

# Point Processes of Seasonal Thunderstorm Rainfall

## 2. Rainfall Depth Probabilities

R. E. SMITH AND H. A. SCHREIBER

USDA Southwest Watershed Research Center, Agricultural Research Service, Tucson, Arizona 85705

As a sequel to an earlier study of the stochastic properties of daily rainfall occurrence in a summer thunderstorm season, the probabilistic nature of daily rainfall depths is examined. Daily rainfall depths are found to be well described by what may be termed a compound exponential distribution. Some 1600 to over 2000 daily rainfalls from three stations are partitioned in various conceptual ways to evaluate homogeneity with respect to the pattern of occurrence with sequences or seasons as well as the annual variance of population properties. Although some small but statistically significant inhomogeneities are found, the statistical description of depths in arbitrary intervals, such as annual total depths, is shown to be treatable as the sum of a random number of (independent) random variables by using the model for rainy day occurrence. Simulations from this model match well the measured data from the stations treated. The effect of truncation of the rainfall sample on both depth distribution and Markov chain dependence is discussed. It is also indicated how daily depths are functions, in turn, of individual storm occurrence probability (number per day) and storm depth distributions.

In a previous paper [Smith and Schreiber, 1973], three sample stations with long (55–73 years) records from the Sonoran-Chihuahuan intermountain basins of southern Arizona–New Mexico were used to evaluate the validity of simple stochastic models for rainy day occurrences in this semiarid region for the summer thunderstorm season. A Markov chain model reproduced the observed occurrence patterns much better than an independent sequence model. This paper presents studies of the daily rainfall depths from the same stations, including a study of possible dependence connecting rainfall occurrence patterns and daily depths, and the combination of the two models (occurrence and depth) to provide information on the probability of the depth of rain in arbitrary periods. The utility of this information in hydrologic and watershed (range forage) studies is indicated.

### TOTAL SAMPLE DISTRIBUTION OF DAILY RAINFALL DEPTH

The three sampling points providing data for this study are located in southeastern Arizona and were illustrated by Smith and Schreiber [1973, Figure 1]. Information about summer point rainfall in this region is expected to provide valuable assistance for studies of areal storm patterns on the heavily instrumented Walnut Gulch watershed at Tombstone. At the same time, this study is intended to represent a large area of the Southwest. Although for some purposes it would be preferable to study the depth of rain within a time period small enough to distinguish separable storms occurring in 1 day, in a thunderstorm-dominated season the day is a natural time cycle for meteorological energy related to thunderstorm phenomena.

If all rainfall day data are lumped and the total sample depth distribution is determined, the general shape of an arithmetic plotting indicates approximately an exponential distribution, given by

$$P(X \leq x) = 1 - e^{-\lambda x} \quad (1)$$

in which  $X$  is the depth of daily rainfall,  $\lambda$  is a single parameter, and  $x$  is a random variable. This suggests plotting the depth distribution data by comparing  $\log(1 - P)$  and  $X$ ,

as has been done for the three sampling stations in Figure 1.

The large sample size for each station defines the distribution accurately, so that the bend shown in the distribution for all stations is taken as a significant feature of the natural rainfall distribution in this climatic area. D. A. Woolhiser (personal communications, 1973) has found the same shape for daily rainfall in eastern Colorado. Nature, of course, does not necessarily conform to a simple algebraic probability distribution. All the curves shown in Figure 1 may be approximated by two segments divided by some point of inflection  $x_c$ , so that

$$P(X \leq x) = 1 - e^{-a - \lambda x} \quad x \geq x_c \quad (2)$$

in which  $a$  is a parameter representing  $P(X = 0)$ . The short segment  $x \leq x_c$  also may be approximated by an exponential distribution, e.g., a straight-line segment on Figure 1.

The total trace may be described also as the sum of two exponential distributions as follows:

$$P(X \leq x) = \alpha e^{-\lambda_1 x} + (1 - \alpha)e^{-\lambda_2 x} \quad (3)$$

in which  $\alpha$  is a weighting factor ( $0 \leq \alpha \leq 1$ ) and  $\lambda_1$  and  $\lambda_2$  are different parameters. Other approximations are, of course, equally possible. Owing to the very skewed shape of the probability density function for this random variable, the uncertainty of the shape of the probability distribution function increases with the depth of rainfall. Very few samples even in 70 years define the frequency of extreme (high) rainfall. The opposite applies to low rainfall depths.

### INVESTIGATION OF HOMOGENEITY OF RAINFALL POPULATIONS

At this point we ask the question, Does this probability distribution represent the sampling of independent, identically distributed events? More formally, we may suggest the hypothesis that the sample at each station represents daily depth distribution, independent of when the rainy day occurs. If such a hypothesis can be supported empirically, a rather complete and powerful stochastic model for rainfall results. The tests performed to study this question are not exhaustive, but they represent several possible (conceptual) ways that rainfall depth might be related to rainfall day occurrence.

*Homogeneity of large-storm occurrence.* One relation has

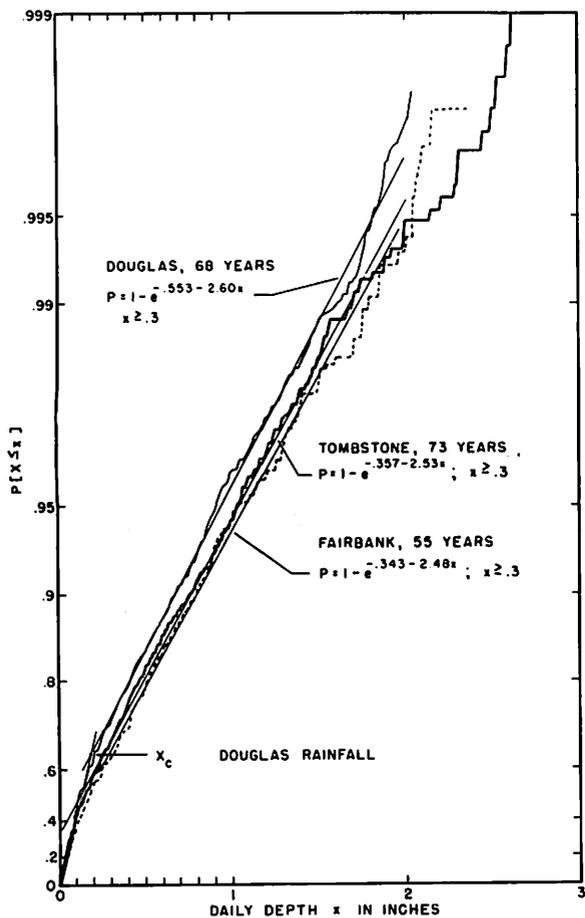


Fig. 1. Distributions of summer seasonal daily rainfall depths for three sampling stations in southeastern Arizona.

been suggested by *Osborn and Renard* [1970]. On the basis of a short but intensive areal record they postulated that there is meteorological significance in the observation that the few largest runoff-producing rainfall events occurred following longer, 'relatively dry' periods in this area. They suggested that perhaps atmospheric energy relations favored larger storms after dry periods.

This idea was studied in three ways. The first test was a simple correlation of the length of all dry periods with the rainfall occurring on the day following that dry period. For a sample size of 1172 the linear correlation coefficient for these two variables was  $-0.0132$ .

It follows from this hypothesis that the first rainy day following a dry sequence might produce more rain regardless of the drought length. Thus for the second test the rainy days were divided into two subsamples: first days of a wet sequence (run length of  $\geq 1$  day) and all others. The sample distributions determined for each are presented in Figure 2 for three sample stations. This graph indicates that for days with less than the mean depth (approximately 0.25 inch) the distributions are practically identical and that storms with depths from about 0.25 inch to 1.5 inches are slightly less probable on the first days of rain than on other days. These differences are not totally consistent and are quite small but by a nonparametric test are statistically significant; the significance level is 86% at Tombstone and  $>99\%$  at Fairbanks and Douglas. Note that the differences are exaggerated by the logarithmic plotting of Figure 2.

In a final test of the large-storm occurrence hypothesis,

only rainfall days with 1.0 inch or more of rain were considered. These large depths occurred on days following dry days or wet days in almost the same proportions as they occurred for the entire sample. Further, when the largest storms occurred following a wet day, the sample distribution of depths on the preceding days was not significantly different from the general population of rainfall depths. Thus there is no general support in this long-term data for the hypothesis of *Osborn and Renard* [1970], which is apparently an interesting example of generalization from a small sample.

*Homogeneity of depth in various wet day runs.* In another test for independence in rainfall depth occurrence the rainfall depth was related to the size of the wet day run in which it occurred. The population was divided into subsamples of  $n$ -day rainy periods, and the resulting depth distributions were compared by a relatively powerful distribution-free test, the Wilcoxon rank sum test, for identity of population [Bradley, 1968]. Under the hypothesis  $H_0$  that the populations are from different samples the test accepted  $H_0$  at the 99% level for the Tombstone and Douglas stations for comparing 1-day rainy period depths with 2-day period depths. In all other comparisons, including all comparisons at the third station, Fairbank,  $H_0$  was rejected at considerably lower confidence levels. The tests are summarized in Table 1.

These general results indicate that there is a slightly higher probability that larger storms will occur in periods of 2 or more days of rain than on single (isolated) rainy days. This might be interpreted as a small carry-over of persistence affecting depth. On the basis of these tests, however, any dependence is apparently slight when it does occur.

*Seasonal homogeneity in relative event occurrence.* In another experiment, each season's record was divided

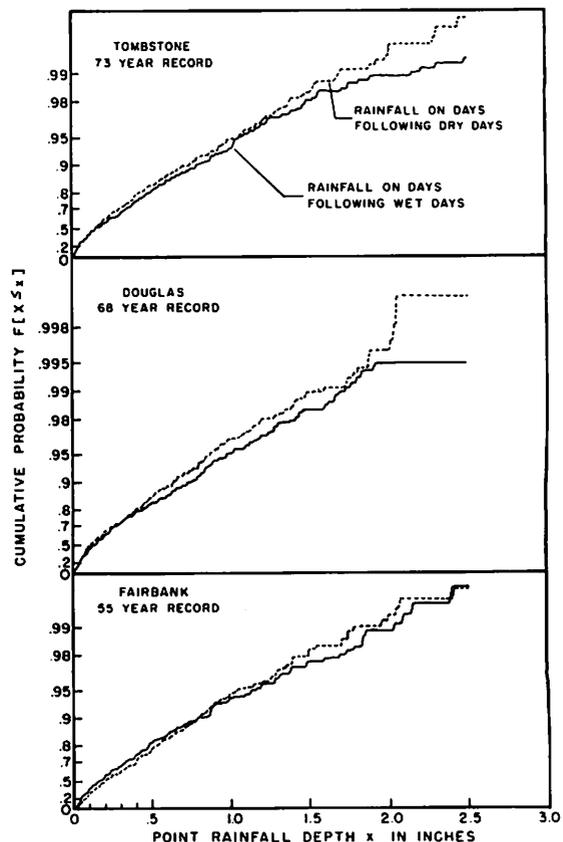


Fig. 2. Comparison of point depth distributions for days preceded by wet days and by dry days for three sample stations.

TABLE 1. Statistical Distribution-Free Test for Daily Rainfall Population Partitioned by Location in Sequence

Comparison of Rainfall From Days in Wet Sequence of Size	Normalized Wilcoxon Parameter	Confidence Level, %	Test Conclusion: Populations
<i>Tombstone</i>			
1 vs. 2	2.80	99.7	different
2 vs. 3	0.335	63.1	same
3 vs. 4	0.044	51.0	same
<i>Douglas</i>			
1 vs. 2	2.71	99.7	different
2 vs. 3	0.530	70.2	same
3 vs. 4	0.372	64.5	same
<i>Fairbank</i>			
1 vs. 2	1.01	84.4	different
2 vs. 3	0.249	59.8	same
3 vs. 4	-0.399	65.5	same

chronologically into near-equal thirds to form three groups by relative occurrence within the season, and the distribution of rainfall depth in each period was compared. As is shown in Figure 3, rainfall in the middle of the thunderstorm season is statistically significantly (99% level) greater than rainfall in the earlier and later thirds of the season. Nevertheless, the maximum difference in probability is only 9% at a rainfall depth of approximately 0.12 inch, and the difference is much smaller for larger depths.

*Annual homogeneity.* As a final test for dependence effects we consider the hypothesis that daily rainfall amounts are drawn from different populations in different years. This hypothesis would perhaps have strong support among cattle ranchers, who speak in terms of 'good' and 'bad' years. The question may be said to deal with annual homogeneity.

The mean daily rainfall for each season is used as a best indicator of the daily rainfall sample distribution for that season. This parameter is plotted in Figure 4 against the total depth for the respective year for each of the 73 sample years at Tombstone. Since  $\bar{X}_i = (\sum_n X) / n_i$  exactly, where  $\sum_n X$  is the

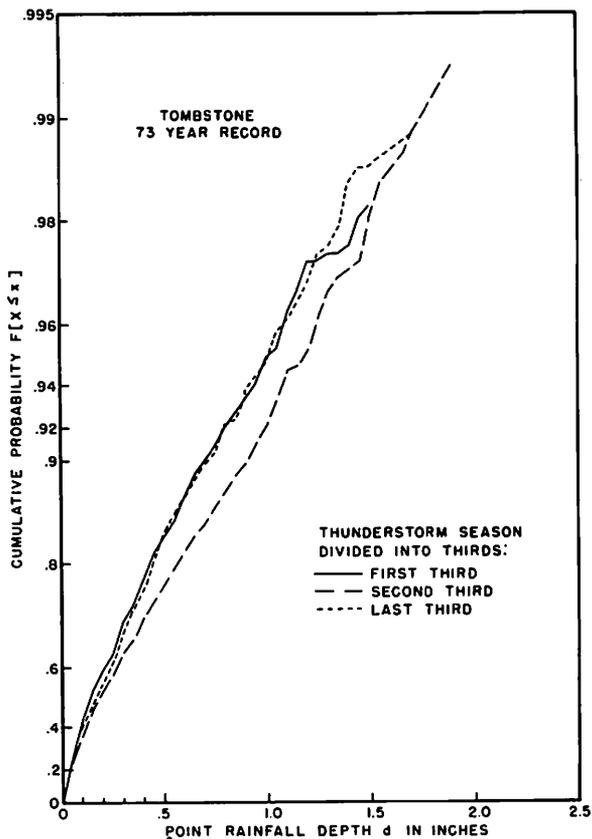


Fig. 3. Comparison of point rainfall depth distributions by location of event day within the thunderstorm season at Tombstone.

$i$ th seasonal total depth,  $n_i$  is the number of storms per  $i$ th season, and  $\bar{X}_i$  is the seasonal mean storm depth; the lines representing the number of storms  $n_i$  are exact.

Two observations may be made from inspection of this figure. First, concerning homogeneity, the general scatter of

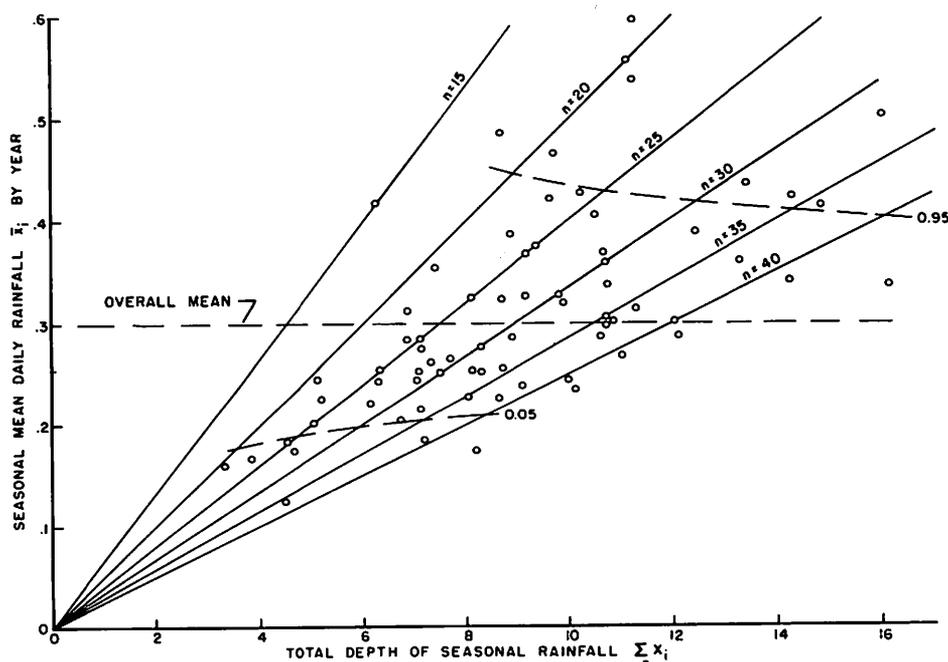


Fig. 4. Plot of variance of mean seasonal daily rainfall as a function of total seasonal depth at Tombstone:  $n$  represents the number of rainy days per season.

samples is symmetrical along lines of equal  $n$  with respect to the overall mean, although there is apparently a slight bias for extreme years. When less than approximately 20 storms occur per season, the mean daily depth is slightly higher. Conversely, there is a weak bias indicating that seasons with more than 40 rainfall days have a slightly lower daily mean. However, two-tail confidence limits of 0.05 and 0.95 as drawn on this graph (obtained by using (8) below) show that 16 years fall outside these limits, as compared with an expected number of 7.3, indicating strongly that a small annual variation in the mean may be occurring or that each year's rainfall may not be drawn from quite the same population. Cross correlation of annual summer rain depths from the three neighboring stations supports this, each depth showing a low (0.42-0.62) but statistically significant correlation. Curiously, similar comparison of annual mean daily depths exhibits insignificant correlation values (0.11-0.27). Apparently, regional relationships are complex and deserving of additional analysis.

Second and more importantly, the sampling variance for the annual sample of daily depths is far higher apparently than the variance of the number of rainfall days. One striking feature is that the most extreme dry years have resulted mostly from the occurrence of an unusually low mean storm depth for that year. Considering general scatter along lines of equal  $n$  in Figure 5, the individual rainfall depth and the annual number of wet days may apparently be considered independent, or  $P(y/n)_i = P(y)_i$ , in which  $i$  refers to the year,  $y = \sum_{j=1}^{n(i)} X$  is the annual total depth, and  $n(i)$  refers to the annual number of  $i$ th seasonal rain days.

CHARACTERISTICS OF TRUNCATED DISTRIBUTIONS

It is of interest to note the effect of truncation on the distributions shown. Several authors have chosen to deal only with rainfall either above some threshold value [Osborn, 1968; Lane and Osborn, 1972] or from storms with maximum depth above a threshold value [Duckstein et al., 1972]. The second case presents a complex picture by effectively defining a probabilistically varied threshold for point depths. The first case is quite easy to treat mathematically. One may show that for a simple exponential distribution of point rainfall depth, truncation by some threshold  $x_T$  simply causes a shift in the

definition of the random variable, so that (1) becomes

$$P(X \leq x) = 1 - e^{-\lambda(x - x_T)} \quad X \leq x_T \quad (4)$$

where  $x_T$  is the threshold value. The result is effectively an ordinate shift, and one could therefore consider the simple exponential distribution as a good approximation for rainfall depths  $\geq 0.2$  or 0.3 inch in the area of sampling discussed in this paper. Osborn [1968] and Lane and Osborn [1972] used a threshold of 0.25 inch, but neither investigated sample depth distributions. Osborn et al. [1974] employed an exponential distribution hypothesis for simulating the 'storm center depth' of storms with more than a 0.25-inch threshold. The relation of this variable to point rainfall depth is discussed in a paper by Smith [1974], indicating on the basis of a small sample that Osborn et al. [1971] have made a good choice.

The use of a threshold for daily rainfall depths has a significant effect also on the rainy day occurrence model studied by Smith and Schreiber [1973]. Use of a 0.25-inch threshold for the stations here will eliminate more than 60% of the sample (the median is  $\approx 0.16$ ). The tests used in the analysis [Smith and Schreiber, 1973] were repeated with a 0.2-inch threshold definition, however, the tests again strongly favoring the Markov chain model of rainfall day occurrence.

The effect of a fixed threshold on the description of daily rainfall occurrence as a Markov chain may be studied explicitly when the probability of exceeding the threshold is known. The original transition probability matrix for a two-state chain with 0 representing dry days and 1 representing wet days [Smith and Schreiber, 1973] is:

	$e_i$		
	0	1	
$e_{i-1}$			
0	$f_d$	$1 - f_d$	(5)
1	$1 - f_w$	$f_w$	

D. A. Woolhiser (personal communication, 1973) has shown how the knowledge of the probability  $P_T$  of an arbitrary threshold level  $T$  from the depth probability distribution may be used first to expand matrix 5 into a three-state chain, including the vent 0' that daily depth is less than  $T$ . Subsequently, states 0 and 0' may be lumped by using

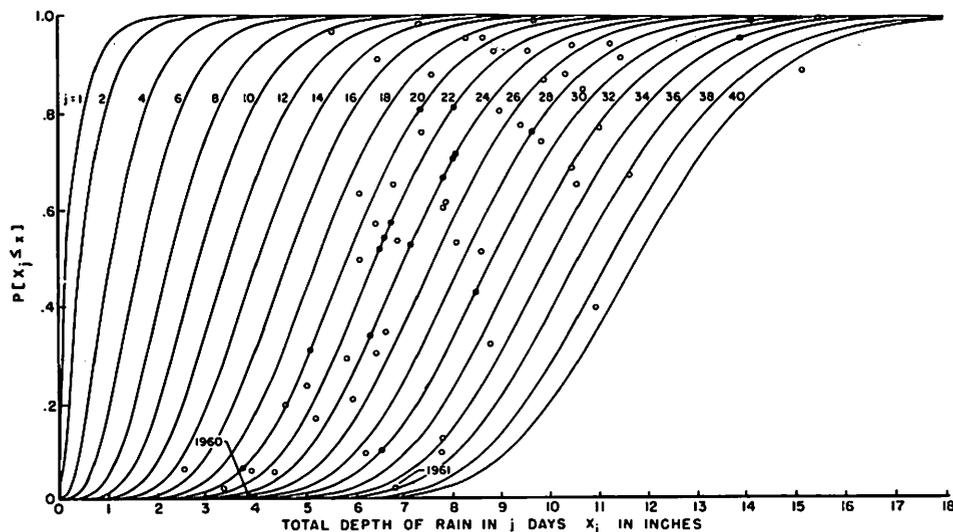


Fig. 5. Computed distribution of the total depth of rainfall from  $n$  rainfall days, based on the measured daily depth distribution ( $j = 1$ , shown). Plotted circles indicate the years of record.

stationary Markov probabilities to produce a new transition probability matrix in which state 1' represents rain  $\geq T$ :

$$\begin{array}{c|cc}
 e_{i-1} & \begin{array}{c} 0' \\ \hline 1' \end{array} & \begin{array}{c} e_i \\ \hline 1' \end{array} \\
 \hline
 0' & 1 - P_{0'0'} & P_{0'1'} = \frac{(1 - P_T)(1 - f_d)[1 - f_w(1 - P_T)]}{1 - f_w + P_T(1 - f_d)} \\
 1' & 1 - f_w(1 - P_T) & P_{1'1'} = f_w(1 - P_T)
 \end{array} \quad (6)$$

Note that as  $P_T \rightarrow 1$ , the matrix approaches

$$\begin{array}{c|cc}
 & 0' & 1' \\
 \hline
 0' & 1 & 0 \\
 1' & 1 & 0
 \end{array}$$

as a limit, which is rationally consistent. Given that  $f_w > 1 - f_d$ , one may show that even as  $P_T \rightarrow 1$ ,  $P_{1'1'} > P_{0'0'}$ , indicating that dependence remains even for large thresholds.

Hershfield [1970] in a study of unconditional and conditional probabilities for several stations in the United States used thresholds from 0.01 inch (implicit in the data used herein) to 0.50 inch. His results were interpreted to indicate that dry-to-dry day transition probabilities approach unconditional dry day occurrence probabilities as the threshold is increased. Wet day dependence was less conclusive. This is consistent in general with the preceding development from Woolhiser. When the threshold is increased sufficiently, it is difficult with such a severely reduced sample size to distinguish between an independent and a dependent process.

**SIMULATION OF PERIOD DEPTH DISTRIBUTIONS**

It was demonstrated above that interdependencies of daily rainfall depth probability and rainfall day occurrence are at most weak. Therefore we proceed under the assumption of independence to investigate rainfall depth occurring in specified periods, presuming such depth to be the sum of a random number of random variables. Properties of such sums have been treated by Wald [1944] and more extensively by Todorovic [1970]. Todorovic treated processes in which  $n$  random sized events occurred in continuous time  $t$ . Daily rainfall allows some simplification. Using notation similar to that of Todorovic, let  $k = 0, 1, 2, \dots, m$  represent the number of rainfall events in a period of length  $m$  and  $E_k^m$  represent the event that  $k$  rainy days occur in a period of length  $m$ . Also, define random variable  $X_k$  as the total rainfall resulting from  $k$  rainy days, or  $X_k = \sum_{i=1}^k x_i$ , where  $x_i$  is the rainfall on the  $i$ th rainy day of the period. Then the cumulative distribution function  $F_m(x)$  for the depth occurring in an  $m$ -day period is as follows [cf. Todorovic and Woolhiser, 1974]:

$$F_m(x) = P(E_0^m) + \sum_{i=1}^m P(E_i^m)P(X_i \leq x | E_i^m) \quad (7)$$

Note that this does not require independence of  $P(X_j \leq x_j)$  and  $P(E_j^m)$ . Under such conditions of independence, however, (7) is simplified to

$$F_m(x) = P(E_0^m) + \sum_{i=1}^m P(E_i^m)P(X_i \leq x) \quad (8)$$

The distribution function  $P(X_j \leq x)$  is simply the  $j$ th convolution of  $P(X_1 \leq x)$ , described for the present region by (2)

with a parabolic transition curve for  $X \leq X_c$ . This compound approximation to the sample distribution requires the use of numerical methods. The only difficulty involved is that any numerical error in a convolution procedure is multiplied  $j$  times for the  $j$ th convolution, so that a precise method is necessary. The distribution functions  $P(X_j \leq x)$  are shown in Figure 5 for a representative range of  $j$  at Tombstone resulting from a numerical procedure employing integration by Simpson's rule.

*Distribution of summer seasonal rainfall.* The  $P(E_j^m)$  is obtained from the 73-year sample presented by Smith and Schreiber [1973, Figures 10-12]. The  $F_m(x)$  as calculated by (8) for the 122-day thunderstorm season for the three stations of the present study are compared with the measured values in Figure 6. The agreement is generally excellent and thus indirectly verifies the assumption of independence. In addition, Figure 6 shows a theoretically developed summer season rainfall distribution for the Tombstone gage using the Markov chain model for the number of rainfall days per season. Comparison gives an idea of sensitivity by observation of the error in using the theoretical Markov chain model to predict the distribution of summer rainfall depth. Other interesting information may easily be developed from the joint use of depth and occurrence distributions for shorter periods of time within the season.

*Relation of storm depth distribution to daily depth distribution.* Equation 5 also can explicitly relate storm depth distribution to the daily depth distribution that we have been

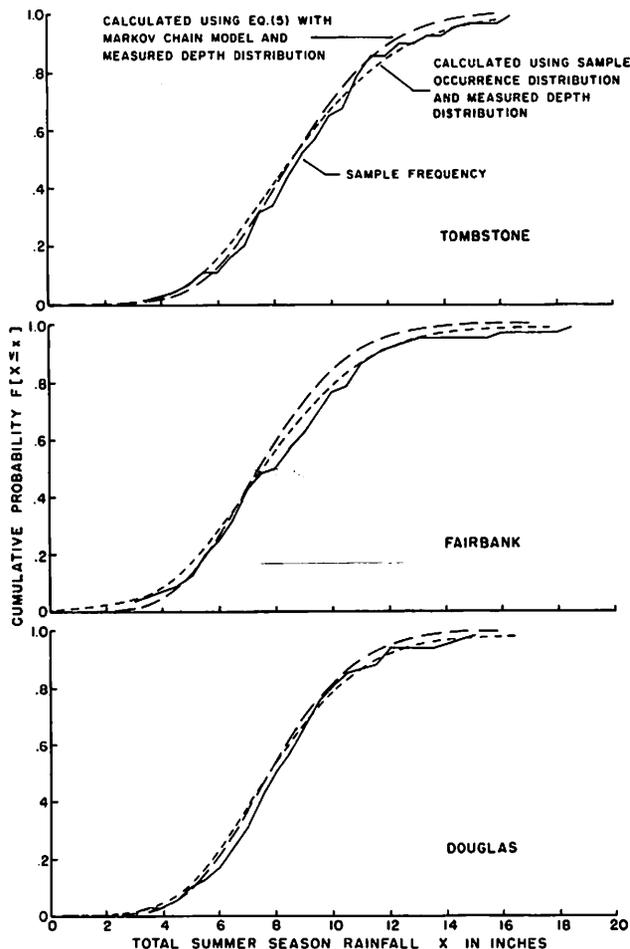


Fig. 6. Measured and simulated distributions of the total summer seasonal rainfall at three sample stations in Southeastern Arizona.

concentrating on here. Data from the Agricultural Research Service Walnut Gulch experimental watershed provided the mean storm depth distribution from approximately 15-year records of five gages in the Tombstone area, which established an approximate distribution for storm depths and an estimate for the distribution of the number of storms per day. Application of equation 8 (specifically limiting  $m$ , possible storms per day) to this information showed decisively that each rainfall event on a multiple-storm day cannot be considered to be coming from the same population. However, the short available storm rainfall record has produced insufficient sample data for the few (<20%) days of multiple storms to predict the distribution of depths in secondary storms. Further work is required, including experiments to deduce characteristics of secondary storm distributions by using the daily and single storm distributions in (9).

#### SUMMARY AND CONCLUSIONS

As a sequel to the first part of this study [Smith and Schreiber, 1973], in which alternative models for the binary daily process of thunderstorm rainfall occurrence were compared, this paper reports results of a study of the conditional distribution of daily rainfall amounts, given the occurrence of rain. Implied in the separate treatment of these two random variables is the assumption that they may be treated as being independent. The first topic considered therefore was the explicit testing of this hypothesis by studying several possible partitionings of the rainfall samples to determine if dependencies connecting rainfall depth samples and pattern of occurrence exist.

Using a sample of more than 2000 rainfall days, it was demonstrated that a tendency apparently exists for storms occurring in the middle of the thunderstorm season to exhibit a slightly higher mean rainfall depth than storms occurring on other rainfall days and for storms following wet days to have slightly higher rainfall depths. Comparison of the rainfall samples by year for the 73-year record at Tombstone shows a general independence of the number of rainy days per season and the average daily rainfall. Variability of the annual rainfall depth is apparently more attributable to large sampling variance in daily rainfall than to variation in the number of rainfall days per season.

On the basis of these tests the total rainfall depth per 122-day summer season apparently may be approximated as a random variable determined as the sum of a random number of random variables. Simulation using (8) produces good agreement with recorded depths, confirming the general independence of the occurrence and depth processes. The sensitivity of the predicted seasonal depth to the bias in the assumed Markov chain occurrence model is apparently relatively small. This model, for example, could be used in

practical watershed management for this hydrologic region to estimate the probability of extreme drought or wet seasonal rainfall, which is indeterminate from short records.

Hopefully, one of the main results is an indication of further useful work. The general result to be emphasized in this study is the analytical utility of the separate treatment of occurrence and depth of rainfall in thunderstorm conditions. Some of the simplifying assumptions employed here were not theoretically necessary and perhaps could be eliminated on the basis of further investigation. Daily depths and individual storm depth distributions are related by a knowledge of the distribution of the number of storms per day and the depth distributions of secondary storms. This relation could be presented more exactly as data for these distributions are accumulated.

#### REFERENCES

- Bradley, J. V., *Distribution Free Statistical Tests*, pp. 105-111, Prentice-Hall, Englewood Cliffs, N. J., 1968.
- Duckstein, L., M. M. Fogel, and C. C. Kisiel, A stochastic model of runoff-producing rainfall for summer type storms, *Water Resour. Res.*, 8(2), 410-420, 1972.
- Hershfield, D. M., A comparison of conditional and unconditional probabilities for wet- and dry-day sequences, *J. Appl. Meteorol.*, 9(5), 825-827, 1970.
- Lane, L. J., and H. B. Osborn, Hypotheses on the seasonal distribution of thunderstorm rainfall in southeastern Arizona, paper presented at 2nd International Symposium in Hydrology, Colo. State Univ., Fort Collins, Sept. 1972.
- Osborn, H. B., Persistence of summer rainy and drought periods on a semiarid rangeland watershed, *Bull. Int. Ass. Sci. Hydrol.*, 13(1), 14-19, 1968.
- Osborn, H. B., and K. G. Renard, Thunderstorm runoff on the Walnut Gulch experimental watershed, U.S.A., paper presented at Symposium on Results of Research on Representative and Experimental Basins, Int. Ass. Sci. Hydrol., Wellington, New Zealand, 1970.
- Osborn, H. B., L. J. Lane, and R. S. Kagan, Stochastic models of spatial and temporal distribution of thunderstorm rainfall, *U.S. Dep. Agr. Misc. Publ. 1275*, in press, 1974.
- Parzen, E., *Modern Probability Theory and Its Applications*, pp. 136-145, John Wiley, New York, 1960.
- Smith, R. E., Point processes of seasonal thunderstorm rainfall, 3, Relation of point rainfall to storm areal properties, *Water Resour. Res.*, 10, this issue, 1974.
- Smith, R. E., and H. A. Schreiber, Point processes of seasonal thunderstorm rainfall, 1, Distribution of rainfall events, *Water Resour. Res.*, 9(4), 871-884, 1973.
- Todorovic, P., On some problems involving random number of random variables, *Ann. Math. Statist.*, 41, 1059-1063, 1970.
- Todorovic, P., and D. A. Woolhiser, Stochastic model of daily rainfall, *U.S. Dep. Agr. Misc. Publ. 1275*, in press, 1974.
- Wald, A., On cumulative sums of random variables, *Ann. Math. Statist.*, 15, 283-296, 1944.

(Received May 30, 1973;  
revised January 11, 1974.)