Appendix 8

WEIGHTING OF INDEPENDENT ESTIMATES

The following procedure is suggested for adjusting flow frequency estimates based upon short records to reflect flood experience in nearby hydrologically similar watersheds, using any one of the various *generalization methods mentioned in V.C.1. The procedure is based upon* the assumption that the estimates are independent, which for practical purposes is true in most situations.

If two independent estimates are weighted inversely proportional to their variance, the variance of the weighted average, z, is less than the variance of either estimate. According to Gilroy (30), if

$$z = \frac{x(V_y) + y(V_x)}{V_y + V_x} \qquad (8\text{-}1)$$

then

$$V_z = \frac{V_x V_y}{(V_x + V_y)^2} \left[V_x + V_y + 2r\sqrt{V_x V_y}\right] \qquad (8\text{-}2)$$

in which V_x, V_y, and V_z are the variances of x, y, and z respectively, and r is the cross correlation coefficient between values of x and values of y. Thus, if two estimates are independent, r is zero and

$$V_z = \frac{V_x V_y}{V_x + V_y} \qquad (8\text{-}3)$$

As the variance of flood events at selected exceedance probabilities computed by the Pearson Type III procedure is inversely proportional to the number of annual events used to compute the statistics (25), equation (8-3) can be written

$$C/N_z = \frac{(C/N_x)\ (C/N_y)}{C/N_x + C/N_y} = \frac{C}{N_x + N_y} \qquad (8\text{-}4)$$

in which C is a constant, N_x and N_y are the number of annual events used to compute x and y respectively, and N_z is the number of events that would be required to give a flood event at the selected exceedance probabilities with a variance equivalent to that of z computed by equation 8-1. Therefore,

$$N_z = N_x + N_y \tag{8-5}$$

From equation 8-1,

$$Z = \frac{xC/N_y + yC/N_x}{C/N_x + C/N_y} = \frac{x(N_x) + y(N_y)}{N_x + N_y} \tag{8-6}$$

Equation 8-6 can be used to weight independent estimates of the logarithms of flood discharges at selected probabilities and equation 8-5 can be used to appraise the accuracy of the weighted average. As a flood frequency discharge estimated by generalization tends to be independent of that obtained from the station data, such weighting is often justified particularly if the stations used in the generalization cover an area with a radius of over 100 miles or if their period of record is long in comparison with that at the station for which the estimate is being made. For generalizations based on stations covering a smaller area or with shorter records, the accuracy of the weighted average given by equation 8-6 is less than given by equation 8-5.

For cases where the estimates from the generalization and from the station data are not independent, the accuracy of the weighted estimate is reduced depending on the cross correlation of the estimates.

Given a peak discharge of 1,000 cfs with exceedance probability of 0.02 from a generalization with an accuracy equivalent to an estimate based on a 10-year record, for example, and an independent estimate of 2,000 cfs from 15 annual peaks observed at the site, the weighted average would be given by substitution in equation 8-6 as follows:

$$\text{Log } Q_{.02} = \frac{10(\log 1000) + 15(\log 2000)}{25} = 3.181$$

from which $Q_{.02}$ is 1,520 cfs. By equation 8-5 this estimate is as good as would be obtained from 25 annual peaks.

If an expected probability adjustment is to be applied to a weighted estimate, the adjustment to probability should be the same as that applicable to samples from normal distributions as described in Appendix 11, but N should be that for a sample size that gives equivalent accuracy. Thus, in the preceding example, the expected probability adjustment would be that for a sample of size 25 taken from a normal distribution.

*

Appendix 9

CONFIDENCE LIMITS

The record of annual peak flows at a site is a random sample of the underlying population of annual peaks and can be used to estimate the frequency curve of that population. If the same size random sample could be selected from a different period of time, a different estimate of the underlying population frequency curve probably would result. Thus, an estimated flood frequency curve can be only an approximation to the true frequency curve of the underlying population of annual flood peaks. To gauge the accuracy of this approximation, one may construct an interval or range of hypothetical frequency curves that, with a high degree of confidence, contains the population frequency curve. Such intervals are called confidence intervals and their end points are called confidence limits.

This appendix explains how to construct confidence intervals for flood discharges that have specified exceedance probabilities. To this end, let X_P^* denote the true or population logarithmic discharge that has exceedance probability P. Upper and lower confidence limits for X_P^*, with confidence level c, are defined to be numbers $U_{P,c}(X)$ and $L_{P,c}(X)$, based on the observed flood records, X, such that the upper confidence limit $U_{P,c}(X)$ lies above X_P^* with probability c and the lower limit $L_{P,c}(X)$ lies below X_P^* with probability c. That is, the confidence limits have the property that

$$\text{Probability} \quad \{U_{P,c}(X) \geq X_P^*\} = c \qquad \text{(9-1a)}$$

$$\text{Probability} \quad \{L_{P,c}(X) \leq X_P^*\} = c \qquad \text{(9-1b)}$$

*

*Explicit formulas for computing the confidence limits are given below; the above formulas simply explain the statistical meaning of the confidence limits.

The confidence limits defined above are called <u>one-sided</u> confidence limits because each of them describes a bound or limit on just one side of the population p-probability discharge. A two-sided confidence interval can be formed from the overlap or union of the two one-sided intervals, as follows:

$$\text{Probability} \quad \left\{ L_{P,c}(X) \leq X_P^* \leq U_{P,c}(X) \right\} \quad = 2c-1 \qquad (9\text{-}2)$$

Thus, the union of two one-sided 95-percent confidence intervals is a two-sided 90-percent interval. It should be noted that the two-sided interval so formed may not be the narrowest possible interval with that confidence level; nevertheless, it is considered satisfactory for use with these guidelines.

It may be noted in the above equations that $U_{P,c}(X)$ can lie above X_P^* if and only if $U_{P,c}(X)$ lies above a fraction (1-P) of all possible floods in the population. In quality control terminology, $U_{P,c}(X)$ would be called an upper <u>tolerance limit</u>, at confidence level c, for the proportion (1-P) of the population. Similarly, $L_{P,c}(X)$ would be a lower tolerance limit for the proportion (P). Because the tolerance limit terminology refers to proportions of the population, whereas the confidence-limit terminology refers directly to the discharge of interest, the confidence-limit terminology is adopted in these guidelines.

Explicit formulas for the confidence limits are derived by specifying the general form of the limits and making additional simplifying assumptions to analyze the relationships between sample statistics and population statistics. The general form of the confidence limits is specified as:

$$U_{P,c}(X) = \overline{X} + S\left(K^U_{P,c}\right) \qquad (9\text{-}3a)$$

$$L_{P,c}(X) = \overline{X} + S\left(K^L_{P,c}\right) \qquad (9\text{-}3b)$$

*

*
in which $\overline{X}$ and S are the logarithmic mean and standard deviation of the final estimated log Pearson Type III frequency curve and $K^U_{P,c}$ and $K^L_{P,c}$ are upper and lower confidence coefficients.

The confidence coefficients approximate the non-central t-distribution. The non-central t-variate can be obtained in tables (41, 32), although the process is cumbersome when G_W is non-zero. More convenient is the use of the following approximate formulas (32, pp. 2-15), based on a large sample approximation to the non-central t-distribution (42):

$$K^U_{P,c} = \frac{K_{G_W,P} + \sqrt{K^2_{G_W,P} - ab}}{a} \qquad (9\text{-}4a)$$

$$K^L_{P,c} = \frac{K_{G_W,P} - \sqrt{K^2_{G_W,P} - ab}}{a} \qquad (9\text{-}4b)$$

in which

$$a = 1 - \frac{z_c^2}{2(N-1)} \qquad (9\text{-}5)$$

$$b = K^2_{G_W,P} - \frac{z_c^2}{N} \qquad (9\text{-}6)$$

and z_c is the standard normal deviate (zero-skew Pearson Type III deviate) with cumulative probability c (exceedance probability 1-c). The systematic record length N is deemed to control the statistical reliability of the estimated frequency curve and is to be used for calculating confidence limits even when historic information has been used to estimate the frequency curve.

The use of equations 9-3 through 9-6 is illustrated by calculating 95-percent confidence limits for $X^*_{0.01}$, the 0.01 exceedance probability flood, when the estimated frequency curve has logarithmic mean, standard deviation, and skewness of 3.00, 0.25, and 0.20, respectively based on 50 years of systematic record.

*

*

$z_c = 1.645$ $\qquad\qquad$ $K_{G_W,P} = 2.4723$

$$a = 1 - \frac{(1.645)^2}{98} = 0.9724$$

$$b = (2.4723)^2 - \frac{(1.645)^2}{50} = 6.058$$

$$K^u_{0.01,\ 0.95} = \frac{2.4723 + \sqrt{(2.4723)^2 - (0.9724)(6.058)}}{0.9724}$$

$$= 3.026$$

$$K^L_{0.01,\ 0.95} = \frac{2.4723 - \sqrt{(2.4723)^2 - (0.9724)(6.058)}}{0.9724}$$

$$= 2.059$$

$$U_{0.01,\ 0.95}(X) = 3.00 + (0.25)(3.026) = 3.756$$

$$L_{0.01,\ 0.95}(X) = 3.00 + (0.25)(2.059) = 3.515$$

The corresponding limits in natural units (cubic feet per second) are 3270 and 5700; the estimated 0.01 exceedance probability flood is 4150 cubic feet per second.

Table 9-1 is a portion of the non-central t tables (43) for a skew of zero and can be used to compute $K^U_{P,c}$ and $K^L_{P,c}$ for selected values of P and c when the distribution of logarithms of the annual peaks is normal (i.e., G_W=0).

An example of using table 9-1 to compute confidence limits is as follows: Assume the 95-percent confidence limits are desired for $X^*_{0.01}$, the 0.01 exceedance probability flood for a frequency curve with logarithmic mean, standard deviation and skewness of 3.00, 0.25 and 0.00, respectively, based on 50 years of systematic record.

*

* $K^{U}_{0.01,\ 0.95} = 2.862$ Found by entering table 9-1 with confidence level 0.05, systematic record length 50 and exceedance probability 0.01.

$K^{L}_{0.01,\ 0.95} = 1.936$ Found by entering table 9-1 with confidence level 0.95, systematic record length 50 and exceedance probability 0.01.

$U_{0.01,\ 0.95}\ (X) = 3.00 + 0.25(2.862) = 3.715$

$L_{0.01,\ 0.95}\ (X) = 3.00 + 0.25(1.936) = 3.484$

The corresponding limits in natural units (cubic feet per second) are 3050 and 5190; the estimated 0.01 exceedance probability flood is 3820 cubic feet per second. *

*

Appendix 9 Notation

$U_{P,c}(X)$ = upper confidence limit in log units

$L_{P,c}(X)$ = lower confidence limit in log units

P = exceedance probability

c = confidence level

X_P^* = population logarithmic discharge for exceedance probability P

$\overline{X}$ = mean logarithm of peak flows

S = standard deviation of logarithms of annual peak discharges

$K_{G_w,P}$ = Pearson Type III coordinate expressed in number of standard deviations from the mean for weighted skew (G_w) and exceedance probability (P).

G_w = weighted skew coefficient

$K_{P,c}^U$ = upper confidence coefficient

$K_{P,c}^L$ = lower confidence coefficient

N = systematic record length

z_c = is the standard normal deviate *

*

TABLE 9-1
CONFIDENCE LIMIT DEVIATE VALUES FOR NORMAL DISTRIBUTION

Confidence Level	Systematic Record Length	EXCEEDANCE PROBABILITY											
	N	.002	.005	.010	.020	.040	.100	.200	.500	.800	.900	.950	.990
.01	10	6.178	5.572	5.074	4.535	3.942	3.048	2.243	.892	-.107	-.508	-.804	-1.314
	15	5.147	4.639	4.222	3.770	3.274	2.521	1.841	.678	-.236	-.629	-.929	-1.458
	20	4.675	4.212	3.832	3.419	2.965	2.276	1.651	.568	-.313	-.705	-1.008	-1.550
	25	4.398	3.960	3.601	3.211	2.782	2.129	1.536	.498	-.364	-.757	-1.064	-1.616
	30	4.212	3.792	3.447	3.071	2.658	2.030	1.457	.450	-.403	-.797	-1.107	-1.667
	40	3.975	3.577	3.249	2.893	2.500	1.902	1.355	.384	-.457	-.854	-1.169	-1.741
	50	3.826	3.442	3.125	2.781	2.401	1.821	1.290	.340	-.496	-.894	-1.212	-1.793
	60	3.723	3.347	3.038	2.702	2.331	1.764	1.244	.309	-.524	-.924	-1.245	-1.833
	70	3.647	3.278	2.974	2.644	2.280	1.722	1.210	.285	-.545	-.948	-1.272	-1.865
	80	3.587	3.223	2.924	2.599	2.239	1.688	1.183	.265	-.563	-.968	-1.293	-1.891
	90	3.538	3.179	2.883	2.561	2.206	1.661	1.160	.250	-.578	-.984	-1.311	-1.913
	100	3.498	3.143	2.850	2.531	2.179	1.639	1.142	.236	-.591	-.998	-1.326	-1.932
.05	10	4.862	4.379	3.981	3.549	3.075	2.355	1.702	.580	-.317	-.712	-1.017	-1.563
	15	4.304	3.874	3.520	3.136	2.713	2.068	1.482	.455	-.406	-.802	-1.114	-1.677
	20	4.033	3.628	3.295	2.934	2.534	1.926	1.370	.387	-.460	-.858	-1.175	-1.749
	25	3.868	3.478	3.158	2.809	2.425	1.838	1.301	.342	-.497	-.898	-1.217	-1.801
	30	3.755	3.376	3.064	2.724	2.350	1.777	1.252	.310	-.525	-.928	-1.250	-1.840
	40	3.608	3.242	2.941	2.613	2.251	1.697	1.188	.266	-.565	-.970	-1.297	-1.896
	50	3.515	3.157	2.862	2.542	2.188	1.646	1.146	.237	-.592	-1.000	-1.329	-1.936
	60	3.448	3.096	2.807	2.492	2.143	1.609	1.116	.216	-.612	-1.022	-1.354	-1.966
	70	3.399	3.051	2.765	2.454	2.110	1.581	1.093	.199	-.629	-1.040	-1.374	-1.990
	80	3.360	3.016	2.733	2.425	2.083	1.559	1.076	.186	-.642	-1.054	-1.390	-2.010
	90	3.328	2.987	2.706	2.400	2.062	1.542	1.061	.175	-.652	-1.066	-1.403	-2.026
	100	3.301	2.963	2.684	2.380	2.044	1.527	1.049	.166	-.662	-1.077	-1.414	-2.040

*

*

TABLE 9-1 (CONTINUED)
CONFIDENCE LIMIT DEVIATE VALUES FOR NORMAL DISTRIBUTION

EXCEEDANCE PROBABILITY

Confidence Level	Systematic Record Length N	.002	.005	.010	.020	.040	.100	.200	.500	.800	.900	.950	.990
.10	10	4.324	3.889	3.532	3.144	2.716	2.066	1.474	.437	-.429	-.828	-1.144	-1.715
	15	3.936	3.539	3.212	2.857	2.465	1.867	1.320	.347	-.499	-.901	-1.222	-1.808
	20	3.743	3.364	3.052	2.712	2.338	1.765	1.240	.297	-.541	-.946	-1.271	-1.867
	25	3.623	3.255	2.952	2.623	2.258	1.702	1.190	.264	-.570	-.978	-1.306	-1.908
	30	3.541	3.181	2.884	2.561	2.204	1.657	1.154	.239	-.593	-1.002	-1.332	-1.940
	40	3.433	3.082	2.793	2.479	2.131	1.598	1.106	.206	-.624	-1.036	-1.369	-1.986
	50	3.363	3.019	2.735	2.426	2.084	1.559	1.075	.184	-.645	-1.059	-1.396	-2.018
	60	3.313	2.974	2.694	2.389	2.051	1.532	1.052	.167	-.662	-1.077	-1.415	-2.042
	70	3.276	2.940	2.662	2.360	2.025	1.511	1.035	.155	-.674	-1.091	-1.431	-2.061
	80	3.247	2.913	2.638	2.338	2.006	1.495	1.021	.144	-.684	-1.103	-1.444	-2.077
	90	3.223	2.891	2.618	2.319	1.989	1.481	1.010	.136	-.693	-1.112	-1.454	-2.090
	100	3.203	2.873	2.601	2.305	1.976	1.470	1.001	.129	-.701	-1.120	-1.463	-2.101
.25	10	3.599	3.231	2.927	2.596	2.231	1.671	1.155	.222	-.625	-1.043	-1.382	-2.008
	15	3.415	3.064	2.775	2.460	2.112	1.577	1.083	.179	-.661	-1.081	-1.422	-2.055
	20	3.320	2.978	2.697	2.390	2.050	1.528	1.045	.154	-.683	-1.104	-1.448	-2.085
	25	3.261	2.925	2.648	2.346	2.011	1.497	1.020	.137	-.699	-1.121	-1.466	-2.106
	30	3.220	2.888	2.614	2.315	1.984	1.475	1.002	.125	-.710	-1.133	-1.479	-2.123
	40	3.165	2.838	2.568	2.274	1.948	1.445	.978	.108	-.726	-1.151	-1.499	-2.147
	50	3.129	2.805	2.538	2.247	1.924	1.425	.962	.096	-.738	-1.164	-1.513	-2.163
	60	3.105	2.783	2.517	2.227	1.907	1.411	.950	.088	-.747	-1.173	-1.523	-2.176
	70	3.085	2.765	2.501	2.213	1.893	1.401	.942	.081	-.753	-1.181	-1.532	-2.186
	80	3.070	2.752	2.489	2.202	1.883	1.392	.935	.076	-.759	-1.187	-1.538	-2.194
	90	3.058	2.740	2.478	2.192	1.875	1.386	.929	.071	-.763	-1.192	-1.544	-2.201
	100	3.048	2.731	2.470	2.184	1.868	1.380	.925	.068	-.767	-1.196	-1.549	-2.207

*

TABLE 9-1 (CONTINUED)
CONFIDENCE LIMIT DEVIATE VALUES FOR NORMAL DISTRIBUTION

*

Confidence Level	Systematic Record Length N	.002	.005	.010	.020	.040	.100	.200	.500	.800	.900	.950	.990
		EXCEEDANCE PROBABILITY											
.75	10	2.508	2.235	2.008	1.759	1.480	1.043	.625	-.222	-1.155	-1.671	-2.104	-2.927
	15	2.562	2.284	2.055	1.803	1.521	1.081	.661	-.179	-1.083	-1.577	-1.991	-2.775
	20	2.597	2.317	2.085	1.831	1.547	1.104	.683	-.154	-1.045	-1.528	-1.932	-2.697
	25	2.621	2.339	2.106	1.851	1.566	1.121	.699	-.137	-1.020	-1.497	-1.895	-2.648
	30	2.641	2.357	2.123	1.867	1.580	1.133	.710	-.125	-1.002	-1.475	-1.869	-2.614
	40	2.668	2.383	2.147	1.888	1.600	1.151	.726	-.108	-.978	-1.445	-1.834	-2.568
	50	2.688	2.400	2.163	1.903	1.614	1.164	.738	-.096	-.962	-1.425	-1.811	-2.538
	60	2.702	2.414	2.176	1.916	1.625	1.173	.747	-.088	-.950	-1.411	-1.795	-2.517
	70	2.714	2.425	2.186	1.925	1.634	1.181	.753	-.081	-.942	-1.401	-1.782	-2.501
	80	2.724	2.434	2.194	1.932	1.640	1.187	.759	-.076	-.935	-1.392	-1.772	-2.489
	90	2.731	2.441	2.201	1.938	1.646	1.192	.763	-.071	-.929	-1.386	-1.764	-2.478
	100	2.739	2.447	2.207	1.944	1.652	1.196	.767	-.068	-.925	-1.380	-1.758	-2.470
.90	10	2.165	1.919	1.715	1.489	1.234	.828	.429	-.437	-1.474	-2.066	-2.568	-3.532
	15	2.273	2.019	1.808	1.576	1.314	.901	.499	-.347	-1.320	-1.867	-2.329	-3.212
	20	2.342	2.082	1.867	1.630	1.364	.946	.541	-.297	-1.240	-1.765	-2.208	-3.052
	25	2.390	2.126	1.908	1.669	1.400	.978	.570	-.264	-1.190	-1.702	-2.132	-2.952
	30	2.426	2.160	1.940	1.698	1.427	1.002	.593	-.239	-1.154	-1.657	-2.080	-2.884
	40	2.479	2.209	1.986	1.740	1.465	1.036	.624	-.206	-1.106	-1.598	-2.010	-2.793
	50	2.517	2.244	2.018	1.770	1.493	1.059	.645	-.184	-1.075	-1.559	-1.965	-2.735
	60	2.544	2.269	2.042	1.792	1.513	1.077	.662	-.167	-1.052	-1.532	-1.933	-2.694
	70	2.567	2.290	2.061	1.810	1.529	1.091	.674	-.155	-1.035	-1.511	-1.909	-2.662
	80	2.585	2.307	2.077	1.824	1.543	1.103	.684	-.144	-1.021	-1.495	-1.890	-2.638
	90	2.600	2.321	2.090	1.836	1.553	1.112	.693	-.136	-1.010	-1.481	-1.874	-2.618
	100	2.613	2.333	2.101	1.847	1.563	1.120	.701	-.129	-1.001	-1.470	-1.861	-2.601

*

TABLE 9-1 (CONTINUED)
CONFIDENCE LIMIT DEVIATE VALUES FOR NORMAL DISTRIBUTION

EXCEEDANCE PROBABILITY

Confidence Level	Systematic Record Length N	.002	.005	.010	.020	.040	.100	.200	.500	.800	.900	.950	.990
.95	10	1.989	1.757	1.563	1.348	1.104	.712	.317	-.580	-1.702	-2.355	-2.911	-3.981
	15	2.121	1.878	1.677	1.454	1.203	.802	.406	-.455	-1.482	-2.068	-2.566	-3.520
	20	2.204	1.955	1.749	1.522	1.266	.858	.460	-.387	-1.370	-1.926	-2.396	-3.295
	25	2.264	2.011	1.801	1.569	1.309	.898	.497	-.342	-1.301	-1.838	-2.292	-3.158
	30	2.310	2.053	1.840	1.605	1.342	.928	.525	-.310	-1.252	-1.777	-2.220	-3.064
	40	2.375	2.113	1.896	1.657	1.391	.970	.565	-.266	-1.188	-1.697	-2.125	-2.941
	50	2.421	2.156	1.936	1.694	1.424	1.000	.592	-.237	-1.146	-1.646	-2.065	-2.862
	60	2.456	2.188	1.966	1.722	1.450	1.022	.612	-.216	-1.116	-1.609	-2.022	-2.807
	70	2.484	2.214	1.990	1.745	1.470	1.040	.629	-.199	-1.093	-1.581	-1.990	-2.765
	80	2.507	2.235	2.010	1.762	1.487	1.054	.642	-.186	-1.076	-1.559	-1.964	-2.733
	90	2.526	2.252	2.026	1.778	1.500	1.066	.652	-.175	-1.061	-1.542	-1.944	-2.706
	100	2.542	2.267	2.040	1.791	1.512	1.077	.662	-.166	-1.049	-1.527	-1.927	-2.684
.99	10	1.704	1.492	1.314	1.115	.886	.508	.107	-.892	-2.243	-3.048	-3.738	-5.074
	15	1.868	1.645	1.458	1.251	1.014	.629	.236	-.678	-1.841	-2.521	-3.102	-4.222
	20	1.974	1.743	1.550	1.336	1.094	.705	.313	-.568	-1.651	-2.276	-2.808	-3.832
	25	2.050	1.813	1.616	1.399	1.152	.757	.364	-.498	-1.536	-2.129	-2.633	-3.601
	30	2.109	1.867	1.667	1.446	1.196	.797	.403	-.450	-1.457	-2.030	-2.515	-3.447
	40	2.194	1.946	1.741	1.515	1.259	.854	.457	-.384	-1.355	-1.902	-2.364	-3.249
	50	2.255	2.002	1.793	1.563	1.304	.894	.496	-.340	-1.290	-1.821	-2.269	-3.125
	60	2.301	2.045	1.833	1.600	1.337	.924	.524	-.309	-1.244	-1.764	-2.202	-3.038
	70	2.338	2.079	1.865	1.630	1.365	.948	.545	-.285	-1.210	-1.722	-2.153	-2.974
	80	2.368	2.107	1.891	1.653	1.387	.968	.563	-.265	-1.183	-1.688	-2.114	-2.924
	90	2.394	2.131	1.913	1.674	1.405	.984	.578	-.250	-1.160	-1.661	-2.082	-2.883
	100	2.416	2.151	1.932	1.691	1.421	.998	.591	-.236	-1.142	-1.639	-2.056	-2.850

Appendix 10

RISK

This appendix describes the recommended procedures for estimating the risk incurred when a location is occupied for a period of years. As used in this guide, risk is defined as the probability that one or more events will exceed a given flood magnitude within a specified period of years.

Two basic approaches may be used to compute risk, nonparametric methods [(e.g., (19)] and parametric methods [(e.g., (20)]. Parametric methods which use the binomial distribution require assuming that the annual exceedance frequency is exactly known. The difference between methods is not great, particularly in the range of usual interest; consequently, use of the binomial distribution is recommended because of ease of comprehension and application.

The binomial expression for estimating risk is:

$$R_I = \frac{N!}{I!\,(N-I)!} P^I \, (1-P)^{N-I} \qquad (10\text{-}1)$$

in which R_I is the estimated risk of obtaining in N years exactly I number of flood events exceeding a flood magnitude with annual exceedance probability P.

When I equals 0 equation 10-1 reduces to:

$$R_o = (1-P)^N \qquad (10\text{-}2)$$

in which R_o is the estimated probability of nonexceedance of the selected flood magnitude in N years. From this the risk R of one or more exceedance becomes

$$R\ (1 \text{ or more}) = 1 - (1-P)^N \qquad (10\text{-}3)$$

Risk of 2 or more exceedances, R (2 or more), is

$$R(2 \text{ or more}) = R - R_1 = R - NP\,(1-P)^{N-1} \qquad (10\text{-}4)$$

* Some solutions are illustrated by the following table and figure * 10-1.

*

BINOMIAL RISK TABLE

TIME	** RISK (PERCENT) ** P=0.100 NONE	ONE OR MORE	TWO OR MORE	** RISK (PERCENT) ** P=0.050 NONE	ONE OR MORE	TWO OR MORE
10	35	65	26	60	40	9
20	12	88	61	36	64	26
30	4	96	82	21	79	45
40	1	99	92	13	87	60
50	1	99	97	8	92	72
60	0	100	99	5	95	81
70	0	100	99	3	97	87
80	0	100	100	2	98	91
90	0	100	100	1	99	94
100	0	100	100	1	99	96
110	0	100	100	0	100	98
120	0	100	100	0	100	98
150	0	100	100	0	100	100
200	0	100	100	0	100	100

TIME	** RISK (PERCENT) ** P=0.040 NONE	ONE OR MORE	TWO OR MORE	** RISK (PERCENT) ** P=0.020 NONE	ONE OR MORE	TWO OR MORE
10	66	34	6	82	18	2
20	44	56	19	67	33	6
30	29	71	34	55	45	12
40	20	80	48	45	55	19
50	13	87	60	36	64	26
60	9	91	70	30	70	34
70	6	94	78	24	76	41
80	4	96	83	20	80	48
90	3	97	88	16	84	54
100	2	98	91	13	87	60
110	1	99	94	11	89	65
120	1	99	96	9	91	69
150	0	100	98	5	95	80
200	0	100	100	2	98	91

NOTE: TABLE VALUES ARE ROUNDED TO NEAREST PERCENT

*

*

BINOMIAL RISK TABLE

TIME	** RISK (PERCENT) ** P=0.010			** RISK (PERCENT) ** P=0.005		
	NONE	ONE OR MORE	TWO OR MORE	NONE	ONE OR MORE	TWO OR MORE
10	90	10	0	95	5	0
20	82	18	2	90	10	0
30	74	26	4	86	14	1
40	67	33	6	82	18	2
50	61	39	9	78	22	3
60	55	45	12	74	26	4
70	49	51	16	70	30	5
80	45	55	19	67	33	6
90	40	60	23	64	36	8
100	37	63	26	61	39	9
110	33	67	30	58	42	11
120	30	70	34	55	45	12
150	22	78	44	47	53	17
200	13	87	60	37	63	26

TIME	** RISK (PERCENT) ** P=0.002			** RISK (PERCENT) ** P=0.001		
	NONE	ONE OR MORE	TWO OR MORE	NONE	ONE OR MORE	TWO OR MORE
10	98	2	0	99	1	0
20	96	4	0	98	2	0
30	94	6	0	97	3	0
40	92	8	0	96	4	0
50	90	10	0	95	5	0
60	89	11	1	94	6	0
70	87	13	1	93	7	0
80	85	15	1	92	8	0
90	84	16	1	91	9	0
100	82	18	2	90	10	0
110	80	20	2	90	10	1
120	79	21	2	89	11	1
150	74	26	4	86	14	1
200	67	33	6	82	18	2

NOTE: TABLE VALUES ARE ROUNDED TO NEAREST PERCENT

*

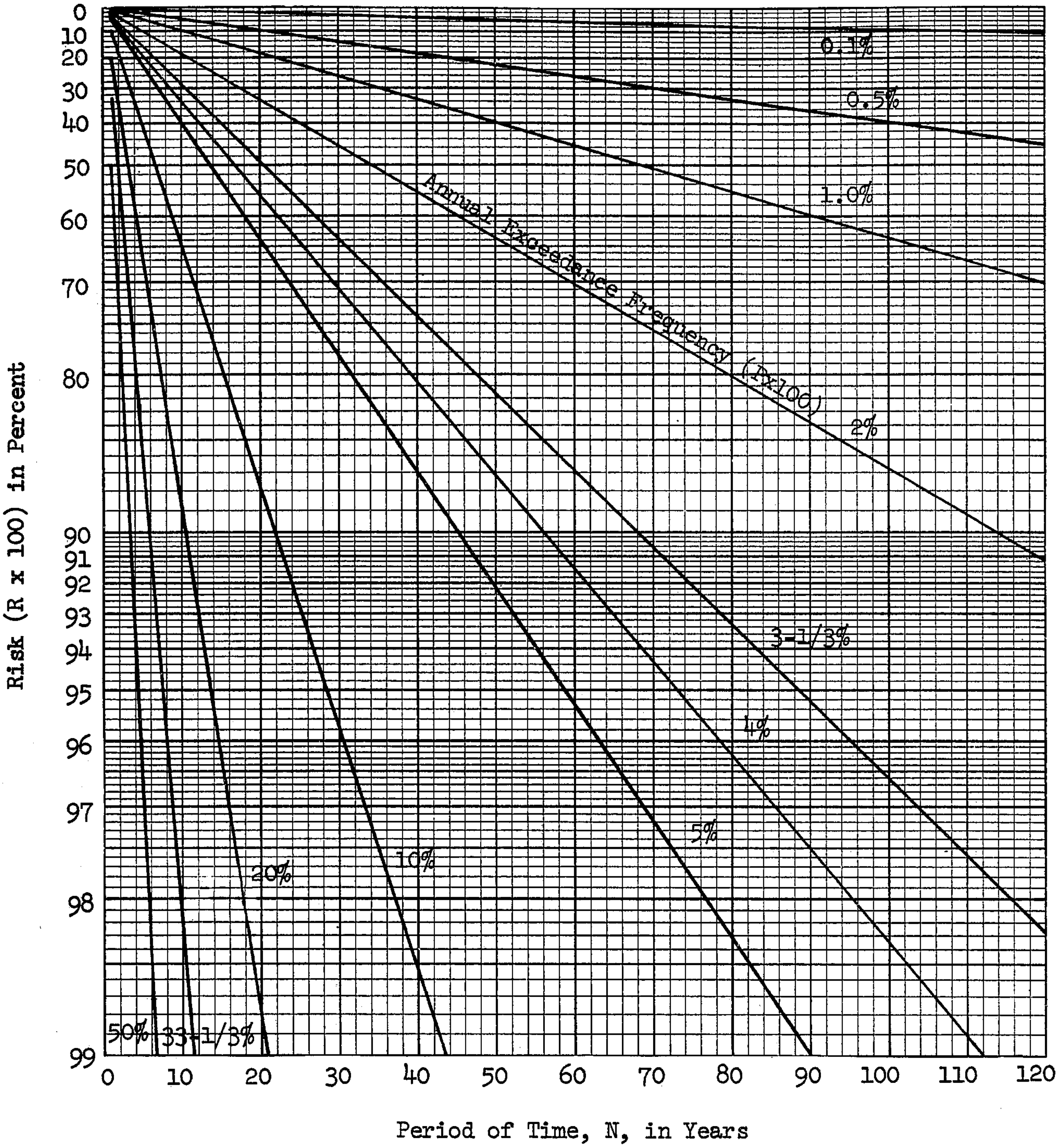


Figure 10-1. RISK OF ONE OR MORE FLOOD EVENTS EXCEEDING A FLOOD OF GIVEN ANNUAL EXCEEDANCE FREQUENCY WITHIN A PERIOD OF YEARS

Appendix 11

EXPECTED PROBABILITY

The principle of gambling based upon estimated probabilities can be applied to water resources development decisions. However, because probabilities must be inferred from random sample data, they are uncertain and mathematical expectation cannot be computed exactly as errors due to uncertainty do not necessarily compensate. For example, if the estimate based on sample data is that a certain flood magnitude will be exceeded on the average once in 100 years, it is possible that the true exceedance could be three or four more times per hundred years, but it can never be less than zero times per hundred years. The impact of errors in one direction due to uncertainty can be quite different from the impact of errors in the other direction. Thus, it is not adequate to simply be too high half the time and too low the other half. It is necessary to consider the relative impacts of being too high or too low.

It is possible to delineate uncertainty with considerable accuracy when dealing with samples from a normal distribution. Therefore, when flood flow frequency curves conform fairly closely to the logarithmic normal distribution, it is possible to delineate uncertainty of frequency or probability estimates of flood flows.

Figure 11-1 is a generalized representation of the range of uncertainty in probability estimates based on samples drawn from a normal population. The vertical scale can represent the logarithm of streamflow. The curves show the likelihood that the true frequency of any flood magnitude exceeds the value shown on the frequency scale. The curve labeled .50 is the curve that would be used for the best frequency estimate of a log-normal population. From this curve a magnitude of 2 would be exceeded on the average 30 times per thousand events. The figure also shows a 5 percent chance that the true frequency is 150 or more times per thousand or a 5 percent chance that the true frequency is two times or less per thousand events.

If a magnitude of 2.0 were selected at 20 independent locations, the best estimate for the frequency is 3 exceedances per hundred years for each location. The estimated total exceedance for all 20 locations would be 60 per 100 years. However, due to sampling uncertainties, true frequencies for a magnitude of 2.0 would differ at each location and total exceedances per 100 years at the 20 locations might be represented by the following tabulation.

Exceedances Per 100 Years at Each of 20 Locations*

20	5	3	.9	
12	5	2	.8	
10	4	2	.5	Total Exceedances = Approximately 90
8	4	2	.3	
7	3	1	.1	

*Determined from Figure 11-1 using 0.05 parameter value increments from .025 through .975.

The total of these exceedances is about 90 per 100 years or 30 more than obtained using the best probability estimate as the true probability at each location. If, however, the mathematically derived expected probability function were used instead of the traditional "best" estimate we could read the expected probability curve of Figure 11-1 to obtain the value of about 4.5 exceedances per 100 events. This value when applied to each of the 20 locations would give an estimate of 90 exceedances per 100 years at all 20 locations. Thus, while the expected probability estimate would be wrong in the high direction more frequently than in the low direction, the heavier impacts of being wrong in the low direction would compensate for this. It can be noted, at this point, that expected probability is the average of all estimated true probabilities.

If a flood frequency estimate could be accurately known--that is, the parent population could be defined--the frequency distribution of observed flood events would approach the parent population as the number of observations approaches infinity. This is not the case where probabilities are not accurately known. However, if the expected probabilities as illustrated in Figure 11-1 can be computed, observed flood frequency for a large number of independent locations will approach the estimated flood frequency as the number of observations approaches infinity and the number of locations approaches infinity.

It appears that the answer to the question as to whether expected probability should be used at a single location would be identical to the answer to the question, "What is a fair wager for a single gamble?" If the gamble must be undertaken, and ordinarily it must, then the answer to the above question is that the wager should be proportional to the expected return. In determining whether the expected probability concepts should apply for a single location, the same line of reasoning would indicate that it should.

It has been shown (21) that for the normal distribution the expected probability P_N can be obtained from the formula

$$P_N = \text{Prob}\left[t_{N-1} > K_n \left(\frac{N}{N+1}\right)^{1/2}\right] \qquad (11\text{-}1)$$

where K_n is the standard normal variate of the desired probability of exceedance, N is the sample size, and t_{N-1} is the Student's t-statistic with N-1 degrees of freedom.

The actual calculations can be carried out using tables of the t-statistic, or the modified values shown in Table 11-1 (31). To use Table 11-1, enter with the sample size minus 1 and read across to the column with the desired exceedance probability. The value read from the table is the corrected plotting position.

The expected probability correction may also be calculated from the following equations (34) which are based on Table 11-1. For selected exceedance probabilities greater than 0.50, and a given sample size, the appropriate P_N value equals 1 minus the value in Table 11-1 or the equations 11-2.

Exceedance Probability	Expected Probability, P_N	
.0001	$.0001\ (1.0 + 1600/N^{1.72})$	(11-2a)
.001	$.001\ (1.0 + 280/N^{1.55})$	(11-2b)
.01	$.01\ (1.0 + 26/N^{1.16})$	(11-2c)
.05	$.05\ (1.0 + 6/N^{1.04})$	(11-2d)
.10	$.1\ (1.0 + 3/N^{1.04})$	(11-2e)
.30	$.3\ (1.0 + 0.46/N^{0.925})$	(11-2f)

For floods with an exceedance probability of 0.01 based on samples of 20 annual peaks, for example, the expected probability of exceedance from equation 11-2c is (.01) (1.0 + 26/32.3) or 0.018. Use of Table 11-1 gives 0.0174. Comparable equations for adjusting the computed discharge upward to give a discharge for which the expected probability equals the exceedance probability are available (22).

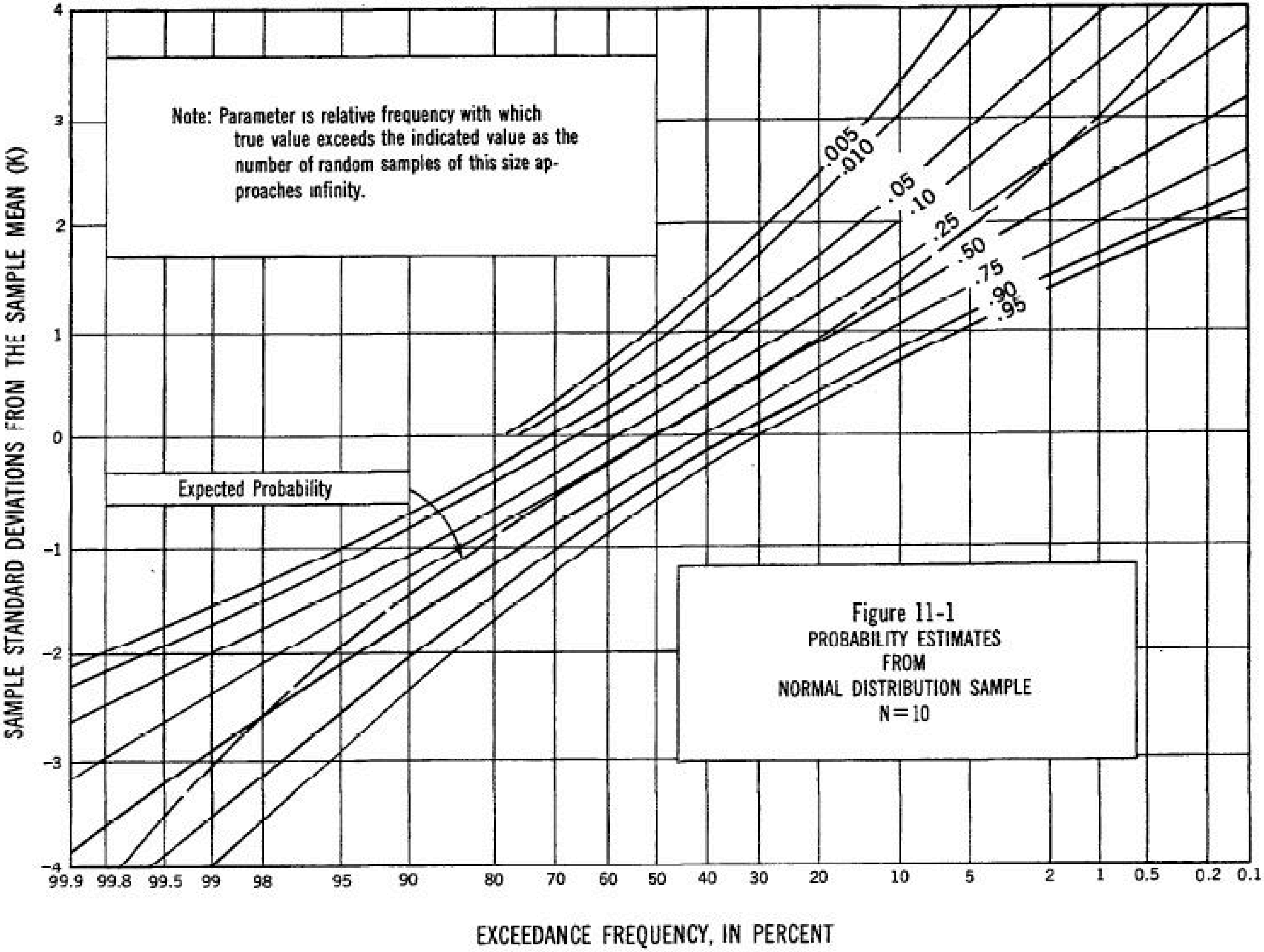


Figure 11-1
PROBABILITY ESTIMATES
FROM
NORMAL DISTRIBUTION SAMPLE
N=10

Table 11-1

TABLE OF P_N VERSUS P_∞

For use with samples drawn from a normal population

N-1 \ P_∞	.50	.30	.10	.05	.01	.001	.0001
1	.500	.372	.243	.204	.154	.121	.102
2	.500	.347	.193	.146	.090	.057	.043
3	.500	.336	.169	.119	.064	.035	.023
4	.500	.330	.154	.104	.050	.024	.0137
5	.500	.325	.146	.094	.042	.0179	.0092
6	.500	.322	.138	.088	.036	.0138	.0066
7	.500	.319	.135	.083	.032	.0113	.0050
8	.500	.317	.131	.079	.029	.0094	.0039
9	.500	.316	.127	.076	.027	.0082	.0031
10	.500	.315	.125	.073	.025	.0072	.0025
11	.500	.314	.123	.071	.023	.0064	.0021
12	.500	.313	.121	.069	.022	.0058	.0018
13	.500	.312	.119	.068	.021	.0052	.0016
14	.500	.311	.118	.067	.020	.0048	.0014
15	.500	.311	.117	.066	.0196	.0045	.0013
16	.500	.310	.116	.065	.0190	.0042	.0012
17	.500	.310	.115	.064	.0184	.0040	.0011
18	.500	.309	.114	.063	.0179	.0038	.0010
19	.500	.309	.113	.062	.0174	.0036	.00091
20	.500	.308	.113	.062	.0170	.0034	.00084
21	.500	.308	.112	.061	.0167	.0033	.00078
22	.500	.308	.111	.061	.0163	.0031	.00073
23	.500	.307	.111	.060	.0161	.0030	.00068
24	.500	.307	.110	.060	.0158	.0029	.00064
25	.500	.307	.110	.059	.0155	.0028	.00060
26	.500	.306	.109	.059	.0153	.0027	.00057
27	.500	.306	.109	.059	.0151	.0026	.00054
28	.500	.306	.109	.058	.0149	.0026	.00051
29	.500	.306	.108	.058	.0147	.0025	.00049
30	.500	.306	.108	.058	.0145	.0024	.00046
40	.500	.304	.106	.056	.0133	.0020	.00034
60	.500	.303	.104	.054	.0122	.0016	.00025
120	.500	.302	.102	.052	.0111	.0013	.00017
∞	.500	.300	.100	.050	.0100	.0010	.00010

NOTE: P_N values above are usable approximately with Pearson Type III distributions having small skew coefficients.

*

Appendix 12

FLOW DIAGRAM AND EXAMPLE PROBLEMS

*

The sequence of procedures recommended by this guide for defining flood potentials (except for the case of mixed populations) is described in the following outline and flow diagrams.

A. Determine available data and data to be used.
   1. Previous studies
   2. Gage records
   3. Historic data
   4. Studies for similar watersheds
   5. Watershed model

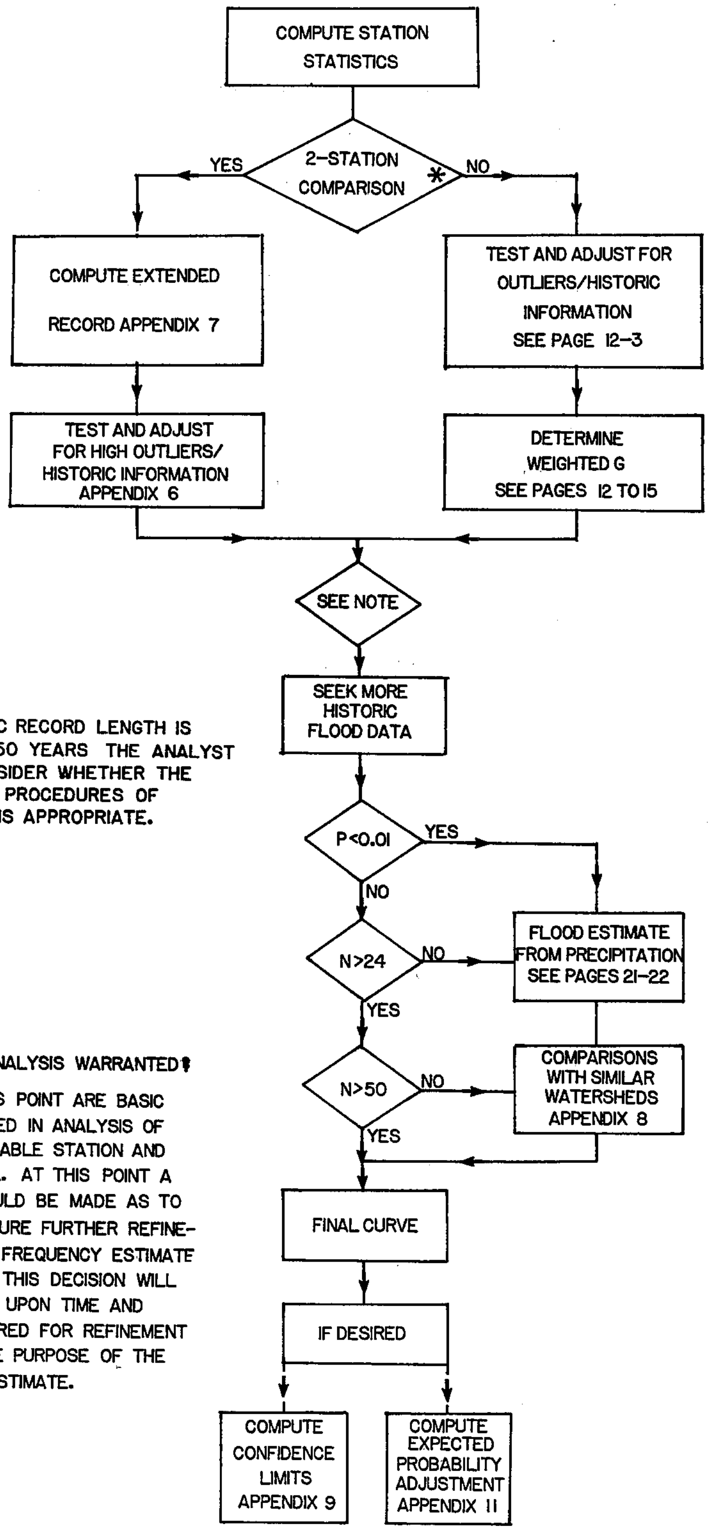
B. Evaluate data.
   1. Record homogeneity
   2. Reliability and accuracy

c. Compute curve following guide procedures as outlined in following flow diagrams. Example problems showing most of the computational techniques follow the flow diagram.

*

ZERO FLOOD OR INCOMPLETE RECORD | COMPLETE RECORD

SEE APPENDIX 5, CONDITIONAL PROBABILITY ADJUSTMENT, FOR OUTLIERS SEE PAGES 17 TO 19 AND APPENDIX 5 AND 6



* IF SYSTEMATIC RECORD LENGTH IS LESS THAN 50 YEARS THE ANALYST SHOULD CONSIDER WHETHER THE USE OF THE PROCEDURES OF APPENDIX 7 IS APPROPRIATE.

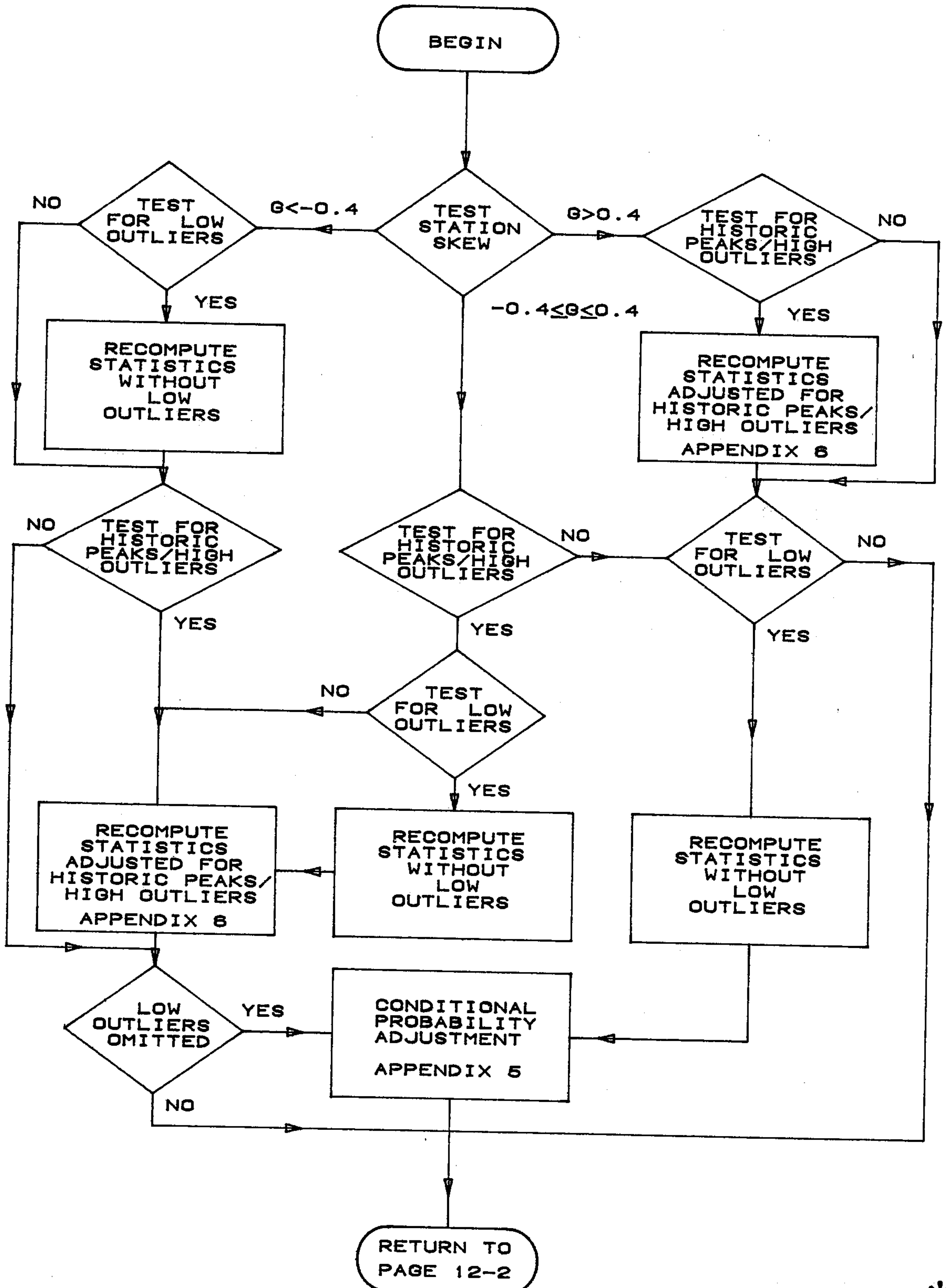
NOTE: IS FURTHER ANALYSIS WARRANTED?

STEPS TO THIS POINT ARE BASIC STEPS REQUIRED IN ANALYSIS OF READILY AVAILABLE STATION AND HISTORIC DATA. AT THIS POINT A DECISION SHOULD BE MADE AS TO WHETHER FUTURE FURTHER REFINEMENT OF THE FREQUENCY ESTIMATE IS JUSTIFIED. THIS DECISION WILL DEPEND BOTH UPON TIME AND EFFORT REQUIRED FOR REFINEMENT AND UPON THE PURPOSE OF THE FREQUENCY ESTIMATE.

FLOW DIAGRAM FOR FLOOD FLOW FREQUENCY ANALYSIS

*

* FLOW DIAGRAM FOR HISTORIC AND OUTLIER ADJUSTMENT



*

The following examples illustrate application of most of the techniques recommended in this guide. Annual flood peak data for four stations (Table 12-1) have been selected to illustrate the following:

1. Fitting the Log-Pearson Type III distribution
2. Adjusting for high outliers
3. Testing and adjusting for low outliers
4. Adjusting for zero flood years

The procedure for adjusting for historic flood data is given in Appendix 6 and an example computation is provided. An example has not been included specifically for the analysis of an incomplete record as this technique is applied in Example 4, adjusting for zero flood years. The computation of confidence limits and the adjustment for expected probability are described in Example 1. The generalized *skew coefficient used in these examples was taken from Plate I. In actual practice, the generalized skew may be obtained from other sources or a special study made for the region. *

Because of round off errors in the computational procedures, computed values may differ beyond the second decimal point.

* These examples have been completely revised using the procedures recommended in Bulletin 17B. Specific changes have not been indicated on the following pages: *

TABLE 12-1

ANNUAL FLOOD PEAKS FOR FOUR STATIONS IN EXAMPLES

Year	Fishkill Creek 01-3735 Example 1	Floyd River 06-6005 Example 2	Back Creek 01-6140 Example 3	Orestimba Creek 11-2745 Example 4
1929			8750	
1930			15500	
1931			4060	
1932				4260
1933				345
1934				516
1935		1460		1320
1936		4050	22000*	1200
1937		3570	-	2180
1938		2060	-	3230
1939		1300	6300	115
1940		1390	3130	3440
1941		1720	4160	3070
1942		6280	6700	1880
1943		1360	22400	6450
1944		7440	3880	1290
1945	2290	5320	8050	5970
1946	1470	1400	4020	782
1947	2220	3240	1600	0
1948	2970	2710	4460	0
1949	3020	4520	4230	335
1950	1210	4840	3010	175
1951	2490	8320	9150	2920
1952	3170	13900	5100	3660
1953	3220	71500	9820	147
1954	1760	6250	6200	0
1955	8800	2260	10700	16
1956	8280	318	3880	5620
1957	1310	1330	3420	1440
1958	2500	970	3240	10200
1959	1960	1920	6800	5380
1960	2140	15100	3740	448
1961	4340	2870	4700	0
1962	3060	20600	4380	1740
1963	1780	3810	5190	8300
1964	1380	726	3960	156
1965	980	7500	5600	560
1966	1040	7170	4670	128
1967	1580	2000	7080	4200
1968	3630	829	4640	0
1969		17300	536	5080
1970		4740	6680	1010
1971		13400	8360	584
1972		2940	18700	0
1973		5660	5210	1510

*Not included in example computations.

EXAMPLE 1

FITTING THE LOG-PEARSON TYPE III DISTRIBUTION

a. Station Description

Fishkill Creek at Beacon, New York

USGS Gaging Station: 01-3735
Lat: 41°30'42", long: 73°56'58"
Drainage Area: 190 sq. mi.
Annual Peaks Available: 1945-1968

b. Computational Procedures

Step 1 - List data, transform to logarithms, and compute the squares and the cubes.

TABLE 12-2
COMPUTATION OF SUMMATIONS

Year	Annual Peak (cfs)	Logarithm (X)	X^2	X^3
1945	2290	3.35984	11.28852	37.92764
1946	1470	3.16732	10.03192	31.77429
1947	2220	3.34635	11.19806	37.47262
1948	2970	3.47276	12.06006	41.88170
1949	3020	3.48001	12.11047	42.14456
1950	1210	3.08279	9.50359	29.29759
1951	2490	3.39620	11.53417	39.17236
1952	3170	3.50106	12.25742	42.91397
1953	3220	3.50786	12.30508	43.16450
1954	1760	3.24551	10.53334	34.18604
1955	8800	3.94448	15.55892	61.37186
1956	8280	3.91803	15.35096	60.14552
1957	1310	3.11727	9.71737	30.29167
1958	2500	3.39794	11.54600	39.23260
1959	1960	3.29226	10.83898	35.68473
1960	2140	3.33041	11.09163	36.93968
1961	4340	3.63749	13.23133	48.12884
1962	3060	3.48572	12.15024	42.35235
1963	1780	3.25042	10.56523	34.34144
1964	1380	3.13988	9.85885	30.95559
1965	980	2.99123	8.94746	26.76390
1966	1040	3.01703	9.10247	27.46243
1967	1580	3.19866	10.23143	32.72685
1968	3630	3.55991	12.67296	45.11459
N=24	--	Σ 80.84043	273.68646	931.44732

Example 1 - Fitting the Log-Pearson Type III Distribution (continued)

Step 2 - Computation of mean by Equation 2:

$$\bar{X} = \frac{\Sigma X}{N}$$

$$= \frac{80.84043}{24} = 3.3684 \qquad (12\text{-}1)$$

Computation of standard deviation by Equation 3b:

$$S = \left[\frac{\Sigma X^2 - (\Sigma X)^2/N}{N-1}\right]^{0.5}$$

$$S = \left[\frac{273.68646 - (80.84043)^2/24}{23}\right]^{0.5} \qquad (12\text{-}2)$$

$$S = \sqrt{\frac{1.38750}{23}} = 0.2456$$

Computation of skew coefficient by Equation 4b:

$$G = \frac{N^2(\Sigma X^3) - 3N(\Sigma X)(\Sigma X^2) + 2\,(\Sigma X)^3}{N(N-1)(N-2)S^3}$$

$$= \frac{(24)^2(931.44732) - 3(24)(80.84043)(273.68646) + 2(80.84043)^3}{24(24-1)\qquad(24-2)\qquad(.24561)^3}$$

$$= \frac{536513.6563 - 1592995.0400 + 1056612.7341}{(24)(23)(22)\qquad(.014816)} \qquad (12\text{-}3)$$

$$= \frac{131.3504}{179.9285} = 0.7300$$

Example 1 - Fitting the Log-Pearson Type III Distribution (continued)

Step 3 - Check for Outliers:

$$X_H = \overline{X} + K_N S$$

$$= 3.3684 + 2.467\ (.2456) \quad = 3.9743 \qquad (12\text{-}4)$$

$$Q_H = \text{antilog}\ (3.9743) \quad = \quad 9425 \text{ cfs}$$

The largest recorded value does not exceed the threshold value. Next, the test for detecting possible low outliers is applied. The same K_N value is used in equation 8a to compute the low outlier threshold (Q_L):

$$X_L = \overline{X} - K_N S$$

$$= 3.3684 - 2.467(.2456) \quad = 2.7625 \qquad (12\text{-}5)$$

$$Q_L = \text{antilog}\ (2.7625) \quad = \quad 579 \text{ cfs}$$

There are no recorded values below this threshold value. No outliers were detected by either the high or low tests. For this example a generalized skew of 0.6 is determined from Plate I. In actual practice a generalized skew may be obtained from other sources or from a special study made for the region. A weighted skew is computed by use of Equation 5. The mean square error of the station skew can be found within Table 1 or computed by Equation 6. Computation of mean-square error of station skew by Eq. 6:

$$MSE_G \simeq 10^{\left[A - B\ [\log_{10}(N/10)]\right]}$$

Where:

$$A = -0.33 + 0.08\ |G| = -0.33 + 0.08(.730) \quad = -.2716 \qquad (12\text{-}6)$$

$$B = 0.94 - 0.26\ |G| \quad = \quad 0.94 - 0.26(.730) = \quad .7502 \qquad (12\text{-}7)$$

$$MSE_G \simeq 10^{\left[-.2716 - .7502\ [\log_{10}(2.4)]\right]} \simeq 10^{-.55683} \simeq 0.277 \qquad (12\text{-}8)$$

The mean-square error of the generalized skew from Plate I is 0.302.

Computation of weighted skew by equation 5:

$$G_W = \frac{MSE_{\overline{G}}\,(G) + MSE_G\,(\overline{G})}{MSE_{\overline{G}} + MSE_G}$$

$$= \frac{.302(.7300) + .277(.6)}{.579} = 0.6678 \qquad (12\text{-}9)$$

$$= 0.7 \text{ (rounded to nearest tenth)}$$

Step 4 - Compute the frequency curve coordinates.

The log-Pearson Type III K values for a skew coefficient of 0.7 are found in Appendix 3. An example computation for an exceedance probability of .01 using Equation 1 follows:

$$\log Q = \overline{X} + KS = 3.3684 + 2.82359(.2456) = 4.0619 \qquad (12\text{-}10)$$

$$Q = 11500 \text{ cfs}$$

The discharge values in this computation and those in Table 12-3 are rounded to three significant figures.

TABLE 12-3

COMPUTATION OF FREQUENCY CURVE COORDINATES

P	$K_{G_W,P}$ for $G_W = 0.7$	log Q	Q cfs
.99	-1.80621	2.9247	841
.90	-1.18347	3.0777	1200
.50	-0.11578	3.3399	2190
.10	1.33294	3.6957	4960
.05	1.81864	3.8150	6530
.02	2.40670	3.9595	9110
.01	2.82359	4.0619	11500
.005	3.22281	4.1599	14500
.002	3.72957	4.2844	19200

The frequency curve is plotted in Figure 12-1.

Step 5 - Compute the confidence limits.

The upper and lower confidence limits for levels of significance of .05 and .95 percent are computed by the procedures outlined in Appendix 9. Nine exceedance probabilities (P) have been selected to define the confidence limit curves. The computations for two points on the curve at an exceedance probability of 0.99 are given below.

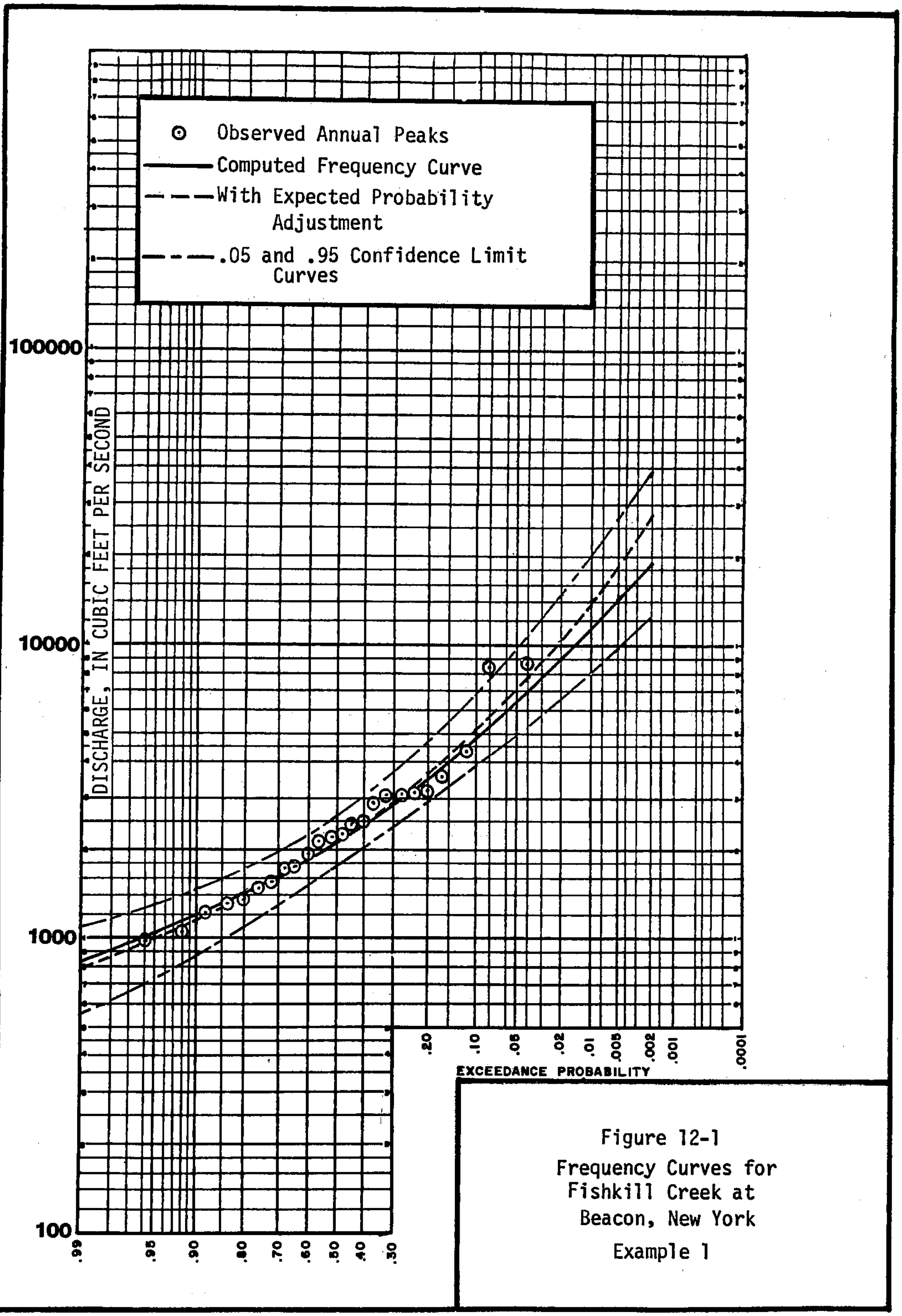


Figure 12-1
Frequency Curves for Fishkill Creek at Beacon, New York
Example 1

Example 1 - Fitting the Log-Pearson Type III Distribution (continued)

Equations in Appendix 9 are used in computing an approximate value for $K_{P,c}$. The normal deviate, z_c, is found by entering Appendix 3 with a skew coefficient of zero. For a confidence level of 0.05, $z_c = 1.64485$. The Pearson Type III deviates, $K_{G_w,P}$ are found in Appendix 3 based on the appropriate skew coefficient. For an exceedance probability of 0.99 and skew coefficient of 0.7, $K_{G_w,P} = -1.80621$.

$$a = 1 - \frac{z_c^2}{2(N-1)} = 1 - \frac{(1.64485)^2}{2(24-1)} = 0.9412 \qquad (12\text{-}11)$$

$$b = K^2_{G_w,P} - \frac{z_c^2}{N} = (-1.80621)^2 - \frac{(1.64485)^2}{24} = 3.1497 \qquad (12\text{-}12)$$

$$K^U_{P,c} = \frac{K_{G_w,P} + \sqrt{K^2_{G_w,P} - ab}}{a} = \frac{-1.80621 + \sqrt{(-1.80621)^2 - (.9412)(3.1497)}}{.9412} \qquad (12\text{-}13)$$

$$= \frac{-1.80621 + .5458}{.9412} = -1.3392$$

The discharge value is:

$$\text{Log } Q = 3.3684 + (-1.3392)(.2456) \qquad (12\text{-}14)$$

$$= 3.0395$$

$$Q = 1100$$

For the lower confidence coefficient:

$$K^L_{P,c} = \frac{K_{G_w,P} - \sqrt{K^2_{G_w,P} - ab}}{a} = \frac{-1.80621 - .5458}{.9412} = -2.4989 \qquad (12\text{-}15)$$

The discharge value is:

$$\text{Log } Q = 3.3684 + (-2.4989)(.2456) \tag{12-16}$$

$$= 2.7546$$

$$Q = 568$$

The computations showing the derivation of the upper and lower confidence limits are given in Table 12-4. The resulting curves are shown in Figure 12-1.

TABLE 12-4

COMPUTATION OF CONFIDENCE LIMITS

P	$K_{G_W,P}$ for $G_W = 0.7$	0.05 UPPER LIMIT CURVE $K^U_{P,c}$	log Q	Q cfs	0.05 LOWER LIMIT CURVE $K^L_{P,c}$	log Q	Q cfs
.99	-1.80621	-1.3392	3.0395	1100	-2.4989	2.7546	568
.90	-1.18347	-0.7962	3.1728	1490	-1.7187	2.9462	884
.50	-0.11578	0.2244	3.4235	2650	-0.4704	3.2528	1790
.10	1.33294	1.9038	3.8359	6850	0.9286	3.5964	3950
.05	1.81864	2.5149	3.9860	9680	1.3497	3.6998	5010
.02	2.40670	3.2673	4.1708	14800	1.8469	3.8220	6640
.01	2.82359	3.8058	4.3031	20100	2.1943	3.9073	8080
.005	3.22281	4.3239	4.4303	26900	2.5245	3.9884	9740
.002	3.72957	4.9841	4.5925	39100	2.9412	4.0907	12300

Step 6 - Compute the expected probability adjustment.

The expected probability plotting positions are determined from Table 11-1 based on N - 1 of 23.

TABLE 12-5

EXPECTED PROBABILITY ADJUSTMENT

P	Q	Expected Probability
.99	841	.9839
.90	1200	.889
.50	2190	.50
.10	4960	.111
.05	6530	.060
.02	9110	.028*
.01	11500	.0161
.005	14500	.0095*
.002	19200	.0049*

*Interpolated values

The frequency curve adjusted for expected probability is shown in Figure 12-1.

EXAMPLE 2

ADJUSTING FOR A HIGH OUTLIER

a. Station Description

Floyd River at James, Iowa

USGS Gaging Station: 06-6005
Lat: 42°34'30", long: 96° 18'45"
Drainage Area: 882 sq. mi.
Annual Peaks Available: 1935-1973

b. Computational Procedures

Step 1 - Compute the statistics.

The detailed computations for the systematic record 1935-1973 have been omitted; the results of the computations are:

Mean Logarithm	3.5553
Standard Deviation of logs	0.4642
Skew Coefficient of logs	0.3566
Years	39

At this point, the analyst may wish to see the preliminary frequency curve based on the statistics of the systematic record. Figure 12-2 is the preliminary frequency curve based on the computed mean and standard deviation and a weighted skew of 0.1 (based on a generalized skew of -0.3 from Plate I).

Step 2 - Check for Outliers.

The station skew is between $\pm$ 0.4; therefore, the tests for both high outliers and low outliers are based on the systematic record statistics before any adjustments are made. From Appendix 4, the K_N for a sample size of 39 is 2.671.

The high outlier threshold (Q_H) is computed by Equation 7:

$$X_H = \overline{X} + K_N S$$

$$= 3.5553 + 2.671(.4642) = 4.7952 \qquad (12\text{-}17)$$

$$Q_H = \text{antilog } (4.7952) \qquad = 62400 \text{ cfs}$$

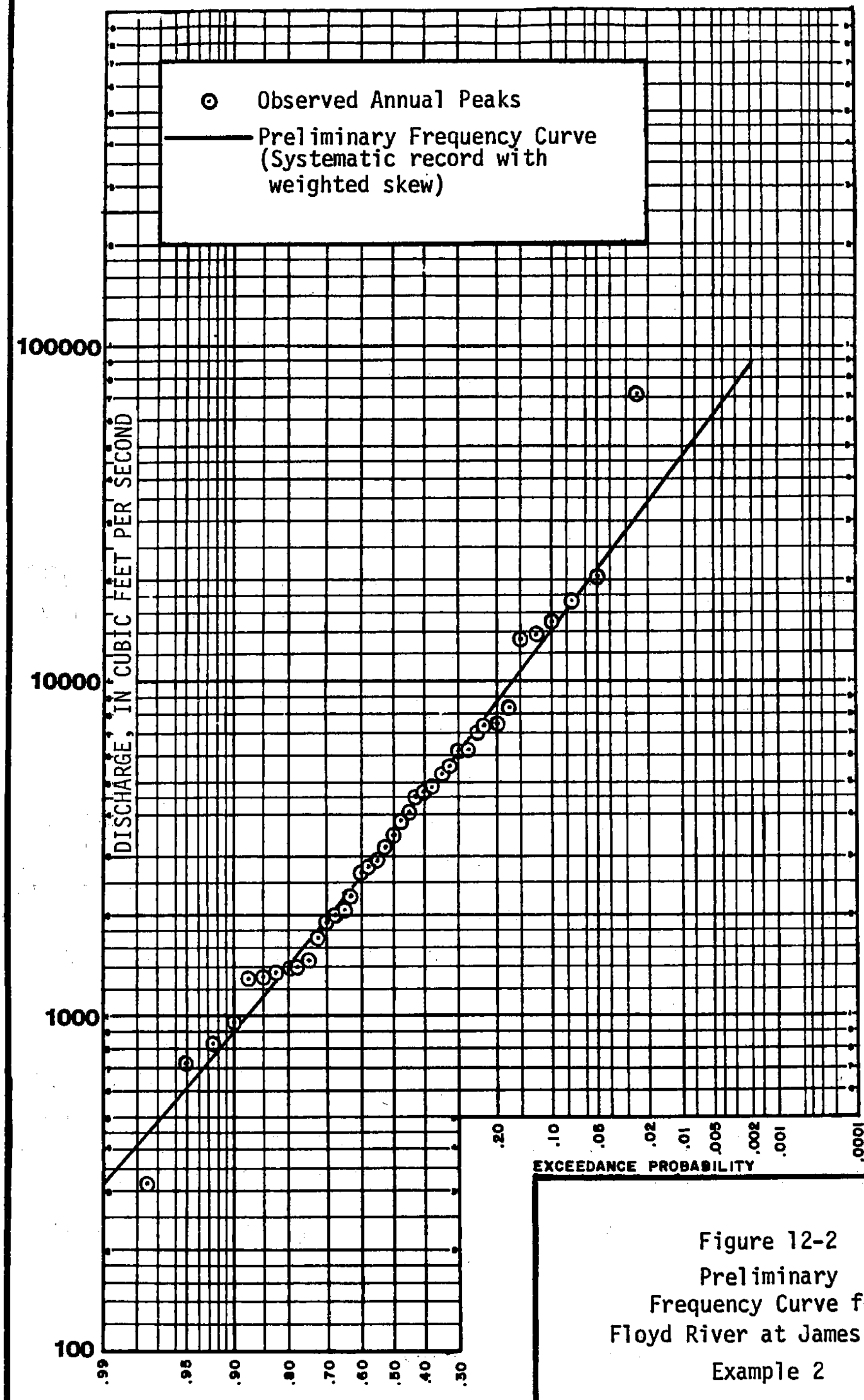


Figure 12-2
Preliminary
Frequency Curve for
Floyd River at James, Iowa
Example 2

The 1953 value of 71500 exceeds this value. Information from local residents indicates that the 1953 event is known to be the largest event since 1892; therefore, this event will be treated as a high outlier. If such information was not available, comparisons with nearby stations may have been desirable.

The low-outlier threshold (Q_L) is computed by Equation 8a:

$$X_L = \overline{X} - K_N S$$

$$= 3.5553 - 2.671(.4642) = 2.3154 \qquad (12\text{-}18)$$

$$Q_L = \text{antilog}\ (2.3154) = 207 \text{ cfs}$$

There are no values below this threshold value.

Step 3 - Recompute the statistics.

The 1953 value is deleted and the statistics recomputed from the remaining systematic record:

Mean Logarithm	3.5212
Standard Deviation of logs	0.4177
Skew Coefficient of logs	-0.0949
Years	38

Step 4 - Use historic data to modify statistics and plotting positions.

Application of the procedures in Appendix 6 allows the computed statistics to be adjusted by incorporation of the historic data.

(1) The historic period (H) is 1892-1973 or 82 years and the number of low values excluded (L) is zero.

(2) The systematic period (N) is 1935-1973 (with 1953 deleted) or 38 years.

(3) There is one event (Z) known to be the largest in 82 years.

(4) Compute weighting factor (W) by Equation 6-1:

$$W = \frac{H - Z}{N + L}$$

$$= \frac{82 - 1}{38 + 0} = 2.13158 \qquad (12\text{-}19)$$

Compute adjusted mean by Equation 6-2b:

$$\tilde{M} = \frac{WNM + \Sigma X_z}{H-WL}$$

$$\overline{X} = M = 3.5212$$

$$WNM = 285.2173$$

$$\Sigma X_z = \underline{4.8543}$$

$$290.0716$$

$$\tilde{M} = 290.0716/(82-0) = 3.5375 \qquad (12\text{-}20)$$

Compute adjusted standard deviation by Equation 6-3b:

$$\tilde{S}^2 = \frac{W(N-1)S^2 + WN(M-\tilde{M})^2 + \Sigma(X_z - \tilde{M})^2}{H-WL-1}$$

$$S = .4177$$

$$W(N-1)S^2 = 13.7604$$

$$WN(M-\tilde{M})^2 = .0215$$

$$\Sigma(X_z-\tilde{M})^2 = \underline{1.7340}$$

$$15.5159$$

$$\tilde{S}^2 = \frac{15.5159}{82-0-1} = .1916 \qquad (12\text{-}21)$$

$$\tilde{S} = .4377$$

Compute adjusted skew:

First compute adjusted skew on basis of record by Equation 6-4b:

$$\tilde{G} = \frac{H - WL}{(H-WL-1)(H-WL-2)\tilde{S}^3}\left[\frac{W(N-1)(N-2)S^3G}{N} + 3W(N-1)(M-\tilde{M})S^2 + WN(M-\tilde{M})^3 + \Sigma(X_z - \tilde{M})^3\right]$$

$$G = -0.0949$$

$$\frac{W(N-1)(N-2)S^3G}{N} = -.5168$$

752

$$3W(N-1)(M-\tilde{M})S^2 = -.6729$$

886

$$WN(M-\tilde{M})^3 = -.0004$$

351

$$\Sigma(X_z-\tilde{M})^3 = \frac{2.2833}{1.0932}$$

281

92

$$\frac{H}{(H-WL-1)(H-WL-2)\tilde{S}^3} = .150907 \qquad (12\text{-}22)$$

$$\tilde{G} = .1509\ (1.0932) = .1650$$

.164985

Next compute weighted skew:

For this example, a generalized skew of -0.3 is determined from Plate I. Plate I has a stated mean-square error of 0.302. Interpolating in Table I, the mean-square error of the station skew, based on H of 82 years, is 0.073. The weighted skew is computed by use of Equation 5:

$$G_W = \frac{.302(.1650) + .073(-.3)}{.302 + .073} = 0.0745 \qquad (12\text{-}23)$$

$G_W = 0.1$ (rounded to nearest tenth)

Example 2 - Adjusting for High Outlier (continued)

Step 5 - Compute adjusted plotting positions for historic data.

For the largest event (Equation 6-6):

$$\tilde{m}_1 = 1$$

For the succeeding events (Equation 6-7):

$$\tilde{m} = W\,E - (W-1)(Z + 0.5)$$

$$\tilde{m}_2 = 2.1316(2) - (2.1316-1)(1 + .5) \qquad (12\text{-}24)$$

$$= 2.5658$$

For the Weibull Distribution a = 0; therefore, by Equation 6-8

$$\tilde{PP} = \frac{\tilde{m}}{H + 1}(100)$$

$$\tilde{PP}_1 = \frac{1}{82 + 1}(100) = 1.20 \qquad (12\text{-}25)$$

$$\tilde{PP}_2 = \frac{2.5658}{83}(100) = 3.09 \qquad (12\text{-}26)$$

Exceedance probabilities are computed by dividing values obtained from Equation 12-26 by 100.

$$\frac{3.09}{100} = .0309$$

TABLE 12-6

COMPUTATION OF PLOTTING POSITIONS

Year	Q	W	Event Number E	Weighted Order $\tilde{m}$	Weibull Plotting Position Percent Chance $\tilde{PP}$	Exceedance Probability $\tilde{PP}$
1953	71500	1.0000	1	1.0000	1.20	.0120
1962	20600	2.1316	2	2.5658	3.09	.0309
1969	17300	2.1316	3	4.6974	5.66	.0566
1960	15100	2.1316	4	6.8290	8.23	.0823
1952	13900	2.1316	5	8.9606	10.80	.1080
1971	13400	2.1316	6	11.0922	13.36	.1336
1951	8320	2.1316	7	13.2238	15.93	.1593
1965	7500	2.1316	8	15.3554	18.50	.1850
1944	7440	2.1316	9	17.4870	21.07	.2107
1966	7170	2.1316	10	19.6186	23.64	.2364

Only the first 10 values are shown for this example

Step 6 - Compute the frequency curve.

TABLE 12-7

COMPUTATION OF FREQUENCY CURVE COORDINATES

P	$K_{G_W,P}$ for $G_W = 0.1$	log Q	Q cfs
.99	-2.25258	2.5515	356
.90	-1.27037	2.9815	958
.50	-0.01662	3.5302	3390
.10	1.29178	4.1029	12700
.05	1.67279	4.2697	18600
.02	2.10697	4.4597	28800
.01	2.39961	4.5878	38700
.005	2.66965	4.7060	50800
.002	2.99978	4.8504	70900

The final frequency curve is plotted on Figure 12-3.

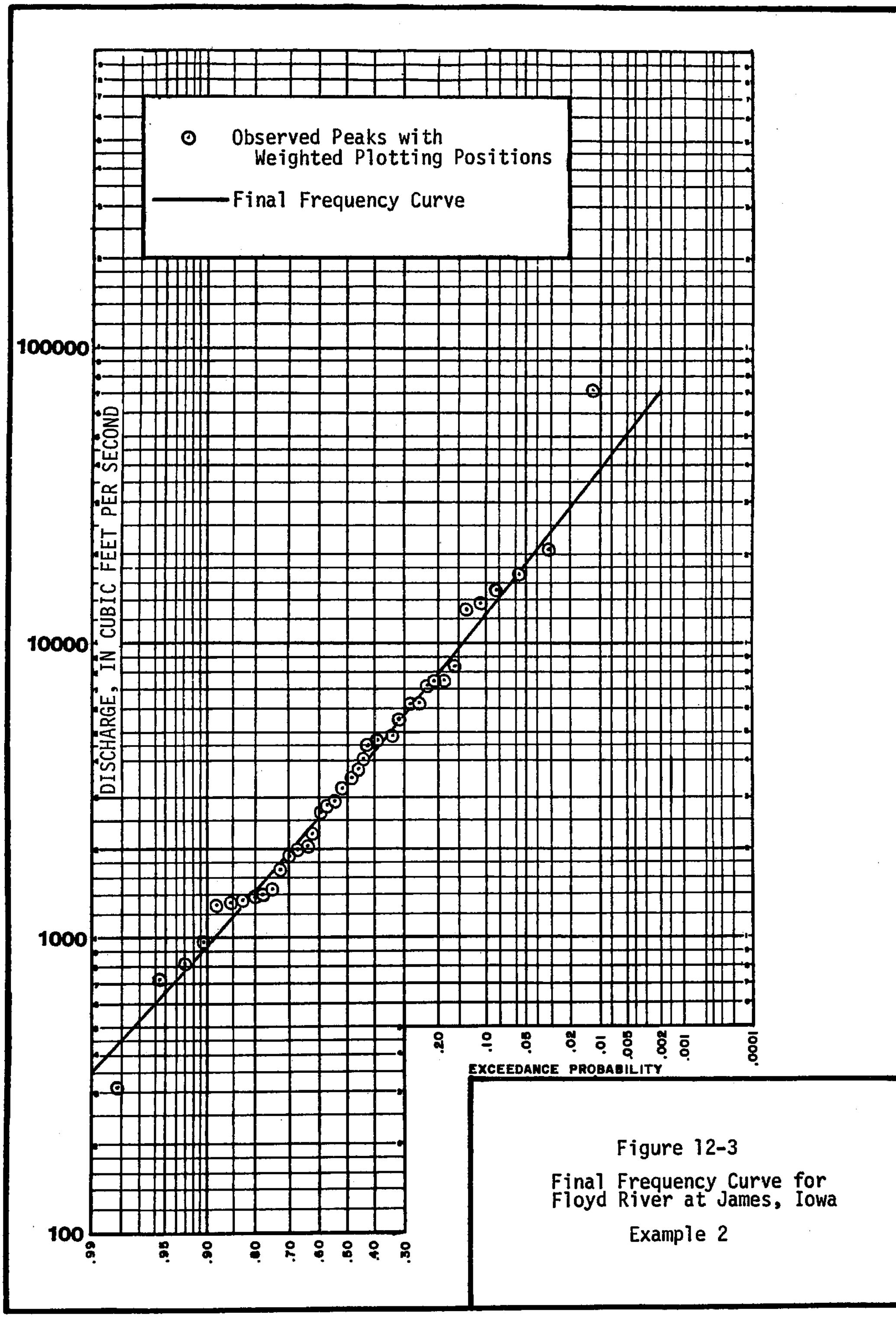


Figure 12-3

Final Frequency Curve for
Floyd River at James, Iowa

Example 2

EXAMPLE 3

TESTING AND ADJUSTING FOR A LOW OUTLIER

a. Station Description

Back Creek near Jones Springs, West Virginia

USGS Gaging Station: 01-6140
Lat: 39°30'43", long: 78°02'15"
Drainage Area: 243 sq. mi.
Annual Peaks Available: 1929-31, 1939-1973

b. Computational Procedures

Step 1 - Compute the statistics of the systematic record.

The detailed computations have been omitted; the results of the computations are:

Mean Logarithm	3.7220
Standard Deviation of logs	0.2804
Skew Coefficient of logs	-0.7311
Years	38

At this point the analyst may be interested in seeing the preliminary frequency curve based on the statistics of the systematic record. Figure 12-4 is the preliminary frequency curve based on the computed mean and standard deviation and a weighted skew of -0.2 (based on a generalized skew of 0.5 from Plate I).

Step 2 - Check for outliers.

As the computed skew coefficient is less than -0.4, the test for detecting possible low outliers is made first. From Appendix 4, the K_N for a sample size of 38 is 2.661.

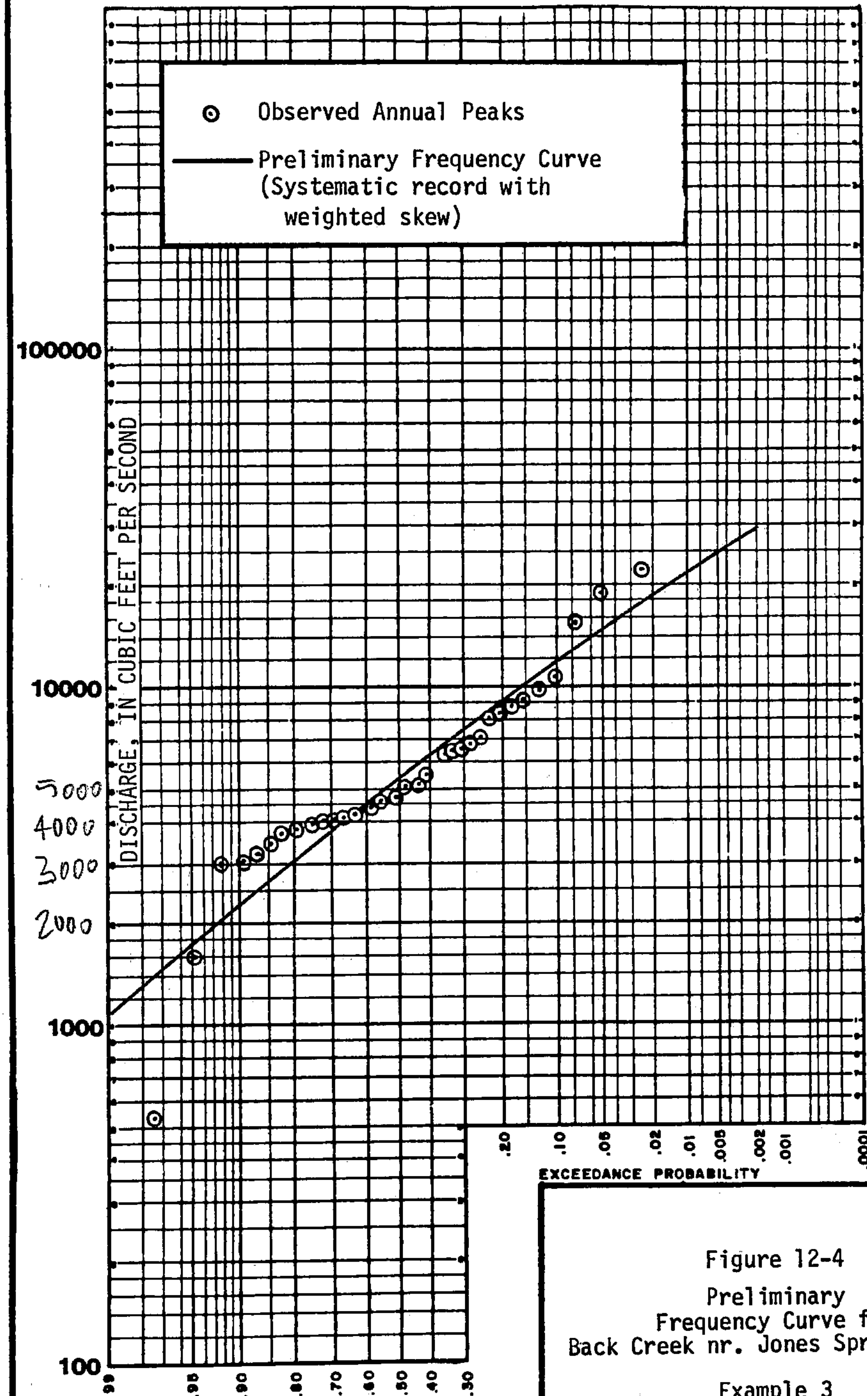


Figure 12-4

Preliminary
Frequency Curve for
Back Creek nr. Jones Springs, W. VA.

Example 3

Example 3 - Testing and Adjusting for a Low Outlier (continued)

The low outlier threshold is computed by Equation 8a:

$$X_L = \overline{X} - K_N S$$

$$= 3.7220 - 2.661\ (.2804) = 2.9759 \qquad (12\text{-}27)$$

$$Q_L = \text{antilog}\ (2.9759) = 946 \text{ cfs}$$

The 1969 event of 536 cfs is below the threshold value of 946 cfs and will be treated as a low outlier.

Step 3 - Delete the low outlier(s) and recompute the statistics.

Mean Logarithm	3.7488
Standard Deviation of logs	0.2296
Skew Coefficient of logs	0.6311
Years	37

Step 4 - Check for high outliers.

The high-outlier threshold is computed to be 22,760 cfs based on the statistics in Step 3 and the sample size of 37 events. No recorded events exceed the threshold value. (See Examples 1 and 2 for the computations to determine the high-outlier threshold.)

Step 5 - Compute and adjust conditional frequency curve.

A conditional frequency curve is computed based on the statistics in Step 3 and then modified by the conditional probability adjustment

(Appendix 5). The skew coefficient has been rounded to 0.6 for ease in computation. The adjustment ratio computed from Equation 5-1a is:

$$\tilde{P} = N/n = 37/38 = 0.9737 \qquad (12\text{-}28)$$

TABLE 12-8

COMPUTATION OF CONDITIONAL FREQUENCY CURVE COORDINATES

P_d	K_{G,P_d} for G = 0.6	log Q	Q cfs	Adjusted Exceedance Probability $(P.P_d)$	
.99	-1.88029	3.3171	2080	.9639	
.90	-1.20028	3.4732	2970	.876	0.5
.50	-0.09945	3.7260	5320	.487	
.10	1.32850	4.0538	11300	.097	
.05	1.79701	4.1614	14500	.049	
.02	2.35931	4.2905	19500	.0195	.01
.01	2.75514	4.3814	24100	.0097	
.005	3.13232	4.4680	29400	.0049	
.002	3.60872	4.5774	37800	.0019	

The conditional frequency curve, along with the adjusted frequency curve, is plotted on Figure 12-5.

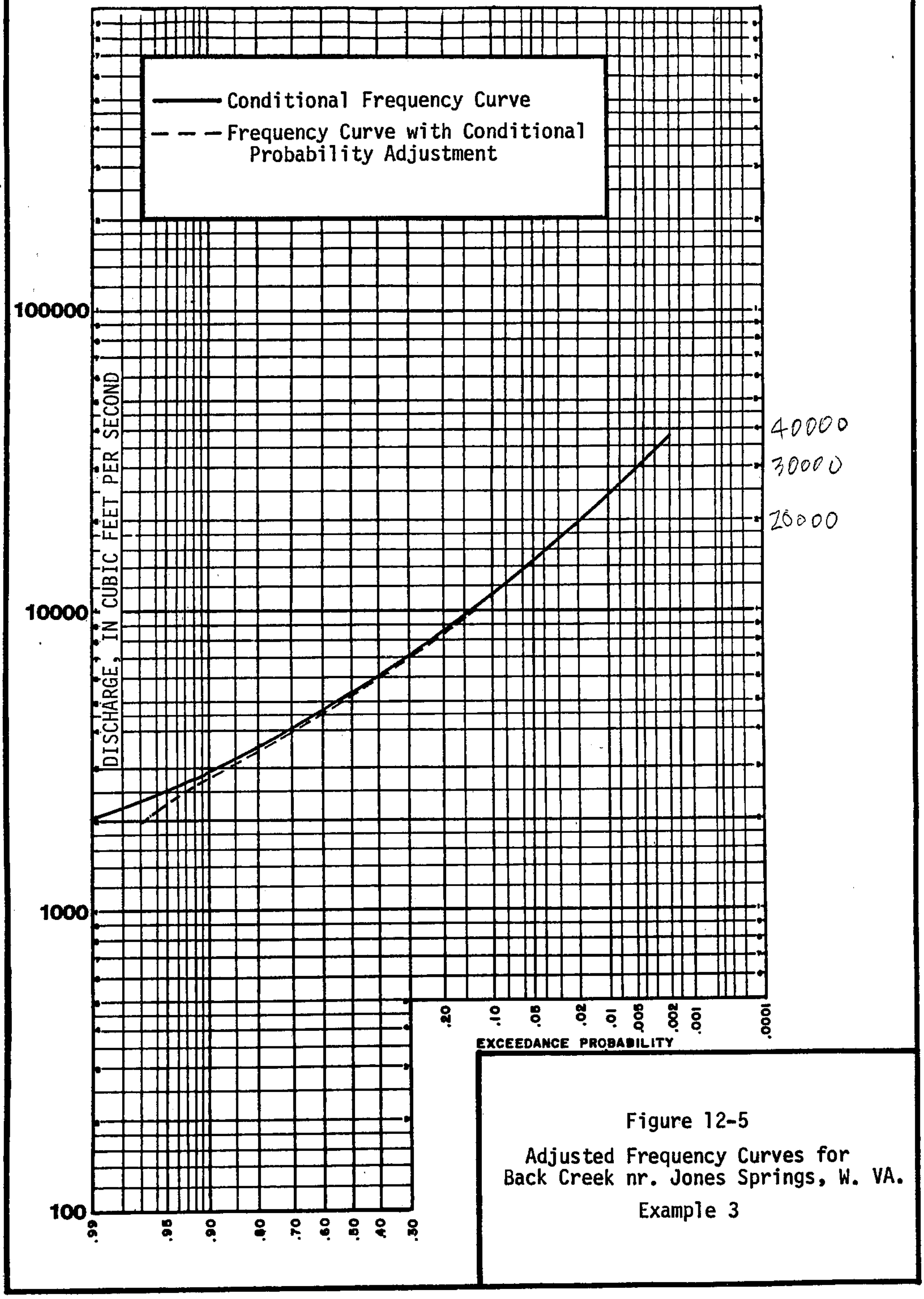


Figure 12-5

Adjusted Frequency Curves for
Back Creek nr. Jones Springs, W. VA.

Example 3

Step 6 - Compute the synthetic statistics.

The statistics of the adjusted frequency curve are unknown. The use of synthetic statistics provides a frequency curve with a log-Pearson Type III shape. First determine the $Q_{.01}$, $Q_{.10}$, and $Q_{.50}$ discharges from the adjusted curve on Figure 12-5.

$Q_{.01}$ = 23880 cfs

$Q_{.10}$ = 11210 cfs

$Q_{.50}$ = 5230 cfs

Next, compute the synthetic skew coefficient by Equation 5-3.

$$G_s = -2.50 + 3.12 \frac{\log(Q_{.01}/Q_{.10})}{\log(Q_{.10}/Q_{.50})}$$

$$= -2.50 + 3.12 \frac{\log(23880/11210)}{\log(11210/5230)} \qquad (12\text{-}29)$$

$$= -2.50 + 3.12 \frac{.32843}{.33110}$$

$$= 0.5948$$

Compute the synthetic standard deviation by Equation 5-4.

$$S_s = \log(Q_{.01}/Q_{.50})/(K_{.01}-K_{.50})$$

$$= \log\ (23880/5230)/\left[2.75514-(-.09945)\right] \qquad (12\text{-}30)$$

$$S_s = .6595/2.8546 = 0.2310$$

Compute the synthetic mean by Equation 5-5.

$$\overline{X}_s = \log\ (Q_{.50}) - K_{.50}(S_s)$$

$$= \log\ (5230) - (-.09945)(.2310) \qquad (12\text{-}31)$$

$$\overline{X}_s = 3.7185 + .0230 = 3.7415$$

Step 7 - Compute the weighted skew coefficient.

The mean-square error of the station skew, from Table 1, is 0.183 based on n = 38 and using G_s for G

$$G_w = \frac{.302(0.5948) + .183(.5)}{.302 + .183} = 0.5590 \qquad (12\text{-}32)$$

$$G_w = 0.6 \text{ (rounded to nearest tenth)}$$

Example 3 - Testing and Adjusting for a Low Outlier (continued)

Step 8 - Compute the final frequency curve.

TABLE 12-9

COMPUTATION OF FREQUENCY CURVE COORDINATES

P	$K_{G_W,P}$ for $G_W = 0.6$	log Q	Q cfs
.99	-1.88029	3.3072	2030
.90	-1.20028	3.4642	2910
.50	-0.09945	3.7185	5230
.10	1.32850	4.0484	11200
.05	1.79701	4.1566	14300
.02	2.35931	4.2865	19300
.01	2.75514	4.3780	23900
.005	3.13232	4.4651	29200
.002	3.60872	4.5751	37600

The final frequency curve is plotted on Figure 12-6

Note: A value of 22,000 cfs was estimated for 1936 on the basis of data from another site. This flow value could be treated as historic data and analyzed by the producers described in Appendix 6. As these computations are for illustrative purposes only, the remaining analysis was not made.

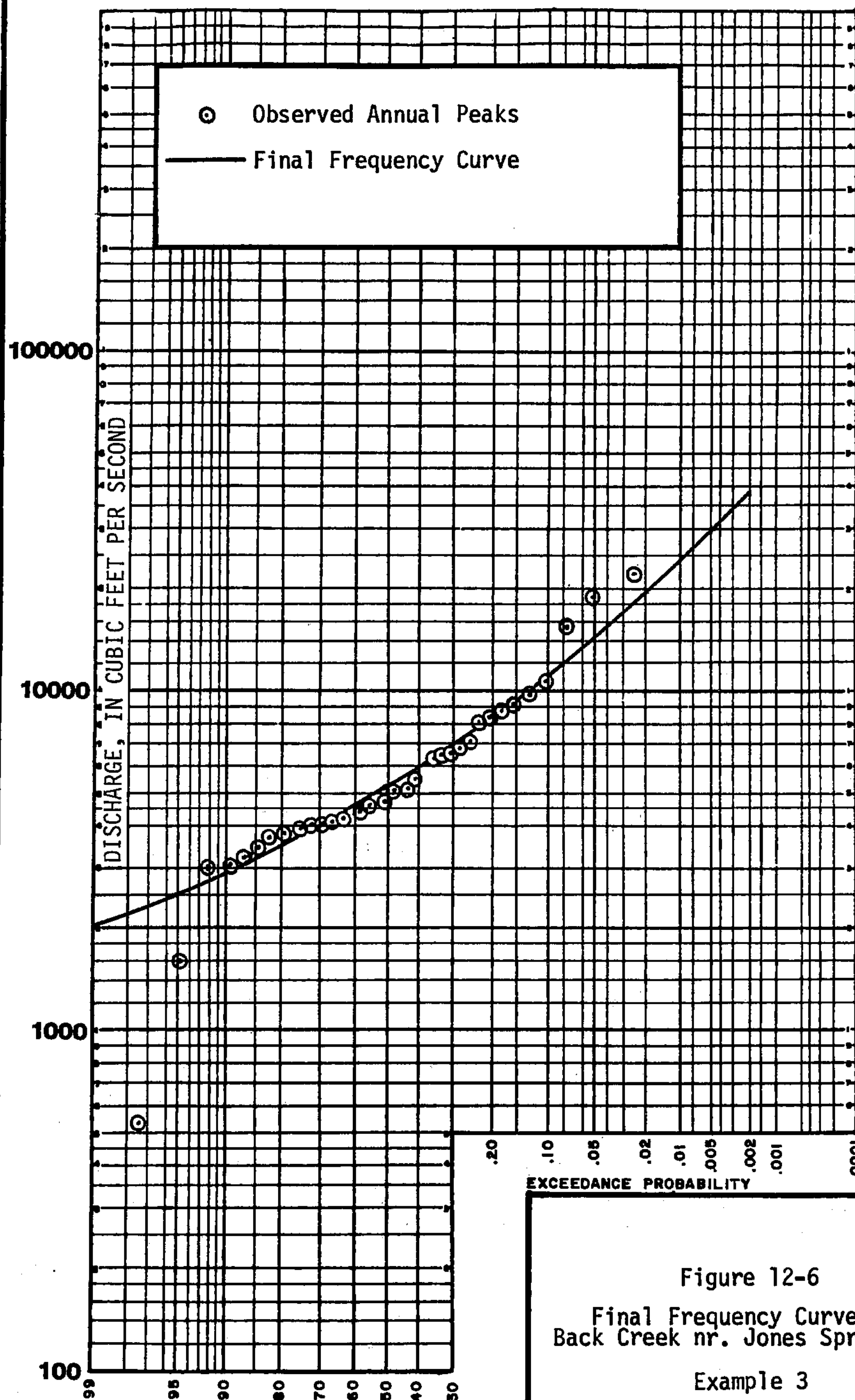


Figure 12-6

Final Frequency Curve for
Back Creek nr. Jones Springs, W. VA.

Example 3

EXAMPLE 4

ADJUSTING FOR ZERO FLOOD YEARS

a. Station Description

Orestimba Creek near Newman, California

USGS Gaging Station: 11-2745
Lat: 37°19'01", long: 121°07'39"
Drainage Area: 134 sq. mi.
Annual Peaks Available: 1932-1973

b. Computational Procedures

Step 1 - Eliminate zero flood years.

There are 6 years with zero flood events, leaving 36 non-zero events.

Step 2 - Compute the statistics of the non-zero events.

Mean Logarithm	3.0786
Standard Deviation of logs	0.6443
Skew Coefficient of logs	-0.8360
Years (Non-Zero Events)	36

Step 3 - Check the conditional frequency curve for outliers.

Because the computed skew coefficient is less than -0.4, the test for detecting possible low outliers is made first. Based on 36 years, the low-outlier threshold is 23.9 cfs. (See Example 3 for low-outlier threshold computational procedure.) The 1955 event of 16 cfs is below the threshold value; therefore, the event will be treated as a low-outlier and the statistics recomputed.

Mean Logarithm	3.1321
Standard Deviation of logs	0.5665
Skew Coefficient of logs	-0.4396
Years (Zero and low outliers deleted)	35

Example 4 - Adjusting for Zero Flood Years (continued)

Step 4 - Check for high outliers

The high outlier threshold is computed to be 41,770 cfs based on the statistics in Step 3 and the sample size of 35 events. No recorded events exceed the threshold value. (See examples 1 and 2 for the computations to determine the high-outlier threshold.)

Step 5 - Compute and adjust the conditional frequency curve.

A conditional frequency curve is computed based on the statistics in step 3 and then adjusted by the conditional probability adjustment (Appendix 5). The skew coefficient has been rounded to -0.4 for ease in computation. The adjustment ratio is 35/42 = 0.83333.

TABLE 12-10

COMPUTATION OF CONDITIONAL FREQUENCY CURVE COORDINATES

P_d	$K_{G,P}$ for G = -0.4	log Q	Q cfs	Adjusted Exceedance Probability ($\tilde{P} \cdot P_d$)
.99	-2.61539	1.6505	44.7	.825
.90	-1.31671 ← -0.278257	2.3862 2.9745	243 1172	.750 ← 0.5
.50	0.06651	3.1698 3.796	1480 6482	.417 ← 0.1
.10	1.23114	3.8295	6750	.083
.05	1.52357	3.9952	9890	.042
.02	1.83361 ← 1.991086	4.1708 4.2600	14800 18260	.017 ← 0.01
.01	2.02933	4.2817	19100	.0083
.005	2.20092	4.3789	23900	.0042
.002	2.39942	4.4914	31000	.0017

Both frequency curves are plotted on Figure 12-7.

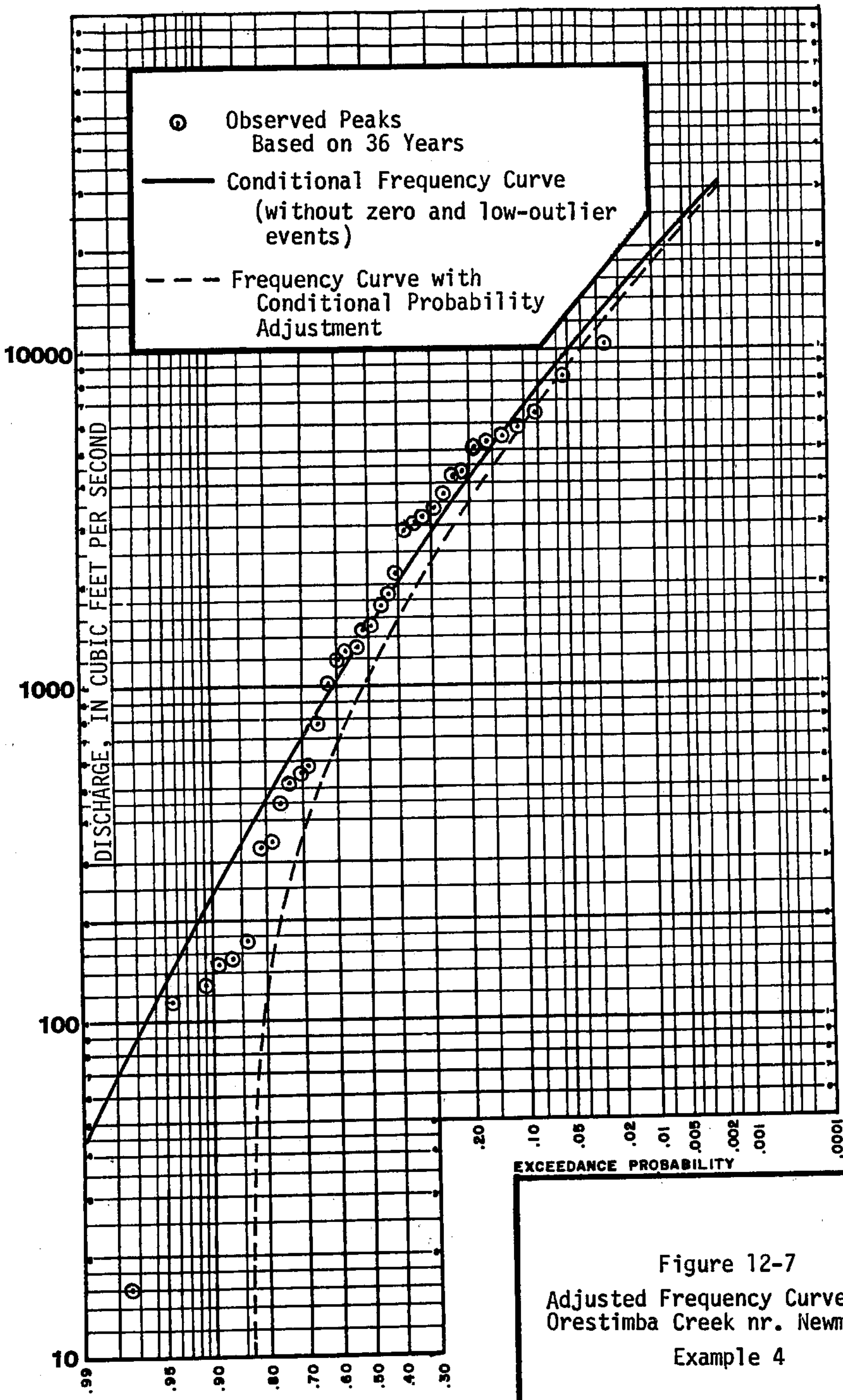


Figure 12-7

Adjusted Frequency Curves for
Orestimba Creek nr. Newman, CA

Example 4

Example 4 - Adjusting for Zero Flood Years (continued)

Step 6 - Compute the synthetic statistics.

First determine the $Q_{.01}$, $Q_{.10}$, and $Q_{.50}$ discharges from the adjusted curve on Figure 12-7.

$Q_{.01}$ = 17940 cfs ← slight variations here lead to big differences below

$Q_{.10}$ = 6000 cfs

$Q_{.50}$ = 1060 cfs

Compute the synthetic skew coefficient by Equation 5-3.

-0.5286867

$$G_s = -2.50 + 3.12 \frac{\log(17940/6000)}{\log(6000/1060)} = -0.5287 \qquad (12\text{-}33)$$

G_s = -0.5 (rounded to nearest tenth)

Compute the synthetic standard deviation by Equation 5-4.

MISTAKE

$$S_s = \log(17940/1060)/(\underbrace{1.95472}_{K_{100}} - \underbrace{\cancel{.08302}}_{K_2}) \qquad (12\text{-}34)$$

-0.27826

S_s = 0.6564 0.5502

Compute the synthetic mean by Equation 5-5.

$$\overline{X}_s = \log(1060) - (\underbrace{.08302}_{K_2})(.6564) \qquad (12\text{-}35)$$

$\overline{X}_s$ = 2.9708 3.178

Step 7 - Compute the weighted skew coefficient by Equation 5.

A generalized skew of -0.3 is determined from Plate I. From Table I, the mean-square error of the station skew is 0.163.

$$G_W = \frac{.302(-.529) + .163(-.3)}{.302 + .163} = -0.4487 \qquad (12\text{-}36)$$

G_W = -0.4 (rounded to nearest tenth)

Example 4 - Adjusting for Zero Flood Years (continued)

Step 8 - Compute the final frequency curve.

TABLE 12-11

COMPUTATION OF FREQUENCY CURVE ORDINATES

P	$K_{G_w,P}$ for G_w = -0.4	log Q	Q cfs
.99	-2.61539	1.2541	17.9
.90	-1.31671	2.1065	128
.50	0.06651	3.0145	1030
.10	1.23114	3.7789	6010
.05	1.52357	3.9709	9350
.02	1.83361	4.1744	14900
.01	2.02933	4.3029	20100
.005	2.20092	4.4155	26000
.002	2.39942	4.5458	35100

This frequency curve is plotted on Figure 12-8. The adjusted frequency derived in Step 4 is also shown on Figure 12-8. As the generalized skew may have been determined from stations with much different characteristics from the zero flood record station, judgment is required to determine the most reasonable frequency curve.

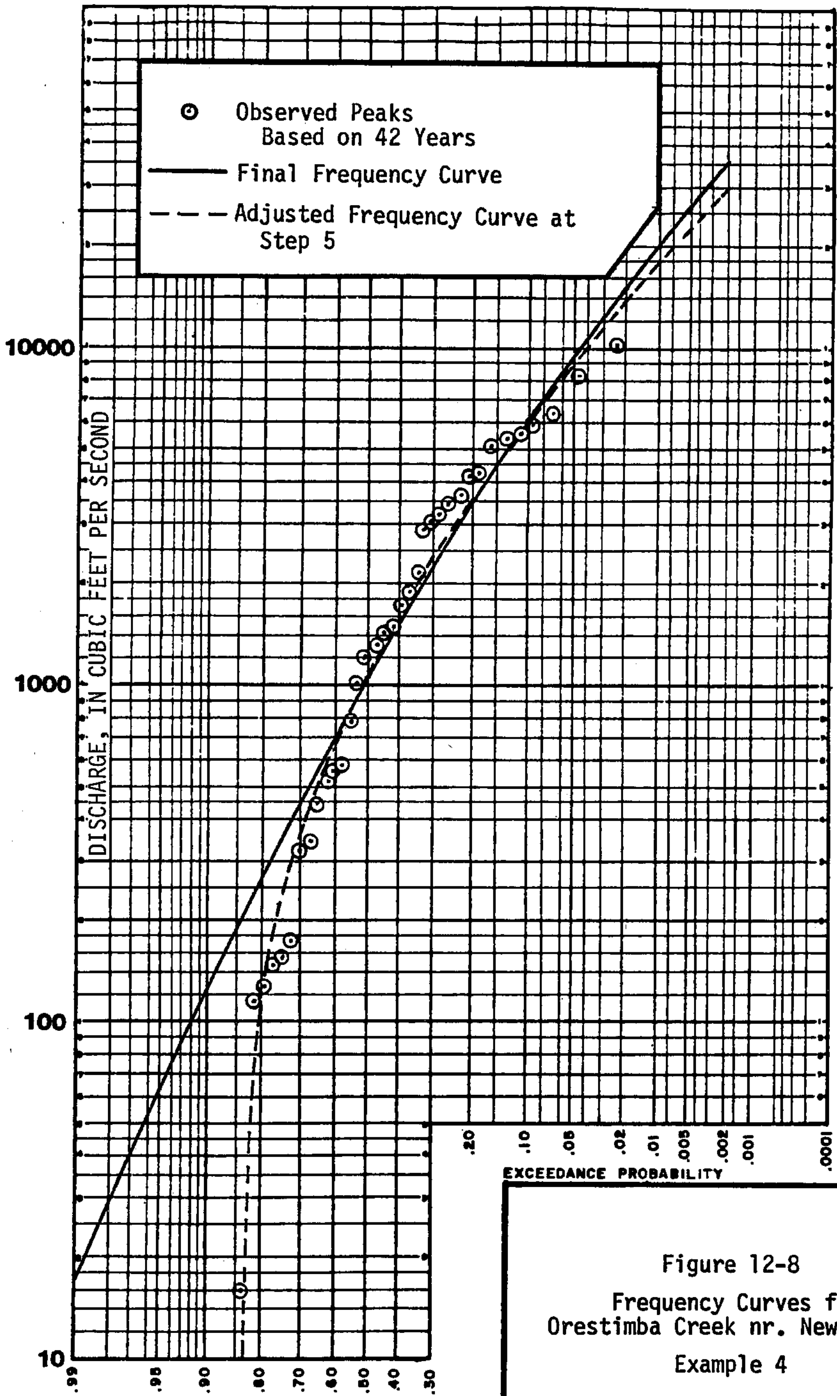


Figure 12-8

Frequency Curves for
Orestimba Creek nr. Newman, CA

Example 4

COMPUTER PROGRAM

+ Programs have been developed that compute a log-Pearson Type III distribution from systematically recorded annual maximum streamflows at a single station -- and other large known events. Special routines are included for managing zero flows and very small flows (outliers) that would distort the curve in the range of higher flows. An option is included to adjust the computed curve to represent expected probability. Copies of agency programs that incorporate procedures recommended by this Guide may be obtained from either of the following:

Chief Hydrologist
U.S. Geological Survey, WRD
National Center, Mail Stop 437
Reston, VA 22092

Phone: (703) 860-6879

Hydrologic Engineering Center
U.S. Army Corps of Engineers
609 2nd Street, Suite I
Davis, CA 95616

Phone: (916) 756-1104

There is no specific recommendation to utilize these particular computer programs. Other federal and state agencies as well as private organizations may have developed individual programs to suit their specific needs. +

*

Appendix 14

"FLOOD FLOW FREQUENCY TECHNIQUES"

REPORT SUMMARY

*

Following is a summary of "Flood Flow Frequency Techniques," a report by Leo R. Beard, Technical Director, Center for Research in Water Resources, The University of Texas at Austin, for the Office of Water Resources Research and the Water Resources Council. Much of the text and a majority of the exhibits are taken directly from the report.

The study was made at the Center for Research in Water Resources of The University of Texas at Austin at the request of and under the general guidance of the Work Group on Flood Flow Frequency, Hydrology Committee, of the Water Resources Council through the auspices of the Office of Water Resources Research. The purpose was to provide a basis for development by the Work Group of a guide for flood frequency analysis at locations where gage records are available which would incorporate the best technical methods currently known and would yield greater reliability and consistency than has heretofore been available in flood flow frequency determinations.

The study included: (a) a review of the literature and current practice to select candidate methods and procedures for testing, (b) selection of long-record station data of natural streamflows in the United States and development of data management and analysis computer programs for testing alternate procedures, (c) testing eight basic statistical methods for frequency analysis including alternate distributions and fitting techniques, (d) testing of alternate criteria for managing outliers, (e) testing of procedures for treating stations with zero flow years, (f) testing relationships between annual maximum and partial-duration series, (g) testing of expected probability adjustment, (h) testing to determine if flood data exhibit consistent long-term trends, and (i) recommendations with regard to each procedure tested and development of background material for the guides being developed by the Work Group.

Data

In all, 300 stations were used in the testing. Flows were essentially unregulated. Record length exceeded 30 years with most stations having records longer than 40 years. The stations were selected to give the best feasible coverage of drainage area size and geographic location and to include a substantial number of stations with no flow for an entire year. Table 14-1 lists the number of stations by size and geographic zone.

Split Record Testing

A primary concern of the study was selection of a mathematical function and fitting technique that best estimates flood flow frequencies from annual peak flow data. Goodness of fit of a function to the data used in the fitting process is not necessarily a valid criterion for selecting a method that best estimates flood frequencies. Consequently, split record testing was used to simulate conditions of actual application by reserving a portion of a record from the fitting computation and using it as "future" events that would occur in practice. Goodness of fit can nevertheless be used, particularly to eliminate methods whose fit is very poor.

Each record of annual maximum flows was divided into two halves, using odd sequence numbers for one half and even for the other in order to eliminate the effect of any general trend that might possibly exist. This splitting procedure should adequately simulate practical situations as annual events were tested and found independent of each other. Frequency estimates were made from each half of a record and tested against what actually happened in the other half.

Development of verification criteria is complicated, because what actually happens in the reserved record half also is subject to sampling irregularities. Consequently, reserved data cannot be used as a simple, accurate target and verification criteria must be probabilistic. The test procedure, however, simulates conditions faced by the planner, designer, or operator of water resource projects, who knows neither that past events are representative nor what future events will be.

The ultimate objective of any statistical estimation process is not to estimate the most likely theoretical distribution that generated the observed data, but rather to best forecast future events for which a decision is formulated. Use of theoretical distribution functions and their attendant reliability criteria is ordinarily an intermediate step to forecasting future events. Accordingly, the split record technique of testing used in this study should be more rigorous and direct than alternative theoretical goodness-of-fit tests.

Frequency Computation Methods

Basic methods and fitting techniques tested in this study were selected by the author and the WRC Work Group on Flood Flow Frequency after careful review of the literature and experience in the various agencies represented; those that were tested are listed below. Numbering corresponds to the identification number of the methods in the computer programs and in the attached tables.

1. Log-Pearson Type III (LP3). The technique used for this is that described in (35). The mean, standard deviation, and skew coefficients for each data set are computed in accordance with the following equations:

$$\overline{X} = \frac{\Sigma X}{N} \tag{14-1}$$

$$S^2 = \frac{\Sigma X^2 - (\Sigma X)^2/N}{N-1} \tag{14-2}$$

$$g = \frac{N^2\Sigma X^3 - 3N\Sigma X\Sigma X^2 + 2(\Sigma X)^3}{N(N-1)\ (N-2)S^3} \tag{14-3}$$

where

X = logarithm of peak flow
N = number of items in the data set
$\overline{X}$ = mean logarithm
S = standard deviation of logarithms
g = skew coefficient of logarithms

Flow logarithms are related to these statistics by use of the following equation:

$$X = \overline{X} + kS \tag{14-4}$$

Exceedance probabilities for specified values of k and values of k for specified exceedance probabilities are calculated by use of the normal distribution routines available in computer libraries and the approximate transform to Pearson deviates given in reference (31).

2. Log Normal (LN). This method uses a 2-parameter function identical to the log-Pearson III function except that the skew coefficient is not computed (a value of zero applies), and values of k are related to exceedance probabilities by use of the normal distribution transform available in computer libraries.

3. Gumbel (G). This is the Fisher-Tippett extreme-value function, which relates magnitude linearly with the log of the log of the reciprocal of exceedance probability (natural logarithms). Maximum likelihood estimates of the mode and slope (location and scale parameters) are made by iteration using procedures described by Harter and Moore in reference (36). The initial estimates of the location and scale statistics are obtained as follows:

$$M = \overline{X} - 0.45005\ S \tag{14-5}$$

$$B = .7797\ S \tag{14-6}$$

Magnitudes are related to these statistics as follows:

$$X = M + B(-\ln(-\ln P)) \tag{14-7}$$

where

M = mode (location statistic)
B = slope (scale statistic)
X = magnitude
P = exceedance probability
S = standard deviation of flows

Some of the computer routines used in this method were furnished by the Central Technical Unit of the Soil Conservation Service.

4. Log Gumbel (LG). This technique is identical to the Gumbel technique except that logarithms (base 10) of the flows are used.

5. Two-parameter Gamma (G2). This is identical to the 3-parameter Gamma method described below, except that the location parameter is set to zero. The shape parameter is determined directly by solution of Nörlund's (37) expansion of the maximum likelihood equation which gives the following as an approximate estimate of α:

$$\alpha = \frac{1 + \sqrt{1 + \frac{4}{3}\left(\ln\overline{Q} - \frac{1}{N}\Sigma\ln Q\right)}}{4\left(\ln\overline{Q} - \frac{1}{N}\Sigma\ln Q\right)} - \Delta\alpha \qquad (14\text{-}8)$$

where

$\overline{Q}$ = average annual peak flow

N = number of items in the data set

Q = peak flow

$\Delta\alpha$ = correction factor

β is estimated as follows:

$$\beta = \frac{1}{\alpha} \cdot \frac{1}{N}\Sigma Q \qquad (14\text{-}9)$$

6. Three-parameter Gamma (G3). Computation of maximum likelihood statistics for the 3-parameter Gamma distribution is accomplished using procedures described in reference (38). If the minimum flow is zero, or if the calculated lower bound is less than zero, the statistics are identical to those for the 2-parameter Gamma distribution. Otherwise, the lower bound, γ, is initialized at a value slightly smaller than the lowest value of record, and the maximum likelihood value of the lower bound is derived by iteration using criteria in reference (38). Then the parateters α and β are solved for directly using the equations above replacing Q with Q-γ. Probabilities corresponding to specified magnitudes are computed directly by use of a library gamma routine. Magnitudes corresponding to specified

probabilities are computed by iteration using the inverse solution.

7. Regional Log-Pearson Type III (LPR). This method is identical to the log-Pearson Type III method, except that the skew coefficient is taken from Figure 14-1 instead of using the computed skew coefficient. Regionalized skew coefficients were furnished by the U.S. Geological Survey.

8. Best Linear Invariant Gumbel (BLI). This method is the same as for the Gumbel method, except that best linear invariant estimates (BLIE) are used for the function statistics instead of the maximum likelihood estimates (MLE). An automatic censoring routine is used for this method only, so there are no altenative outlier techniques tested for this method. Statistics are computed as follows:

$$M = \Sigma(X(I) \cdot U(N,J,I)) \qquad (14\text{-}10)$$

$$B = \Sigma(X(I) \cdot V(N,J,I)) \qquad (14\text{-}11)$$

where

U = coefficient UMANN described in reference (39)
V = coefficient BMANN described in reference (39)
J = number of outliers deleted plus 1
I = order number of flows arranged in ascending-magnitude order
N = sample size as censored.

Since weighting coefficients U and V were made available in this study only for sample sizes ranging from 10 to 25, 5-year samples are not treated by this method, and records (or half records) of more than 25 years are divided into chronological groups and weighted average coefficients used in lieu of coefficients that might otherwise be obtained if more complete sets of weighting coefficients were available. Up to two outliers are censored at the upper end of the flow array. Each one is removed if sequential tests show that a value that extreme would occur by chance less than 1 time 10 on the basis of the BLIE statistics. Details of this censoring technique are contained in reference (40). Weighting coefficients and most of the routines used in this method were furnished by the Central Technical Unit of the Soil Conservation Service.

Outliers

Outliers were defined for purpose of this study as extreme values whose ratio to the next most extreme value in the same (positive or negative) direction is more extreme than the ratio of the next most extreme value to the eighth most extreme value.

The techniques tested for handling outliers consisted of

a. keeping the value as is,

b. reducing the value to the product of the second largest event and the ratio of the second largest to eighth largest event,

c. reducing the value to the product of the second largest event and the square root of that ratio, and

d. discarding the value.

In the cases of outliers at the low end, the words largest in (b) and (c) should be changed to smallest.

Zero Flow

Two techniques were tested for handling stations with some complete years of no flow as follows:

(a) Adding 1 percent of the mean magnitude to all values for computation purposes and subtracting that amount from subsequent estimates, and

(b) removing all zeros and multiplying estimated exceedance frequencies of the remaining by the ratio of the number of non-zero values to the total number of values. This is the procedure of combining probabilities described in reference (27).

Partial-Duration Series

A secondary concern of the study was the relationship between annual maximum flow frequencies and partial-duration flow frequencies.

Because a partial-duration series consists of all events above a specified magnitude, it is necessary to define separate events. The definition normally depends on the application of the frequency study as well as the hydrologic characteristics of the stream. For this study separate events were arbitrarily defined as events separated by at least as many days as five plus the natural logarithm of the square miles of drainage area, with the requirement that intermediate flows must drop below 75 percent of the lower of the two separate maximum daily flows. This is considered representative of separation criteria appropriate for many applications.

Maximum daily flows were used for this part of the study, because there were insufficient readily available data on instantaneous peak flows for events smaller than the annual maximum. There is no reason to believe that the frequency relationship would be different for peak flows than for daily flows.

The relationship between the maximum annual and partial-duration series was expressed as a ratio of partial-duration to annual event frequencies at selected annual event frequencies. In order to develop partial-duration relationships independent of any assumptions as to frequency functions, magnitudes corresponding to annual-maximum event exceedance probabilities of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, and 0.7 are established for complete records at each station by linear interpolation between expected probability plotting positions ($M/(n+1)$) for the annual maximum events. Corresponding frequencies of partial-duration flows are established simply by counting the total number of independent maximum daily flows at each station above each magnitude and dividing by the total number of years at that station. Ratios of partial-duration to annual event frequencies were averaged for all stations in each USGS zone and compared with ratios derived for certain theoretical conditions by Langbein (9).

Expected Probability Estimation

The expected probability is defined as the average of the true probabilities of all magnitude estimates for any specified flood frequency that might be made from successive samples of a specified size. For any specified flow magnitude, it is considered to be the most appropriate estimate of probability or frequency of future flows for water resources planning and management use.

It is also a probability estimate that is theoretically easy to verify, because the observed frequencies in reserved data at a large number of stations should approach the computed probability or frequency estimates as the number of stations increases. Accordingly, it was considered that expected probability estimates should be used in the split record tests.

A method of computing expected probabilities has been developed for samples drawn from a Gaussian normal distribution as described in (21).

Similar techniques are not available for the other threoretical distribution functions. Consequently, an empirical transform is derived for each distribution. To do this a calibration constant was determined which, when multiplied by the theoretical normal transform adjustment, removed the observed average bias in estimating probabilities for the 300 stations used in this study. This empirical transform was used in making the accuracy tests that are the main basis for judging the relative adequacy of the various methods tests.

Trends and Cycles

There is some question as to whether long-term trends and cycles (longer than 1 year) exist in nature such that knowledge of their nature can be used to improve forecasts of flood flow frequencies for specific times in the future. As a part of this research project, lag 1 autocorrelation coefficients of annual peak flows for all stations were computed. If trends or cycles exist in any substantial part of the data, there should be a net positive average autocorrelation for all stations. A statistically significant positive average autocorrelation was not found.

Accuracy and Consistency Tests

Criteria used in judging the adequacy of each method for fitting a theoretical distribution were as follows:

Accuracy tests consisted of the following comparisons between computed frequencies in one-half the record with frequencies of events that occurred in the reserved data.

a. Standard deviation of observed frequencies (by count) in reserved data for magnitude estimates corresponding to exceedance probabilities of 0.001, 0.01, 0.1, and 0.5 computed from the part of the record used. This is the standard error of a frequency estimate at individual stations that would occur if a correction is made for the average observed bias in each group of stations for each selected frequency and method.

b. Root-mean-square difference between expected probability plotting position (M/(n+1)) of the largest, upper decile and median event in a half record and the computed expected probability exceedance frequency of that respective event in the other half. This is the standard error of a frequency estimate at individual stations without any bias adjustment for each method and for the frequency of each selected event.

c. Root-mean-square difference between 1.0 and the ratio of the computed probability of flow in the opposite half of a record to the plotting position of the largest, upper decile and median event (in turn) in a half record. This criterion is similar to that of the preceding paragraph except that methods that are biased toward predicting small frequencies are not favored.

Consistency tests involved the following comparisons between computed frequencies in each half of the record with the total record.

a. Root-mean-square difference between computed probabilities from the two record halves for full record extreme, largest, upper decile and median events, in turn. This is an indicator of the relative uniformity of estimates that would be made with various random samples for the same location.

b. Root-mean-square value of 1.0 minus the ratio of the smaller to the larger computed probabilities from the two record halves for full record extreme, largest, upper decile and median events, in turn. This is essentially the same as the preceding criterion, except that methods that are biased toward predicting small frequencies are not favored.

The extreme event used in the consistency tests is an arbitrary value equal to the largest multiplied by the square root of the ratio of the largest to the median event for the full record.

It should be recognized that sampling errors in the reserved data are as large or larger for the same sample size as are sampling errors of computed values. Similarly, sampling errors are comparable for estimates based on opposite record halves used for consistency tests. Consequently, a great number of tests is necessary in order to reduce the uncertainty due to sampling errors in the reserved data. Further, a method that is biased toward estimating frequencies too low may have a small standard error of estimating frequencies in comparison with a method that is biased toward high frequencies, if the bias is not removed. The latter may have smaller percentage errors. Accordingly, consideration of the average frequency estimate for each of the eight methods must be a component of the analyses.

As a further means of evaluating alternate procedures the complete record results, computed curve without any expected probability adjustment, and the plotted data point were printed out.

Evaluation of Distributions

Table 14-2 shows for each method and each USGS zone the number of stations where an observed discharge exeeeded the computed 1,000-year discharge. With 14,200 station-years of record, it might be expected that about 14 observed events would exceed true 1,000-year magnitudes. This comparison indicates that the log-Pearson Type III (method 1), log normal (method 2), and log-Pearson Type III with generalized skew (method 7), are the most accurate.

Table 14-3 shows average observed frequencies (by count) in the reserved portions of half records for computed probabilities of 0.001, 0.01, 0.1, and 0.5 and the standard deviations (accuracy test a) of the observed frequencies from their averages for each computed frequency. It is difficult to draw conclusions from these data. Figure 14-2 shows a plotting of the results for the 0.01 probability estimates which aids in comparison. This comparison indicates that the log normal and log-Pearson Type III methods with generalized skew have observed frequencies closest to those computed and the smallest standard deviations except for method 4.

Table 14-4 shows the average results for all stations of accuracy tests b and c. Results are not definitive, but again the log normal (method 2) and log-Pearson Type III with generalized skew (method 7) show results as favorable as any other method as illustrated for test b in Figure 14-3.

Table 14-5 shows the results of the consistency tests. Figure 14-4 displays the results graphically for test a. The consistency test results are not substantially different from or more definitive than the accuracy results. From Figure 14-4 it appears that the log-Pearson Type III method with generalized skew yields considerably more consistent results than the log normal.

Results of Outlier Testing

Table 14-6 shows results for all stations of the accuracy and consistency tests for the four different outlier techniques. Results of these tests show that for the favorable methods [log normal (method 2) and log-Pearson Type III with generalized skew (method 7)], outlier techniques a and b are most favorable. Unfortunately, no discrimination was made in the verification tests between treatment of outliers at the upper and lower ends of the frequency arrays. Outliers at the lower end can greatly increase computed frequencies at the upper end. Average computed frequencies for all half records having outliers at the upper or lower end are generally high for the first three outlier techniques and low for the fourth.

It is considered that this is caused primarily by outliers at the lower end. Values observed are as follows:

Average plotting position of maximum flow	0.042
Average computed probability, method a	0.059
Average computed probability, method b	0.050
Average computed probability, method c	0.045
Average computed probability, method d	0.038

Until more discriminatory outlier studies are made, method a appears to be the most logical and justifiable to use.

Results of Zero Flow Testings

Table 14-7 shows the average for all stations of the results of accuracy and consistency tests for the two different zero flow techniques.

These test comparisons indicate that for the favorable methods [log normal (method 2) and log-Pearson Type III with generalized skew (method 7)], technique b is slightly better than a.

Results of Partial-Duration Studies

Results of partial-duration studies are shown in Table 14-8. It can be seen that there is some variation in values obtained for different zones and that the average of all zones is somewhat greater than the theoretical values developed by Langbein. The theoretical values were based on the assumption that a large number of independent (random) events occur each year. If the number of events per year is small, the average values in Table 14-8 would be expected to be smaller than the theoretical values. If the events are not independent such that large events tend to cluster in some years and small events tend to cluster in other years, the average values in Table 14-8 would be expected to be larger than the theoretical values.

It was concluded that values computed for any given region (not necessarily zones as used in this study) should be used for stations in that region after smoothing the values such that they have a constant relation to the Langbein theoretical function.

Expected Probability Adjustment Results

The ratios by which the normal expected probability theoretical adjustment must be multiplied in order to compute average probabilities equal to those observed for each zone are shown in Tables 14-9, 14-10, and 14-11. It will be noted that these vary considerably from zone to zone and for different exceedance intervals. Much of this variation, however, is believed due to vagaries of sampling. Average ratios for the 100-year flood shown on the last line in Table 14-10 were adopted for each distribution for the purpose of comparing accuracy and the various methods. These are as follows:

1.	Log-Pearson Type III	2.1
2.	Log Normal	0.9
3.	Gumbel, MLE	3.4
4.	Log Gumbel	-1.2
5.	2-parameter gamma	3.4

6.	3-parameter gamma	2.3
7.	Regional log-Pearson Type III	1.1
8.	Gumbel, BLIE	5.7

Results of this portion of the study indicate that only the log normal (method 2) and log-Pearson Type III with regional skew (method 7) are free of substantial bias because zero bias should correspond approximately to a coefficient of 1.0 as would be the case if the distribution characteristics do not greatly influence the adjustment factor. The following tabulation for log-Pearson Type III method with regional skew indicates that the theoretical expected probability adjustment for the normal distribution applies approximately for this method. Coefficients shown range around the theoretical value of 1.0 and, with only one exception, do not greatly depart from it in terms of standard-error multiples. It is particularly significant that the most reliable data (the 100-year values) indicate an adjustment factor near 1.0.

Expected Probability Adjustment Ratios for All Zones

Sample	10-Yr		100-Yr		1000-Yr	
Size	Avg.	Std. Err.	Avg.	Std. Err.	Avg.	Std. Err.
5	0.81	0.17	0.94	0.12	1.01	0.13
10	0.60	0.22	1.12	0.20	1.45	0.27
23	0.17	0.27	1.14	0.23	1.68	0.28

Results of Test for Trends and Cycles

Results of lag I autocorrelation studies to test for trends are shown in Table 14-12. It is apparent that there is a tendency toward positive autocorrelation, indicating a tendency for flood years to cluster more than would occur in a completely random process. The t values shown are multiples of the standard error of the lag I correlation coefficient, and it is obvious that extreme correlation coefficients observed are not seriously different from variations that would occur by chance. It is considered that annual peak flows approximate a random process in streams used in this study.

Conclusions

Although split record results were not as definitive as anticipated, there are sufficient clearcut results to support definite recommendations. Conclusions that can be drawn are as follows:

a. Only method 2 (log normal) and method 7 (log-Pearson Type III with regional skew) are not greatly biased in estimating future frequencies.

b. Method 7 gives somewhat more consistent results than method 2.

c. For methods 2 and 7, outlier technique "a" (retaining the outlier as recorded) is more accurate in terms of ratio of computed to observed frequencies than methods that give less weight to outliers.

d. For methods 2 and 7, zero flow technique "b" (discarding zero flows and adjusting computed frequencies) is slightly superior to zero flow technique "a."

e. Streamflows as represented by the 300 stations selected for this study are not substantially autocorrelated; thus, records need not be continuous for use in frequency analysis.

f. Partial-duration frequencies are related to annual event frequencies differently in different regions; thus, empirical regional relationships should be used rather than a single theoretical relationship.

Of particular significance is the conclusion that frequencies computed from theoretical functions in the classical manner must be adjusted to reflect more frequent extreme events if frequencies computed in a great number of cases are to average the same as observed frequencies. For the recommended method, adjustment equal to the theoretical adjustment for estimates made from samples drawn from a normal population is approximately correct.

Of interest from a research standpoint is the finding that split record techniques require more than 300 records of about 50 events each to be definitive. This study showed that random variations in the reserved data obscure the results to greater degree than would be the case if curve-fitting functions could reduce uncertainty to a greater degree than has been possible.

In essence, then, regardless of the methodology employed, substantial uncertainty in frequency estimates from station data will exist, but the log-Pearson type III method with regional skew coefficients will produce unbiased estimates when the adjustment to expected probability is employed, and will reduce uncertainty as much as or more than other methods tested.

Recommendations for Future Study

It is considered that this study is an initial phase of a more comprehensive study that should include

a. Differentiation in the treatment of outliers at the upper and lower ends of a frequency curve;

b. Treatment of sequences composed of different types of events such as flood flows resulting from rainfall and those from snowmelt, or hurricane and nonhurricane floods;

c. Physical explanation for great differences in frequency characteristics among streams in a given region;

d. Development of systematic procedures for regional coordination of flood flow frequency estimates and applications to locations with recorded data as well as to locations without recorded data;

e. Development of procedures for deriving frequency curves for modified basin conditions, such as by urbanization;

f. Development of a step-by-step procedure for deriving frequency curves for locations with various amounts and types of data such that progressively reliable results can be obtained on a consistent basis as the amount of effort expended is increased; and

g. Preparation of a text on flood flow frequency determinations for use in training and practical application.

FIGURE 14-1

GENERALIZED SKEW COEFFICIENTS OF ANNUAL MAXIMUM STREAMFLOW LOGARITHMS

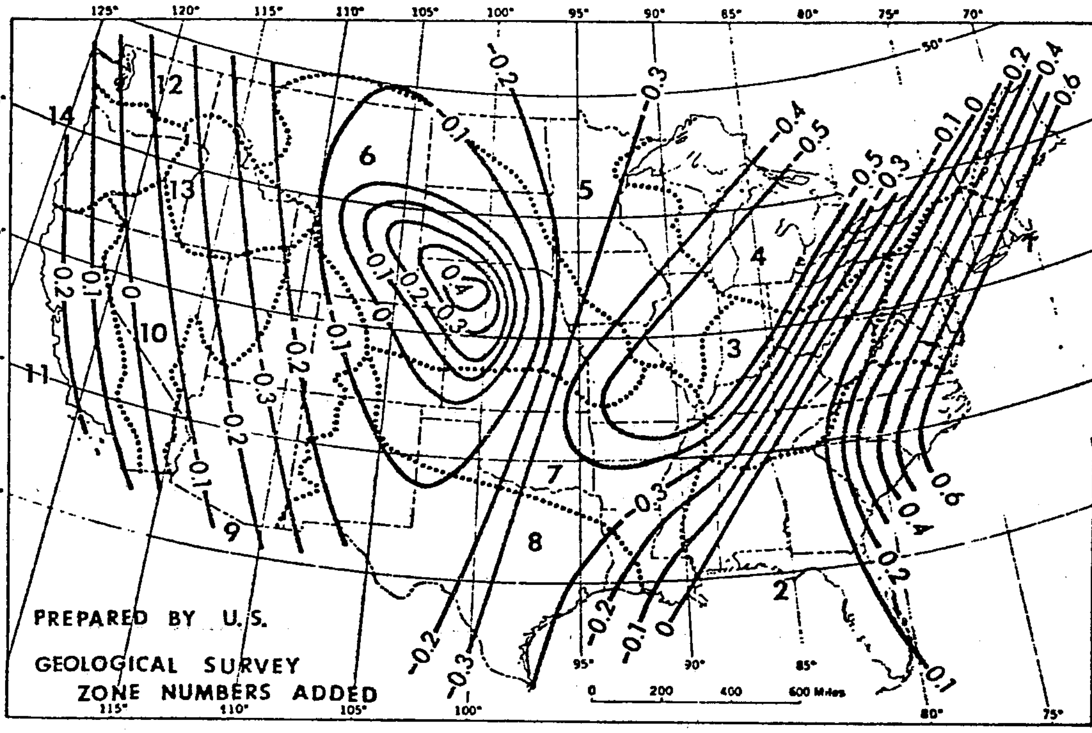
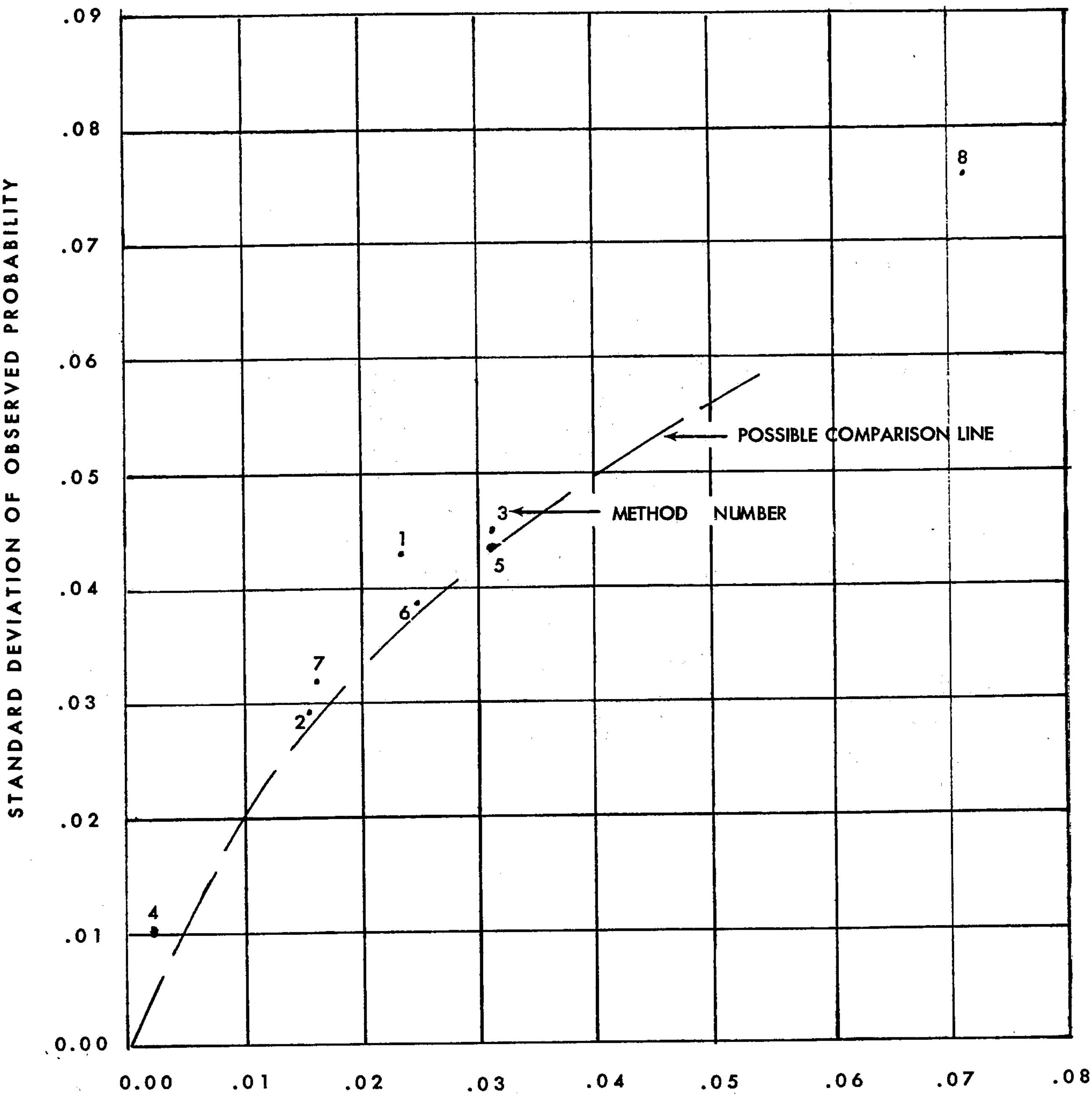


FIGURE 14-2

ACCURACY COMPARISON FOR 0.01 PROBABILITY ESTIMATE (TABLE 14-3)



AVERAGE OBSERVED PROBABILITY IN TABLE 14-3

FOR 0.01 COMPUTED PROBABILITY

FIGURE 14-3

ACCURACY COMPARISON FOR MAXIMUM OBSERVED FLOW

(TABLE 14-4, TEST B)

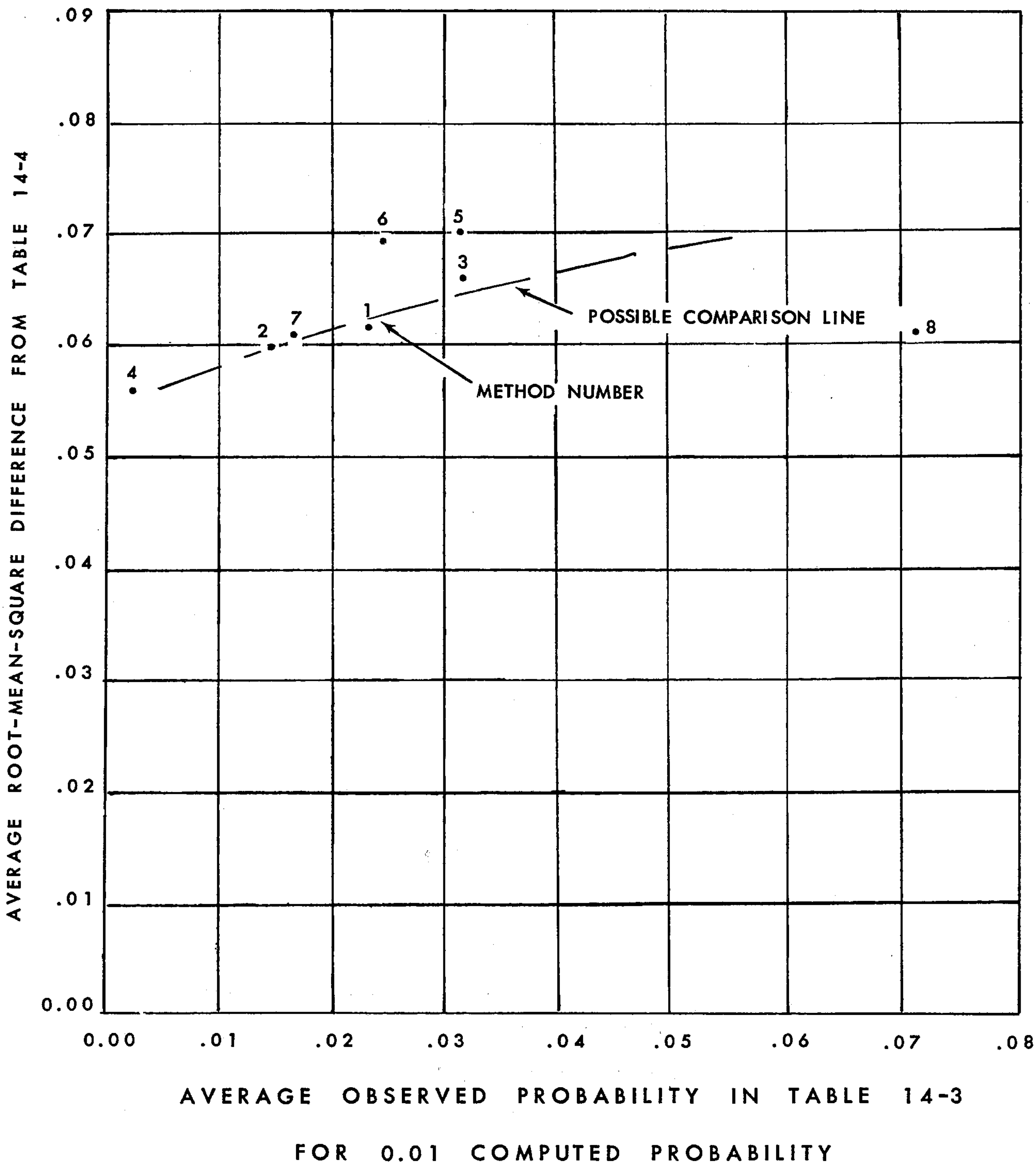


FIGURE 14-4

CONSISTENCY COMPARISON FOR MAXIMUM OBSERVED FLOW

(TABLE 14-5, TEST A)

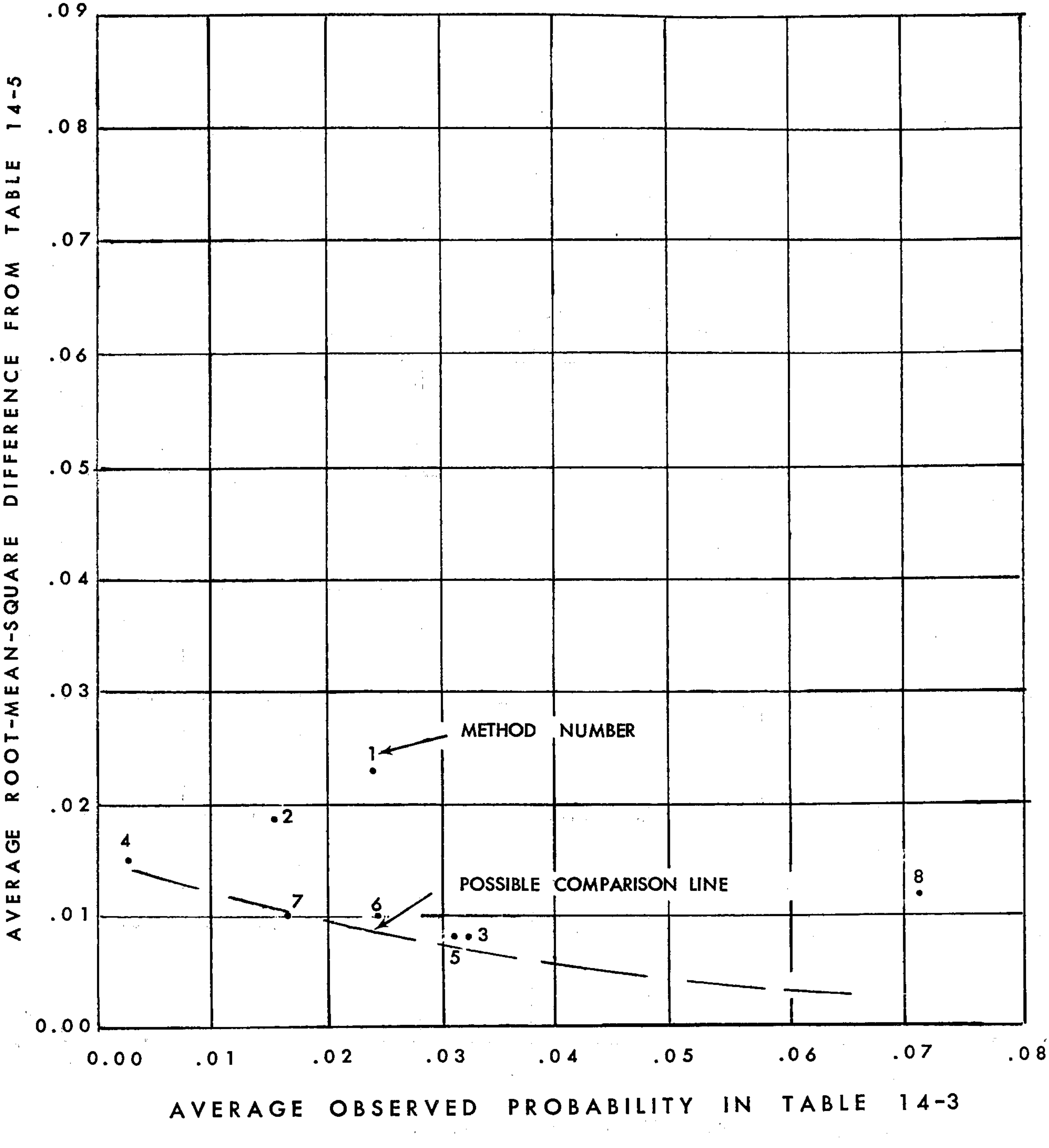


Table 14-1

Numbers of Verification Stations by Zones and Area Size

USGS ZONE	Drainage area category (sq. mi.) 0-25	25-200	200-1000	1000+	Total
1	4	8	10	5	27
2	2	5	12	5	24
3	5	3	16	1	25
4	1	6	8	0	15
5	3	2	14	1	20
6	4	3	13	4	24
7	5	2	12	2	21
8	8	2	11	2	23
9	1	7	8	2	18
10	0	8	4	0	12
11	2	5	6	0	13
12	0	5	9	3	17
13	0	2	10	5	17
14	0	6	8	1	15
15	2	1	0	0	3
16	12	1	0	0	13
*	4	7	1	1	13
Total	53	73	142	32	300

*Zero-flow stations (zones 8, 10 & 11 only)

Table 14-2

NUMBER OF STATIONS WHERE ONE OR MORE OBSERVED FLOOD EVENTS EXCEEDS THE 1000-YR FLOW COMPUTED FROM COMPLETE RECORD

ZONE	STATION-YEARS OF RECORD	METHOD 1	2	3	4	5	6	7	8
1	1414	0	1	8	0	10	7	2	26
2	1074	0	3	9	0	10	7	1	19
3	1223	1	3	7	0	9	8	4	22
4	703	1	2	3	0	3	3	2	12
5	990	2	1	7	0	4	4	0	19
6	1124	0	2	4	0	4	4	1	18
7	852	1	2	5	1	3	4	3	17
8	969	1	1	10	0	3	3	1	19
9	920	3	0	4	0	3	3	1	16
10	636	1	0	2	0	1	1	0	10
11	594	1	1	6	0	4	4	0	11
12	777	0	2	2	0	2	2	2	9
13	911	1	0	1	0	4	2	2	14
14	761	0	0	3	0	4	1	1	15
15	120	0	0	0	0	0	0	0	2
16	637	1	0	4	0	4	3	0	12
*	495	1	0	2	0	0	0	0	12
TOTAL	14,200	14	18	77	1	68	56	20	253

Based on the 14,200 station-years of record, it might be expected that about 14 observed events would exceed the true 1000-year magnitudes.

*Zero-flow stations

Table 14-3

STANDARD DEVIATION COMPARISONS

AVERAGE FOR ZONES 1 TO 16

COMPUTED PROBABILITY	METHOD 1	2	3	4	5	6	7	8
	AVERAGE OBSERVED PROBABILITIES							
.001	.0105	.0041	.0109	.0001	.0110	.0092	.0045	.0009
.01	.0232	.0153	.0315	.0023	.0309	.0244	.0170	.0015
.1	.1088	.1007	.1219	.0707	.1152	.1047	.1020	.0029
.5	.5090	.5149	.4576	.6152	.4713	.4950	.5108	.0037
	STANDARD DEVIATION OF OBSERVED PROBABILITIES FOR SPECIFIED COMPUTED PROBABILITIES							
.001	.0290	.0134	.0244	.0025	.0239	.0218	.0150	.0222
.01	.0430	.029	.045	.010	.043	.039	.032	.035
.1	.086	.084	.089	.074	.089	.084	.084	.067
.5	.132	.131	.142	.133	.133	.141	.130	.123

Note: Averages and standard deviations are of observed frequencies in the reserved portion of each record corresponding to computed mangitudes based on half records. Low standard deviations in relation to averages indicate more reliable estimates.

Table 14-4

Evaluation of Alternative Methods

Accuracy Tests b and c, Average Values, All Stations

Test b--Root mean square difference between plotting position and computed probability in other half of record.

	Method							
	1	2	3	4	5	6	7	8
Maximum	.062	.060	.067	.056	.070	.069	.061	.061
Decile	.084	.080	.097	.063	.098	.094	.081	.082
Median	.254	.105	.657	.193	.518	.295	.120	.727

Test c--Root mean square difference bewteen 1.0 and ratio of computed probability of flow in opposite half of record to plotting position. A zero value would indicate a perfect forecast.

	Method							
	1	2	3	4	5	6	7	8
Maximum	.53	.51	.56	.45	.56	.56	.51	.59
Decile	.37	.34	.38	.27	.37	.37	.34	.40
Median	.40	.12	.65	.19	.59	.44	.14	.52

Table 14-5

Evaluation of Alternative Methods

Consistency Tests a and b, Average Values, All Stations

Test a--Root mean square difference between computed probabilities from the two record halves for full record extreme, largest, upper decile and median events. A zero value would indicate perfect consistency.

Event	Method 1	2	3	4	5	6	7	8
Extreme	.003	.006	.001	.010	.001	.002	.003	.002
Maximum	.023	.019	.008	.016	.008	.010	.010	.012
Upper Decile	.072	.047	.043	.025	.037	.033	.025	.048
Median	.119	.076	.072	.047	.049	.045	.041	.131

Test b--Root mean square value of (1.0 minus the ratio of the smaller to the larger computed probabilities from the two record halves) for full record extreme, largest, upper decile and median events. A zero value would indicate perfect consistency.

Event	Method 1	2	3	4	5	6	7	8
Extreme	.87	.54	.46	.26	.39	.35	.29	.75
Maximum	.74	.45	.41	.21	.34	.30	.24	.72
Upper Decile	.50	.32	.31	.16	.24	.21	.17	.58
Median	.21	.14	.12	.10	.08	.08	.07	.24

Table 14-6
Evaluation of Outlier Techniques
Average Values, All Stations

Method

Accuracy Test b

Outlier Technique	1	2	3	4	5	6	7
a	.061	.062	.071	.057	.074	.073	.062
b	.056	.055	.060	.053	.063	.062	.055
c	.052	.050	.054	.048	.057	.055	.051
d	.047	.045	.048	.044	.051	.050	.045

Accuracy Test c

Outlier Technique	1	2	3	4	5	6	7
a	.53	.55	.57	.47	.58	.58	.54
b	.57	.59	.59	.49	.62	.60	.58
c	.58	.61	.60	.52	.64	.63	.60
d	.65	.65	.64	.38	.68	.65	.64

Consistency Test a

Outlier Technique	1	2	3	4	5	6	7
a	.002	.005	.001	.009	.000	.002	.002
b	.002	.004	.001	.008	.000	.002	.002
c	.003	.003	.000	.007	.000	.002	.002
d	.003	.003	.000	.007	.000	.002	.001

Consistency Test b

Outlier Techniques	1	2	3	4	5	6	7
a	.87	.56	.46	.27	.39	.36	.30
b	.86	.56	.45	.28	.38	.35	.30
c	.85	.56	.45	.29	.38	.35	.30
d	.88	.59	.45	.31	.38	.35	.32

A zero value would indicate perfect consistency.

Method 8 includes its unique technique for outliers and was, therefore, not included in these tests.

Table 14-7

Evaluation of Zero Flow Techniques

Average Values, All Stations

Accuracy Test b

Technique	Method 1	2	3	4	5	6	7
a	.057	.057	.059	.057	.062	.055	.059
b	.064	.060	.070	.057	.068	.061	.061

Accuracy Test c

Technique	Method 1	2	3	4	5	6	7
a	.46	.32	.59	.32	.40	.40	.32
b	.51	.30	.59	.30	.40	.41	.31

Consistency Test a

Technique	Method 1	2	3	4	5	6	7
a	.007	.012	.000	.014	.001	.000	.006
b	.007	.008	.000	.012	.000	.001	.004

Consistency Test b

Technique	Method 1	2	3	4	5	6	7
a	.89	.43	.44	.21	.39	.34	.24
b	.86	.43	.44	.19	.40	.38	.23

Method 8 was not tested because logarithms are not used in its fitting computations and therefore zero flows are not a problem.

Table 14-8

Summary of Partial-Duration Ratios

Zone	Partial-duration frequencies for annual-event frequencies of						
	.1	.2	.3	.4	.5	.6	.7
1 (21 sta)	.094	.203	.328	.475	.641	.844	1.10
2 (17 sta)	.093	.209	.353	.517	.759	1.001	1.30
3 (19 sta)	.094	.206	.368	.507	.664	.862	1.18
4 (8 sta)	.095	.218	.341	.535	.702	.903	1.21
5 (17 sta)	.093	.213	.355	.510	.702	.928	1.34
6 (16 sta)	.134	.267	.393	.575	.774	1.008	1.33
7 (9 sta)	.099	.248	.412	.598	.826	1.077	1.42
8 (12 sta)	.082	.211	.343	.525	.803	1.083	1.52
9 (15 sta)	.106	.234	.385	.553	.765	.982	1.26
10 (12 sta)	.108	.248	.410	.588	.776	1.022	1.34
11 (12 sta)	.094	.230	.389	.577	.836	1.138	1.50
12 (12 sta)	.103	.228	.352	.500	.710	.943	1.21
13 (16 sta)	.095	.224	.372	.562	.768	.986	1.30
14 (14 sta)	.100	.226	.371	.532	.709	.929	1.22
15 (3 sta)	.099	.194	.301	.410	.609	.845	1.05
16 (13 sta)	.106	.232	.355	.522	.696	.912	1.27
Average	.099	.243	.366	.532	.733	.964	1.28
Langbein	.105	.223	.356	.510	.693	.917	1.20

Note: Data limited to 226 stations originally selected for the study.

TABLE 14-9

ADJUSTMEMT RATIOS FOR 10-YEAR FLOOD

SAMPLE SIZE	ZONE 1		27 STATIONS			AVG 1/2 RECORD = 26 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.54	.38	.76	.29	.82	.57	.28	-1.85
10-YR	.75	.45	1.02	-.27	.95	.37	.34	4.56
1/2-REC	1.21	1.11	2.21	-1.04	2.01	1.01	1.03	4.49
	ZONE 2		24 STATIONS			AVG 1/2 RECORD = 22 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.48	.42	1.06	.64	1.03	.93	.41	-1.85
10-YR	1.01	.94	1.91	.68	1.60	1.31	.80	5.70
1/2-REC	1.33	1.33	2.76	-1.58	1.90	.49	.54	7.14
	ZONE 3		25 STATIONS			AVG 1/2 RECORD = 24 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.41	1.32	1.92	1.02	1.95	1.79	1.40	-1.85
10-YR	1.41	.81	1.80	.00	1.87	.96	1.01	5.39
1/2-REC	.98	.14	1.65	-1.88	1.17	.21	.39	4.80
	ZONE 4		15 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.05	.94	1.20	.85	1.29	1.15	.94	-1.85
10-YR	-.52	-.50	.12	-.85	-.01	-.54	-.45	3.68
1/2-REC	.45	.02	1.63	-3.07	1.63	.46	.25	5.57
	ZONE 5		20 STATIONS			AVG 1/2 RECORD = 25 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.55	.35	1.03	.15	.98	.88	.47	-1.85
10-YR	.40	-.03	1.40	-.96	.61	.42	.19	7.37
1/2-REC	.81	-.40	2.91	-3.61	1.42	.99	.67	6.23
	ZONE 6		24 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.80	.36	1.19	.15	1.11	.95	.45	-1.85
10-YR	1.43	.18	2.26	-.98	1.78	.96	.33	5.64
1/2-REC	1.08	-.45	2.94	-3.93	1.94	.07	-.04	6.14
	ZONE 7		21 STATIONS			AVG 1/2 RECORD = 20 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.15	1.19	1.69	1.29	1.62	1.59	1.29	-1.85
10-YR	1.58	1.36	2.34	.12	1.99	1.62	1.57	5.78
1/2-REC	1.97	1.00	2.45	-.74	2.07	.92	1.17	7.11
	ZONE 8		23 STATIONS			AVG 1/2 RECORD = 21 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.89	.79	1.71	.79	1.41	1.36	.79	-1.85
10-YR	-.66	-1.02	.29	-2.04	-.35	-.43	-1.02	4.52
1/2-REC	-.13	-.87	2.28	-3.08	.74	.66	-.87	7.88

TABLE 14-9 CONTINUED

	ZONE 9		18 STATIONS			AVG 1/2 RECORD = 25 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.38	1.02	2.05	.96	1.96	1.78	1.10	-1.85
10-YR	1.95	1.54	2.54	.75	2.49	2.22	1.69	5.76
1/2-REC	.45	-.36	.97	-3.36	.45	-.07	-.27	4.07
	ZONE 10		12 STATIONS			AVG 1/2 RECORD = 26 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	-.79	-.80	-.41	-.83	-.43	-.43	-.77	-1.85
10-YR	-.03	-.42	.90	-1.16	.71	.35	-.22	4.24
1/2-REC	.08	-1.27	1.24	-5.10	.58	-.27	-1.27	2.97
	ZONE 11		13 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.29	1.21	1.89	1.20	1.93	1.75	1.11	-1.85
10-YR	1.11	1.03	2.21	.04	1.87	1.25	1.03	6.78
1/2-REC	.04	-.23	1.99	-2.93	1.20	1.20	-.23	5.32
	ZONE 12		17 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.34	.73	1.34	.57	1.51	1.03	.80	-1.85
10-YR	.79	.41	.86	-.45	.92	-.44	.57	4.06
1/2-REC	.19	-.31	.54	-2.94	.92	-.35	-.19	2.81
	ZONE 13		17 STATIONS			AVG 1/2 RECORD = 26 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.27	1.16	1.65	.96	1.77	1.52	1.19	-1.85
10-YR	.26	.22	.88	-.83	.67	.42	.38	4.60
1/2-REC	-.31	-1.52	.21	-4.89	.17	-.97	-1.12	2.88
	ZONE 14		15 STATIONS			AVG 1/2 RECORD = 25 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.72	1.65	2.12	1.61	2.19	2.00	1.65	-1.85
10-YR	2.60	2.50	3.17	1.88	2.82	1.87	2.56	6.80
1/2-REC	.51	.61	1.83	-1.47	1.30	.29	.75	5.22
	ZONE 15		3 STATIONS			AVG 1/2 RECORD = 20 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	2.47	2.47	2.74	2.55	2.66	2.28	2.28	-1.85
10-YR	1.27	1.27	1.58	1.27	1.58	1.58	1.27	2.65
1/2-REC	3.29	3.29	3.29	2.79	3.29	1.90	3.29	6.33
	ZONE 16		13 STATIONS			AVG 1/2 RECORD = 24 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.69	.75	1.03	.66	1.09	1.05	.75	-1.85
10-YR	.58	.42	.83	-.21	.76	.07	.42	4.24
1/2-REC	1.41	.07	1.68	-3.43	1.25	.64	.07	5.29
	ALL ZONES		287 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.94	.79	1.38	.71	1.37	1.21	.81	-1.85
10-YR	.87	.52	1.52	-.29	1.26	.72	.60	5.27
1/2-REC	.77	.04	1.93	-2.66	1.34	.40	.17	5.36

Values shown are ratios by which the theoretical adjustment for Gaussian-distribution samples must be multiplied in order to convert from the computed 0.1 probability to average observed probabilities in the reserved data. See note table 14-11.

TABLE 14-10

ADJUSTMENT RATIOS FOR 100-YEAR FLOOD

SAMPLE SIZE	ZONE 1		27 STATIONS			AVG 1/2 RECORD = 26 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.35	1.11	1.27	.39	1.61	1.12	.88	-.25
10-YR	1.50	1.10	2.05	-.25	2.42	1.73	.73	3.42
1/2-REC	2.83	2.84	3.90	-1.06	4.89	3.67	1.66	5.28
	ZONE 2		24 STATIONS			AVG 1/2 RECORD = 22 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.91	.79	1.05	.31	1.27	1.13	.63	-.25
10-YR	1.44	1.40	2.48	.63	2.41	2.07	1.37	5.40
1/2-REC	1.00	1.08	3.69	-.82	2.97	2.46	.14	7.16
	ZONE 3		25 STATIONS			AVG 1/2 RECORD = 24 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.80	1.18	1.76	.41	2.05	1.86	1.29	-.25
10-YR	2.42	1.15	2.43	-.04	2.84	1.62	1.32	4.79
1/2-REC	2.90	1.41	3.36	-1.12	3.71	2.76	2.30	5.53
	ZONE 4		15 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.67	1.48	1.45	.59	2.27	2.02	1.64	-.25
10-YR	.67	.35	.56	-.48	1.07	.46	.42	1.50
1/2-REC	1.86	.48	1.54	-1.15	2.83	.88	1.03	3.81
	ZONE 5		20 STATIONS			AVG 1/2 RECORD = 25 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.03	.64	1.37	.24	1.19	1.12	.82	-.25
10-YR	1.22	.57	1.42	-.29	1.27	1.09	.80	5.65
1/2-REC	2.97	.21	4.38	-1.24	2.97	2.39	1.68	7.25
	ZONE 6		24 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.15	.67	1.02	.04	1.17	.88	.76	-.25
10-YR	2.30	.55	1.67	-.27	1.78	1.10	.66	4.43
1/2-REC	1.20	-.23	3.22	-1.24	2.45	.79	.46	5.09
	ZONE 7		21 STATIONS			AVG 1/2 RECORD = 20 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.04	1.07	2.23	.28	2.20	2.16	1.20	-.25
10-YR	1.18	1.09	2.66	-.19	2.54	2.20	1.53	5.40
1/2-REC	3.10	.47	3.92	-.80	2.99	2.29	1.74	8.33
	ZONE 8		23 STATIONS			AVG 1/2 RECORD = 21 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.57	.27	2.08	.01	1.66	1.52	.27	-.25
10-YR	1.30	.14	1.59	-.35	1.15	.93	.14	4.17
1/2-REC	.82	-.32	4.36	-1.13	2.16	2.16	-.32	8.49

TABLE 14-10 CONTINUED

	ZONE 9		18 STATIONS			AVG 1/2 RECORD = 25 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.07	1.33	1.90	.72	2.11	2.11	1.50	-.25
10-YR	2.45	2.23	3.21	.90	3.75	3.55	2.57	4.39
1/2-REC	1.07	.39	2.90	-1.72	3.78	2.38	.66	4.49
	ZONE 10		12 STATIONS			AVG 1/2 RECORD = 26 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	-.10	-.10	.27	-.25	.29	.29	-.06	-.25
10-YR	.21	-.15	.96	-.59	1.06	.75	.15	2.55
1/2-REC	3.29	-.27	1.63	-1.79	2.42	1.32	-.27	4.40
	ZONE 11		13 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.68	.70	1.79	.11	1.58	1.54	.66	-.25
10-YR	2.41	1.51	4.14	.17	3.76	3.43	1.28	6.64
1/2-REC	.30	.79	5.40	-1.08	3.05	2.43	.50	9.77
	ZONE 12		17 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.81	1.10	1.16	.44	1.56	1.19	1.19	-.25
10-YR	1.99	1.93	1.55	.13	2.27	1.04	2.11	2.60
1/2-REC	3.77	1.65	2.12	-1.33	4.39	2.57	1.86	1.82
	ZONE 13		17 STATIONS			AVG 1/2 RECORD = 26 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.63	.87	1.12	.50	1.63	1.26	1.04	-.25
10-YR	.58	.37	1.27	-.28	1.41	1.25	.60	3.28
1/2-REC	1.01	-.07	2.20	-1.81	2.57	1.61	.81	2.69
	ZONE 14		15 STATIONS			AVG 1/2 RECORD = 25 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.54	1.44	1.79	.65	2.43	2.21	1.44	-.25
10-YR	2.92	2.22	2.58	.23	3.53	1.98	2.32	5.16
1/2-REC	2.11	2.80	3.76	-1.52	4.40	3.10	2.80	5.37
	ZONE 15		3 STATIONS			AVG 1/2 RECORD = 20 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	2.09	2.24	2.24	1.24	2.76	1.98	1.50	-.25
10-YR	.26	.26	.26	-.59	1.84	1.84	.26	1.72
1/2-REC	1.80	1.80	.93	-1.31	4.37	3.16	.93	.93
	ZONE 16		13 STATIONS			AVG 1/2 RECORD = 24 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.61	.55	.90	.18	1.30	1.22	.62	-.25
10-YR	1.87	1.23	1.63	-.59	1.83	.99	1.33	3.64
1/2-REC	4.21	1.17	3.96	-1.27	4.41	2.90	2.13	4.46
	ALL ZONES		287 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.16	.90	1.45	.32	1.65	1.45	.94	-.25
10-YR	1.64	1.03	2.01	-.07	2.20	1.62	1.12	4.25
1/2-REC	2.12	.87	3.40	-1.23	3.35	2.30	1.14	5.66

Values shown are ratios by which the theoretical adjustment for Gaussian-distribution samples must be multiplied in order to convert from the computed 0.01 probability to average observed probabilities in the reserved data. See note table 14-11.

TABLE 14-11

ADJUSTMENT RATIOS FOR 1000-YEAR FLOOD

SAMPLE SIZE								
	ZONE 1		27 STATIONS			AVG 1/2 RECORD = 26 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	2.03	1.10	1.19	.21	2.12	1.44	.85	-.04
10-YR	2.30	.88	2.21	-.14	2.98	1.87	.52	4.06
1/2-REC	5.01	4.13	6.94	-.56	10.11	8.16	1.66	8.54
	ZONE 2		24 STATIONS			AVG 1/2 RECORD = 22 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.31	.83	1.18	.15	1.57	1.35	.68	-.04
10-YR	1.98	2.85	3.85	.64	4.45	3.66	2.07	7.41
1/2-REC	1.93	2.11	4.47	-.45	3.56	3.56	1.58	8.81
	ZONE 3		25 STATIONS			AVG 1/2 RECORD = 24 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	2.42	1.22	2.18	-.01	2.54	2.08	1.24	-.04
10-YR	6.06	2.20	3.06	-.14	3.89	1.82	2.20	7.11
1/2-REC	7.41	2.44	6.77	-.51	7.06	4.82	2.77	11.16
	ZONE 4		15 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.88	1.50	1.46	.30	2.48	2.05	1.63	-.04
10-YR	1.24	.54	.47	-.14	1.13	.36	.71	1.33
1/2-REC	2.86	.80	2.11	-.48	3.60	3.60	2.40	2.81
	ZONE 5		20 STATIONS			AVG 1/2 RECORD = 25 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.84	.94	1.36	.49	1.92	1.45	1.32	-.04
10-YR	2.75	.56	2.90	-.14	2.43	2.00	.91	6.02
1/2-REC	5.51	1.39	5.76	-.52	5.89	5.30	3.22	11.70
	ZONE 6		24 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.91	.61	1.08	.07	1.54	1.13	.79	-.04
10-YR	3.99	.57	1.73	-.06	2.33	1.57	1.12	4.53
1/2-REC	2.88	1.38	2.47	-.48	2.06	1.63	1.24	8.92
	ZONE 7		21 STATIONS			AVG 1/2 RECORD = 20 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.19	.82	1.91	.19	2.18	1.89	1.40	-.04
10-YR	2.33	.96	3.58	.13	3.25	2.15	1.53	6.52
1/2-REC	5.99	1.48	5.36	.16	3.90	3.90	2.34	12.51
	ZONE 8		23 STATIONS			AVG 1/2 RECORD = 21 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.83	.09	1.28	-.01	.83	.83	.14	-.04
10-YR	2.79	.42	2.68	-.14	1.78	1.78	.42	5.90
1/2-REC	2.70	.84	7.62	-.41	3.54	3.54	1.32	13.61
	ZONE 9		18 STATIONS			AVG 1/2 RECORD = 25 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.90	1.30	1.37	.49	2.33	2.33	1.55	-.04
10-YR	3.61	3.59	3.22	.42	5.85	5.85	3.90	6.24
1/2-REC	3.59	.59	3.97	-.53	2.68	1.04	1.07	6.92

TABLE 14-11 CONTINUED

	ZONE 10		12 STATIONS			AVG 1/2 RECORD = 26 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.02	-.04	.25	-.04	.22	.22	-.04	-.04
10-YR	.44	-.14	.70	-.14	.67	.43	-.14	3.79
1/2-REC	7.21	.27	3.04	-.56	1.95	1.95	.27	4.50
	ZONE 11		13 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.13	1.01	2.15	.20	2.13	1.78	.94	-.04
10-YR	4.31	2.44	5.95	.72	5.06	3.58	1.90	10.41
1/2-REC	1.74	.91	6.38	-.46	5.01	4.24	.91	15.65
	ZONE 12		17 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	2.84	1.22	1.31	.45	2.03	1.51	1.27	-.04
10-YR	4.30	2.17	2.52	.10	4.27	1.40	2.17	3.37
1/2-REC	8.58	.75	.75	-.46	2.20	1.34	.75	4.59
	ZONE 13		17 STATIONS			AVG 1/2 RECORD = 26 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.89	1.21	1.11	.32	1.92	1.79	1.21	-.04
10-YR	1.27	.36	1.39	-.14	1.77	1.77	.53	3.56
1/2-REC	4.01	-.57	2.83	-.57	3.65	2.43	.55	4.96
	ZONE 14		15 STATIONS			AVG 1/2 RECORD = 25 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.91	1.45	1.56	.47	2.66	2.03	1.45	-.04
10-YR	5.41	2.35	2.81	-.14	4.63	2.17	2.35	5.56
1/2-REC	3.45	1.04	5.12	-.53	9.90	6.99	1.04	6.69
	ZONE 15		3 STATIONS			AVG 1/2 RECORD = 20 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	2.67	3.00	2.54	-.04	3.51	1.25	1.77	-.04
10-YR	-.14	-.14	-.14	-.14	1.87	1.87	-.14	-.14
1/2-REC	2.17	2.17	-.38	-.38	6.15	6.15	-.38	-.38
	ZONE 16		13 STATIONS			AVG 1/2 RECORD = 24 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	.69	.62	1.15	-.04	1.40	1.18	.69	-.04
10-YR	4.02	1.56	3.05	-.14	3.90	1.97	2.01	4.46
1/2-REC	8.74	2.37	7.24	-.51	8.30	6.21	3.76	7.24
	ALL ZONES		287 STATIONS			AVG 1/2 RECORD = 23 YRS		
METHOD	1	2	3	4	5	6	7	8
5-YR	1.60	.95	1.40	.21	1.89	1.54	1.01	-.04
10-YR	3.13	1.40	2.66	.04	3.22	2.19	1.45	5.36
1/2-REC	4.66	1.49	4.81	-.45	4.99	4.02	1.68	8.80

Values shown are ratios by which the theoretical adjustment for Gaussian-distribution samples must be multiplied in order to convert from the computed 0.001 probability to average observed probabilities in the reserved data.

Values in table 14-11 are obtained as follows:

a. Compute the magnitude corresponding to a given exceedance probability for the best-fit function.

b. Count proportion of values in remainder of record that exceed this magnitude.

c. Subtract the specified probability from b.

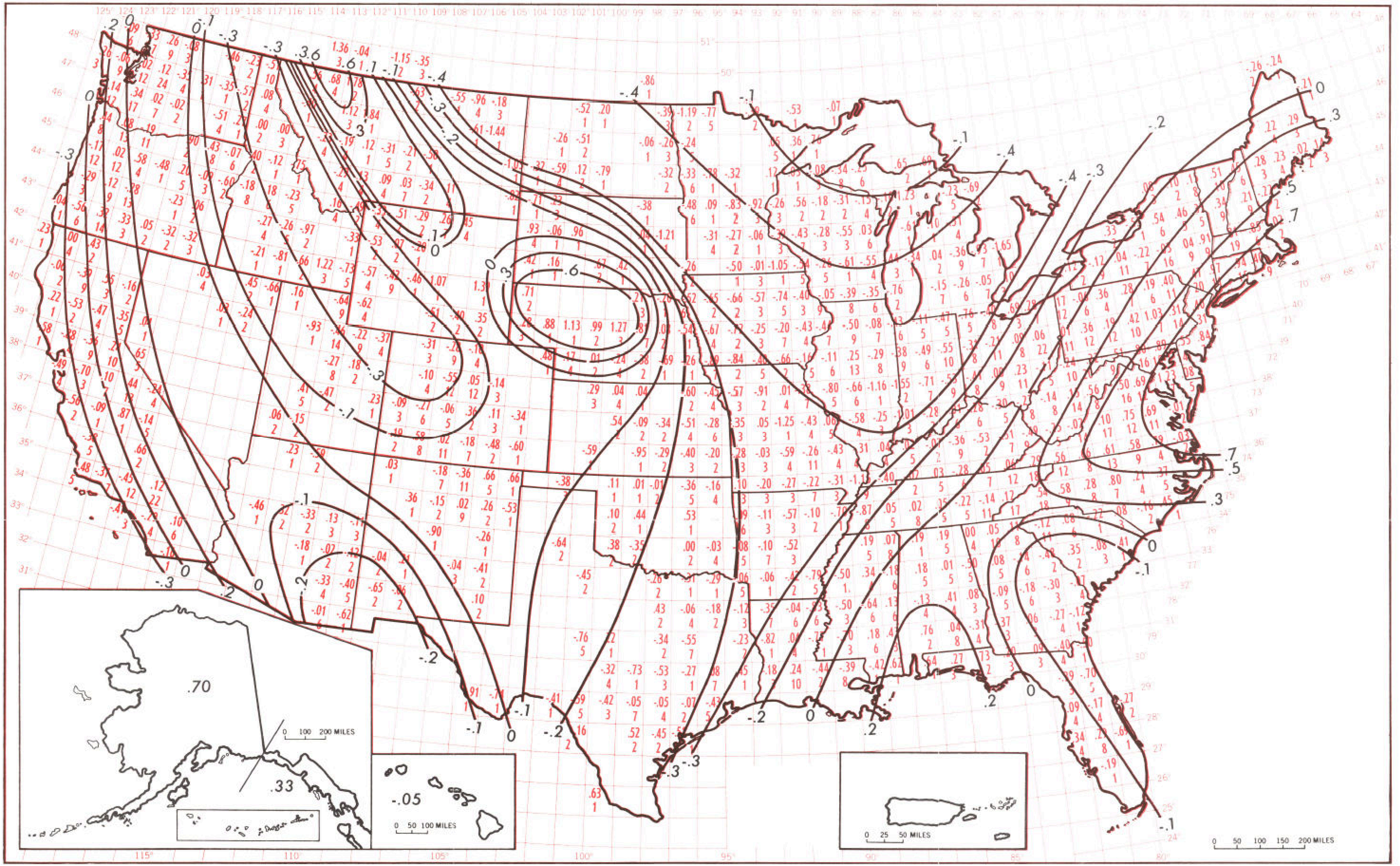
d. Compute the Gaussian deviate that would correspond to the specified probability.

e. Compute the expected probability for the given sample size (record length used) and the Gaussian deviate determined in d.

f. Subtract the specified probability from e.

g. Divide f by c.

☆U.S. GOVERNMENT PRINTING OFFICE:1983- 381-614/209



GENERALIZED SKEW COEFFICIENTS OF LOGARITHMS OF ANNUAL MAXIMUM STREAMFLOW

AVERAGE SKEW COEFFICIENT BY ONE DEGREE QUADRANGLES

Lower number in each quadrangle is number of stream gaging stations for which the average shown above it was computed

GENERALIZED SKEW COEFFICIENTS OF ANNUAL MAXIMUM STREAMFLOW LOGARITHMS

AUGUST 1975 EDITION

The generalized skew map was developed for those guide users who prefer not to develop their own generalized skew relationships. The map was developed from readily available data. Users are encouraged to make detailed studies for their region of interest using the procedures outlined in Section V,B-2. It is expected that Plate I will be revised as more data become available and more extensive studies are completed.

The map is of generalized logarithmic skew coefficients of annual peak discharge. It is based on skew coefficients at 2,972 stream gaging stations. These are all the stations available on USGS tape files with drainage areas equal to or less than 3,000 square miles that had 25 or more years of essentially unregulated annual peaks through water year 1973. Periods when the annual peak discharge likely differed from natural flow by more than about 15 percent were not used. At 144 stations the lowest annual peak was judged to be a low outlier by equation 5 using $\overline{G}$ from figure 14-1 and was not used in computing the skew coefficient. At 28 stations where the annual peak flow for one or more years was zero, only the remaining years were used in computing the low outlier test and in computing the logarithmic skew coefficients. No attempt was made to identify and treat high outliers, to use historic flood information, or to make a detailed evaluation of each frequency curve.

The generalized map of skew coefficients was developed using the averaging technique described in the guide. Preliminary attempts to determine prediction equations relating skew coefficients to basin characteristics indicated that such relations would not appreciably affect the isopleth position. Averages used in defining the isopleths were for groups of 15 or more stations in areas covering four or more one-degree quadrangles of latitude and longitude.

The average skew coefficients for all gaging stations in each one-degree quadrangle of latitude and longitude and the number of stations are also shown on the map. Average skew coefficients for selected groups of one-degree quadrangles were computed by weighting averages for one-degree quadrangles according to the number of stations. The averages for various groups of quadrangles were used to establish the maximum and minimum values shown by the isopleths and to position the intermediate lines.

Because the average skew for 15 or more stations with 25 or more years of record is subject to time sampling error, especially when the stations are closely grouped, the smoothed lines are allowed to depart a few tenths from some group averages. The standard deviation of station values of skew coefficient about the isopleth line is about 0.55 nationwide.

Only enough isopleths are shown to define the variations. Linear interpolation between isopleths is recommended.

The generalized skew coefficient of -0.05 shown for all of Hawaii is the average for 30 stream gaging stations. The generalized skew coefficient of 0.33 shown for southeastern Alaska is the average for the 10 stations in that part of the State. The coefficient of 0.70 shown for the remainder of Alaska is based on skew coefficients at nine stations in the Anchorage-Fairbanks area. The average skew of 0.85 for these nine stations was arbitrarily reduced to the maximum generalized skew coefficient shown for conterminous United States in view of the possibility that the average for the period sampled may be too large.

*This generalized skew map was originally prepared for Bulletin 17 published in 1976. It has not been revised utilizing the techniques recommended in Bulletin 17B.