JOURNAL OF THE
HYDRAULICS DIVISION

Flood-Frequency Analyses with Prerecord Information

By Gary D. Tasker,[1] A. M. ASCE and Wilbert O. Thomas, Jr.[2]

Introduction

Prerecord flood information is often useful for flood-frequency analysis of systematic stream-gaging records. In the past, such historic information has been evaluated in a rather subjective graphic manner. Recently the Hydrology Committee of the Water Resources Council (1) has provided analytical guidelines for treating prerecord historic flood information in flood-frequency analyses. However, the Committee includes the study of alternative procedures for treating historic information in their list of needed additional studies.

Monte Carlo experiments were performed on each of several procedures for treating historic flood information to gain some insight into the usefulness of alternative procedures. The procedures are similar in that they follow general guidelines provided by the Hydrology Committee. The procedures differ in the sequence in which sample skew is weighted with a generalized skew and adjustment for historic information is made. The procedures also differ in the weight given the sample skew coefficient.

Experimental Design

Description of Estimating Procedures.—For each method of treating historic data, sample statistics were defined as in Ref. 1 as follows:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \quad \ldots \ldots \ldots \quad (1)$$

Note.—Discussion open until July 1, 1978. To extend the closing date one month, a written request must be filed with the Editor of Technical Publications, ASCE. This paper is part of the copyrighted Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers, Vol. 104, No. HY2, February, 1978. Manuscript was submitted for review for possible publication on August 2, 1977.

[1]Hydrologist, U.S. Geological Survey, Reston, Va.
[2]Hydrologist, U.S. Geological Survey, Reston, Va.

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

$$G = \frac{N \sum_{i=1}^{N} (X_i - \bar{X})^3}{(N - 1)(N - 2)\, S^3} \qquad (3)$$

in which X_i = the ith observation for sequence of length N; $\bar{X}$ = sample mean; S = sample standard deviation; and G = sample skew coefficient. Equations for adjusting statistics for historic data were defined as in Ref. 1 as follows:

$$W = \frac{H - Z}{N} \qquad (4)$$

$$\tilde{M} = \frac{W \sum_{i=1}^{N} X_i + \sum_{j=1}^{Z} \hat{X}_j}{H} \qquad (5)$$

$$\tilde{S}^2 = \frac{W \sum_{i=1}^{N} (X_i - \tilde{M})^2 + \sum_{j=1}^{Z} (\hat{X}_j - \tilde{M})^2}{H - 1} \qquad (6)$$

$$\tilde{G} = \frac{H}{(H - 1)(H - 2)\, \tilde{S}^3} \left[\frac{W(N - 1)(N - 2)\, S^3\, G'}{N} + 3W(N - 1)(M - \tilde{M})\, S^2 + WN(M - \tilde{M})^3 + \sum_{j=1}^{Z} (\hat{X}_j - \tilde{M})^3 \right] \qquad (7)$$

in which W = the weighting factor to be applied to the N events (excluding any high outliers); H = the number of years in the historic period; $\tilde{M}$ = the historically adjusted mean; $\tilde{S}$ = the historically adjusted standard deviation; $\tilde{G}$ = the historically adjusted skew coefficient; Z = the number of historic peaks and high outliers with historic information; $\hat{X}_j$ = the jth observed historic peak or high outlier, or both; and G' = either the sample skew coefficient or a weighted average of sample skew and generalized skew, as specified herein.

Five methods of treating historic information were used to estimate statistics of the underlying Pearson Type III distribution. The methods tested for treating historic information were as follows.

Method A.—Sample statistics were computed from Eqs. 1, 2, and 3. Sample skew, G, was weighted with a generalized skew coefficient, $\bar{G}$, by

$$G' = wG + (1 - w)\, \bar{G} \qquad (8)$$

in which $w = 0$ when $N \leq 25$; $w = (N - 25)/75$ when $25 < N < 100$; and $w = 1$ when $N \geq 100$. The estimates of statistics are then adjusted for historic information by Eqs. 4–7 to obtain the final estimate of statistics.

Method B.—Sample statistics are computed from Eqs. 1, 2, and 3. The resulting estimates of statistics are adjusted for historic information by Eqs. 4–7 with $G' = G$. The historically adjusted skew coefficient is weighted with a generalized skew coefficient, $\bar{G}$, by Eq. 8.

Method C.—Same as method B except w is defined as follows: $w = 0$ when $(N + H)/2 \leq 25$; $w = \{[(N + H)/2] - 25\}/75$ when $25 < (N + H)/2 < 100$; and $w = 1$, when $(N + H)/2 \geq 100$.

Method D.—Same as method B except $w = 0$ when $H \leq 25$; $w = (H - 25)/75$ when $25 < H < 100$; and $w = 1$ when $H \geq 100$. This is the method intended for use in Ref. 1 to adjust for historic information in a flood frequency analysis.

Method E.—Sample statistics are computed from Eqs. 1, 2, and 3. Sample skew is weighted with a generalized skew coefficient by Eq. 8. No historical adjustment is made. This method is included for comparison.

Monte Carlo Test Plan.—The Hydrology Committee assumes that a series of logarithms of annual peaks is adequately described by a Pearson Type III distribution. Therefore, a series of logarithms of annual peaks including one or more historic peaks is simulated by a series of Pearson Type III random numbers. The number generator uses the modified Wilson-Hilferty transformation described by Kirby (2). It produces Pearson Type III random numbers with correct mean, variance, skew, lower bound for skews up to at least 9.0, and upper bound for skews down to at least −9.0.

A record with historical information was simulated by generating a series of H numbers, the largest of which was considered the logarithm of the historic peak. The first N of these numbers were considered the logarithms of systematically recorded annual peaks. If any simulated historic peak occurred within the systematic record (high outlier) it was removed from the systematic record, the value of N decreased by one, and the historic peak treated as if it occurred outside the systematic record. This simulates the method recommended in Ref. 1 for treating high outliers. In order to simulate a record with more than one historic peak, the experiment was repeated with the three largest of the H numbers considered three historic peaks.

For each method, for values of $\bar{G} = -0.3$, 0.0, 0.3, and 0.6, and for each combination of N and H, 200 samples were generated with population statistics $\mu = 3.0$, $\sigma = 0.25$, and γ equal to −0.3, 0.0, and 0.6. Seventeen combinations of N and H were considered: (N, H) = (15, 25), (15, 55), (15, 95), (15, 135), (25, 35), (25, 65), (25, 105), (25, 145), (35, 45), (35, 75), (35, 115), (50, 60), (50, 90), (50, 130), (65, 75), (65, 105), and (65, 145).

Note that this method of testing procedures for treating historic information assumes that the historic information is certain. That is, both historic period of record, H, and the discharge associated with a historic peak are known without error.

Experimental Results

The performance of each method was judged by how well the 100-yr peak discharges determined by each method approximate the true value of the 100-yr peak discharge of the Pearson Type III distribution underlying the generated data.

The estimated 100-yr peak discharge using the jth method, $Q_{100,j}$, is

$$\log (Q_{100,j}) = M_j + K_j S_j \quad (9)$$

in which M_j and S_j = the historically adjusted sample mean and standard deviation; and K_j = the Pearson Type III deviate for exceedance probability of 0.01 obtained from tables of K values in Appendix 3 of Ref. 1 using the historically adjusted sample skew. The true 100-yr peak discharge, $Q_{100,T}$, is

$$\log (Q_{100,T}) = \mu + K\sigma \quad (10)$$

in which μ and σ = population mean and standard deviation; and K is obtained from tables of K values using population skew.

Three criteria are used to evaluate the performance of each method: (1) The root-mean-square error of the logarithm of the 100-yr peak; (2) the probability of overestimation of the 100-yr peak; and (3) the expected opportunity loss incurred in choosing one method over a method that gives the population value of the 100-yr peak discharge. The latter is an economic criterion that requires some knowledge of the economic loss incurred when an overestimate or underestimate is made.

TABLE 1.—Optimal Choice of Method Based on Root-Mean-Square Error of 100-yr Peak Discharge as Function of N, H, $\hat{G}$, and γ

		Values of γ											
		−0.3				0.0				0.6			
		Values of $\hat{G}$											
N	H	−0.3	0.0	0.3	0.6	−0.3	0.0	0.3	0.6	−0.3	0.0	0.3	0.6
15	25	A	B,C,D	A	A	A	B,C,D	A	A	A	A	B,C,D	B,C,D
	55	A	A	D	D	A	A	A	D	A	A	A	A
	95	A	C	D	D	A	A	C	D	D	A	A	A
	135	A	C	C	D	A	A	C	D	D	A	A	A
25	35	D	D	D	D	A	D	D	D	A	A	A	B
	65	D	D	D	D	A	A	D	D	D	A	A	A
	105	C	C	D	D	A	A	C	C	D	A	A	A
	145	A	C	C	D	A	A	C	C	C	A	A	A
35	45	C	D	D	D	A	B	D	D	A	A	A	B
	75	A	D	D	D	A	A	C	D	D	A	A	B
	115	A	C	D	D	A	A	C	D	D	A	A	A
50	60	B	D	D	D	A	B	D	D	A	D	D	B
	90	A	C	D	D	A	B	C	D	D	D	A	B
	130	A	C	C	D	A	A	C	C	D	A	A	A
65	75	B	D	D	D	A	B	C	D	D	A	B	B
	105	B	C	D	D	A	B	C	D	D	C	A	B
	145	B	C,D	C,D	C,D	A	A	B	C,D	C,D	C,D	B	A

Results in Terms of Root-Mean Square Error.—Let $Y_{i,j}$ denote the logarithmic value of the 100-yr peak estimate for the ith replication using the jth method, and Y_{100} denote the logarithmic value of the 100-yr peak determined from the population statistics of the underlying distribution. Root-mean-square error, rmse$_j$, is

$$\mathrm{rmse}_j = \left[\sum_{i=1}^{200} \frac{(Y_{i,j} - Y_{100})^2}{200} \right]^{1/2} \quad (11)$$

The optimal method for each combination of N, H, $\hat{G}$, and γ (Table 1) is the method that yielded the smallest root-mean-square error. The computed root-mean-square error for method E (no historic adjustment used) was higher than the computed root-mean-square error for the other four methods. Table 2 shows that, generally, inclusion of historic information by any of the tested methods, A to D, improves the accuracy of an estimate of the 100-yr peak discharge. Even with historic record lengths only 10 yr longer than systematic record lengths, some improvement in accuracy is indicated.

TABLE 2.—Root-Mean-Square Error of 100-yr Peak Discharge as Function of N, H, $\hat{G}$, and γ

			Values of γ											
			−0.3				0.0				0.6			
			Values of $\hat{G}$											
Method	N	H	−0.3	0.0	0.3	0.6	−0.3	0.0	0.3	0.6	−0.3	0.0	0.3	0.6
A	15	25	.10	.12	.14	.18	.12	.12	.13	.16	.18	.17	.16	.17
		95	.08	.10	.13	.16	.10	.09	.10	.13	.17	.15	.14	.14
	35	45	.06	.09	.12	.16	.08	.09	.10	.13	.14	.13	.11	.11
		115	.06	.08	.11	.15	.07	.07	.09	.12	.13	.11	.09	.09
	50	60	.06	.08	.10	.13	.07	.07	.08	.10	.13	.12	.12	.12
		130	.05	.07	.10	.13	.06	.06	.07	.09	.12	.10	.09	.09
B	15	25	.10	.11	.15	.19	.13	.12	.13	.16	.22	.19	.16	.16
		95	.10	.10	.13	.17	.13	.11	.11	.13	.22	.19	.17	.16
	35	45	.06	.08	.12	.16	.09	.08	.10	.13	.18	.14	.12	.10
		115	.06	.08	.11	.15	.08	.07	.08	.12	.16	.13	.10	.09
	50	60	.06	.07	.10	.13	.07	.06	.07	.10	.14	.12	.11	.10
		130	.05	.07	.09	.12	.07	.06	.07	.09	.14	.12	.10	.09
C	15	25	.10	.11	.15	.19	.13	.12	.13	.16	.22	.19	.16	.16
		95	.09	.09	.11	.13	.11	.10	.10	.11	.18	.17	.16	.15
	35	45	.06	.08	.11	.15	.09	.08	.09	.12	.17	.14	.11	.11
		115	.06	.07	.07	.07	.07	.07	.07	.08	.12	.11	.10	.10
	50	60	.06	.07	.09	.12	.07	.06	.07	.09	.14	.12	.11	.11
		130	.06	.06	.06	.07	.07	.07	.07	.07	.11	.11	.10	.10
D	15	25	.10	.11	.15	.19	.13	.12	.13	.16	.22	.19	.16	.16
		95	.10	.10	.10	.10	.10	.10	.10	.10	.16	.16	.16	.16
	35	45	.06	.08	.10	.14	.09	.08	.09	.12	.16	.13	.11	.11
		115	.07	.07	.07	.07	.08	.08	.08	.08	.11	.11	.11	.11
	50	60	.06	.07	.09	.11	.07	.07	.07	.09	.13	.12	.11	.11
		130	.06	.06	.06	.06	.07	.07	.07	.07	.11	.11	.11	.11
E	15		.11	.13	.16	.20	.14	.12	.14	.17	.23	.20	.18	.18
	35		.07	.08	.12	.16	.09	.08	.10	.13	.18	.14	.12	.12
	50		.06	.07	.10	.13	.08	.07	.08	.10	.15	.13	.11	.11

In addition, Tables 1 and 2 show that: (1) When $\gamma = \hat{G}$, method A has the smallest rmse for small samples and method B has the smallest rmse for the large samples; and (2) when γ is significantly different from $\hat{G}$, method D has the smallest rmse and method B has the largest rmse.

Results in Terms of Probability of Overestimation.—To some designers a desirable property of an estimate may be that the chance of the estimate being too large is equal to the chance of the estimate being too small. That is, it may be desirable to have

$$P(Y_{i,j} < Y_{100}) = P(Y_{i,j} > Y_{100}) = 0.5 \quad \text{. . . . . (12)}$$

in which $Y_{i,j}$ = the estimated 100-yr peak discharge. For each method and each combination of N, H, γ, and $\bar{G}$, the proability of overestimation was calculated by

$$P(Y_{i,j} > Y_{100}) = \frac{\text{number of overestimates}}{\text{total number of estimates (200)}} \quad \text{. . . . . (13)}$$

The optimal method (Table 3) in terms of probability of overestimation is the method that yields the value of $P(Y_{i,j} > Y_{100})$ nearest 0.50. Table 4 reveals

TABLE 3.—Optimal Choice of Method Based on Probability of Overestimation of 100-yr Peak Discharge as Function of N, H, $\bar{G}$, and γ

		Values of γ											
		−0.3				0.0				0.6			
		Values of $\bar{G}$											
N	H	−0.3	0.0	0.3	0.6	−0.3	0.0	0.3	0.6	−0.3	0.0	0.3	0.6
15	25	A	B, C, D	B, C, D	E	A	E	D	D	A	A	A	A
	55	A	D	D	D	A	A	D	D	A	A	A	A
	95	A	C	D	D	D	A	E	D	D	D	D	A, E
	135	D	C, D	D	D	D	A	D	D	D	D	D	A
25	35	D	D	D	D	A	A	D	D	A	A	A	A
	65	B	D	D	D	A	E	D	D	D	A	A	A
	105	D	D	D	D	D	A	D	D	D	D	D	A
	145	D	D	D	D	D	A, E	D	D	D	D	D	E
35	45	D, E	D	D	D	A	D	D	D	A	A	A	A
	75	E	D	D	D	D	A	D	D	D	D	D	A
	115	E	D	D	D	D	E	D	D	D	D	D	A
50	60	D	D	D	D	A	E	D	D	A	A	D	D
	90	D	D	D	D	D	A	D	D	D	D	D	E
	130	E	D	D	D	D	E	D	D	D	D	D	A
65	75	E	D, E	D	D	A	A	D	D	A	A	A	A
	105	E	D	D	D	D	E	D	D	D	D	D	A
	145	B	C, D	C, D	C, D	C, D	E	C, D	C, D	C, D	C, D	C, D	A

a tendency for overestimation using method A and a tendency for underestimation using the other four methods. Part of the tendency for underestimation is due to the fact that S and G are biased estimates of σ and γ. Wallis, Matalas, and Slack (5) show that S and G underestimate σ and γ for a Pearson Type III distribution with positive γ. When γ is negative, S underestimates σ and G overestimates γ. Method E in Table 4 shows that these two biases work against each other to produce an approximately unbiased estimate of the 100-yr peak when $\bar{G} = \gamma = -0.3$. Also, observed in Table 4 is that the tendency for method A for overestimation apparently negates the bias in S and G when γ is about 0.6. Method A is then the optimal method in terms of probability of overestimation. For the general case of $\gamma \neq \bar{G}$, method D is clearly optimal

in terms of probability of overestimation.

Results in Terms of Expected Opportunity Loss.—In decision theory penalties for making incorrect decisions are taken into account. This leads to a third criterion for choosing the method of treating historic data in which penalties for incorrect decisions are minimized. Assume that minimum loss, L_{min}, is incurred when the correct value of the 100-yr peak is used for a design. The difference between L_j, the loss associated with the jth method, and L_{min} is the opportunity

TABLE 4.—Probability of Overestimation of 100-yr Peak as Function of N, H, $\bar{G}$, and γ

			Values of γ											
			−0.3				0.0				0.6			
			Values of $\bar{G}$											
Method	N	H	−0.3	0.0	0.3	0.6	−0.3	0.0	0.3	0.6	−0.3	0.0	0.3	0.6
A	15	25	.51	.66	.80	.89	.39	.57	.64	.72	.28	.34	.42	.50
		95	.44	.67	.82	.88	.27	.44	.61	.74	.18	.28	.36	.46
	35	45	.58	.81	.92	.97	.38	.58	.72	.82	.18	.28	.36	.50
		115	.54	.81	.93	.97	.26	.56	.77	.91	.10	.20	.36	.52
	50	60	.59	.79	.91	.96	.36	.58	.73	.86	.24	.31	.40	.55
		130	.58	.81	.94	.98	.30	.56	.72	.89	.14	.24	.34	.48
B	15	25	.40	.57	.76	.90	.24	.44	.61	.70	.09	.18	.29	.40
		95	.37	.62	.78	.86	.19	.34	.54	.72	.10	.20	.30	.42
	35	45	.44	.76	.91	.97	.26	.48	.67	.80	.04	.14	.24	.42
		115	.46	.77	.90	.97	.22	.44	.71	.90	.04	.12	.28	.44
	50	60	.54	.74	.88	.96	.28	.50	.70	.84	.15	.22	.30	.50
		130	.50	.77	.92	.98	.24	.48	.70	.84	.08	.16	.27	.38
C	15	25	.40	.57	.76	.90	.24	.44	.61	.70	.09	.18	.29	.40
		95	.40	.54	.68	.77	.25	.35	.50	.64	.21	.28	.34	.42
	35	45	.44	.74	.90	.97	.30	.48	.66	.78	.06	.16	.25	.40
		115	.53	.63	.77	.82	.33	.46	.60	.66	.25	.32	.34	.40
	50	60	.51	.72	.88	.92	.29	.50	.68	.83	.18	.24	.32	.50
		130	.56	.60	.63	.70	.38	.44	.51	.54	.34	.36	.37	.40
D	15	25	.40	.57	.76	.90	.24	.44	.61	.70	.09	.18	.29	.40
		95	.44	.44	.44	.46	.35	.36	.38	.40	.39	.39	.39	.40
	35	45	.46	.72	.88	.95	.31	.48	.64	.78	.08	.19	.26	.40
		115	.54	.54	.54	.54	.46	.46	.46	.46	.38	.38	.38	.38
	50	60	.50	.70	.84	.92	.30	.52	.66	.80	.18	.26	.34	.48
		130	.56	.56	.56	.56	.43	.43	.43	.43	.40	.40	.40	.40
E	15		.43	.62	.81	.90	.28	.47	.61	.73	.11	.21	.34	.45
	35		.48	.76	.88	.96	.26	.49	.68	.82	.06	.16	.28	.44
	50		.50	.77	.90	.94	.27	.48	.67	.81	.12	.20	.30	.45

loss incurred in choosing method j rather than a method that estimates the value of the 100-yr peak correctly. For a large number of samples the average difference, $\bar{L}_j - L_{min}$, approaches the expected opportunity loss (3). Assuming that loss is linearly related to the difference between the estimated and true 100-yr peak discharges, expected opportunity loss, $E(L)$, is calculated from

200 replications for each method and each combination of N, H, $\bar{G}$, and γ by

$$E(L) = \frac{k^{+}\Sigma w^{+} + k^{-}\Sigma w^{-}}{(k^{+} + k^{-})\,200} \quad (14)$$

in which $w^{+} = \max\,(Y_{i,j} - Y_{100}, 0)$; $w^{-} = \min\,(Y_{i,j} - Y_{100}, 0)$; and k^{+} and k^{-} = slopes of the overdesign and underdesign loss functions, respectively. This method of evaluating relative performance of the methods was suggested by the work of Slack, Wallis, and Matalas (4).

TABLE 5.—Optimal Choice of Method Based on Expected Opportunity Loss as Function of k^{+}/k^{-}, N, H, G, and γ

			Values of γ											
			−0.3				0.0				0.6			
			Values of $\bar{G}$											
k^{+}/k^{-}	N	H	−0.3	0.0	0.3	0.6	−0.3	0.0	0.3	0.6	−0.3	0.0	0.3	0.6
10/1	15	25	B,C,D	B,C,D	B,C,D	A	B,C,D	B,C,D	B,C,D	B,C,D	B,C,D	B,C,D	B,C,D	B,C,D
		55	D	D	D	D	B	B	D	D	B	B	B	D
	35	45	D	D	D	D	B	D	D	D	B	B	B	B
		115	B	D	D	D	B	B	D	D	B	B	B	B
	65	75	B	D	D	D	B	B	D	D	B	B	B	B
		105	B	D	D	D	B	B	D	D	B	B	B	B
10/2	15	25	B,C,D	B,C,D	B,C,D	A	B,C,D	B,C,D	B,C,D	B,C,D	B,C,D	B,C,D	B,C,D	B,C,D
		55	D	D	D	D	C	B	D	D	A	C	B	D
	35	45	D	D	D	D	B	D	D	D	D	B	B	B
		115	B	D	D	D	B	C	D	D	B	B	B	B
	65	75	B	D	D	D	B	B	D	D	B	B	B	B
		105	B	D	D	D	B	B	D	D	B	B	B	B
10/5	15	25	B,C,D	B,C,D	B,C,D	A	B,C,D	B,C,D	B,C,D	A	A	B,C,D	B,C,D	B,C,D
		55	D	D	D	D	A	D	D	D	A	A	A	D
	35	45	D	D	D	D	B	D	D	D	A	D	B	B
		115	B	D	D	D	A	C	D	D	B	B	B	B
	65	75	B	D	D	D	B	B	D	D	D	B	B	B
		105	B	D	D	D	B	B	C	D	C	B	B	B
10/10	15	25	A	A	B,C,D	A	A	B,C,D	B,C,D	A	A	A	A	A
		55	A	D	D	D	A	A	D	D	A	A	A	A
	35	45	D	D	D	D	A	B	D	D	A	A	A	B
		115	A	C	D	D	A	A	C	D	D	A	A	A
	65	75	B	D	D	D	B	B	D	D	D	A	B	B
		105	B	C	D	D	A	B	C	C	D	C	A	A
5/10	15	25	A	A	A	A	A	A	B,C,D	A	A	A	A	A
		55	A	A	A	D	A	A	A	D	A	A	A	A
	35	45	A	D	D	D	A	A	C	D	A	A	A	A
		115	A	A	D	D	D	A	A	C	D	D	A	A
	65	75	A	B	D	D	A	A	B	D	A	A	A	A
		105	A	A	C	C	A	A	A	B	D	D	A	A
2/10	15	25	A	A	A	A	A	A	A	A	A	A	A	A
		55	A	A	A	D	A	A	A	A	A	A	A	A
	35	45	A	A	D	D	A	A	A	A	A	A	A	A
		115	A	A	B	A	D	A	A	A	D	D	A	A
	65	75	A	A	B	D	A	A	A	A	A	A	A	A
		105	A	A	A	B	D	A	A	A	D	D	A	A
1/10	15	25	A	A	A	A	A	A	A	A	A	A	A	A
		55	A	A	A	A	A	A	A	A	A	A	A	A
	35	45	A	A	D	D	A	A	A	A	A	A	A	A
		115	A	A	A	A	D	A	A	A	D	D	A	A
	65	75	A	A	A	B	A	A	B	A	A	A	A	A
		105	A	A	B	A	D	A	A	A	D	D	A	A

Nine combinations of k^{+} and k^{-} were considered: (k^{+},k^{-}) = (10,0), (10,1) (10,2), (10,5), (10,10), (5,10), (2,10), (1,10), and (0,10). The combination (10,0) is the case where no loss is associated with underdesign, and the combination (0,10) is the case where no loss is associated with overdesign. The other combinations give various weights to losses associated with overdesign and underdesign. Table 5 shows the optimal method in terms of expected opportunity loss as the method that yields the smallest $E(L)$. The combinations of k^{+} and k^{-} are shown as the ratio k^{+}/k^{-} in Table 5 because it is the relative value of k^{+} and k^{-} that is important in computing $E(L)$. The partial results shown in Table 5 indicate that method A is usually the optimal method when overdesign losses are small when compared to underdesign losses. Also, method D is the predominant optimal method in terms of expected opportunity loss when underdesign losses are small when compared to overdesign losses.

Tests with Three Historic Peaks.—The Monte Carlo experiments were repeated using the three largest numbers in H years taken as three historic peaks. Resulting root-mean-square errors were slightly smaller than those for one historic peak for methods A, B, C, and D.

The probabilities of overestimation of the 100-yr peak discharge for methods B, C, and D were about the same when three historic peaks were simulated as when one historic peak was simulated. The tendency for method A to overestimate the 100-yr peak showed a slight increase when three historic peaks were simulated instead of one. No calculations were made for expected opportunity losses.

Conclusions

Monte Carlo experiments were performed to evaluate several alternative methods of treating historic information in flood-frequency analyses. The tested methods are similar in that generally they follow the guidelines provided by the Hydrology Committee of the Water Resources Council. The methods differ in the sequence in which sample skew and generalized skew are weighted and adjustment for historic information is made, or they differ in the weight given sample skew.

Results, based on the assumption that a series of logarithms of annual peaks are adequately described by a Pearson Type III distribution, are sufficient to make several observations:

1. Inclusion of accurate prerecord information in a flood-frequency analysis by any of the methods tested generally improves the estimate of 100-yr peak discharge. The uncertainty of the historic period of record or the discharge of the historic peak was not considered in this analysis.
2. When weighting historically adjusted sample skew with a generalized skew as specified in Ref. 1, generally it is better to compute the weight factor using the historic record length (method D) rather than the systematic record length (method B) or an average of the systematic record length and historic record length (method C).
3. Adjusting for historical information after weighting with a generalized skew coefficient (method A) tends to overestimate the 100-yr peak when population skew is less than about 0.3. In terms of expected opportunity loss, method A tends to be the optimal method when losses resulting from overdesign are small compared to losses resulting from underdesign.
4. Methods B, C, and D tend to underestimate the 100-yr peak when population skew is positive. This is probably due, in part, to the fact that S and G are biased estimates of σ and γ. In terms of expected opportunity loss, method

D tends to be optimal when underdesign losses are small compared to overdesign losses.

Appendix I.—References

1. "Guidelines for Determining Flood Flow Frequency," Hydrology Committee of the Water Resources Council, *Bulletin No. 17*, Washington, D.C., 1976.
2. Kirby, W., "Computer-oriented Wilson-Hilferty Transformation that Preserves the First Three Moments and the Lower Bound of the Pearson Type III Distribution," *Water Resources Research*, Vol. 8, No. 5, Sept., 1972, pp. 1251–1254.
3. Raffia, H., *Decision Analysis: Intro. Lectures on choices under uncertainty*, Addison-Wesley Publishing Co. Inc., New York, N.Y., 1970.
4. Slack, J. R., Wallis, J. R., and Matalas, N. C., "On the Value of Information to Flood Frequency Analysis," *Water Resources Research*, Vol. 11, No. 5, Oct., 1975, pp. 629–647.
5. Wallis, J. R., Matalas, N. C., and Slack, J. R., "Just a Moment!," *Water Resources Research*, Vol. 10, No. 2, Apr., 1972, pp. 211–219.

Appendix II.—Notation

The following symbols are used in this paper:

$E(L)$ = expected opportunity loss (3);
G = sample skew coefficient, or historically weighted skew coefficients;
G' = sample skew coefficient or weighted average of sample skew coefficient and generalized skew coefficient;
$\bar{G}$ = generalized skew coefficient;
H = historic record length, in years;
K_j or K = Pearson Type III deviate for exceedance probability of 0.01 and skew coefficient G or γ, respectively;
k^+ and k^- = slopes of overdesign and underdesign loss functions, respectively;
L_{min} = economic loss incurred when correct value of design discharge is used for design;
L_j = economic loss incurred when value of design discharge is obtained using jth method and is used for design;
$\tilde{M}$ = historically weighted mean;
N = systematic record length, in years;
rmse_j = root-mean-square error of 100-yr peak using jth method;
S = sample standard deviation;
$\tilde{S}$ = historically weighted standard deviation;
W = weighting factor to be applied to N events in adjusting for historical information;
w = weighting factor to be applied to sample skew coefficient when weighting sample skew coefficient with generalized skew coefficient;
w^+ = $Y_{i,j} - Y_{100}$, when $Y_{i,j} > Y_{100}$;
w^- = $Y_{i,j} - Y_{100}$, when $Y_{i,j} < Y_{100}$;
X_i = logarithm of ith simulated, systematically recorded annual peak;

$\tilde{X}_j$ = logarithm of jth simulated historic peak;
$\bar{X}$ = sample mean;
$Y_{i,j}$ = logarithmic value of 100-yr peak estimated for ith replication using jth method;
Y_{100} = logarithmic value of 100-yr peak estimated from population statistics of underlying distribution;
Z = number of historic peaks;
μ = population mean;
σ = population standard deviation; and
γ = population skew coefficient.