#100

MHD-1

A STOCHASTIC MODEL FOR RUNOFF EVENTS FOR A SEMIARID WATERSHED IN SOUTHEASTERN ARIZONA [1/]

M. H. Diskin and L. J. Lane [2/]

INTRODUCTION

Stochastic hydrologic models are models used for the generation of synthetic hydrologic data in such a way that the statistical properties of the original data observed in the watershed are preserved in the synthetic data generated by the model. The parameters describing the various distributions associated with the original data serve as input to the stochastic models. The output of the models is in the form of sequences of hydrologic data that may be used for design purposes. Goodness of fit of the model is judged in terms of the agreement between the values of statistical parameters derived from the synthetic data and the corresponding parameters of the input data, including some parameters that were not used as input.

Systems represented by stochastic models are usually assumed to be stationary. If this is not the case, any nonstationarity, such as a trend or a cyclical variation, is removed from the original data. It is again superimposed on the output of the model. Stochastic models are usually employed for generation of either rainfall data or runoff data. There are only very few cases where stochastic models were used to represent the rainfall-runoff relationship.

1/ Contribution of the Southwest Watershed Research Center, USDA, Agricultural Research Service, Soil and Water Conservation Research Div., in cooperation with the Arizona Agricultural Experiment Station, Tucson.

2/ Research Hydraulic Engineer (visiting scientist on leave from Technion-Israel Institute of Technology, Haifa, Israel), and Hydrologist, respectively, Southwest Watershed Research Center, Agricultural Research Service, USDA, SWC, 442 E. Seventh Street, Tucson, Arizona 85705.

The purpose of the present paper is to describe a stochastic model that has been developed for runoff events on a subwatershed of the Walnut Gulch watershed in southeastern Arizona. The subwatershed considered is that contributing to Flume No. 6 (W-6). It is a brush- and grass-covered watershed with a well developed stream network. The area of the watershed is 36.7 square miles (Fig. 1). Flow in the tributary streams and in the main channel out of the watershed is intermittent with relatively long periods of no flow and short periods of runoff following local thunderstorm rainfall.

Discharge from the watershed is measured with a large supercritical-depth flume (Gwinn, 1964), and the measurements are considered to be of good quality. Runoff records for the watershed are available for the period 1962 through 1968. The great majority of runoff events occur in the summer season from the beginning of July to the end of September with occasional events in other times of the year. The model developed herein is based on data pertaining to the summer events only, and its purpose is to generate sequences of such summer runoff events that may be used in subsequent studies of the hydrologic behavior of the watershed considered.

STATISTICAL PROPERTIES OF RUNOFF DATA

The characteristics of runoff data from the watershed considered are shown in Figure 2, giving the time distribution of runoff volumes for the 7 years of available data. Examination of the runoff records of the watershed indicated that the runoff season can be described in terms of the starting date of the season, the number of runoff events that occur per season, and the mean time interval between runoff events. The seven years of records for Watershed 6 indicated that there is some negative correlation between the starting date and the number of events per season (Fig. 3). No correlation was found between the mean time interval between runoff events and either of the other two variables mentioned above.

The starting date of the runoff season for the 7 years of record was between July 2 and July 28 with a mean value of July 15 and a standard deviation of 8.7 days. The range of number of events per

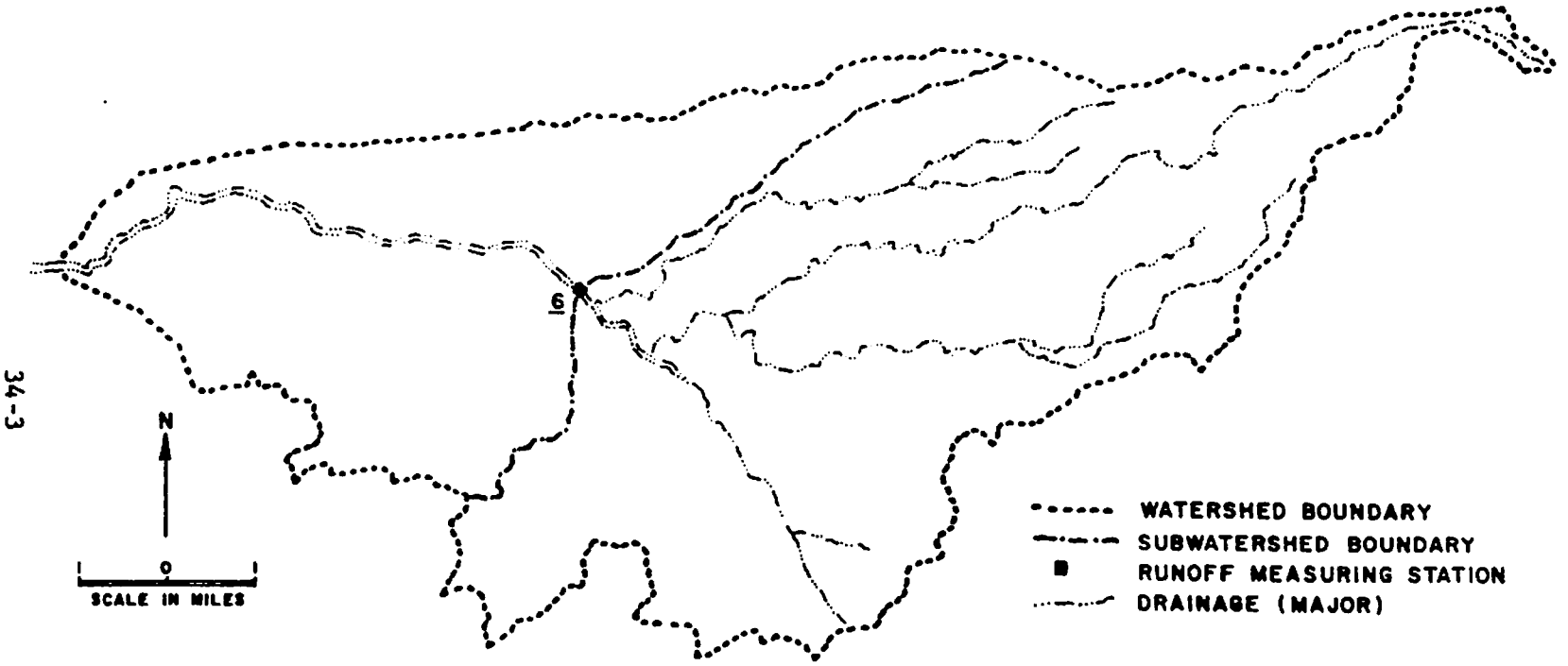

FIGURE 1. WALNUT GULCH EXPERIMENTAL WATERSHED

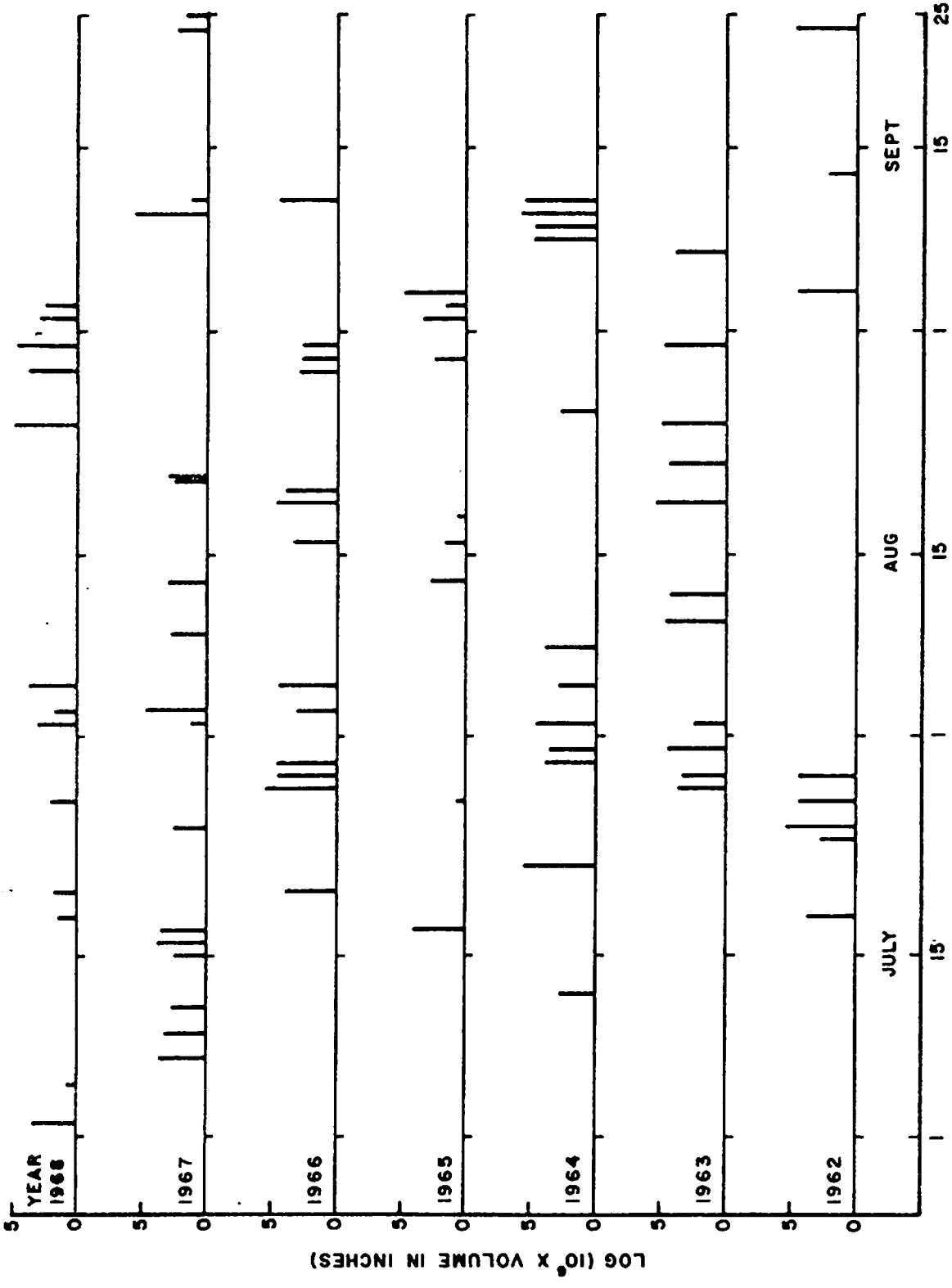

FIGURE 2. TIME DISTRIBUTION OF RUNOFF VOLUMES, OBSERVED DATA FOR WATERSHED 6.

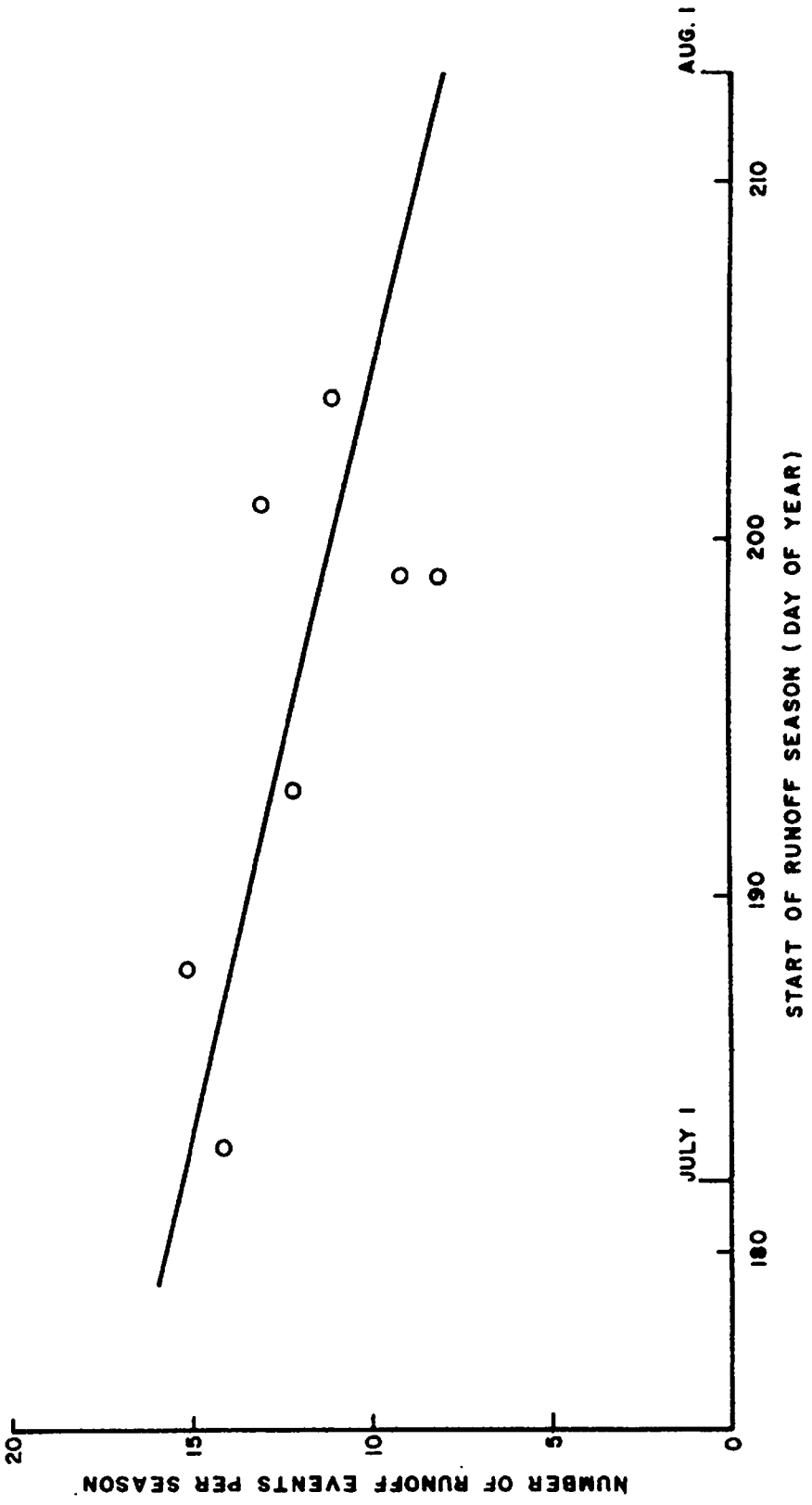

FIGURE 3. LINEAR RELATION BETWEEN NUMBER OF EVENTS PER SEASON AND START OF SEASON.

season was 8 to 17 with a mean value of 12.0 and a standard deviation of 3.1. The seven years of record were insufficient to establish accurately the shape of the distribution curves for the starting date and for the number of events per season. Observations from an ARS experimental watershed in Safford, Arizona, where 30 years of records were available, indicated that the assumption of normal (Gaussian) distributions for these two variables is acceptable. The agreement between the measured values of starting date and number of events per season and the assumed normal distributions is shown in Figures 4 and 5.

The distribution of the time interval between runoff events was based on 77 values of this interval obtained by combining the number of intervals for all 7 years of record. The mean time interval was 5.38 days, and the standard deviation was 6.10 days. The negative exponential distribution was found to give a fairly good approximation of the actual distribution of the intervals. The agreement between the observed data and the negative exponential distribution adopted to describe the data is shown in Figure 6. Data from the Safford watershed mentioned above also supported the choice of the negative exponential distribution for the interval between runoff events.

Runoff events on Watershed 6, as well as on other watersheds in the Walnut Gulch area, tend to occur, because of local meteorological conditions, in late afternoon. Analysis of time of start of the 84 runoff events observed in Watershed 6 indicated that it may be assumed to be normally distributed with a mean value of 18.0 (6:00 p.m.) and a standard deviation of 3.53 hours (Fig. 7).

The above variables describe adequately the time of occurrence of the runoff events. To complete the description of the runoff data, more elements were considered. These were the volume of runoff and the peak discharge during the runoff event. The two variables were found to be highly correlated ($r = 0.98$), and their distribution curves were similar. As expected, the distributions of the volumes and the peaks were skewed with small values occurring more frequently than the high values. The mean volume of runoff was 49.0 acre-feet, and the mean peak discharge was 562 cfs. The standard deviations of the two variables were 98.4 acre-feet and 1165 cfs, respectively.

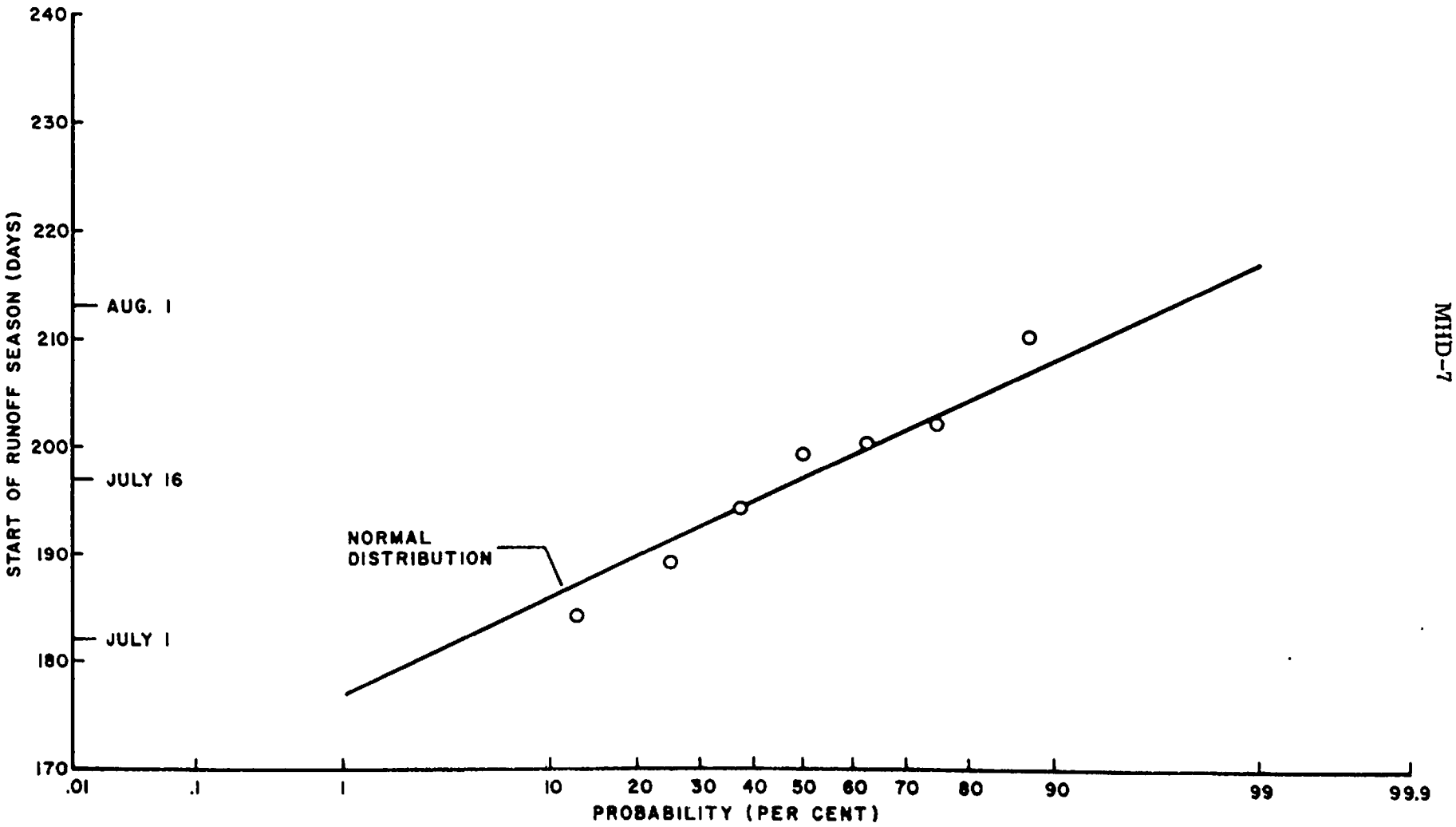

FIGURE 4. WATERSHED 6 DISTRIBUTION OF DATE OF FIRST RUNOFF EVENT OF SEASON.

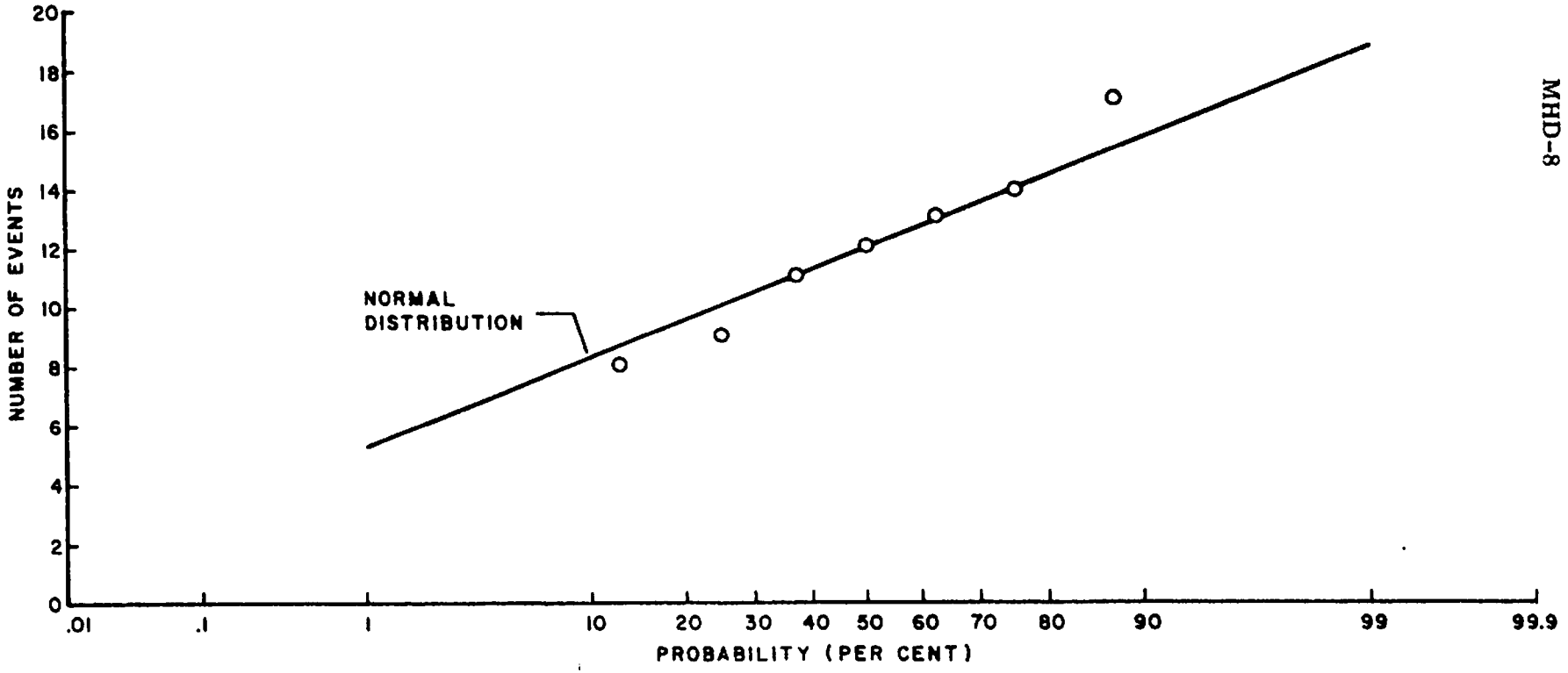

FIGURE 5. WATERSHED 6 DISTRIBUTION OF NUMBER OF EVENTS PER SEASON.

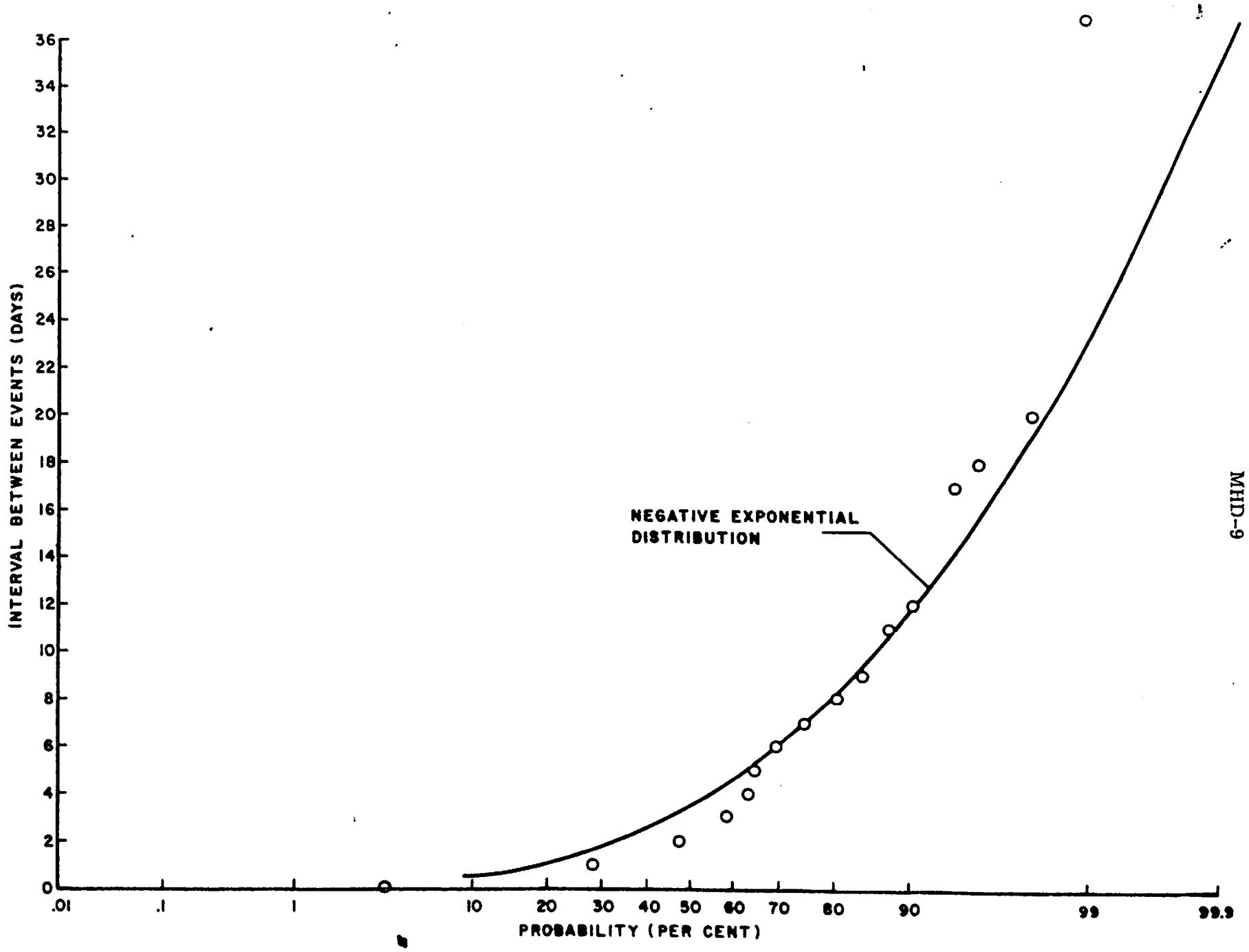

FIGURE 6. WATERSHED 6 DISTRIBUTION OF INTERVAL BETWEEN EVENTS.

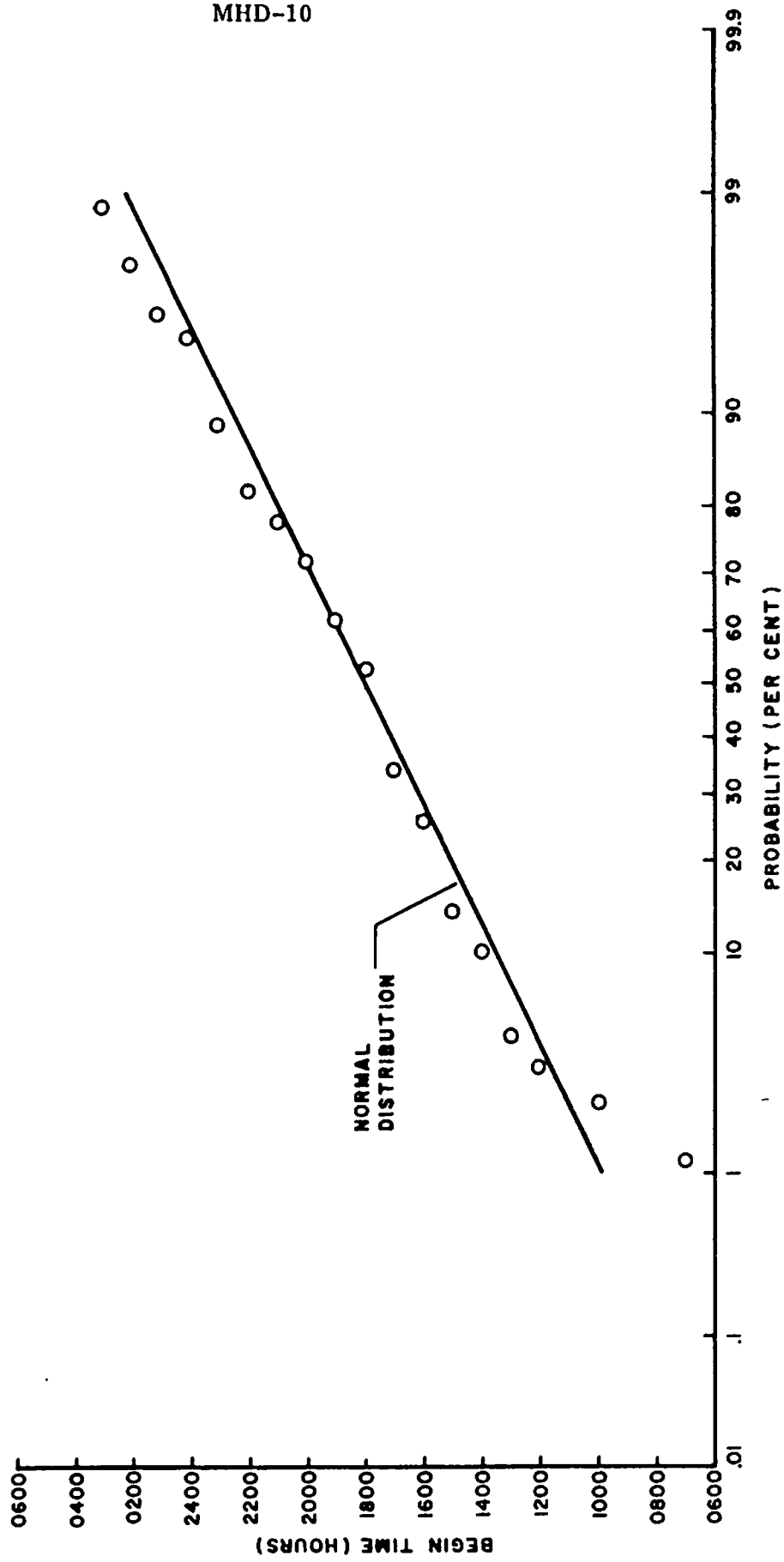

FIGURE 7. WATERSHED 6 DISTRIBUTION OF BEGIN TIME OF RUNOFF.

Adopting a logarithmic transformation for the volumes of runoff, it was found that the logarithms of the volume (expressed in 10^{-6}-inch units) were approximately normally distributed (Fig. 8), with a mean value of 3.30 and a standard deviation of 1.27. A better approximation could be obtained if a two-peaked distribution composed of two separate normal distributions is assumed.

The above description of runoff events in terms of the times of occurrence of the events, their volumes and peak discharges, is summarized in Table 1 which gives the mean and standard deviation for each variable.

DESCRIPTION OF THE STOCHASTIC MODEL

The stochastic model developed for generation of synthetic runoff data produces, as its output, sets of runoff events for the watershed considered. The runoff events are described in terms of the variables discussed above. The input to the model consists of the set of means and standard deviations, essentially as given in Table 1, as well as the parameters of the regression equations for the two variables that were found to be dependent on other variables.

A flow chart showing the structure of the model and the sequence of computations is given in Figure 9. A detailed description of the various steps involved is given below with reference to the operations enumerated in the flow chart.

The first step consists of reading in the values of the parameters used for generation of the synthetic data, reading the number of years for which data are required, and setting to "one" the sequential number of the year for which data are generated.

Using a subroutine for the generation of a reduced normal variate Z (with zero mean and unit standard deviation), the model first generates the starting date of the season. This is done by converting the reduced variate to the variable concerned

$$S = Z \cdot \sigma_S + \bar{S} \qquad (1)$$

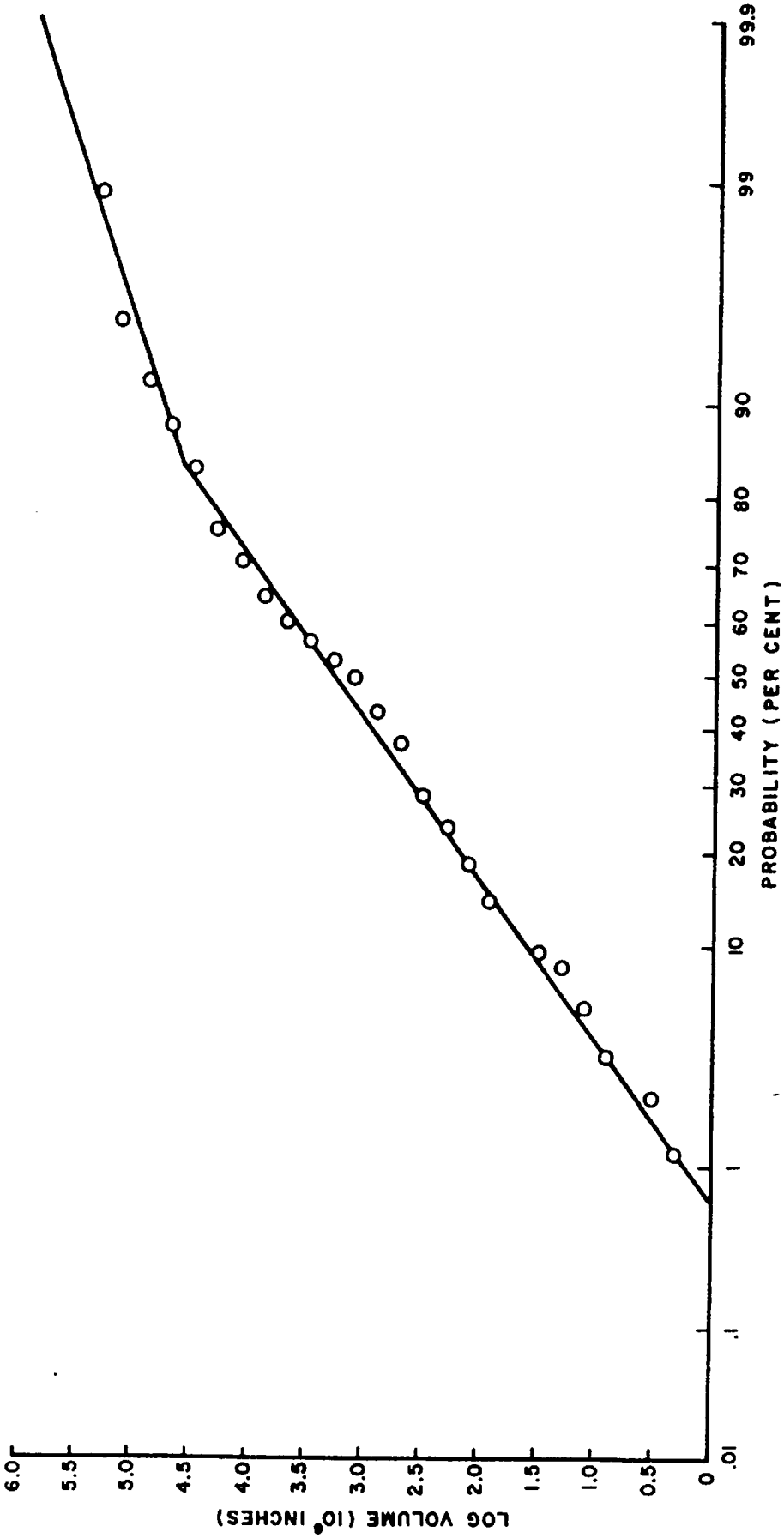

FIGURE 8. WATERSHED 6 DISTRIBUTION OF LOG (10^6 X VOLUME OF RUNOFF).

TABLE 1.--Mean and standard deviations for variables used in model

Variable	Property	Mean value	Standard deviation
S	Day of year of first runoff event (date)	196.0	8.70
N	Number of runoff events per season	12.0	3.06
T	Time of start of runoff event (hour)	18.0	3.53
V	Volume of runoff in acre-feet	49.0	98.4
L	Log_{10} of volume of runoff in 10^{-6} inches	3.30	1.27
P	Peak discharge in runoff event (cfs)	562	1165
D	Number of days between runoff events	5.38	6.10

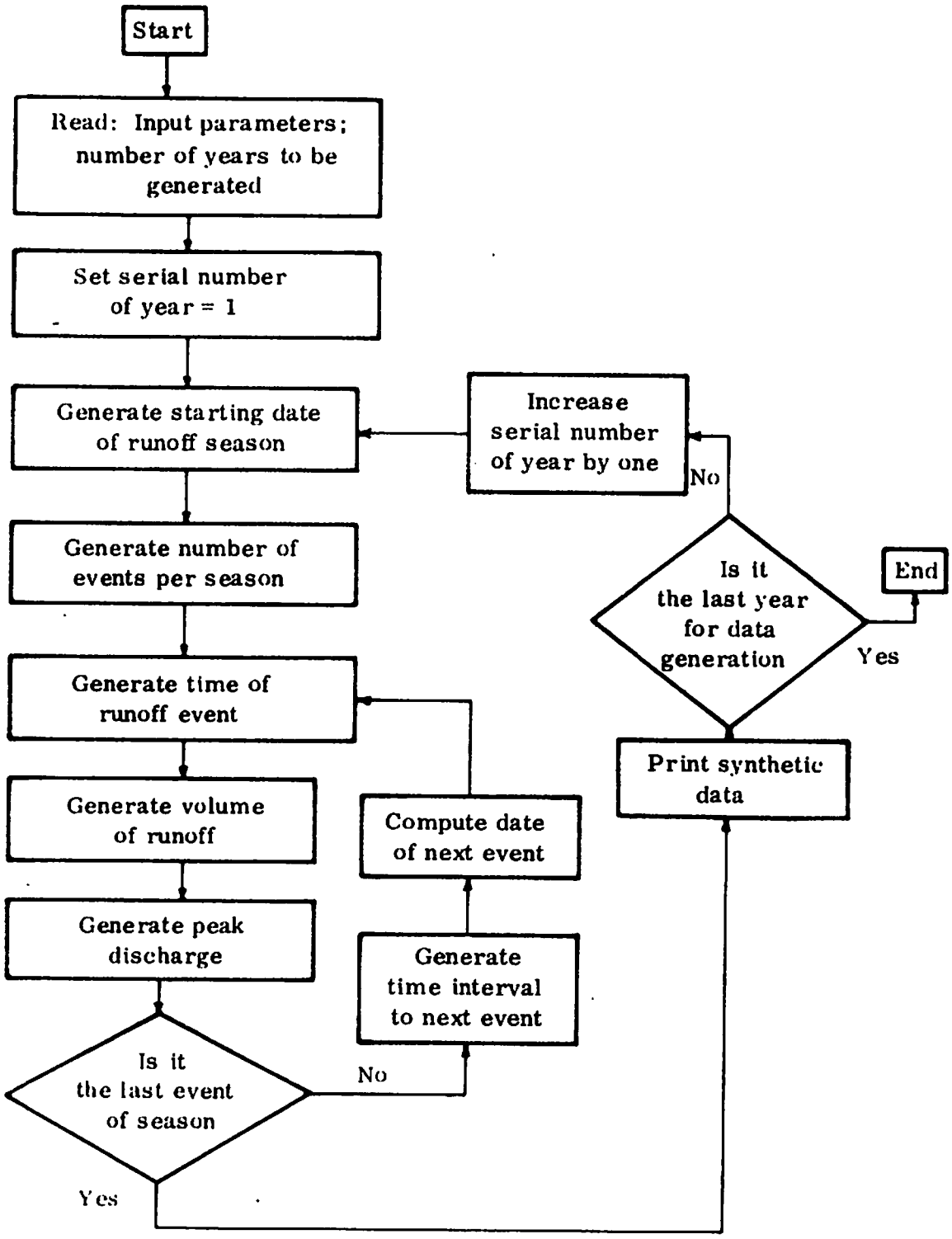

Fig. 9.-- Flowchart of stochastic model for runoff events.

where $\bar{S}$ and σ_S are the mean and standard deviation of the starting date being generated, as given in Table 1.

The number of events N is generated with the aid of the correlation found between the number of events and the starting date S. A random deviation, normally distributed with mean zero, was added to values computed by the regression line. The generating equation used was

$$N = 59.0 - 0.24\,S + \varepsilon_n \qquad (2)$$

where ε_n is the deviation computed from a value of the reduced normal variate Z by

$$\varepsilon_n = 2.46\,Z \qquad (3)$$

2.46 being the root mean square deviation from the regression line.

With the number of events N generated, the program now generates for each event the time of day of the event. The values are generated with the aid of the reduced normal variate subroutine

$$T = Z \cdot \sigma_t + \bar{T} \qquad (4)$$

where $\bar{T}$ and σ_t are the mean and standard deviation given in Table 1. The starting date and the number of events per season are rounded to the nearest integer, and a check is made to ensure that the number of events is at least one. The time of start of the runoff event is checked and if it happens to be larger than 24.0 hours, a value of 24 is subtracted from the computed time, and the date of the event is advanced by one. The time is then converted from a decimal value to a value expressed in hours and minutes.

The next item to be generated is the volume of runoff of each event. As mentioned above the distribution of the logarithms of volumes of runoff appears to be double-peaked, and the cumulative curve is therefore not a straight line on normal probability paper (Fig. 8). The method adopted for generation of this distribution was by approximating the curved cumulative distribution by two straight lines as shown in Figure 8. Each of these two lines is specified in terms of two parameters corresponding to a mean and a standard

deviation. A reduced normal variate Z was generated by the subroutine mentioned above and was converted to a value of the logarithm of the volume (expressed in 10^{-6}-inch units) by one of the following two equations:

$$L = 1.27Z + 3.30 \tag{5}$$

or

$$L = 0.60Z + 4.00 \tag{6}$$

depending on whether the probability of the reduced variate Z was less than or greater than 0.84.

Values of peak discharge for each runoff event were computed from the value of logarithm of volume of runoff generated in the previous step and the regression equation developed for the relationship between this variable and the peak discharge P,

$$P = 17000\ V^{0.90} \tag{7}$$

where P is in cfs and V is in inches. A random deviation between observed and computed data was also included so that the equation used in the program was

$$\log P = 4.23 + 0.90 \log V + \varepsilon_p \tag{8}$$

where ε_p is the random deviation term generated from the reduced normal variate Z by

$$\varepsilon_p = 0.21Z \tag{9}$$

The constant 0.21 used is the standard deviation of the logarithms of the peak about the regression line.

The final step in the computations for each year is to generate for each event, except the last of the season, the interval in days until the next runoff event. The interval is generated from the negative exponential distribution, using, for the parameter of the

distribution, the value of

$$\lambda = 1/5.38 \tag{10}$$

which is the reciprocal of the mean interval between events. The values obtained are rounded to the nearest integer. This rounding includes assigning a value of zero if the value generated was less than 0.5. If the interval between events is zero, there is a provision in the model to make sure that the time of start of the two runoff events falling on the same day are separated by at least 3.0 hours.

After completion of a series of computations for one year, the program prints the synthetic data for the current year, checks if it is the last year for which synthetic data are required, and if it is not, starts a new cycle of computations by computing the starting date of the new season of runoff.

RESULTS

A typical example of the data generated by the model is shown in Figure 10. The data presented are for a seven-year period which is the same number of years shown in Figure 2. It is, of course, not expected that a one-to-one correspondence exists between the data in the two figures, but the general appearance of the data is similar.

As a test of the adequacy of the model, it was used to generate a number of sets of data for 7-year periods, comparable to the original seven years of data available. The parameters of the synthetic data thus generated were computed and compared to the corresponding parameters of the original data. The results are shown in Table 2. As expected, the values obtained from any one such set of synthetic data are not exactly equal to the original values of the parameters, but the range of values obtained is in good agreement with the original data.

Included in the table are also two parameters that were not used in the generation of synthetic data. These are the mean length of the runoff season and the mean yearly volume of runoff. Values given are the means for the seven years of record, as well as the means for each of the seven years of synthetic data generated by the model.

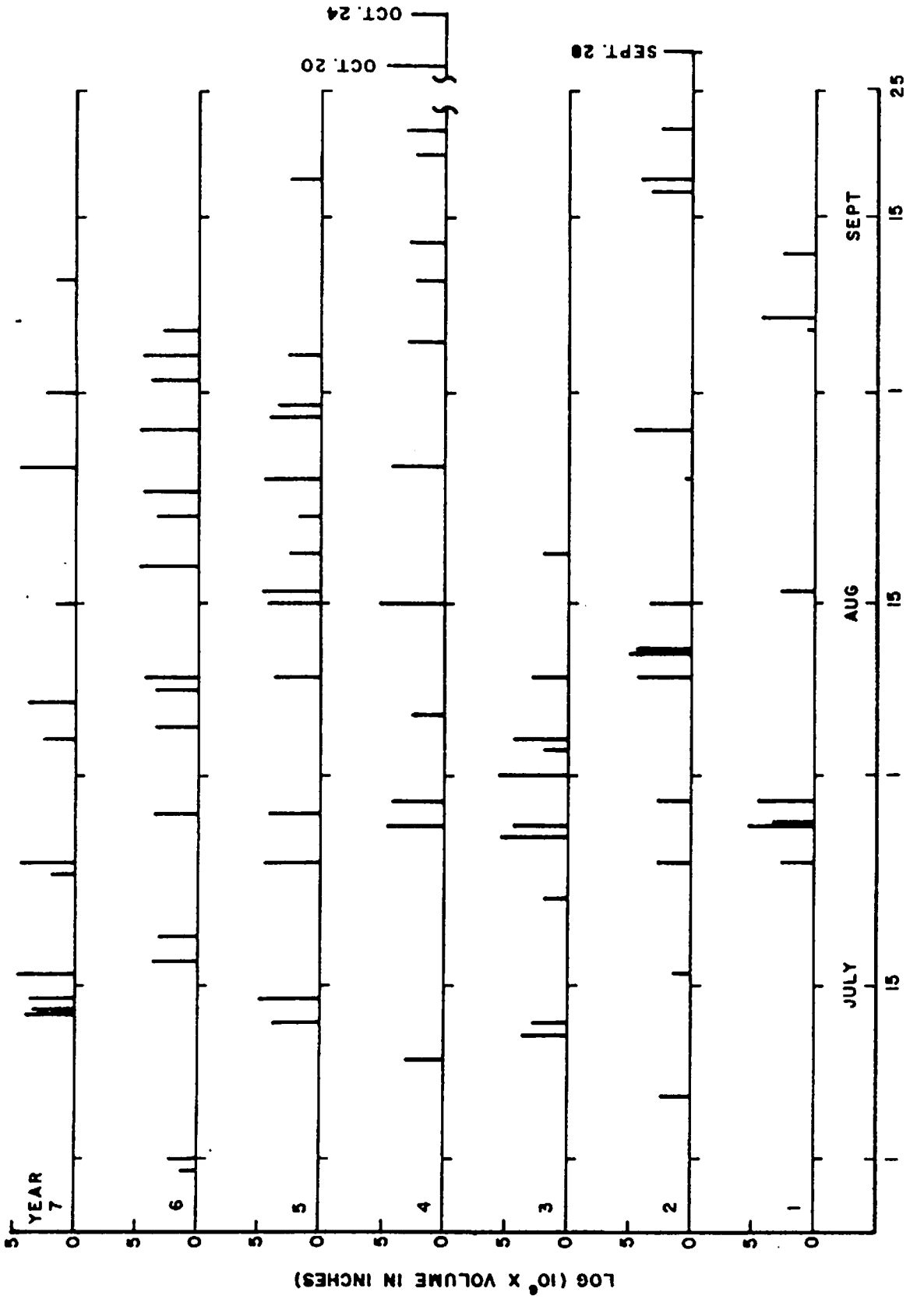

FIGURE 10. TIME DISTRIBUTION OF RUNOFF VOLUMES, SYNTHETIC DATA FOR WATERSHED 6.

TABLE 2.--Comparison of parameters for 7 years of actual data and five 7-year periods of synthetic data

Variable	Parameters	Data set					
		Observed	F-1	F-2	F-3	F-4	F-5
S 1/	mean	196.0	201.3	200.7	195.7	191.9	194.9
	s.d. 2/	8.70	7.57	6.18	7.45	7.65	8.80
N	mean	12.0	12.6	11.1	13.7	12.3	10.4
	s.d.	3.06	3.46	1.57	3.77	2.50	2.07
T (hours)	mean	18.0	18.0	18.5	17.7	17.4	18.0
	s.d.	3.53	3.92	3.66	3.43	3.57	3.78
V (inches)	mean	0.025	0.029	0.034	0.015	0.024	0.016
	s.d.	0.050	0.099	0.090	0.035	0.062	0.031
L(10^{-6}inches)	mean	3.30	3.39	3.17	3.25	3.37	3.23
	s.d.	1.27	1.11	1.34	1.09	1.17	1.30
P (cfs)	mean	562.	668.	803.	518.	705.	514.
	s.d.	1165.	1655.	1999.	1107.	1776.	991.
D (days)	mean	5.38	5.20	5.42	5.84	6.00	5.29
	s.d.	6.10	4.76	5.33	5.11	5.39	5.23
Mean annual volume (inches)		0.300	0.367	0.380	0.209	0.298	0.168
Mean length of season (days)		59.1	60.1	55.0	74.3	67.7	49.9

1/ Variables defined in Table 1.

2/ s.d. = standard deviation.

Two other comparisons between synthetic data and the original data are given in Figures 11 and 12. Figure 11 shows the cumulative yearly volume of runoff for the seven years of record compared to corresponding cumulative curves for the 5 sets of 7-year periods of synthetic data. Figure 12 gives plots of values of maximum yearly peak discharge on Gumbel's extreme value probability paper. The lines shown are for the seven years of record and the enveloping lines for the annual peak distributions for 5 sets of 7 years of synthetic data.

CONCLUSIONS

Runoff data in southeastern Arizona and in other semiarid regions where runoff events are intermittent and independent of each other may be generated with the aid of a stochastic model described herein. The model starts by generating the starting date of the season. Using this value, the number of events per season is generated using a correlation between the two variables. For each event, the model generates independently the time of occurrence of the event, the time interval to the next event, and the volume of runoff. The peak discharge of the event generated is based on the correlation between the peak and volume of runoff.

The time distribution of runoff or the shape of the hydrograph was not included in the present model. The main reason for this omission is the short duration of the runoff events in the watershed concerned. Runoff events in the watershed last usually less than 6.0 hours and, in many cases, less than 4.0 hours with the discharge reaching a peak value some 0.3 to 0.6 hours after the start of flow in the previously dry channel.

Future improvements of the model will include some considerations of the time element of the hydrograph. Another development considered is to apply models of a similar structure to other watersheds in southeastern Arizona so as to obtain the changes in the parameters with the size of the watersheds and other characteristics which may be significant. This type of investigation can lead to a regional stochastic model of runoff in semiarid areas which may be used for generation of synthetic data in ungaged watersheds.

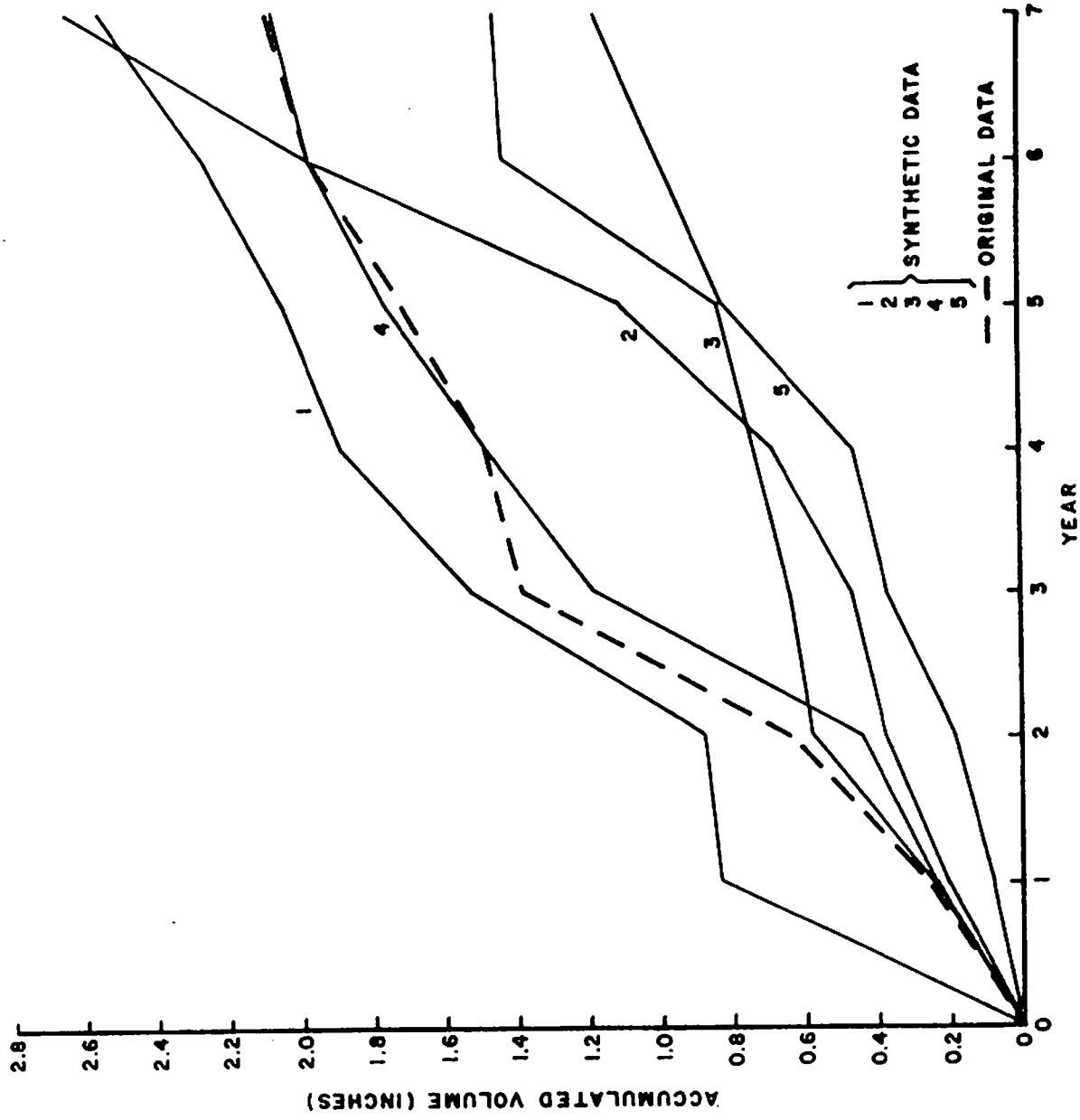

FIGURE 11. ACCUMULATIVE VOLUME OF RUNOFF FOR HISTONIC AND SYNTHETIC DATA.

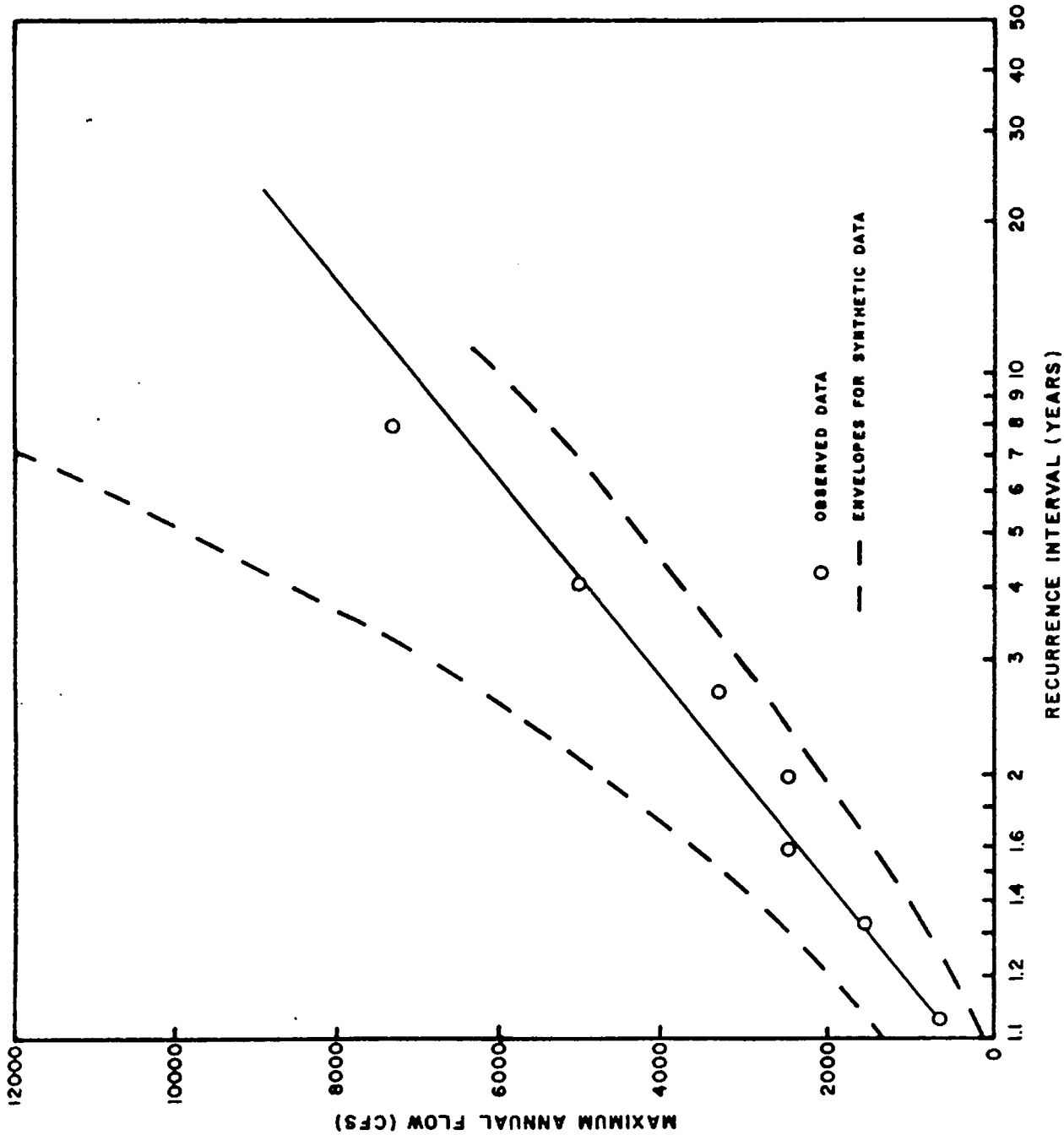

FIGURE 12. DISTRIBUTION OF MAXIMUM ANNUAL DISCHARGE FOR HISTONIC AND SYNTHETIC DATA.

The stochastic model described herein will also be related to a stochastic model of convective precipitation now under development at the Southwest Watershed Research Center in Tucson, Arizona.

REFERENCE CITED

Gwinn, W. R. Walnut Gulch supercritical measuring flume. Trans. ASAE 7(8):197-199, 1964.