Evaluation of a Basin-Wide Stochastic Model for Ephemeral Runoff from Semiarid Watersheds

L. J. Lane and K. G. Renard
MEMBER
ASAE

SYNTHETIC hydrology techniques permit a long series of hydrologic events to be generated using statistical parameters evaluated from a short sample. These techniques are most useful when only short records are available. Longer records are really necessary. A problem from such methods occurs because sampling errors are preserved in the synthetic data. Unrealistic design may result. The authors believe this uncertainty can be partly overcome by generating several sequences of synthetic data and comparing the results to obtain a range of answers.

The model used by the authors contains five statistical distributions to generate individual flow which, in turn, are added to provide annual data. Although some of the statistical distributions are based on limited samples others have many events and should not be subject to the customary short-sample problems. Maass, et al. (1962), recognized the problems of short samples because sequences of high or low flows are generated, which are not encountered in the original record, the synthetic results may be an improvement over the original record. Benson and Matalas (1967) explained another weakness of synthetic series and proposed a solution. Most models are based on actual records at a gaged location. Benson and Matalas (1967) used generalized statistical parameters estimated from physical and climatic characteristics of the drainage basins to provide data at ungaged locations. Diskin and Lane (1970) used essentially the same approach except the parameters of the individual frequency distributions were related to drainage area only. The model parameters were related only to watershed area, because on subareas within the Walnut Gulch basin other characteristics are relatively homogeneous. However, most of the parameters were highly correlated with watershed area. Therefore, the model can be adapted to ungaged watersheds with similar physical characteristics in similar climatic provinces.

Most synthetic runoff models are designed to fulfill specific uses such as the "annual peak discharges" procedure developed by Getty and McHughes (1962). As a contrast, Chow and Kareliotis (1970) have formulated a comprehensive model which can be used to generate stochastic streamflows for the analysis of water resource systems. Their model, unlike the one evaluated here, involved both precipitation and runoff. The model in this paper is versatile in its uses. For example, because it generates all the runoff events in a year and in a sequence of several years, it can be used: (a) in water yield studies where drought sequences within and between years are important; and (b) in flood peak discharge conditions such as are required for bridge, culvert, and detention reservoir designs. Thus the single model can be used to satisfy design criteria in several alternative projects.

The model was tested on data from the Walnut Gulch Experimental Watershed in southeastern Arizona (Fig. 1). This 58-sq-mile watershed is an ephemeral tributary of the San Pedro River. The watershed, which is operated by the U. S. Agricultural Research Service, includes a dense network of rain gages and runoff-measuring flumes. Osborn and Hickok (1968) reported 70 percent of the annual rainfall of 14 in. and almost all runoff result from short-duration, convective thunderstorms during the summer months. The channels are dry 99 percent of the time, and flow events from consecutive thunderstorms are generally independent.

Article was submitted for publication on April 23, 1971; reviewed and approved for publication by the Soil and Water Division of ASAE on October 11, 1971.
Contribution of the Soil and Water Conservation Research Division ARS, USDA, in cooperation with the Arizona Agricultural Experiment Station, Tucson, Arizona.
The authors are: L. J. LANE, Hydrologist, and K. G. RENARD, Research Hydraulic Engineer, Southwest Watershed Research Center, Tucson, Ariz.

THE MODEL

A stochastic model for generation of intermittentent and independent runoff events proposed by Diskin and Lane (1970) was used in this study. Using

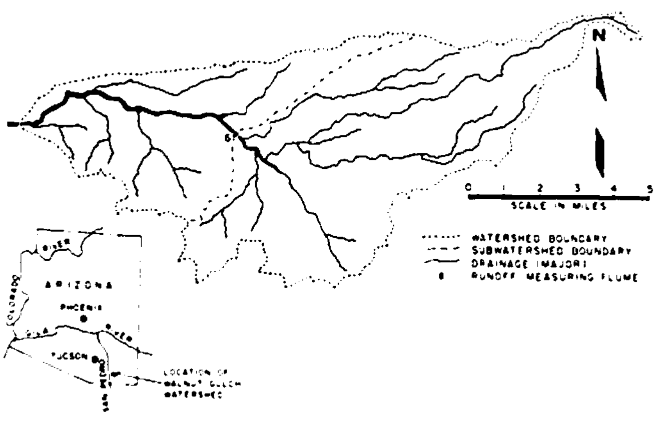

FIG. 1 Drainage map of Walnut Gulch watershed showing flow measuring station No. 6.

TABLE 1. ASSUMED PROBABILITY DISTRIBUTIONS FOR VARIABLES DESCRIBING RUNOFF EVENTS

Runoff variable	Symbol used	Theoretical distribution	Parameters
Start of runoff season	S	Normal	Mean, standard deviation
Number of events at outlet per runoff season	N	Normal	Mean, standard deviation
Begin time of each event	T	Normal	Mean, standard deviation
Interval between events	D	Negative exponential	Mean
Logarithm of volume of runoff for each event	L	Normal	Mean, standard deviation (of logarithms)

This article is reprinted from the TRANSACTIONS of the ASAE (Vol. 15, No. 2, pp. 280, 281, 282, 283, 1972)
Published by the American Society of Agricultural Engineers, St. Joseph, Michigan

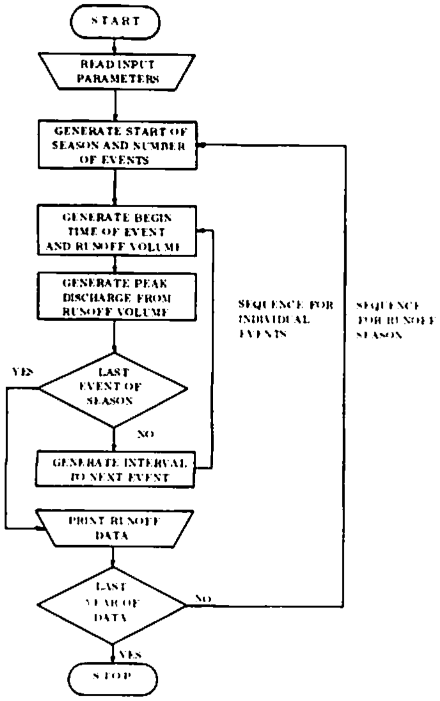

FIG. 2 Flow chart of synthetic data generation process.

TABLE 2. COMPARISON OF PARAMETERS USED IN GENERATION OF SYNTHETIC DATA

Runoff variable	Parameters	Historic data (93 events)	Synthetic data: Set 50-1 (605 events)	Set 50-2 (567 events)	Set 50-3 (578 events)
S*	Mean	196.2‡	196.7	197.4	196.5
	s.d.†	8.08	8.07	8.73	7.96
N	Mean	11.6	12.1	11.3	11.6
	s.d.	3.46	3.28	3.59	2.57
T, hr	Mean	17.8	17.9	18.0	17.9
	s.d.	4.16	4.16	4.20	4.07
D, days	Mean	5.58	5.82	5.86	5.73
	s.d.	6.44	5.90	5.95	5.34
L (10^{-6} in.)	Mean	3.17	3.26	3.20	3.17
	s.d.	1.28	1.15	1.17	1.13

*Variables defined in Table 1.
†s.d. = standard deviation.
‡Days since beginning of the calendar year.

data from the Walnut Gulch watershed, Diskin and Lane found that two variables could describe the runoff season: (a) the starting date of the summer runoff season and (b) the number of thunderstorm runoff events recorded at the watershed outlet per season. Since the starting and ending dates of the runoff season were independent, a linear relation between the starting date of a season and the number of events in that season was assumed even though the correlation was low. The form of the equation was

$$N = -BS + C \quad [1]$$

where N is the number of events, S is the starting date, and B and C are positive constants. Therefore, for a late starting date, fewer events might occur.

Two variables were necessary to describe the temporal position of runoff events: (a) time of day of the runoff event, and (b) interval between intermittent events. The runoff volume and peak discharge for each event were the final variables used to characterize ephemeral runoff. Runoff volume and peak discharge were highly correlated (r = 0.95), and thus the peak discharge was generated from the runoff volume. The probability distributions assumed for these variables are shown in Table 1. A normalizing logarithmic transformation was made on the runoff volumes. These log values are used within the model, and the inverse transformation is taken as the last step before the sequences of synthetic runoff data are listed in the computer output. The Kolmogorov-Smirnov one-sample goodness-of-fit test was used as the objective method for choosing the distributions listed in Table 1.

Using the model described above, generation of synthetic runoff data can be considered as two sequential operations. The first operation generates the starting date of the season, S, and the number of events in that season, N. The second operation generates beginning time and runoff volume for N individual events. Peak discharge is generated as a function of the volume. Finally, the length of the interval to the next event is specified. The process is continued until the season is complete and the specified number of years of data is produced.

To generate a random variable with a specified probability distribution, a uniform random variable, u, is equated to the cumulative distribution function F (x) and the inverse F^{-1} (u) specifies a random variable x. In practice a random variable is produced by generating pseudo-random numbers using a random numbers generator on a computer and then transforming the values so that the desired probability distribution is obtained.

Fig. 2 is a flow chart of the data generation scheme used. The input parameters consist of the means and standard deviations of the necessary runoff variables shown in Table 1. The inner sequence in Fig. 2 is repeated N times for N individual events in each season, and the outer sequence is repeated as specified by the number of years of data required.

CHARACTERISTICS OF OBSERVED DATA

Watershed 6 (63.006) comprises the upper 36.7 sq miles of the Walnut Gulch Experimental Watershed (Fig. 1). Renard (1970) and others have described the Walnut Gulch watershed and the hydrology of semiarid watersheds. Runoff records from 1962 to 1969 form the eight years of actual data used in this study.

For these eight years of data, the average starting date of the summer runoff season was July 15, with a standard deviation of approximately eight days (Table 2). There were 12 events per season with a standard deviation of 3.5. The average beginning time of runoff events within a day was 1750 hr (military time), with a standard deviation of 4.2 hr. The average interval between events was 5.6 days, producing a mean length of runoff season of [illegible] days. The length of runoff season ranged from 41 to 80 days.

For 93 individual runoff events, the mean runoff volume for the basin area was 0.022 in., with a standard deviation of 0.050 in. The mean peak discharge was 495 cfs, with a standard deviation of 1100 cfs. Mean annual runoff volume was 0.25 in. and ranged from 0.009 in. to 0.75 in. Maximum peak discharge for an individual event during the 8 years was 7300 cfs.

COMPARISON OF ACTUAL AND SYNTHETIC DATA

Two comparisons are necessary to judge the degree of correspondence between synthetic and actual runoff data. The first comparison requires preserving the input parameters used for the stochastic model in the synthetic data. Parameters of the actual or historic data and of the three 50-year sets of synthetic data are shown in Table 2. Correspondence is good between the actual and synthetic means and standard deviations of each of the five variables shown. No stochastic model can preserve all properties of observed runoff data. However, the authors feel that if a model is to be

TABLE 3. COMPARISON OF PARAMETERS NOT USED IN GENERATION OF SYNTHETIC DATA

Runoff variable	Parameters	Historic data (93 events)	Synthetic data Set 50-1 (605 events)	Synthetic data Set 50-2 (567 events)	Synthetic data Set 50-3 (578 events)
Peak of individual discharges, cfs	Mean	495.	614.	545.	509.
	s.d.	1114.	1699.	1322.	1483.
	range	0–7300.	0–23,800.	0–11,700.	0–21,300.
Max.* annual discharge, cfs	Mean	2845.	4152.	3399.	3486.
	s.d.	2390.	4212.	2732.	3656.
	range	64.–7300.	417.–23,800	26.–11,700.	184.–21,300.
Annual runoff volume, in.	Mean	0.25	0.27	0.22	0.21
	s.d.	0.24	0.21	0.18	0.16
	range	0.009–0.750	0.019–0.825	0.0002–0.670	0.017–0.852
Length of runoff season, days	Mean	59.2	64.6	60.6	60.5
	s.d.	12.1	28.0	28.0	19.6
	range	41–80	16–137	1–150	23–118

*Annual values based on 8 years of historic data and 50-year sets of synthetic data.

useful for flood discharge and water yield studies, it must preserve reasonable values for the four runoff variables listed in Table 3. The second comparison, therefore, involves parameters not used as direct input to the model. While the means and standard deviations for the same variable should have approximately the same values (Table 3) for all data sets, it seems reasonable to expect wider ranges in the longer synthetic sequences. Peak discharges and lengths of runoff season satisfy this hypothesis, but annual runoff volumes compare very closely with the observed data. The only serious differences in Table 3 are between the standard deviation of the runoff season length in the actual data and in the three synthetic data sets. This suggests a possible need for an added constraint on the length of runoff season within the structure of the model. The difference between standard deviations of the runoff season length in the actual and synthetic data was not apparent until analysis of the long sequences of synthetic data was complete. That these differences were not discovered in the short (8 years) sequence used when developing the model as reported by Diskin and Lane (1970) is justification for generating long sequences of synthetic data.

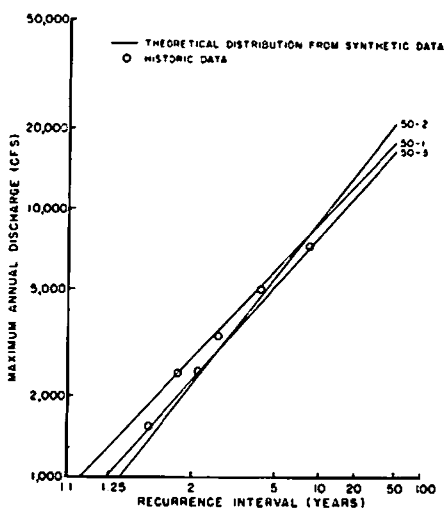

FIG. 3 Walnut Gulch watershed 6, maximum annual discharge, using a log-normal distribution.

Determination of design discharge values for small drainage basins subject to thunderstorm rainfall is a primary concern of engineers and hydrologists. Renard, et al. (1970) tested several distributions for maximum annual discharge on a semiarid rangeland watershed. The log-normal distribution was found to be an acceptable model for maximum annual discharge. Recurrence intervals for maximum annual discharge based on the three 50-year sets of synthetic data are shown in Fig. 3. The points plotted in Fig. 3 represent the actual data from the 36.7-sq-mile watershed, with two years having peak discharges of less than 1,000 cfs. These points indicate visually the good correspondence between the actual data and the theoretical straight-line (log-normal) distributions derived from the synthetic data. Differences in the mean and standard deviation shown in Table 3 are not apparent in this figure.

Sequential maximum annual peak discharge and annual runoff volumes from the actual and synthetic data are shown in Fig. 4. This graph was made to show possible trends or unreasonable values. No trends or "impossible" high values are obvious in any of the sequences of synthetic data. Osborn (1971) states that the "maximum expected" discharge from Watershed 6 is 23,000 cfs. This independent effort was made by centering on the watershed the maximum expected precipitation from an air-mass thunderstorm. The highest peak from each synthetic data set is 23,850 cfs, 21,300 cfs, and 11,740 cfs

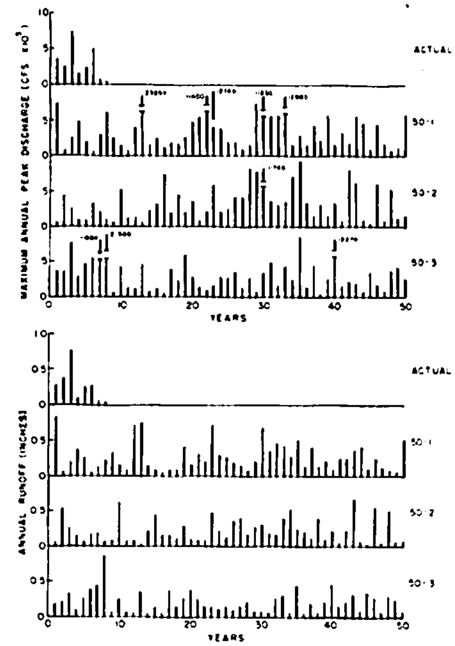

FIG. 4 Sequences of actual and synthetic runoff data for watershed 6.

which are of the same order as Osborn's maximum expected peak discharge. If the log-normal assumption for maximum annual discharge is valid, then Osborn's "maximum expected" discharge would not represent an extremely rare event. However, the maximum expected precipitation event is based on 15 years of record at Walnut Gulch and on longer records on USWB stations in Arizona. Therefore, the log-normal assumption or other characteristics of the model may lead to somewhat conservative (higher) estimates of maximum annual discharge.

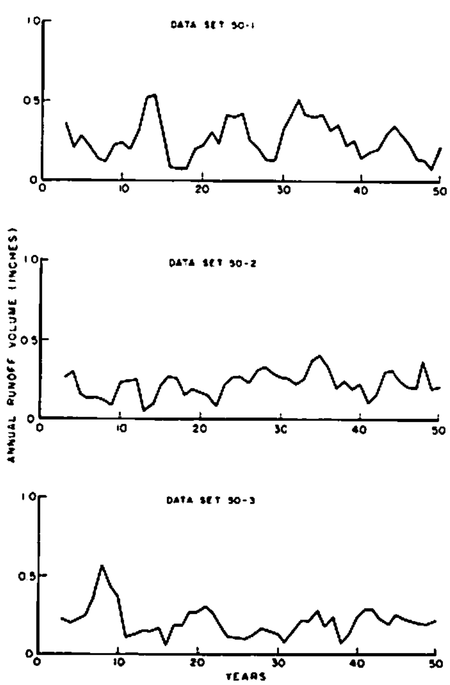

FIG. 5 Walnut Gulch 3-year running mean of generated annual runoff volumes.

Fig. 5 shows three-year running means of annual runoff for the three 50-year sequences. Runoff averages for the data shown in the lower half of Fig. 4 were computed and plotted at the end of each three years. Engineers and hydrologists designing water supply facilities must often provide storage to carry water through low flow years. Thus the synthetic data from this model provide interesting insight to the problem. Each of the three sequences had a minimum yield value of approximately 0.05 to 0.07 in. per year. From this minimum three-year period, one could only expect to have about 0.15 in. of runoff from the 36.7-sq-mile watershed or 294 acre-ft of runoff. Thus in a high evaporation area such as this (Renard, 1970, reported pan evaporation of 100 in.), a storage reservoir could undoubtedly evaporate more than the inflow.

The maximum average 3-year runoff from the three synthetic sequences ranged from 0.4 in. per year for set 50-2 to 0.55 in. per year for sets 50-1 and 50-3. This means that the maximum runoff expected from this particular drainage basin in any 3-year sequence would be about 3,200 acre-feet.

CONCLUSIONS

Methods have been presented for the analysis and comparison of actual synthetic streamflow from semiarid watersheds. The frequency of occurence of independent runoff events defines a summer runoff season. Synthetic data generation based on the concept of a runoff season enables comparison of actual and synthetic data both on an event and on an annual basis.

Two comparisons are necessary to evaluate the performance of a stochastic model for generating synthetic runoff data. First, the model must preserve the parameters used as input. The second comparison involves parameters not used as model input. Choice of the parameter not used as input, but important enough to be preserved, is dependent upon the intended uses of the synthetic data. However, maximum annual discharge, annual runoff volume, and length of runoff season are variables that should be preserved.

The comparisons between actual and synthetic data given in this paper indicate that the proposed model generates data that compare favorable with the actual data except for one parameter. The length of the summer runoff season is too variable in the synthetic data, although the mean values correspond. Adoption of a different distribution for the number of events per season may result in a less variable season length. Not considered in this paper, but worthy of future research is the relation between the number of events per season, interval between events, and length of season, when different distributions are adopted for each variable. Research is now under way to determine the effect of a Poisson distribution for the number of events per season on the season length while keeping the negative exponential distribution for the interval between events.

Significant progress has been made in developing a basin-wide model for ephemeral runoff. Further work will include extending the model to provide a regional stochastic model, allowing synthetic runoff generation for other ungaged basins where thunderstorm runoff predominates. Before a regional model could be developed, it was necessary to demonstrate that the model would generate reasonable data on gaged watersheds.

Another approach which might be investigated to improve the model was proposed by DeCoursey (1970) when he examined the problem of extrapolating probability distributions to produce extreme values. This work involved a method of adjusting the distribution so that it conformed better with observed data. DeCoursey (1970) operated on a skewed random normal variate with a linear transformation which accomodated the upper end of the probability distribution. This approach, along with an effort to adjust to statistics of the short-term record to what might be expected for a long-term record when comparing watersheds with different record lengths, needs further investigation.

References

1 Benson, M. A. and N. C. Matalas. 1967. Synthetic hydrology based on regional statistical parameters. Water Resources Res. 3(4):931-935.

2 Chow, V. T. and S. J. Kareliotis. 1970. Analysis of stochastic hydrologic systems. Water Resources Res. 6(6):1529-1582.

3 DeCoursey, D. G. 1970. Use of multiple discriminant analysis to evaluate the effects of land use change on the simulated yield of a watershed. School of Civil Engineering Doctoral dissertation, George Institute of Technology, Atlanta, 222pp.

4 Diskin, M. H. and L. J. Lane. 1970. A stochastic model for runoff events for a semiarid watershed in southeastern Arizona. Proc. ARS-SCS Watershed Modeling Workshop, Tucson, Ariz., 23 pp.

5 Getty, H. C. and J. H. McHughes. 1962. Synthetic peak discharges for design criteria. J. Hydraulics Div., ASCE 88(HY-5):1-13.

6 Maass, Arthur, et al. 1962. Design of water-resource systems. Harvard Univ. Press, Cambridge, Mass. 620 pp.

7 Osborn, H. B. and R. B. Hickok. 1968. Variability of rainfall affecting runoff from a semiarid rangeland Watershed. Water Resources Res. 4(1):199-203.

8 Osborn, H. B. 1971. Thunderstorm runoff in southeastern Arizona. Dept. of Civil Engineering Doctoral Dissertation, University of Arizona, Tucson, 161 pp.

9 Renard, K. G. 1970. The hydrology of semiarid rangeland watersheds. USDA Pub. ARS 41-162, 26 pp.

10 Renard, K. G. J. C. Drissel and H. B. Osborn. 1970. Flood peaks from a small southwest rangeland watershed. Proc. ASCE 96(HY-3):773-785.