration in the same way it is possible to arrive at a general formula:

\[
\begin{align*}
P^-\{L=0\} &= P^-\{L=0\} P(\cdot) \\
P^+\{L=j\} &= \sum_{m=0}^{k_1} P^-\{L=m\} P(\cdot) P^m(\cdot) \\
P^-\{L=j\} &= \sum_{m=1}^{k_2} P^-\{L=m\} P(\cdot) P^m(\cdot)
\end{align*}
\]

where \( k_1 = \min(i-j-1,j-1) \) and \( k_2 = \min(i-j-1,j-1) \). The probabilities on the longest drought duration can be evaluated from Equation 11 provided that \( P(\cdot, \cdot) \), \( P(\cdot, \cdot) \), \( P(\cdot, \cdot) \), and \( P(\cdot, \cdot) \) are interchanged by \( P(\cdot, \cdot) \), \( P(\cdot, \cdot) \), \( P(\cdot, \cdot) \), and \( P(\cdot, \cdot) \).
and $P(+|+)$, respectively. Of course, in the above equation the probability of the longest run-length can be written as

$$P(L=j) = P^+(L=j) + P^-(L=j)$$

(12)

The mean and the variance of the longest run-length can be evaluated through

$$E(L) = \sum_{j=1}^{n} jP(L=j)$$

(13)

and

$$V(L) = \sum_{j=1}^{n} j^2P(L=j) - E^2(L)$$

(14)

respectively.

**NUMERICAL SOLUTION**

The analytical probability distribution functions given in Equation 11 are solved at $q=0.4$ truncation level, for sample size $n=25$ and for serial correlation coefficients $\rho=0.1$, 0.3 and 0.7. This range of correlation coefficients covers the situations encountered in practice for annual flow sequences. The transition probabilities for these correlation coefficients are present in Table 1.

| $\rho$ | $P(++)$ | $P(+|-)$ | $P(-|+)$ | $P(--)$ |
|--------|--------|----------|----------|--------|
| 0.1000 | 0.4400 | 0.3800   | 0.5600   | 0.6200 |
| 0.3000 | 0.5200 | 0.3200   | 0.4800   | 0.6800 |
| 0.7000 | 0.6900 | 0.2000   | 0.3100   | 0.8000 |

The numerical solutions of Equation 11 are given in Fig. 2 together with the Monte Carlo simulation results. An inspection of this figure is sufficient to prove that the developed methodology for evaluating the longest run-lengths yields results which are practically identical to the simulation results. This agreement between the two approaches is satisfactory even for big correlations such as $\rho=0.7$. It is obvious from the same figure that for a given probability of exceedance (risk) the longest run-length to be exceeded becomes more severe for large correlations provided that the sample length and truncation level remain the same. Therefore, persistent annual flows require larger reservoir capacities than independent processes. On the other hand, the longest run-length is quite sensitive to the correlation coefficient which must be estimated from a given finite observations sequence with a great care and accuracy.

The expected longest run-lengths are given in Fig. 3 which shows that the increase in the longest drought duration is rather slow at large sample sizes. However, in small sample sizes the increase is quite rapid. Figure 4 represents the change of variance of the longest drought duration with the sample size. The same arguments as in $E(L)$ apply to $V(L)$. 
Evaluation of extreme wet and dry periods

**FIG. 2.** Distribution of the longest positive run-length in sample of size $n$: ---, analytical; ----, simulation.

**FIG. 3.** Expected longest run-length versus sample length.

**FIG. 4.** Variance of the longest run-length versus sample length.
Although the above formulation is exact for independent processes it is a very good approximation for Markov chains and yields practically satisfactory results. A prerequisite for the application of these formulations in an actual design situation is the calculation of transition probabilities from observed data. Hence, the probabilities of extreme wet or extreme dry durations that are likely to occur at the same site can be easily evaluated and with the concept of the risk a design value can be found.

CONCLUSIONS

A general methodology for evaluating the probability of the longest drought or wet period has been presented on the basis of an enumeration technique. The formulae developed are applicable to independent as well as dependent processes.

REFERENCES


