

A Stochastic Model of Dimensionless Thunderstorm Rainfall

D. A. WOOLHISER AND H. B. OSBORN

U.S. Department of Agriculture, Agricultural Research Service, Tucson, Arizona

A concise stochastic model for the nondimensional thunderstorm rainfall process at a point is proposed. The accumulated precipitation process for individual thunderstorms is nondimensionalized by dividing the precipitation at any time by the total precipitation and the elapsed time by the total duration. The dimensionless process is divided into 100 equal time increments, and the depth increments are rescaled to range between 0 and 1. The sequence of rescaled increments Z_1, Z_2, \dots, Z_n are assumed to represent a nonhomogeneous Markov process in discrete time with continuous state space. The expected value of the k th rescaled increment, given the $k-1$ st increment, is assumed to be a linear function of that increment, and the marginal distribution of the first increment and the conditional distributions are assumed to be described by the beta distribution. An analyses of data for 275 thunderstorms observed at the Walnut Gulch Experimental Watershed in southeastern Arizona showed that the proposed model structure is a good approximation for this region. The number of model parameters can be reduced from 26 to a minimum of 10 by approximating the 2 parameters in the conditional expectation function and the conditional beta parameter as polynomial functions of the dimensionless time. Likelihood ratio tests and the Akaike information criterion suggest that the dependence parameters are independent of storm amount and duration, but the conditional beta parameter α_k is larger for short-duration storms than for long-duration storms. A 13-parameter model is recommended for disaggregating thunderstorm rainfall in southeastern Arizona.

INTRODUCTION

Greater emphasis is being placed on the use of physically based infiltration models to estimate direct surface runoff, and techniques have been developed to estimate infiltration model parameters from soil texture and other factors [Brakensiek *et al.*, 1981; Rawls *et al.*, 1983]. Such models are quite sensitive to the distribution of total storm rainfall within time increments as short as 5 min. Where recording rain gage data are available in computer compatible form, they can be used to provide input to the infiltration models. Scarcity of such data and the difficulty and expense involved in working with "breakpoint" rainfall data have prevented the widespread adoption of physically based infiltration models by professionals in resource management agencies. For example, CREAMS (field scale model for chemicals, runoff, and erosion from agricultural management systems) [Knisel, 1980], a model developed by the Agricultural Research Service, U.S. Department of Agriculture to simulate the transport of chemicals, water, and sediment from field-size areas, has two options. The first, which is used when only daily rainfall data are available, uses the soil conservation service curve number procedure. The second option requires short-period rainfall data and uses the Green and Ampt [1911] infiltration relation. Although the second option allows a significantly improved prediction of infiltration and runoff [Smith and Williams, 1980], its practical use has been limited, primarily because of the shortage of short-period rainfall data.

Recent work [Richardson and Wright, 1984] promises that very good simulation models for daily rainfall will soon be readily available which could be used effectively on microcomputers. This development suggests that we consider the following questions.

1. Can parameter efficient techniques be developed for disaggregating daily rainfall into the intermittent rainfall process within the day?
2. Are the structure and parameters of these disaggregation models (relatively) spatially invariant so that the more abundant daily precipitation data can be used to describe the important geographical variability of the daily process while fewer stations can be used to identify the precipitation process within a day?

The following sequence of research tasks must be completed before these questions can be answered.

1. Techniques must be developed for disaggregating daily rainfall amounts into individual "storm" amounts, durations, and time of occurrence.

2. Techniques must be developed to disaggregate the amounts of significant storm rainfalls into shorter period rainfall amounts.

3. The seasonal and spatial variability of model structure and parameters must be examined for both the daily and storm disaggregation models.

If this approach proves to be feasible, disaggregation parameters must be estimated for a sufficiently large number of stations so that the model is useful in practice.

Several investigators have developed stochastic models of storm rainfall at a single point [Pattison, 1965; Grace and Eagleson, 1966; Raudkivi and Lawgun, 1973; Knisel and Snyder, 1975; Nguyen and Rousselle, 1981] or at a network of stations [Sinha and Khanal, 1971; Sorman and Wallace, 1972; Wilkinson and Valaderes-Tavares, 1972; Bras and Rodriguez-Iturbe, 1976; Corotis, 1976; Amorcho and Wu, 1977]. A review of this work reveals that the models either were not designed to accommodate intervals less than an hour or they require a large number of empirically determined parameters that make them difficult to utilize in other locations. Austin and Claborn [1974] developed a model for short-duration storms which included a method for distributing the rainfall depth during the storm. Their method appears to be somewhat oversimplified in that it ignores the serial correlation between 4 min rainfall intensities and assumes that the 4-min intensities are described by the same distribution, irrespective of position within the storm. Woolhiser *et al.* [1984] and Her-shenhorn [1984] have made progress in disaggregating daily rainfall into storm rainfall for an Arizona station (the first research task).

This paper is not subject to U.S. copyright. Published in 1985 by the American Geophysical Union.

Paper number 4W1545.

In this paper we address the second research task listed above. Our objective is to describe a stochastic model that can be used to disaggregate storm rainfall at a point in space and to determine if the process parameters are independent of storm amount and duration for thunderstorm rainfall in southeastern Arizona.

APPROACH (NOTATION AND DEFINITIONS)

The basic process under consideration is precipitation intensity at a given point in space. Denote by $\zeta(s)$ the value of this process at time s . Then if we conveniently choose that our observation begins at time $t = 0$, we can write $\zeta_{R_+} = \{\zeta(s); s \geq 0\}$ (here, $R_+ = (0, \infty)$). In general, ζ_{R_+} is a nonnegative intermittent stochastic process positive on random time intervals [Todorovic and Yevjevich, 1969]. In Figure 1 a sample function of this process is presented. Due to the nature of the phenomenon, physical intuition is not violated by assuming ζ_{R_+} to be stochastically continuous.

The total amount of precipitation in the time interval $(0, t)$ is

$$X(t) = \int_0^t \zeta(s) ds \tag{1}$$

If τ_v^* is the time of the beginning of the v th precipitation event, and τ_v is the time of ending, then the total amount of precipitation for the v th event is

$$U_v = \int_{\tau_v^*}^{\tau_v} \zeta(s) ds = X(\tau_v) - X(\tau_v^*) \tag{2}$$

and the storm duration is

$$D_v = \tau_v - \tau_v^* \tag{3}$$

We are interested in various structural properties of the intrastorm intensity process $\{\zeta(t); \tau_v^* < t < \tau_v\} v = 1, 2, \dots$. In general, this process will depend on geographic location and time of year. As an initial assumption, let us suppose that the seasonal and geographical variability of U_v, D_v , and their joint dependence structure can be described adequately, perhaps by mapping Fourier coefficients describing seasonal variations of parameters. Now we define the dimensionless process:

$$U_v^*(t) = \frac{1}{U_v} \int_{\tau_v^*}^t \zeta(s) ds \quad \tau_v^* \leq t \leq \tau_v \tag{4}$$

Let us now fix v and consider the following process:

$$t^* = (t - \tau_v^*)/D_v \quad \tau_v^* \leq t \leq \tau_v \tag{5}$$

It is quite clear that $0 \leq t^* \leq 1$. By virtue of this, the intensity pattern within a storm can be described by the dimensionless stochastic process $\{\bar{U}_v^*(t^*); 0 \leq t^* \leq 1\}$ where

$$\bar{U}_v^*(t^*) = U_v^*(t^* D_v + \tau_v^*) \quad v = 1, 2, \dots \tag{6}$$

It follows from this and (4) that $\bar{U}_v^*(0) = 0$ and $\bar{U}_v^*(1) = 1$. Possible realizations of this process are shown in Figure 2.

Since the process $\bar{U}_v^*(t^*)$ is strictly increasing and bounded, it will be enough to consider a finite subfamily of $\{\bar{U}_v^*(t^*); 0 \leq t^* \leq 1\}$, say,

$$\{\bar{U}_v^*(k/m); k = 0, 1, \dots, m-1, m\}$$

We now define the rescaled increments as

$$z_v(t^*) = \frac{\bar{U}_v^*(t^*) - \bar{U}_v^*(t^* - 1/m)}{1 - \bar{U}_v^*(t^* - 1/m)} \tag{7}$$

where $t^* = 1/m, \dots, (m-1)/m, 1$. From this it follows that

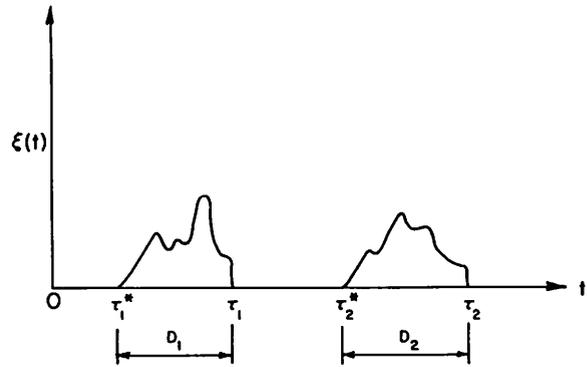


Fig. 1. Rainfall intensity process.

$z_v(0) = 0, z_v(1) = 1$, and $0 \leq z_v(t^*) \leq 1$ (here we assume that $\bar{U}_v^*(x) = 0$ for $x < 0$).

Wilkinson and Valaderes-Tavares [1972] used a similar approach to describe short time increment storm sequences. They used a dimensionless approach and developed a procedure to simulate the octal depths $0_1, 0_2, \dots, 0_7$. Their procedure started by preserving the marginal distribution of accumulated precipitation at the median duration and then proceeded in both directions, allowing the process to reach zero or one before the dimensionless time reached zero or one. They did not account for dependence between increments.

It can be shown that $\bar{U}_v^*(t^*)$ can be expressed as a function of $z_v(t^*)$

$$\bar{U}_v^*(k/m) = 1 - \prod_{i=1}^k [1 - z_v(i/m)] \tag{8}$$

$k = 1, 2, \dots, (m-1)/m$

The expected value of (8) is then given by

$$E[\bar{U}_v^*(k/m)] = 1 - E\left\{ \prod_{i=1}^k [1 - z_v(i/m)] \right\} \tag{9a}$$

$k = 1, 2, \dots, (m-1)/m$

For example, if $k = 2$, we have

$$E[\bar{U}_v^*(2/m)] = E[z_v(1/m)] + E[z_v(2/m)] - E[z_v(1/m)z_v(2/m)] \tag{9b}$$

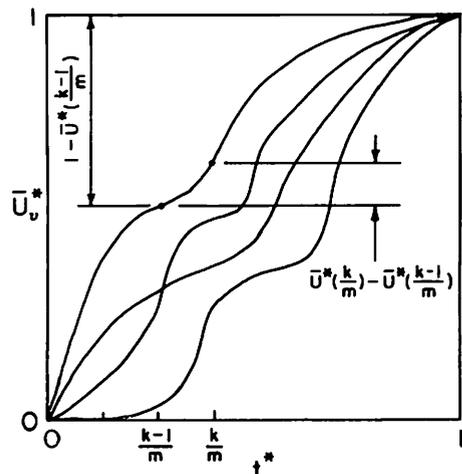


Fig. 2. Sample functions of the dimensionless rainfall accumulation process $\bar{U}_v^*(t^*)$.

ANALYSES OF DATA

We initially decided to use two types of data obtained from the Walnut Gulch Experimental Watershed in southeastern Arizona to explore the important characteristics of the joint distribution function of rescaled dimensionless rainfall increments. The first data set, which we will call "storm centered," was obtained from the rain gage in the 90-gage network that recorded the maximum amount of rainfall for 33 storms of 25.4 mm (1 inch) or greater occurring in the summer thunderstorm season from July through September. The second set, which we will call "geographically centered," was obtained by compositing the total number of storms greater than 6.4 mm (0.25 inches) that occurred from 1961 through 1976 at each of three rain gages located more than 9.66 km (6 miles) apart on the Walnut Gulch watershed. It has been demonstrated [Osborn et al., 1979; Osborn, 1982] that records from gages greater than 5.64 km (3.5 miles) apart can be considered independent. Once storm intensities drop below 3 mm/hour (0.12 in/hour), there is very little likelihood of additional bursts of higher intensity. Usually, rainfall ends shortly thereafter. However, on occasion, particularly during late evening storms, low-intensity rainfall can last for an hour or more. These long-duration "tails" seldom amount to more than 5% of the total storm rainfall but can introduce a misleading bias in storm duration. Therefore we truncated by cutting off storms when intensities dropped below 3 mm/hour (0.12 in/hour). This data set had a total of 242 storms.

Storm-Centered Data

A preliminary analysis was made to investigate the independence of successive increments $z_v(t^*)$ and $z_v(t^* + 1/m)$. The variable m was set equal to 10; thus a 1-hour rainfall would be divided into ten 6-min increments. Linear regression revealed a significant positive correlation between successive increments. Regression coefficients for the linear equation

$$y = a + bx \tag{10}$$

are shown in Table 1.

All but one of the slope values were significantly different from zero at the 0.01 level according to Student's t test. It should be noted that the t test is an approximation in this case, because the increments are not normally distributed. However, the evidence for dependence between successive increments is sufficiently strong that it must be included in the model structure. Because the increments $z_v(\)$ range between zero and one, it is appropriate to consider the beta distribution to describe the unconditional distribution of $z_v(0.1)$ and the conditional distribution of $z_v(k/m)$ given $z_v(k-1/m)$, $k = 2, 3, \dots, 9$. The beta density function for the first increment is

$$h(z(1/m)) = \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} z^{\alpha_1-1}(1-z)^{\beta_1-1} \tag{11}$$

Now the joint density function of the k th and $(k+1)$ th increments can be written

TABLE 1. Statistics for Lag-One Regressions of Rescaled Increments

y	x	a	b	r^2	Student's t
$z(0.2)$	$z(0.1)$	0.0683	0.924	0.234	3.08†
$z(0.3)$	$z(0.2)$	0.0486	1.042	0.574	6.4†
$z(0.4)$	$z(0.3)$	0.118	0.703	0.408	4.62†
$z(0.5)$	$z(0.4)$	0.131	0.593	0.320	3.81†
$z(0.6)$	$z(0.5)$	0.143	0.591	0.370	4.26†
$z(0.7)$	$z(0.6)$	0.292	0.286	0.140	2.24*
$z(0.8)$	$z(0.7)$	0.131	0.882	0.402	4.57†
$z(0.9)$	$z(0.8)$	0.421	0.377	0.241	3.14†

*Here, b is significantly different from zero at 0.05 level.
†Here, b is significantly different from zero at 0.01 level.

$$f_{k,k+1}(Z_k, Z_{k+1}) = g_{k,k+1}(Z_k)h_k(Z_{k+1}) \tag{12}$$

where $g_{k,k+1}(Z_{k+1}|Z_k)$ is the conditional density function. We assume that this conditional density function is also beta distributed:

$$g_{k,k+1}(Z_{k+1}|Z_k) = \frac{\Gamma(\alpha_{k+1} + \beta_{k+1})}{\Gamma(\alpha_{k+1})\Gamma(\beta_{k+1})} \cdot z_{k+1}^{\alpha_{k+1}-1}(1-z_{k+1})^{\beta_{k+1}-1} \tag{13}$$

where α_{k+1} and β_{k+1} may be functions of z_k , and $\alpha_{k+1} > 0$ and $\beta_{k+1} > 0$. Let us assume a linear dependence structure between successive increments, i.e.,

$$E\{Z((k+1)/m)|Z(k/m) = z_k\} = a_{k+1} + b_{k+1}z_k \tag{14}$$

The mean of the conditional density function equation (13) is

$$E\{Z((k+1)/m)|Z(k/m) = z_k\} = \alpha_{k+1}/(\alpha_{k+1} + \beta_{k+1}) \tag{15}$$

Equating the right-hand sides of (14) and (15) and solving for β_{k+1} , we obtain

$$\beta_{k+1} = \alpha_{k+1} \left[\frac{1}{a_{k+1} + b_{k+1}z_k} - 1 \right] \tag{16}$$

Note that by making this substitution we assume that β_{k+1} is a function of the previous increment, z_k , and that α_{k+1} is a constant. (It may also be of interest to consider the case where, for all $k = 1, 2, \dots$)

$$g_{k+1,k}(Z_{k+1}|Z_k) = g(Z_{k+1}|Z_k)$$

in this case,

$$E\{Z((k+1)/m)|Z(k/m) = z_k\} = a + bz_k$$

and the further simplification

$$\alpha_k = \alpha; k = 1, 2, \dots; \beta_{k+1} = \beta(z_k).$$

Substituting the expression for β_{k+1} (equation (16)) into (13), we obtain

$$g_{k+1,k}(Z_{k+1}|Z_k) = \frac{\Gamma\left(\frac{\alpha_{k+1}}{a_{k+1} + b_{k+1}z_k}\right) z_{k+1}^{\alpha_{k+1}-1} (1-z_{k+1})^{\alpha_{k+1}[1/(a_{k+1} + b_{k+1}z_k)] - 1}}{\Gamma(\alpha_{k+1}) \Gamma\left(\frac{a_{k+1}}{a_{k+1} + b_{k+1}z_k} - \alpha_{k+1}\right)} \tag{17}$$

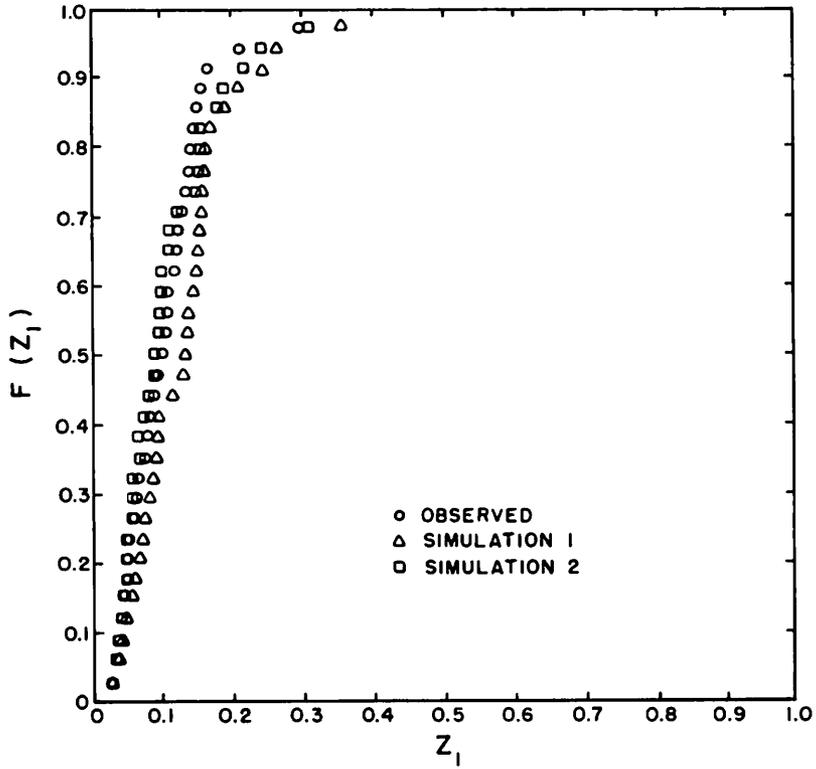


Fig. 3a. Observed and simulated marginal distributions of rescaled increments Z_1 for Walnut Gulch storm-centered data (26 parameters).

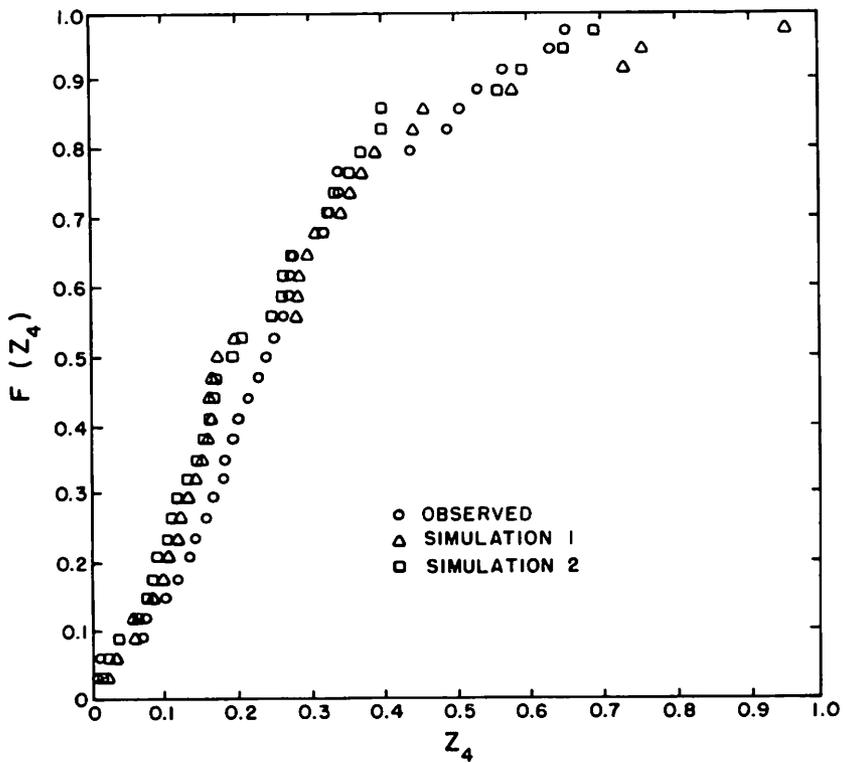


Fig. 3b. Observed and simulated marginal distributions of rescaled increments Z_4 for Walnut Gulch storm-centered data (26 parameters).

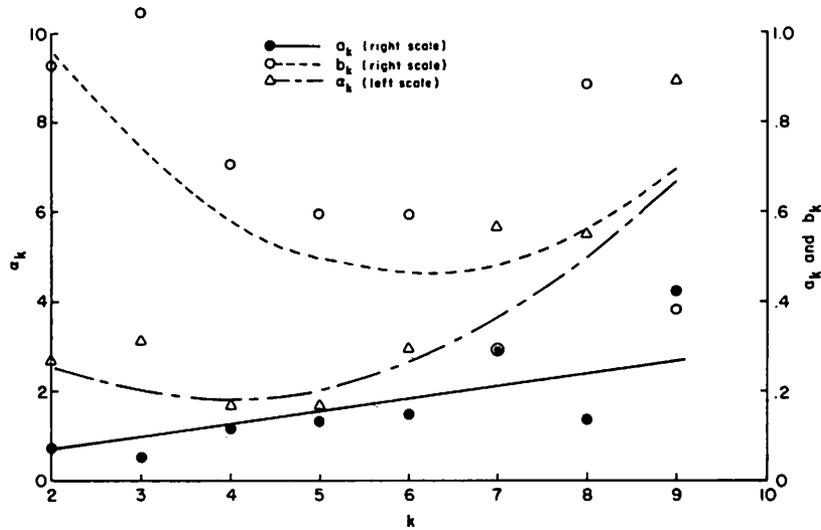


Fig. 4. Parameters of conditional distributions $f(Z_{k+1}|Z_k)$ for storm-centered data.

The random variables Z_1, Z_2, \dots, Z_9 have a joint density function $f(z_1, z_2, \dots, z_9)$. Let us also suppose that the random variables Z_1, Z_2, \dots, Z_9 represent a first order nonhomogeneous Markov chain such that

$$f(Z_k|Z_{k-1} = z_{k-1}, Z_{k-2} = z_{k-2}, \dots, Z_1 = z_1) = g_{k,k-1}(Z_k|Z_{k-1}) \quad (18)$$

The right-hand side of (18) is the one-step transition probability density function [Cox and Miller, 1965].

Now the joint density function can be written as

$$f(z_1, z_2, \dots, z_9) = h_1(z_1) \prod_{k=2}^9 g_{k,k-1}(z_k|z_{k-1}) \quad (19)$$

If (11) and (17) are substituted into (19), it is apparent that 26 parameters must be estimated, α_1 and β_1 for $h_1(z_1)$ and $a_k, b_k,$ and α_k for each conditional density, $g_{k,k-1}(z_k|z_{k-1})$.

Suppose we have a set of n storms; the log-likelihood function can be written as

$$\begin{aligned} \log L &= \log \{f(z_{11}, z_{21}, \dots, z_{91})f(z_{12}, z_{22}, \dots, z_{92}) \\ &\quad \dots f(z_{1n}, z_{2n}, \dots, z_{9n})\} \\ &= \sum_{j=1}^n \log h_1(z_{1j}) + \sum_{k=2}^9 \sum_{j=1}^n \log g(z_{kj}|z_{k-1}, j) \quad (20) \end{aligned}$$

By differentiating (20) with respect to each parameter, it is apparent that the optimization problem involved in the maximum likelihood (ML) estimates of the parameters of the multivariate density $f(z_1, z_2, \dots, z_9)$ can be reduced to a set of suboptimization problems. The parameters α_1 and β_1 were obtained by numerical optimization of the likelihood function of the univariate density $h_1(z_1)$. The parameters $(a_k, b_k); k = 2, 3, \dots, 9$ were estimated by least squares, and the parameters $\alpha_k; k = 2, 3, \dots, 9$ were estimated by numerical optimization of the log-likelihood functions

$$\log L[g(z_{k+1}|z_k)] = \sum_{j=1}^n \log \frac{\Gamma\left(\frac{\alpha}{a + bz_{kj}}\right)}{\Gamma(\alpha)\Gamma\left(\frac{\alpha}{a + bz_{kj}} - \alpha\right)}$$

$$\begin{aligned} &+ (\alpha - 1) \sum_{j=1}^n \log z_{k+1,j} \\ &+ \sum_{j=1}^n \left[\alpha \left(\frac{1}{a + bz_{kj}} - 1 \right) - 1 \right] \log (1 - z_{k+1,j}) \quad (21) \end{aligned}$$

where the parameter subscripts have been omitted, but it is understood that three parameters $a, b,$ and α will be required for each joint distribution.

Estimates of the parameters $\alpha_k, k = 2, 3, \dots, 9$ that maximized the likelihood function equation (21), obtained by a numerical univariate optimization procedure, are shown below. The α_k values for joint distributions of rescaled increments z_k, z_{k-1} are as follows.

k	α_k
2	2.66
3	3.12
4	1.67
5	1.67
6	2.93
7	5.65
8	5.46
9	8.90

The parameters listed in Table 1 (columns 3 and 4) and above, along with the parameters for $Z(0.1), \alpha_1 = 2.509, \beta_1 = 22.32,$ and (11), (17), and (19) specify the dimensionless rainfall process for the storm-centered data.

One test of the suitability of this model structure is to compare the marginal distributions of the $z(t^*)$ obtained by simulation with those of the original data set. Monte Carlo techniques were used to generate two sets of 33 dimensionless storms. The rejection technique [Naylor et al., 1966] was used to generate a beta distributed random variable for $z(0.1)$. Increment $z(0.2)$ was then generated using the rejection technique with the conditional beta distribution specified by (17). The procedure was repeated until $z(0.9)$ had been generated. Note that by definition (equation (7)), $z_{10} = 1$. The observed and simulated cumulative marginal distributions for $z(0.1)$ and $z(0.4)$ are shown in Figure 3. The simulated observed marginal distributions for the other increments show similar agreement.

The good agreement between the simulated and observed marginal distribution functions suggests that the stochastic

structure proposed is a reasonable first approximation. However, it seems useful to investigate methods to reduce the number of parameters required.

The parameters a_k , b_k , and α_k were plotted as a function of k to see if any regularities exist (Figure 4). From these data it appears that the parameters can be approximated by equations of the form

$$\begin{aligned} a_k &= B_1 + B_2k \\ b_k &= B_3 + B_4k + B_5k^2 \\ \alpha_k &= B_6 + B_7k + B_8k^2 \end{aligned} \tag{22}$$

where $K = 2, 3, \dots, 9$. If this representation provides a satisfactory fit, the parameter set can be reduced from 26 parameters to 10, i.e., $\theta = (B_1, B_2, \dots, B_8, \alpha_1, \beta_1)$.

By substituting the relationships for a_k , b_k , and α_k given by (22) into (17), and utilizing (11) and (20), the log-likelihood function can be written as

$$\begin{aligned} \log L &= \sum_{j=1}^n \log \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} z_{1j}^{\alpha_1-1} (1 - z_{1j})^{\beta_1-1} \\ &+ \sum_{k=2}^9 \sum_{j=1}^n \log \left\{ \left[\frac{\Gamma(B_6 + B_7k + B_8k^2)}{B_1 + B_2k + (B_3 + B_4k + B_5k^2)z_{k-1,j}} \right] \right. \\ &\cdot \left. \left[\Gamma(B_6 + B_7k + B_8k^2) \Gamma\left(\frac{B_6 + B_7k + B_8k^2}{B_1 + B_2k + (B_3 + B_4k + B_5k^2)z_{k-1,j}} - (B_6 + B_7k + B_8k^2) \right) \right]^{-1} \right\} \\ &+ \sum_{k=2}^9 (B_6 + B_7k + B_8k^2 - 1) \sum_{j=1}^n \log z_{kj} \\ &+ \sum_{k=2}^9 \sum_{j=1}^n \left[(B_6 + B_7k + B_8k^2) \left(\frac{1}{B_1 + B_2k + (B_3 + B_4k + B_5k^2)z_{k-1,j}} - 1 \right) - 1 \right] \log(1 - z_{kj}) \end{aligned} \tag{23}$$

The maximum likelihood estimators of the parameter set $\theta = (B_1, B_2, \dots, B_8, \alpha_1, \beta_1)$ can be obtained by numerical maximum likelihood techniques. It is worth noting that the parameters α_1 and β_1 can be obtained independently, because they appear only in the first term of (23); therefore the optimization involved a maximum of 8 parameters. A pattern search optimization method, PATSEAR [Green, 1970], was used to obtain approximate maximum likelihood parameters. In general, the gamma functions in (23) have noninteger arguments. They were evaluated in the following manner:

$$\Gamma(x + 1) = x! = x(x - 1)(x - 2) \cdots (y + 1)y! \tag{24}$$

where $0 \leq y \leq 1$, and $y!$ is evaluated by a power series

$$y! = 1 + a_1y + a_2y^2 + a_3y^3 + a_4y^4 + a_5y^5 + \epsilon(y) \tag{25}$$

for the values of the coefficients used $|\epsilon(y)| \leq 5 \times 10^{-5}$ [Hastings, 1955].

Because $0 < z_{k+1} < 1$, the parameters a_{k+1} and b_{k+1} in (14) must be constrained according to the relationships

$$\begin{aligned} a_{k+1} &> 0 \\ 0 &< a_{k+1} + b_{k+1} < 1.0 \end{aligned} \tag{26}$$

Keeping in mind expression (22), (B_1, B_2, \dots, B_8) must satisfy the inequalities

$$\begin{aligned} 0 &< B_1 + B_2k \quad k = 2, 3, \dots, 9 \\ B_1 + B_2k + B_3 + B_4k + B_5k^2 &< 1.0 \quad k = 2, 3, \dots, 9 \end{aligned} \tag{27}$$

These constraints were imposed in the program by adding a penalty function to the objective function, equation (23), when either expression (27) was within 0.001 of the constraint.

Four alternative forms of (22) were compared with each other and with the 26-parameter model using the Akaike minimum information theoretical criterion (AIC) [Akaike, 1974]. According to this criterion, we select the model which has the minimum AIC where

$$\text{AIC} = (-2) \log(\text{maximum likelihood}) + 2m \tag{28}$$

where m is the number of optimized parameters. The values of the coefficients and the test statistics for the alternative functional relationships are shown in Table 2. From an examination of this table we see that the model in which a_k is a linear

TABLE 2. Comparison of Alternative Functional Relationships for Parameters for Storm-Centered Data

a_{k+1}		b_{k+1}			α_{k+1}			Maximum Log Likelihood ¹	AIC	Number of Optimized Parameters, m	
B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8				
0.0269	0.226	0.654	0	0	1.005	3.375	0	167.21	-324.43	5	
0.0284	0.247	0.600	0.0534	0	4.556	-14.42	18.64	174.26	-334.53	7	
0.0244	0.253	0.638	0	0	4.560	-14.47	18.70	174.20	-336.40	6	
0.0127	0.283	1.546	-3.517	2.853	4.684	-14.58	18.64	179.26	-342.53*	8	
<i>26-Parameter Model</i>											
								194.81	-341.62	24	

Here, $\sum_{j=1}^n \log f_1(z_{ij}; \alpha\beta)$ was omitted because it is the same for all cases (503.721).
*Minimum AIC.

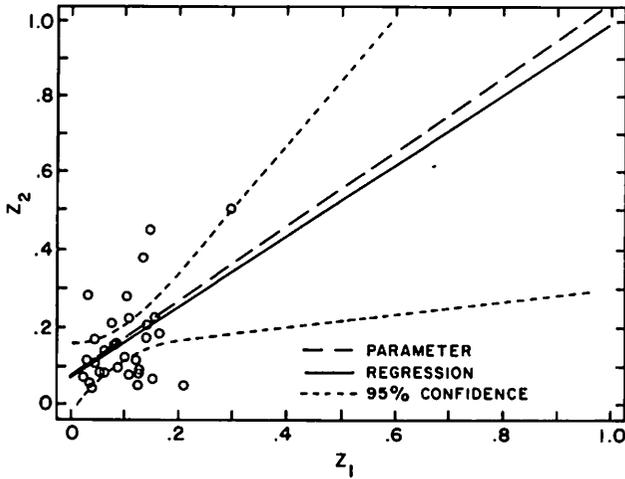


Fig. 5a. Regression relationship between Z_2 and Z_1 for storm-centered data.

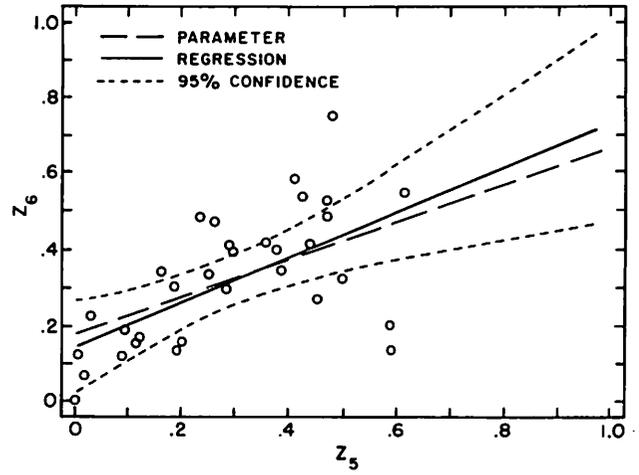


Fig. 5c. Regression relationship between Z_6 and Z_5 for storm-centered data.

function of t_* , and b_k and α_k are quadratic functions of t_* is superior to the other parametric forms, and also to the 26-parameter model (although it must be noted that the (a_k, b_k) for the 26-parameter model were estimated by least squares).

The linear expected value functions, calculated from (14) with values of a_{k+1} and b_{k+1} , calculated from (22) using coefficients B_1 through B_5 from row 4 of Table 2, are plotted along with the data for selected increment pairs in Figures 5a through 5d. The linear regression equations and the 95% confidence bands for the regression lines are shown for comparison. In all cases, the linear expected value functions obtained from the 8-parameter model were within the 95% confidence band for the regression lines. The confidence bands are approximate, since the distributions are not bivariate normal.

Geographically Centered Data

The data for 242 summer thunderstorms over 0.64 cm (0.25 inches) were obtained from three rain gages spaced over 9.66 km (6 miles) apart on the Walnut Gulch Experimental Watershed in southeastern Arizona. With the minimum 9.66-km spacing, the gages sample different thunderstorm cells. As stated earlier, storms were truncated when intensities dropped below 3 mm/hour (0.12 in/hour). These data were separated into four files based on storm amount and duration (see Figure 6).

Because we have a relatively large sample of storms in each subset ($D_1 \cap P_1$), for example, it is possible to test the hypotheses "the dimensionless precipitation process is independent of actual storm duration," or "the dimensionless precipitation process is independent of total storm precipitation."

Numerical optimization techniques were used to estimate the parameter set for each of the four classes of data shown in Figure 6 and for the two-way classification based upon storm duration. The maximum likelihood parameter set for storms with greater than the median amount of rainfall and equal to, or less than, the median duration ($D_1 \cap P_1$) is designated as θ_{11} . The parameter set for all storms greater than the median duration is designated as θ_B .

The dependence parameters, i.e., the intercept a_k , slope b_k , and conditional parameter α_k , obtained by numerical optimization for each data subset, are shown as a function of the time index k in Figure 7. An examination of Figures 7a and 7b reveals that the dependence structure is remarkably similar for each data subset. However, from Figure 7c we see a substantial difference between the α_k functions for data subsets, with α_k taking on the largest values for subset $D_1 \cap P_2$ and the lowest values for subset $D_2 \cap P_2$. It appears that the duration of the storms may be a significant factor for this parameter.

To demonstrate the differences between the conditional density function $g(z_9|z_8)$ for the range of α_9 values shown on

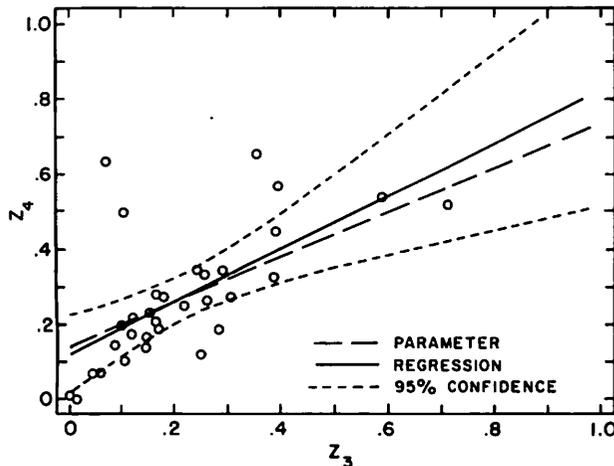


Fig. 5b. Regression relationship between Z_4 and Z_3 for storm-centered data.

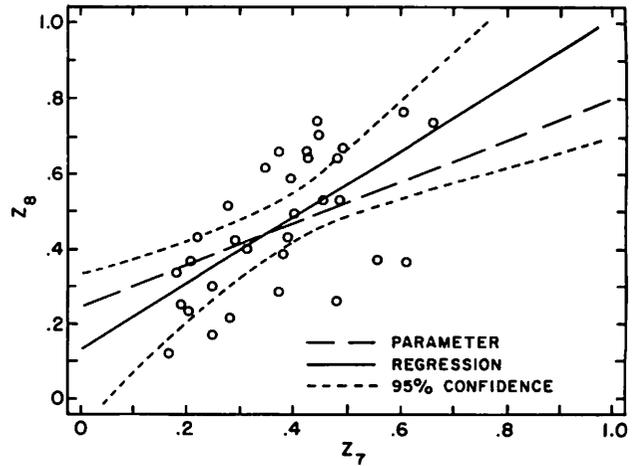


Fig. 5d. Regression relationship between Z_8 and Z_7 for storm-centered data.

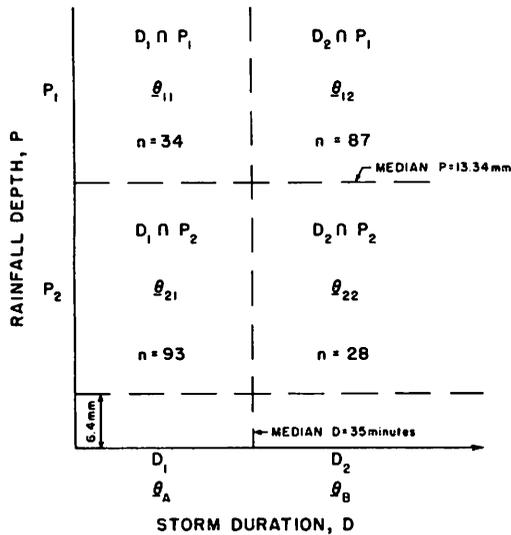


Fig. 6. Sample domains and parameter set definitions for geographically centered data.

Figure 7c, the density functions were calculated for two values of z_8 and for $\alpha_9 = 8$ and $\alpha_9 = 16$. These density functions are shown in Figure 8. For the short-duration data sets the values of z_9 are more closely clustered around the linear expected value function than for the longer-duration storms.

The likelihood ratio test and the Akaike information criterion were used to select the appropriate model. For example, we might wish to test the null hypothesis that for storms with durations equal to, or less than, the median duration, there is no difference between the dimensionless precipitation process for storms having depths greater than the median depth as compared to storms with depths equal to, or smaller than, the median depth.

Symbolically, the null hypothesis is

$$H_0: \omega' = (\theta_A) = (\alpha_A, \beta_A, B_{1A}, B_{2A}, \dots, B_{8A})$$

and the alternative is

$$H_1: \omega = (\theta_{11}, \theta_{21}) = (\alpha_{11}, \beta_{11}, B_{11}, B_{21}, \dots, B_{81}, \alpha_{12}, \beta_{12}, B_{12}, B_{22}, \dots, B_{82})$$

If the likelihood function under the null hypotheses is denoted by L_0 and under the alternative hypotheses by L_1 , then for large sample size, the statistic

$$-2 \log \lambda = -2 \log (L_0/L_1) \tag{29}$$

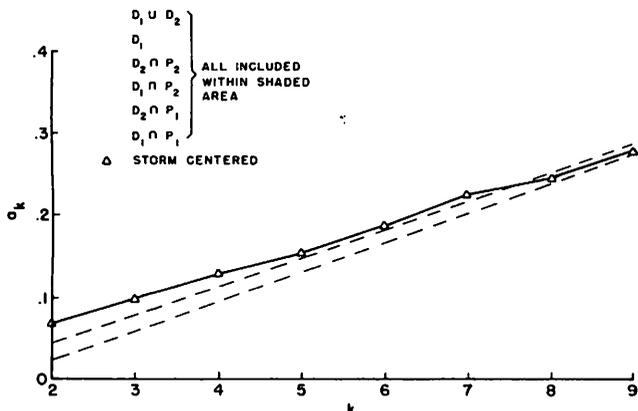


Fig. 7a. Parameter α_k (intercept) for geographically centered data.

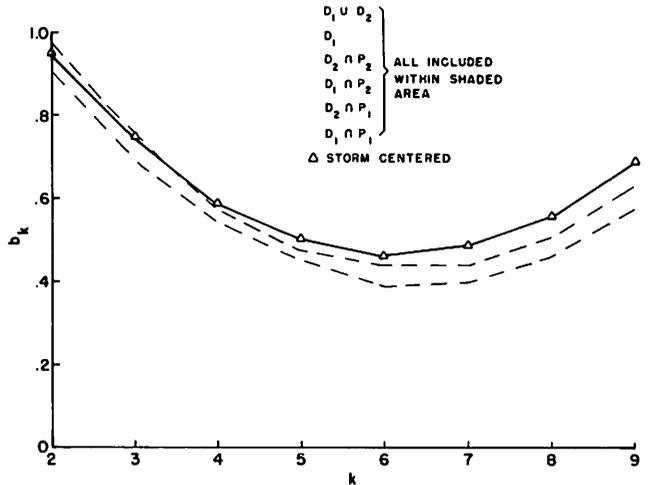


Fig. 7b. Parameter b_k (slope) for geographically centered data.

is approximately χ^2 with degrees of freedom equal to the difference between the number of parameters in the alternate parameter sets (in this case, 10) [c.f. Hoel, 1971; Mielke and Johnson, 1973]. A choice can also be made between competing models by selecting the model which provides the minimum information theoretical criterion AIC [Akaike, 1974].

The test statistics are shown in Table 3. From an examination of this table we find that for the short-duration class D_1 we cannot reject the hypotheses that the dimensionless precipitation process is independent of rainfall amount. The probability of obtaining a larger χ^2 statistic is approximately 0.04, but the model under the null hypothesis gives the minimum AIC. However, for sample D_2 the hypothesis of a common set of parameters θ_B for both subsets $D_2 \cap P_1$ and $D_2 \cap P_2$ is rejected by both the likelihood ratio test and the AIC.

Similarly, the null hypothesis of a common parameter set θ was tested against the alternative of four sets, one for each subset of data ($\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}$), and was rejected.

Finally, the observation from Figure 7 that the dependence parameters were similar for all data subsets but that α_k was dependent on duration prompted the test of a 13-parameter model against a 30-parameter model, as is shown in the last row of Table 3. Here it is hypothesized that the parameters for the beta distribution for z_1, α , and β and the linear dependence parameters B_1 through B_5 are common to the whole data set, but that the parameters describing the within storm variability of α_k are dependent on duration. Thus 13 parameters were

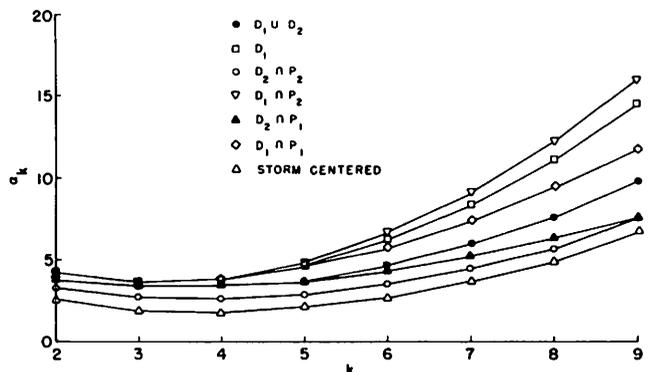


Fig. 7c. Parameter α_k for geographically centered data.

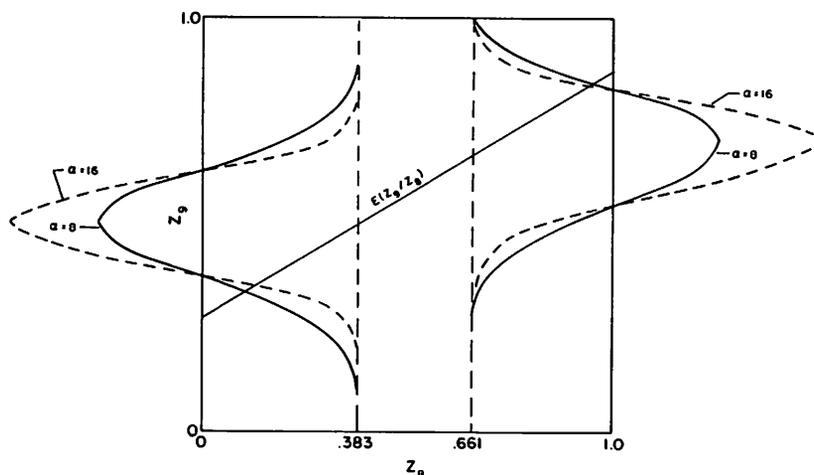


Fig. 8. Effect of variations of α_9 on the conditional density function $g(Z_9/Z_8)$.

optimized. The parameters $(\alpha, \beta, B_1, B_2, \dots, B_5)$ were optimized using all storms. The parameters (B_{6A}, B_{7A}, B_{8A}) were optimized over data set D_1 , and (B_{6B}, B_{7B}, B_{8B}) were optimized over data set D_2 . The alternative model utilized the parameter set $\omega = (\theta_A, \theta_{12}, \theta_{22})$, consisting of 30 parameters. The probability of obtaining a larger χ^2 statistic, if the null hypothesis is true, is slightly less than 0.05. However, the 13-parameter model leads to the smallest AIC, so it appears to be superior to the more complicated model. In this particular application it can be seen that the minimum AIC criterion is equivalent to application of the log-likelihood ratio test, with a level of significance of approximately 0.04.

The parameter values for the 13-parameter model in Table 3 are shown in Table 4. These values can be used in a simulation program to generate dimensionless accumulated precipitation realizations for areas where thunderstorms are similar to those at the Walnut Gulch watershed. Some appreciation for the parsimonious nature of this model may be obtained by considering the number of parameters required by alternative disaggregation schemes. For example, the method proposed by Valencia and Schaake [1973] would require 110 parameters, and the procedure used by Bras and Rodriguez-Iturbe [1976] would require 21.

BIVARIATE DISTRIBUTION OF STORM AMOUNT AND DURATION

The bivariate distribution of rainfall amount and duration has usually been determined by defining the marginal distribution of storm duration and the conditional distribution of rainfall amount given duration. This, of course, is the natural approach when the rainfall process is simulated by generating

storm interarrival times and storm durations, followed by the storm amount, conditioned on the storm duration. Because the objective of this research is to disaggregate daily rainfall, we first describe the marginal distribution of storm amounts, $g_Y(y)$, and then the conditional distribution of storm duration $h_D(d|Y = y)$. For the geographically centered rainfall amount data (>6 mm) we estimated the parameters for the exponential, mixed exponential, log-normal and Kappa III distributions using maximum likelihood techniques. The exponential distribution, with mean $1/\lambda = 0.426$ (inches) (10.8 mm), had the greatest likelihood function (the mixed exponential converged to a single exponential), and thus provided the best fit to these data. The theoretical and empirical distribution functions were virtually indistinguishable, and the hypothesis that the sample came from an exponential distribution could not be rejected at the 0.05 level using the χ^2 test ($\chi^2 = 6.2768, \chi_{0.05,7}^2 = 14.0720$). These results are approximate, because we used the data to estimate the form of the distribution function as well as the parameters.

The conditional distribution of storm duration, given amount, was described by a log-normal distribution with expected value function

$$E\{\log D|Y' = y'\} = 0.0272y' + 3.255 \tag{30}$$

where $y' = y - 6.35$ mm. The standard error of estimate was 0.423. To test the hypothesis of a normal distribution of errors about the regression line, the data were subdivided into three approximately equal classes of storm amounts. The hypothesis that the residuals from regression for each class was described by the normal distribution with zero mean, and the standard deviation of 0.423 could not be rejected at the 0.05 level using the χ^2 test.

TABLE 3. Likelihood Ratio Statistics and AIC for Alternative Models

Sample	Hypothesis	$-2 \log \lambda$	d.f.	$\chi^2_{0.95}$	AIC	$\log L$	m
D_1	$H_0: \omega' = \theta_A$	19.02	10	18.3	-2445.56*	1232.78	10
	$H_1: \omega = (\theta_{11}, \theta_{21})$				-2444.58	1242.29	20
D_2	$H_0: \omega' = \theta_B$	26.26	10	18.3	-1741.48	880.74	10
	$H_1: \omega = (\theta_{12}, \theta_{22})$				-1747.74*	893.87	20
$D_1 \ D_2$	$H_0: \omega' = (\theta)$	106.4	30	43.8	-4145.92	2082.96	10
	$H_1: \omega = (\theta_{11}, \theta_{21}, \theta_{12}, \theta_{22})$				-4192.32*	2136.16	40
$D_1 \ D_2$	$H_0: \omega' = (\alpha, \beta, B_1, \dots, B_5, B_{6A}, B_{6B}, \dots, B_{8A}, B_{8B})$	27.72	17	27.6	-4199.58*	2112.79	13
	$H_1: \omega = (\theta_A, \theta_{12}, \theta_{22})$				-4193.3	2126.65	30

*Minimum AIC.

TABLE 4. Parameters for Model With Minimum AIC.

Domain	α	β	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8
D_1	2.058	12.981	-0.028	0.343	1.538	-3.421	2.607	7.139	-20.179	30.954
D_2	2.058	12.981	-0.028	0.343	1.538	-3.421	2.607	5.30	-10.75	13.912

EXTREME INTENSITIES FOR FIXED DURATIONS

Techniques for disaggregating storms described in this paper will certainly result in simulated storms that have lower short period rainfall intensities than natural storms because of the requirement for breaking each storm into 10 equal time periods. However, it is difficult to appreciate how significant this bias might be without a comparison of simulated storms with the original data. Consequently, 242 storms were generated using the bivariate distribution of amount and duration described above. These storms were disaggregated using both the 13 parameters identified for the geographically centered data (Table 4) and the 10 parameters identified for the storm-centered data (Table 2). The maximum intensities for 5, 10, 30, 60, and 120 min were then identified for each simulated storm and for the original data. Empirical distribution functions of these maximum intensities are shown in Figures 9 and 10.

Except for the 5-min duration, there is a tendency for the simulated intensities to be biased low as compared with the empirical distribution function from the data. It is noteworthy that the simulated distributions, using the "storm-centered" parameters, actually fit better than those using parameters from the geographically centered data. This appears to be related to the parameter α , which is lower for all k for the storm-centered data. From Figure 8 it is apparent that this leads to greater variance in the conditional distribution of z_{k+1} given z_k . Considering the simplifications and assumptions

involved in this disaggregation model, the extreme intensity statistics are preserved very well.

DISCUSSION

The nature of thunderstorm rainfall, the procedures used to tabulate recording raingage records, and the method of re-scaling the dimensionless rainfall increments used in this paper combine in a manner that suggests that the differences in the parameters α_k for data sets D_1 and D_2 may be more apparent than real. Near the end of a thunderstorm there frequently is a period of low-intensity rainfall for a period of up to 0.3 of the total storm duration. Because this intensity is often nearly constant, or only 0.5-0.76 mm (0.02-0.03 inches) of rain are involved, it will be represented in the dimensionless function $U_v^*(t^*)$ as a straight line. Under these circumstances, the method of defining $z(t^*)$ will lead to a significant number of cases where $z(0.7) = 0.25$, $z(0.8) = 0.33$, and $z(0.9) = 0.5$. The empirical distribution functions of these increments will then appear to be partly continuous and partly discrete. Because this is more likely to occur for the shorter storms, it contributes to the higher values of α_k estimated for set D_1 , as compared with set D_2 .

Although the 13-parameter representation of the dimensionless stochastic precipitation process reproduces many of the important characteristics of the original data and is parsimonious, it does introduce some bias into the marginal distri-

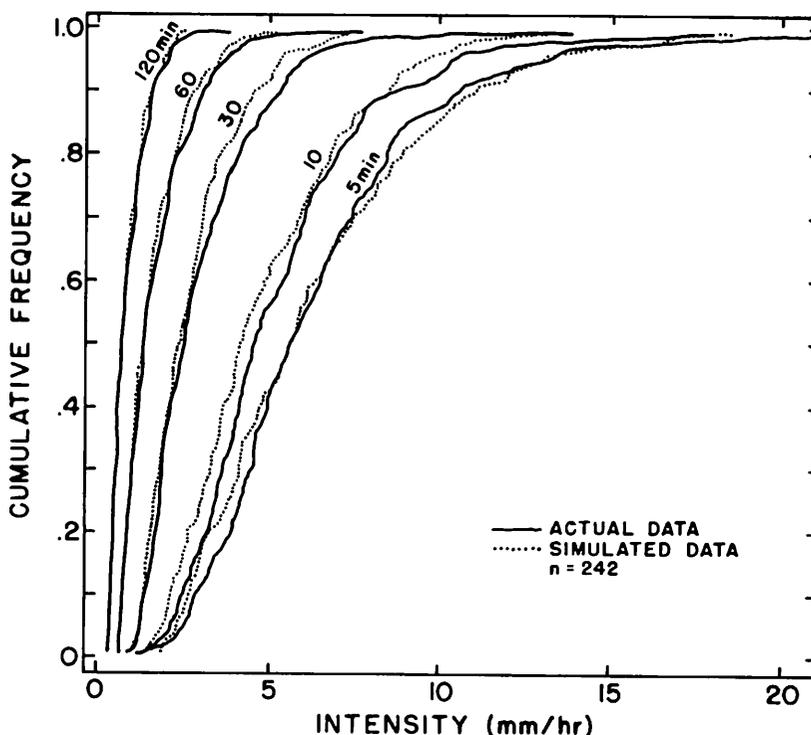


Fig. 9. Distributions of maximum intensity for time periods of 5, 10, 30, 60, and 120 min for geographically centered data (13-parameter simulation model).

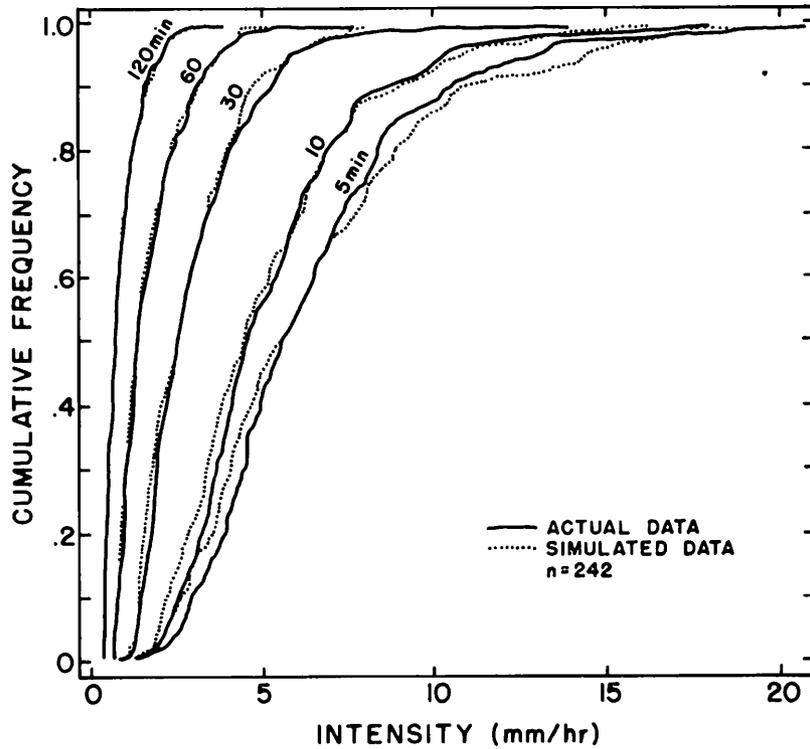


Fig. 10. Distributions of maximum intensity for time periods of 5, 10, 30, 60, and 120 min (10-parameter storm-centered simulation model).

butions of the increments $z(t^*)$ and the marginal distributions of the dimensionless accumulated precipitation process $\bar{U}_v^*(k/m)$ shown in Figure 11. Clearly, using the functional forms of the parameters specified by (22) results in a less precise description of the process than can be achieved with the use of 26 parameters, as is demonstrated by Figure 3. Although the AIC could be used to compare the 10- or 13-parameter model with the 26-parameter model, it would be more appropriate, from the hydrologic standpoint, to determine if there are significant differences in derived statistics such as rates and volumes of runoff obtained from infiltration models.

The proposed model does an acceptable job of maintaining the correlation structure between increments. The correlation coefficient $r_{k,k-j}$ is shown as a function of the lag j for $z_9, z_7, z_5,$ and z_3 in Figure 12.

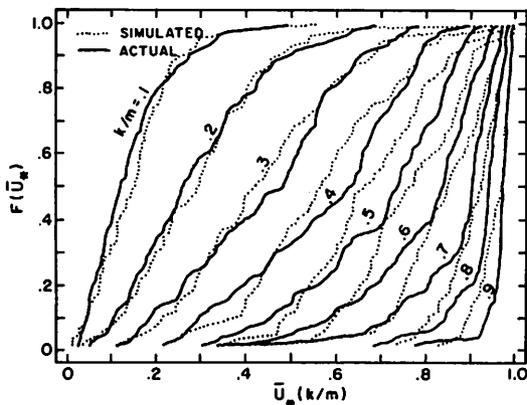


Fig. 11. Cumulative distribution functions of the dimensionless process $\bar{U}^*(k/m)$ for long-duration, high-intensity storms $D_2 \cap P_1$.

As was stated in an earlier section, the joint distribution of rainfall amount and duration for thunderstorm rainfall will most certainly depend on geographic location and time of year. However, it is possible that the dimensionless stochastic rainfall accumulation process $\bar{U}_v^*(k/m)$ may be (relatively) spatially invariant. Although we have not thoroughly examined this possibility, the expected value function $E(\bar{U}_v^*(k/m))$ for a small sample of thunderstorms in the southeastern United States [Sorman and Wallace, 1972] is very similar to that for the storms investigated in this paper. We plan to investigate the seasonal and geographic variability of $\bar{U}_v^*(k/m)$ in a subsequent paper.

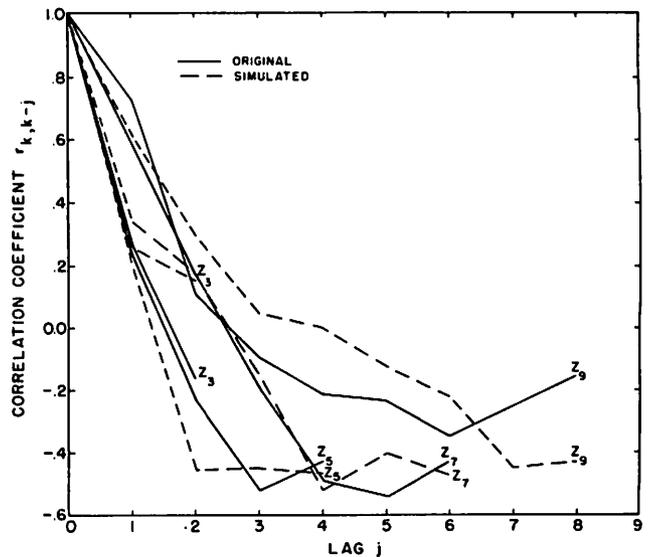


Fig. 12. Correlation versus lag for long-duration, high-intensity storms.

SUMMARY AND CONCLUSIONS

A concise stochastic model for the dimensionless accumulated thunderstorm rainfall process at a point is proposed. The dimensionless process is divided into 10 equal time increments, and the depth increments are rescaled to a range of 0 to 1. The sequence of rescaled increments Z_1, Z_2, \dots, Z_9 are assumed to represent a nonhomogeneous Markov chain in discrete time with continuous state space. The expected value function of the k th rescaled increment, given the $(k-1)$ th increment, is assumed to be linear, and the marginal distribution of the first increment and the conditional distributions of the k th increment, given the $(k-1)$ th increment, are assumed to be described by the beta distribution.

An analysis of 275 storms observed at the Walnut Gulch Experimental Watershed in southeastern Arizona suggests that the proposed model structure provides an acceptable approximation for summer thunderstorm rainfall. Simulated marginal distributions of the rescaled increments z_k and the accumulated dimensionless precipitation \bar{U}_k agreed well with the data. The correlation structure between the rescaled increments was also maintained reasonably well. The distributions of maximum intensities for periods of 5, 10, 30, 60, and 120 min obtained from simulated storms were also very similar to those obtained from the data.

The number of model parameters can be reduced from 26 to a minimum of 10 by approximating the two dependence parameters and the conditional beta parameter as power series functions of the time increment. The coefficients of the power series for each parameter can be estimated by numerical maximum likelihood techniques. According to the Akaike information criterion, the model with 10 parameters is superior to the 26-parameter model for the storm-centered data.

Likelihood ratio tests and the Akaike information criterion suggest that the dependence parameters are independent of storm amount and duration, but the parameter α_k is larger for short duration storms than for long duration storms. A 13-parameter model, in which the parameter α_k is dependent on storm duration, is recommended for summer rainfall in southeastern Arizona.

Acknowledgments. T. Econopouly did most of the computer programming for the analyses presented. E. D. Shirley and P. Todorovic made several helpful suggestions.

REFERENCES

- Akaike, H., A new look at the statistical model identification, *IEEE Trans. Autom. Contr.*, 19(6), 716-723, 1974.
- Amocho, J., and B. Wu, Mathematical models for the simulation of cyclonic storm sequences and precipitation fields, *J. Hydrol.*, 32, 329-345, 1977.
- Austin, T. A., and B. J. Claborn, Statistical model of short-duration precipitation events, Proceedings, symposium on statistical hydrology, *Misc. Publ. 1275*, pp. 356-365, U.S. Dep. Agric. Res. Serv., Tucson, Ariz., 1974.
- Brakensiek, D. L., R. L. Engleman, and W. J. Rawls, Variation within texture classes of soil water parameters, *Trans. Am. Soc. Agric. Eng.*, 24(2), 335-339, 1981.
- Bras, R. L., and I. Rodriguez-Iturbe, Rainfall generation: A non-stationary time-varying multidimensional model, *Water Resour. Res.*, 12(3), 450-456, 1976.
- Corotis, R. B., Stochastic considerations in thunderstorm modeling, *J. Hydraul. Div. Am. Soc. Civ. Eng.*, 102(HY7), 865-878, 1976.
- Cox, D. R., and H. D. Miller, *The Theory of Stochastic Processes*, Chapman and Hall, London, 1972.
- Grace, R. A., and P. S. Eagleson, The synthesis of short-time increment rainfall sequences, *Hydrodyn. Lab. Rep. 91*, Mass. Inst. of Technol., Cambridge, Mass., 1966.
- Green, R. F., Optimization by the pattern search method, *Res. Pap. 7*, 72 pp. Div. Water Contr. Plann., Tenn. Val. Auth., Knoxville, Tenn., 1970.
- Green, W. A., and G. A. Ampt, Studies on soil physics, I, The flow of air and water thru soils, *J. Agric. Sci.*, 4, 1-24, 1911.
- Hastings, C., Jr., *Approximations for Digital Computers*, Princeton University Press, Princeton, N. J., 1955.
- Hershendorff, J. S., Disaggregation of daily rainfall, M.S. thesis, 146 pp., Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson, 1984.
- Hoel, F. G., *Introduction to Mathematical Statistics*, 4th ed., John Wiley, New York, 1971.
- Knisel, W. G., Jr. (Ed.), CREAMS: A field-scale model for chemicals, runoff, and erosion from agricultural management systems, 640 pp., *Conserv. Res. Rep. 26*, U.S. Dep. Agric., Tucson Ariz., 1980.
- Knisel, W. G., Jr., and W. M. Snyder, Stochastic time distribution of storm rainfall, *Nord. Hydrol.*, 6, 242-262, 1975.
- Mielke, P. W., and E. S. Johnson, Three-parameter kappa distribution maximum likelihood estimates and likelihood ratio tests, *Mon. Weather Rev.*, 101(9), 701-707, 1973.
- Naylor, T. H., J. L. Balintfy, D. S. Burdick, and K. Chu, *Computer Simulation Techniques*, John Wiley, New York, 1966.
- Nguyen, V.-T.-V., and J. Rousselle, A stochastic model for the time distribution of hourly rainfall depth, *Water Resour. Res.*, 17(2), 399-409, 1981.
- Osborn, H. B., K. G. Renard, and R. Simanton, Dense networks to measure convective rainfall in the southwestern United States, *Water Resour. Res.*, 15(6), 1701-1711, 1979.
- Osborn, H. B., Quantifiable differences between air mass and frontal-convective thunderstorm rainfall in the southwestern United States, in *Statistical Analyses of Rainfall and Runoff*, Water Resources Publishers, Littleton, Colo., 1982.
- Pattison, A., Synthesis of hourly rainfall data, *Water Resour. Res.*, 1(4), 489-498, 1965.
- Raudkivi, A. J., and N. Lawgun, Simulation of rainfall sequences, *J. Hydrol.*, 22, 271-294, 1974.
- Rawls, W. J., D. L. Brakensiek, and N. Miller, Predicting Green and Ampt infiltration parameters from soils data, *J. Irrig. Drain. Div. Am. Soc. Civ. Eng.*, 109(IR1), 62-70, 1983.
- Richardson, C. W., and D. A. Wright, WGEN: A model for generating daily weather variables, *Rep. ARS-8*, 83 pp., U.S. Dep. Agric. Res. Serv., Tucson, Ariz., 1984.
- Sinha, L. K., and N. N. Khanal, Estimation of rainfall for Kissimmee River Basin, *Pap. Am. Soc. Agric. Eng.*, 71-728, 25 pp., 1971.
- Smith, R. E., and J. R. Williams, Simulation of the surface water hydrology, in CREAMS: A field-scale model for chemicals, runoff, and erosion from agricultural management systems, *Conserv. Res. Rep. 26*, U.S. Dep. Agric., Tucson, Ariz., 1980.
- Sorman, U. A., and J. R. Wallace, Digital simulation of thunderstorm rainfall, *Rep. ERC-0972*, Sch. of Civ. Eng., Ga. Inst. of Tech., Atlanta, Ga., 1972.
- Todorovic, P., and V. Yevjevich, Stochastic process of precipitation, *Hydrol. Pap. 35*, Colo. State Univ., Fort Collins, 1969.
- Valencia, D. R., and J. C. Schaake, Jr., Disaggregation processes in stochastic hydrology, *Water Resour. Res.*, 9(3), 580-585, 1973.
- Wilkinson, J. C., and L. Valaderes-Tavares, A methodology for the synthesis of spatially-distributed short-time increment storm sequences, *J. Hydrol.*, 16, 307-315, 1972.
- Woolhiser, D. A., H. B. Osborn, and J. Hershendorff, Disaggregation of daily rainfall, in Proceedings, Natural Resources Modeling Workshop, Pingree Park, Colo., USDA-ARS, Tucson, Ariz., in press.

H. B. Osborn and D. A. Woolhiser, U.S. Department of Agriculture, Agricultural Research Service, 2000 E. Allen Road, Tucson, AZ 85719.

(Received May 7, 1984;
revised December 5, 1984;
accepted December 17, 1984.)