Journal of Hydrology, 95 (1987) 299-322
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

[1]

DISAGGREGATION OF DAILY RAINFALL

J. HERSHENHORN and D.A. WOOLHISER

Pima County Department of Transportation and Flood Control, 1313 S. Mission Road, Tucson, AZ 85713 (U.S.A.)
Aridland Watershed Management Research, Agricultural Research Service, U.S. Department of Agriculture, 2000 E. Allen Road, Tucson, AZ 85719 (U.S.A.)

(Received April 6, 1987; revised and accepted April 23, 1987)

ABSTRACT

Hershenhorn, J. and Woolhiser, D.A., 1987. Disaggregation of daily rainfall. J. Hydrol., 95: 299-322.

A parameter-efficient model for disaggregating daily rainfall into individual storms is presented. This model allows simulation of the number of rainfall events (storms) in a day, and the amount, duration, and starting time of each event, given only the total rainfall on that day and on the preceding and following days. Twenty-three years of data for July and August, from a gage on the Walnut Gulch Experimental Watershed, were used to find the appropriate model structure and to estimate parameters. Statistical tests indicate that simulated sequences of storms compare favorably with observed sequences, and that the disaggregation model structure and parameters identified for one gage provide a satisfactory fit for three stations within a 121 km radius where elevation differs by as much as 244 m, and mean annual rainfall differs by up to 76 mm.

INTRODUCTION

Greater emphasis is being placed on the use of physically-based infiltration models to estimate surface runoff, and these models are sensitive to the distribution of rainfall amounts in time periods as short as 5 min. "Breakpoint data", the observed time patterns of rainfall intensity, can be used to provide input to infiltration models, however, such data are not always readily available.

One method of obtaining short-time period rainfall data is to disaggregate daily rainfall amounts into individual showers, and to further disaggregate the showers into intensity patterns. Daily rainfall data are readily available for many locations, and simulation procedures are well advanced. If spatial and seasonal variability can be accounted for in the daily process, and the disaggregation processes are relatively spatially invariant, they can be applied over climatologically-similar areas, subject to minor modifications.

In this paper, we describe a parameter-efficient model to simulate the number of rainfall events (showers) in a day, and the amount, duration, and starting time of each event, given only the total amount of rainfall on that day and on the preceding and following days. When used in conjunction with a daily

NOTATION

D (or d)	Duration of a storm event (min)
D' (or d')	$\log_e d$
D_c	Duration of a complete event
D_{cn}	Duration of the nth complete event
D_k	Duration of the kth event in a day
$\hat{D}_k$	Simulated duration of the kth event in a day
D_n	Duration of the nth event
D_p	Duration of a partial event
$F(\cdot)$, $G(\cdot)$, $H(\cdot)$	Cumulative distribution functions
i	Year index
j	Day index
k	Total number of complete and partial events in a day
M_j	Random variable with values of 0 or 1 indicating occurrence of a partial event on day j
N_c	Number of complete events in a day
N_p	Number of partial events in a day
N_{cij}	Number of complete events on day j of year i
N_{pij}	Number of partial events on day j of year i
p	Parameter in shifted negative binomial distribution
r	Parameter in shifted negative binomial distribution
R_1	The ratio of the amount of the first event of the day to the daily total amount on days in which two events occur
R_2	The ratio of the sum of the amounts of the second plus third events to the daily total on days in which three events occur
R_3	The ratio of the amount of the second event to the sum of the amounts of the second plus third events on days in which three events occur
R_4	The ratio of the amounts of the sum of the third plus fourth events to the daily total amount on days in which four events occur
R_5	The ratio of the amount of the first event to the sum of the amounts of the first plus second events on days in which four events occur
R_6	The ratio of the amount of the third event to the sum of the amounts of the third plus fourth events on days in which four events occur
s	Dummy variable
t	Time in cumulative hours from an arbitrary t 0
T_c	Ratio of event starting time to 24 h
$U(t)$	Process representing time of day (h)
U_{cl}	The starting time of the lth complete event (h)
U_k	Starting time of kth event in a day (h)
$[x]$	Integer part of the number x
X_n	Total amount of rainfall in the nth event
X_k	Amount of the kth event in a day (mm)
$\hat{X}_k$	Simulated amount of the kth event in a day (mm)
Y_c	The amount of a complete storm event
Y_c' (or y')	Transformed event amount: $Y_c' = Y_c - 0.229$ (mm)
Y_c (or y'')	$\log_{10} y'$
Y_{cn}	The amount of the nth complete event
Y_p	The amount of a partial event
Y_p'	$Y_p - 0.229$ (mm)
Z_c	Daily total rainfall due to complete events
Z_c'	Transformed daily total rainfall for complete events (Z_c' $Z_c - 0.229$ mm)
Z_{ij}	Total rainfall on day j of year i

Greek

α	Parameter in Weibull distribution and beta distribution
γ	Parameter in Weibull distribution
β	Parameter in beta distribution
θ	Parameter in beta Fourier distribution
$\hat{\zeta}(t)$	Simulated rainfall process of storm amounts and durations
$\xi(t)$	Precipitation intensity process
$\hat{\xi}(t)$	Simulated precipitation intensity process
τ_n	Time of beginning of nth event
τ_n^*	Time of ending of nth event
ε	Normally distributed random variable
ω	Weighting parameter in the mixed beta distribution

occurrence model and an intrastorm intensity model, this model enables the simulation of short-time period ("breakpoint") rainfall data.

BACKGROUND

Little research has been carried out on rainfall disaggregation modeling. Betson et al. (1980) described a model to disaggregate daily rainfall into hourly rainfall, and hourly rainfall into 5 min amounts. Their approach requires the estimation of a large number of transition probabilities, however, and does not explore the possibility of a dependency between storm duration and amount, or between the amounts of successive rainfall events on a given day. Valencia and Schaake's linear disaggregation model (1973) does not address the intermittency associated with the daily process and, therefore, is not directly applicable to the disaggregation of daily rainfall. Srikanthan and McMahon (1985) developed a model for generating rainfall at 6 min intervals based upon a daily rainfall model and an hourly model. A time-dependent, two-state second order Markov Chain, conditioned on the amount of daily rainfall, describes the occurrence of hourly rainfall, and up to a seven-state transition probability matrix is used to generate hourly amounts. The 6 min model is conditioned upon four types of wet hours, and requires up to a seven-state transition probability matrix. Both the hourly and 6 min models require separate parameter estimates for each month. The number of parameters required for this approach is very large (5 000–6 000), demonstrating the need for a more parameter-efficient approach.

While storm models based on aggregation techniques are available to simulate short-time period rainfall sequences (Pattison, 1965; Grace and Eagleson, 1966; Todorovich and Yevjevich, 1969; Austin and Claborn, 1974; Raudkivi and Lawgun, 1974; Bras and Rodriguez-Iturbe, 1976; Nguyen and Rousselle, 1981), it would be difficult to adjust the parameters of these models to maintain daily statistics at locations other than those for which the models were developed.

DESCRIPTION OF THE RAINFALL PROCESS

Let $\xi(t)$ denote the continuous process of precipitation intensity at a point in space. $\xi(t)$ will be equal to, or greater than, zero, and will take on positive values over random time intervals (Fig. 1).

If the starting time of the nth shower is denoted by τ_n in cumulative hours from some arbitrary time, $t = 0$, and the ending time by τ_n^*, then the total amount of rainfall in the nth event is X_n, where:

$$X_n = \int_{\tau_n}^{\tau_n^*} \xi(s) \mathrm{d}s$$

The duration of the nth event is $D_n = \tau_n^* - \tau_n$. Define a complete event as one which begins and ends on the same day, and a partial event as one which begins on one day and ends on the next (Fig. 2). Let k represent the total number of events in the day. Then $k = N_c + N_p$, where N_c is the number of complete events, and N_p is the number of partial events ($N_p = 0, 1, 2$). To represent phenomena occurring within the day, let t represent time, in hours, from some arbitrary midnight, $t = 0$. Define the random variable $U(t)$ such that:

$$U(t) = t - \{[t/24]24\}$$

where $[t/24]$ represents the integer part of the number $t/24$. This function represents the beginning time of a shower in hours from midnight and is shown in Fig. 3. Note that, for each time of beginning τ_n, there is a corresponding $U(\tau_n)$, where $0 \leqslant U(\tau_n) \leqslant 24$.

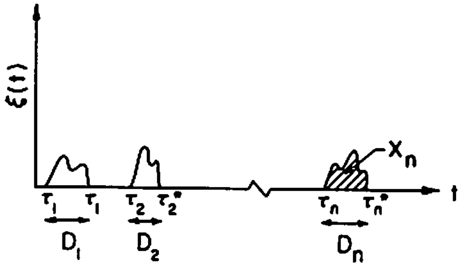

Fig. 1. Possible sample function of the rainfall intensity process.

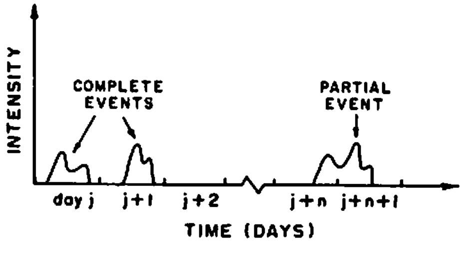

Fig. 2. Definition sketch of complete and partial events.

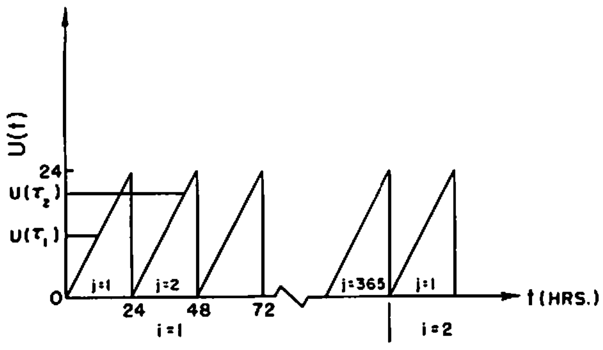

Fig. 3. The function $U(t)$.

The rainfall process is assumed to be stationary, with possible cyclic variations on a seasonal basis. That is, the sample functions of the rainfall process for each year are assumed to be repeated samples of the same underlying process.

THE DISAGGREGATION PROCESS

Let Z_{ij} represent the total rainfall on day j of year i:

$$Z_{ij} = \int_{24(j-1)}^{24_j} \xi(s)ds$$

Then Z_{ij} for $i = 1, 2, 3, \ldots; j = 1, 2, \ldots, 365$ represents a sequence of daily rainfall amounts.

Associated with each day that rain is observed are the random variables Z_{ij}, $N_{c_{ij}}$, $N_{p_{ij}}$, U_k; $k = 1, 2, 3, \ldots, (N_{c_{ij}} + N_{p_{ij}})$, and the amount X_k and duration D_k of each complete or partial event in the day. The model described in this paper provides a method of simulating the number of events per day, as well as the amount, duration, and starting time of each event given a simulated or observed sequence of daily rainfall amounts. The simulated rainfall process $\hat{\zeta}(t)$ will then consist of a sequence of pulses taking on positive values over random time intervals, as shown in Fig. 4. The intensity of each pulse is $\hat{X}_k/\hat{D}_k$, and:

$$\sum_{k=1}^{(N_{c_{ij}}+N_{p_{ij}})} \hat{X}_k = Z_{ij}$$

The final step in the process is to disaggregate the storms into intensity patterns leading to the process $\xi(t)$, as shown in Fig. 5. An intrastorm model that would transform the process $\hat{\zeta}(t)$ into $\hat{\xi}(t)$ for thunderstorm rainfall has been developed by Woolhiser and Osborn (1985). When combined with an intrastorm intensity model, the disaggregation model, described herein, provides a reasonable estimate of short-time period rainfall.

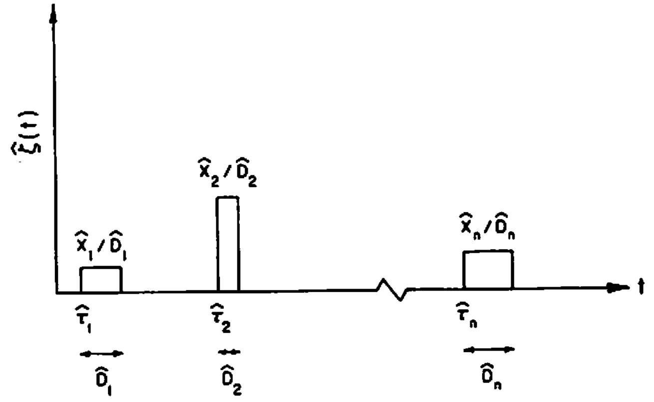

Fig. 4. The simulated storm rainfall process.

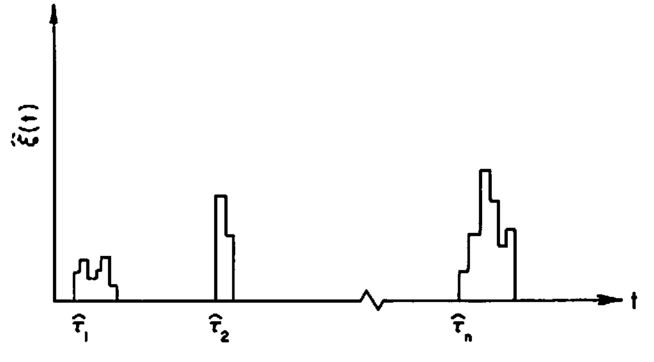

Fig. 5. The simulated rainfall intensity process.

DATA ANALYSIS

Twenty-three years of summer breakpoint data, for the months of July and August, 1954–1976, were obtained from gage number 5, a weighing-recording rain gage on the Walnut Gulch Experimental Watershed operated by the U.S. Department of Agriculture, Agricultural Research Service (USDA-ARS). The watershed is located in southeastern Arizona, and includes the city of Tombstone (Fig. 6). The elevation of gage number 5 is 4200 ft (1280 m).

A shower is arbitrarily defined as any period in which the total rainfall is $\geqslant$0.01 in (0.254 mm), which contains no intervening periods of zero intensity exceeding 10 min in duration. Any period of greater than 10 min, in which no increase in precipitation is measured, signifies an event ending. A day is defined as that 24 h period beginning at midnight. This definition of a shower was chosen so that the model would provide input suitable for simulation on very small watersheds where a 10 min cessation of rainfall is highly significant. It should be noted that independence of shower amounts is not required in this model.

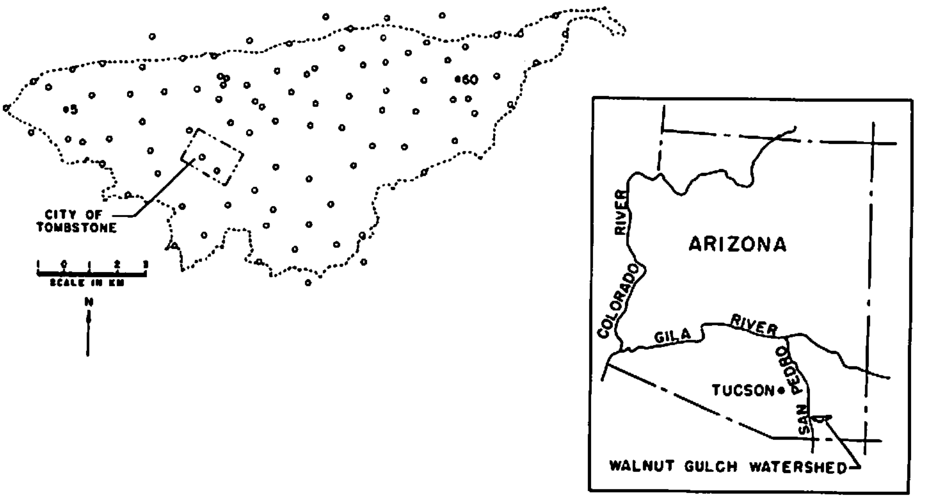

Fig. 6. Location map of the Walnut Gulch Experimental Watershed and rain gage locations.

The data set consisted of 695 storms. Of these, 659 (94.82%) were complete events, and 36 (5.18%) were partial events. Some general statistics are presented in Table 1. As the two groups differ statistically, they were analyzed separately.

MODEL OF COMPLETE EVENTS

The 659 complete events were distributed over 450 rainy days. The majority (308) occurred as a single storm in a day, and the largest number of occurrences observed in a day was six.

TABLE 1

Statistics on duration and amount for summer thunderstorms on the Walnut Gulch Watershed (23 years of data)

Type of event:	Complete		Partial	
Number:	659		36	
	in	mm	in	mm
Amount				
Mean	0.199	5.06	0.442	11.24
Standard deviation	0.307	7.80	0.413	10.48
Range	0.01–2.99	0.25–75.94	0.02–1.36	0.50–34.54
Duration (min)				
Mean	66.38		226.36	
Standard deviation	71.26		157.42	
Range	1.99–510.0		70.02–668.99	

Joint distribution of number of events per day and daily amount

The joint distribution of the number of events per day, and the daily amount, can be written as a product of the conditional and marginal distributions:

$$H_{N_c, Z_c'}(n, z) = G_{N_c|Z_c'}(n|z) F_{Z_c'}(z) \tag{1}$$

As the lower threshold of observation is 0.01in (0.254 mm), all calculations were performed on a transformed daily amount, $Z_c' = Z_c - 0.009$ ($Z_c - 0.229$ mm). The value of 0.009 was used, instead of 0.01, because it was desired to maintain these values in the data set as occurrences of small, positive rainfall amounts. Five distributions were compared for the marginal distribution of daily amount, $F_{Z_c'}(z)$: the exponential, lognormal, Weibull, mixed exponential (Rider, 1961) and beta-kappa (Mielke and Johnson, 1974). Parameters were estimated using analytic or numerical maximum likelihood techniques. The Weibull distribution had the maximum likelihood function value and minimum Akaike Information Criterion, or AIC (Akaike, 1974), and was selected as the best choice. The marginal distribution of daily rainfall amount is written as:

$$F_{Z_c'}(z) = 1 - \exp\left[-\left(\frac{z}{\alpha}\right)^{\gamma}\right] \tag{2}$$

where $\alpha = 0.2399$ in (7.617 mm), $\gamma = 0.7364$, and $z \geqslant 0$. A chi-squared goodness-of-fit test indicates that this hypothesis cannot be rejected at the 95% level. This significance level is approximate, because the data were used to choose the distribution function, and, therefore, some uncertainty is involved in estimating the degree of freedom.

It now remains to determine an expression for the conditional distribution of number of events per day given a daily amount, $G_{N_c|Z_c'}(n|z)$. The truncated Poisson (truncated at one, as we are not interested in the occurrence of zero events) and geometric distributions were tried. Both tended to overpredict the occurrence of one, and underpredict the occurrence of two events per day. However, the geometric worked well in the range of lower rainfall amounts; thus, it is desirable to find a distribution that approaches the geometric at the lower limit. The shifted negative binomial distribution (SNBD) possesses this property (Buishand, 1977). The probability mass function is written as:

$$P(N_c = n) = \binom{n + r - 2}{n - 1} p^r (1 - p)^{n-1}, \quad n = 1, 2, \ldots \tag{3}$$

When $r = 1$, this reduces to the geometric distribution.

The parameters p and r were allowed to vary with daily amount. The lower limits were set such that, given the minimum observable daily rainfall amount (0.01 in, or 0.254 mm), the probability of the occurrence of one event in that day is a certainty. To determine the upper limits, p and r were estimated over the two classes of highest daily rainfall amount using numerical maximum likelihood techniques. An investigation showed that exponential curves are appropriate for points between, and these parameters were estimated using a modified version of Rosenbrock's parameter optimization method (Palmer, 1969; Fig. 7). The final forms for p and r as functions of daily rainfall amount z (mm) are:

$$\left.\begin{aligned} p &= 0.7228 + 0.2772\exp(-0.2281\,z) \\ r &= 2.3097 - 1.3097\exp(-0.3776\,z) \end{aligned}\right| \quad (4)$$

and the conditional probability is written as:

$$P(N_c = n|Z = z) = \binom{n+r-2}{n-1} p^r(1-p)^{n-1}, \quad n = 1, 2, \ldots \quad (5)$$

with parameters as given in eqn. (4).

With the functional forms for p and r given by eqn. (4), the expected value function $E\{N_c/Z\}$ begins at one for $z = 0$ and asymptotically approaches a constant. Although this form is satisfactory for the Walnut Gulch rainfall data an analysis of data from three midwestern states shows that this is a special case and the more general form of $E\{N_c|Z\}$ would asymptotically approach a straight line with positive slope.

Utilizing eqns. (1)–(3), the joint distribution function of number of storms and transformed daily amount can be written as:

$$H_{N_c,Z_c}(n, z) = \sum_{i=1}^{n} \int_0^z \gamma/\alpha(z/\alpha)^{\gamma-1}\exp[-(z/\alpha)^\gamma] \binom{n+r-2}{n-1} p^r(1-p)^{n-1}\mathrm{d}z \quad (6)$$

where p and r are given by expression (4). Equation (6) was integrated numerically to obtain theoretical probabilities for 19 joint classes of N_c and Z_c'. A chi-squared goodness of fit test indicated that the null hypothesis could not be rejected at the 0.05 level:

$$\chi^2 = 15.0365\,(\chi^2_{0.05,10} = 18.307)$$

A comparison of the histograms of the marginal distribution of number of

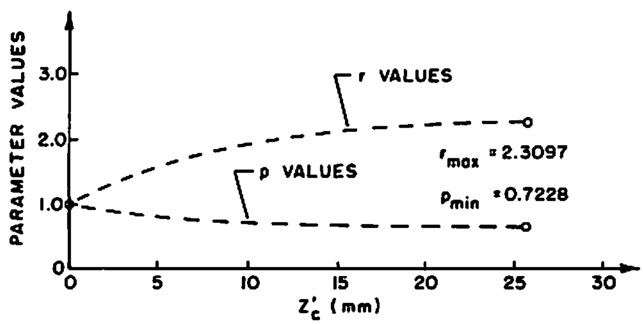

Fig. 7. Functional relationship between the parameters in the shifted negative binomial distribution and the transformed daily rainfall, Z_c'.

events per day also showed close agreement between the observed and simulated data.

Disaggregating daily rainfall into individual storm amounts

Given a daily rainfall amount Z_c and the number of storms N_c where $N_c = 2, 3, \ldots, 6$, the next step is to disaggregate Z_c into N_c amounts of $Y_{c_1}, Y_{c_2}, \ldots, Y_{c_{N_c}}$ such that:

$$\sum_{i=1}^{N_c} Y_{c_i} = Z_c \tag{7}$$

Note that if only one rainfall event occurs, further storm event disaggregation is unnecessary.

Define the random variables R_1 through R_2 as follows:

(a) two events per day:

$$R_1 = \frac{\text{amount of first event}}{\text{daily total}} = \frac{Y_{c_1}}{Z_c} \tag{8}$$

(b) three events per day:

$$R_2 = \frac{Y_{c_2} + Y_{c_3}}{Z_c}, \quad R_3 = \frac{Y_{c_2}}{Y_{c_2} + Y_{c_3}} \tag{9}$$

(c) four events per day:

$$R_4 = \frac{Y_{c_3} + Y_{c_4}}{Z_c}, \quad R_5 = \frac{Y_{c_1}}{Y_{c_1} + Y_{c_2}}, \quad R_6 = \frac{Y_{c_3}}{Y_{c_3} + Y_{c_4}} \tag{10}$$

Distributions are fit to the cumulative distribution functions (CDF's) of the storm ratios defined above. To simplify matters, it is assumed that R_2 and R_3 are independent for days with three events, and $R_2, R_3 \ldots R_6$ are independent for days with four events. While this assumption may not be strictly valid, it makes the problem tractable, and gives reasonable results. Note that although the ratios defined above are assumed to be independent, the showers obtained by the disaggregation process will be dependent due to eqns. (7)–(10).

A beta distribution was fit to the CDF's of R_1 and R_2 using maximum likelihood techniques. The fit, while satisfactory, was significantly improved by the addition of a Fourier series term:

$$f_R(r) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1 - r)^{\beta-1} + \theta \sin 2\pi r, \quad \alpha, \beta > 0, \quad 0 \leqslant r \leqslant 1 \tag{11}$$

θ must be defined so that:

$$|\theta \sin 2\pi r| < \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1 - r)^{\beta-1}$$

Under this constraint $f_R(r) \geqslant 0$, $0 \leqslant r \leqslant 1$. This distribution has three parameters, α, β, and θ, and will be referred to as the beta-Fourier. The likelihood ratio test (Hoel, 1971) was used to test the hypothesis of a common parameter set for $f_{R_1}(r)$ and $f_{R_2}(r)$. At $\alpha = 0.05$, the hypothesis is not rejected.

Thus, the density functions of R_1 and R_2 can be described by eqn. (11) with parameters as follows: $\alpha = 1.2514$, $\beta = 0.9045$, and $\theta = 0.0819$. The CDF is obtained by numerically integrating eqn. (11). It was also hypothesized that the CDF's of the random variables R_3, R_4, R_5, and R_6 are uniform. In each case, at $\alpha = 0.05$, a Kolmogorov–Smirnov (K–S) test indicates the hypothesis cannot be rejected.

Due to a lack of data for the classes of five and six events per day (three and one occurrences, respectively), some ad hoc assumptions are necessary to describe the disaggregation process. The first is that, on days in which five events occur, the distribution of the ratio of the sum of the amounts of the first and second events to the daily total is uniform. It is further assumed that the sum of the first and second amounts can be disaggregated in the same manner as the first and second amounts on days in which two events occur, and that the third, fourth, and fifth amounts follow the scheme developed for the first, second, and third amounts on days in which three events occur. For days in which six events occur, it is assumed that the distribution of the ratio of the sum of the first three events to the daily total amount is uniform. Each set of three events is then disaggregated according to the scheme already developed for three events per day. A maximum of six events was observed in any day at Walnut Gulch gage number 5, so there is no historical basis for disaggregating Z_c into more than six events. In Monte Carlo simulations from eqn. (5), if $N_c > 6$ was generated, it was set equal to 6.

The joint distribution of event duration and amount

It is now necessary to simulate the durations $D_{c_1}, D_{c_2}, \ldots, D_{c_{N_c}}$ associated with each event amount, $Y_{c_1}, Y_{c_2}, \ldots, Y_{c_{N_c}}$. The joint density function of event amount and duration is written as a product of the conditional and marginal density functions:

$$h_{Y_c,D_c}(y, d) = g_{D_c|Y_c}(d\,|\,y) f_{Y_c}(y) \tag{12}$$

where Y_c' is the transformed event amount: $Y_c' = Y_c - 0.009(Y_c - 0.229\,\text{mm})$. The random variable Y_c is obtained from the storm ratios and the daily rainfall amounts. In identifying a form for the conditional distribution, it is assumed that the distribution of the duration, given an amount, is the same for all complete events.

Define two new random variables: let $y'' = \log_e y'$, and $d' = \log_e d$. Assume that the conditional density of d', given y'', is normal, with an expected value function which is a linear function of y'':

$$E[d'|y''] = A + By'' \quad (13)$$

The following parameter estimates were obtained by linear regression: $A = 4.646$ (d in inches); $A = 3.415$ (d in mm); $B = 0.3785$; standard error of estimate = 0.8885; and $r^2 = 0.2478$. A correlation ratio test (Kendall and Stuart, 1979) was used to test the hypothesis of linearity of the regression. The hypothesis cannot be rejected ($\alpha = 0.05$). To test the hypothesis that the conditional density of d', given y'', is normal, the values of the y'' were separated into four classes based on magnitude, and a chi-squared test was run on the residuals, $d' - E[d'|y'']$, for each class. The hypothesis could not be rejected ($\alpha = 0.05$) in each case.

Thus, the conditional density is written as:

$$g_{D_c|Y_c}(d'|y'') = \begin{cases} 0.3785y'' + 4.646 + \varepsilon(0, 0.8885)\,(\text{in}) \\ 0.3785y'' + 3.415 + \varepsilon(0, 0.8885)\,(\text{mm}) \end{cases} \quad (14)$$

A transformation is performed to obtain the duration D_c as follows:

$$D_c = \exp(D_c') \quad (15)$$

Distribution of event starting times

The time interval between storms is controlled by the time distribution of event occurrence within the day, and is an important input to infiltration models, because antecedent soil moisture conditions affect infiltration rates. Let U_{c_l}, $l = 1, 2, \ldots, N_c$, represent the starting time of the lth complete event in a day in military hours. Then the ratio:

$$T_c = \frac{U_{c_l}}{24} \quad (16)$$

takes on the values zero through one, inclusive, representing this starting time. As summer thunderstorm rainfall in southeast Arizona exhibits a high frequency of occurrence in the late afternoon and just before midnight, a flexible distribution is needed to describe event starting times. The beta distribution was tried, but the fit is highly unsatisfactory (Fig. 8), and it appears that more than two parameters will be needed. The mixed beta distribution has five parameters, and is written as:

$$F_{T_c}(t) = \omega\left[\frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)}\, t^{\alpha_1 - 1}(1 - t)^{\beta_1 - 1}\right] + (1 - \omega)\left[\frac{\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)}\, t^{\alpha_2 - 1}(1 - t)^{\beta_2 - 1}\right],$$

$$\alpha_1, \alpha_2, \beta_1, \beta_2, > 0;\ 0 < \omega < 1;\ 0 \leqslant t \leqslant 1 \quad (17)$$

Parameter estimates were obtained for the class of all complete event starting times using numerical maximum likelihood techniques, (Hershenhorn, 1984)

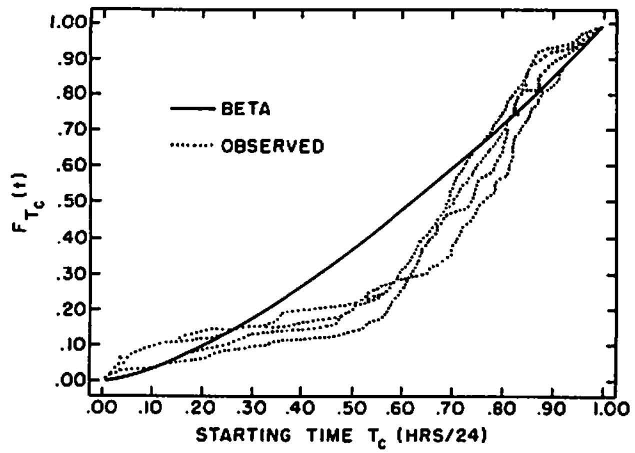

Fig. 8. Best fit of beta distribution to storm starting times.

and are as follows:

$\alpha_1 = 0.6389$

$\beta_1 = 3.2895$

$\alpha_2 = 6.2318$

$\beta_2 = 2.3816$

$\omega = 0.1483$

The fit obtained is much better (Fig. 9).

We wish to test the hypothesis H_0 that the mixed beta distribution describes the observed CDF's of starting times for the classes of one, two, three, and four

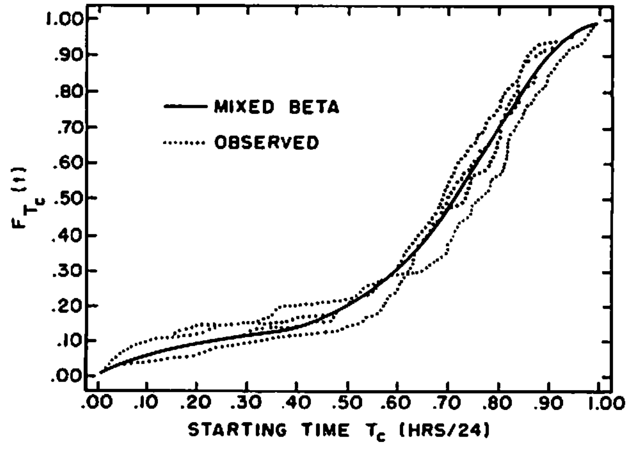

Fig. 9. Best fit of mixed beta distribution to storm starting times.

events per day. The hypothesis could not be rejected for two, three, and four events per day using the K–S goodness-of-fit test at $\alpha = 0.05$. Although the hypothesis is rejected at this level for the class of one event per day, it is not rejected at $\alpha = 0.01$. It was decided not to reject H_0 under these conditions.

The second hypothesis H_0 we wish to test is that on a day in which n events occur, the starting times are independent, and can be obtained using order statistics. Let $F_{T_r}(t)$ represent the CDF of the starting time of the rth event on a day in which n events occur. The general form is:

$$F_{T_r}(t) = \frac{n!}{(r-1)!(n-r)!}\left[\sum_{i=0}^{n-r} \frac{\binom{n-r}{i}(-1)^i}{i+r}(F_{T_c})^{i+r}\right] \tag{18}$$

where F_{T_c} is the CDF of all event starting times as given by eqn. (17) (Kendall and Stuart, 1979, p. 347).

Given that n events occur in a day, it is only necessary to generate n values of the random variable r distributed according to eqn. (17), order these from lowest to highest, and multiply by 24 to get the appropriate starting times. The empirical distribution functions were compared with the theoretical distribution given the assumption of independence, eqn. (18). See Table 2 for mathematical expressions for the distributions of starting times obtained using order statistics.

K–S goodness-of-fit tests were run to test the hypothesis that event starting times are independent. With the exception of the first event on days in which three events occur, this hypothesis is not rejected at $\alpha = 0.05$. However, the hypothesis is not rejected at $\alpha = 0.01$ for the first event of three per day. Therefore, it was decided not to reject the hypothesis.

Not only is there a diurnal variation in the time of occurrence of showers, there also is a diurnal variation in the expected depth of a shower given the time of occurrence. This variability is not accounted for explicitly in the

TABLE 2

Distributions of event starting times obtained by using order statistics

Number of events per day	Order of occurrence	Cumulative distribution function
2	1st	$F_{T1} = 2F - F^2$
	2nd	$F_{T2} = F^2$
3	1st	$F_{T1} = 3F - 3F^2 + F^3$
	2nd	$F_{F2} = 3F^2 - 2F^3$
	3rd	$F_{T3} = F^3$
4	1st	$F_{T1} = 4F - 6F^2 + 4F^3 - F^4$
	2nd	$F_{T2} = 6F^2 - 8F^3 + 3F^4$
	3rd	$F_{T3} = 4F^3 - 3F^4$
	4th	$F_{T4} = F^4$

F is F_{T_c}, as given in eqn. (16).

procedure described above. Some variability in depth given time of occurrence is implicit for days with two or more showers. For example, the distribution of R_1 (eqn. 11) was not symmetrical. Therefore, the distributions of the first and second showers of the day given the daily rainfall are not identical.

MODEL OF PARTIAL EVENTS

It has been shown that the occurrence of a partial event at the end of a day is independent of the sum and product of daily amounts on two consecutive wet days for summer thunderstorm data from gage 5 at Walnut Gulch (Hershenhorn, 1984). Let $P\{M_j = m\}$ be the probability mass function of the random variable M_j where:

$$m = \begin{cases} 0, & \text{if no partial event occurs at the end of the day } j \\ 1, & \text{if a partial event occurs at the end of day } j \end{cases}$$

As 36 of the 217 observed W–W (wet–wet) day transitions contained partial events, $P\{M = 1\} = 36/217 = 0.1659$ given two consecutive rainy days. The number of partial events in a day is limited to zero, one, or two.

The mean amount of a partial event (0.4425 in, or 11.25 mm) is more than twice as great as the mean amount of a complete event (0.1992 in, or 5.06 mm). However, each partial event can be separated into two events based on day of occurrence, producing a new set of 72 events with a smaller mean.

If Y_p represents the amount of a partial event (after separation), then let $Y_p' = Y_p - 0.009$ represent the transformed event amount. The likelihood ratio test is u sed to test the hypothesis, H_0, that the marginal distribution of event amount for partial events, $F_{Y_p'}(y)$, is not different from the marginal distribution of event amount for complete events, $F_{Y_c'}(y)$. At $\alpha = 0.05$, H_0 is not rejected, and it is assumed that the random variable Y_p can be obtained from the storm ratios in the same manner as the amounts for complete events, Y_c see eqns. (8)–(10).

The joint density function of event amount and duration is also obtained in the same manner as for complete events. A linear regression between $\log_e(D_p)$ and $\log_e(Y_p')$, where D_p is the duration of the partial (separated) event yielded the following statistics: slope = 0.3296; intercept = 5.1624 (Y_p' in inches) (4.096 for Y_p' in mm); standard error of estimate = 0.7755; and $r^2 = 0.3066$. A correlation ratio test was run to test the hypothesis of linearity of the regression. At $\alpha = 0.05$, the hypothesis cannot be rejected. The residuals were tested for normality about the regression, and a chi-squared test indicates this hypothesis is not rejected.

The conditional density function is written as follows:

$$g_{\log_e D_p | \log_e Y_p'}(\log_e d \,|\, \log_e y') = 5.1624 + 0.3296 \log_e Y_p' + \varepsilon(0, 0.7755)$$

A transformation is performed to obtain D_p. A partial event has been defined as either ending or beginning at midnight, so once the duration has been simulated, the time of occurrence within the day is specified.

SIMULATION

Input to the simulation program consists of a sequence of daily rainfall amounts. For each rainy day j the number of events N and the amount, duration, and starting time of each event are simulated.

The method examines rainfall amounts on days $j - 1$, j and $j + 1$, simultaneously. If days $j - 1$ and j are wet and day $j + 1$ is dry, Monte Carlo techniques are used to simulate partial event occurrence at the end of day $j - 1$, and the results are stored. If all three days are wet, a partial event may also occur at the end of day j with probability 0.1659 unless the daily rainfall was 0.01 inch (0.254 mm), in which case only one shower can occur. Next, the *total* number of events on day j (NUM_T) is generated according to the SNBD, eqn. (3). The number of complete events is then either $NUM_c = NUM_T - NUM_p$, where NUM_p is the number of partial events simulated on day j if $NUM_T > NUM_p$, or zero. If NUM_T was smaller than NUM_p another random number was generated until $NUM_T \geqslant NUM_p$. Thus NUM_T was distributed as the probability mass function $P\{N = n|Z = z \text{ and } n \geqslant NUM_p\}$. Although the parameters for the SNBD were estimated using only the complete storm set, it should be a good approximation if partial storms were included as well because of the small number of days with partial storms (36). This small sample size precluded a separate estimate of parameters for days with partial storms.

For each of the NUM_T events on day j, the storm ratios defined in eqns. (8)–(10) are used to disaggregate the daily amount into individual storm amounts. Equation (14) is then used to generate an associated storm duration for complete storms, and the starting times are simulated from the mixed beta distribution, eqn. (17), and ordered accordingly. The form of eqn. (14), with appropriate parameters, is used to generate durations for partial storms. In each case, the inverse transformation method is applied to transform a uniform random variable to a variable with the desired distribution.

The program was run on a Digital Equipment Corporation PDP-11/34 computer at the USDA-ARS Southwest Rangeland Watershed Research Center in Tucson. It took 130 s of CPU time to simulate 23 yr of summer storm data for the months of July and August. The observed and simulated data are shown in Figs. 10–12. Two sample Kolmogorov–Smirnov tests indicate that the hypothesis that the observed and simulated data come from the same population can not be rejected at $\alpha = 0.05$ (Benjamin and Cornell, 1970).

EXTENSION OF THE MODEL TO OTHER LOCATIONS

We have shown that various properties of the point rainfall process can be modeled, and that simulations yield rainfall sequences maintaining the desired properties, for that location. To be useful, however, two additional criteria must be satisfied: (1) the model must be applicable over climatologically similar areas; and (2) the simulated rainfall obtained using this model, and an intra-storm intensity model, must yield runoff hydrographs which more closely resemble observed hydrographs than those obtained using other types of

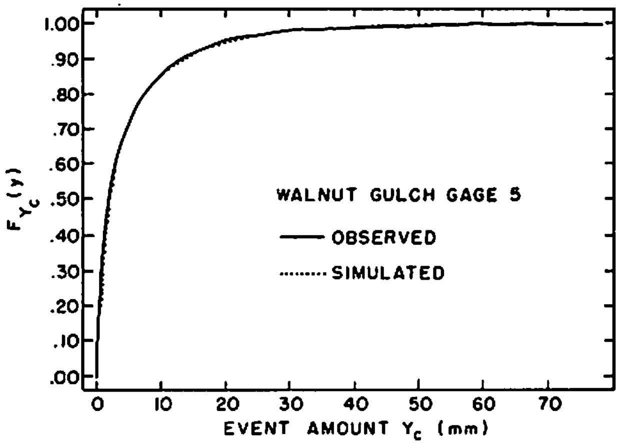

Fig. 10. Observed and simulated distributions of storm rainfall amounts for complete events.

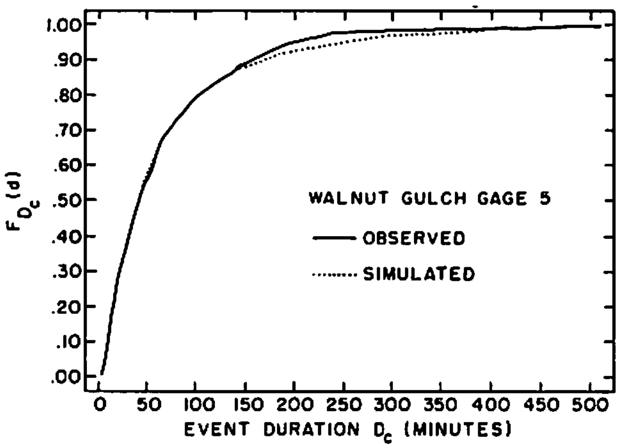

Fig. 11. Observed and simulated distributions of storm duration for complete events.

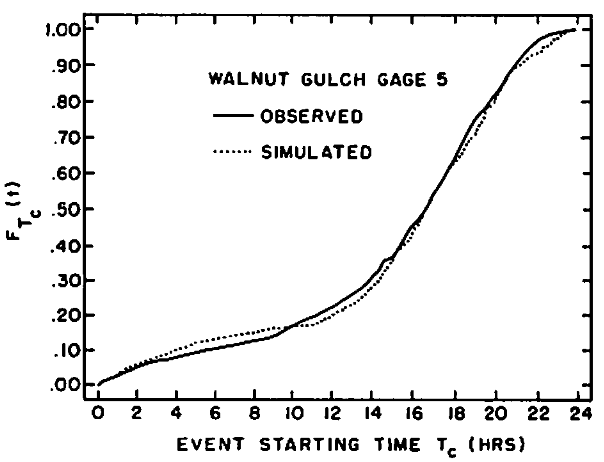

Fig. 12. Observed and simulated distributions of event starting time for complete events.

simulated rainfall input. Research has been carried out concerning the extension of the model to other locations, and the results are presented below. Further research to measure runoff hydrograph sensitivity is underway.

Application of the model to Walnut Gulch rain gage no. 60

Gage no. 60 is located approximately 10.5 mi (16.9 km) east of gage no. 5, at an elevation of 4990 ft (1520 m). Osborn, Renard, and Simanton (1979) have shown that storm rainfall data obtained from gages located greater than 3.7 mi (6 km) apart on the watershed are independent. Disaggregation model parameters, estimated from breakpoint data at gage no. 5, were used with observed daily rainfall at gage no. 60 to simulate storm amounts, durations, and starting times.

A comparison of percentages of single and multiple-event days to total rainy days shows good agreement between the observed and simulated data. The observed and simulated distributions of event amount, duration, and starting time are also in close agreement (Figs. 13–15). In each case, the hypothesis of a common population for the observed and simulated data cannot be rejected using a K–S two-sample test at $\alpha = 0.05$.

Application of the model to Safford gage no. 3

The average annual precipitation at Safford, Arizona, located approximately 75 mi (121 km) north of Walnut Gulch at an elevation of 3350 ft (1021 m), is 7.17 in (182 mm), as compared with 11.0 in (279 mm) at Walnut Gulch gage no. 5. In this case, a comparison of percentages of single and multiple-event days to total rainy days shows that a greater number of multiple-event days was simulated than was observed. Also, the hypotheses of common observed and simulated distributions of amount and duration for the entire set of daily data were rejected using the two-sample K–S test at $\alpha = 0.05$.

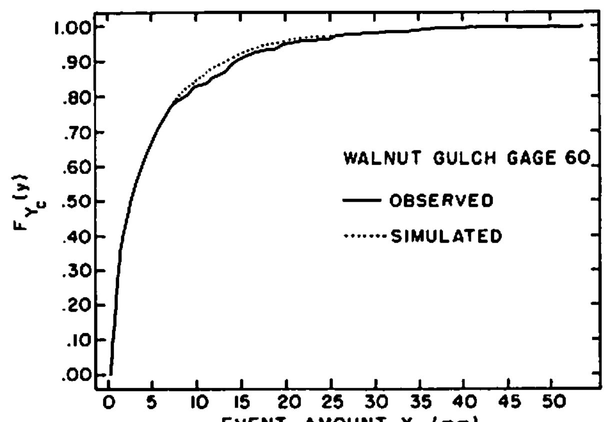

Fig. 13. Observed and simulated distributions of storm rainfall amount for complete events, Walnut Gulch gage 60.

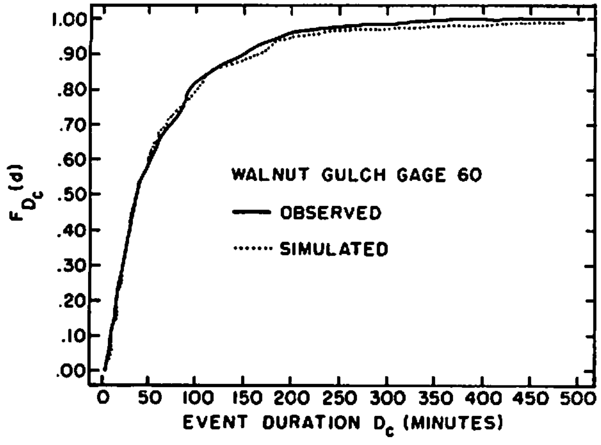

Fig. 14. Observed and simulated distributions of event duration for complete events, Walnut Gulch gage 60.

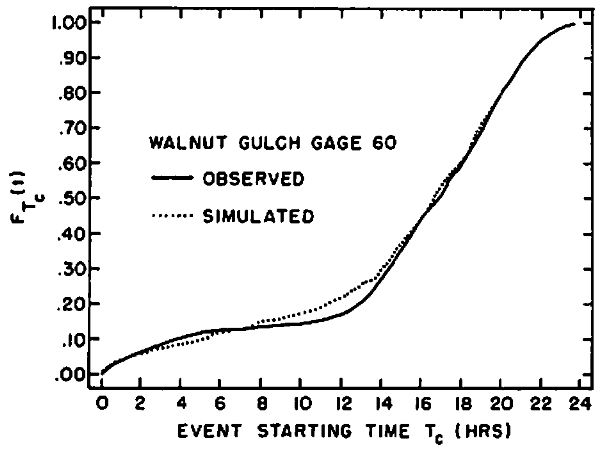

Fig. 15. Observed and simulated distributions of event starting time for complete events, Walnut Gulch gage 60.

Because there is little runoff associated with events in which the total rainfall is less than 0.25 in (6.35 mm), comparisons were also made using subsets of observed daily data for days in which the rainfall total equaled or exceeded this threshold. A two-sample K–S test at $\alpha = 0.05$ indicates that hypotheses of common populations for the observed and simulated distributions of event amount, duration, and starting time for days in which the total rain was > 0.25 in cannot be rejected (Figs. 16–18).

Application of the model to Santa Rita gage no. 2

Santa Rita gage no. 2 is located at the Santa Rita Range Experiment Station, located 31.5 mi (50.7 km) west of the Walnut Gulch Experimental Watershed at

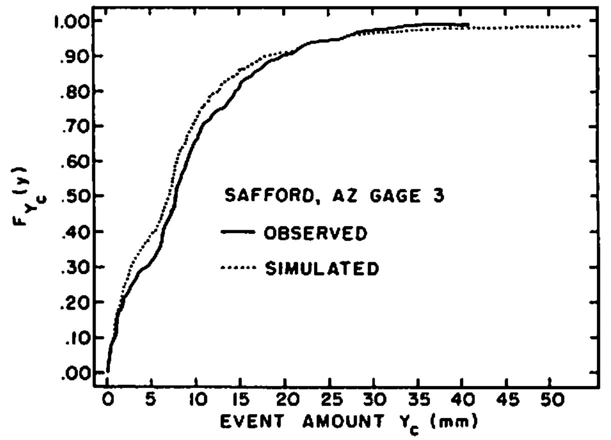

Fig. 16. Observed and simulated distributions of storm rainfall amount for complete events, Safford, AZ gage 3.

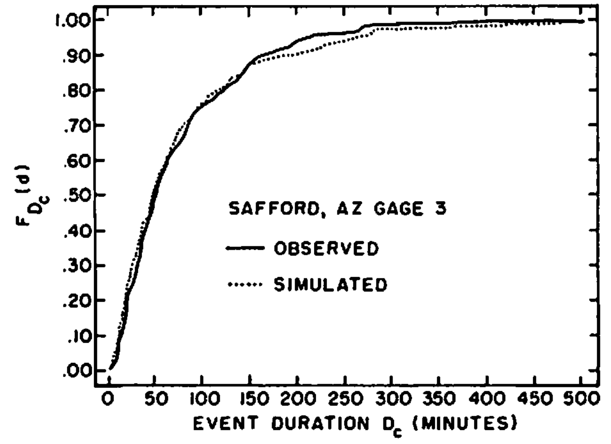

Fig. 17. Observed and simulated distributions of storm duration for complete events, Safford, AZ gage 3.

an elevation of 3400 ft (1036 m). The average annual precipitation is 14.35 in (364 mm). Data were available for the 8 yr period 1975–1982. Simulated data, using the parameter estimates from Walnut Gulch gage no. 5, were compared with the observed data (Figs. 19–21). The distribution of the number of simulated storms per day was very close to the observed distribution, and hypotheses of common distributions for the observed and simulated event amounts, durations, and starting times cannot be rejected at $\alpha = 0.05$ using the two-sample K–S test.

DISCUSSION AND CONCLUSIONS

The objective of this research was to develop a parsimonious model to simulate the number of storm rainfall events per day and the amount, duration,

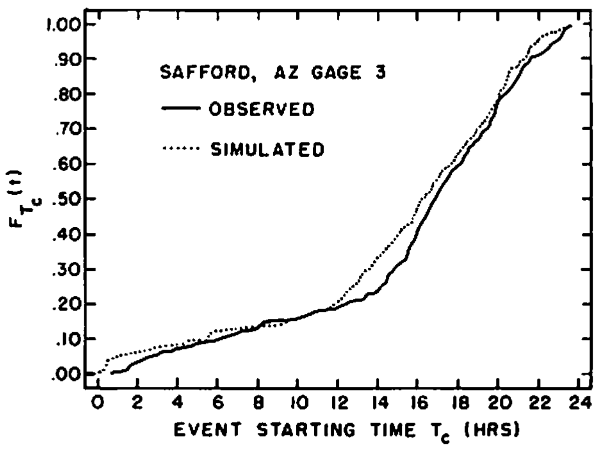

Fig. 18. Observed and simulated distributions of event starting time for complete events, Safford, AZ gage 3.

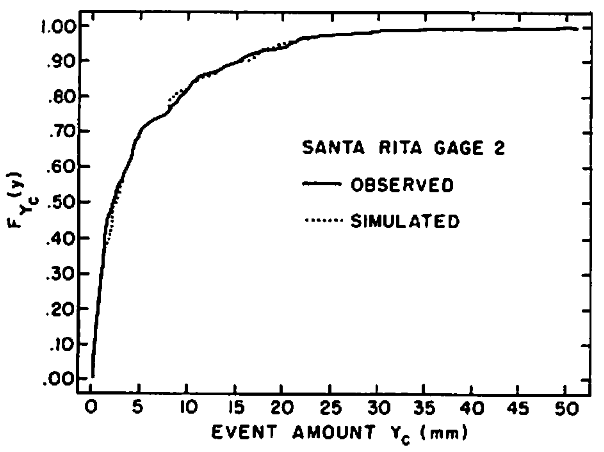

Fig. 19. Observed and simulated distributions of event amounts, Santa Rita gage 2.

and starting time of each event, given only the daily rainfall amount and the amounts on the preceding and following days. A new disaggregation methodology is proposed to accomplish this objective.

Simulated distributions of number of storms per day and storm amounts, durations, and starting times compared well with observed data for Walnut Gulch gage no. 5. The spatial invariance of the disaggregation model structure and parameter values was investigated by disaggregating daily data for three additional stations using parameter values identified for Walnut Gulch gage no. 5 and comparing the simulated distributions with the distributions of measured data. For Walnut Gulch gage no. 60 and Santa Rita gage no. 2, the simulated distributions also compared well with the observed distributions. Null hypotheses of common distributions could not be rejected at $\alpha = 0.05$ by the two-sample K–S test. These null hypotheses were rejected for Safford gage

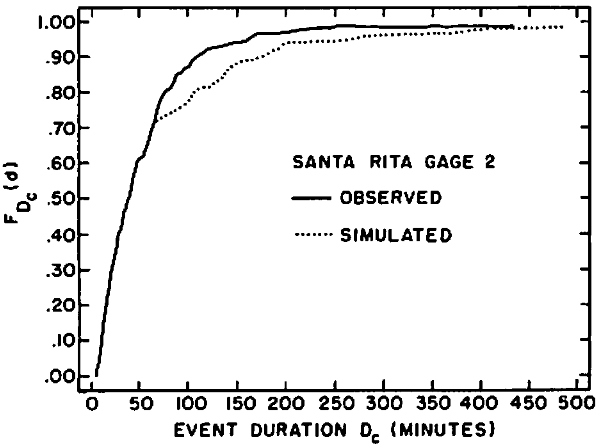

Fig. 20. Observed and simulated distributions of event durations, Santa Rita gage 2.

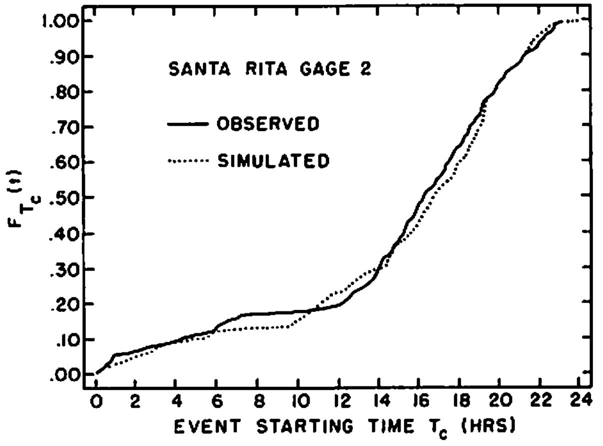

Fig. 21. Observed and simulated distributions of event starting times, Santa Rita gage 2.

no. 3 data, but could not be rejected if the comparison was restricted to days with rainfall ≥ 0.25 in (6.35 mm). Although this does not constitute a conclusive demonstration of low spatial variability of disaggregation structure and parameter values, the results are encouraging, because the differences between stations in elevation of approximately 800 ft (244 m), and in mean annual rainfall of over 3 in (76 mm), are greater than would be encountered over rather large areas in less-mountainous regions.

Some limitations and inconsistencies in the model are apparent. First, the distribution of the number of events per day, and the sums of event amounts, should be consistent with the distribution of the daily amount. That is, in principle, it should be possible to derive the distribution of daily amounts Z_{ij} from the distributions of the random variables N_{cij}, N_{pij}, U_k, and X_k, $k = 1$, 2, . . . Although we attempted this approach, the mathematical problem was intractable. Also the procedure cannot preserve the conditional distribution of shower depth, given daily rainfall amount, for days with only one shower and does not explicitly preserve this distribution for multi-event days.

Considering the complex nature of the rainfall process, the disaggregation model described in this paper is parsimonious. However, the model could be simplified considerably if disaggregation is attempted only for days with rainfall above some threshold greater than 0.01 in (0.254 mm). For example, an examination of Fig. 7 reveals that the parameters r and p, in the shifted negative binomial distribution, are only weakly dependent on daily rainfall amounts over 0.5 in (12.7 mm), suggesting that the number of parameters in the conditional probability mass function of the number of events, given daily amount, could be reduced from 6 to 2. Research on the sensitivity of derived quantities, such as peak runoff rate or volumes to the structure of disaggregation models, may lead to further simplifications. For example, does the time distribution of storm occurrence within the day have a significant effect on calculated runoff? If not, the distribution of starting times might be approximated with a uniform distribution, reducing the number of parameters by 5.

Further research is underway to examine regional and seasonal variations in disaggregation structure and parameters, and will be reported in a subsequent paper. However, it has been found that the general approach described here is appropriate for more humid regions.

ACKNOWLEDGEMENTS

We thank Fatima Lopez for her assistance with the precipitation data.

REFERENCES

Akaike, H., 1974. A new look at the statistical model identification, IEEE Trans. Autom. Control, 19(6): 716–723.

Austin, T.A., and Claborn, B.J., 1974. Statistical model of short duration precipitation events. Proc. Symp. Statistical Hydrology, Tucson, Ariz., U.S. Dep. Agric., Misc. Publ. No. 1275, pp. 147–153.

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

Betson, R.P., Bales, J. and Pratt, H.E., 1980. User's Guide to TVA-HYSIM. U.S. EPA, EPA-600/7-80, pp. 40–44.

Bras, R.L., and Rodriguez-Iturbe, I., 1976. Rainfall generation: A nonstationary time-varying multidimensional model. Water Resour. Res., 12(3): 450–456.

Buishand, T.A., 1977. Stochastic modeling of daily rainfall sequences. Meded. Landbouwhogesch., Wageningen, 77-3.

Grace, R.A., and Eagleson, P.S., 1966. Synthesis of short-time increment rainfall sequences. Hydrodyn. Lab. Rep. No. 91, Mass. Inst. Technol.

Hershenhorn, J.S., 1984. Disaggregation of daily rainfall. M.S. Thesis, Univ. of Ariz., Tucson, Ariz., 146 pp.

Hoel, P.G., 1971. Introduction to Mathematical Statistics. Wiley New York, N.Y., 409 pp.

Kendall, M., and Stuart, A., 1979. The Advanced Theory of Statistics. Vol. 2, Griffin, London, 748 pp.

Mielke, Jr., P.W., and Johnson, E.S., 1974. Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology. Water Resour. Res., 10(2): 223–226.

Nguyen, V., and Rousselle, J., 1981. A stochastic model for the time distribution of hourly rainfall depth. Water Resour. Res., 17(2): 394–409.

Osborn, H.B., Renard, K.G. and Simanton, J.R., 1979. Dense networks to measure convective rainfall in the southwestern United States. Water Resour. Res., 15(6): 1701–1711.

Palmer, J.R., 1969. An improved procedure for orthogonalizing the search vectors in Rosenbrock's and Swann's direct search optimization methods. Comput. J., 12: 69–71.

Pattison, A., 1965. Synthesis of hourly rainfall data. Water Resour. Res., 1(4): 489–498.

Raudkivi, A.J., and Lawgun, N., 1974. Simulation of rainfall sequences. J. Hydrol., 22: 271–294.

Rider, P.R., 1961. The method of moments applied to a mixture of two exponential distributions. Ann. Math. Stat., 32: 143–147.

Srikanthan, R., and McMahon, T.A., 1985. Stochastic generation of rainfall and evaporation data. Aust. Water Resour. Counc. Tech. Pap. No. 84, 301 pp.

Todorovic, P., and Yevjevich, V., 1969. Stochastic process of precipitation. CSU Hydrol. Pap. No. 35, Colo. State Univ., Ford Collins, Colo.

Valencia, D.R., and Schaake, Jr., J.F., 1973. Disaggregation processes in stochastic hydrology. Water Resour. Res., 9(3): 580–585.

Woolhiser, D.A., and Osborn, H.B., 1985. A stochastic model of thunderstorm rainfall intensities. Water Resour. Res., 21(4): 511–522.