### Title

Stochastic renewal model of low-flow streamflow sequences




### Journal

Stochastic Environmental Research and Risk Assessment, 10(1)

### ISSN

1436-3240



### Authors

Loaiciga, HA

Leipnik, RB



### Publication Date

1996-02-01

### DOI

10.1007/bf01581794

Peer reviewed

Stochastic Hydrology and Hydraulics 10, 65-85 

# Stochastic renewal model of low-flow streamflow sequences

**H. A. Loaiciga**
Dept. of Geography and Environmental Studies, University of California, Santa Barbara, CA 93106, USA

**R. B. Leipnik**
Department of Mathematics, University of California, Santa Barbara, CA 93106, USA

**Abstract**: It is shown that runs of low-flow annual streamflow in a coastal semiarid basin of Central California can be adequately modelled by renewal theory. For example, runs of below-median annual streamflows are shown to follow a geometric distribution. The elapsed time between runs of below-median streamflow are geometrically distributed also. The sum of these two independently distributed geometric time variables defines the renewal time elapsing between the initiation of a low-flow run and the next one. The probability distribution of the renewal time is then derived from first principles, ultimately leading to the distribution of the number of low-flow runs in a specified time period, the expected number of low-flow runs, the risk of drought, and other important probabilistic indicators of low-flow. The authors argue that if one identifies drought threat with the occurrence of multiyear low-flow runs, as it is done by water supply managers in the study area, then our renewal model provides a number of interesting results concerning drought threat in areas historically subject to inclement, dry, climate. A 430-year long annual streamflow time series reconstructed by tree-ring analysis serves as the basis for testing our renewal model of low-flow sequences.

**Key words**: Streamflow, drought, tree-ring data, renewal model, geometric variables.

## 1 Introduction

In a study of natural hazards around the world, that included droughts, tropical cyclones, floods, earthquakes, volcanic eruptions, and, virtually, any conceivable natural hazard, Bryant (1991; see Table 1.1) concluded that droughts had the greatest adverse impacts among all surveyed hazards. In spite of their environmental and economic significance, progress made on the mathematical/statistical study of droughts has been slow. This can be attributed to (1) the scanty long-term hydroclimatic data available to describe spatial and temporal drought patterns, and (2) the theoretical puzzle of placing drought phenomena within a tractable probabilistic framework (for an early probabilistic study of droughts see, e.g., Blumenstock, 1943).

Before proceeding further into the subject matter, the concept of drought deserves more precise definition. Climatologists, meteorologists, hydrologists, water resources planners, and irrigation analysts have somewhat different interpretations of what a drought is (see, e.g., Yevjevich, 1964; 1967; Palmer, 1965; Loaiciga et al., 1992a). This is a result of the gamut of spatial and temporal scales with which professionals in various disciplines operate. It is also rooted in the diversity of economic and environmental impacts that reduced levels of surface moisture are perceived to have by different professional disciplines. To illustrate, meteorologists use an index that combines precipitation and potential evapotranspiration to describe drought severity, the well-known Palmer drought severity index (Palmer, 1965). Typical time-scales to define the onset and duration of meteorological drought so described are days and weeks. These time scales are relevant to irrigation scheduling, where a few days of sustained crop-stressful conditions (e.g., low precipitation and high evapotranspirative rates) can determine the crop yield. On the other extreme of the spectrum of time scales, water resources planners consider much longer time scales in defining drought incidence. In California, for example, a state that supports the largest water resources infrastructure in the world (Marino and Loaiciga, 1985), the initial state of water storage in reservoirs and the duration (in years) of dry spells (say, measured in terms of below-median annual precipitation) are the major determinants on the onset of drought conditions (Loaiciga, 1988). In regards to the spatial scales of droughts, previous meteorologic (e.g., Stockton and Meko, 1975) and hydrologic (e.g., Loaiciga et al., 1992a; 1993)) studies have focused attention on regional-scale drought coverage (i.e., on the order of $10^4$ to $10^5$ $km^2$). This is consistent with empirical evidence suggesting that protracted drought conditions are typically associated with sustained synoptic-scale climatic anomalies and with teleconnections affecting climatic conditions over broad areas (Loaiciga et al., 1993).

In this study, we equate drought with sustained low-streamflow conditions. The annual streamflow time series used in this work, spanning over 430 years, was obtained from a previous tree-ring reconstruction of annual streamflow (Michaelsen and Haston, 1988) for the Santa Ynez River (Santa Barbara County, California). Turner (1992) has shown that the Michaelsen and Haston (1988) streamflow reconstruction reproduces well the persistence pattern of low- and high-flow annual streamflow runs, based on a detailed analysis of the Hurst coefficient (Hurst, 1951) of both observed and reconstructed streamflows. Turner (1992) also shows the excellent predictive skill of the cited streamflow reconstruction. Following other authors (Yevyevich, 1964, 1972; Zektser and Loaiciga, 1993), streamflow is chosen in this work as the indicator for surface hydrologic conditions, and thus, as an indicator for droughts. Streamflow integrates other major fluxes in the hydrologic cycle (precipitation, snowmelt, evapotranspiration, and baseflow), and, therefore, from the perspective of water supply analysis, it is uniquely suited as a drought indicator. In fact, Loaiciga et al. (1992a, 1993) used annual streamflow in the study of regional-scale drought in the Sacramento river basin of California and the Upper and Lower Colorado river basins. In California, a state with a vast water resource infrastructure, for example, it was found that sequences of below-median annual stream flow lasting three or more years almost inevitably led to hydrologic drought conditions for any initial water storage and water demand level (Loaiciga et al., 1992a, 1993). Following these studies, sustained low (annual) stream flow is the basis for the definition of hydrologic drought in this paper. We present a probabilistic theory of droughts based on the fundamental renewal processes of stochastic analysis.

## 2 Low-flow runs as a renewal process

Suppose that a time series of annual runoff is given by $Q_1$, $Q_2$, ..., $Q_t$, where time ranges from year 1 to year t. In attempting to characterize runs of runoff years, we prefer to classify runoff into two exclusive categories, such as, say, below-median runoff or above-median runoff. Then, a run of, say, below-median years can be of length D = 1, 2, .... Similarly, one could define the first 40-percentile, or $Q_{0.40}$, for which annual streamflow falls below with a 40% probability, as the threshold for classifying runoff years. In that instance, a runoff year is either equal to or less than $Q_{0.40}$, or it is larger than it. Exactly where the classification threshold lies is entirely the analyst's decision. That decision, certainly, must be based on an understanding of the hydrologic consequence of dry runs in the area of study (Klemes, 1974), as well as on water storage characteristics, water level demands, water transfers and the like, all factors important to hydrologic drought, that, if considered jointly would render the mathematical-probabilistic analysis of drought too complex (this issue is treated further later in this paper). For our study, the severity of drought is measured in terms of the level of streamflow through the specification of a streamflow threshold (e.g., median streamflow); recurrence and the probable duration of drought are derived by probabilistic analysis.

It may appear unreasonable in some instances to consider one-year runs of "dry" years to be relevant to droughts. This would be the case in areas where there is multiyear carryover storage capacity of reservoirs. For the purpose of mathematical analysis, however, this is a convenient convention imposed on the probabilistic model of droughts. It is possible to introduce a threshold constraint whereby the onset of a drought is subject to a minimum length of runs of dry runoff years, as shown later in this work.

The concept of droughts as a renewal process is illustrated in Figure 1. Starting at time t = 0, there is an elapsed period of time, $\tau_1$, until the first drought (i.e., a run of dry years) occurs. At that point, the number of events occurred in time t, denoted by N(t), takes the value of 1. The first dry run lasts $D_1$ years, and is followed by an interarrival time of $\tau_2$ years until the next drought is initiated at time $t = \tau_1 + D_1 + \tau_2$. N(t) increases to 2 at that point The second drought lasts $D_2$ years, and the renewal "cycle" is initiated again. In general, a renewal cycle time, or more concisely, renewal time, is the time R comprising the duration of a drought D plus the subsequent interarrival time $\tau$; therefore, $R = D + \tau$. The term "renewal" (see, e.g., Blackwell, 1953; Feller, 1957; Parzen, 1964) chosen to describe this type of stochastic counting process conveys the idea that the phenomenon in question (droughts in this instance) reappears with statistical regularity over time (i.e., drought recurrence is probabilistically stationary). There need be no cyclic behavior in the sense of a perfectly repetitive deterministic process with known period. Statistical regularity means in this instance that the elapsed time between the initiation of a drought and the beginning of the next one (i.e., the renewal time R of Figure 1) has a time-independent distribution with constant expected value.

Renewal processes are a generalization of the fundamental Poisson stochastic process. Poisson processes have been considered mainly in hydrology in the study of floods or other extreme hydrologic phenomena (e.g., Todorovic, 1978). In the Poisson model the recurring phenomena has an instantaneous duration (i.e., the duration D in Figure 1 is zero) with independent, exponentially distributed, interarrival times

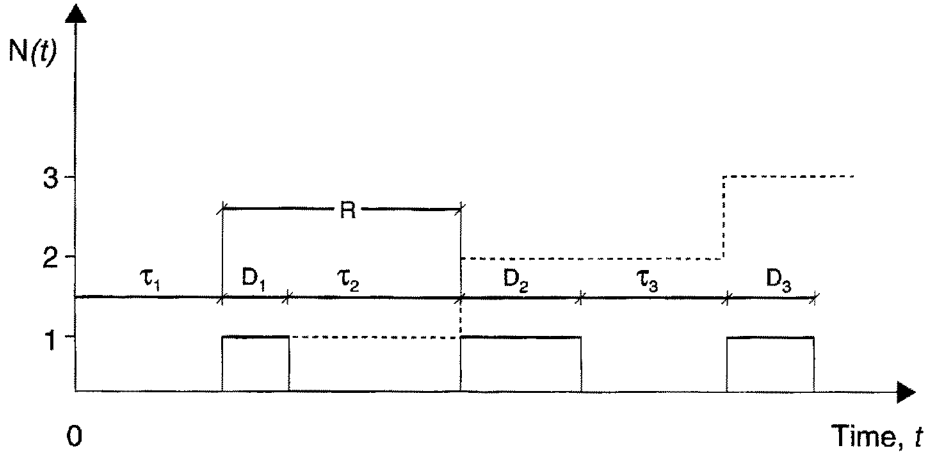


**Figure 1.** The concept of drought as a recurrence process. D is the duration of drought, $\tau$ is the duration of nondrought conditions, R is the renewal time, and N(t) is the number of droughts in period t.

($\tau$ in Figure 1). In the renewal process, the recurring phenomenon has a stochastic duration D, and, the interarrival time $\tau$ need not be exponentially distributed. The critical requirement, which we claim to hold for drought recurrence (at least in the hydroclimatic regimes considered in this study), is that the times D and $\tau$ must be independent (although not necessarily identically distributed). In the context of our discussion, this means that the duration of a drought does not influence the waiting time until the next one. This assumption has been established in (Loaiciga et al., 1992a,b; and Loaiciga et al., 1993).

It is worth indicating that the study of runs of streamflow conditions, defined classically in terms of deviations about the mean or median of a time series, has a long tradition in hydrologic analysis (Hurst, 1951; Yevjevich, 1964; Mandelbrot and Wallis, 1969; Klemes, 1974). Statistics such as the run-length, run-sum, maximum run-length, and the like, have been studied extensively. In this study we aim at the probability distribution of the number of droughts in a period t, interarrival times, risk of droughts, expected number of events in an interval t, etc., using the renewal model to explain the underlying probabilistic process. This perspective on the probabilistic study of droughts appears novel.

## 3 The distribution of the renewal time

Figure 2 shows histograms of below- (solid line) and above-median (dashed line) annual flow durations in the Santa Ynez river of Central California. Figure 2 was developed from a 430-year streamflow time series constructed by dendrochronological analysis (Michaelsen and Haston, 1988) and tested for accuracy and hydrologic consistency by Turner (1992). (The hydroclimatology of the Santa Ynez river basin has been described by Upson and Thomasson, 1951). These exponentially-shaped histograms suggest (this will be corroborated in a latter section) that the underlying

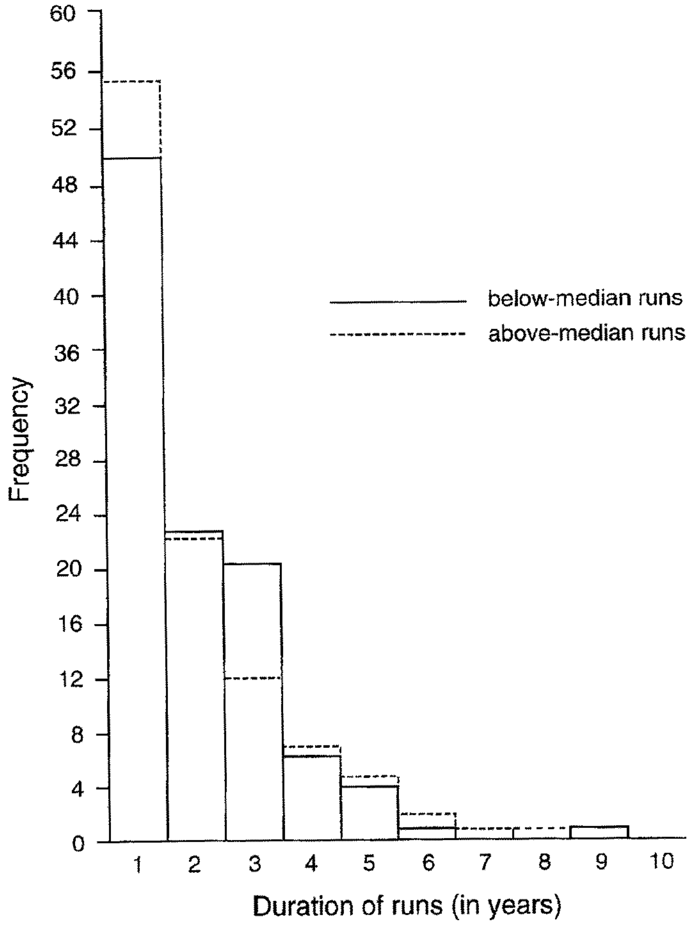


**Figure 2.** Histograms of below-median (solid line) and above-median (dashed line) annual flow durations for the Santa Ynez river of Central California.

probability model of below-median and above-median durations may be adequately described by a geometric distribution (Loaiciga et al., 1992b). Assume, in general, that the runs of dry and non-dry years (of which a special case would be runs of below and above-median flows) are governed by geometric distributions with parameters $p_1$ and $p_2$, respectively, i.e.,

$$P(D = r) = (1 - p_1)\, p_1^{r-1}, \qquad r = 1, 2, \ldots \tag{1}$$

$$P(\tau = r) = (1 - p_2)\, p_2^{r-1}, \qquad r = 1, 2, \ldots \tag{2}$$

in which r represents the length of the runs in either case. (Note that the lengths of dry and non-dry runs are denoted by D and $\tau$, respectively.) The parameters $p_1$ and $p_2$ satisfy the relation $p_1 + p_2 = 1$, since $p_1$ defines the probability of a dry streamflow condition and $p_2$ is the probability of a non-dry streamflow. In particular, when below-median flow defines a dry year, $p_1 = p_2 = p = 1/2$.

The renewal time R, i.e., the time from the beginning of a dry run to the initiation of the next dry run, is equal to the sum of the two independent geometric variables defined in equations (1) and (2). The distribution of the renewal time R is then (see Appendix A for a proof):

$$P(R = r) = \frac{(1-p_1)(1-p_2)}{p_1 - p_2}\left[p_1^{r-1} - p_2^{r-1}\right] \quad (3)$$

where r = 2, 3, .... This distribution becomes $P(R = r) = (1-p)^2(r-1)p^{r-2}$, when $p_1 = p_2 = p$. The distribution of equation (3) is of fundamental importance here. From (3), it is possible to define (i) the risk of a drought, (ii) the probability of occurrence of a number of droughts r within a time interval t, and (iii) the expected number of droughts within a time interval t. This provides a pretty complete probabilistic description of drought incidence, though other statistics could be asked for, and many could be calculated efficiently. Incidentally, the expected value of the renewal time R in Eq. (3), $\mu$, is equal to 4 when $p_1 = p_2 = 0.50$, i.e., when the median streamflow is the threshold that divides dry and wet streamflow runs. This theoretical value of $\mu$ agrees well with the calculated value for $\mu$ obtained from the histogram of below-median streamflow runs shown in Figure 2, which is obtained by dividing the length of the time series, t, by the number of runs observed in time-period t, N(t), or 430/105 = 4.09. The similarity between the theoretical renewal time $\mu$ and its calculated analog follows from a rather general theorem of stochastic processes (Feller, 1957; Parzen, 1964; Ross, 1985) which states that for sufficiently large t, $N(t)/t \rightarrow \mu^{-1}$. In reference to this theorem, it is possible to estimate from Figure 2, for example, that the average recurrence time of 5-year droughts is approximately 108 years (after rounding 430/4).

## 4 Drought probabilities and related results

### *4.1 The probability of the number of droughts N(t)*

It has been established that the interarrival time for the renewal drought process conceptualized in Figure 1, has a distribution function given by Eq. (3). The waiting time until the r-th drought, $W_r$, is $W_r = \sum_{j=1}^{r} R_j$, where the interarrival times $R_j$ are independent and identically distributed according to the distribution in Eq. (3). A basic relation between the number of droughts in a period t, N(t), and the waiting time to the r-th event is that (see, e.g., Parzen, 1964) $N(t) \leq r$ if and only if $W_{r+1} > t$. Upon reflection on this basic relation, it follows that $N(t) = r$ if and only if $W_r \leq t$ and $W_{r+1} > t$. It is concluded at once that (see, e.g., Parzen, 1964):

$$P[N(t) = r] = P[W_r \leq t] - P[W_{r+1} \leq t] \quad (4)$$

Equation (4) can be expressed in terms of the renewal parameters of interest, i.e., the geometric parameters $p_1$ and $p_2$, and the rth- and r+1st-order convolutions of the renewal time R introduced by the waiting times $W_r$ and $W_{r+1}$, respectively. Appendix B establishes the following result:

$$P[N(t) = r] = \sum_{n=0}^{t-2r} \left\{ (1-p_1)^r (1-p_2)^r \sum_{s=0}^{n} \left[ b_{r,s} b_{n,r,s} p_1^s p_2^{n-s} \right] \right\} - \sum_{n=0}^{t-2r(r+1)} \left\{ (1-p_1)^{r+1} (1-p_2)^{r+1} \sum_{s=0}^{n} \left[ b_{r+1,s} b_{n,r+1,s} p_1^s p_2^{n-s} \right] \right\} \quad (5)$$

for r = 0, 1, 2,... [(t/2)-1], where [(t/2)-1] is the largest integer not larger than (t/2)-1. In equation (5):

$$b_{r,s} = (r+s-1)!/[s!(r-1)!] \quad (6a)$$

$$b_{n,r,s} = (n+r-s-1)!/[(n-s)!(r-1)!] = b_{r,n-s} \quad (6b)$$

Equation (5) for the probability of the number of events N(t) in a period t based on geometric probability distributions as construed herein appears novel in the stochastic analysis of droughts. (Similar treatments have been presented using the negative binomial distribution Ross (1972) and the gamma distribution Parzen (1964) for the renewal times in problems other than hydrology).

From the previous relations between the counting process N(t) and the waiting time $W_r$, it follows that the cumulative probability of N(t) satisfies $P[N(t) \leq r] = P[W_{r+1} > t]$, or (see details in Appendix B):

$$P[N(t) \leq r] = 1 - \sum_{n=0}^{t-2(r+1)} \left\{ (1-p_1)^{r+1} (1-p_2)^{r+1} \sum_{s=0}^{n} \left[ b_{r+1,s} b_{n,r+1,s} p_1^s p_2^{n-s} \right] \right\} \quad (7)$$

for r = 1, 2, ..., [(t/2)-1].

The risk of drought, Π, is the probability of having at least one occurrence in t years, or $P[N(t) \geq 1]$, and, therefore:

$$\Pi = \sum_{n=2}^{t} \frac{(1-p_1)(1-p_2)}{(p_1-p_2)} (p_1^{n-1} - p_2^{n-1}) \quad (8)$$

where the relation $P[N(t) \geq 1] = P(R \leq t)$, was used in establishing equation (9), and R is the renewal time (see equation (3)). (It is noted that the renewal time R and the time to the first occurrence, $W_1$, are, by definition, identically distributed.)

### *4.2 The expected number of droughts*

The count of droughts in a period t, N(t), is nonnegative and integer valued, therefore, its expected value, E[N(t)], can be written in terms of its probabilities as $E[N(t)] = \sum P[N(t) \geq r]$ (see, e.g., Ross, 1985) where the sum is over $r \geq 1$. In addition, from the relation $N(t) \geq r$ if and only if $W_r \leq t$, it follows then that the expected number of droughts in a period t is given by:

$$E[N(t)] = \sum_{r=1}^{[t/2]} \sum_{n=2}^{t-2r} \left\{ (1-p_1)^r (1-p_2)^r \sum_{s=0}^{n} \left[ b_{r,s} b_{n,r,s} p_1^s p_2^{n-s} \right] \right\} \quad (9)$$

where [t/2] is the largest integer not larger than t/2.

Equations (5) through (9) summarize the theoretical results of posing drought recurrence as a renewal process with geometric distributions. These equations provide the probability distribution (equation (5)), the cumulative distribution function (equation (7)), the risk of drought (equation (8)), and the expected value of the number of droughts (equation (9)). All other distributional properties, such as the variance and higher moments, and, more generally, the characteristic function of the renewal process are derivable from the fundamental results presented previously. A peculiar aspect of the probabilistic results just derived in Eqs. (5) through (9) is that they do not involve parameter estimations. Model validation in this work is based on testing the goodness-of-fit of the geometric (Eqs. (1) and (2)), renewal (Eq. (3)) and counting (Eq. (5)) distribution, as shown below. A very important special ease, i.e., when a dry year is classified as below-median flow, is treated next

## 5 The case of below-median runs

In a number of previous studies (Loaiciga et al., 1992a; Loaiciga et al., 1993), it was demonstrated that runs of below-median annual streamflow can lead to hydrologically significant droughts. When the duration of dry runs is defined by below-median annual streamflow, the parameters $p_1$ and $p_2$ of the geometric distributions for dry and non-dry year runs (see equations (1) and (2), respectively) are equal, i.e., $p_1 = p_2 = p = 1/2$. As a result, simplifications occur in the probabilistic equations of the previous sections dealing with the general case of unequal geometric probabilities. The following results are shown in Appendix C.

### *5.1 The probability of the number of droughts*

The probability of the number of (below-median) droughts in a period t, P[N(t)], is given (where $p_1 = p_2 = p = 1/2$) by the following expression:

$$P[N(t) = r] = \sum_{n=0}^{t-2r} p^{2r+n} \binom{2r+n-1}{2r-1} - \sum_{n=0}^{t-2(r+1)} p^{2(r+1)+n} \binom{2r+n+1}{2r+1} \quad (10)$$

where r = 0, 1, ..., [(t/2)-1]. The terms in parentheses denote binomial coefficients (that is, in general, for integers n and r such that $0 \leq r \leq n$,

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

defines the binomial coefficient).

The cumulative probability for the number of below-median droughts in a period t is:

$$P[N(t) \leq r] = 1 - \sum_{n=0}^{t-2(r+1)} p^{2(r+1)+n} \binom{2r+n+1}{2r+1} \quad (11)$$

with r = 0, 1, ..., [(t/2)-1].

The risk of drought, $\Pi$, for the below-median case takes the simple form:

$$\Pi = 1 - p^t (t+1) \quad (12)$$

*5.2 The expected number of droughts*

The expected number of (below-median) droughts in a period t becomes:

$$E[N(t)] = \sum_{r=1}^{[t/2]} \sum_{n=0}^{t-2r} p^{2r+n} \binom{2r+n-1}{2r-1} \quad (13)$$

Equations (5)-(13) permit a complete characterization of drought recurrence for the cases when $p_1 \neq p_2$ and $p_1 = p_2$. In the former case, the calculation of probabilities, risk and expected values is computationally intensive. This is particularly true for large time periods t. Simplifications arise when $p_1 = p_2 = p$. The binomial coefficients appearing in equations (10) to (13) satisfy computationally convenient recursions. Let

$$a_n = p^{2r+n} \binom{2r+n-1}{2r-1} \quad (14)$$

and

$$b_n = p^{2r+n+2} \binom{2r+n+1}{2r+1} \quad (15)$$

The following recursions hold:

$$a_n = a_{n-1}\, p\, \frac{2r+n-1}{n} \quad (16)$$

$$b_n = b_{n-1}\, p\, \frac{2r+n+1}{n} \quad (17)$$

for n = 1, 2, ..., with $a_0 = p^{2r}$, and $b_0 = p^{2r+2}$.

Using the previous recursions, equation (10), the expression for the probability of below-median droughts, becomes:

$$P[N(t) = r] = \sum_{n=0}^{t-2r} a_n - \sum_{n=0}^{t-2(r+1)} b_n \quad (18)$$

The sums in equation (18) are then calculated via the recursions in (16) and (17). This greatly facilitates programming of the probability calculations, and avoids loss of accuracy by sidestepping the computation of large binomial coefficients that arise when the time period t is large.

## 6 Truncation of drought duration and other complicating matters

*6.1 Truncated drought duration*

It was previously discussed that in areas with water resources infrastructure, primarily reservoirs, there may be multiyear carryover water storage that helps mitigate the impacts of short runs of dry streamflow years. Loaiciga et al. (1993) established in a survey of drought studies that, typically, runs of three or more years of below-median stream flow almost unequivocally lead to some sort of drought impact. Considering then that runs of dry streamflow years are likely to trigger a hydrologic drought if their length exceed, say, $\theta$ years, it is still possible to model drought recurrence as a renewal process. In this instance, drought duration becomes a truncated random variable in the sense of (Loaiciga et al., 1992b). For the case of a geometric drought duration (see equation (1)), the distribution gets modified when the duration D is larger than $\theta$. The truncated geometric distribution is known to be (see Loaiciga et al., 1992b):

$$P[D = r] = (1 - p_1)\, p_1^{r-\theta-1}, \qquad r \geq \theta + 1 \tag{19}$$

The renewal time in this instance is the sum of D, the truncated drought duration (larger than $\theta$ years), plus the duration of the elapsed time, $\tau$, from the end of a drought to the initiation of the next one. The latter elapsed-time distribution must be identified from data, and wouldn't be a sample geometric distribution with parameter $p_2 = 1 - p_1$ as in equation (2), since it now includes runs of non-dry years (non-dry years occur with probability $p_2$ in any year) and runs of dry years that last $\theta$ or less years. Upon identification of the renewal time so defined, probability results analogous to those embodied in Eqs. (5)-(13) are obtainable by the methods of Appendices A C.

*6.2 Length of streamflow time series*

In order to make reliable probabilistic inferences on hydrologic droughts based on stream flow time series, the length of the data set and its accuracy are primordial. In the Santa Ynez river streamflow time series analyzed herein, in 430 years of data, there were 32, 6, and 2 below-median runs of durations $\geq$3, $\geq$5, and $\geq$6 years, respectively (see Figure 2). Clearly, the ability to conduct sound probability analyses is limited when the truncation threshold for drought duration increases, say, from 3 to 6 years, by the limited sample sizes available for long droughts. The renewal-theory approach of this paper, in conjunction with the empirical criterion discussed previously, namely, to consider runs of dry years lasting at least 3 years, seems to represent well drought incidence in regions with high hydroclimatic variability and recurrent droughts. The important Sacramento and Colorado river basins are examples in point (Loaiciga et al., 1993).

The renewal method of drought occurrence requires reasonably long hydrologic time series. A role of thumb might be that time series be no less than one hundred years for studies based on either annual streamflow or precipitation (Loaiciga et al., 1993). Annual precipitation typically shows less temporal correlation than annual streamflow (Michaelsen et al., 1987). Therefore, it is likely to see a larger proportion of short-duration runs of dry years in annual precipitation records than in streamflow records. The threshold for defining hydrologic drought can arguably be reduced to two or more - as opposed to three or more- consecutive dry years when working with annual precipitation. The implication here is that for annual streamflow and precipitation time series of equal length, the precipitation records are likely to be more representative in modeling short-duration droughts. If drought is triggered by relatively larger dry runs, the streamflow records are likely to be more suitable for drought modeling purposes.

It is desirable to pool instrumental records with well verified reconstructed proxy records, such as tree-ring based reconstructions, when available. It is then possible to develop centuries-long time series leading to highly reliable results by the methods presented herein. It is significant that in drought analysis based on binary classification of droughts (i.e., dry vis-a-vis non-dry conditions), as done in this work, reconstructed streamflow and precipitation (say, from tree-rings) can be very successful in identifying dry years. (Stockton and Meko, 1975; Michaelsen et al., 1988). Therefore, the predictive skill of those reconstructions can be very high when it comes to classifying streamflow and precipitation into dry and non-dry categories, i.e., up to or above 90% of years correctly classified in either category (Haston, 1992; Turner, 1992). This fortunate situation does not occur when one attempts to identify extreme high precipitation and streamflow based on tree rings. It is known that tree-rings cease growing after ambient humidity exceeds high levels, so the ability to discern the extremely wet events from tree rings is greatly impeded (Haston, 1992). Therefore, in tree-ring based precipitation and stream flow reconstructions, accuracy is greater in drought analysis than in the analysis of extremely wet events. For an extensive and recent account of dendrohydrologic applications to drought analysis see Loaiciga et al. (1993).

## 7 Results based on the renewal theory

A study of below-median drought recurrence was conducted based on a 430-year long reconstruction of annual streamflow for the Santa Ynez river in Santa Barbara County, California. This data set has been shown to have excellent predictive skill by Michaelsen and Haston (1988) and Turner (1992), providing an unusually representative time series for the purpose of model testing. The Santa Ynez river basin's hydroclimatology has been described in Upson and Thomasson (1951). Characterized by extreme climatic variability, this 2,000 $km^2$ river basin of Mediterranean-like climate has average annual precipitation of about 35.6 cm on the coast to 88 or 102 cm in the headwater mountains. The histograms of (unimpaired) below-median and above-median annual stream flow run durations at Bradbury Dam located at about 1/3 of the way of the stream course from its headwaters to the ocean) are shown in Figures 2. At this site, the drainage area is approximately 1050 $km^2$, with an annual median flow of 36.62 million cubic meters (29.7 thousand acre-feet per year).

### *7.1 Testing of geometric and renewal distributions*

The geometric distribution was proposed as a suitable distribution for modeling the duration of below-median and above-median run durations (see equations (1) and (2)). Figure 3 shows the observed (solid line) and expected (dashed line) frequencies of below-median run durations. The histogram of expected frequencies in Figure 3 was calculated by letting the frequency f be equal to f = m P, where m is the total number of observed runs (i.e., 105, as can be verified in the histogram of Figure 2) and P is the theoretical probability as given by equation (1), with $p_1 = p = 1/2$ for any duration $r = 1, 2, \ldots$. The chi-squared test (Loaiciga et al., 1992b) was conducted to assess the goodness of fit of the geometric model to the observed distribution of below-median run durations. Since we are testing for the goodness-of-fit of theoretical probability models to empirical frequencies, the chi-squared test is the most robust test for this purpose (Pearson, 1914; Rao, 1989). The chi-squared statistic was found to be equal to 8.89, well below the critical test statistic (with eight degrees of freedom) $\chi_{(0.05;8)} = 15.5$, indicating that the geometric distribution is a suitable model at a 5% significance level.

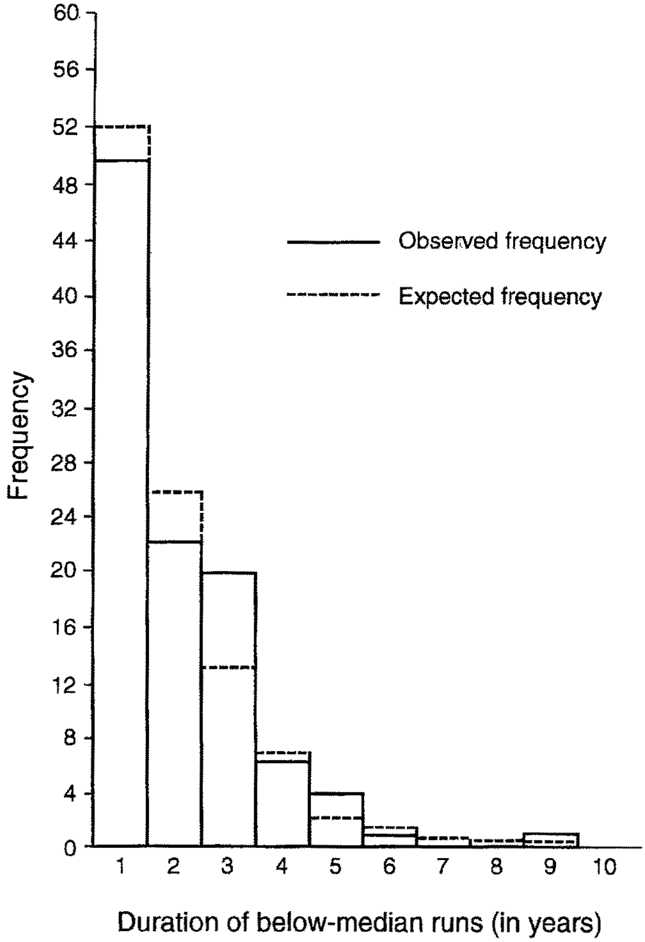


**Figure 3.** Results of chi-squared goodness-of-fit test for the distribution of the duration of below-median droughts. The theoretical geometric frequency fits the observed frequency of drought durations at a 5% significance level.

Figure 4 shows the observed (solid line) and expected (dashed line) frequencies of above-median run durations. Proceeding in an analogous manner as explained for the case of below-median runs, the chi-squared test yielded a chi-squared statistic of 2.80, below the critical test statistic (with seven degrees of freedom in this case) $\chi_{(0.05;7)} = 14.1$, indicating that the geometric distribution is a suitable model for above-median run durations at a significance level of 5%.

In Figure 5 are shown the observed (solid line) and expected (dashed line) frequencies of renewal time durations (see definition implied by Figure 1). There were $m = 106$ renewal time runs with durations ranging from $r = 2$ to $r = 11$. The distribution of expected frequencies in Figure 5 was calculated by letting the frequency f be $f = m\ P$, in which the probability P is given by the theoretical model of equation (3) applicable to renewal durations $r = 2, 3 \ldots$, after letting $p_1 = p_2 = p = 1/2$.

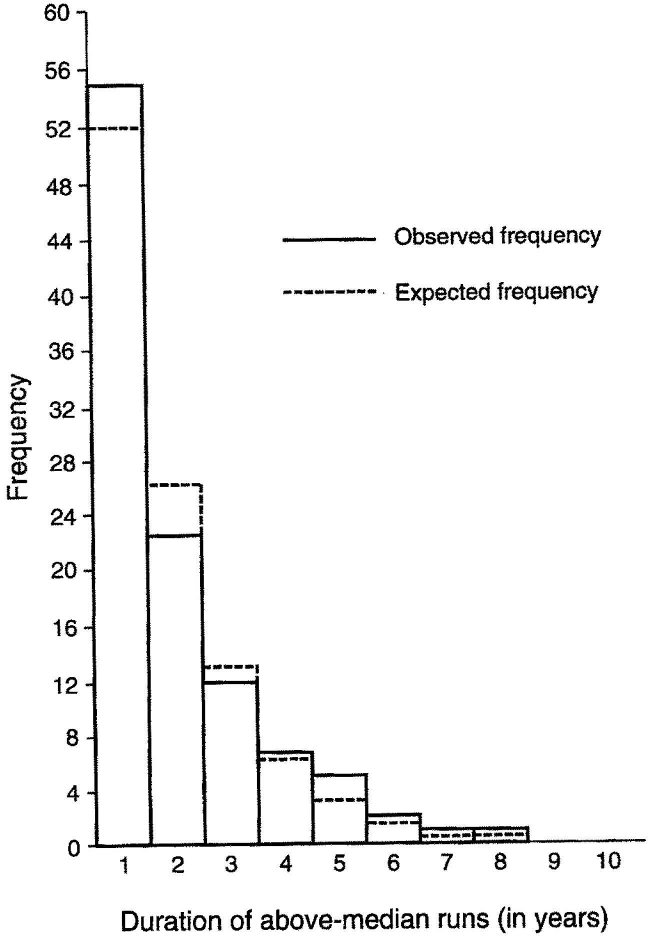


**Figure 4.** Results of chi-squared goodness-of-fit test for the distribution of the duration of above-median droughts. The theoretical geometric frequency fits the observed frequency of drought durations at a 5% significance level.

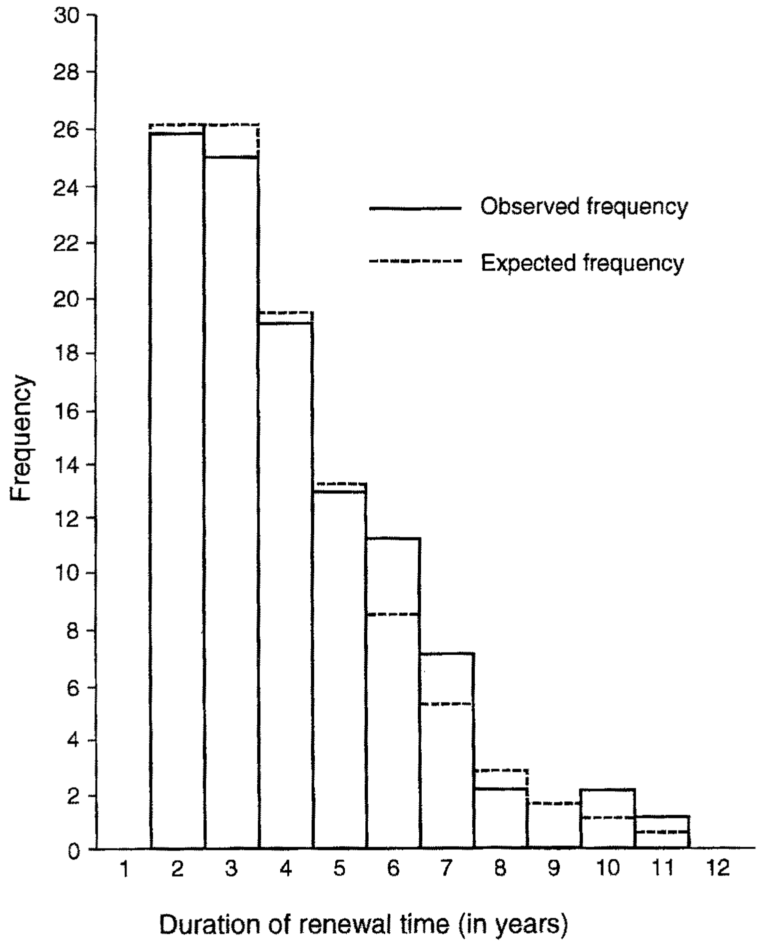


**Figure 5.** Results of chi-squared goodness-of-fit test for the distribution of the duration of the renewal time. The theoretical distribution of equation (3) (i.e., the observed frequency in the figure) fits the observed frequency of renewal times at the 5% significance level.

The chi-squared test for goodness of fit of the probability model of equation (3) to the observed renewal durations yielded a chi-squared statistic of 2.03, well below the critical test statistic (with six degrees of freedom in this case) $\chi_{(0.05;6)} = 12.7$. The theoretical probability of equation (3) is therefore adequate for observed renewal time durations at a 5% significance level.

The quality of the goodness of fit for below-median, above-median, and renewal time durations by the proposed (equations (1) and (2)) and derived (equation (3)) probability models is remarkable. Therefore, a high degree of confidence can be attached to the probability results of equations (5) to (13). Calculations related to these probability results follow.

### *7.2 Probability calculations*

Based on equation (10), the probabilities of observing a specified number of droughts in t = 430 years, P[N(430) = r] for r = 100 through r = 117 were calculated. This interval of renewal time durations account for approximately 91% of the mass of the probability of the number of droughts when the total observation time is 430 years. The 430 interval is relevant since it is the length of the Santa Ynez river streamflow time series. It was indicated previously that there were m = 106 observed renewal time runs in the Santa Ynez river time series. The calculated probability P [N(430)= 106] is approximately 7.5%, which is the third most likely number of droughts in 430 years, and, as shown in Figure 6, only exceeded by P[N(430) = 107] = 0.077 (this is the mode of the probability distribution) and P [N(430) = 108] = 0.0761.

### *7.3 Expected value calculations*

Equation (12) for the expected number of droughts in a period t = 430 was used to calculate that E[N(430)] = 107, in remarkable agreement with the distribution of Figure 6 where the mode is r = 107, and with the fact that the number of observed droughts was r = 106. In the same vein, it was previously established that the average renewal time $\mu$ in the Santa Ynez time series is approximately 430/105 = 4.09 years. This is in excellent agreement with the theoretical expected renewal time obtained from equation 0), with $p_1 = p_2 = 1/2$, that is, $\mu = 4$.

Expected values of the number of droughts in periods of other lengths, i.e., t = 100, 200, 500, 1000, were calculated via equation (13) to be equal to 25, 50, 125 and 250, respectively. This is in perfect agreement with the asymptotic result $N(t)/t \rightarrow \mu^{-1}$, where $\mu = 4$ for the case of below-median droughts. (Strictly speaking, this last statement implies a slight extension of the previous asymptotic statement to $m(t)/t \rightarrow \mu^{-1}$, where m(t), the expected number of droughts in a period t, replaces N(t), the number of droughts in that same period t).

## 8 Summary and conclusions

The theory of droughts as a renewal process has been developed in this paper. Our analysis was based on identifying the distribution of the renewal time, that is, the time elapsing from the initiation of a drought to the beginning of the next drought. In this study, that renewal time turned out to be the sum of two independent geometric variables. Once the renewal time distribution was derived, a number of important

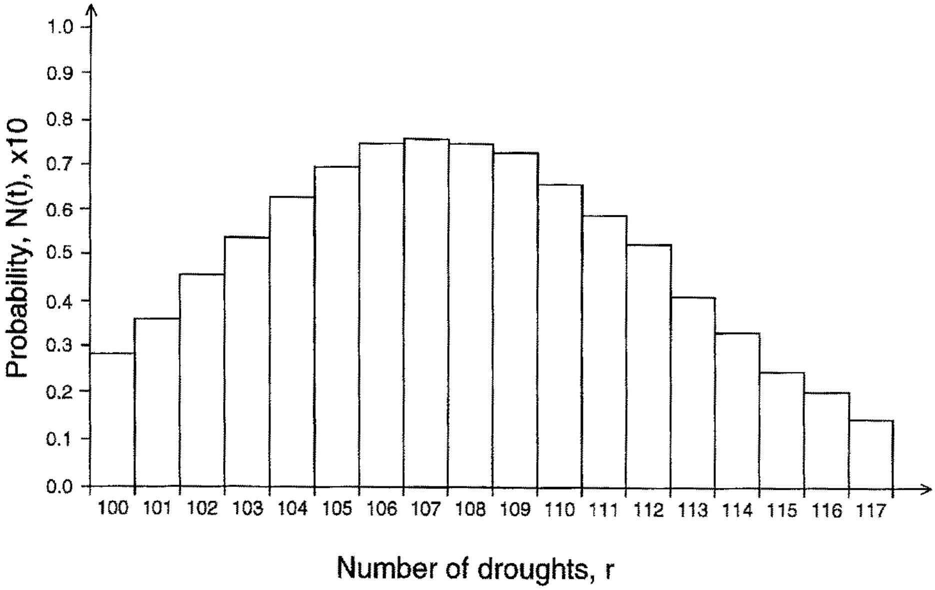


**Figure 6.** Calculated probabilities of the number of droughts in a period t = 430 years. The interval $100 < r < 117$ comprises about 91% of the probability mass. The observed number of droughts was 106, close to the theoretical mode of 107.

drought-related questions were solved by standard methods, such as the probability of having so many droughts in a time period, the risk of drought, the expected number of droughts per time period, and the like.

The theory presented here is not restricted to pairs of independent geometric distributions as building blocks for the renewal time. These can be in fact any pair of independent distributions for drought duration D and the intervening time $\tau$ (given by equations (1) and (2), respectively). Even though this study is based on discrete probabilities, the concepts and methods apply as well to continuous and mixed distributions.

A long and statistically representative streamflow time series was key to the proper identification of the renewal time's distribution. When drought is defined to have a threshold period of dry conditions, then one must work with a suitably truncated distribution for the drought duration. As restrictions on the basic drought definitions presented herein are added, e.g., thresholds, contemporaneous states of water storage, water demand and river flow, data needs become vexing and difficult to meet in practice. Also, the renewal model will no longer apply to these specially conditioned data, even if available. These man-made considerations will result in much less tractable models, although the conclusions would be more important to interested planners with such elaborations.

The methods of this paper were tested with a representative annual streamflow time series. Model testing basically consists of (i) verifying the basic distributions for drought duration D and intervening time $\tau$, (ii) verifying the derived distribution for the renewal time, and (iii) calculating expected values and probability of droughts and comparing those calculated values with observed ones and with asymptotic estimates (see, e.g., equation (4)) available from renewal theory. Chi-squared tests were implemented in this work to establish an excellent agreement between postulated and observed distributions. Probability calculations, using accurate and efficient recursive computational algorithms (see equations (14)-(18)), were remarkably close to the observed data values and theoretical asymptotic results.

In conclusion, a complete probabilistic characterization of drought recurrence is feasible with the methods of this paper. Our theory does not require estimation of parameters; instead, the basic geometric distributions and derived probability distribution must be verified against observed frequencies. The analyst simply specifies the streamflow threshold differentiating dry from non-dry years. For geometric drought durations, the case worked out here, closed-form expressions in terms of elementary functions yield all pertinent drought probabilities and expected drought values, which were easily calculated after adequate provision was given to ensure numerical accuracy.

## Acknowledgement

This research was supported by the California Water Resources Center through grant W-791 and by the National Science Foundation, grant NSF-SES-88-109-17 to the National Center for Geographic Information and Analysis, at the Department of Geography, University of California at Santa Barbara. Kate Rees and Rachid Labgaa, both from the Department of Geography at the University of California at Santa Barbara, assisted with the data analysis. Ken Turner, of the California Department of Water Resources, Sacramento Office, provided seminal inspiration in the initiation of this work.

## Appendix A

To show equation (3), one must realize that the characteristic function of a geometrically distributed variable with parameter $p_j$ is given by $\phi_j(\Psi) = e^{i\Psi}(1-p_j)/(1-p_j\, e^{i\Psi})$, where $i^2$ = -1. Therefore, the characteristic function of the sum of two independent geometric variables with parameters $p_1$ and $p_2$), $\phi_R(\Psi)$, is equal to the product $\phi_1(\Psi)\ \phi_2(\Psi)$:

$$\phi_R(\Psi) = \frac{(1-p_1)}{(\xi - p_1)}\frac{(1-p_2)}{(\xi - p_2)} \tag{A1}$$

where $\xi = e^{-i\Psi}$. The characteristic function in equation (A1) can be rewritten as a sum of partial fractions as follows:

$$\phi_R(\Psi) = \frac{(1-p_1)(1-p_2)}{(p_1-p_2)(\xi-p_1)} + \frac{(1-p_1)(1-p_2)}{(p_2-p_1)(\xi-p_2)} \tag{A2}$$

The denominators in the right-hand side of equation (A2) can be expanded as a geometric series on $p_j e^{i\Psi}$, $j = 1, 2$, and after factoring common terms, the characteristic function can be rewritten as:

$$\phi_R(\Psi) = \frac{(1-p_1)(1-p_2)}{(p_1-p_2)} \sum_{\nu=0}^{\infty} e^{i\Psi(\nu+1)}(p_1^{\nu} - p_2^{\nu}) \tag{A3}$$

By definition, the characteristic function $\phi_R$ is equal to:

$$\phi_R(\Psi) = \sum_{r=0}^{\infty} P(R=r)e^{i\Psi r} \tag{A4}$$

Letting $r = \nu + 1$ in equation (A3), and matching term by term the right-hand sides of equations (3) and (4) yields that:

$$P(R=r)\ \frac{(1-p_1)(1-p_2)}{(p_1-p_2)}\ \left[p_1^{r-1} - p_2^{r-1}\right] \tag{A5}$$

where $r = 2, 3$ ..... as proposed in equation (3), the expression to be proved.

## Appendix B

To prove equation (5), it suffices to derive the distribution of the time till the rth arrival, $W_r$, as it is clear from equation (4). It was previously stated that the time till the rth arrival is the sum of the renewal time, R, r times, i.e., $W_r = \sum_{j=1}^{r} R_j$. In turn, the renewal time is the sum of two independent geometric variables as seen in Appendix A. The characteristic function of $W_r$ is, therefore:

$$\phi_{W_r}(\Psi) = \frac{(1-p_1)^r}{(1-p_1e^{i\Psi})^r}\,\frac{(1-p_2)^r}{(1-p_2e^{\Psi})^r}\ e^{2ir\Psi} \tag{B1}$$

By differentiating $r-1$ times the geometric series $\sum_{k=0}^{\infty} \left[p_j\ e^{i\Psi}\right]^k = (1-p_j\ e^{i\Psi})^{-1}$, $j = 1, 2$, with respect to $p_j\ e^{i\Psi}$, it is readily established that:

$$\frac{1}{(1-p_1\ e^{i\Psi})^r} = \sum_{s=0}^{\infty} b_{r,s}p_1^s\ e^{i\Psi s} \tag{B2}$$

where $b_{r,s}$ was defined in equation (6a). Also:

$$\frac{1}{(1-p_2\ e^{i\Psi})^r} = \sum_{\nu=0}^{\infty} b_{r,\nu}p_2^{\nu}\ e^{i\Psi\nu} \tag{B3}$$

where $b_{r,\nu}$ is defined by equation (6a) with $s = \nu$. Combining equations (B1)-(B3), the characteristic function of $W_r$ becomes:

$$\phi_{W_r}(\Psi) = (1-p_1)^r(1-p_2)^r \sum_{s,\nu=0}^{\infty} b_{r,s}b_{r,\nu}p_1^s p_2^{\nu}\ e^{i(s+\nu)\Psi}\ e^{2ir\Psi} \tag{B4}$$

Using the well-known formula

$$\sum_{s,\nu=0}^{\infty} f_{s,\nu} = \sum_{n=0}^{\infty}\sum_{s=0}^{n} f_{s,n-s} \tag{B5}$$

for any argument f, the characteristic function $\phi_{W_r}(\Psi)$ becomes:

$$\phi_{W_r}(\Psi) = \sum_{n=0}^{\infty} \left[(1-p_1)^r(1-p_2)^r \sum_{s=0}^{n} b_{r,s}b_{r,n-s}\ p_1^s p_2^{n-s}\right]\ e^{i\Psi(n+2r)} \tag{B6}$$

in which $b_{r,n-s}$ was given in equation (6b).

The characteristic function $\phi_{W_r}(\Psi)$ is also, by definition, equal to:

$$\phi_{W_r}(\Psi) \;=\; \sum_{n=0}^{\infty} P[W_r = n + 2r] \; e^{i\Psi(n+2r)} \tag{B7}$$

Comparing equations (B6) and (B7) term by term within the summation from n = 0 to n = ∞, yields at once the probability for the distribution of the time till the rth arrival:

$$P[W_r \;=\; n + 2r] \;=\; (1 - p_1)^r(1 - p_2)^r \sum_{s=0}^{n} b_{r,s}b_{r,n-s} \; p_1^s p_2^{n-s} \tag{B8}$$

for $n \geq 0$; for $n < 0$ the probability is zero.

Equation (B8) is the fundamental probability, since, based on it and on the relations between the counting process N(t) and the waiting times $W_r$ and $W_{r+1}$ (see, e.g., equation (4)), construction of the probabilities in equations (5), (7), and (8), a s well as the expected value of equation (9), is direct.

## Appendix C

The probability results of equations (10)-(13) for the case when $p_1 \;=\; p_2 = p$, i.e., the classification criterion for dry years is that annual stream flow be below median, require essentially the probability of the waiting time $W_r$ for this special case. Once this probability is available, other results follow easily from previously established relations between the count process N(t) and the waiting time to the rth arrival.

The characteristic function of the time to the rth arrival for below-median droughts is a limiting case ($p_1 \to p_2 \;=\; p$) of the result in equation (B1), and is given by:

$$\phi_{W_r}(\Psi) \;=\; \frac{(1 - p)^{2r}}{(1 - p\; e^{i\Psi})^{2r}} \; e^{2ir\Psi} \tag{C1}$$

By taking 2r-1 derivatives with respect to $p\; e^{i\Psi}$ on both sides of the geometric series $(1\text{-}p\; e^{i\Psi})^{-1} \;=\; \sum_{k=0}^{\infty} [p\; e^{i\Psi}]^k$ it is established that:

$$\frac{1}{(1 - p\; e^{i\Psi})^{2r}} \;=\; \sum_{n=0}^{\infty} \binom{2r + n - 1}{2r - 1} \; p^n \; e^{i\Psi n} \tag{C2}$$

Combining equations (C1) and (C2) yields the characteristic equation of the time to the rth event:

$$\phi_{W_r}(\Psi) \;=\; \sum_{n=0}^{\infty} p^{2r+n} \binom{2r + n - 1}{2r - 1} \; e^{(2r+n)i\Psi} \tag{C3}$$

The characteristic function $\phi_{W_r}(\Psi)$ is, by definition, equal to:

$$\phi_{W_r}(\Psi) \;=\; \sum_{n=0}^{\infty} P[W_r = 2r + n] \; e^{(2r+n)i\Psi} \tag{C4}$$

Equating terms within the summations of equations (C3) and (C4) yields the probability of the rth arrival time:

$$P[W_r = 2r + n] = p^{2r+n} \binom{2r + n - 1}{2r - 1} \tag{C5}$$

for $n \geq 0$, and for $n < 0$ the probability is zero.

Equation (C5) is then the basis for deriving expressions (10)-(13) based on the relations among the arrival time and the count process N(t).

## References

Blackwell, D. 1953: Extension of a renewal theorem, Pacific Journal of Mathematics 3, 315-320

Blumenstock, D.L. 1942: Drought in the United States analyzed by means of the theory of probability, United States Dept. Agriculture Technical Bulletin 819

Bryant, E.A. 1991: Natural Hazards, Cambridge University Press, Cambridge

Feller, W. 1957: An Introduction to Probability Theory and Its Applications, 2nd. ed., John Wiley and Sons, New York

Haston, L. 1992: Reconstruction of Spatial and Temporal Patterns of Precipitation Variability in Southern California Using Tree-Ring Chronologies, Ph.D. Dissertation, University of California, Santa Barbara

Hurst, H.E. 1951: Long-term storage capacity of reservoirs, Trans. Am. Soc. Engrs. 116, 770-799

Klemes, V. 1974: The Hurst phenomenon: A puzzle?, Water Resour. Res. 10(4), 675-688

Loaiciga, H.A. 1988: On the use of chance constraints in reservoir design and operation modeling, Water Resour. Res. 24, 1221-1222

Loaiciga, H.A.; Michaelsen, J.; Haston, L.; Garver, S.; Leipnik, R.B. 1992a: Droughts in river basins of the western United States, Geophys. Res. Lett. 19, 2051-2054

Loaiciga, H.A.; Michaelsen, J.; Hudak, P.F. 1992b: Truncated distributions in hydrologic analysis, Water Resour. Bull. 28, 853-863

Loaiciga, H.A.; Haston, L.; Michaelsen, J. Dendrohydrology and long-term hydrologic phenomena, Rev. Geophys. 31(2), 150-171

Mandelbrot, B.; Wallis, J.R. 1969: Some long-run properties of geophysical records, Water Resour. Res. 5, 321-340

Marino, M.A.; Loaiciga, H.A. 1985: Dynamic model for multireservoir operation, Water Resour. Res. 21, 619-630

Michaelsen, J.; Haston, L.; Davis, F. 1987: 400 years of central California precipitation variability reconstructed from tree rings, Water Res. Bull. 23, 809-818

Michaelsen, J.; Haston, L. 1988: Reconstruction of the flows of the Santa Ynez river at Bradbury Dam, Santa Barbara County, California, 1537-1966, California Department of Water Resources Contract No. B56369, University of California, Santa Barbara, California

Palmer, W.C. 1965: Meteorological Drought, U.S. Weather Bureau Research Paper 45, U.S. Department of Commerce, Washington, D.C.

Parzen, E. 1964: Stochastic Processes, Holden-Day, Inc., San Francisco, California

Pearson, K. 1914: On the probability that two independent distributions of frequency are really samples of the same population, with special reference to recent work on the identity of Trypanosome strains, Biometrika 10, 85-154

Rao, C.R. 1989: Statistics and Truth, International Cooperative Publishing House, Fairland, Maryland, USA

Ross, S.M. 1985: Introduction to Probability Models, Academic Press, Orlando, Florida, 1972, 1st. ed., and 3rd. ed.

Stockton, C.W.; Meko, D.M. 1975: A long-term history of drought recurrence in western United States as inferred from tree rings, Weatherwise, 28, 245-249

Todorovic, P. 1978: Stochastic model of floods, Water Resour. Res. 14(2), 345-356

Turner, K.M. 1992: Reliability of storage schemes, tree rings, and Hurst phenomena, Proceedings of the American Geophysical Union Twelfth Annual Hydrology Days, March 31-April 3, Colorado State University, Fort Collins, Colorado, pp. 230242, Hydrology Days Publications, Atherton, California

Upson, J.E.; Thomasson, H.G. Jr. 1951: Geology and water resources of the Santa Ynez river basin, Santa Barbara County, California, U.S. Geological Survey Water Supply Paper 1107, U.S. Government Printing Office, Washington, D.C.
Yevjevich, V.M. 1964: Fluctuations of wet and dry years, Part II, analysis by serial correlation, Hydrology Paper 4, Colorado State University, Fort Collins, Colorado
Yevjevich, V.M. 1967: An objective approach to definitions and investigations of continental hydrologic droughts, Colorado State Univ. Hydrology Paper No. 23, Fort Collins, Colorado
Yevjevich, V.M. 1972: Stochastic Processes in Hydrology, Water Resources Publications, Fort Collins, Colorado
Zektser, I.S.; Loaiciga, H.A. 1993: Groundwater fluxes in the global hydrologic cycle, Journal of Hydrology 144, 405-427.