REGIONAL SKEW COEFFICIENTS FOR FLOOD-FREQUENCY ANALYSIS OF MISSISSIPPI STREAMS

Mark N. Landers
U. S. Geological Survey, Mississippi District

INTRODUCTION

Analyses of annual peak-flow records provide the empirical basis for estimating flood frequency. Statistical methods of analysis are well suited to the random nature of annual flooding. Statistical methods may be used to estimate flood frequency from a sample of recorded annual peak flows using the assumption that the recorded sample represents the population of all the recorded and unrecorded annual peak flows at that stream site. The Pearson type III distribution has been recommended as the appropriate probability distribution for log-transformed annual peak-flow data by the Interagency Advisory Committee on Water Data (IACWD, 1982). The Pearson type III distribution requires estimates of the population mean, variance, and skew. The population parameters are estimated by computing the sample moments which correspond to these population moments as follows:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{1}$$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})^2 \tag{2}$$

$$G_s = \frac{N}{(N-1)(N-2)S^3}\sum_{i=1}^{N}(X_i - \overline{X})^3 \tag{3}$$

where $\overline{X}$ is the sample mean,
S^2 is the sample variance,
Gs is the sample skew,
Xi is the log-transformed annual peak flow, and
N is the sample size, that is the number of year I peak-flow record for the stream site being analyzed.

Previous studies of the sampling distribution of sample skew (Gs) have shown that Gs is a biased estimator of the population skew and is subject to large sampling variances as compared with X and S^2. Empirical bias-correction factors were computed by Wallis and others (1974) based on Monte Carlo experiments. Tasker and Stedinger (1986) describe a bias-correction equation which is based on record length and defined as:

$$C_b = (1 + 6/N) \tag{4}$$

where Cb is the bias-correction coefficient. Tasker and Stedinger (1986) showed only minor differences between bias-correction coefficients from this equation and from the empirical results of Wallis and others (1974), when N > 20 and |Gs| < 1.0 . The magnitude of error of estimated population skew may be reduced using the weighted average of Gs and an independent regional skew estimate for that site, as recommended by the IACWD (1982). The sample skew is weighted inversely to its mean-square error (MSEs), and regional skew (Gr) is weighted inversely to an estimate of its sampling variance. The IACWD (1982) uses mean square error (MSEr) as an estimate of the sampling variance of regional skew. Population skew then is estimated by:

$$G = \frac{(MSEr * Gs) + (MSEs * Gr)}{MSEr + MSEs} \tag{5}$$

where G is an estimate of the population skew coefficient, and Gr is assumed to be unbiased so that MSEr is equal to the sampling variance of Gr.

Population skew estimates are improved when computed from the weighted average of the sample and unbiased regional skew estimates. Further improvements in estimated population skew are obtained by using weighted methods to estimate regional skew.

Regional skew coefficients typically are estimated from the sample skews of long-term annual peak-flow record stations in the study area using regression, mapping, or simple averaging methods. The IACWD (1982) provides a skew isoline map of the United States for regional skew estimates; however, because of the limited accuracy of the map and subsequent improvements in estimating methods, the Committee suggests that separate regional skew analyses may be appropriate. Ordinary mean, isoline mapping, and least-squares regression methods assume the sampling distribution has uniform variance; that is, skew coefficients computed from log-transformed annual peak-flow records of different gaging stations all are assumed to have equal accuracy. However, previous studies have shown the sample variance of skew to vary with record length. Therefore, methods that include a weighting function to account for nonuniform sampling variance estimate regional skew with greater accuracy. Tasker and Stedinger (1986) used weighted least-squares regression procedures to estimate regional skew coefficients and showed improved results over ordinary least squares. The weighted mapping procedure used in this report positions isolines according to weighted grid-node values.

WEIGHTING FUNCTIONS

The sampling variance of skew varies with the length of record and must be estimated to define a weighting function. Because the exact form of the probability distribution of sample skew (Gs) is unknown, estimates of the sampling variance require either an assumed distribution for Gs or a nonparametric approach. Sampling variances of Gs assuming a log-Pearson distribution, and based on Monte Carlo experiments, may be obtained from Wallis and others (1974). Where a normal distribution is assumed for Gs, several parametric equations have been used to estimate the sample variance of skew coefficients. The weighted mapping procedure in this investigation uses the parametric method of Fisher (1931) corrected for bias and defined as:

$$Vs = \frac{6N(N-1)[1+(6/N)]^2}{(N-2)(N+1)(N+3)} \qquad (6)$$

This equation was used by Tasker and Stedinger (1986). Comparisons of skew sample variance estimating methods by Tung and Mays (1981) indicate that nonparametric approaches provide greater accuracy than parametric ones; however, for this investigation, the improvements shown over Fisher's method did not warrant the greater computational requirements of nonparametric procedures. Sample skew is weighted inversely to its estimated sample variance, so the weighting function is defined by:

$$W = \frac{1}{Vs} \qquad (7)$$

where W is the weight given to Gs.

This weighting function was applied to mapping methods used to estimate regional skew coefficients for Mississippi streams.

WEIGHTED MAPPING OF SKEW

The spatial variability of skew suggests description by regionalization and isoline mapping. Automatic mapping techniques have been developed to eliminate the subjectivity of hand-drawn isoline maps. Automatic mapping generally is simplified by initially gridding the study area. Gridding consists of estimating the value of the study variable at each node of a regular grid over the study area. Isolines are drawn based on the grid-node values by a cubic spline or similar fitting process. Grid-node values may be estimated by a two-step procedure. First a spatial search is made to select the subset of sample data points to be used in estimating each node. Various search procedures may be used; selection of the nearest n data points for each grid node is the simplest. Second, the grid-node estimate is computed from the selected data point values by a distance-weighted mean, where the weights are a function of distance from the grid node and uniform sampling variance is assumed. Nonuniform variance of sample data point values may be accounted for by using a weighting function in the grid-node estimator. In this analysis, grid-node estimates were weighted for error of sample skew and sample point distance from the grid node by the following equation:

$$Zi = \frac{\sum_{j=1}^{n_i} Gs_j(W_j)(1/d_j)}{\sum_{j=1}^{n_i} (W_j)(1/d_j)} \qquad (8)$$

where Zi is the estimated skew at grid node i,
ni is the number of sample points selected to estimate Zi,
dj is the distance from the grid node to the centroid of the basin whose records define Gsj, and
wj is the weight given to Gs at station j.

Weighting for sampling error increases the accuracy of the isoline map by eliminating the assumption of uniform sampling error. This weighted mapping method does assume sample skews to be independent. The use of weighted grid mapping methods to estimate regional skew coefficients for Mississippi streams are discussed in the following section.

DISCUSSION

Regional skew coefficients (table 1) for flood-frequency analysis of Mississippi streams were determined using sample skews from 171 stream-gaging stations located in Mississippi, Arkansas, Tennessee, Alabama, and Louisiana (fig. 1). The sample skews were computed by equation (3) and corrected for bias using equation (4) from systematic record periods. The systematic record periods for the data set average more than 30 years, and there are more than 10 years for every gaging station in the data set. Skew characteristics were tested for heterogeneity between distinct regions and particularly between large basins. Boxplots of unbiased sample skew coefficients and boundaries of the three homogenous skew coefficient regions which were defined are shown in figure 2. The null hypothesis that the mean of the subregion is equal to the whole sample mean was rejected for regions 1 and 2 using the two-sample t-test. The computed p-values of 0.0001 and 0.011, for regions 1 and 2, respectively, indicate that the mean skew values for these regions are statistically different from the statewide average skew at about the 1 percent confidence level.

Table 1. -- Gaging stations used in regional skew coefficient analysis

Map No.	Station Number	Unbias. Skew	Region	Map No.	Station Number	Unbias. Skew	Region	Map No.	Station Number	Unbias. Skew	Region
1	2430000	0.166	3	58	2479000	0.253	1	115	7077920	-0.959	2
2	2430500	.534	3	59	2479165	.139	1	116	7077940	-1.366	2
3	2431000	-.134	3	60	2479180	.204	1	117	7077950	-1.424	2
4	2432900	.271	3	61	2479190	.530	1	118	7078000	-.750	2
5	2433000	.014	3	62	2479300	.658	1	119	7078170	-1.841	2
6	2433500	.080	3	63	2479500	.708	1	120	7263860	-1.523	2
7	2434000	.202	3	64	2480150	-.246	1	121	7264000	.092	2
8	2434500	.040	3	65	2480500	.826	1	122	7264100	-1.233	2
9	2435300	.547	3	66	2481130	.826	1	123	7266000	-.298	3
10	2435400	.165	3	67	2481400	1.273	1	124	7268000	-.250	3
11	2435500	.390	3	68	2481450	1.732	1	125	7269990	-.398	3
12	2435800	.235	3	69	2482000	-.429	3	126	7271000	-.250	3
13	2435920	-.573	3	70	2482100	-.367	3	127	7275000	.105	3
14	2435930	.991	3	71	2482310	-.105	3	128	7275500	.033	3
15	2436500	.852	3	72	2482500	-.340	3	129	7282000	-.614	3
16	2437000	.355	3	73	2483890	-.442	3	130	7283490	-.511	3
17	2437300	-.023	3	74	2484000	.113	3	131	7285700	.076	3
18	2437500	.187	3	75	2484500	-.003	3	132	7286000	-.748	2
19	2437550	.038	3	76	2484750	-.136	3	133	7286047	.757	2
20	2437600	.050	3	77	2485380	-.234	3	134	7286520	-.198	3
21	2439800	-.316	3	78	2485392	-.300	3	135	7287165	1.219	2
22	2439980	.753	3	79	2485900	.885	3	136	7287170	.590	2
23	2440400	.488	3	80	2486000	-.450	3	137	7287480	-.648	3
24	2440600	.459	3	81	2486690	-.295	3	138	7288500	-.138	2
25	2440800	-.396	3	82	2487300	.362	3	139	7288570	.456	2
26	2441000	-.926	3	83	2487500	.263	3	140	7288650	-1.067	2
27	2441220	-.455	3	84	2487620	.459	3	141	7288770	-1.083	2
28	2441300	-.246	3	85	2487670	-.242	3	142	7289350	.021	3
29	2441500	.256	3	86	2487710	-.006	3	143	7289530	.835	3
30	2443000	.099	3	87	2487770	-.145	3	144	7289600	-.508	3
31	2443700	.046	3	88	2488340	-.037	3	145	7289641	-1.349	3
32	2444000	.332	3	89	2488500	.105	3	146	7290000	.056	3
33	2447500	-.035	3	90	2488510	.377	3	147	7290005	.424	3
34	2447800	.332	3	91	2488680	.342	3	148	7290525	.289	3
35	2448000	.156	3	92	2488700	.069	3	149	7290650	.255	3
36	2467500	.702	3	93	2489000	.404	3	150	7290690	-.045	3
37	2471100	-.166	1	94	2489030	.235	3	151	7290870	-.393	3
38	2471500	-.103	1	95	2489160	.339	3	152	7291000	-.634	3
39	2472000	.339	1	96	2490000	-.816	3	153	7291250	.498	3
40	2472500	.829	1	97	2490105	.154	3	154	7291260	-.337	3
41	2473000	.502	1	98	2490500	-.515	3	155	7294400	.713	3
42	2473480	-.084	1	99	2490550	.425	3	156	7295000	-.990	3
43	2473500	-.198	1	100	2491500	-.399	3	157	7364120	-.771	2
44	2473850	-.640	1	101	2492360	-.027	3	158	7364150	-.597	2
45	2474500	.405	1	102	3592800	.001	3	159	7364190	-.170	2
46	2474740	.699	1	103	3593010	-.481	3	160	7367740	-.496	2
47	2475000	.384	1	104	7029270	.111	3	161	7367800	.799	2
48	2475050	-.007	1	105	7029300	.109	3	162	7369250	1.140	2
49	2475220	.602	1	106	7029400	.138	3	163	7369500	-.640	2
50	2475500	.105	1	107	7030500	-.730	3	164	7369700	-1.722	2
51	2476500	.085	1	108	7047200	-.451	2	165	7370000	.017	2
52	2477000	-.059	1	109	7047600	-.078	2	166	7373500	-.418	3
53	2477050	.298	1	110	7047924	.255	2	167	7373550	-.197	3
54	2477090	.638	1	111	7047942	-.791	2	168	7375800	.152	3
55	2477350	.307	1	112	7077500	.196	2	169	7376760	.001	3
56	2477500	.497	1	113	7077700	-.294	2	170	7377000	-.499	3
57	2478500	.696	1	114	7077860	-1.371	2	171	7377400	-.115	3

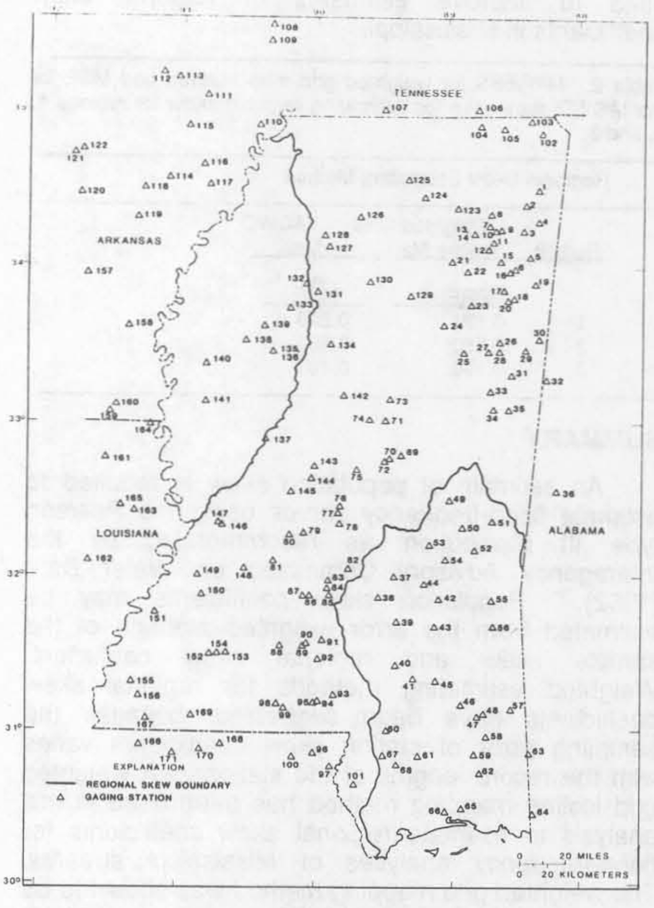

Figure 1.--Location of gaging stations used to estimate regional skew and to define regions of homogenous skew.

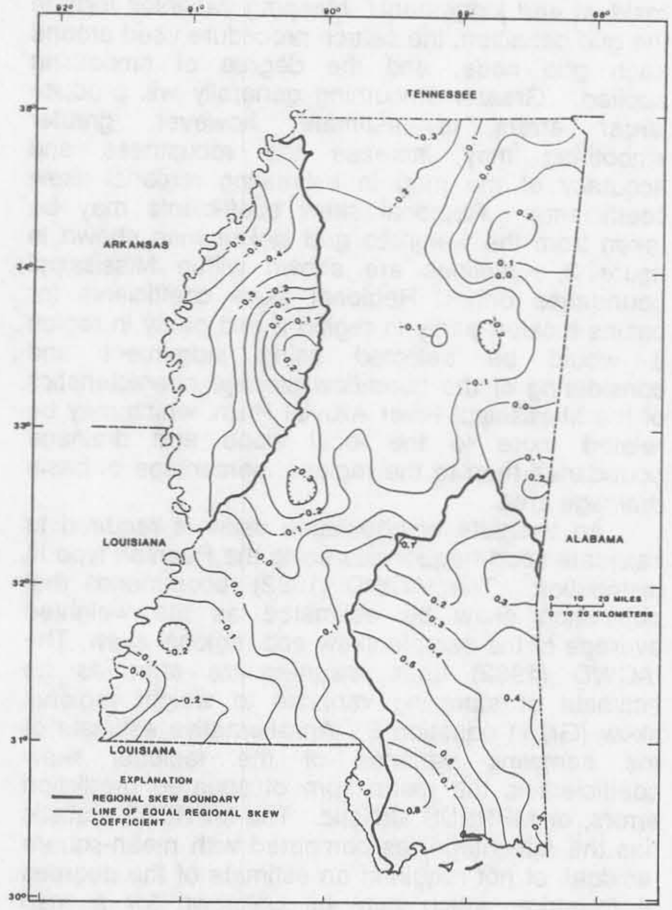

Figure 3.--Weighted grid isolines of unbiased regional skew of log-transformed annual peak flow.

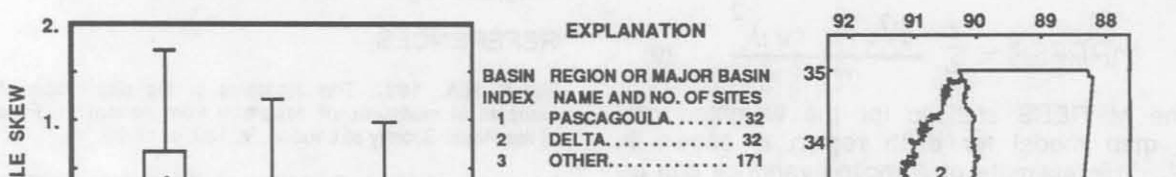

Figure 2.- - Skew coefficient characteristics and boundaries of homogenous skew regions in Mississippi.

The best weighted grid isoline map for each region was selected based on least mean-square residual and judgement. Mapping variables include the grid definition, the search procedure used around each grid node, and the degree of smoothing applied. Greater smoothing generally will produce larger errors of estimate; however, greater smoothing may increase the robustness and accuracy of the map in estimating regional skew coefficients. Regional skew coefficients may be taken from the weighted grid isoline map shown in figure 3. Isolines are shown within Mississippi boundaries only. Regional skew coefficients for basins located partly in region 2 and partly in region 3 would be selected using judgement and considering of the flood-flow storage characteristics of the Mississippi River Alluvial Plain, which may be related more to the local slope and drainage boundaries than to the regional percentage of basin drainage area.

An estimate of population skew is required to calculate flood-frequencies using the Pearson type III distribution. The IACWD (1982) recommends that population skew be estimated as the weighted average of the sample skew and regional skew. The IACWD (1982) uses mean-square error as an estimate of sampling variance to weight regional skew (Gr) in equation 5. An alternative estimate of the sampling variance of the regional skew coefficient is the mean sum of squared prediction errors, or MPRESS statistic. The MPRESS statistic has the advantage, as compared with mean-square residual, of not requiring an estimate of the degrees of freedom, which may be unknown for a map estimator. The MPRESS statistic is calculated by splitting the original data set of m points into two sets: a calibration set of size m-1, and a validation set of size one. The estimator then is computed from the calibration set and used to determine the estimated value ($\hat{Y}vi$) for the validation point (Yvi). The predictive error is computed by ($\hat{Y}vi$ - Yvi). This is done for each observation in the original sample so the mean sum of squared prediction errors is simply:

$$MPRESS = \sum_{i=1}^{m} \frac{(\hat{Y}v_i - Yv_i)^2}{m} \qquad (9)$$

The MPRESS statistic for the weighted grid isoline map model for each region is shown in table 2. This estimate of sampling variance can be computed for any of the skew estimators and included in equation 5 to estimate population skew when computing flood-frequency using the Pearson type III distribution. The estimated mean-square error ($\hat{MSE}$) of the IACWD skew map (1982) was determined (assuming m-2 degrees of freedom) for the stations in each region and also is shown in table 2. The $\hat{MSE}$ of skew estimates determined from the IACWD map is larger than the MPRESS of estimates determined from the weighted map method in regions 1 and 2 and is only slightly smaller than the MPRESS of the weighted map method in region 3. Thus the weighted map method can generally be used to improve estimates of regional skew coefficients in Mississippi.

Table 2. MPRESS for weighted grid map method and MSE for the IACWD skew map for estimating regional skew for regions 1, 2, and 3.

Regional Skew Estimating Method		
Region	Weighted Grid Isoline Map	IACWD Map
	MPRESS	$\hat{MSE}$
1	0.191	0.275
2	0.550	0.801
3	0.196	0.187

SUMMARY

An estimate of population skew is required to estimate flood-frequency curves using the Pearson type III distribution as recommended by the Interagency Advisory Committee on Water Data (1982). Population skew coefficients may be estimated from the error weighted average of the sample skew and regional skew coefficient. Weighted estimating methods for regional skew coefficients have been suggested, because the sampling error of station skew coefficients varies with the record lengths of the stations. A weighted grid isoline mapping method has been used in this analysis to estimate regional skew coefficients for flood-frequency analyses of Mississippi streams. The weighted grid mapping method was shown to be more accurate than the IACWD skew map (1982) in two of three homogenous skew regions in Mississippi, based on a comparison of the MPRESS and $\hat{MSE}$ of the two estimators.

Flood-frequency information forms an essential basis for managing development in flood plains. The accuracy of flood-frequency estimates from records of annual peak flow is improved by correcting for bias in sample skew coefficients and by using weighted regional skew estimating methods.

REFERENCES

Fisher, R.A., 1931, The moments of the distribution of normal samples of measures of departure from normality: Proceedings of the Royal Society of London, V. 130, p. 16-28.

Interagency Advisory Committee on Water Data, 1982, Bull. 17b, Guidelines for determining floodflow frequency: U.S. Geological Survey, Office of Water Data Coordination, Reston, Va., 28 p.

Tasker, G.D., and Stedinger, J.R., 1986, Regional skew with weighted LS regression: Journal of Water Resources Planning and Management, v. 112, no. 2, p. 225-237.

Tung, Y.K., and Mays, L.W., 1981, Generalized skew coefficients for flood frequency analysis: Wtr Rscs Bull., v. 17, no. 2, p. 262-269.

Wallis, J.R., Matalas, N.C., and Slack, J.R., 1974, Just a moment!: Water Resources Research, v. 10, no. 2, p. 217-219.