COMPARISON OF FLOOD FREQUENCY FROM SYNTHETIC AND OBSERVED DATA ON SMALL DRAINAGE AREAS IN MISSISSIPPI

B. E. Colson
U.S. Geological Survey
Reston, Virginia

INTRODUCTION

The U. S. Geological Survey and the Mississippi State Highway Department have a long-standing cooperative program of water-resources investigations, one element of which is the continuing investigation of the flood frequency of streams in Mississippi. In 1955, the Survey began the systematic collection of peak-flow data on small drainage areas in Mississippi.

In 1964 the Survey entered into a nationwide program sponsored by the Federal Highway Administration to study flood frequency of small drainage areas. The Mississippi District of the Survey, in cooperation with the State Highway Department, expanded the streamgaging network for small drainage areas to a total of 95 sites to collect rainfall and runoff data. Gages were relocated or discontinued for various reasons resulting in fluctuations of the number of stations operated during the study.

During this time, the nation was involved in a massive highway construction program which required hydraulic design of stream crossings. It was deemed undesirable to wait for the collection of long-term flood records on small streams for frequency analysis.

The Survey developed a rainfall-runoff model (Dawdy and others, 1972) to generate flood peaks using available long-term rainfall data. The model was calibrated for each of the gaging sites by optimizing 10 basin parameters using concurrent precipitation and runoff records (table 1). It was determined during the calibration process that four of these parameters -- DRN, EVC, RR, TP/TC (table 1) -- could be held constant. Subsequently, each set of parameters was used in the model with long-term climatic records collected by the National Weather Service at Meridian and Vicksburg, Miss., Memphis, Tenn., and New Orleans, La., and pan-evaporation data from Mississippi State University to generate synthetic flood records.

The frequency analyses of these synthetic records were used to prepare the statewide report, "Flood Frequency of Mississippi Streams" (Colson and Hudson, 1976). Upon completion of this report, it was recommended that the small streams network be continued until at least 20 to 25 years of record was available to assess the validity of the synthetic flood-peak data. In 1977, a technique described by Moss and Karlinger (1974) was applied to the streamgaging network. This analysis indicated that 80 percent of the information could be obtained from 45 sites chosen to represent a broad range of stream characteristics. The small basin gaging station network was reduced to 45 sites during 1977. The reduced network was operated through the water year ending September 1984. The Survey agreed to make a comparison of the synthetic data with results from observed data through 1984.

This report provides a comparison of the synthetic flood-frequency values used in the first report "Flood Frequency of Mississippi Streams" and the values computed from observed annual peaks through the 1984 water year. Fifty-one sites that have 16 years or longer record are used in this report for comparison with the synthetic results (fig. 1).

Table 1.–The ten model parameters and their applications in the modeling process

Parameter	Units	Definition and Application
		Antecedent moisture component
EVC----------	------------------------	Coefficient to convert pan evaporation to potential-evapotranspiration values.
RR------------	------------------------	Proportion of daily rainfall than infiltrates the soil.
BMSM-------	Inches--------------	Soil moisture storage volume at field capacity.
DRN----------	------------------------	Drainage parameter for redistribution of soil moisture (fraction of KSAT).
		Infiltration component
PSP-----------	Inches---------------	Product of moisture deficit and suction at the wetted front for soil moisture at field capacity.
KSAT--------	Inches per hour-	The minimum (saturated) hydraulic conductivity used to determine infiltration rates.
RGF-----------	------------------------	Ratio of the product of moisture deficit and suction at the wetted front for soil moisture at wilting point to that at field capacity.
		Surface runoff component (routing)
KSW----------	Hours---------------	Time characteristic for linear reservoir routing.
TC-------------	Minutes------------	Length of the base of the triangular translation hydrograph.
TP/TC--------	------------------------	Ratio of time to peak to base length of the triangular translation hydrograph.

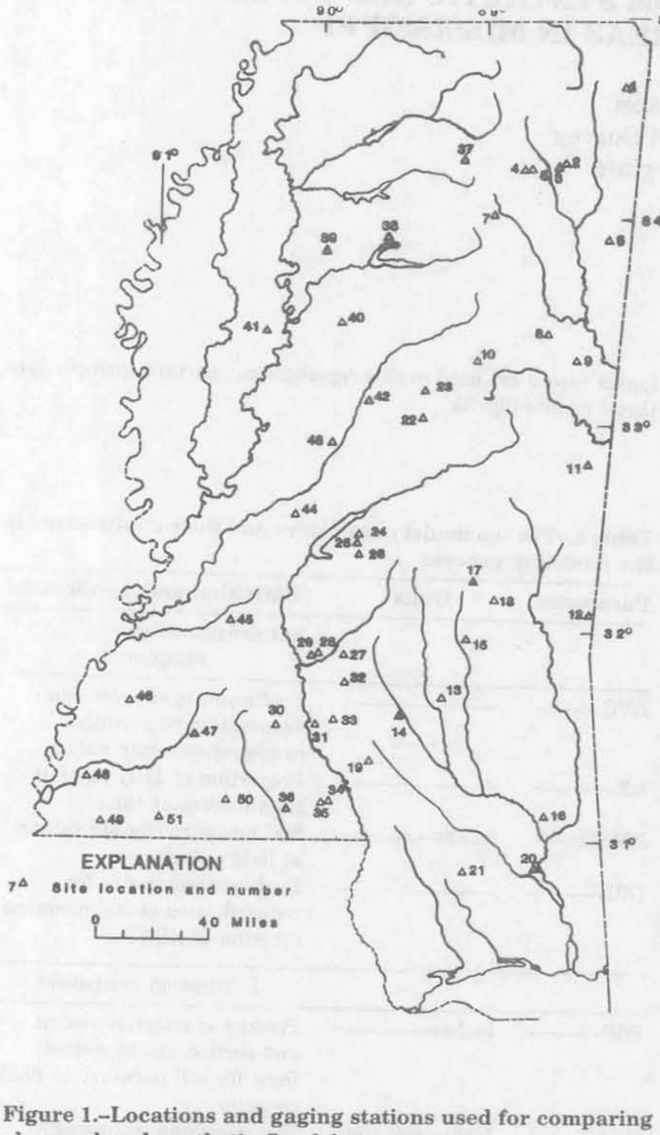

Figure 1.–Locations and gaging stations used for comparing observed and synthetic flood frequency.

FLOOD-FREQUENCY ANALYSIS

The synthetic values used in the 1976 report were computed from the annual peaks generated by using long-term rainfall data in a calibrated model for each of 89 sites. This produced four sets of annual peak discharges at each gage site, based on rainfall for Meridian and Vicksburg, Miss., Memphis, Tenn., and New Orleans, La.

A frequency curve for each of these four synthetic data sets for each site was computed following procedures outlined in U.S. Water Resources Council Bulletin 15 (1967). A weighted average of the four frequency curves was obtained for each site based on an isohyetal map of Mississippi and the site location. The average rainfall for the period of record used was 53 inches at Meridian, Miss., 51 inches at Vicksburg, Miss., 48 inches at Memphis, Tenn., and 62 inches at New Orleans, La. The procedure was to determine the mean annual rainfall at each site from an isohyetal map of the state, then select the appropriate weighting of the discharges from the following formulas.

Sites having less than 51 inches	$\frac{\text{Vicksburg, Miss.} + \text{Memphis, Tenn.}}{2}$
Sites having 51 to 53 inches	$\frac{\text{Meridian} + \text{Vicksburg}}{2}$
Sites having 54 to 56 inches	$\frac{(2\text{x Meridian}) + \text{Vicksburg} + \text{New Orleans, La.}}{4}$
Sites having 57 to 59 inches	$\frac{\text{Meridian} + \text{Vicksburg} + (3\text{x New Orleans, La})}{5}$
Sites having greater than 59 inches	New Orleans, La.

The discharge values of the 2-, 5-, 10-, 25-, 50-, and 100-year recurrence intervals from the log-Pearson Type III analyses were weighted using these formulas. For example, if the mean annual rainfall at a site was 52 inches, the average of the frequency values obtained from synthetic discharges generated by the Meridian and Vicksburg long term rainfall was used. Therefore the synthetic curves were usually a combination of more than one flood-frequency curve.

ANALYSIS OF OBSERVED DATA

Observed annual peak-flow data for 16 years or longer are available at 51 sites for which synthetic data were used in the 1976 flood-frequency report. The average length of record is just over 21 years and some of these sites how have 30 years of observed annual peak discharges. The basin area of the sites ranges from 0.07 to 4.35 mi^2 (table 2). The channel length, which ranged from 0.34 to 3.69 miles, was measured upstream from the site to the basin divide. The average slope, in feet per mile, was computed between points 10 and 85 percentile of the length upstream and ranged from 10.6 to 195. These three basin parameters were found to be significant in regression analysis of flood magnitudes (Colson and Hudson, 1976).

The observed data were analyzed by fitting a Pearson Type III distribution to the logarithms of the annual peak discharges following procedures outlined in U.S. Water Resources Council Bulletin 17B, "Guidelines for Determining Flood-Flow Frequencies" (1981). When logarithms of the discharges are used the distribution is usually referred to as a log-Pearson Type III distribution. The mean, standard deviation, and skew completely define the log-Pearson Type III distribution. Using these values the flood magnitude for any recurrence interval may be computed.

The synthetic data (Colson and Hudson, 1976) were analyzed in accordance with methodology described in Bulletin 15 of the U.S. Water Resources Council (1967). Both Bulletins 15 and 17B use a log-Pearson Type III distribution but differ in the weighting procedure for determining skew and the detection of outliers. Bulletin 15 did not weight the generalized skew and station skew as proposed in Bulletin 17B. Bulletin 17B recommends that the skew computed from the annual peaks be given a weight inversely proportional to the variance computed from the station data. No outliers were observed in the synthetic data.

High outliers were detected in 3 of the observed records and 1 low outlier was detected in each of 13 observed records. These were all treated according to procedures given in Bulletin 17B. Data for one station, Goines Draw near Prentiss, appears to be anomalous. The presence of five abnormally low annual peak discharges distorted the standard deviation. These low peak discharges were truncated by excluding peaks less than 20 ft^3/s and the standard deviation was reduced from 0.603 to 0.445 log units -- still considerably greater than the standard deviation of the other 50 sites.

COMPARISON OF RESULTS

One of the most effective ways to compare results is by graphical plots. The 2- and 100-year flood magnitudes of the synthetic results versus those obtained from observed data are compared in figures 2 and 3. These are typical of the results obtained. A visual inspection indicates that, in general, there is good agreement between the synthetic and observed results.

A closer study indicates that values computed from the observed record tend to be slightly greater than values obtained from the synthetic record. For statistical analysis the data were transformed by taking the logarithms of the observed and synthetic discharge values. The RMSE (root mean square error) and the mean of the residuals between observed and synthetic discharge values were computed for the 2-, 5-, 10-, 25-, 50-, and 100-year recurrence interval floods. The results, expressed as a percent of the observed data value, show that the average RMSE ranges from 27 to 38 percent and the average difference from -9 to -18 percent at the 2- and 100-year recurrence intervals respectively. Student's t statistic was computed from the residuals for each of the recurrence interval. The values of Student's t ranged from -2.56 to -4.26, which at the 5 percent significance level, indicate that the synthetics results are biased.

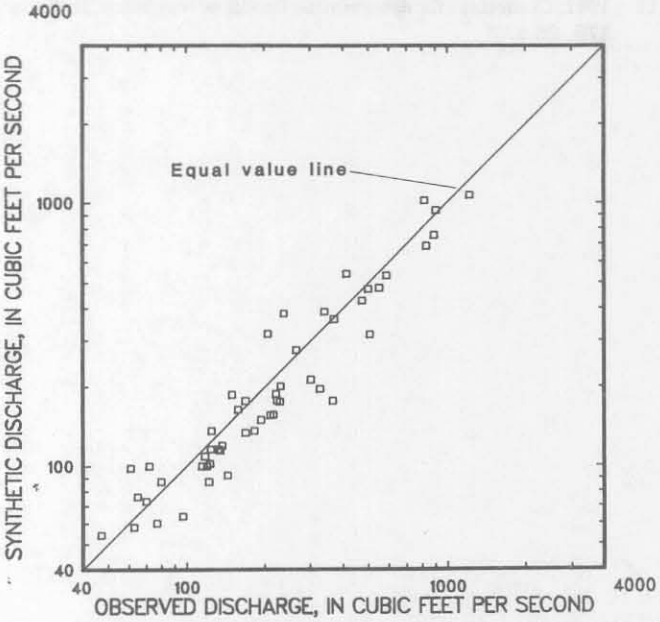

Figure 2.–Comparison of synthetic and observed 2-year flood discharge.

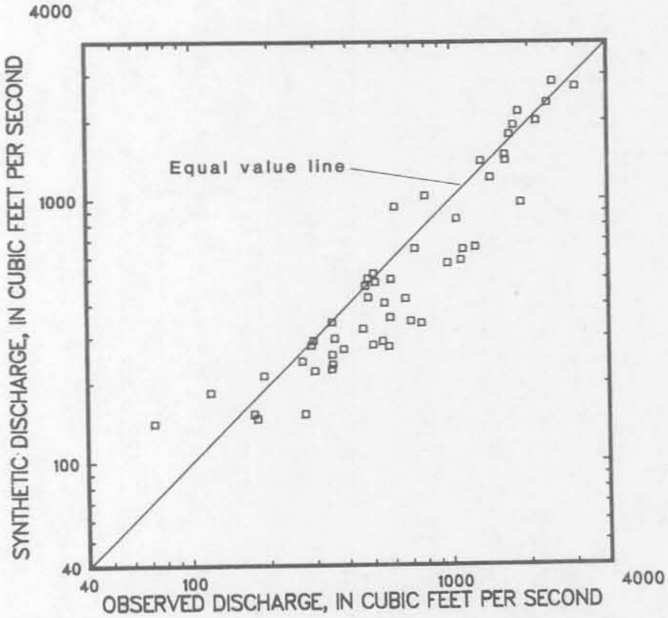

Figure 3.–Comparison of synthetic and observed 100-year flood discharge.

The results of this analysis are shown below:

Recurrence Interval	2-year	5-year	10-year	25-year	50-year	100-year
RMSE difference In percent	27	27	28	32	35	38
Average difference in percent (BIAS)	-9	-11	-13	-15	-16	-18

The flood-frequency curves for the 51 sites were plotted for both synthetic and observed data. Many of the observed data curves crossed, indicating a marked difference in slope among the frequency curves. Comparing the synthetic frequency curves, fewer of the curves crossed, indicating less variation in slope. Both sets of frequency curves span about the same range of discharges but the observed data curves exhibit a slightly wider variation.

The reduced variance in the synthetic results is partially due to model limitations. Many basin parameters were averaged or lumped into the model. Only four sets of long-term rainfall records were available to generate the synthetic flood peaks. The rainfall was implicitly distributed over the basin in the same manner as for the storms used in calibration of the model.

The greater variance in the frequency curves that were based on observed data (fig. 8) is due partly to shorter record length as well as the increased freedom for variation in basin characteristics that affect the distributions of flood peaks. Benson (1952) showed that within a known homogeneous record that the random error for the 100-year flood could range from about 30 percent less than to more than 50 percent greater than the true frequency curve when based on 25 years of annual peaks.

Benson concluded that accuracy within 10 percent is rarely attainable and that accuracy within 25 percent should be considered satisfactory. Considering the greater variation that is to be expected from the relatively short length (21 years) of observed record, the 27 to 38 percent RMSE difference between the synthetic record (67 years) is considered to be a satisfactory agreement.

SUMMARY

The immediate need for flood-frequency information on small drainage areas led the U.S. Geological Survey to synthesize flood peaks using a rainfall driven digital model. An average of 21 years of annual peaks have been collected for 51 sites in Mississippi for which synthetic records were developed. The synthetic records were analyzed using techniques described in Bulletin 15 of the Water Resources Council (1967). These results are those used in the report "Flood Frequency of Mississippi Streams" (Colson and Hudson 1976). The observed annual peaks were analyzed under guidelines presented in Bulletin 17B of the Water Resources Council (1981).

The differences between observed and synthetic peak records for each of the 2-, 5-, 10-, 25-, 50-, and 100-year recurrence-interval floods are evaluated. The average difference indicated as bias ranges from -9 to -18 percent. The average root-mean-square error ranges from 27 to 38 percent. Plots of the synthetic versus the observed peak discharges were prepared for the 2- and 100-year recurrence-interval floods. Plots of the entire frequency curve for each of the 51 sites indicated more variation among frequency curves based on observed record than for the synthetic record. This is as expected due to the averaging effect of the model on the synthetic data and the longer record. The difference from the observed record is about the same as might be expected for 21 years of record. The results are considered satisfactory.

REFERENCES

1. Benson, M. A., 1952, Characteristic of frequency curves based on theoretical 1,000-year record: U.S. Geological Survey Open-File Report, 21 p.
2. 1962, Factors influencing the occurrence of floods in a humid region of diverse terrain: U.S. Geological Survey Water Supply Paper 1580-B, 64 p.
3. Colson, B. E., and Hudson, J. W., 1976, Flood frequency of Mississippi streams: Mississippi State Highway Department Report No. 76-014-PR, 34 p.
4. Dawdy, D. R., Lichty, R. W., and Bergman, J. M., 1972, A rainfall-runoff simulation model for estimation of flood peaks for small drainage basins: U.S. Geological Survey Professional Paper 506-B, 28 p.
5. Hardison, C. H., 1971, Prediction error of regression estimates of streamflow characteristics at ungaged sites: U.S. Geological Survey Professional Paper 750-C, p. 228-236.
6. Maddock, Thomas, III, 1974, An optimum reduction of gauges to meet data program constraints: Bulletin, International Association Hydrologic Sciences, V. 19, no. 3, p. 337-345.
7. Matalas, N. C., 1967, Optimum gaging station location: Proceedings IBM Scientific Computing Symposium on Water and Air Resource Management, October 23-25, 1967, Yorktown Heights, New York, p. 85-94.
8. Moss, M. E., and Karlinger, M. R., 1974, Surface-water network design by regression analysis simulation: Water Resources Research, v. 10, no. 3, p. 427-433.
9. Riggs, H. C., 1968b, Frequency curves: U.S. Geological Survey Techniques of Water Resources Investigations, book 4, chapter A1, 39 p.
10. U.S. Water Resources Council, 1967, A uniform technique for determining floodflow frequencies: U.S. Water Resources Council Bulletin 15, 15 p.
11. 1981, Guidelines for determining floodflow frequency, Bulletin 17B, 28 p.