Symposium on Statistical Hydrology, August, 1971, Tucson, AZ.
. USDA·Misc. Pub. #1275. 1974. # 123

STOCHASTIC MODELS OF SPATIAL AND TEMPORAL DISTRIBUTION OF THUNDERSTORM RAINFALL [1]

By H. B. Osborn, L. J. Lane, and R. S. Kagan [2]

Abstract

A simplified stochastic model based on airmass thunderstorm rainfall data from the 58-square-mile Walnut Gulch Experimental Watershed in southeastern Arizona is being developed at the Southwest Watershed Research Center, Tucson, Ariz. Records from the 95 rain gage network on this watershed provide valuable information on airmass thunderstorm rainfall in the Southwestern United States. Probability distributions are being used to model random variables—number of cells, spatial distribution of the cells, and cell center depths—of thunderstorms in a summer rainy season. A computer program produces synthetic thunderstorm rainfall based on these distributional assumptions. The synthetic data are compared, with respect to storm center depths and isohyetal map characteristics, with data from the dense rain gage network on Walnut Gulch.

The daily and hourly chances of occurrence of seasonal airmass thunderstorm rainfall are modeled. Efforts are being made to model the temporal distribution of rainfall from individual cells within the airmass thunderstorm.

Finally, the question of model transferability to other regions and locations is tied to defining regional meteorology and local topography.

Introduction

Chow (*1*) and others have defined and differentiated among deterministic, stochastic, and probabilistic processes and models in hydrology, pointing out that stochastic processes follow probabilistic laws and are time dependent, whereas purely probabilistic models are time independent. In short, stochastic modeling in hydrology is the sequential generation of hydrologic information considered wholly or partly random in nature. In this paper, stochastic models of the spatial and temporal distribution of thunderstorm rainfall are considered, and an example based on airmass thunderstorm rainfall is formulated. The authors are more familiar with thunderstorms of the Southwestern United States than for any other region, so the discussion and analyses will be based on southwestern thunderstorms.

Thunderstorms

Thunderstorms are an important source of rainfall in the Southwest. Because of the extreme variability in thunderstorm rainfall both in time and space and the difficulty in measuring this variability, most publications on the subject have been more qualitative than quantitative. Among the publications of interest are those by MacDonald (*9, 10*), Sellers (*21*), Woolhiser and Schwalen (*22*), Osborn and Reynolds (*18*), Osborn (*14, 15*), Osborn and Hickok (*16*), Drissel and Osborn (*3*), Fogel (*5*), and Fogel and Duckstein (*6*). The last nine of these publications also contain attempts at quantifying thunderstorm rainfall as well as containing qualitative description.

Petterssen (*19*) made the following distinction between thunderstorm types.

> Outside the intertropical belt, thunderstorms are observed to occur in three easily recognized patterns. (1) When an air mass is convectively unstable, sufficiently warm and moist, thunderstorms will be released in the upglide motion associated with frontal zones. Although the storms may be widely scattered, the general pattern moves along with the fronts with which they are associated. They are usually referred to as *frontal thunderstorms*. (2) Within more or less uniform air masses one finds an irregular pattern of individual storms, or clusters of such storms. These, which are usually referred to as *air-mass thunderstorms*, show a pronounced diurnal variation with a maximum in the afternoon

[1] Contribution of the Agricultural Research Service, USDA, in cooperation with the Arizona Agricultural Experiment Station, Tucson, Ariz.

[2] Research hydraulic engineer, hydrologist, and engineering technician, respectively, Southwest Watershed Research Center, 442 East Seventh Street, Tucson, Ariz. 85705.

or early evening. (3) Analyses of radar scopes show that thunderstorms not associated with fronts often have a tendency to be arranged in lines or bands more or less along the direction of the wind at low levels. These are called *line thunderstorms*.

Unfortunately, the delineation between thunderstorm types is not always so easily recognized. Regions subject to significant numbers of airmass thunderstorms, such as the Southwestern United States, also may be subject to varying degrees of frontal activity, and sometimes these fronts are difficult to detect. Also, different types of thunderstorms may be dominant in different parts of a region, and frontal activity may vary within a region.

For example, on the arid and semiarid rangelands of southeastern Arizona and southwestern New Mexico (as well as many other regions), airmass thunderstorms produce well over one-half of the average annual precipitation and almost all of the annual surface runoff. In other arid and semiarid regions of the Southwest, frontal activity is more common, and either frontal, airmass thunderstorms, or both, are common. At higher elevations in the Southwest, winter rain and snow are more important sources of water yield to the valleys below.

The Southwest is a region where fronts tend to dissipate or disappear from weather maps; yet they may still influence thunderstorm activity. Thunderstorm buildup will vary with the amount and distribution of moist air aloft, temperatures at various levels, and the winds aloft. In southern Arizona, for example, airmass thunderstorms result from a combination of convective heating and moist air moving into the region from the south, generally from the Gulf of Mexico, but occasionally from the Pacific Ocean. Moist air from the Gulf of Mexico usually is drier than that from the Pacific (because of distance traveled and mountains crossed), and when thunderstorm activity on a given day or during a few consecutive days is more prolonged and the thunderstorms are more closely spaced than usual, the source of moisture is usually the Pacific Ocean. However, there are "in-between" regions where, without meteorological information, one cannot guess the origin of the moisture. Also, if atmospheric conditions are such that the flow of moist air from the Gulf of Mexico continues uninterrupted for a long enough period, thunderstorm activity may be similar to that which occurs when moist air moves into the region from the Pacific. In southeastern Arizona, almost all runoff-producing rainfall on watersheds of 100 square miles or less appears, at least from analysis of recording rain gage records, to result from airmass thunderstorms.

Sellers (*21*) described occasional September storms as "rampaging" across southern Arizona. These storms develop as warm, moist air is pushed into southern Arizona from the Pacific by tropical storms. A combination of one or more of three conditions—orographic lifting, convective heating, and colder air pushing from the north under the advancing warm, moist air—produces more general rains with thunderstorm activity throughout the period, rather than just in the afternoon and evening hours. In reality, these storms probably should be a subclass under frontal thunderstorms because convective heating is an important part of much of the thunderstorm activity within the overall storm period. Possibly, they should be classified as *frontal-convective thunderstorms.*

In general, the occurrence of a thunderstorm at a particular point or over a particular small area within a climatic region appears purely random, and the depth and intensity of rainfall and the area covered by varying depths and intensities of rainfall appear, within limits, to be random. Therefore, thunderstorm rainfall appears to fit very neatly the definition of a stochastic process in hydrology. However, there should be considerable latitude in the assumptions and mathematical representations of such thunderstorms depending upon the amount and accuracy of available information and the proposed use of the model.

Stochastic Thunderstorm Models

Storm systems producing thunderstorms are difficult to classify without simplification; yet simplification is necessary both in definition and classification of the systems and in the eventual modeling of the systems. Rosenblueth and Wiener (*20*) stated:

> No substantial part of the universe is so simple that it can be grasped and controlled without abstraction. Abstraction consists in replacing the part of the universe under consideration by a model of similar but simpler structure. Models, formal or intellectual on one hand, or material on the other, are thus a central necessity of scientific procedure.

In general, stochastic thunderstorm rainfall models are either physically based, data based, or both. Ideally, models based on atmospheric and topographic conditions might be preferred, but realistically, most models are based on data collected at the ground surface and are developed without atmospheric parameters. Since thunderstorm rainfall is highly variable both spatially and temporally in space and this variability is difficult to measure, any great degree of sophistication of thunderstorm rainfall models not based on atmospheric data may be suspect. This is particularly true if the end result is to predict runoff, where uncertainties in watershed response add to the uncertainty of the output and may limit runoff models to rather simple inputs and "black box" techniques.

LeCam (*8*) developed a theoretical model for rainfall as a random phenomenon incorporating yearly periodicity. The model was described as a clustering process of the type presented by Neyman and Scott (*12*, *13*). LeCam's lucid description of rainfall occurrence and his comments on validating or testing such models are especially relevant as the complexity of models increases.

Airmass Thunderstorm Rainfall Model

As an example, a simplified stochastic model incorporating the spatial and temporal distribution of thunderstorm rainfall was developed from rain gage records of airmass thunderstorm rainfall on the Walnut Gulch Experimental Watershed, Tombstone, Ariz. (fig. 1). The Southwest Watershed

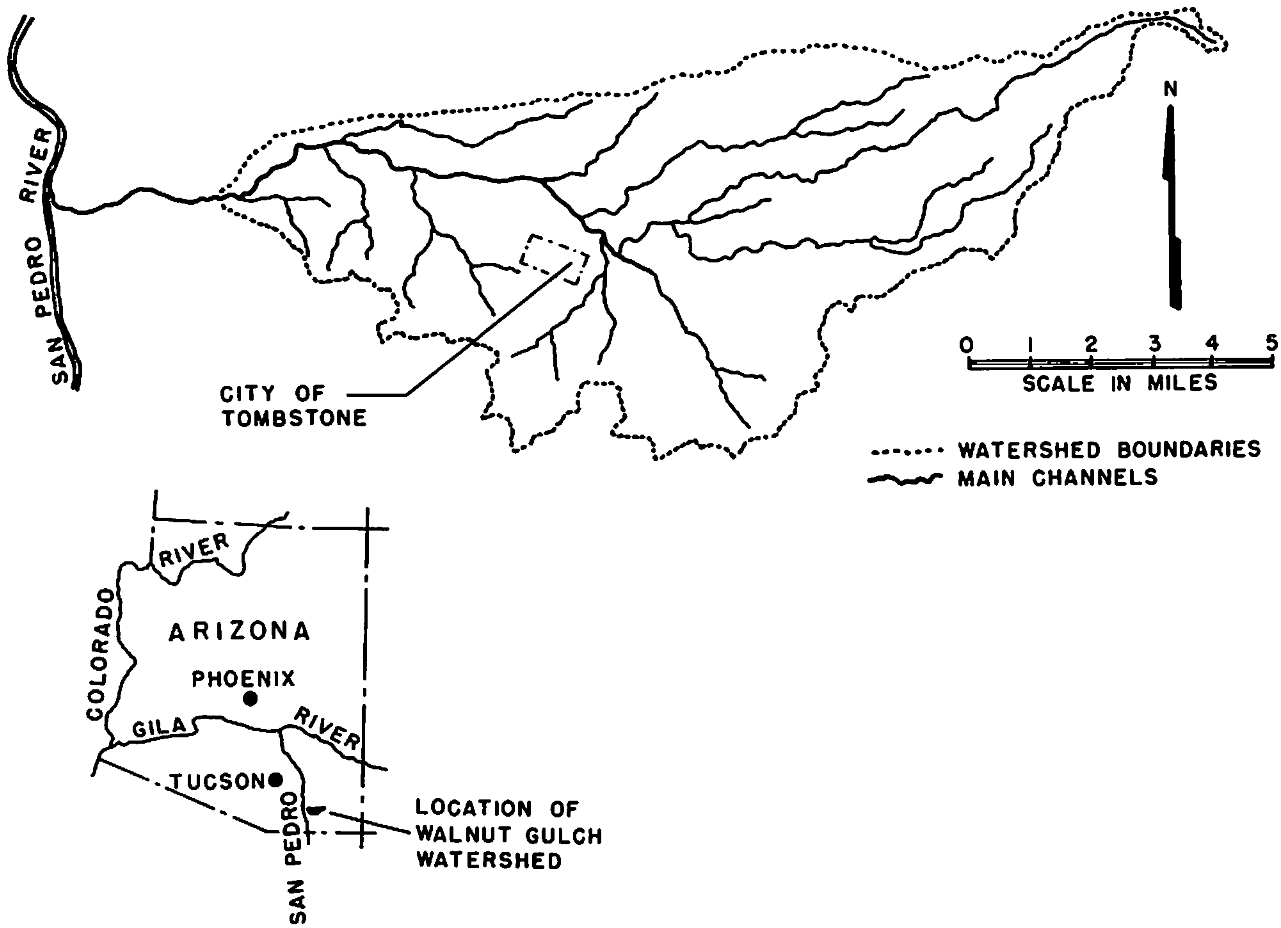


FIGURE 1.—The Walnut Gulch Watershed.

Research Center of the Agricultural Research Service operates this 58-square-mile experimental rangeland watershed. The watershed is representative of semiarid rangelands throughout much of the Southwest. Of the 95 recording rain gages on the watershed, 80 have been in continuous operation for over 10 years. The means and ranges of the variables used in this model were determined from the records from Walnut Gulch. Total storm rainfall for eight selected events is shown to illustrate the variability of thunderstorm rainfall and to indicate visually the difficulties in modeling such rainfall (fig. 2(A–H)).

A stochastic model of thunderstorm rainfall for Walnut Gulch is being developed in three parts. The first part, or routine, in the model determines the chance of daily and hourly occurrence of a significant event. Included in this routine is the chance of more than one event occurring on the same day. The second part of the model generates the total storm rainfall through addition of individual synthetic storm cells regardless of time of occurrence within the storm. Significant progress has been made on the first two parts of the model.

The final part of the model involves generating the cells sequentially and continuously, possibly describing the storm with a series of isohyetal maps of short duration (possibly 10 minutes). Development of the third part of the model will continue after possible modifications and final verification of the first two parts.

Occurrence of an Airmass Thunderstorm Event

An initial attempt at modeling the probability of a thunderstorm occurring during the summer rainy season involved assuming a probability distribution for the start of the rainy season. Once the season had started, the occurrence or nonoccurrence of an event was modeled as a Bernoulli variable with constant parameter throughout the season. However, considering the assumptions about moist air movement stated in the previous section, the assumption about constant probability of occurrence (Bernoulli parameter) throughout the season was not consistant. Analysis of rainfall data from the Walnut Gulch Experimental Watershed indicates that the probability of storm occurrence varies considerably within the rainy season.

Therefore, a variable probability of occurrence of significant thunderstorm rainfall based on 10 years

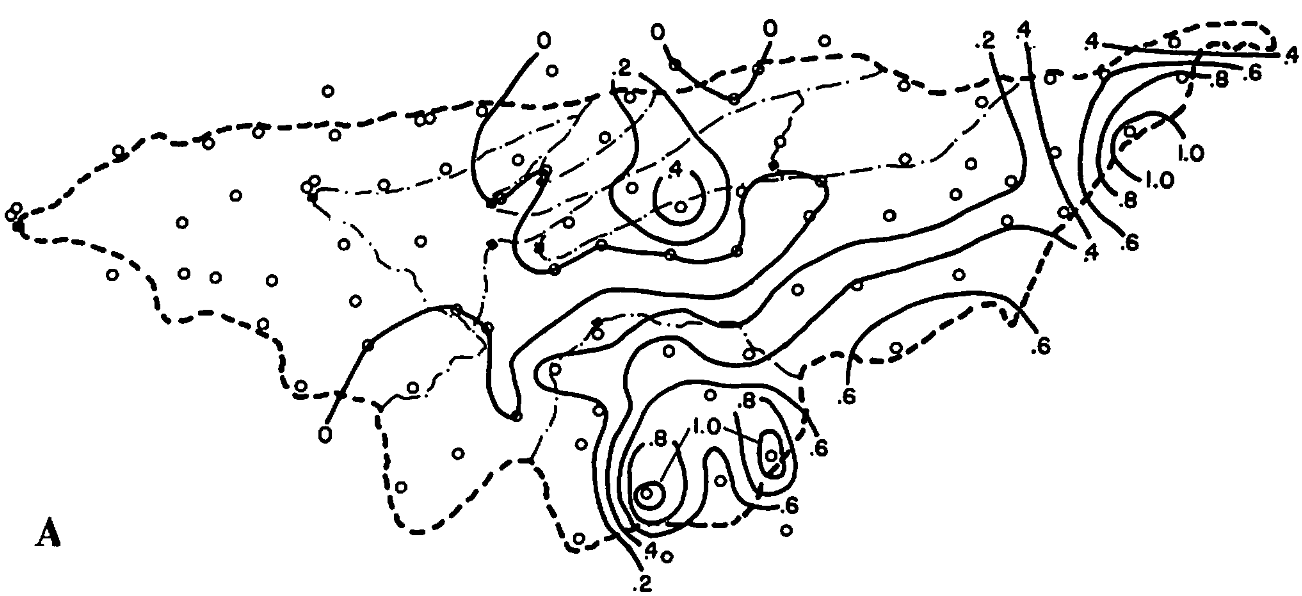
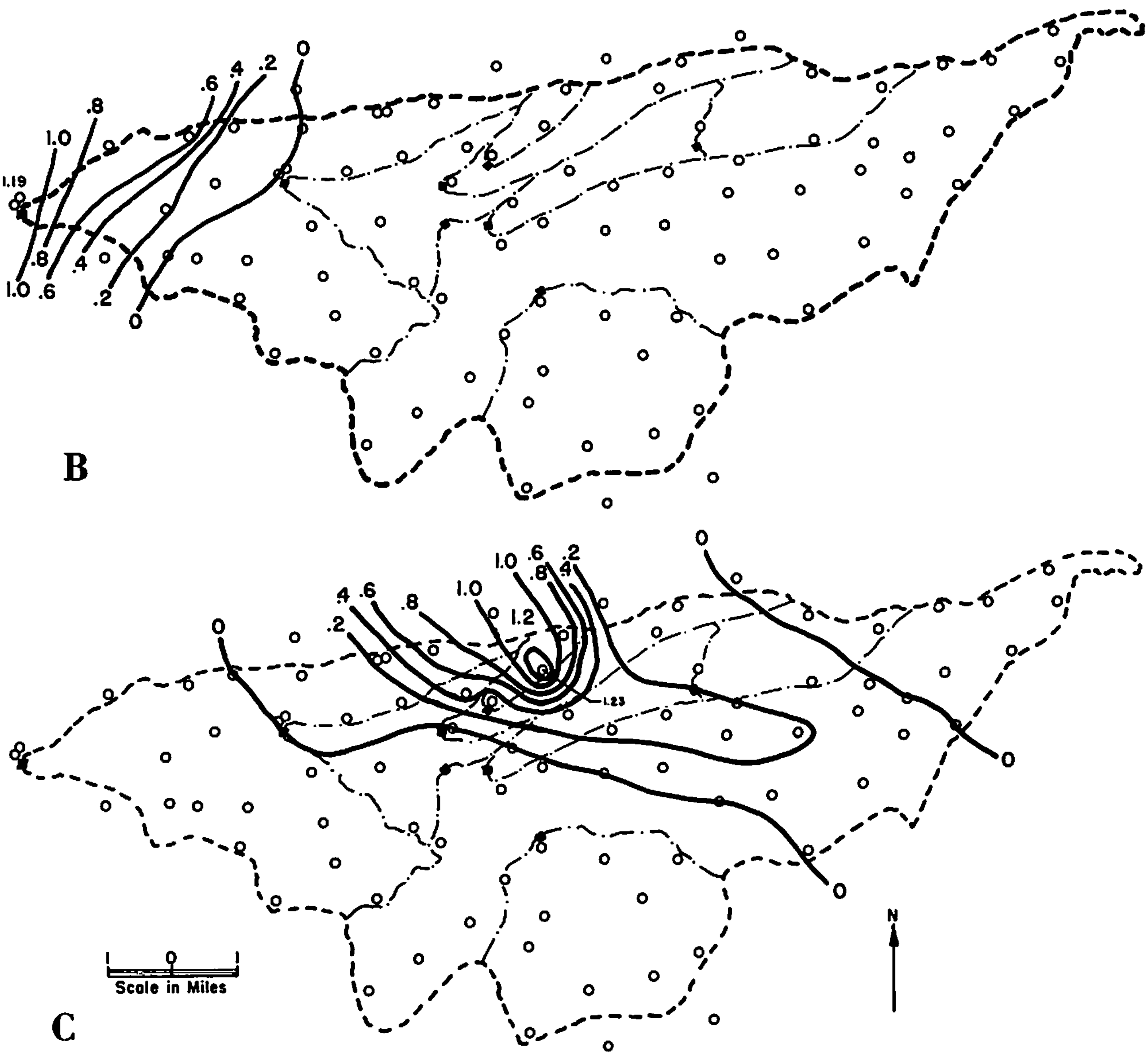


FIGURE 2.—Isohyetal maps of selected thunderstorm rainfall. Walnut Gulch watershed precipitation (in inches): A, Storm of August 12, 1963 (1200). B, Storm of August 16, 1963 (1640). C, Storm of July 13, 1964 (1600). D, Storm of September 11, 1964 (1700). E, Storm of July 29, 1966 (1830). F, Storm of July 7, 1967 (1500). G, Storm of August 3, 1967 (1700). H, Storm of August 13, 1967 (1400).

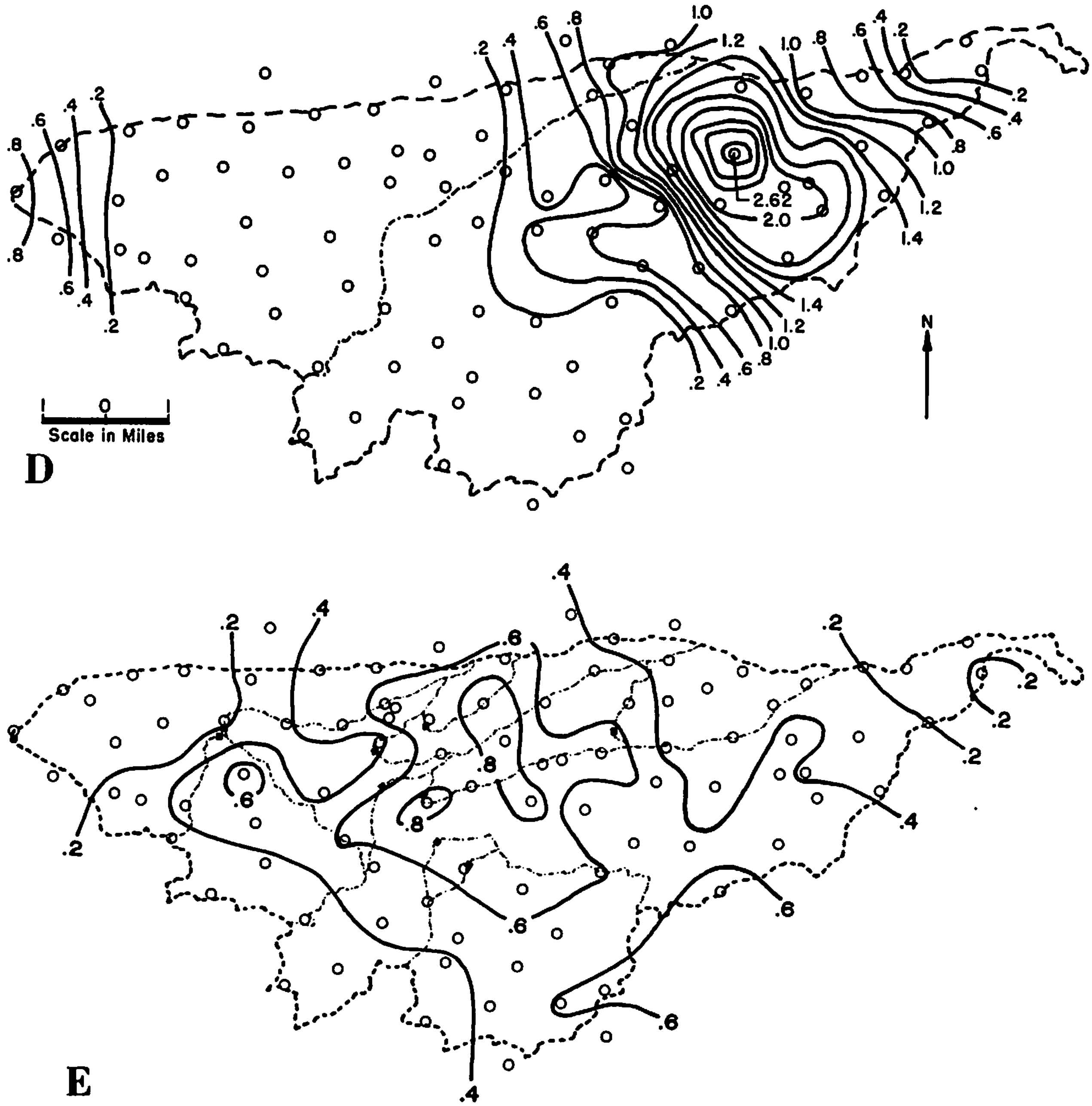


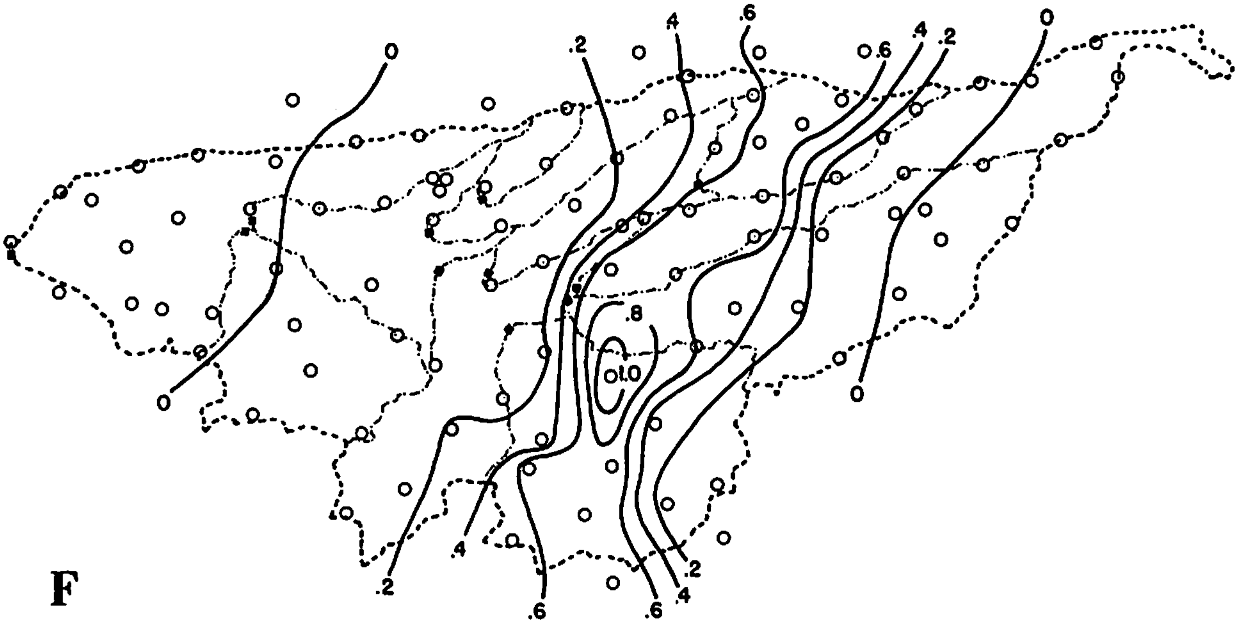
of precipitation data from Walnut Gulch is used to estimate the probability of a significant storm occurring somewhere over the 58-square-mile watershed. A significant storm is specified as one with at least 0.25 inch of rainfall recorded on at least two adjacent rain gages. Effects of modeling the seasonal distribution of daily rainfall by incorporating a varying Bernoulli parameter are discussed in Appendix I. Figure 3 shows the 5-day running mean for the number of significant storms recorded (1960–69) on the Walnut Gulch watershed. The smoother curve shown in figure 3 is arbitrarily adopted in the model. The curve is similar in shape to the point frequency value from long-term U.S. Weather Bureau records from Tombstone, Ariz.

Additional work is in progress to facilitate extrapolating point frequencies from Weather Bureau and other data to provide storm frequencies for finite-sized watersheds throughout the Southwest. The relationship between point and areal frequency on finite sized watersheds for different climates and topographies is essential in regionalizing such a model (*25*).

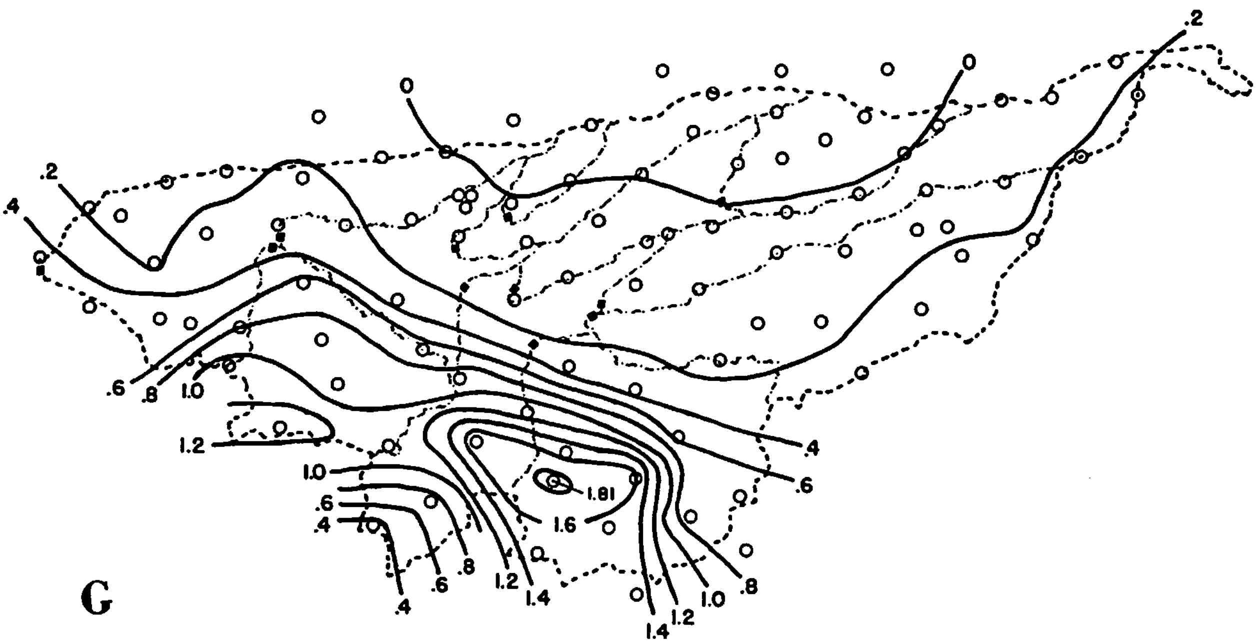
The procedure used here for generating synthetic airmass thunderstorm rainfall data over a finite-sized area is summarized in figure 4. A table of probabilities derived from the "smoothed" curve in figure 3 is used as the Bernoulli parameter p, that is, the probability of a significant storm occurring anywhere on the watershed on a given day, in the sequential generation of a Bernoulli variable. If the Bernoulli variable is equal to zero (a failure), then there is no significant storm on the given day. The date is then indexed and the next Bernoulli variable simulated. If there is a significant storm, then the beginning time of the storm (0000 to 2400 in military time) is generated as a truncated normally distributed random variable with mean starting time 1700 (5:00 p.m.) and a standard deviation of 3.5 hours (2).

The next step is to simulate the airmass thunderstorm, (described in the next section) and to print





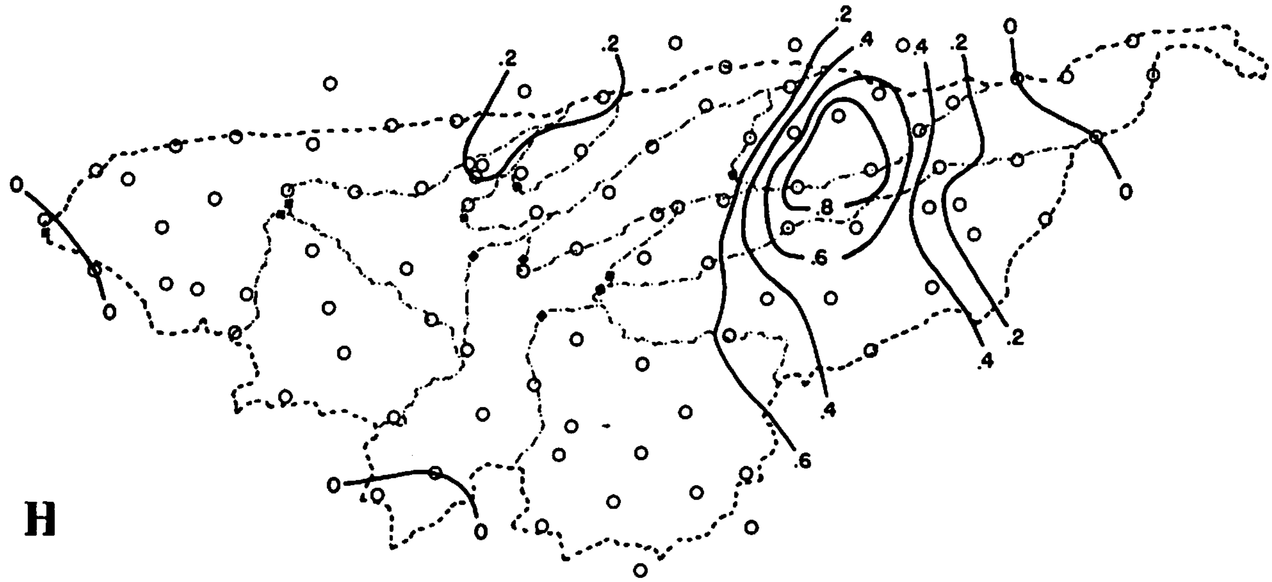
F



G

the necessary data as to date, time, location, and magnitude. The model also allows for multiple storms occurring on the same day. If a storm occurs between 0500 and 1700, there is a reduced probability of another storm occurring 3 hours or more after the beginning of the first storm. If the second storm also occurs before 1700, the same reduced probability, determined by trial and error as one-fifth the original rainfall chance, is used to predict a third storm, and so on. That is, the occurrence of subsequent storms is also modeled as a Bernoulli variable.

Logically, the model should allow for persistence—the tendency for wet days to follow wet days and dry days to follow dry days. However, persistence was not included in this simplified model.



H

Simulation of Airmass Thunderstorms

Once the date and starting hour have been determined, the synthetic storm itself is generated through a group of equations implemented by a program referred to as CELTH–4 (fig. 5). CELTH–4 consists of a unit cell model coupled with a technique of randomly grouping these cells as a means of describing thunderstorm rainfall shapes and distributions. The model development was a combination of simplification, abstraction, physical arguments, and trial and error, with the output as the final test of the combination of distributions.

The unit cell, building block of this rainfall model, was initially chosen to be circular, with the rainfall (D) at any point within the cell dependent only on distance (r) from the center. Individual cells appear more elliptical than circular with a long-short axis ratio of $1\frac{1}{2}$:1 (*6*, *15*), and such a further refinement might be justified depending upon the stated use of the model. Analysis of rainfall data collected from the Walnut Gulch Experimental Watershed (*17*) indicates that the approximate relationship between the distance r and corresponding rainfall D (in inches) is:

$$D = 0.9D_0 \, [1 - K \ln (\sqrt{\pi}\, r)] \qquad (1)$$

for

$$r \geqslant 1/\sqrt{\pi} \text{ miles and}$$

$$D = D_0 \, [1 - \sqrt{\pi}\, r/10] \qquad (2)$$

for

$$r \leqslant 1/\sqrt{\pi} \text{ miles}$$

where D_0 is the center depth (in inches) of the unit cell, $K = 1/\ln (\sqrt{\pi}R)$; R is the cell radius; and ln is the $\log_e$.

Total storm radius R and the center depth D_0 are considered constant within each cell. To insure flexibility in both shape and rainfall distributions between unit cells, D_0 is randomly generated for each cell. Rainfall records on Walnut Gulch suggest that individual cell center depths can be approximated by a negative exponential distribution generated by the equation:

$$D_0 = \bar{D}_0 \ln (1 - U) \qquad (3)$$

where $\bar{D}_0$ is the mean cell center depth obtained from rainfall data and U is a uniform random variable, $0 < U \leqslant 1$.

In this distribution and in subsequent distributions, U is approximated by pseudo-random numbers from a random numbers generator. By keeping R constant and varying D_0 in this manner, a variety of rainfall configurations are obtained, and the rainfall at any point within the cell, is determined in terms of the generated parameter, D_0, and the variable, r.

The choice of the exponential distribution for individual cell center depths also arises from assuming a multicellular model with total storm rainfall modeled as a gamma variable. Specifically, the summation of N exponential variables produces a random variable with a gamma distribution since the gamma densities are closed under convolution (*4*).

The next step was to describe shapes and rainfall distributions of entire storms by grouping these cells.

One might assume that cell occurrence is uniformly random across the rain gage network. However, although the center of each storm has an equal chance of occurring at a given point on the grid, the clustering of cells and analysis of rainfall with respect to time on recorded rainfall isohyets suggests that only the location of the *first* cell is truly random. The remaining cells of the storm tend to group around the first cell and at the same time preserve a direction of storm movement. These observations motivated the introduction of two basic storm parameters: the average number of cells per storm, $\bar{N}$, and the preferred direction of cell placement, θ_0.

$\bar{N}$, determined from rainfall data, is used to govern the number of cells generated per storm (N). Since N is a discrete random variable, and the occurrence rate of cells within the duration of rainfall is assumed constant, it is assumed to have a Poisson distribution limited at the lower end by three cells as suggested by Petterssen (*19*). The average number of cells was determined roughly from Walnut Gulch data, and the chance of having more than seven cells in one storm was very small.

θ_0, on the other hand, is used to locate the direction of the next cell generated. It is the direction of the second cell from the first and is altered by an amount $\Delta\theta$ for each additional cell so that θ_i is the direction of movement, in degrees, after the i'th cell, or:

$$\theta_i = \theta_{i-1} + \Delta\theta_{i-1} \tag{4}$$

where i goes from 1 to N and $\Delta\theta_0 = 0°$.

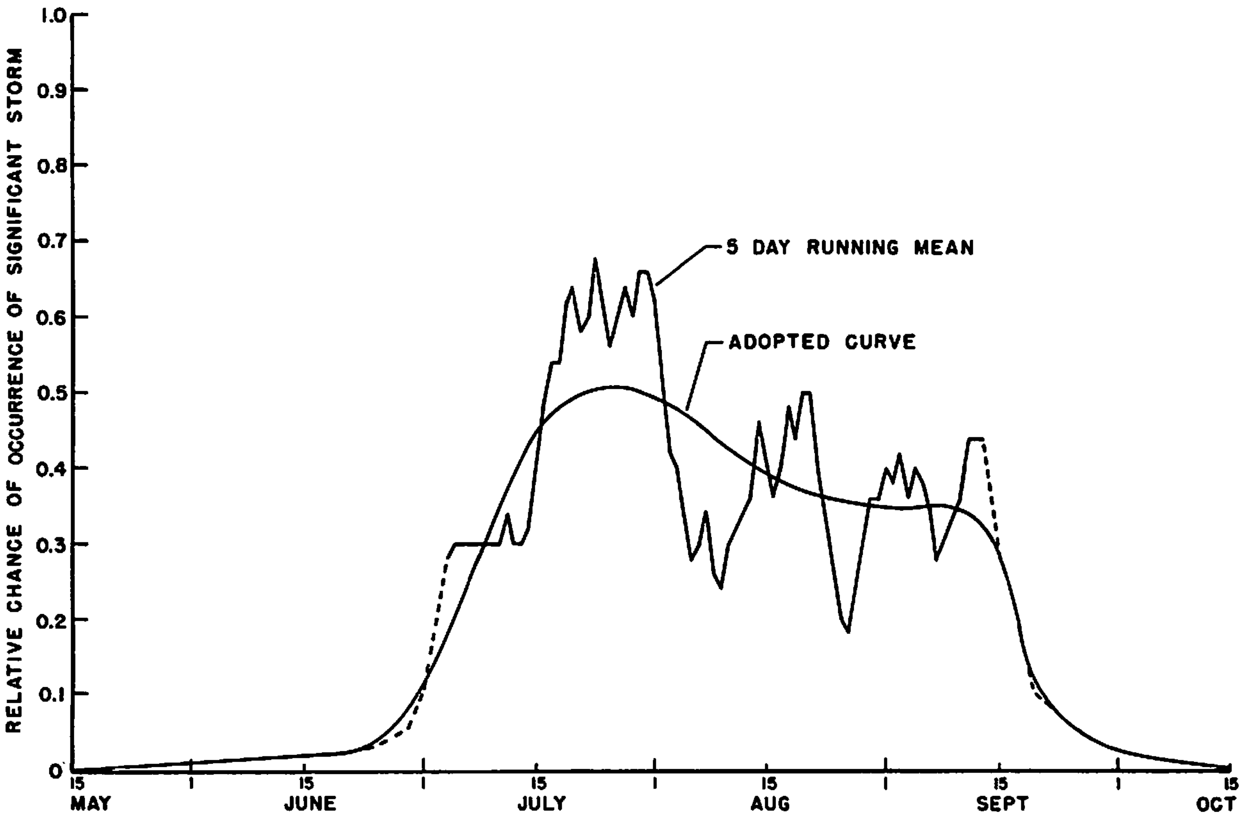


FIGURE 3.—Empirically derived curve for the probability of significant storms on Walnut Gulch watershed.

Although θ_0 has an equal chance of being in any direction (and so, is a uniform random variable), $\Delta\theta_0$ has a directional component, and was initially abitrarily assigned a normal distribution about a mean of 0° with a standard deviation of about 60°. Although arbitrary, this did tend to sustain the direction of storm movement in a manner similar to that observed in real events.

The next step involves determining the distance between cell centers. This is the third storm parameter d, governed by its corresponding mean d as calculated from rainfall information. However, unlike N, θ, and $\Delta\theta$, its distribution is more difficult to determine from available storm data. The distribution is approximated by two lines and generated by:

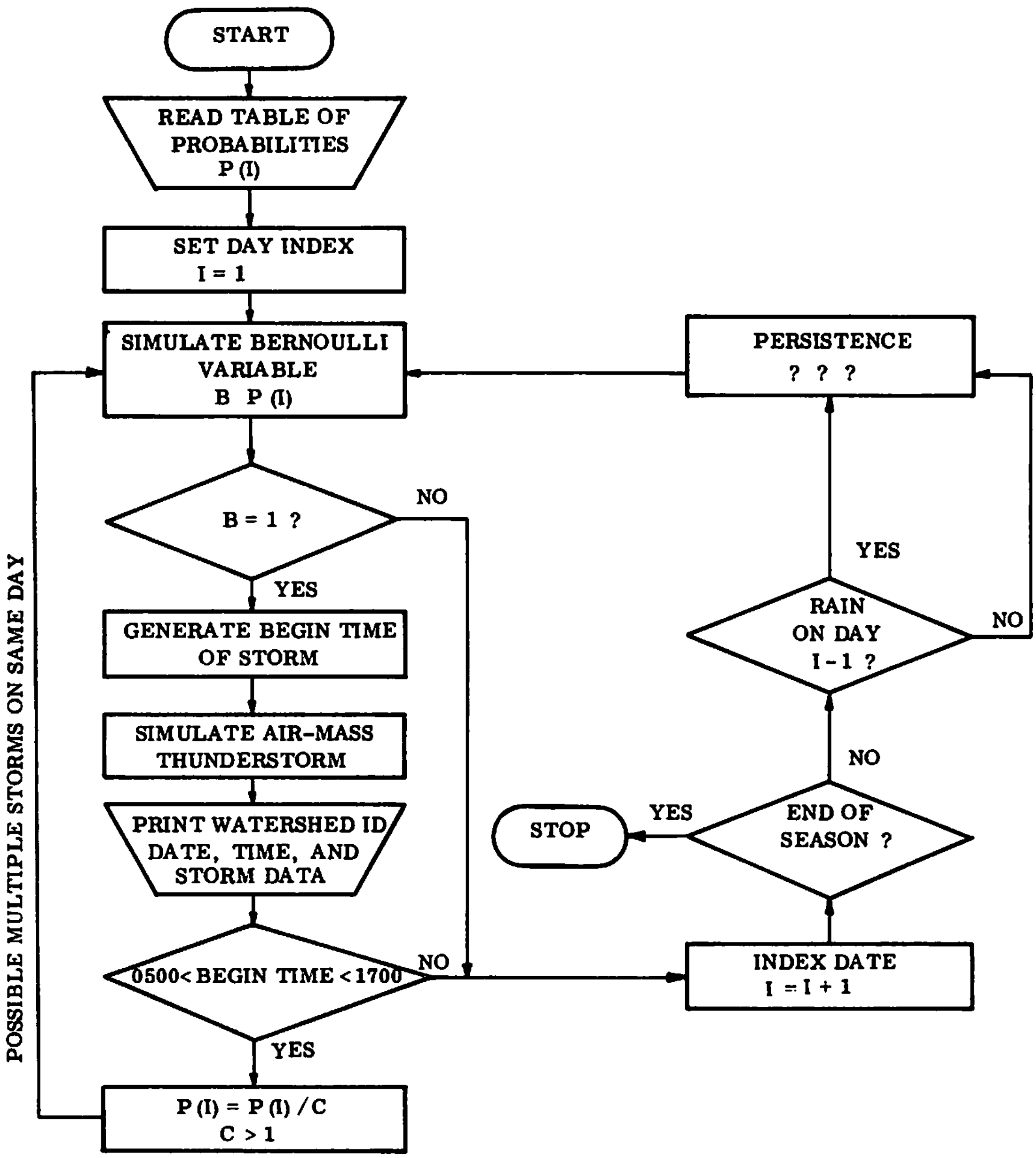


FIGURE 4.—Flow chart for generation of seasonal synthetic rainfall data.

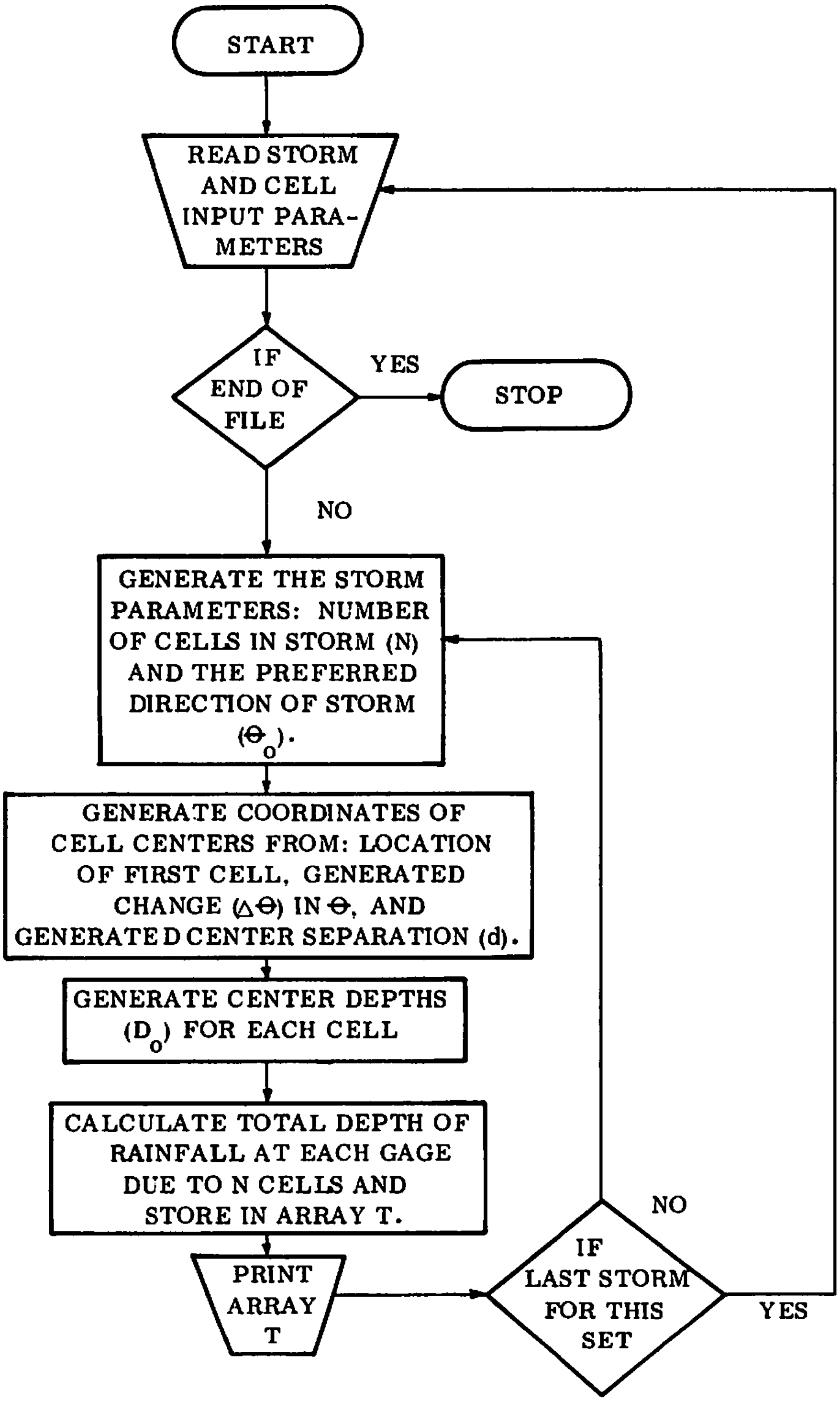


FIGURE 5—Flow chart for simulation of individual airmass thunderstorm rainfall.

$$d = \sqrt{10U} \qquad \text{for } U \leq 0.4 \tag{5}$$

$$d = 0.5\ (10 - \sqrt{60 - 60U}) \qquad \text{for } U \geq 0.4 \tag{6}$$

where U is a uniform random variable, $0 < U \leq 1$, such that the total storm rainfall covers an average of about 40 square miles and a maximum of about 90 square miles with the long axis never greater than about 18 miles (*17*).

The final step in the synthesis involves calculating the location of the first storm rain cell on the Walnut Gulch network. This is done by using a uniform random variable to generate a rain gage number. For example, on a 100-rain-gage grid, the equation used is:

$$I = 100U + 1 \tag{7}$$

where I is the rain gage number and U is a uniform random variable, $0 < U < 1$.

For Walnut Gulch, the first cell of each storm has an equal chance of occurrence at any rain gage on or immediately adjacent to the watershed. This step completes the processes involved in storm generation from cell definition to method of placement. Any number of storms can be generated from the cell and storm parameters in CELTH-4.

Results

Ten years of thunderstorm rainfall were generated for the Walnut Gulch watershed. The first 4 years of synthetic record are shown in figure 6, along with 1 year of actual data, 1963, which was chosen randomly from the 1960 through 1969 records. The horizontal scale in figure 6 represents the summer rainy season from June 15 through October 15. The vertical scales represent the maximum total depth of point rainfall for each storm. The average annual number of storms exceeding given depths and ranges in annual values for the 1960–69 data and the synthetic data are compared in table 1. The location within the season, the length of season, the number of events per season, and the maximum storm depths for the actual and synthetic data appeared to correspond fairly well. However, comparison of the synthetic and real data suggest that persistence is a significant factor in thunderstorm rainfall and should be added to the model. Also, the maximum recorded depth in 10 years of record (1960–69) was 3.45 inches compared to 2.57 inches in the 10 years of synthetic data. For the real data, the range of annual maximum rainfall depths was 1.63 inches to 3.45 inches with a mean of 2.48 inches. For the synthetic data, the range was 1.70 inches to 2.57 inches with a mean of 2.13 inches. Further analysis is necessary to determine whether the 3.45-inch storm has a recurrence interval greater than 10 years, whether the model underestimates maximum depths, or both.

Frequency plots of storm beginning times do not contradict the normality assumption for beginning time, and the mean and standard deviation values of 1700 and 3.5 hours appear reasonable. Other studies of precipitation (*11*) and of runoff (*2*) also point to a preponderance of late afternoon and early evening storms in southeastern Arizona.

As an example, eight synthetic events from year 5 (fig. 7 A–H) were chosen to compare with the real events in figure 2 A–H. Comparison of eight isohyetal maps, as well as the full 10 years of synthetic data with real rainfall maps, suggests that while the synthetic storms compare to some real events,

TABLE 1.—*Comparison of maximum storm depths between 10 years of Walnut Gulch data (1960–69) and 10 years of synthetic data*

Item	Number of events annually equal to or exceeding given depths of—											
	0.6 inch			1.0 inch			1.4 inches			1.8 inches		
	Max.	Min.	Ave.	Max.	Min.	Ave.	Max.	Min.	Ave.	Max.	Min.	Ave.
Actual data (1960–69)	25	10	19	13	6	10	8	2	5	6	0	3
Synthetic data (10 years)	27	13	20	14	6	10	8	3	5	4	0	2

the actual thunderstorm rainfall is far more complex than this simplified model. Further sophistication such as elliptical cells might improve the model, but within the limited ability to test the accuracy of the model such a refinement might not be justified at present.

Discussion

The synthetic data produced were compared both on seasonal characteristics and on individual storm center depths and isohyetal map characteristics with data from the dense rain gage network on Walnut Gulch. These comparisons indicated

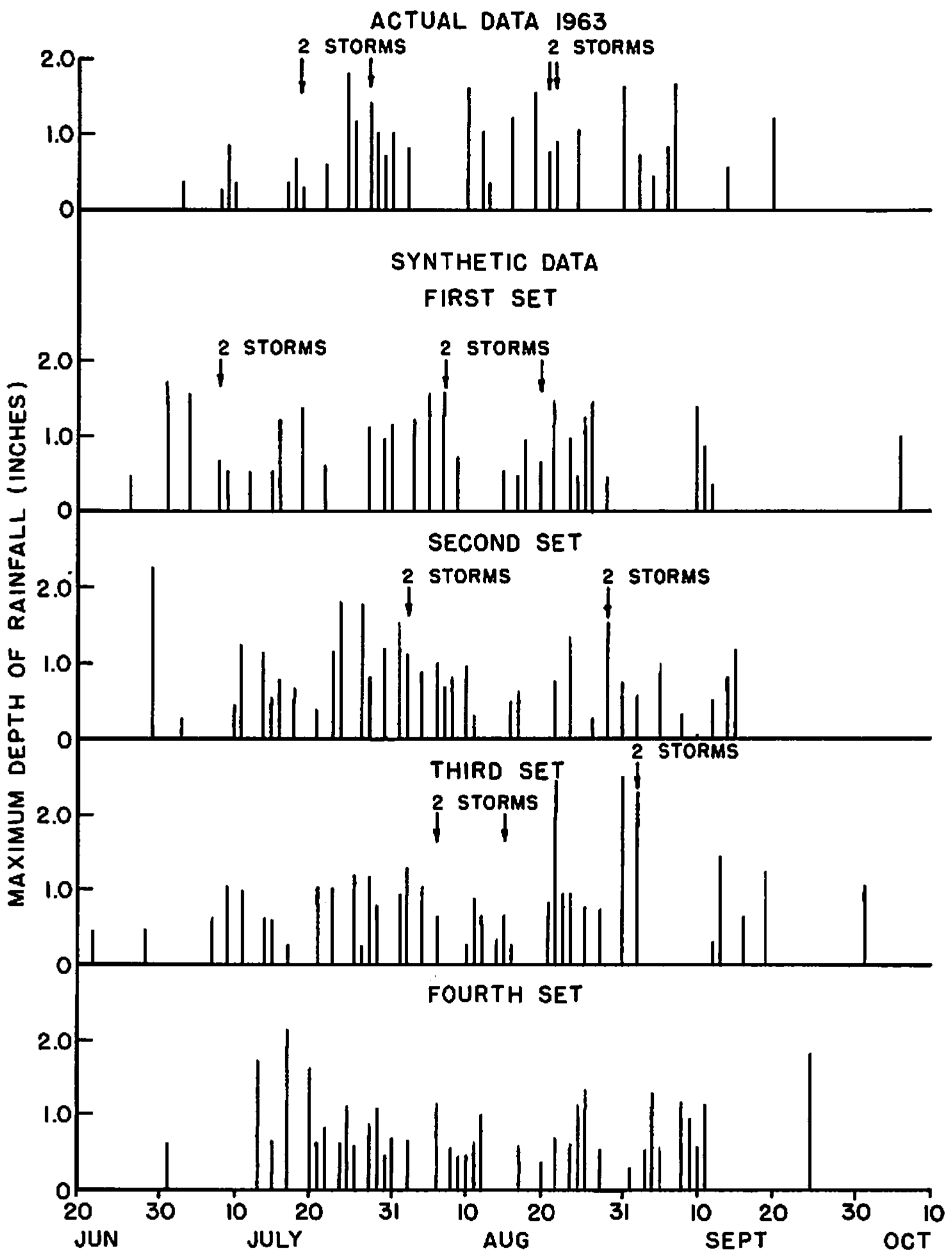
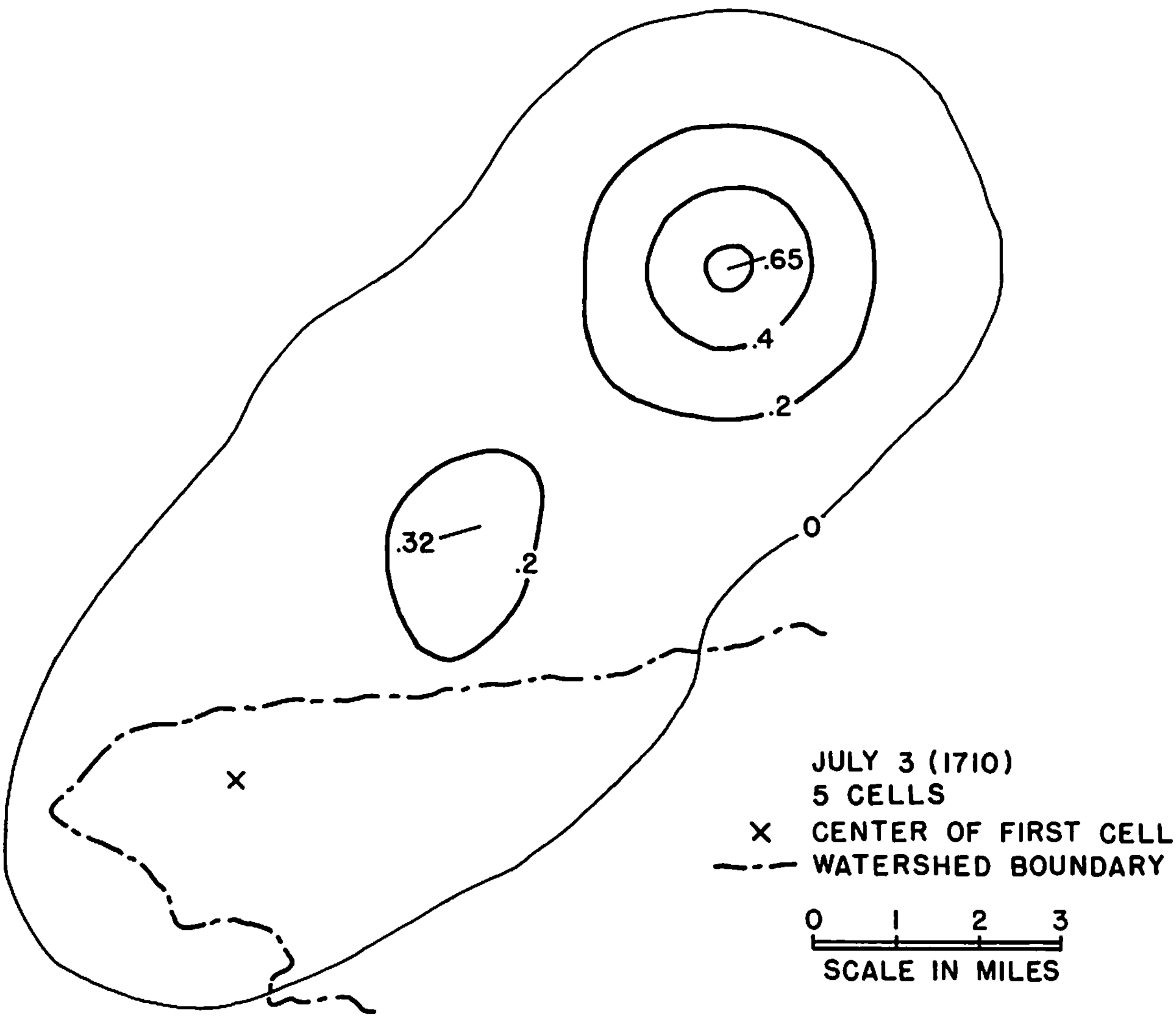


FIGURE 6.—Seasonal distribution of significant airmass thunderstorm events; 1 year of actual data, 4 years of synthetic data.

that such a model may generate simplified events which have some uses, such as runoff prediction, but it is a rather crude approximation and does not represent all of the observed variability in thunderstorm rainfall.

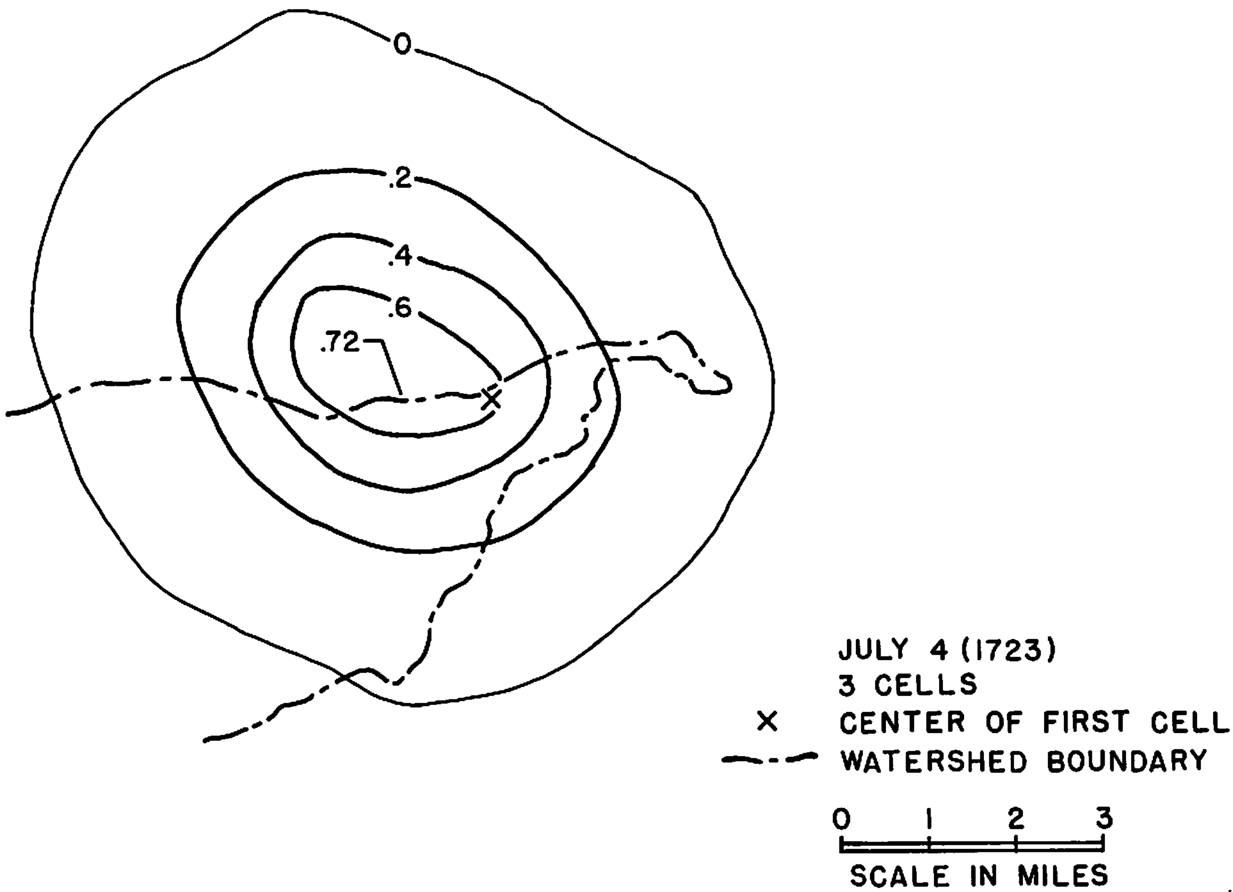
The generated data correspond roughly to individual thunderstorms occurring over a finite-sized area. The storms could be superimposed over a smaller area to completely cover it, or on to larger areas (greater than 100 square miles) in groups to simulate the areal distribution of several multicellular storms over an entire region. Potential uses of the synthetic data would determine their application. However, additional research is needed to determine if better, simpler, or more complex models can be developed from available information.

LeCam (*8*), in referring to his complex precipitation model, stated, "The main difficulty in such circumstances is that, in a model of this complexity, it becomes more and more difficult to estimate or test anything through purely statistical



A

FIGURE 7.—Isohyetal maps of selected synthetic thunderstorm events: A, July 3 (1710) 5 cells center of first cell watershed boundary. B, July 4 (1723) 3 cells center of first cell watershed boundary. C, July 9 (1733) 4 cells center of first cell watershed boundary. D, Aug. 15 (1300) 5 cells center of first cell watershed boundary. E, Aug. 16 (1957) 5 cells center of first cell watershed boundary. F, Aug. 29 (1641) 5 cells center of first cell watershed boundary. G, Sept. 11 (1641) 5 cells center of first cell watershed boundary. H, Sept 13 (1835) 5 cells center of first cell watershed boundary.



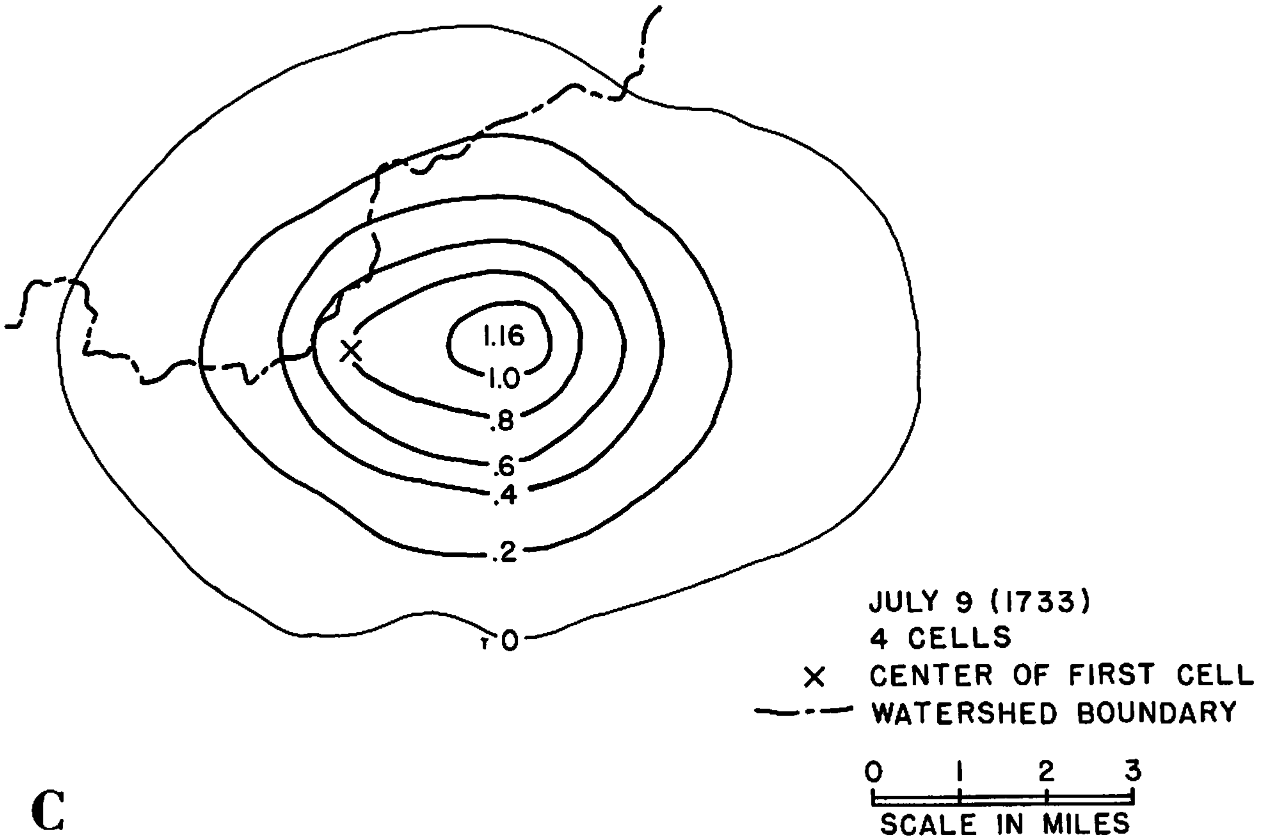
B

methods." He continued that, "it is then necessary to specify some of the elements through purely physical arguments." These same observations became apparent to the authors during the development of the Walnut Gulch stochastic thunderstorm rainfall model.

The question of accuracy must also be answered before such models can be put to practical use. Although the model is based on measurement of thunderstorm rainfall from a dense network of recording gages, there are still considerable areas for error. The statistical distributions have been chosen largely for their simplicity since more involved distributions may not be justified from known information. To date, the overall test of the model has been somewhat subjective; that is, within aforementioned limits it looks good visually. More objective methods, if possible, would be valuable. These could include investigations of persistence of wet and dry periods throughout the region, and possibly direct comparison of volumes of rainfall above specified depths between the real and synthetic data. Furthermore, work is in progress to facilitate using point frequencies from long-term precipitation data to estimate storms on a finite-sized area. Thus, frequencies of storm occurrence could be predicted from point frequencies which are more widely known.

Preliminary evaluations reported here suggest that the model may generate usable synthetic airmass thunderstorm rainfall data, depending on what is wanted from the model. However, extensive evaluation procedures, such as those developed by Lane and Renard (7), need to be implemented. Such procedures, allowing for large sample tests of the model, would allow for a more comprehensive evaluation and are being investigated.

For general use, the variables in this model, revised models, or other similar thunderstorm rainfall models, must be tied to meteorological and topographical differences locally and between regions. For example, in the Southwest, east of the



C

Continental Divide, more intense, longer lasting thunderstorms have been recorded than in southeastern Arizona. These storms have added components for frontal activity and added moisture aloft. Variables representing frontal "strength" and available moisture could be added to the model for airmass thunderstorms. The chance of a front moving across a specific watershed when airmass thunderstorms are expected to develop can be assigned a seasonal probability, just as pure airmass thunderstorms are assigned probabilities within a season. Available moisture would increase or decrease the magnitude of the event. There is a certain chance that pure airmass thunderstorms will occur, with the magnitude conditional on available moisture, along with a chance that a front also will add to the magnitude of rainfall for specific events.

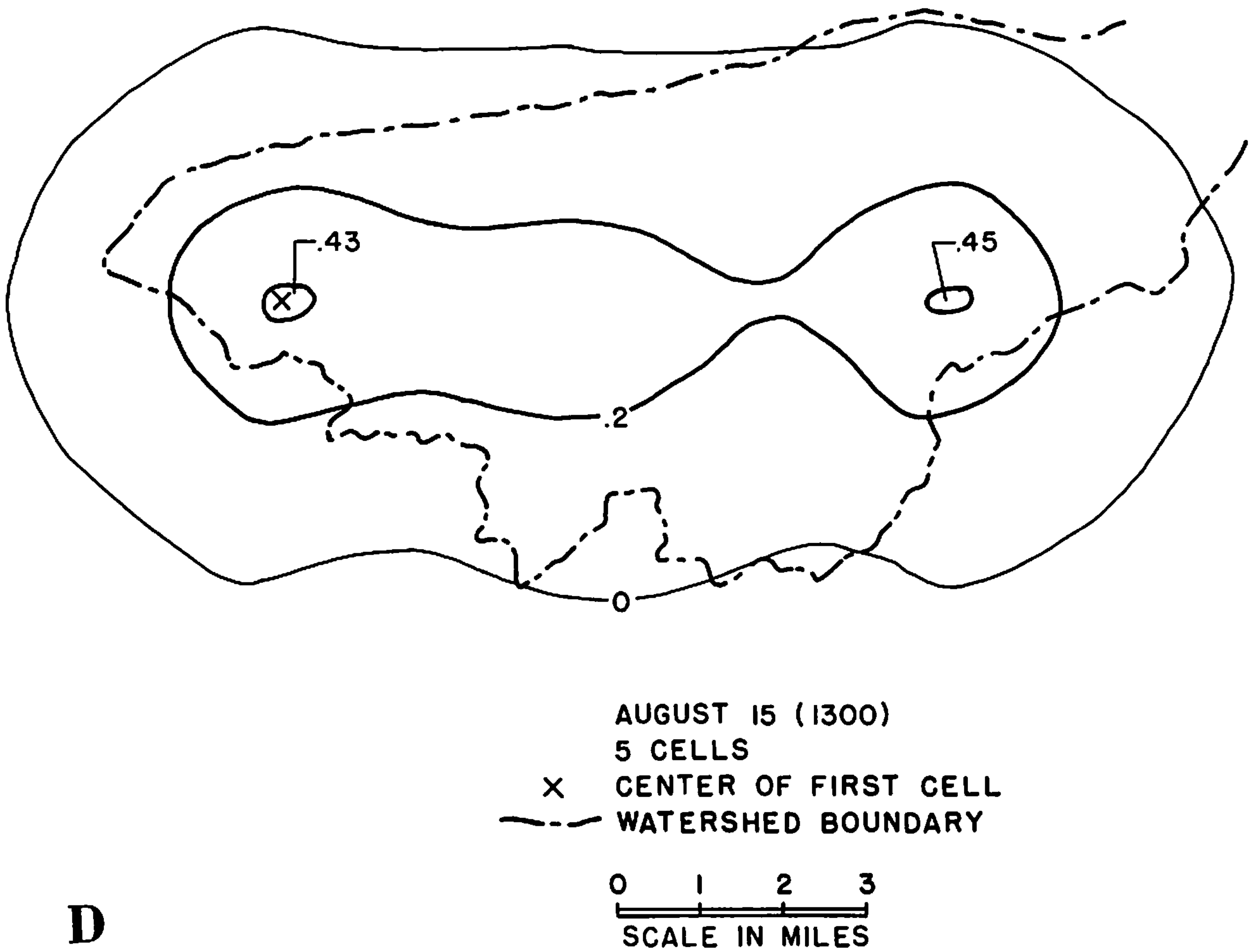
References

(1) CHOW, V. T.
1964. HANDBOOK OF APPLIED HYDROLOGY. Chapter 8, Statistical and Probability Analysis of Hydrologic Data. Pp. 8-1 to 8-97. McGraw-Hill Book Co., Inc. New York.

(2) DISKIN, M. H., and LANE, L. J.
1970. A STOCHASTIC MODEL FOR RUNOFF EVENTS FOR A SEMIARID WATERSHED IN SOUTHEASTERN ARIZONA. Joint ARS-SCS Watershed Modeling Workshop Proc., USDA, pp. 34-1 to 34-23.

(3) DRISSEL, J. C., and OSBORN, H. B.
1968. VARIABILITY IN RAINFALL AFFECTING RUNOFF FROM A SEMARID RANGELAND WATERSHED, ALAMOGORDO CREEK, NEW MEXICO. Jour. Hydrol. 6:194-201.

(4) FELLER, W.
1966. AN INTRODUCTION TO PROBABILITY THEORY AND ITS APPLICATION. V. II, 669 pp. John Wiley & Sons, Inc. New York.



D

(5) FOGEL, M. M.
1968. THE EFFECT OF THE SPATIAL AND TEMPORAL VARIATIONS OF RAINFALL ON RUNOFF FROM SMALL SEMI-ARID WATERSHEDS. 90 pp. Ph.D. Dissertation, Dept. of Watershed Mangt., Univ. Ariz. Tucson.

(6) ——— and DUCKSTEIN, L.
1969. TRANSFORMATIONS OF CONVECTIVE STORM PATTERNS INTO POINT RAINFALL FREQUENCIES. Water Resources Res. 5(6):1229–1237.

(7) LANE, L. J., and RENARD, K. G.
1971. EVALUATION OF A BASINWIDE STOCHASTIC MODEL FOR EPHEMERAL RUNOFF. Presented at the Spring Meeting of the Pacific Region ASAE, Las Vegas, Nev., and pending pub. Trans. ASAE.

(8) LECAM, L.
1961. A STOCHASTIC DESCRIPTION OF PRECIPITATION. Proceedings of the 4th Berkeley Symposium in Mathematical Statistics and Probability. 3: 165–185. Univ. Calif. Berkeley.

(9) MCDONALD, J. E.
1956. VARIABILITY OF PRECIPITATION IN AN ARID REGION. Technical Reports on the Meteorology and Climatology of Arid Regions, No. 1. Univ. Ariz. Inst. Atmos. Physics. Tucson.

(10) ———
1960. VARIABILITY FACTORS IN MOUNTAIN-WATERSHED HYDROMETEOROLOGY IN ARID REGION. Jour. Ariz. Acad. Sci. 1(3):89–98.

(11) NEYMAN, J., and OSBORN, H. B.
1971. EVIDENCE OF WIDESPREAD EFFECTS OF CLOUD SEEDING AT TWO ARIZONA EXPERIMENTS. Natl. Acad. Sci. Proc. 68(3):649–652.

(12) ——— and SCOTT, E. L.
1957. ON A MATHEMATICAL THEORY OF POPULATIONS CONCERNED AS CONGLOMERATIONS OF CLUSTERS. Cold Spring Harbor Symposium on Quantitative Biology, 22:109–120

(13) ——— and SCOTT, E. L.
1959. STOCHASTIC MODELS OF POPULATION DYNAMICS. Science, V. 130, pp. 303–308

(14) OSBORN, H. B.
1968. PERSISTENCE OF SUMMER RAINY AND DROUGHT PERIODS ON A SEMIARID RANGELAND WATERSHED. Bul. IASH 13(1):14–19.

(15) ———
1971. SOME REGIONAL DIFFERENCES IN RUNOFF-PRODUCING THUNDERSTORM RAINFALL IN THE SOUTHWEST. Proc. 1971 Joint Meeting Arizona Section American Water Resources Assoc. and Ariz. Acad. Sci., V. 1: 13–26.

(16) ——— and HICKOK, R. B.
1968. VARIABILITY OF RAINFALL AFFECTING RUNOFF FROM A SEMIARID RANGELAND WATERSHED. Water Resources Res., American Geophysical Union 4(1):199–203.

(17) ——— and LANE, L. J.
1971. DEPTH-AREA RELATIONSHIP FOR THUNDERSTORM RAINFALL IN SOUTHEASTERN ARIZONA. Presented ASAE, Las Vegas, Nev., April 1971, and accepted for publication Trans. ASAE.

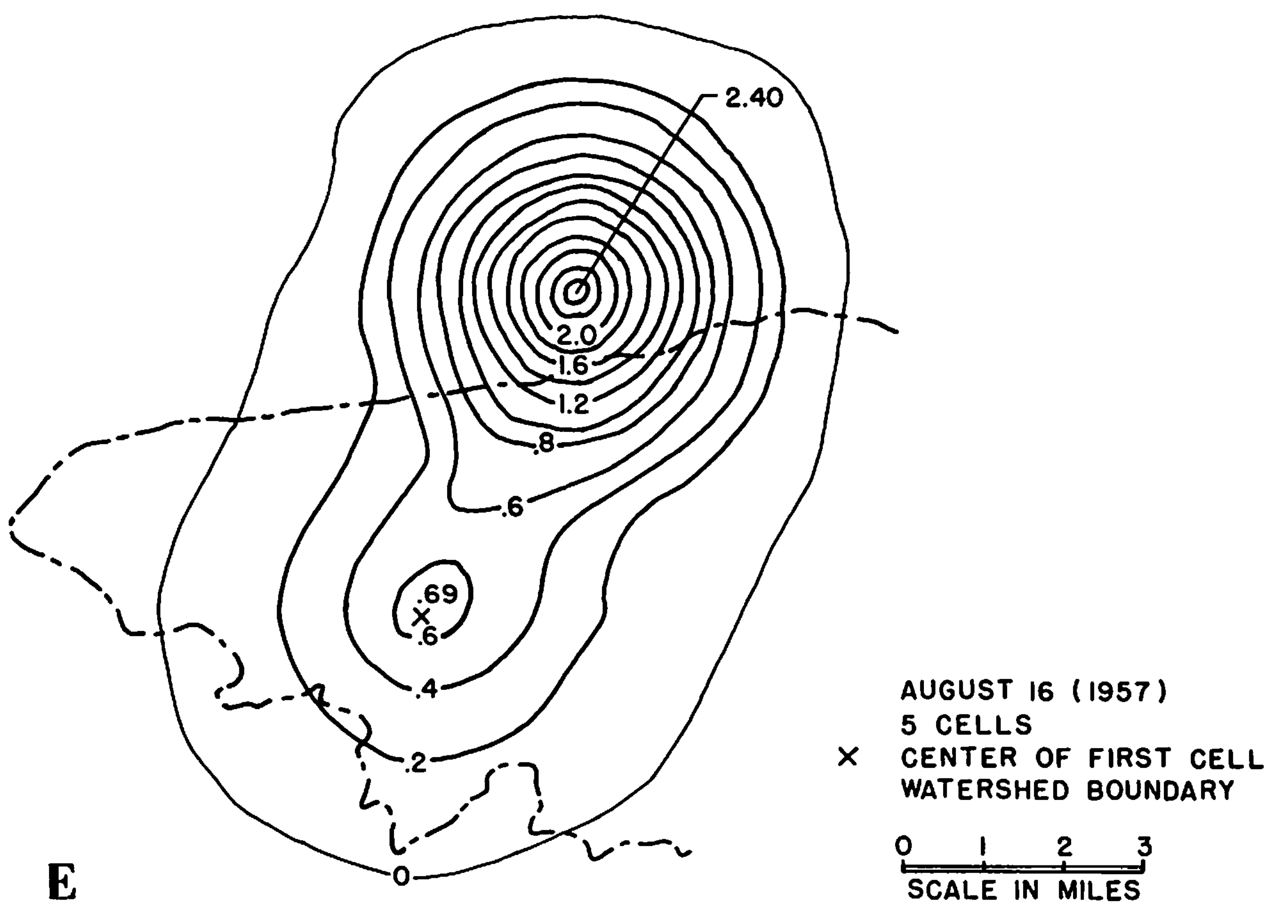
(18) ——— and REYNOLDS, W. N.
1963. CONVECTIVE STORM PATTERNS IN THE SOUTHWESTERN UNITED STATES. Bul. IASH 8(3):71–83.

(19) PETTERSSEN, S.
1956. WEATHER ANALYSIS AND FORECASTING. V. 2, pp. 156–165. McGraw-Hill Book Co., Inc. New York.

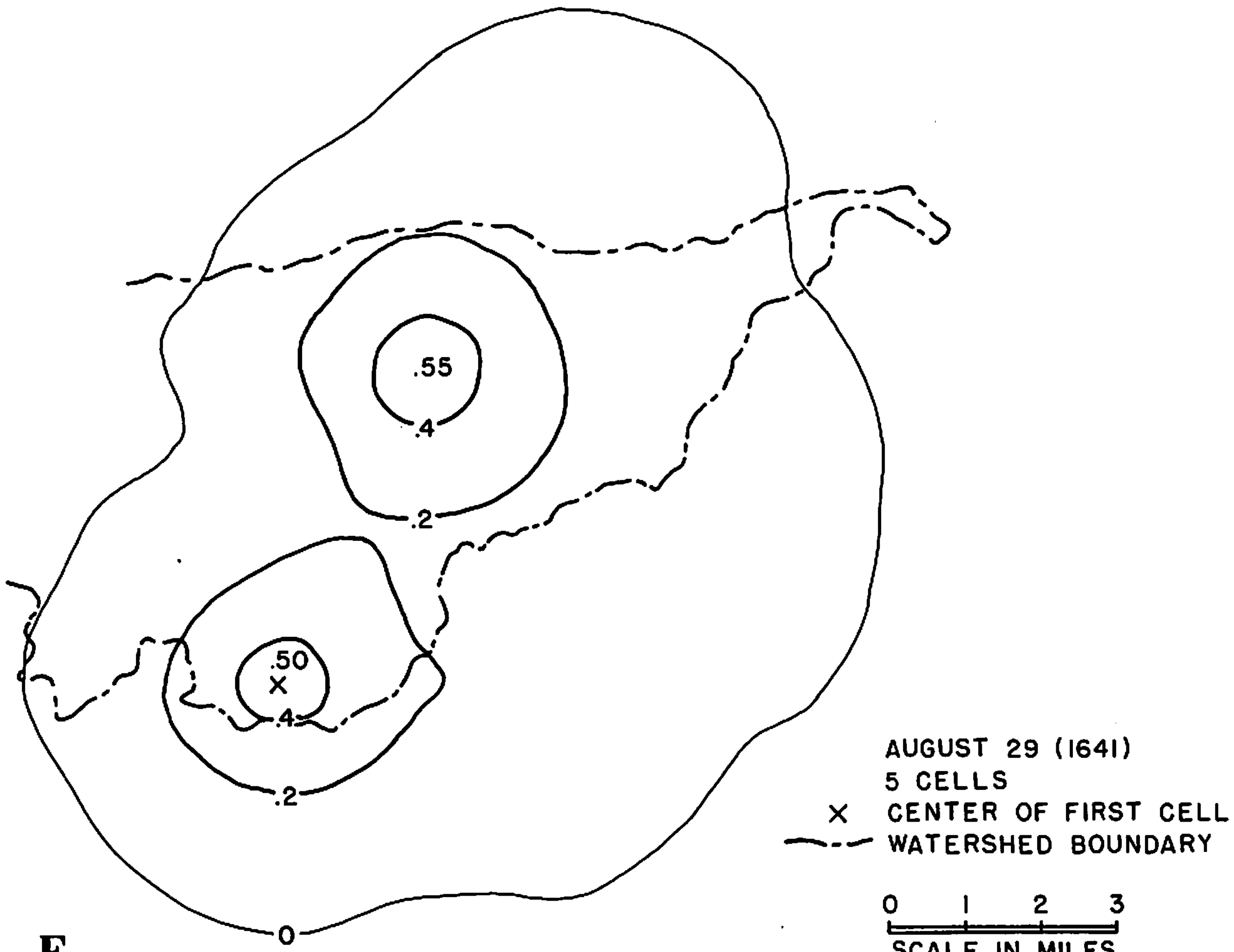
(20) ROSENBLUETH, A., and WEINER, N.
1945. THE ROLE OF MODELS IN SCIENCE. Philosophy of Science, XII(4):316–321. *Quoted* by Woolhiser, D. A., 1970, Deterministic Approach to Watershed Modeling. Proc. Joint ARS-SCS Workshop, USDA, pp. 7–1 to 7–29.

(21) SELLERS, W. D.
1960. THE CLIMATE OF ARIZONA. *In* Greene, C. R., and Sellers, W. D. Arizona Climate. 265 pp. University of Arizona Press. Tucson.

(22) WOOLHISER, D. A., and SCHWALEN, H. A.
1959. AREA-DEPTH FREQUENCY RELATIONS FOR THUNDERSTORM RAINFALL IN SOUTHERN ARIZONA. Univ. Ariz. Agr. Exp. Sta. Tech. Paper 527.



E



F

Appendix

List of Variables and Parameters Used To Describe Multicellular Airmass Thunderstorms

K Constant used in rainfall relationship and dependent only on cell radius.

d Distance between cell centers generated using a "triangular" distribution and used in grouping cells.

D Rainfall in inches with each cell; a function of center depth D_0 and distance from the center r.

D_0 Cell center depth generated using a negative exponential distribution with a mean of $\bar{D}_0$.

$\bar{D}_0$ Mean cell center depth estimated from Walnut Gulch rainfall data.

I Rain gage number where the first cell is located; generated from a uniform distribution where I is an integer.

N Number of cells per storm generated from a Poisson distribution using $\bar{N}$ as the mean.

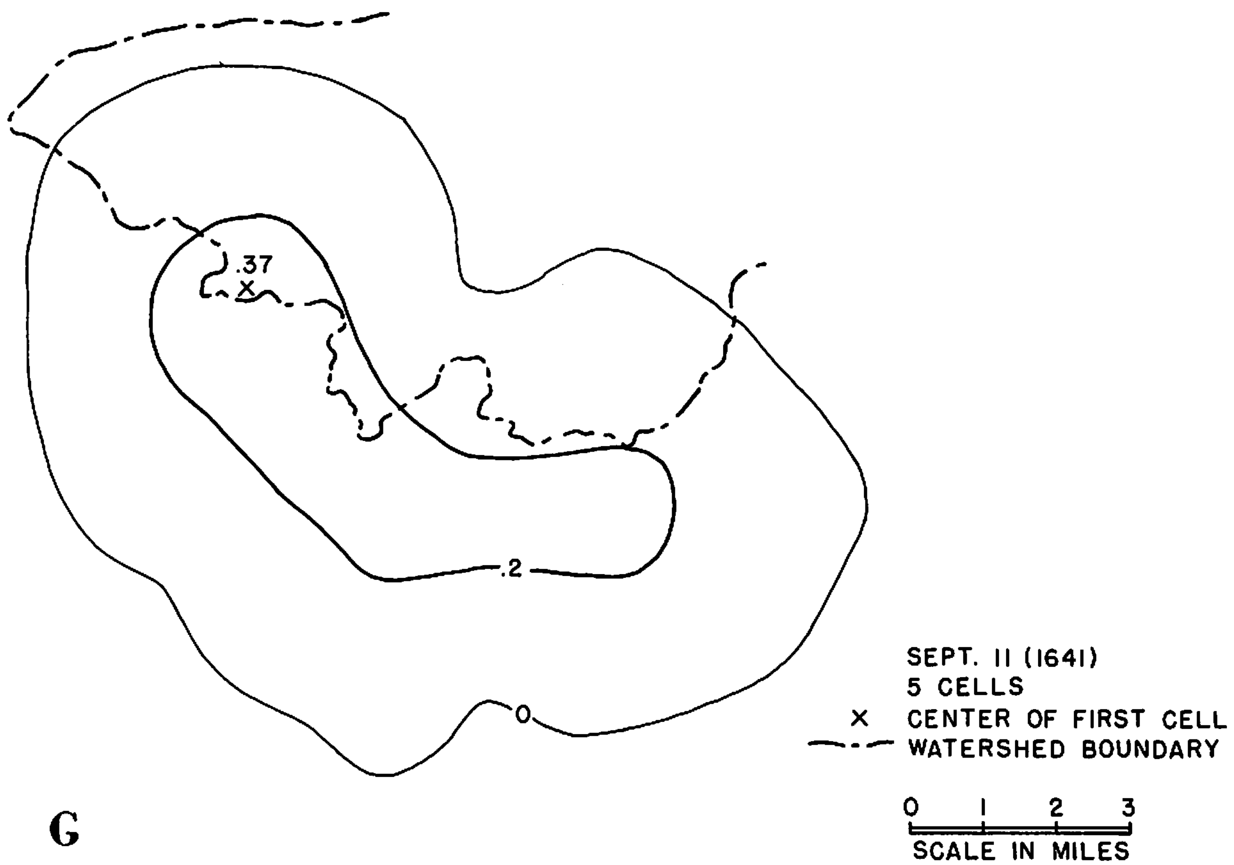
$\bar{N}$ Mean number of cells per storm estimated from Walnut Gulch rainfall data.

r Distance from the cell center.

R Radius of unit cell estimated from isohyetal maps of Walnut Gulch data.

θ_0 Direction of the second cell as measured in degrees from the first cell where east is defined to be 0°; generated from a uniform distribution.

θ_i Direction of cell number $i+1$ as measured in degrees from the ith cell where east is defined to be 0° and calculated from the equation $\theta_i = \theta_{i-1} + \Delta\theta_{i-1}$.



G

$\Delta\theta_i$ Change in the direction of cell placement generated from a normal distribution with a mean of 0° and a standard deviation of 60°.

U Uniform random variable approximated by pseudo-random numbers from a random number generator ($0 < U \leq 1$).

A Note on the Variability of the Number of Storms in a Season Where the Occurrence of Storms is Modeled as a Bernoulli Variable

In the absence of persistence in daily rainfall, the occurrence of a storm can be modeled as a Bernoulli variable where:

$$X_k = \begin{cases} 1 \text{ if there is a storm on day } k \\ 0 \text{ otherwise} \end{cases} \tag{1}$$

and

$$P[X_k = 1] = P_k. \tag{2}$$

With the above definitions,

$$S_n = \sum_{k=1}^{n} X_k \tag{3}$$

will be the number of rainy days in a period of length n days.

Of interest in this discussion is the expected value of S_n and the variability of S_n. Mathematical expectation leads to

$$E(S_n) = n\bar{p}, \tag{4}$$

and

$$\text{Var } (S_n) = n\bar{p} - \sum_{k=1}^{n} p_k^2,$$

where Var denotes variance, and $\bar{p}$ is the "average" probability of rain, such that

$$\bar{p} = (1/n) \sum_{k=1}^{n} p_k. \tag{6}$$

For example, see Feller (*4*).

The assertion is that the variance of S_n is maximum in the absence of strong seasonality. That is, for the same $\bar{p}$, the variance of S_n is maximum when the p_k values do not vary in the season. The proof of this assertion is complete if we can show that in equation 5, the second term is minimum when $p_k = \bar{p}$ for all k.

Let T_{n_1} be the sum when all $p_k = \bar{p}$, then

$$T_{n_1} = \sum_{k=1}^{n} p_k^2 = n\bar{p}^2. \tag{7}$$

Let T_{n_2} be the sum when all but two of the $p_k = \bar{p}$, and the other items are $p_j = \bar{p} + \epsilon$ and $p_{j+1} = \bar{p} - \epsilon$ for some j, then:

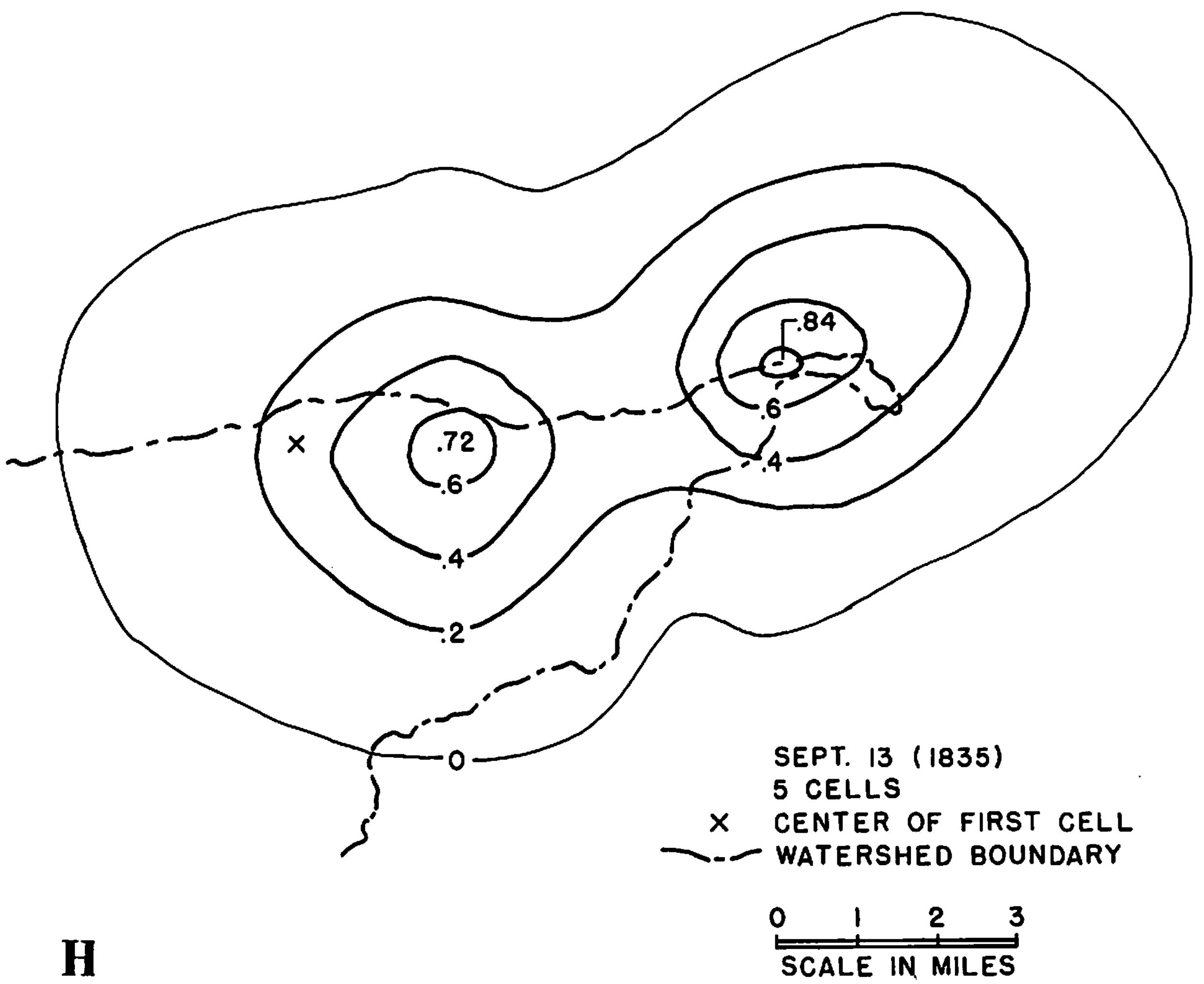
$$T_{n_2} = (\bar{p} + \epsilon)^2 + (\bar{p} - \epsilon)^2 + \sum_{k=1}^{n-2} p_k^2 \tag{8}$$

and thus,

$$T_{n_2} = 2\bar{p}^2 + 2\epsilon^2 + (n-2)\bar{p}^2 \tag{8a}$$

$$= n\bar{p}^2 + 2\epsilon^2. \tag{8b}$$

Clearly, $T_{n_2} > T_{n_1}$ for all $\epsilon > 0$, and the proof for more than two of the $p_k = \bar{p}$ follows by induction.



H

STOCHASTIC MODEL OF DAILY RAINFALL [1]

By P. Todorovic and D. Woolhiser [2]

Abstract

An application of stochastic processes for description and analysis of daily values of precipitation is presented. The total amount of precipitation $S(n)$ during an n-day period is a discrete parameter stochastic process such that $0 \leq S(n) \leq S(n+1)$. The most general form of the distribution function, mathematical expectation, and variance of $S(n)$ are determined. The following special cases for the sequence of daily rainfall occurrences are considered: (1) Sequence of independent identically distributed random variables, (2) sequence of independent random variables, and (3) Markov chain. In addition, assuming that certain regularity conditions hold, it has been proved that $S(n)$ is asymptotically normal. The first passage time of $S(n)$ and the corresponding distribution function are also considered. A numerical example for cases (1) and (3) is presented assuming that the daily rainfall amounts are exponentially distributed.

1. Introduction

In this paper, an attempt is made to develop a stochastic model for description and analysis of certain aspects of the rainfall phenomenon utilizing daily precipitation records. The primary reason for constructing such a model is the fact that daily rainfall data are the most readily available and are sufficient for many hydrological problems.

In this report, we are concerned with merely a probabilistic treatment of the observed record. Various climatological and other factors, such as temperature, winds, and origin of airmasses, are not taken into account. Therefore, the model can not provide physical explanations of features of rainfall phenomena.

Consider a certain period of time which, for example, consists of n days. To each day of the n-day period is associated a random variable η_j which assumes only two values, 0 and 1, defined as follows:

$$\eta_j = \begin{cases} 1 \text{ if } j\text{th day is wet} \\ 0 \text{ if } j\text{th day is dry} \end{cases}$$

where $j=1, 2, \ldots, n$. According to this definition, the number of rainy days N_n in this period is obviously equal to the following sum:

$$N_n = \sum_{j=1}^{n} \eta_j \tag{1.1}$$

$$(N_n = 0, 1, \ldots, n).$$

Let ξ_ν, $\nu = 1, 2, \ldots, n$ denote the daily value of precipitation of νth rainy day, then the total amount of precipitation $S(n)$ of this n-day period is given by

$$S(n) = \sum_{\nu=0}^{N_n} \xi_\nu \qquad n = 1, 2, \ldots \tag{1.2}$$

where by definition $S(n) = 0$ if $N_n = 0$. Since $\xi_\nu > 0$ for all $\nu = 1, 2, \ldots, n$ it follows that

$$0 = S(0) \leq S(1) \leq S(2) \leq . \tag{1.3}$$

Provided $\{\xi_\nu\}_1^\infty$ is a sequence of random variables for which the central limit theorem holds, then if certain regularity conditions are satisfied $S(n)$ is asymptotically normal.

Finally, in connection with random variables N_n and $S(n)$, we will consider the following two

---

[1] Contribution from the Colorado State University Experiment Station and the Agricultural Research Service, USDA.

[2] Associate professor civil engineering, Colorado State University, and research hydraulic engineer, USDA, Fort Collins, Colo.