PB-223 943

A STOCHASTIC MODEL OF RUNOFF-PRODUCING RAINFALL FOR SUMMER TYPE STORMS

Lucien Duckstein, et al

Arizona University
Tucson, Arizona

26 October 1971

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

A Stochastic Model of Runoff-Producing Rainfall for Summer Type Storms

LUCIEN DUCKSTEIN, MARTIN M. FOGEL, AND CHESTER C. KISIEL

University of Arizona, Tucson, Arizona 85721

Abstract. Modification of watersheds occurs either through natural processes, such as erosion, or human influences, such as urbanization. In either case the rainfall input must be properly modeled before the runoff output can be predicted as the modifications take place. The paper considers runoff-producing summer precipitation of short duration and high spatial variability as an intermittent stochastic phenomenon. The probability distribution of seasonal total point or areal rainfall is obtained by convoluting a Poisson number of events with a geometric or negative binomial probability of rainfall amount. Close agreement with the experimental data is found. Next the probability of various combinations of rainfall amounts, given the seasonal total and the number of events, is computed. With these results, the theoretical seasonal water yield distribution can be obtained by using a simple rainfall-runoff relationship, such as the Soil Conservation Service formula. The possibility of using regional input parameters to study the distribution of the output of poorly gaged small watersheds is discussed. In particular, extreme total flows can be computed.

This paper concerns the runoff caused by summer type precipitation, namely, thunderstorms as found in a continental climate or local convective storms as found in the southwestern United States. This type of localized rainfall activity is pertinent to urban or small watersheds. Note that the exact meteorologic origin of the runoff-producing rainfall is of no concern here; anyway, to the best of our knowledge there is no accepted classification of thunderstorms. Eventually, this classification will be necessary in terms of atmospheric parameters, but it is seemingly impossible at present because of a lack of detailed physical data on storms. The definitions of a runoff-producing event given later are based on a substantial precipitation with a short duration and a high spatial variance; this we have called summer type precipitation for the sake of brevity.

The design of flood control structures, such as storm sewers, culverts, levees, or small dams, depends on correct identification of rainfall inputs. More precisely, in this paper, input to a given watershed will be defined by stochastic rainfall models; then runoff events can be generated by using known and simple rainfall-runoff relationships that have some empirical support [*Fogel and Duckstein*, 1970]. Such simple models are further justified by a foreknowledge of the mathematical difficulties connected with more involved stochastic transformations.

The stochastic rainfall models used in this study can be calibrated with 10–15 years of rainfall record; with as few as 5 years of data, useful rainfall estimates can be obtained [*Fogel et al.*, 1971].

PROCESS DESCRIPTION

Equispaced and event-based hydrologic models. For reasons of convenience or tradition, hydrologic data are usually taken at or averaged over equispaced time intervals. For example, maps of x year-y minute rainfall are given by meteorologists [*Hershfield*, 1961]; also average and peak daily, weekly, monthly, or yearly flows of rivers are traditional hydrologic quantities. Although such equispaced information constitutes useful data, especially for storms causing evenly distributed rainfall over space or perennial flows, it is of dubious value for summer type precipitation or intermittent flows. This limitation is indirectly demonstrated by *Hershfield* [1965, 1967], who could not find consistent isocorrelation contours in the Walnut Gulch Watershed, Arizona, with the concepts of either 2 year–1 hour or 2 year–2 hour rainfall. Acceptable results were obtained, however,

in climates where the summer precipitation plays a much less important role than it does in southern Arizona and where the sample sizes were larger. Another demonstration of the difficulty of using equispaced data is given by *Kisiel et al.* [1971] in the statistics of the Rillito Creek flow. The same reference contains an event-based analysis of the flow of the Tucson arroyo that provides a methodology for the present investigation and demonstrates the necessity of separating seasons.

Such an approach based on naturally occurring events is warranted whenever well-defined events, such as a point precipitation greater than 1 inch, occur relatively infrequently, so that their effects are usually separated by a time interval that contains the 'null event.' Most hydrologic phenomena in semiarid countries certainly satisfy this criterion, as flood occurrences in any climatic condition do [*Zelenhasic*, 1970]. For example, the considerable scatter of the correlation coefficients between x year–y minute point precipitation, on the one hand, and distance between gages, on the other hand, found by *Hershfield* [1967] in southern Arizona is due to the nature of convective storm events, which can easily hit one gage repeatedly and miss the next one. Possible definitions of summer rainfall events will now be given.

Definition 1. In the study of the convective storm precipitation over the dense rain gage network of the Atterbury Experimental Watershed near Tucson, Arizona, an event was defined as the occurrence of at least one storm center (point of maximum rainfall) over the 20 mi^2 of the watershed [*Fogel and Duckstein*, 1969].

Such a definition, necessitating a dense network of rain gages, includes values of rainfall that do not produce runoff. *Huff* [1967, 1968] suggests several definitions of runoff-producing storms based on several point rainfall measurements a few miles apart. *Fogel et al.* [1971] show that the distribution function characterizing point events is insensitive to the definition of an event, provided the definition reflects a substantial difference in measurements between gages. These precipitation events generally produce the peak flows on small watersheds. Furthermore, the records (in Chicago, for example) show that high values of point rainfall are not evenly spread over space.

Definition 2. Let us consider n gages, 1, 2, $\cdots$, n, with total precipitations, respectively, of $R_1, R_2, \cdots, R_n$ for a given day as given by the U.S. Weather Bureau records. By definition, an event is said to occur at any of the gages if the mean precipitation $(1/n)\ \Sigma R_j, j = 1, 2, \cdots, n$, is greater than 0.5 inch and one gage records more than 1 inch. Most urban areas have a sufficient number of gages to use this definition.

Event-based or intermittent process models are characterized by at least two random variables and their distribution function: (1) the random number N of events per unit time (season) including the interarrival time between events during the season and (2) the random variable(s) of interest in an event, e.g., the rainfall magnitude R, a time factor T, such as the maximum 15-minute rainfall intensity [*Fogel and Duckstein*, 1970], the runoff volume Q, the runoff duration [*Kisiel et al.*, 1971], the peak runoff, or the like; if more than one random variable is of interest, then a joint distribution is needed.

In this investigation, both discrete and continuous representation of the distribution function of certain random variables, such as rainfall R or total runoff Q, will be used. The use of the discrete form generally simplifies convolutions and computations; however, the continuous form is more convenient for transforming random variables. The symbols, conventions, and abbreviations to be used for any random variable V are defined in the notation list. The first task will be to determine the probability mass functions (pmf) $f_N(j)$ of N and $f_R(j)$ of R. The runoff Q will then be obtained by transforming the random variable R by means of the lumped rainfall–runoff relationship; the probability density function (pdf) of Q is denoted by $f_Q(y)$, $y > 0$. The pdf of the seasonal water yield will be calculated by adding a random number N of random runoff events.

RAINFALL MODEL

Number of events N per warm season. Thunderstorms seem to occur in an independent manner in time and space such that the number of rainfall events N per season follows a Poisson pmf [*Brooks and Carruthers*, 1953; *Todorovic and Yevjevich*, 1969]. This observation holds for a variety of conditions: for convective storm rainfall in Tucson, Arizona, when

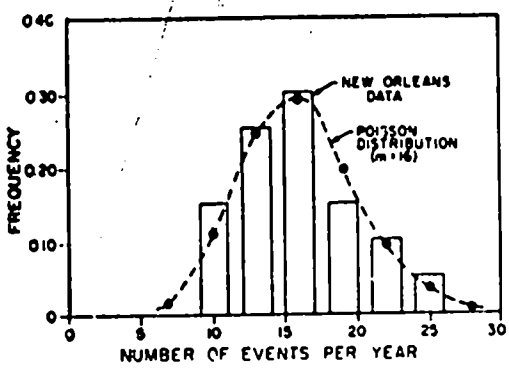

Fig. 1. Distribution of occurrences of warm season rainfall in which the areal mean of five gages exceeded 0.50 inch and at least one gage recorded more than 1.0 inch.

definition 1 is used [*Fogel and Duckstein*, 1969]; for summer type rainfall in Chicago when definition 2 is used [*Fogel et al.*, 1971]; and for summer type rainfall in New Orleans when definition 2 is used (Figure 1). Furthermore the Poisson pmf seems to hold for more complex runoff phenomena: for the number of summer runoff events in the Tucson arroyo, whose watershed is less than 50 km² [*Kisiel et al.*, 1971], and for the number of summer runoff events in the Rillito Creek, whose watershed is more than 2300 km² [*Baran et al.*, 1971].

We can thus write that N is distributed as

$$P(N = j) = f_N(j) = e^{-m}\frac{m^j}{j!} \qquad (1)$$

$$j = 0, 1, \cdots$$

where m is the mean number of events per season.

Of later use is the generating function of pmf (1),

$$F_N(s) = e^{-m+ms} \qquad (2)$$

Magnitude R of point rainfall. In earlier papers, it had been established by using Tucson [*Fogel and Duckstein*, 1969] and Chicago [*Fogel et al.*, 1971] data that a geometric pmf could not be rejected for point rainfall amount R:

$$f_R(j) = (1 - p)p^j \qquad j = 0, 1, 2, \cdots \qquad (3)$$

The generating function of pmf (3) is

$$F_R(s) = (1 - p)/(1 - ps) \qquad (4)$$

For the Tucson data, definition 1, which requires a dense rain gage network, was used, and a value of $p = 0.48$ was found as the probability that the point receives one or more units of rain. For the Chicago data, definition 2, which can be applied to National Weather Service records, was used with $n = 5$, and a value of $p = 0.35$ was found. New Orleans Weather Bureau data for five gages were then analyzed for warm season rainfall (May–September) under the conditions for the Chicago data. A value of $p = 0.40$ was found (Figure 2). Thus we see empirically from Figure 2 that the probability of point rainfall R, given the presence of rain over the area considered, is geometric on the basis of the heuristic argument that summer storms are not persistent. Recall that a geometric distribution arises from a sequence of independent Bernoulli trials. The parameter p (Figure 2) may also be characteristic of a region. There is, however, a need to evaluate the regional homogeneity of p in areas where dense rain gage networks exist, but the authors do not have access to this information.

Total seasonal precipitation. Let Z be the total number of units of rainfall during one summer season:

$$Z = R_1 + R_2 + \cdots + R_N \qquad (5)$$

where $R_1, R_2, \cdots, R_N$ are mutually independent, identically distributed random variables.

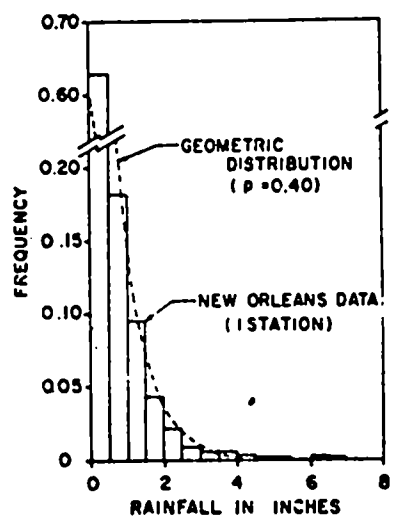

Fig. 2. Distribution of warm season rainfall depths per rainy day.

The generating function of Z is [*Feller*, 1957, chapter 12]

$$F_Z(s) = F_N[F_R(s)]$$

$$= \exp\left[-m + m\left(\frac{1-p}{1-ps}\right)\right] \qquad (6)$$

The pdf of Z is obtained from the generating function (6) by successive differentiations as follows:

$$f_Z(j) = \frac{1}{j!}\frac{d^j}{ds^j} F_Z(s)\Big|_{s=0} \qquad (7)$$

The mean or expectation $E(Z)$ and variance $\mathrm{var}(Z)$ can be calculated directly without determining the pmf (7):

$$E(Z) = E(N)E(R) \qquad (8)$$

$$\mathrm{var}(Z) = \mathrm{var}(N)[E(R)]^2 + [E(N)]^2\,\mathrm{var}(R)$$

(See *Benjamin and Cornell* [1970] for proof.)

For example, in the Atterbury precipitation model [*Fogel and Duckstein*, 1969], $E(N) = m = 5.33$ and $p = 0.48$, so that

$$E(N) = 5.33 \qquad \mathrm{var}(N) = 5.33$$

$$E(R) = 0.92 \qquad \mathrm{var}(R) = 1.77$$

$$E(Z) = 4.90 \qquad \mathrm{var}(Z) = 55.0$$

The pmf of Z when the above values of m and p are used is given in the second row of Table 1. The purpose of Table 1 is to generate possible combinations of amounts of point rainfall per event and the number of events during one season. The rows correspond to the number of events per year N; the columns correspond to the yearly total Z. Cell (j, k) in the table represents the joint probability of $j \geq 0$ units of rain and $k \geq 0$ events per year. For example, the probability that $Z = 7$ units of rain will occur in $N = 4$ events is

$$P(Z = 7, N = 4)$$

$$= P(Z = 7 \mid N = 4)P(N = 4)$$

$$= \binom{4+7-1}{7}(1-p)^4 p^7 f_N(4)$$

$$= 0.0513 \times 0.163$$

$$= 0.00835$$

Within each cell, different occupancy distributions of storms are possible. For example, when the goal is to estimate water yield in a season

TABLE 1. Generation of Seasonal Sets of Rainfall Events

k	j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	P(Z=j)	.077	.103	.118	.120	.113	.101	.085	.070	.055	.043	.032	.024	.018	.013	.009	.007	.005	.003	.002
	P(N=k)																			
0	0.005																			
1	0.026																			
2	0.069																			
3	0.122																			
4	0.162								.0083											
5	0.174																			
6	0.153																			
7	0.117																			
8	0.079																			
9	0.048																			
10	0.026																			
11	0.013																			
12	0.006																			

$\underline{Z}$ = total number of units of rain per season (in half inches)

$\underline{N}$ = number of rainfall events per season

$$P(\underline{Z} = j, \underline{N} = k) = P(\underline{Z} = j \mid \underline{N} = k)\,P(\underline{N} = k)$$

$$= \binom{k+j-1}{j}(1-p)^k p^j f_{\underline{N}}(k)$$

irrespective of the order in which the amounts occur, 7 units can be combined into 4 events in three possible ways: 4-1-1-1, 3-2-1-1, and 2-2-2-1. When classical combinatorial analysis techniques are used, the probability of each of these occupancies can be calculated [*Feller*, 1957, chapters 2, 3].

A Monte Carlo simulation can now be set up to generate an unordered succession of yearly combinations of events. This synthetic rainfall set may be useful, for example, in evaluating runoff modification practices and in studying separately the effects of random fluctuations and control. The order of the occurrence of the events may be important here, in which case it can be incorporated into the simulation, and a time series of events can then be generated. In any event, transforming the rainfall into runoff by using existing simple watershed models is necessary. This transformation is the subject of the next section.

TRANSFORMING RAINFALL INTO RUNOFF

Two models are considered to obtain the runoff volume Q from a rainfall event: a linear model with a random proportionality parameter and the Soil Conservation Service formula with constant parameters. The seasonal water yield is then calculated by using these results.

Linear model with random parameter. The following formula seems to yield good coefficients of determination (squared correlation coefficients) for small watersheds [*Fogel and Duckstein*, 1970]:

$$Q = C(R - A) \tag{9}$$

where A represents the initial abstractions depending on the watershed and C is a function of the rainfall characteristics for a given watershed, in particular, a time factor such as the maximum 15-minute intensity.

We shall assume on the basis of data presented in *Fogel and Duckstein* [1970] that C is a random variable with a gamma pdf

$$f_C(a, b; x) = \frac{1}{\Gamma(a)} b^a x^{a-1} e^{-bx} \qquad x > 0 \tag{10}$$

and that the rainfall amount R and the coefficient C are statistically independent.

These assumptions are not unreasonable in view of the earlier cited work [*Fogel and Duckstein*, 1970] where a linear relationship between C and a time factor T (the maximum 15-minute intensity) was postulated. Typical values found experimentally were $C = 0.3 + 0.072\ (T - 3)$.

The pdf of T may be taken as gamma with $E(T) = 3$ and $\mathrm{var}(T) = 0.8$. Thus $E(C) = 0.3 = a/b$ and $\mathrm{var}(C) = 0.005 = a/b^2$; hence $a = 18$ and $b = 60$.

Such high values of a and b yield a flat distribution of C. Truncation of pdf (10) for $C \geq 1$ may be necessary for accurate numerical calculations; however, from the relationship between C and T, we find that $P(C \geq 1) = P(T \geq 9.72)$. This probability is very small (less than 10^{-4}) and makes truncation unnecessary for practical purposes.

Since A is assumed to be a constant for a given watershed, we can define a shifted or effective rainfall as

$$\begin{aligned} P &= R - A \qquad & R > A \\ P &= 0 \qquad & \text{otherwise} \end{aligned} \tag{11}$$

The pmf of R is geometric (3); the continuous equivalent is an exponential pdf

$$f_R(x) = ue^{-ux} \qquad x > 0 \tag{12}$$

where u is approximately given by equating the means of pmf (3) and pdf (12): $1/u = p/(1 - p)$. When (11) and (12) are used, the distribution function of P is

$$\begin{aligned} \Phi_P(x) &= 1 - e^{-u(x+A)} \qquad & x \geq 0 \\ \Phi_P(x) &= 0 \qquad & x < 0 \end{aligned} \tag{13}$$

Note that the pdf of P, strictly speaking, does not exist, since it has a discontinuity (or atom) at the origin. Numerical approximations may be obtained, but they will not be useful in the sequel, and we shall deal with only distribution functions. Although the transformation of the random variables represented by (11) is nonlinear, the kth moment of P can be obtained without difficulty as follows:

$$\begin{aligned} E(P^k) &= \int_0^A (0) f_R(x)\, dx + \int_A^\infty (x - A)^k f_R(x)\, dx \\ &= \int_A^\infty u(x - A)^k e^{-ux}\, dx \\ &= e^{-uA} E(R^k) \end{aligned}$$

In particular, the mean and variance of P are found to be

$$E(P) = e^{-uA}/u \tag{14}$$

$$\text{var}(P) = E(P^2) - [E(P)]^2 = \frac{2e^{-uA} - e^{-2uA}}{u^2}$$

$$u = (1/p) - 1$$

Other moments of the distribution of P can be calculated in a similar manner. Note that the natural tendency to use 'mean values' of (11) would lead to $E(P) = E(R) - A$, which is an erroneous result.

We have thus obtained $Q = CP$. If we assume that C and P are independent, the mean and variance of Q can readily be calculated as follows:

$$E(Q) = E(C)E(P) \tag{15}$$

$$\text{var}(Q) = \text{var}(C)[E(P)]^2 + [E(C)]^2 \text{var}(P)$$

The latter equation explains why the coefficient of variation of the flow K_Q is higher than that of the precipitation K_P. From (10) and (14) the following relations are obtained:

$$K_C^2 = \frac{\text{var } C}{[E(C)]^2} = \frac{a}{b^2}\left(\frac{b}{a}\right)^2 = \frac{1}{a}$$

$$K_P^2 = \frac{\text{var } P}{[E(P)]^2} = 2e^{uA} - 1$$

so that

$$K_Q^2 = K_C^2 + K_P^2 = \left(\frac{1}{a} + 2\right)e^{uA} - 1$$

Because of the numerical value $a = 18$ found earlier, the difference between K_Q^2 and K_P^2 is less than 5%; however, other cases in which the contribution of K_C^2 is important may be found.

The distribution function $Q = CP$ is obtained by randomizing C [*Feller*, 1967, chapter 2]:

$$\Phi_Q(y) = \int_0^\infty \Phi_P\left(\frac{y}{x}\right) f_C(x)\, dx \tag{16}$$

Using (10) and (13) yields

$$\Phi_Q(y) = 1 - \frac{b^a}{\Gamma(a)} \int_0^\infty x^{a-1} e^{-u[(y/x)+A]-bx}\, dx \tag{17}$$

Numerical methods are needed to compute this function, whose usefulness depends on the accuracy obtainable on the parameters a, b, A, and u (or p).

Soil Conservation Service formula. An alternate method of obtaining the pdf of Q will be examined next. The Soil Conservation Service (SCS) has established an empirical rainfall-runoff relationship whose coefficients depend on the physical characteristics of the watershed [*Kent*, 1968]. This relationship can be written as

$$Q' = \frac{(R - A)^2}{(R - A) + S} \tag{18}$$

where A represents, as before, the initial abstractions and S is a watershed factor. Equation 18 does not account for the effect of rainfall intensity as (9) does; on the other hand it shows a nonlinear increase of Q with R.

The effective rainfall is the same as before (equation 11), so that (18) becomes

$$Q' = \frac{P^2}{P + S} \tag{19}$$

The distribution function of Q', denoted $\Phi_{Q'}(y)$, is obtained by a classical transformation of random variables [*Benjamin and Cornell*, 1970]. Let y represent Q' and x represent P; then

$$\Phi_{Q'}(y) = \Phi_P[x(y)] \tag{20}$$

$$y = \frac{x^2}{x + S} \quad \text{or} \quad x = \tfrac{1}{2}[y + (y^2 + 4Sy)^{1/2}]$$

Since $\Phi_P(x)$ is given by (13), we can rewrite (20) as

$$\Phi_{Q'}(y) = 1 - \exp\{-u[y + 2A + (y^2 + 4Sy)^{1/2}]\} \tag{21}$$

$$u = (1/p) - 1$$

As was true for (17), a numerical evaluation of (21) is possible, but it is warranted only if the parameters p, A, and S are known with sufficient accuracy. For purposes of illustration, the distribution function (21), which is conditioned on the occurrence of an event, is represented in Figure 3 for a given watershed with $A = 0.6$, $S = 3$, and different values of the rainfall parameter $p = 0.35$, $p = 0.40$, and $p = 0.45$.

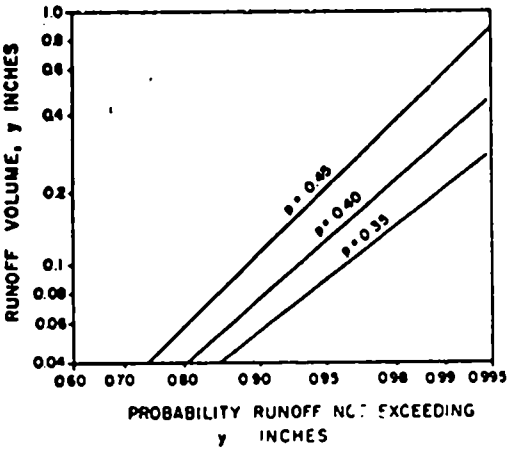

Fig. 3. Distribution function of storm runoff volumes per event using the SCS formula for a given watershed ($A = 0.6$ and $S = 30$).

Seasonal water yield from small watersheds. From the rainfall record, the point rainfall pmf (3) can be obtained; hence with A known the distribution function (13) can be evaluated.

A simple procedure to obtain the mean seasonal water yield W can be defined as follows:

1. Obtain a simulated set of point rainfalls $R_1, R_2, \cdots, R_N$.
2. Transform the NR_j into MP_j: $P_j = 0$ if $R_j \leq A$ and $P_j = R_j - A$ if $R_j > A$.
3. Use $Q_j = E(C)P_j$ or $Q_j' = P_j^2/(P_j + S)$ to obtain the set $Q_1, \cdots, Q_N$ or $Q_1', \cdots, Q_N'$. Then $W = Q_1 + \cdots + Q_N$ or $W = Q_1' + \cdots + Q_N'$. In this procedure, the moments of the yield could be calculated from the simulation. In the equations to follow, only the symbol Q will be used to denote runoff volume.

Another method of obtaining not only the mean and variance but also the pdf of the seasonal water yield consists of the following steps.

1. Obtain the pmf of the number of runoff-producing events M. Using definition 2, we have $f_M(j) = f_N(j)$ (equation 1), since definition 2 considers only runoff-producing events. Using definition 1, we must first evaluate the probability that any given rainfall produces runoff: $1 - \Phi_N(A) = e^{-\omega A}$. Then we obtain

$$f_M(k) = \sum_j \binom{j}{k} e^{-\omega A k}(1 - e^{-\omega A})^{j-k} f_N(j) \qquad (22)$$

This formula is derived under the assumptions that each rainfall is equivalent to a Bernoulli trial with probability $e^{-\omega A}$ of success (runoff) and that there is a random number j of rainfall events per season. Precipitation is measured at a single rain gage that is taken as representative of the area.

2. The seasonal water yield W is the sum of M identically distributed, mutually independent variables Q_j:

$$W = Q_1 + Q_2 + \cdots + Q_M$$

Hence the mean and variance of W as given by (8) are

$$E(W) = E(M)E(Q) \qquad (23)$$

$$\text{var}(W) = \text{var}(M)[E(Q)^2] + [E(M)]^2 \text{var}(Q)$$

3. If we use the linear rainfall-runoff relationship, then (14), which gave $E(Q)$ and $\text{var}(Q)$, can be substituted into (23). Thus we have an explicit expression for the mean and variance of the seasonal water yield. The Fourier transform of the distribution of W (also called the characteristic function) is

$$F_W(\omega) = F_M[F_Q(\omega)] \qquad (24)$$

where $F_Q(\omega)$ is the characteristic function of Q obtained from (17).

4. If the SCS formula (18) for rainfall-runoff is used, the mean and variance of Q' given by (21) can be substituted into (23) to obtain the explicit expression of $E(W)$ and $\text{var}(W)$. The characteristic function of W is then

$$G_W(\omega) = F_M[G_Q(\omega)] \qquad (25)$$

where $G_Q(\omega)$ is the characteristic function of Q' obtained from (21). The moments of W can be obtained by taking the successive derivatives of (25) evaluated at $\omega = 0$.

The above considerations concern situations in which point rainfall measurements at only one station are quite representative of the areal pattern; they would not be applicable to the larger watersheds, for which runoff must be calculated from areal rainfall. In the next sections, generalization of the proposed model to areal rainfall estimated from point measurements, the problem of estimating runoff in

poorly gaged watersheds, and the forecast of extreme flow events using point rainfall measurements are discussed.

REGIONALIZATION

Extension to area models. The areal rainfall B corresponding to a given event may be considered as the average of a fixed number of n point rainfall measurements $R_1, R_2, \cdots, R_n$. The pmf of the sum ΣR_k and of the average $(1/n) \Sigma R_k$, $k = 1, \cdots, n$, differ only by a constant scaling factor, so that we shall deal with the sum, which has a simpler expression than the average. If the point rainfall variates $R_1, \cdots, R_n$ are mutually independent and have an identical geometric pmf (equations 3 and 4), then $B = \Sigma R_k$ has a negative binomial pmf with parameters p and $r = n$:

$$f_B(j) = \binom{r+j-1}{j}(1-p)^r p^j \tag{26}$$

The generating function of this pmf is

$$F_B(s) = \left[\left(\frac{1-p}{1-ps}\right)\right]^r \tag{27}$$

The validity of (26) was demonstrated for Tucson and Chicago data by *Fogel et al.* [1971]. Although the fit at low values was not outstanding, the tail did fit well.

If the gage measurements are correlated, then the parameter r is less than the number of gages. At the limit, if a uniform rainfall occurs over the n gages, the result reduces to the information given by a single point rainfall; hence $r = 1$. In other words

$$1 \leq r \leq n \tag{28}$$

To find the parameter r for the New Orleans data, the method of moments was used, and a value $r = 2$ was found for $n = 5$ gages. Note that five gages were available for the area under consideration and these gave reasonable estimates of the mean areal rainfall. As Figure 4 illustrates, the theoretical and empirical distributions are closer together for the New Orleans data than for the Chicago data. The null hypothesis that the mean rainfall was negative binomial ($p = 0.40$, $r = 2$) could not be rejected at any level of significance by using a Kolmogorov-Smirnov test for goodness of fit. All the calculations presented using point rainfall R can be repeated step by step by using areal rainfall B, defined by (26) and (27). For simulation purposes, a table of rainfall combinations similar to Table 1 can be used to generate seasonal sets of areal rainfall, which are transformed into sets of runoff by using the linear formula (9) or the SCS formula (18). Analytically, the mean and variance of Q are given by formulas similar to (15). If B is the runoff-producing area rainfall, then we have

$$E(Q) = E(C)E(B) \tag{29}$$

$$\text{var}(Q) = \text{var}(C)[E(B)]^2 + [E(C)]^2 \text{var}(B) + \text{var}(C)\text{var}(B)$$

The seasonal water yield W has a mean and variance given by (23) when $E(Q)$ and var(Q) are calculated by (29). As we have seen above, areal rainfall B requires a point parameter p and a spatial parameter r. If several gages are present within a 250–400 km² watershed, as is the case for Tucson, Chicago, and New Orleans, p and r can be determined from the record. The estimation problem in the absence of such data is discussed in the next section.

Poorly gaged watersheds. Assume we have a 250-km² watershed with a short record of summer rainfall, say 5 years, at a single gage. How can the pdf of the areal rainfall B be estimated? The following steps are suggested.

1. Obtain a prior value of p from regional considerations. Thus around Chicago or New Orleans, when definition 2 of an event is used, $p = 0.35$ or $p = 0.40$, respectively.
2. Improve the estimate of p using the 5 years of single point data over the watershed by using, for example, Bayes' formula [*De Groot*, 1970, p. 55; *McGilchrist et al.*, 1970].
3. Obtain a prior value of r based on the size of the watershed. When definition 1 of an event is used, r is approximately equal to the mean number of cells centered over the watershed per event (1.23 for 50 km² near Tucson). When definition 2 is used, r is approximately equal to the mean number of independent runoff-producing storm events over the area ($r = 2$ or $r = 3$ for 150 km² near Chicago or New Orleans). If a uniform distribution of thunderstorm cells is assumed, r can be expected to vary linearily with the watershed area; then for 250 km² a value of r between 4 and 5 seems reasonable. The data that have been examined to date do not warrant further extrapolation.

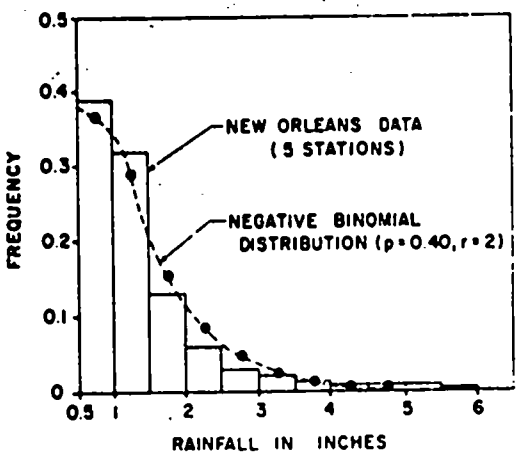

Fig. 4. Distribution of areal mean depth of warm season rainfall per event.

4. Improve the estimate of r by using an areal observation method, if feasible, to count the number of cells producing precipitation over a certain area at a given time.

5. Quantify the uncertainty on the pdf f_B of B (equation 26) as a function of the uncertainty on the parameters p and r. Note that the mean and variance characterize completely the pmf of B (26), since we are dealing with a two-parameter distribution.

Further research is necessary to validate and eventually to implement such a method. Once the areal rainfall distribution is determined, runoff can be calculated by one of the methods presented earlier.

Extreme events. Let the areal rainfall distribution function be given as $\Phi_B(k)$. For runoff-producing rainfall (definition 2), by using the pmf of B (26), we obtain

$$\Phi_B(k) = \sum_j \binom{r+j-1}{j}(1-p)^r p^j \qquad (30)$$

$$j = 0, 1, \cdots, k$$

Two methods can then be used to obtain the distribution function of the maximum seasonal runoff $\Phi_Q^*(k)$ called the maximum (or extreme) distribution.

The first method consists of using the Monte Carlo simulation to generate seasonal areal rainfall sets from which the runoff values are computed. The relative cumulative frequency of yearly floods of magnitude k or less is, by definition, $\Phi_Q^*(k)$.

The second method is analytic and uses the pmf $f_M(j)$ of the yearly number of runoff-producing events M. Either $M = N$ or (22) gives f_M for point rainfall. For areal rainfall, f_M is obtained as the pmf of a random number of events over the watershed; that is, f_M is a compound Poisson pmf if the independence assumption holds in both space and time. At any rate, the distribution function of the seasonal maximum areal rainfall per event is given by the equation

$$\Phi_B^*(k) = \sum_j [\Phi_B(k)]^j f_M(j) \qquad (31)$$

$$j = 0, 1, 2, \cdots$$

If we assume that the pmf of M is Poisson (1) with mean m, then (31) can be written as

$$\Phi_B^*(k) = \exp - \{m[1 - \Phi_B(k)]\} \qquad (32)$$

Distribution function (32) is plotted in Figure 5 for $m = 10$; $p = 0.35$, 0.40, and 0.45; and $r = 3$. As can be expected, the probability that the areal rainfall reaches a certain magnitude increases with r, since r represents the number of independent point events that may add up to cause runoff. The maximum distribution function of seasonal runoff $\Phi_Q^*(k)$ can be obtained from (31) by using a rainfall-runoff relationship such as (9) or (19).

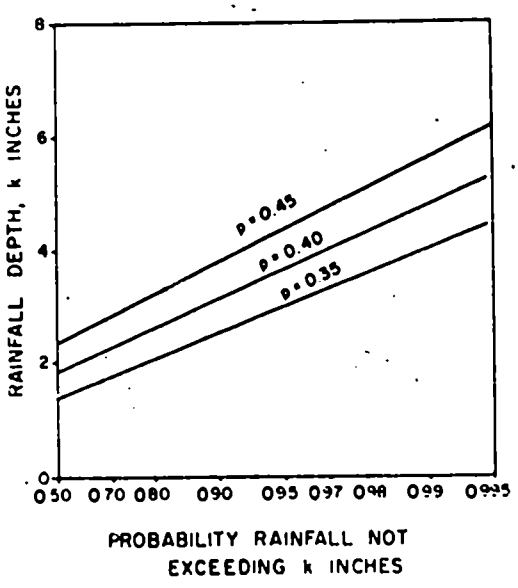

Fig. 5. Maximum distribution function of areal mean rainfall for $r = 3$ and $m = 10$.

One of the problems that may develop with this procedure concerns the reproduction of the extreme or rare event. Parameter estimation is part of the problem, but so is the selection of the distribution function itself. Many of the hydrologic processes have exceedingly long tails that are difficult to characterize with the more commonly used distribution function. A closer look at this aspect is currently being taken. Possible techniques to be investigated will involve the use of a combination of distributions and parameter randomization.

DISCUSSION AND CONCLUSIONS

The proposed models have several shortcomings. For example, the influence of antecedent moisture conditions on runoff has not been considered. For adequate consideration, a time series of rainfall events would have to be simulated, and such a simulation would require many more data than are available. Independence assumptions have been made between events, between the number and magnitude of events, and within an event (independence of time factors and total rainfall). Distribution functions have been hypothesized. The effects of parameter uncertainty on the results have not been thoroughly assessed. To assess these effects thoroughly, a managerial goal and the notion of economic risk must be defined, when the effect of parameter uncertainty can be ascertained using Bayesian decision theory [*Davis*, 1971].

Another restrictive aspect of this paper is that only summer rainfall has been considered. Winter precipitation exhibits different characteristics, such as lower intensity, longer duration, more uniform areal distribution, and persistence from one event to another. Ultimately, if and when validated models of summer and winter runoff become available, the yearly yield of water can be obtained by adding the two corresponding random variables [*Kisiel et al.*, 1971].

In conclusion, we have demonstrated the following points.

1. The description of thunderstorm type rainfall as an intermittent process is advantageous because of the temporarily independent, infrequent, irregularly spaced occurrences of precipitation.
2. Rainfall-runoff models can be based on either of two definitions of an event, one conditioned on an area, the other conditioned on adjacent point measurements.
3. Whether the location is Tucson, Chicago, or New Orleans and whether definition 1 or definition 2 is used, the following distribution functions cannot be rejected for summer rainfall: (*a*) Poisson, for the number of events per season; (*b*) geometric, for the rainfall quantity at a point; and (*c*) negative binomial, for the areal rainfall. As a result, at most three parameters are necessary for the analysis.
4. Monte Carlo simulation can be set up to obtain seasonal sets of rainfall events.
5. The distribution function for runoff volumes is obtained by randomizing the multiplicative constant in a linear rainfall-runoff relationship to account for a randomly varying time factor of the hyetograph. An alternate method is a transformation of random variables using the Soil Conservation Service formula.
6. The total seasonal water yield can be computed either by analytical methods or Monte Carlo simulation.
7. Areal rainfall models have been obtained. For poorly gaged watersheds, the parameter estimation is facilitated by regional considerations and the use of Bayes' formula.

The approach taken in this paper is based on empirical and heuristic considerations. Given the stochastic hypotheses about rainfall and runoff patterns, the use of probability theory leads to analytical solutions for the probability distributions of seasonal rainfall totals and areal rainfall. Such solutions have advantages over simulated probability distributions to the extent that underlying hypotheses are acceptable to the user and are consistent with conditions in the geographic area of application.

NOTATION

a,	parameter of gamma distribution;
A,	initial abstraction, inches;
b,	parameter of gamma distribution;
$E(\ )$,	expected value of ( );
$f_V(\ ), g_V(\ )$,	probability of mass (or density) function of discrete (or continuous) random variable V;
$F_V(\ ), G_V(\ )$,	generating (or characteristic) function of discrete (or continuous) random variable V;

j, discrete index, dummy variable in the physical domain;
k, discrete index, dummy variable in the physical domain;
K_V, coefficient of variation of V;
m, average number of events per season;
n, number of rain gages over area considered;
p, parameter of the geometric distribution;
r, parameter of negative binomial distribution;
s, dummy variable in the generation function transform domain corresponding to discrete functions of j or k, $0 \leq s \leq 1$;
S, constant in Soil Conservation Service rainfall-runoff formula;
ν, mean interarrival time between events;
x, continuous dummy variable in the physical domain;
y, continuous dummy variable in the physical domain;
$\Phi_V(\ )$, distribution function of V;
Φ_V^*, maximum (or extreme) distribution function of V;
ω, dummy variable in characteristic function, in the Fourier transform domain corresponding to the continuous function of x or y, $\omega > 0$;
$\Gamma(\ )$, gamma function;

Random variables:

B, areal rainfall;
C, rainfall-runoff proportionality factor;
M, number of runoff-producing rainfalls per season;
N, number of rainfall events per season;
P, total effective rainfall per event, inches ($P = R - A$);
Q, Q', total runoff per event, inches;
R, total rainfall per event, inches;
$R(i)$, precipitation on day i,
R_j, precipitation at point j;
T, time factor of hydrograph;
V, random variable;
W, seasonal water yield, inches;
Z, total number of units of precipitation per season.

Acknowledgments. The research presented in this paper was partially supported by allotment grant A-020-Ariz from the Water Resource Research Center of the University of Arizona. The help of Jerry Sanders and Marshall Moss in clarifying and correcting certain theoretical points and of Sam Kao, Linda Dahlke, and John Waits in setting up and performing numerical calculations is gratefully acknowledged. Some of the material in this paper was presented at the 52nd Annual Meeting of the American Geophysical Union in Washington, D. C., April 1971.

REFERENCES

Baran, N. E., C. C. Kisiel, and L. Duckstein, A stochastic analysis of flows of Rillito Creek, *Hydrol. Water Resour. Ariz. Southwest, 1*, 171–210, 1971.

Benjamin, J. R., and C. A. Cornell, *Probability, Statistics, and Decision for Civil Engineers*, p. 109, McGraw-Hill, New York, 1970.

Brooks, C. E. P., and N. Carruthers, *Handbook of Statistical Methods in Meteorology*, Her Majesty's Stationery Office, London, 1953.

Davis, D. R., Decision making under uncertainty in systems hydrology, *Rep. 2*, Hydrol. and Water Resour. Interdisciplinary Program, Dep. of Hydrol. and Dep. of Syst. Eng., Univ. of Ariz., Tucson, 1971.

De Groot, M. A., *Optimal Statistical Decision*, p. 55, McGraw-Hill, New York, 1970.

Feller, W., *An Introduction to Probability Theory and Its Applications*, vol. 1, John Wiley, New York, 1957.

Feller, W., *An Introduction to Probability Theory and Its Applications*, vol. 2, John Wiley, New York, 1967.

Fogel, M. M., and L. Duckstein, Point rainfall frequencies in convective storms, *Water Resour. Res.*, *5*(6), 1229–1237, 1969.

Fogel, M. M., and L. Duckstein, Prediction of convective storm runoff in semi-arid regions, *Int. Assoc. Sci. Hydrol. Publ. 96*, 465–478, 1970.

Fogel, M. M., L. Duckstein, and C. C. Kisiel, Space-time validation of a thunderstorm rainfall, *Water Resour. Bull.*, *7*(2), 309–316, 1971.

Hershfield, D. M., Rainfall frequency atlas for the United States for durations from 30 minutes to 24 hours and return periods from 1 to 100 years, *Tech. Pap. 40*, 115 pp., Weather Bureau, U.S. Dep. of Commer., 1961.

Hershfield, D. M., On the spacing of rain gages, *Proc. Int. Assoc. Sci. Hydrol.*, *67*, 72–79, 1965.

Hershfield, D. M., Rainfall input for hydrologic models, *Trans. Int. Union Geod. Geophys.*, 177–188, 1967.

Huff, F. A., Time distribution of rainfall in heavy storms, *Water Resour. Res.*, *3*(4), 1007–1019, 1967.

Huff, F. A., Spatial distribution of heavy storm rainfalls in Illinois, *Water Resour. Res.*, *4*(1), 47–54, 1968.

Kent, K. M., A method for estimating volume and rate of runoff in small watersheds, *Soil Conserv. Serv. TP-149*, U.S. Dep. of Agr., Soil Conserv. Serv., 1968.

Kisiel, C. C., L. Duckstein, and M. M. Fogel, Stochastic analysis of ephemeral flow in arid-

lands, *J. Hydraul. Div., Amer. Soc. Civil Eng.*, *97*(HY10), 1699–1717, 1971.

McGilchrist, C. A., T. G. Chapman, and K. D. Woodyer, Recurrence intervals between exceedances of selected river levels, 3, Estimation and use of a prior distribution, *Water Resour. Res.*, *6*(2), 499–504, 1970.

Todorovic, P., and V. Yevjevich, Stochastic processes of precipitation, *Hydrol. Pap. 35*, Colo. State Univ., Fort Collins, 1969.

Zelenhasic, E., Theoretical probability distribution for flood peaks, *Hydrol. Pap. 42*, Colo. State Univ., Fort Collins, 1970.

(Manuscript received July 27, 1971;
revised October 26, 1971.)