VOL. 10, NO. 4 WATER RESOURCES RESEARCH AUGUST 1974

Generalized Skew Coefficients of Annual Floods in the United States and Their Application

CLAYTON H. HARDISON

U.S. Geological Survey, Reston, Virginia 22092

Generalized skew coefficients for use in defining flood-frequency curves that follow log Pearson type 3 distributions are shown by isopleths on a map of conterminous United States. The generalized logarithmic skew coefficients range from 0.6 along the eastern seaboard to −0.5 in Indiana and Illinois. West of the one-hundredth meridian the coefficients range from −0.3 to 0.2 except for a small area in Nebraska where the generalized skew goes as high as 0.4. The validity of the map values is verified by a split-sampling procedure. In the west the discharge of 50- and 100-yr peaks computed by using map values of skew is more accurate than that computed by using the observed skew of a sample of 30 annual peaks. East of the Mississippi River the accuracy is even higher and approaches the equivalent of 60 annual peaks along the east coast. An equation gives the adjustment by which a *T*-yr peak computed by using map skew would have to be increased to give a discharge that has an average exceedance probability equal to $1/T$.

Flood-frequency curves are commonly computed by fitting a Pearson type 3 distribution to the logarithms of annual peak discharges observed over a period of years. Such a curve depends on the mean, the standard deviation, and the skew coefficient computed from the logarithms of the annual peaks and is subject to the error inherent in estimating the population parameters from the sample statistics. The *Water Resources Council* [1967] recognized that the skew coefficient has greater variability between samples than the mean and the standard deviation do and suggested the possibility of using a regional value of skew coefficient in place of that based on a short record of annual peaks.

The map of generalized skew coefficients developed in this paper for conterminous United States evaluates areal variation in generalized skew coefficients without abrupt changes from region to region. Skew coefficients taken from the map may be used directly in the computation of *T*-yr peaks, or they may be used in an analysis of observed skew minus map skew against basin characteristics to obtain a better estimate of skew coefficients to use at a given site. Logarithmic skew coefficients computed from the statistics of an observed sample of annual peaks are called observed skew in this paper.

The applicability of the map of skew is tested by a split-sample procedure. In that procedure the available flood record for each site is split; one part is then used to compute a frequency curve based upon mapped skew, and the other part is used to compare the expected and the observed number of exceedances. Previous studies have shown that for records of finite length there is a bias in the computed frequency curves, so that the average probability of exceedance is not $1/T$. It was necessary therefore to adjust the expected number of exceedances or to adjust the computed frequency curves before comparing the observed and the expected number of exceedances. A by-product of this split-sample testing is a relation that adjusts *T*-yr values based upon map skew to obtain values that have an average exceedance probability of $1/T$.

In this report the techniques used in developing the generalized map of skew coefficients are described first. Following in order are a description of the split-sample procedure, an explanation of the two methods used to adjust frequency curves computed from sample statistics in the split-sampling tests, a comparison of the accuracy of *T*-yr peaks computed by using map skew and observed skew, and a discussion of the merits of using the mapped skew coefficients.

Map of Generalized Skew Coefficients

The isopleths of generalized logarithmic skew coefficients shown in Figure 1 are based on records of annual peak discharge at about 1450 stream gaging stations in conterminous United States. These stations were selected by the district offices of the Water Resources Division of the U.S. Geological Survey as having a record of at least 25 annual peaks through 1967 not seriously affected by regulation, diversion, or urbanization. The size of drainage area was limited to a maximum of 500 mi² for stations in the *U.S. Geological Survey Water-Supply Papers*, parts 1–4, and to a maximum of 1000 mi² elsewhere. The average value of the logarithmic skew coefficient for the stations used in each state is shown in Figure 2, together with the number of stations used in computing each average. No average skew coefficient is shown for Delaware, Nevada, South Dakota, and Rhode Island because of an insufficient number of stations that satisfy the criteria used. The observed skew coefficient for one station in each of seven other states was not used in computing the state average because it differed from the state mean by more than 3 standard deviation units.

The observed skew coefficient for each station used in computing the state averages was computed by the equation

$$g^* = N(x_i - \langle x\rangle)^3/[(N-1)(N-2)s^3] \qquad (1)$$

in which x_i represents the logarithms of annual peak discharge, $i = 1, \cdots, N$; $\langle x\rangle$ is the mean of the x_i; and s is the logarithmic standard deviation, computed as

$$s = [(x_i - \langle x\rangle)^2/(N-1)]^{1/2} \qquad (2)$$

The arithmetic average of the skew coefficients at the stations in each state was multiplied by 1.26 to remove the bias in the average skew coefficients computed by the method of moments. This adjustment was obtained from the equation $1 + 8.5/N$ given by *Hazen* [1930], using N as 33 yr, the average length of record at the stations used. Subsequent work with random samples, however, indicates that a ratio of 1.23 would have been a more accurate adjustment. The use of this ratio in place of 1.26 would change some of the large

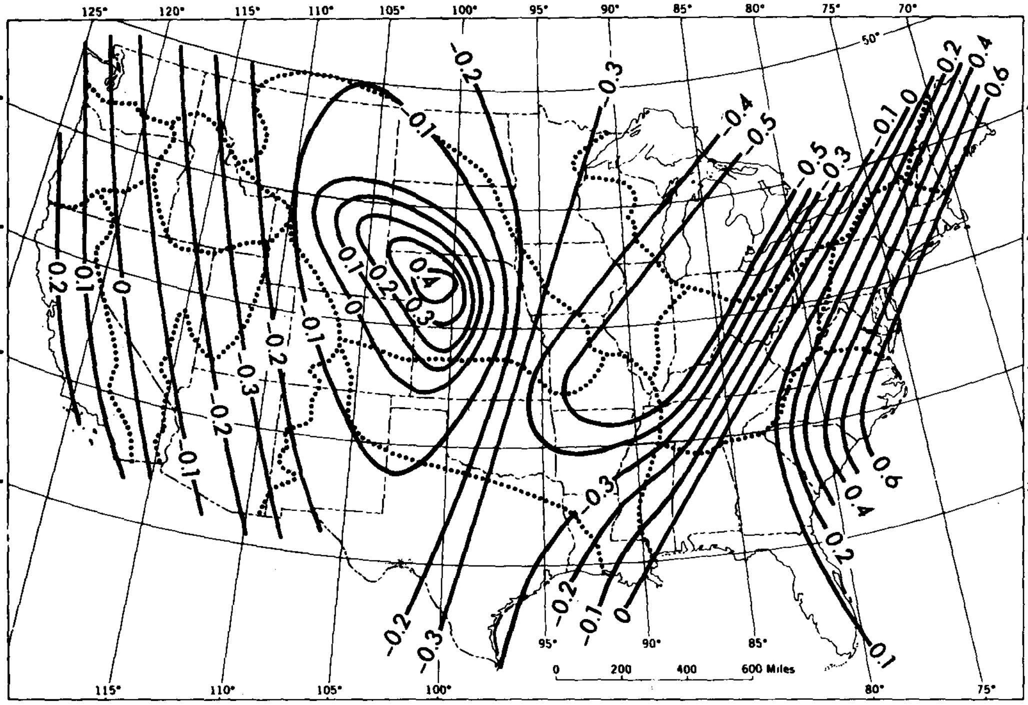

Fig. 1. Isopleths of generalized logarithmic skew coefficients of annual peak discharge. The dotted lines outline the areas in *U.S. Geological Survey Water-Supply Papers*, parts 1–14.

positive and negative averages shown on Figure 2 by 0.02 but would have little effect on the position of the lines shown in Figure 1.

A preliminary isopleth map of logarithmic skew coefficients was prepared primarily on the basis of the state averages, secondary consideration being given to the skew coefficients for individual stations in areas smaller than a state. For the vicinity around the Sand Hill area of Nebraska, skew coefficients for many additional stations that have less than 25 yr of record were used to help define the lines of positive skew. Skew coefficients taken from the preliminary isopleths were used in split sampling of the annual peak discharges at about 1350 stations, using records through 1971 in most states. The preliminary map was then revised where revision was necessary to make the probability of exceedance of T-yr peaks computed by using the map skew more consistent with a

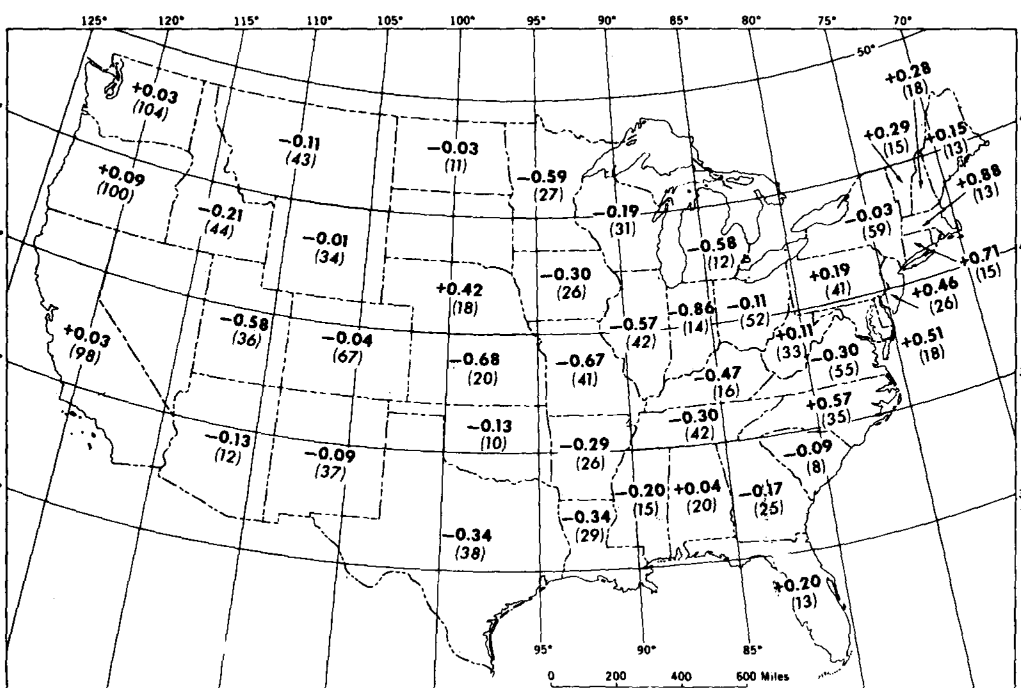

Fig. 2. State average values of logarithmic skew coefficients of annual peak discharge. The number of gaging stations used in computing each average is shown in parentheses.

probability estimated from known parameters. The logarithmic skew coefficients obtained from Figure 1 are called map skew in this paper.

Analysis of the variance of the observed skew coefficients indicates that the standard error of skew coefficients taken from the map is about 0.3 east of the Mississippi River and about 0.4 west of there.

There was no detailed study of areal variations of logarithmic skew coefficients for streams in Alaska and Hawaii. The statewide average skew coefficients are 0.33 for Alaska, based on 27 stations, and −0.08 for Hawaii, based on 46 stations.

SPLIT-SAMPLING VERIFICATION

A split-sampling procedure was used to check and to modify the map of generalized skew coefficients. Each flood record was split into samples of 10 and 20 yr beginning with the first year of record and additionally into other 10- and 20-yr samples by starting with years 5, 9, 13, and 17 for 20-yr samples and with years 3, 5, 7, and 9 for 10-yr samples.

These 10- and 20-yr samples were used as follows: to define 10-, 50-, and 100-yr floods from each sample by the log Pearson type 3 method using map skew and then to count how many times these peak discharges were exceeded in the remainder of the record from which the sample was taken. The total number of exceedances divided by the total number of years from which they were observed is the observed probability of exceedance. This observed probability was compared with the expected probability, computed by adjusting $1/T$ for bias, as is evaluated in the section on risk adjustments.

In a second method of verification, each flood record was split into samples of 10 and 20 annual peaks, as was done in the previous method, but the T-yr peak discharges computed from the samples were adjusted for bias before the exceedances were counted, and the observed probability was then compared directly with $1/T$. This method was used also in a split-sampling verification of the use of observed skew. The adjustment for bias used in this method is given in the section on risk adjustments.

Results of split sampling from 4255 station years of record at 111 stream gaging stations in the northeastern part of the United States are shown in Table 1. The area (designated 1N) covered by the table is that part of the North Atlantic slope region in New York and east. When the observed number of exceedances (column 3) is divided by the number of peaks (column 4), the resulting average exceedance probabilities (column 5) are considerably greater than the probabilities computed from the reciprocal of the recurrence interval ($1/T$) (column 2). The 569 exceedances shown for 50-yr peaks, for example, give an average exceedance probability of 3.42%, as compared with the 2.0% given by the reciprocal of the 50-yr recurrence interval. The 3.42%, however, is reasonably consistent with the 3.09% shown as the estimate of the average exceedance probability in column 6. This estimate, which is based on equations developed later in the section on risk adjustment, represents the average percent chance of exceedance that would be expected if the lines of generalized skew coefficients for the area were correct.

The apparent bias shown in column 8 of Table 1 is based on the difference between the percentages in columns 5 and 6 and is a measure of the bias in T-yr peak discharge, computed by using generalized skew coefficients taken from Figure 1. It was computed by subtracting *Harter*'s [1969] k value for the probability given in column 6 from that for the probability given in column 5, using a skew coefficient of 0.342, which is the average of the map skew coefficients for the stations used in the split sampling for this area. Thus the biases shown in columns 8 and 9 are in standardized log units.

The bias in standardized log units should be multiplied by the logarithmic standard deviation to obtain the bias in log units. If the logarithmic standard deviation (to the base 10) were 0.25, for example, the bias for 50-yr peaks shown in column 8 of Table 1 would indicate that the 50-yr peaks would average 0.25 times −0.055, or −0.014 log unit, which is 3.0% too low. To remove this apparent bias, the map skew given for this area in Figure 1 would have to be increased somewhat.

The bias shown in column 9 of Table 1 is from the results of split sampling in which the discharge of all computed T-yr peaks was increased by a predetermined factor before the exceedances were counted. The computation of this factor is given in the section on risk adjustment. Part of the inconsistency in the apparent bias shown in columns 8 and 9 can be attributed to sample error, and part is due to the difference in the way the bias is computed.

The split-sampling test for samples of size 20 in area 1N was repeated by using samples of 10 annual peaks, and then the tests for both sample sizes were rerun after the observed arrays of annual peaks had been rearranged randomly. Results for the four runs for area 1N are given in Table 2. The results given for $N = 20$ are from the last two columns of Table 1.

Split-sampling tests for 18 other areas of conterminous United States were conducted in the same manner as they were for area 1N, and the results are summarized in Table 3. Ten of the areas shown in column 1 are the same as those with the corresponding part number used in the *U.S. Geological Survey Water-Supply Papers* series, and the other nine areas are arbitrary subdivisions for the purpose of this investigation. The apparent bias was obtained as is shown in Tables 1 and 2, the averages of values for $T = 50$ and $T = 100$ shown

TABLE 1. Split-Sampling Results, Area 1N, for Samples of 20 Annual Peaks Using Map Skew

Recurrence Interval T, yr	$1/T$, %	Number of Exceedances	Number of Annual Peaks Compared	Average Exceedance Probability, %	Estimated Average Exceedance Probability, %	Difference	Apparent Bias Using Adjusted Probability*	Apparent Bias Using Adjusted Discharge*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
10	10.0	1744	16,625	10.49	11.25	−0.76	0.046	0.044
50	2.0	569	16,625	3.42	3.09	0.33	−0.055	−0.064
100	1.0	341	16,625	2.05	1.95	0.10	−0.027	−0.095

*Bias is in standardized log units.

TABLE 2. Apparent Bias in T-Yr Peaks, Area 1N, by Four Split-Sampling Runs

	By Adjusted Probability of T-yr Peaks			By Adjusted Discharge of T-yr Peaks		
Run N	T = 10	T = 50	T = 100	T = 10	T = 50	T = 100
20	0.05	−0.05	−0.03	0.04	−0.06	−0.09
10	−0.05	−0.08	−0.06	−0.03	−0.06	−0.02
20*	0.02	−0.04	−0.03	0.05	−0.03	−0.06
10*	−0.02	−0.05	−0.03	−0.01	0.03	0.03
Average†	0	−0.06	−0.04	0.01	−0.03	−0.04

Bias is in standardized log units of T-yr peaks computed by using map skew.
* Random array of annual peaks.
† The average for use in Table 3 is −0.05 by adjusted probability and −0.03 by adjusted discharge.

in the last note of Table 2 being shown for area lN in columns 7 and 8 of Table 3. Positive values of apparent bias for map skew in columns 7 and 8 of Table 3 are the average amounts in standardized log units by which the computed frequency curves would have to be lowered to make the average probability of exceedance agree with theoretical estimates ($1/T$). Negative values of bias indicate that the curves would have to be raised. For any area the bias in log units would be the standardized bias (column 8) times the logarithmic standard deviation to the base 10 (column 6). Thus for area lN the bias in log units is −0.03 times 0.222, or −0.0067, which is −2.0% in discharge. The bias in percent thus computed for each part is shown in column 9. The data on average logarithmic skew and logarithmic standard deviation in columns 3–6 show no obvious relation to this bias.

Of the 19 areas shown in Table 3, only area 10, which is centered on Nevada, has a bias in column 9 greater than 3%. The average skew coefficients for states surrounding Nevada indicate that the map skew for this area could not be lowered enough to decrease this positive bias appreciably without undue distortion in the pattern of the lines of generalized skew coefficients. The split-sample results for this area are based on only 35 stations, none of which are in Nevada. Although the small bias for some of the other areas could be reduced by making further changes in the map of skew coefficients, no revision is considered justified at this time. One reason for such a conclusion is that further refinement should consider the effect of basin characteristics such as size and slope of the drainage basin.

Columns 10 and 11 in Table 3 show the effect of adjusting T-yr peaks computed by using the observed skew by the same percentages used to adjust T-yr peaks computed by using map skew. The predominately negative bias indicates that a larger adjustment would be required to give peak discharges for which the probability of exceedance would average $1/T$. Preliminary investigation of the source of the bias, however, indicates that the need for additional adjustment is primarily in areas where the map skew is positive.

The computation of average exceedance probability by the split-sampling procedure used in this paper is not affected by

TABLE 3. Summary of Apparent Bias in T-Yr Peaks Indicated by Split-Sampling Results

						Apparent Bias in 50- and 100-yr Peaks*				
						Map Skew by Adjusted Probability,	Map Skew by Adjusted Discharge		Observed Skew by Adjusted Discharge	
Area (1)	Number of Stations (2)	Average of Observed Logarithmic Skew (3)	Average of Map Logarithmic Skew (4)	Map Skew Minus Observed Skew (5)	Average Logarithmic Standard Deviation (6)	$\Delta k/\sigma$ (7)	$\Delta k/\sigma$ (8)	% (9)	$\Delta k/\sigma$ (10)	% (11)
1N	111	0.382	0.342	−0.04	0.222	−0.05	−0.03	−2	−0.13	−10
1S	81	0.517	0.335	−0.18	0.247	−0.06	0.02	1	−0.10	−9
2	85	0.004	0.106	0.08	0.278	0.05	0.06	3	−0.05	−3
3NE	71	0.141	−0.086	−0.27	0.224	−0.03	−0.02	−1	0.01	1
3SE	70	−0.038	0.028	0.07	0.228	0.06	0.06	3	−0.06	−3
3W	23	−0.622	−0.433	0.24	0.270	0.08	0.05	3	0.03	2
4	50	−0.184	−0.250	−0.06	0.199	0.01	−0.02	−1	−0.04	−2
5E	68	−0.502	−0.451	0.05	0.280	0.03	−0.02	−1	−0.04	−3
5W	62	−0.324	−0.320	0.01	0.372	0.02	0	0	−0.01	1
6 plus	36	0.406	0.167	−0.34	0.313	−0.02	0.04	3	−0.04	−3
6 neg	78	−0.455	−0.181	0.27	0.340	0.03	0.04	3	0.02	2
7	100	−0.296	−0.307	0	0.338	0.06	0.02	1	0.01	1
8	79	−0.179	−0.190	−0.01	0.418	0.07	0.02	2	−0.02	−2
9	80	−0.224	−0.145	0.08	0.224	0.04	0	0	−0.02	−1
10	35	−0.369	−0.117	0.26	0.278	0.12	0.07	5	−0.06	−4
11	81	−0.001	−0.103	0.11	0.426	0.02	−0.01	0	−0.04	−1
12	98	−0.053	−0.084	−0.03	0.183	0.02	0.01	0	0.01	1
13	41	−0.154	−0.233	−0.08	0.200	0.07	−0.03	−1	−0.04	−2
14	98	0.134	0.026	−0.11	0.211	−0.01	−0.06	−3	−0.07	−3

The numerical part of the area number is from the *U.S. Geological Survey Water-Supply Papers;* letters are compass direction for the breakdown of the part except that part 6 is divided into an area of positive skew and an area of negative skew.
* The bias shown as $\Delta k/\sigma$ is in standardized log units.

the possibility of cross correlation of annual peaks at some of the stations or by the fact that the replications give results that are not independent. If for a given area, for example, one highly correlated station were to be added for each station that was used, the number of exceedances would be doubled, but since the number of annual peaks would also be doubled, the exceedance probability would be exactly as it is computed. Similarly, if for a given station the replications for samples of size 20 were to be doubled by starting only 2 yr later each time instead of 4 yr later, the number of annual peaks would be doubled, and the number of exceedances would be approximately doubled. The exceedance probability computed by using the 2-yr increments would have slightly less sampling error than would that computed by using the 4-yr increment, but the improvement would not be worth doubling the computation time. Results by either increment, however, would be unbiased.

RISK ADJUSTMENTS

Because of sampling errors the average probability of exceedance of T-yr peaks computed from a finite data sample by a log Pearson type 3 analysis will be greater than $1/T$. For example, *Hardison and Jennings* [1972] showed that 50-yr peaks computed by log Pearson type 3 procedure from 10 annual peaks will be exceeded in 4% of the years on the average even if there is no sample error in the skew coefficient used. They also proposed that a discharge-probability relation could be defined so that the T-yr peak computed from a finite sample does have an average exceedance probability of $1/T$, and they suggested that such a relation be called a flood-risk curve to distinguish it from the flood-frequency curve defined by the usual log Pearson type 3 analysis. Such a flood-risk curve can be defined from a computed flood-frequency curve either by adjusting the probability of exceedance or by adjusting the flood discharge. Both types of adjustment were computed and used in the split-sampling tests. Average exceedance probability as used here is sometimes called 'expected probability,' and the difference between it and $1/T$ is sometimes called 'small-sample bias.'

The differences between flood-frequency curves and flood-risk curves defined by data in the Hardison-Jennings paper are analyzed in the following two sections to define the equations that were used to adjust either the exceedance probabilities or the discharges in the section on split-sampling verification. An analysis of difference in probability is followed by an analysis of difference in discharge. Readers who are not interested in statistical details should skip to the section on use of map skew.

Adjustment to exceedance probability. The estimated average exceedance probability (column 6, Table 1) used in the split-sample testing was determined by adjustment of the probability scale of flood-frequency relations defined from sample mean, sample variance, and the true skew coefficient of the population. The adjustments that are needed to remove bias introduced by error in estimated skew coefficients were computed by the equation

$$D = D_K(R)^{2.15+0.3(\langle g'\rangle)} \qquad (3)$$

in which D is the adjustment $\langle P\rangle - 1/T$ needed to correct the biased value of the average number of exceedances computed from the sample mean, the sample variance, and the map skew to the desired unbiased value $1/T$; D_K is a similar adjustment value defined by Hardison and Jennings to correct the biased value of the average number of exceedances computed from the sample mean, the sample variance, and the population skew (known skew) to the desired unbiased value $1/T$ (equations that give values of D_K comparable to those given in the Hardison-Jennings paper are available from the author); and $(R)^{2.15+0.3(\langle g'\rangle)}$ is a term to adjust for the additional bias introduced by use of the map skew rather than population skew, where R is the ratio by which the standard error of T-yr peaks computed by using the known skew should be multiplied to obtain the standard error of T-yr peaks computed by using the map skew, $\langle g'\rangle$ is the average of the map skew for a region, and the constants 2.15 and 0.3 were defined by log-log plots of $\langle P\rangle - 1/T$ from Table 2 of the Hardison-Jennings paper against S_{x_T}/σ from Table 3 of that paper. Values for R in (3) were computed as

$$R = 1 + \left[\left(\frac{S_{g'}}{0.3}\right)^{1.87}\right] \cdot [0.785 + 0.106(\log N) - 0.054g' + 0.10(\log T) - 1.0] \qquad (4)$$

in which $S_{g'}$ is the standard error of the map skew coefficients, N is the number of annual peaks used in the split samples, and T is the recurrence interval in years. The equation is not applicable when $S_{g'}$ is larger than 0.4, N is outside the range of 10–25, g' is outside the range -0.5 to 0.5, or T is greater than 100.

Values of D obtained by (3) were increased by $1/T$ and multiplied by 100 to obtain the values given in column 6 of Table 1. In computing the 3.09% shown for $T = 50$ in column 6 of Table 1, for example, a D_K of 0.93 interpolated from Table 2 of the Hardison-Jennings paper for a skew of 0.34 and an N of 20 was increased to 1.09 by multiplying it by an R of 1.074 raised to the power of 2.25. The R of 1.074 was obtained from (4) by using an $S_{g'}$ of 0.3, which is the estimated standard error of map skew east of the Mississippi River.

Equation (4) was defined by data obtained from the standard deviation of T-yr peaks computed from random samples. The samples were drawn from populations selected randomly from a family of populations that has skew coefficients that are normally distributed with a given mean and standard deviation. On the basis of the accuracy appraisal of map skews, $S_{g'}$ in (4) was used as 0.3 east of the Mississippi River and 0.4 west of there.

Equation (4) can be used also to estimate the standard error of T-yr peaks computed by using skew coefficients from Figure 1. The standard error would be obtained by multiplying R from (4) by the standard error for known skew given by Table 4 of the Hardison-Jennings paper.

Adjustment to T-yr peak discharge. The discharge of flood-frequency relations computed from sample statistics by the log Pearson type 3 procedure has to be increased to obtain a discharge for which the risk is $1/T$. Discharge from the adjusted relation is called flood risk in this paper.

For T-yr peaks computed by using the known skew coefficient of the population in the log Pearson type 3 computations, the adjustment to T-yr peaks in standardized log (to the base 10) units can be estimated by

$$\log Q_{\langle P\rangle} - \log Q_T = \log (Q_{\langle P\rangle}/Q_T) = s[a + b(g' + 0.5)^{1.62}]/N^{1.2} \qquad (5)$$

in which g' is the generalized skew coefficient, $Q_{\langle P\rangle}$ is the flood risk discharge, Q_T is the T-yr peak discharge, N is the number of annual peaks, and s is the logarithmic standard deviation and in which a and b vary with T as follows:

T	a	b
10	2.5	0.5
20	4.0	1.0
50	6.6	2.0
100	8.8	3.4
500	14.8	6.8

A solution of (5) for a record length of $N = 25$ yr is given in Figure 3. The ordinate scale on the left-hand side of the figure gives the adjustment in standardized log units, $\sigma = 1.0$, and that on the right-hand side gives the ratio by which the computed T-yr peak discharge should be multiplied if the logarithmic standard deviation of the annual peaks is 0.3. Ratios for any other value of the logarithmic standard deviation can be computed by multiplying it by values from the standardized log unit scale and taking the antilog. For sample sizes other than 25 the adjustment in standardized log units given by Figure 3 should be multiplied by $(25/N)^{1.2}$, which gives 3.0 for $N = 10$ and 0.44 for $N = 50$.

Factors $Q_{(P)}/Q_T$, computed from (5), were used to adjust flood frequency discharge Q_T and to obtain the flood risk discharge $Q_{(P)}$, used in computing the bias by adjusted discharge shown in Tables 1, 2, and 3. For a station with a logarithmic standard deviation of 0.3 and a map skew of 0.45, for example, 50-yr peaks computed from samples of size 20 were multiplied by 1.20 before the exceedances were counted.

Equation (5) was defined by data for known skew given in Table 2 of the Hardison-Jennings paper augmented by similarly computed data for T values of 2000 and 10,000 yr. Since data were not available to define similar equations for the case in which skew coefficients have a standard error, (5) was used to adjust T-yr discharge in the split-sample tests for map skew from which the bias shown in columns 8 and 9 of Table 3 was computed. Although such use of this equation is valid only when there is no error in the skew coefficient, the results of the split sampling shown in Table 3 indicate that it is suitable for use with the map skews shown in Figure 1 despite the estimated standard errors at 0.3 and 0.4. Any increase in the adjustment given by (5) would give smaller average exceedance probabilities in the split-sample result and would thus increase the $\Delta k/\sigma$ values in column 8 of Table 3, which already average 0.01 too high. No logical explanation can be given for why an adjustment for error in map skew that was necessary when a risk adjustment was applied to exceedance probabilities (column 7) does not seem to be necessary when a risk adjustment is applied to T-yr peak discharge (column 8).

In the split-sampling tests using observed skew, T-yr peak discharge given by the frequency curves was adjusted by (5) before the exceedances were counted. The predominately negative bias shown for these results in columns 10 and 11 of Table 3 indicates that when observed skew is used, the adjustments given by (5) have to be increased to give discharges for which the probability of exceedance averages $1/T$.

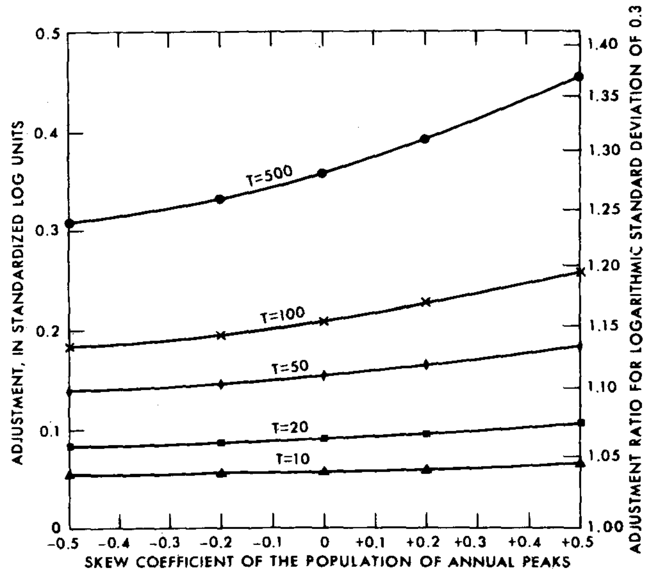

Fig. 3. Risk adjustment to T-yr peak discharge computed from 25 annual peaks by using known logarithmic skew coefficients in log Pearson type 3 computation. The adjustment ratio is the antilog of the product of the standardized log units times the logarithmic standard deviation.

ACCURACY COMPARISON

By using a procedure for separating time-sampling and space-sampling variance [*Hardison*, 1971], the standard error of map skew east of the Mississippi River was found to be 0.3, which is about as accurate as skew coefficients computed from 60 annual peaks. West of the Mississippi River the standard error of map skew was found to be 0.4, which is about as accurate as skew coefficients computed from 35 annual peaks. Because of cross correlation between the statistics of samples taken from skewed distributions, however, this appraisal does not necessarily carry through to the accuracy of T-yr peaks computed by the log Pearson type 3 procedure. Since the effect of cross correlation cannot be evaluated theoretically for the case in which sample skew is used, the effect on T-yr peaks was evaluated empirically by using random number experiments. The results of these experiments show how the standard errors of T-yr peaks computed by using map skew compare with those of peaks computed by using sample skew, as summarized in Table 4. Applying these results to the map skew shown in Figure 1 indicates that when less than about 30 yr of records are available, the standard error of 50- and 100-yr peaks computed by using map skew is less than that for peaks computed by using observed skew except in Missouri. If the standard error of the map skew in Missouri were 0.3, however, as it is east of the Mississippi River, there would be no exception. In areas where map skew is 0.2 or greater the standard errors of 50-, 100-, and 500-yr peaks are less than those of peaks computed by using a skew coefficient based on 60 annual peaks. For small recurrence intervals such as 10 and 20 yr, observed skews frequently give T-yr peaks that are more accurate than those computed by using map skews, especially in areas where the map skew is negative.

The standard error of the T-yr peaks used to define each ratio shown in Table 4 was computed from the statistics of 1000 samples of the indicated number of annual peaks. To simulate natural conditions, for which the true skew coefficient is never known but can be assumed to vary randomly about a given value of map skew, each sample was drawn from a population selected randomly from a family of Pearson type 3 populations. Each family of populations had a mean skew coefficient equal to an assumed map skew, and the skew coefficients of the populations were assumed to be normally distributed about this mean with a standard deviation equal to an assumed standard error in map skew of 0.3 or 0.4. The means and the standard deviations of the given skew

TABLE 4. Ratio of Standard Errors of T-Yr Peaks Computed by Using Map Skew to That Computed by Using Sample Skew

		For $S_{g'} = 0.3$				For $S_{g'} = 0.4$			
g	N	$T = 25$	$T = 50$	$T = 100$	$T = 500$	$T = 25$	$T = 50$	$T = 100$	$3T = 500$
−0.5	10	1.015	0.950	0.881	0.754	1.026	0.968	0.909	0.798
−0.5	25	1.012	0.965	0.920	0.840	1.045	1.021	0.996	0.948
−0.5	50	1.092	1.080	1.062	1.023	1.179	1.205	1.213	1.210
−0.3	10	0.983	0.920	0.857	0.740	0.992	0.937	0.882	0.782
−0.3	25	0.976	0.930	0.888	0.816	1.003	0.977	0.953	0.912
−0.3	50	1.043	1.034	1.021	0.992	1.116	1.142	1.154	1.161
0	10	0.951	0.890	0.833	0.728	0.960	0.905	0.854	0.763
0	25	0.938	0.889	0.847	0.778	0.960	0.925	0.898	0.856
0	50	0.983	0.964	0.948	0.921	1.037	1.048	1.055	1.062
+0.3	10	0.938	0.878	0.824	0.727	0.945	0.890	0.842	0.757
+0.3	25	0.920	0.866	0.821	0.748	0.935	0.892	0.860	0.811
+0.3	50	0.944	0.911	0.885	0.847	0.982	0.974	0.968	0.961
+0.5	10	0.936	0.879	0.827	0.734	0.942	0.889	0.842	0.760
+0.5	25	0.916	0.859	0.812	0.737	0.928	0.881	0.845	0.790
+0.5	50	0.930	0.888	0.855	0.807	0.960	0.938	0.924	0.904

Here g' is map skew, $S_{g'}$ is the standard error of map skew, N is the number of annual peaks in the sample, and T is the recurrence interval in years.

coefficients were selected to encompass the range covered by the map skews presented in this paper. The populations of annual peaks from which the samples were drawn all had a mean of zero and a standard deviation of unity. For a map skew of 0.3 with a standard error of 0.3, for example, a random selection of population skew might give 0.2, which would be used as the true skew, and the skew of the sample drawn from such a population might be 0.6. The error in a 50-yr peak computed by using the sample skew with sample mean and sample standard deviation was computed as the difference between that peak and the true 50-yr peak given by the population curve with a skew coefficient of 0.2. The standard error in a 50-yr peak using map skew was computed as the difference between the true 50-yr peak thus obtained and the 50-yr peak computed by using the sample mean, the sample standard deviation, and the map skew coefficient of 0.3. The next sample would have a different true 50-yr peak with which to compare the 50-yr peak computed by using sample skew and that computed by using the map skew of 0.3.

USE OF MAP SKEW

The map skew shown in Figure 1 is recommended for use in computing T-yr peaks from the usual length record of annual peaks not only because the standard error shown by Table 4 is generally smaller but also because the use of map skew has been found to minimize the seriousness of other problems associated with flood frequency analysis. Other investigations being made by the writer indicate that the use of map skew minimizes the need for considering historic peaks, for removing outliers at the low end of the frequency distribution, or for extending short records of annual peaks by correlation with longer records except at stations close together on the same stream. Although these indications are based on experiments with samples of random numbers, the results are consistent with what might logically be expected in using observed records. In the case of low outliers, for example, it is the effect on the observed skew coefficient that distorts the upper end of the frequency curve. Similarly, it is the unreliability of skew coefficients based on small samples that provides the main justification for extending a short record. Any investigation into these and other problems associated with flood frequency analysis should at least consider the use of map skew such as that given in Figure 1.

The map skew shown in Figure 1 can be considered to be a compromise between those who prefer the nationwide use of a log normal or some other type of two-parameter distribution and those who prefer the use of a three-parameter distribution in which the third parameter is estimated from the statistics of the sample. It is obvious from Figure 1, for example, that the use of a log normal distribution along the east coast, where the average skew coefficient is definitely greater than zero, would be inconsistent with flood experience at gaging stations in the area. The use of map skew taken from lines of equal skew eliminates the problems associated with generalizing the skew coefficients for various selected regions. The question of how large a region to use and of what to do with a station near the boundary between two regions does not occur when map skew is used.

Another possible use of map skew is in investigations dealing with how differences in basin characteristics affect skew coefficients at individual stations. Attempts to generalize skew coefficients by relating observed skew to basin characteristics over wide areas have had little success, possibly because the combined effect of time-sampling error and areal differences in skew coefficients tends to mask any significant relation. If the areal differences are removed by use of map skew before further generalization is attempted, the chance of finding a significant index to explain part of the differences should be improved. At least the investigation could cover a wider geographic area without the problem that areal differences might become the predominant source of observed differences.

For records of annual peaks of a length such that the standard error of T-yr peaks computed by using observed skew is practically equal to that of T-yr peaks computed by using map skew, an average of the two estimates may provide the best estimate of the T-yr peak. Similar averaging might be justified even when the standard errors of the two estimates are somewhat different, but in that case the two estimates should be weighted inversely in proportion to the square of their standard errors. If, for example, the standard error of a 50-yr peak of 10,000 ft^3/s computed by using map skew is shown by Table 4 to have a standard error only 90% of that of a 50-yr peak of 12,000 ft^3/s computed by using observed skew, the 10,000-ft^3/s figure should be given a weight of $(1/0.9)^2$, or 1.24, and the 12,000-ft^3/s figure should be given a weight of 1.00 to give a weighted average of 10,900 ft^3/s.

It is important to point out that if the estimated peaks are to be related to basin characteristics by regression analysis as a means of information transfer to ungaged sites, the T-yr peaks given by the flood-frequency curve should be used rather than the corresponding flood risk discharge. The application of a risk adjustment to T-yr peaks computed by the log Pearson type 3 procedure prior to the regression analysis would tend to destroy the normality about the regression line that is required if the results are to be analyzed statistically. Also, such adjustment would increase the discharge of T-yr peaks based on short records more than those based on long records and would thus tend to increase the standard error of estimate of the regression.

CONCLUSIONS

The lines of generalized logarithmic skew coefficient developed in this paper are based on the average of skew coefficients computed from all suitable stream gaging records in each state. Split-sampling tests using the logarithmic mean and the standard deviation of small samples show that the use of the map values of skew coefficients in log Pearson type 3 computations gives T-yr peaks that have actual probabilities of exceedance in keeping with those to be expected. West of the Mississippi River the accuracy of 50- and 100-yr peaks computed by using map skew is better than that of 50- and 100-yr peaks computed by using observed skew when less than 30 annual peaks are available. East of the Mississippi River the corresponding accuracy is even higher and approaches the equivalent of 60 yr of annual peaks along the east coast.

In addition to giving more accurate 50- and 100-yr peaks, the use of map skew in place of observed skew tends to minimize the need for considering historic peaks, for extending short records, and for removing outliers at the low end of frequency curves. Furthermore, such use is a partial concession to those who prefer the use of a log normal distribution nationwide.

The map skew coefficients given by Figure 1 are sufficiently accurate to use in computing T-yr peaks to be used in regressions or other procedures for generalizing T-yr peaks. Such generalization, however, should include regression of the difference between observed skew and map skew so that the best estimate of a generalized skew coefficient will be available when the need arises for a hydrologic design at or near a gaged site. The discharge of T-yr peaks used in generalization studies should be the discharge from a flood-frequency curve; adjustment for the difference between flood frequency and flood risk should be withheld until a discharge is to be used for a specific project. The necessity of making such adjustment, however, should be mentioned in flood frequency reports. The decision concerning how much weight, if any, should be given to the T-yr peak computed by using the skew coefficient computed from the annual peaks observed at a site or to T-yr peaks estimated in other ways can also be deferred until the design stage or until several agencies are to agree on a value for a T-yr peak at a specific site.

NOTATION

- a coefficient that varies with T in (5) for use in computing $Q\quad/Q_T$.
- b coefficient similar to a.
- D amount by which the average probability of exceedance of a T-yr peak is greater than $1/T$.
- D_K D for known skew, amount by which $\langle P \rangle$ would exceed $1/T$ if there were no error in the estimated skew coefficient used to compute Q_T.
- g^* observed skew, logarithmic skew coefficient of a record of annual peaks computed from the statistics of the sample.
- g' map skew, generalized logarithmic skew coefficient given by Figure 1.
- k distance above the mean in standard deviation units for a given skew coefficient, i.e., standardized units.
- N number of annual peaks used in computing the statistic of the sample.
- $\langle P \rangle$ average probability of exceedance of a T-yr peak.
- Q_T T-yr peak discharge in cubic feet per second, equals antilog of $\langle x \rangle + ks$.
- $Q_{\langle P \rangle}$ flood risk discharge for the average exceedance probability $\langle P \rangle = 1/T$.
- R ratio of the standard error of T-yr peaks computed by using map skew to the standard error of T-yr peaks computed by using known skew.
- s observed logarithmic standard deviation of a record of annual peaks computed from the statistics of the sample.
- $S_{g'}$ standard error of estimate of g'.
- x_i logarithm to the base 10 of the ith peak in a record of annual peaks.
- $\langle x \rangle$ mean of x_i.
- Δk increment of k between the exceedance probabilities.
- σ logarithmic standard deviation of a population of annual peaks of which s is sample statistic.

REFERENCES

Hardison, C. H., Prediction error of regression estimates of streamflow characteristics at ungaged sites, Geological Survey Research 1971, *U.S. Geol. Surv. Prof. Pap. 750-C*, C228–C236, 1971.

Hardison, C. H., and M. E. Jennings, Bias in computed flood risk, *J. Hydraul. Div. Amer. Soc. Civil Eng.*, *98*(HY3), 415–427, 1972.

Harter, H. L., New tables of percentage points of the Pearson type 3 distribution, *Technometrics*, *11*(1), 177–187, 1969.

Hazen, A., *Flood Flows; A Study of Frequencies and Magnitudes*, p. 46, John Wiley, New York, 1930.

Water Resources Council, Hydrology Committee, A uniform technique for determining flood flow frequencies, *Bull. 15*, 15 pp., Washington, D. C., 1967.

(Received February 14, 1974;
accepted April 19, 1974.)