AGENCY FOR INTERNATIONAL DEVELOPMENT
WASHINGTON, D. C. 20523

BIBLIOGRAPHIC INPUT SHEET

FOR AID USE ONLY
0145

1. SUBJECT CLASSIFICATION

A. PRIMARY
Agriculture AF10-0000-0000

B. SECONDARY
Meteorology and climatology

2. TITLE AND SUBTITLE
A probabilistic model of winter rainfall

3. AUTHOR(S)
Kao,S.B.; Duckstein,Lucien; Fogel,M.M.

4. DOCUMENT DATE	5. NUMBER OF PAGES	6. ARC NUMBER
1971	25p.	ARC

7. REFERENCE ORGANIZATION NAME AND ADDRESS
Ariz.

8. SUPPLEMENTARY NOTES *(Sponsoring Organization, Publishers, Availability)*
(Presented at Am.Geophysical Union,fall national meeting,San Francisco)

9. ABSTRACT

10. CONTROL NUMBER
PN-RAA-099

11. PRICE OF DOCUMENT

12. DESCRIPTORS
Probability theory
Rainfall
Seasonal variations

13. PROJECT NUMBER

14. CONTRACT NUMBER
CSD-2457 211(d)

15. TYPE OF DOCUMENT

AID 590-1 (4-74)

71-11

A PROBABILISTIC MODEL OF WINTER RAINFALL

by

Samuel E. Kao,

Lucien Duckstein

and

Martin M. Fogel*

* Respectively,

Graduate Associate
Department of Hydrology
and Water Resources,

Professor of Systems
& Industrial Engineering,

and

Professor of Watershed Management

UNIVERSITY OF ARIZONA
TUCSON, ARIZONA 85721

Paper for presentation before the Fall National Meeting, American Geophysical Union, San Francisco, California, December 6-9, 1971.

For Discussion Purpose Only.

A PROBABILISTIC MODEL OF WINTER RAINFALL

ABSTRACT

A framework has been constructed for the probability distribution of winter rainfall in southwestern United States. For convenience sake, a storm group was defined as a sequence of days consisting of no two consecutive dry days with a total amount of rainfall equal to or greater than a certain threshold value. Assuming that the storm groups are independent and uncorrelated, a Poisson distribution could adequately describe the number of storm groups per season, and a negative binomial distribution could represent the density function for storm duration. The results of a sensitivity analysis indicate that the parameters of both distributions change with threshold values. Given that a storm group has occurred, the frequencies of rainfall amount per group were fitted to a mixed distribution to account for extreme values. The seasonal maximal distribution and the expected return intervals of storm group depth were also obtained. The summation of random number of random variables was used to check the model by obtaining total seasonal rainfall.

INTRODUCTION

There are two distinct seasons of the year when precipitation is especially common in Arizona (Jurwitz, 1953). The first rainy season, principally resulting from local convective storms, lasts approximately from mid July to mid September. The second rainy season, mainly caused by cyclonic storms, extends from December through the middle of March.

Generally, winter storms are less intense, longer, but more widespread than those of summer. In the southwestern United States the annual surface runoff from smaller semi-arid watersheds is basically caused by convective storms. However, historical records show that the accumulation of winter precipitation over several days could also create flood damage in the urban and rural area.

After examination of the record of long term monthly mean in Tucson, shown in Figure 1, the interval of winter rainy season was set from October through March, since the possibility of rain in October and November cannot be excluded.

The station known as Tucson, University of Arizona, was chosen to represent the rainfall characteristics in the metropolitan area of Tucson because its period of record was the longest. Note that the wide spread winter storms are fairly uniformly distributed in this area. The daily rainfall

data were taken between 1914 and 1970. It is believed that 57-year data is long enough for the analysis of frequency distribution (Wiesner, 1970). In this analysis, a wet day was defined as a day which has an amount of rainfall equal to or greater than a threshold value c. For convenience sake, the daily data were lumped together as storm groups because many of the storms occurred continuously over several days. A storm group was defined as a sequence of days consisting of no two consecutive dry days with a total amount of rainfall equal to or above a certain threshold value. The minimum threshold value was chosen to be 0.01 inch.

The objectives of this study are first to derive a probability model for predicting the occurrence, duration and magnitude of winter rainfall in semi-arid regions, and second, to analyze the sensitivity of the model to different values of the threshold c.

NUMBER OF OCCURRENCES OF STORM GROUPS

The occurrences of storms have been successfully described as a Poisson process (Borgman, 1963). Todorovic and Yevjevich (1969) also found that the number of storms within a given interval is Poisson distributed. An inherent assumption of this finding is that storms occur independently from one another. During the colder months in Arizona, precipitation may be obtained under rather unusual circumstances due to a fast-moving air current originating in the north Pacific Ocean. Sellers (1964) points out that once this flow pattern is established, it tends to persist or re-occur, with the results that several storms in succession may pass over the state at intervals of about one week. However, since storms of this type represents less than 5% of the occurrences, it is possible to assume that the storm groups occur in an independent manner.

Theoretically, the occurrence of independent events follows a Poisson process if the interarrival times are exponentially distributed. Several authors have found that the dry spells may follow an exponential or geometric distribution (Longley, 1953, Gabriel and Neumann, 1962; Weiss, 1964). In the present analysis, the interarrival times between two storm groups also tend to be exponentially distributed. As shown in Figure 2, the 57-year record

indicates a trend towards a Poisson distribution for the number of storm groups per season.

The chi-square goodness-of-fit test was used to test this hypothesis. At the 10 percent level of significance, the observed value for each threshold was much less than the tabulated critical value. Thus, there is no reason to reject the null hypothesis. The frequency distribution for the number of storm groups per season is therefore

$$f_{\underline{N}}(n) = \frac{e^{-m}\, m^{n}}{n!} \qquad n = 0,1,2,\ldots \tag{1}$$

where the parameter m represents the mean number of occurrences of storm groups per season.

Figure 3 is an illustration of the relationship between the value of m and the threshold. As it can be expected, the mean number of storm groups per season n decreases when the threshold value c is increased. An approximate relationship between n and c (in inches) is

$$\log n = 11.2 - 28\, c$$

DURATION OF STORM GROUPS

The shape of histograms suggests empirically that the distribution of storm group duration may follow a negative binomial, as shown in Figure 4. Comparing the theoretical frequencies to the observed ones, the chi-square statistics shows that this hypothesis cannot be rejected at the 10 percent level of significance for all values of the threshold. Mathematically expressed, the frequency density function for t-days duration is

$$f_2(t) = \binom{r+t-1}{t}(1-q)^r q^t$$

$$t = 0, 1, 2, \ldots \quad (2)$$

where the parameters q and r were estimated by the method of moments from the theoretical mean $qr/(1-q)$ and variance $qr/(1-q)^2$. As shown in Figures 5 and 6, the value of q decreases and that of r increases when the threshold value c is increased. At the minimum threshold (0.01 inch), r is almost equal to unity. In other words, the distribution of duration is geometric at that threshold.

THE AMOUNT OF RAINFALL PER STORM GROUP

Let the amount of rainfall be expressed as an integral number of half-inch increments. As shown in Figure 7, the amount $\underline{x}$ of rainfall per storm group, given that a storm group has occurred, appears to fit closely a geometric distribution.

$$f_{\underline{X}}(x) = (1 - p)\, p^{x-1} \qquad x = 1, 2, \ldots \tag{3}$$

where the parameter p is the probability that the area receives an amount of rain equal to or greater than a certain threshold c. At the minimum threshold, using the method of moments on grouped data, p was found to be 0.31. For other values of the threshold c, the distribution function in the continuous formulation becomes

$$F_{\underline{X}}(x) = 1 - e^{-(1-p+c)} \tag{4}$$

At the tail of the distribution, however, for rainfall amount greater than three inches, the observed frequencies were found to be higher than the theoretical ones. This may be due to the existence of extreme events which belong to another population. Since the tail of the distribution has important implications in the design of water control structures, a mixed distribution was employed in order to obtain a better

distribution. Let the mixed distribution be

$$H_{\underline{X}}(x) = F_{\underline{X}}(x) + (1 - \alpha)\, G_{\underline{X}}(x) \qquad x = 1, 2, \ldots \tag{5}$$

where α is a weighing factor lying between 0 and 1 and $G_{\underline{X}}$ is a uniform distribution function:

$$G_{\underline{X}}(x) = \frac{x}{k} \qquad x = 1, 2, \ldots, k \tag{6}$$

As no known rain gage in this area has ever recorded more than eleven inches for a single storm group, the value of k was chosen to be 22, i.e., eleven inches. The value of α, on the other hand, was chosen subjectively as 0.98.

The distribution function of $\underline{X}$ is then

$$H_{\underline{X}}(x) = \alpha(1 - p^{x}) + (1 - \alpha)\frac{x}{k} \tag{7}$$

Other mixed distributions are being investigated to avoid truncating a uniform density at an arbitrary level. Also, physically based models that can predict extremes are being developed.

SEASONAL MAXIMAL RAINFALL

Maximum events are of utmost importance for the design of engineering projects involving hydrologic factors. Success or failure of a hydraulic structure strongly depends on its ability to withstand extreme values of rainfall (or runoff) to which it is subjected. If $\underline{Q}$ is the maximum of n random variables $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$, then

$$F_{\underline{Q}}(x) = P\{\underline{Q} \leq x\}$$

$$= P\{\text{all n of the } \underline{X}_i \leq x\} \tag{8}$$

Assuming that the $\underline{X}_i$'s are mutually independent,

$$F_{\underline{Q}}(x) = P\{\underline{X}_1 \leq x\}\ P\{\underline{X}_2 \leq x\} \ldots P\{\underline{X}_n \leq n\}$$

$$= H_{\underline{X}_1}(x)\ H_{\underline{X}_2}(x) \ldots H_{\underline{X}_n}(n) \tag{9}$$

Also, if we assume that the $\underline{X}_i$ are identically distributed,

$$F_{\underline{Q}}(x) = [H_{\underline{X}}(x)]^n \tag{10}$$

If there is a random number $\underline{N}$ of storm events per season with a frequency distribution $f_{\underline{N}}(n)$, then the maximal distribution can be written as (Benjamin and Cornell, 1970);

$$F_{\underline{Q}}(x) = \sum_n \{[H_{\underline{X}}(x)]^n\ f_{\underline{N}}(n)\} \tag{11}$$

Substituting equations 1 and 7 into equation 11,

$$F_Q(x) = e^{-m} \sum_n \left\{ \left[\alpha(1-p^x) + (1-\alpha)\frac{x}{k} \right]^n \frac{m^n}{n!} \right. \tag{12}$$

Equation 12 gives the probability that the maximum rainfall depth in any season be less than or equal to x units. The frequency densities determined from both the model and the historical records are compared in Figure 8. The expected recurrence interval T_r for rainfall equal to or greater than x units is simply

$$T_r = \frac{1}{1 - F_Q(x)} \tag{13}$$

THE TOTAL SEASONAL RAINFALL

The total amount of rainfall $\underline{Y}$ per season can be obtained by means of the convolution

$$\underline{Y}_{\underline{N}} = \underline{X}_1 + \underline{X}_2 + \ldots + X_{\underline{N}} \tag{14}$$

If $\underline{N} = n$, then

$$j = \underline{X}_1 + \underline{X}_2 + \ldots \underline{X}_n \tag{15}$$

where $X_i (i = 1,2, \ldots, n)$, the amount of rainfall per each storm group, are assumed to be identically distributed. We found previously that $\underline{N}$ has a Poisson distribution, and therefore $\underline{Y}_{\underline{N}}$ must have a compound Poisson distribution. Benjamin and Cornell (1970) given an example of deriving the compound distribution of the total annual rainfall. If the number of storm groups is known, then the sum of n (a positive integer) independent and identically distributed geometric random variables has a negative binomial (Pascal) distribution.

$$h(j, n) = \binom{j-1}{n-1} (1-p)^n p^{j-n} \qquad j = n, n+1, \ldots \tag{16}$$

where p is the same parameter of geometric distribution as that in the preceding section. Since n is Poisson distributed, hence for $j>0$.

$$h_{\underline{Y}}(j) = \sum_{n=1}^{\infty} \frac{(j-1)!}{(n-1)!(j-n)!} (1-p)^n p^{j-n} \frac{e^{-m} m^n}{n!} \qquad j = n, n+1, \ldots \tag{17}$$

The distribution of $\underline{Y}$ is a mixed distribution. Equation 17 only gives the major part. At $j = 0$, we must include a spike, which represents the probability of exactly zero rainfall, given that there is a storm event. The magnitude of this spike is

$$P(\underline{N} = 0) = e^{-m} \tag{18}$$

The final form of this distribution function is then

$$H_{\underline{Y}}(j) = e^{-m}\left(1 + \sum_{k=1}^{j} \sum_{n=1}^{\infty} \frac{[m(1-p)]^n}{n!} \left[\frac{(j-1)!}{(n-1)!(j-n)!}\right] p^{j-n}\right) \tag{19}$$

As shown in Figure 9, this compound distribution is quite acceptable. The use of a Kolmogorov-Smirnov test shows that the null hypothesis cannot be rejected at the 10 percent level of significance. In a way we have thus verified that the assumptions of the model such as independence between storm groups and identical distribution of rainfall amount per groups, have not been violated. A Monte Carlo simulation can now be set up to generate an unordered succession of seasonal combinations of events. Further research also includes the extension of this investigation to other geographic areas.

CONCLUSIONS

A framework has been constructed for the stochastic analysis of winter precipitation in semi-arid lands. For convenience sake, a storm group was defined as a sequence of days consisting of no two consecutive dry days with a total amount of rainfall equal to or greater than a certain threshold value. In view of the results, the noticeable features of the winter rainfall in this area are:

(1) The number of occurrence of storm group is Poisson distributed, regardless of threshold values. The logarithm of the parameter of the distribution decreases linearly as the magnitude of threshold is increased.

(2) The storm group duration may be described by a negative binomial distribution. The results of the sensitivity analysis indicate that the parameter q decreases and the parameter r increases with increasing threshold.

(3) Given that a storm group has occurred, the frequency of rainfall amount per group was found to have a mixed distribution, composed for example, of a geometric and a uniform distribution. There is little correlation between storm group duration and its depth.

(4) The distribution of seasonal maximal rainfall and the recurrence interval of a given precipitation amount can be computed from this model.

(5) The summation of a random number of random variables yields the distribution of total seasonal rainfall, which appears to follow a compound Poisson distribution.

ACKNOWLEDGMENTS

Part of the research results reported in this paper have been supported by matching grant project B-012-ARIZ, under the direction of Professor Sol Resnick, on "Hydrological and Environmental Controls on Water Management in Semi-Arid Urban Areas", from the Office of Water Resources Research, U.S. Department of the Interior. The authors would like to thank Professor Theodore Roefs and Mr. Vijay Gupta for their valuable suggestions.

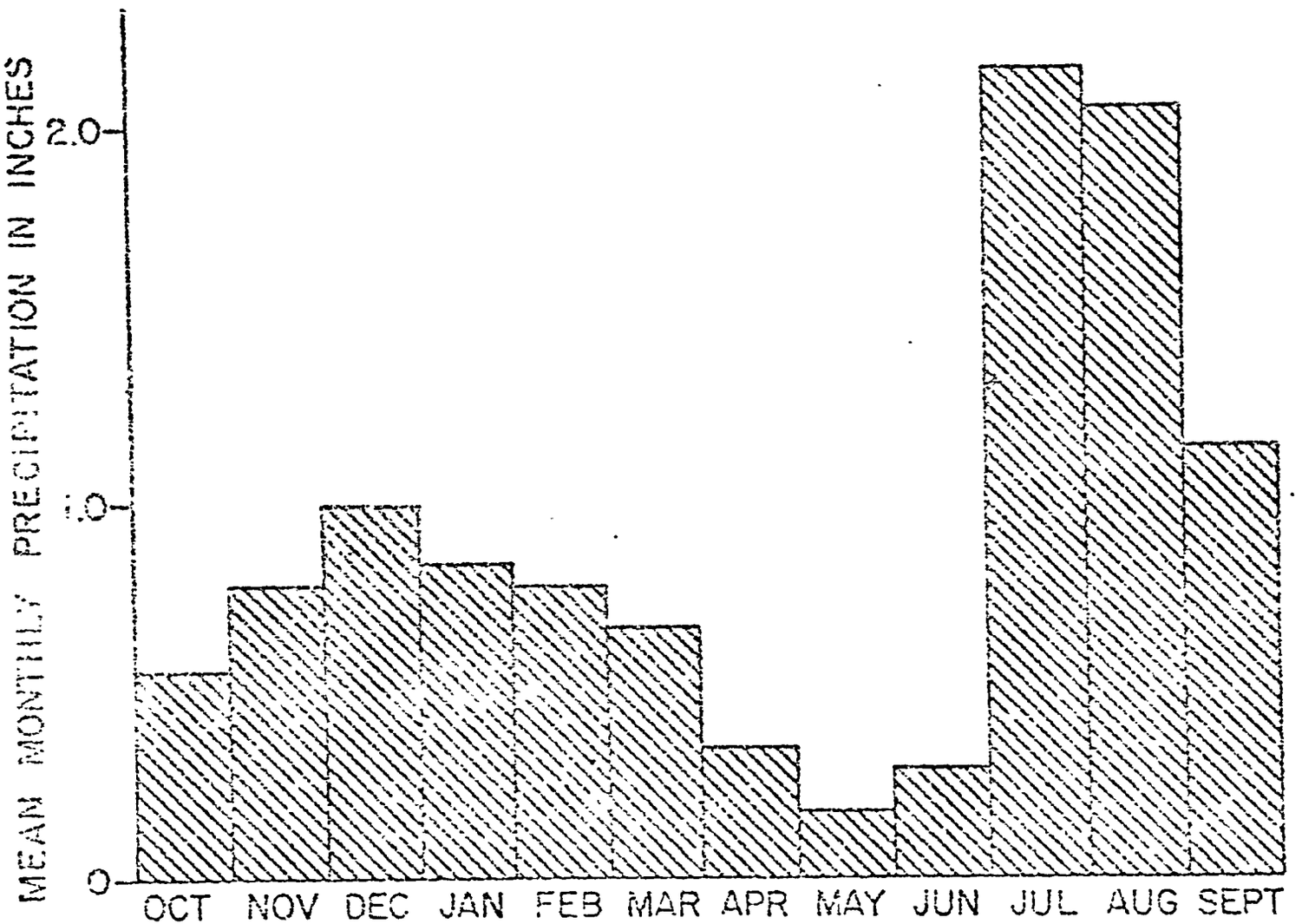

FIG. 1 THE RECORD OF LONG TERM MONTHLY MEAN AT THE STATION OF TUCSON, UNIVERSITY OF ARIZONA (1914 - 1970)

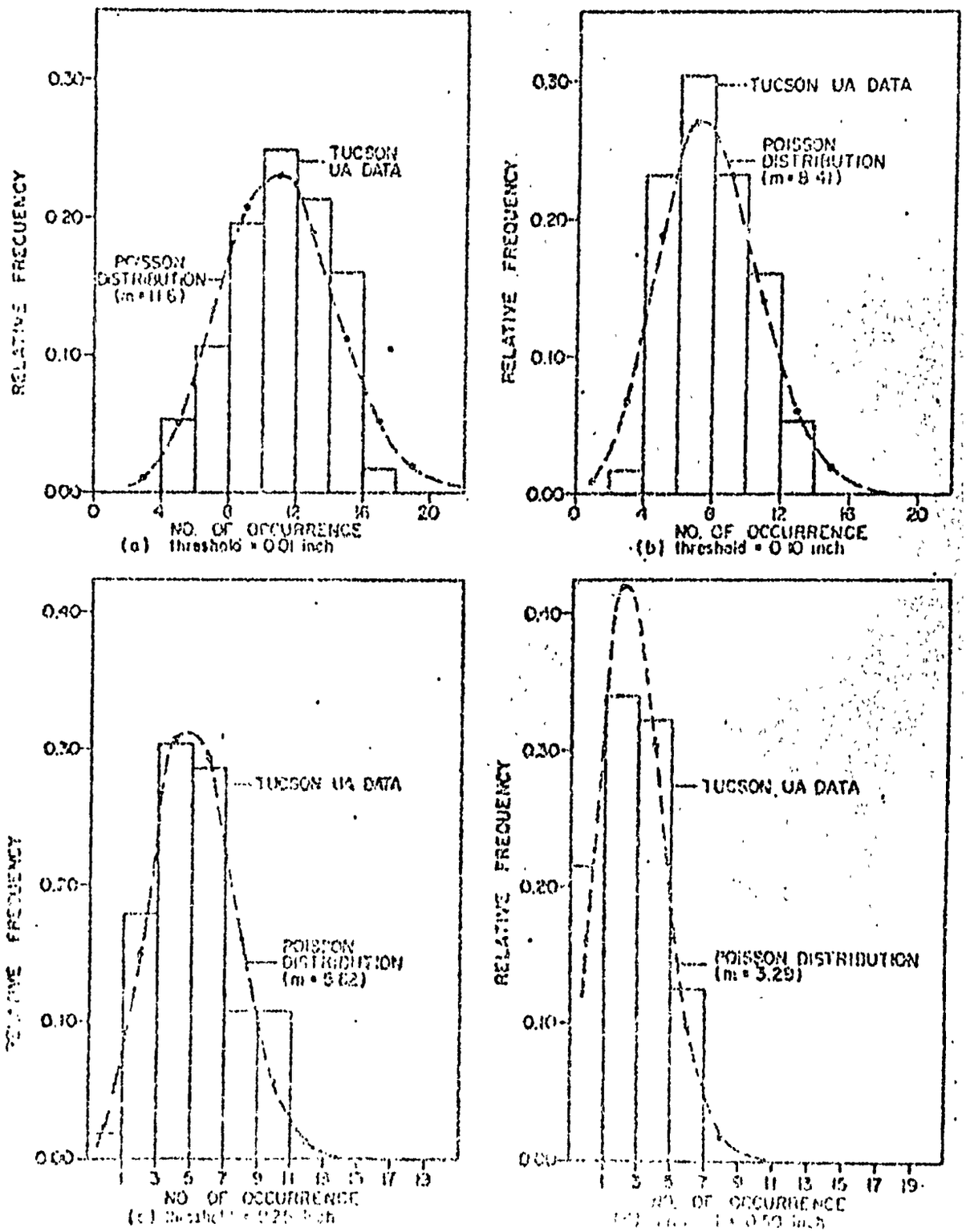

FIG. 2 DISTRIBUTION OF NO. OF OCCURRENCE OF STORM GROUP PER SEASON

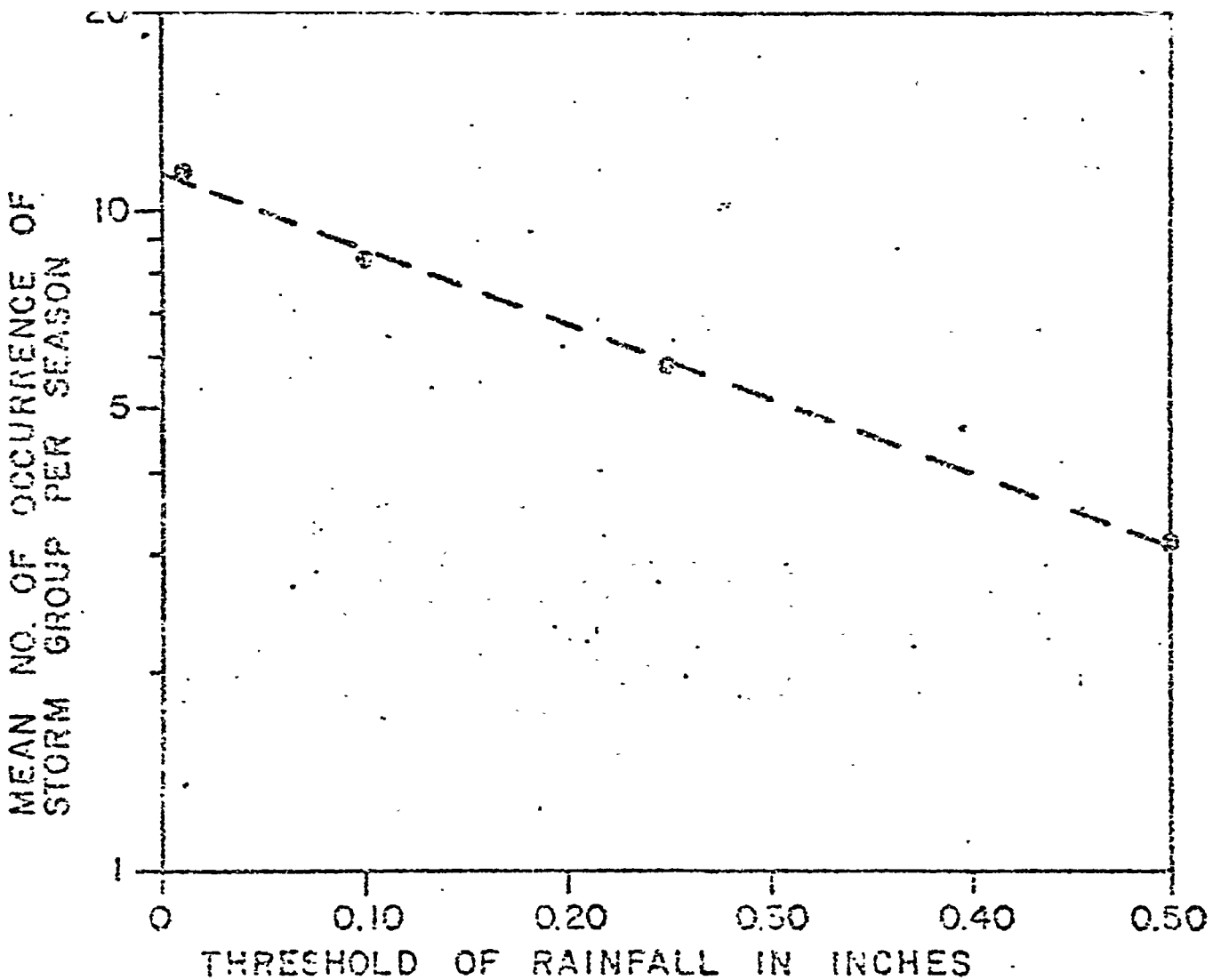

FIG. 3 RELATIONSHIP BETWEEN PARAMETER m AND THRESHOLD OF RAINFALL

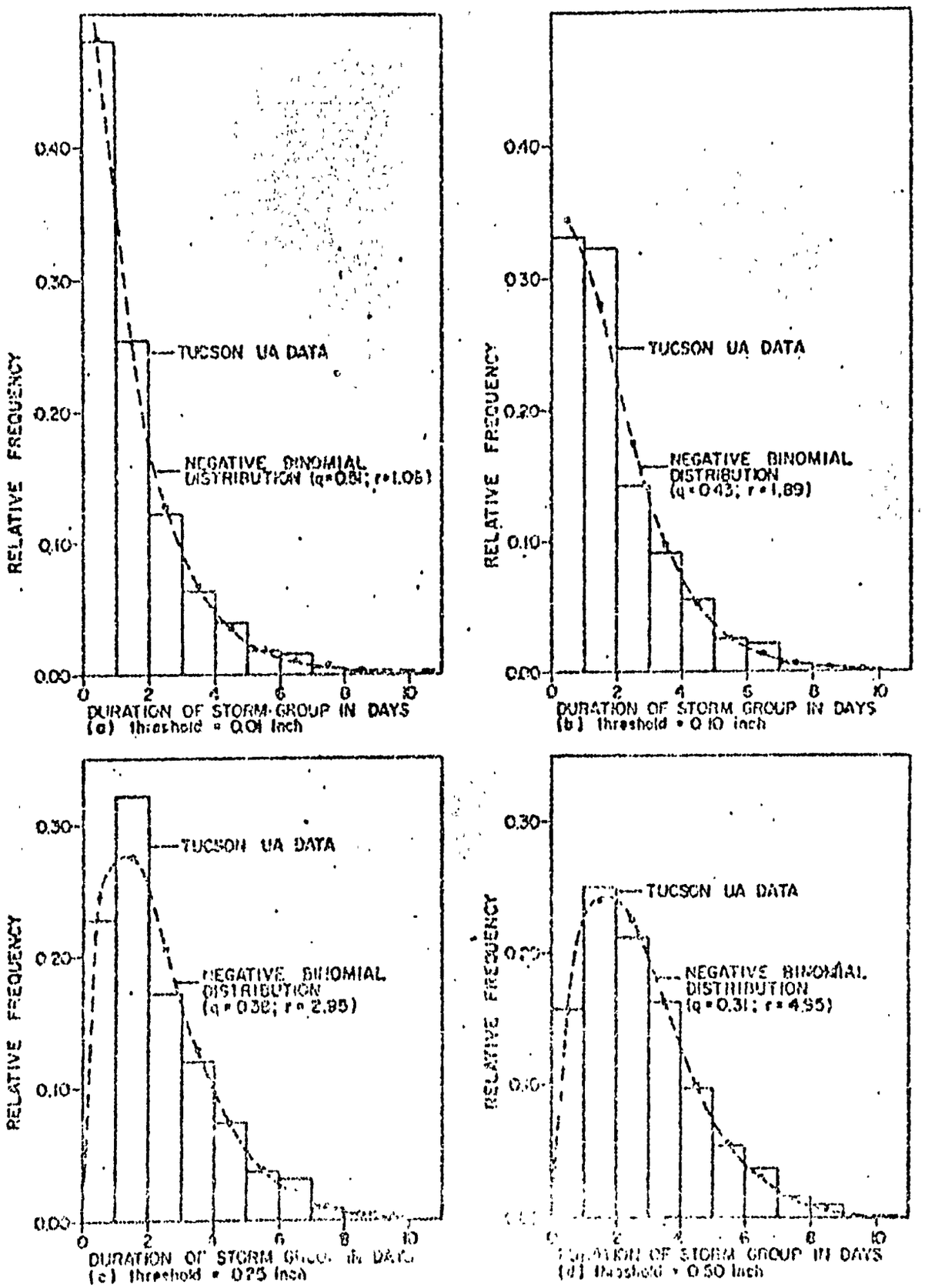

FIG. 4 DISTRIBUTION OF STORM GROUP DURATION

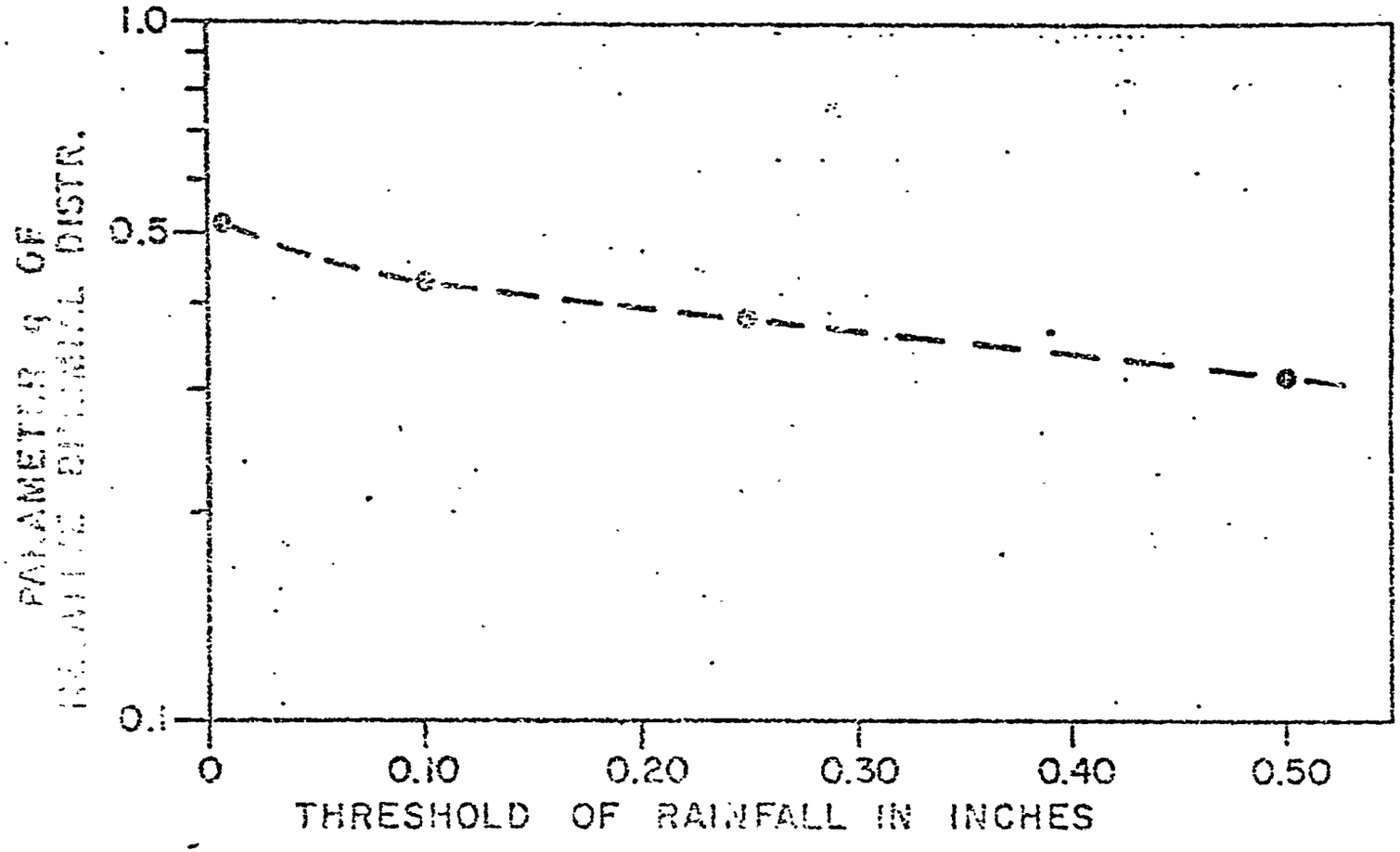

FIG. 5 RELATIONSHIP BETWEEN PARAMETER q AND THRESHOLD OF RAINFALL

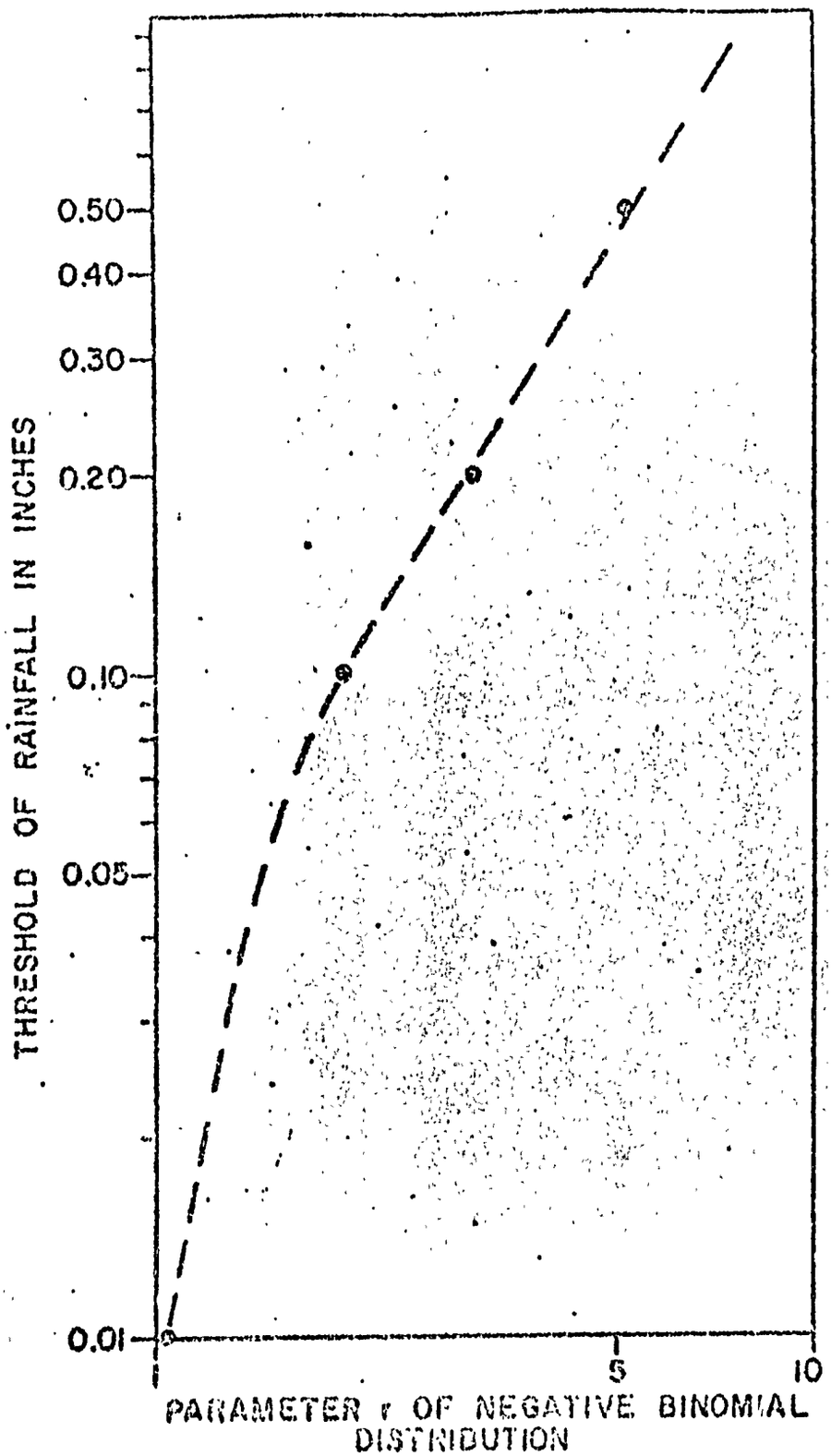

FIG. 6 RELATIONSHIP BETWEEN PARAMETER r AND THRESHOLD OF RAINFALL

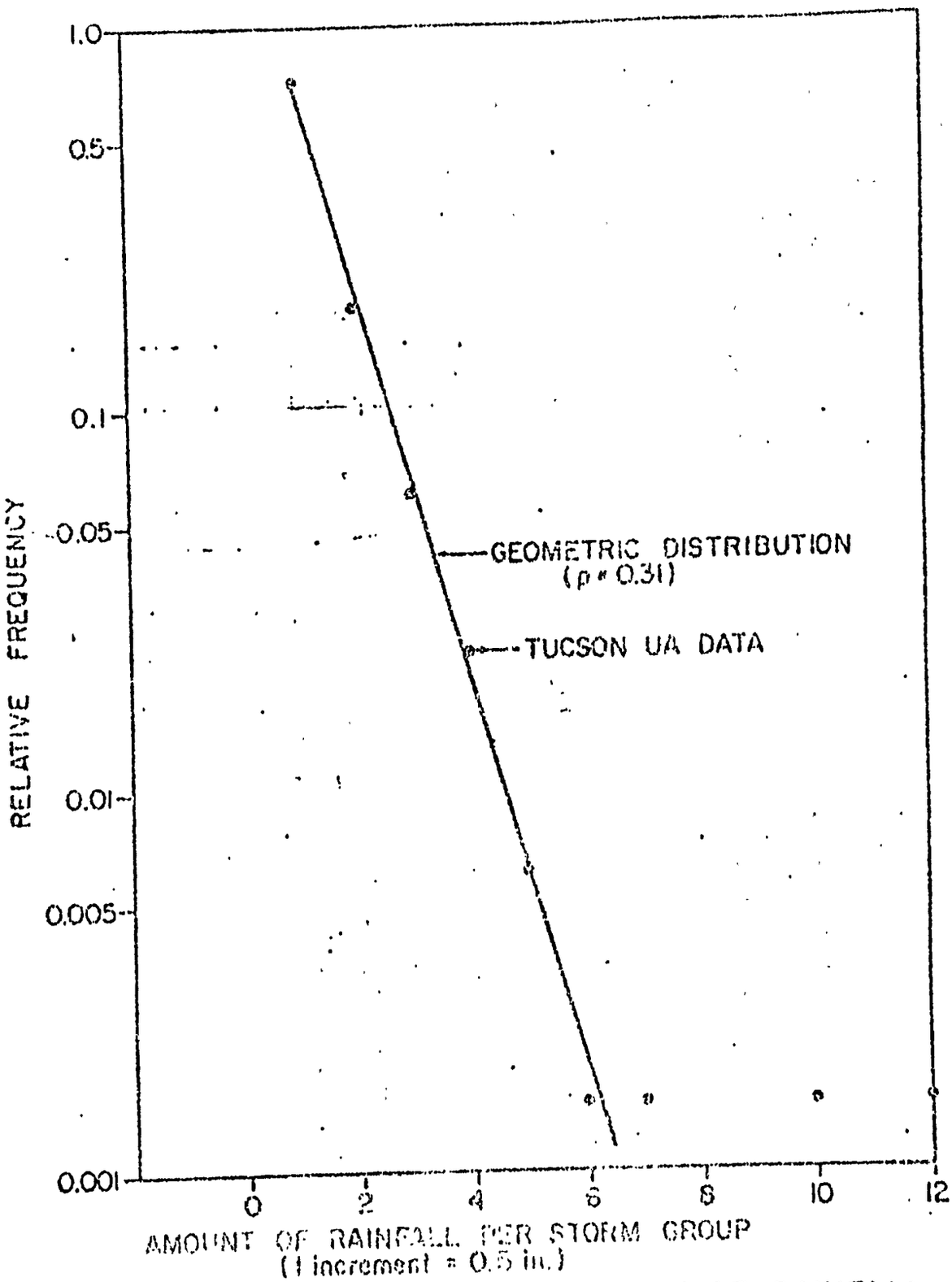

FIG. 7 DISTRIBUTION OF AMOUNT OF RAINFALL PER STORM GROUP

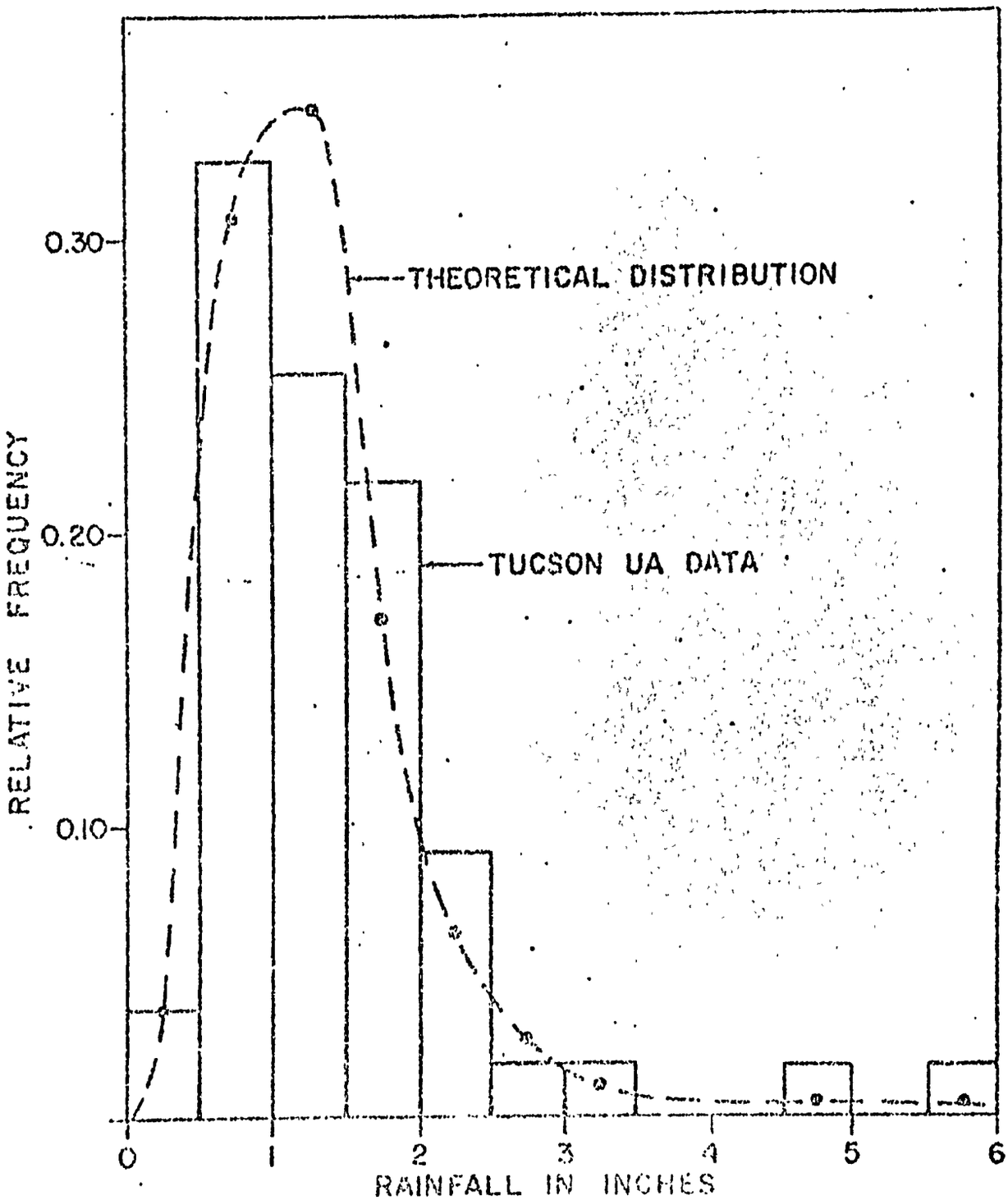

FIG. 6 DISTRIBUTION OF SEASONAL MAXIMUM EVENTS

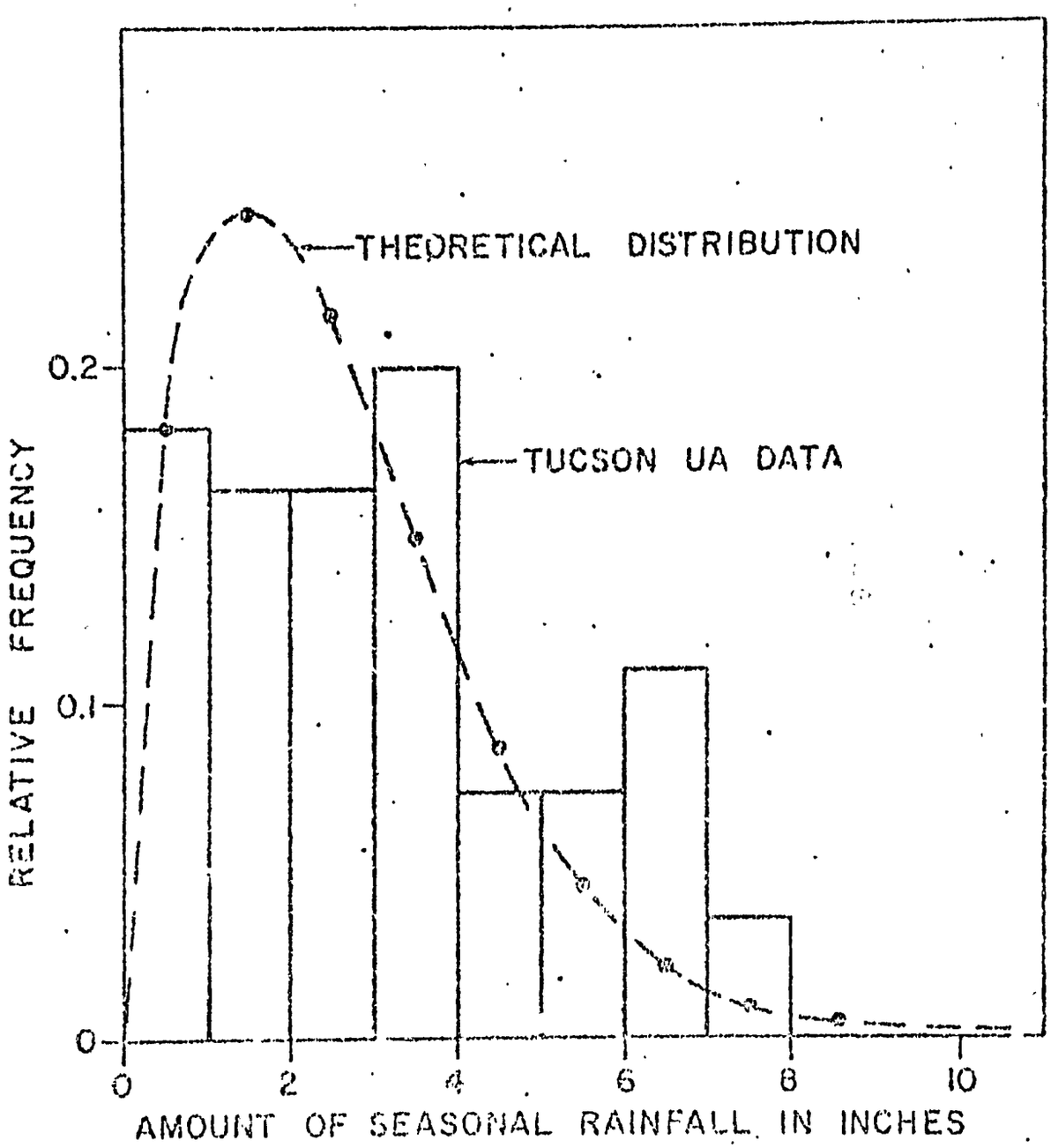

FIG. 9 DISTRIBUTION OF TOTAL SEASONAL RAINFALL

REFERENCES

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

Borgman, L.E., Risk Criteria, J. ASCE, WW3, Vol. 89, Aug. 1-35, 1963.

Feller, W., An Introduction to Probability Theory and Its Applications, Vol. I, John Wiley and Sons, Inc., New York, 1950.

Fogel, M.M. and L. Duckstein, Point Rainfall Frequencies in Convective Storms, Water Resources Res., 5(6): 1229-1237, 1969.

Gabriel, K.R., and J. Neumann, A Markov Chain Model for Daily Rainfall Occurrence at Tel Aviv, Quart. J. Royal Meteo. Soc., Vol. 88, No. 375, 90-95, January 1962.

Jurwitz, L.R., Arizona's Two-Season Rainfall Pattern, Weatherwise, Vol. 6, 96-99, 1953.

Longley, R.W., The Length of Dry and Wet Periods, Quart. J. Royal Meteo. Soc., Vol. 79, No. 342, 520-527, October, 1953.

Sellers, W.D., The Climate of Arizona, in Arizona Climate, by C.R. Green and W.D. Sellers, University of Arizona Press, Tucson, 1964.

Todorovic, P. and V. Yevjevich, 1969, Stochastic Process of Precipitation, Colorado State Univ. Hydr. Paper No. 35, Fort Collins, Colorado.

Weiss, L.L., Sequences of Wet or Dry Days Described by a Markov Chain Probability Model, Monthly Weather Review, Vol. 92, No. 4, 169-176, Apr. 1964.

Wiesner, C.J., Hydrometeorology, Chapman and Hall, Ltd. London, 1970.