BULLETIN 17B Computations

Workshop on Determining Flood Frequencies Using Tools from the U.S. Geological Survey

Presented at the 88th Annual Meeting of the Transportation Research Board

January 11, 2009

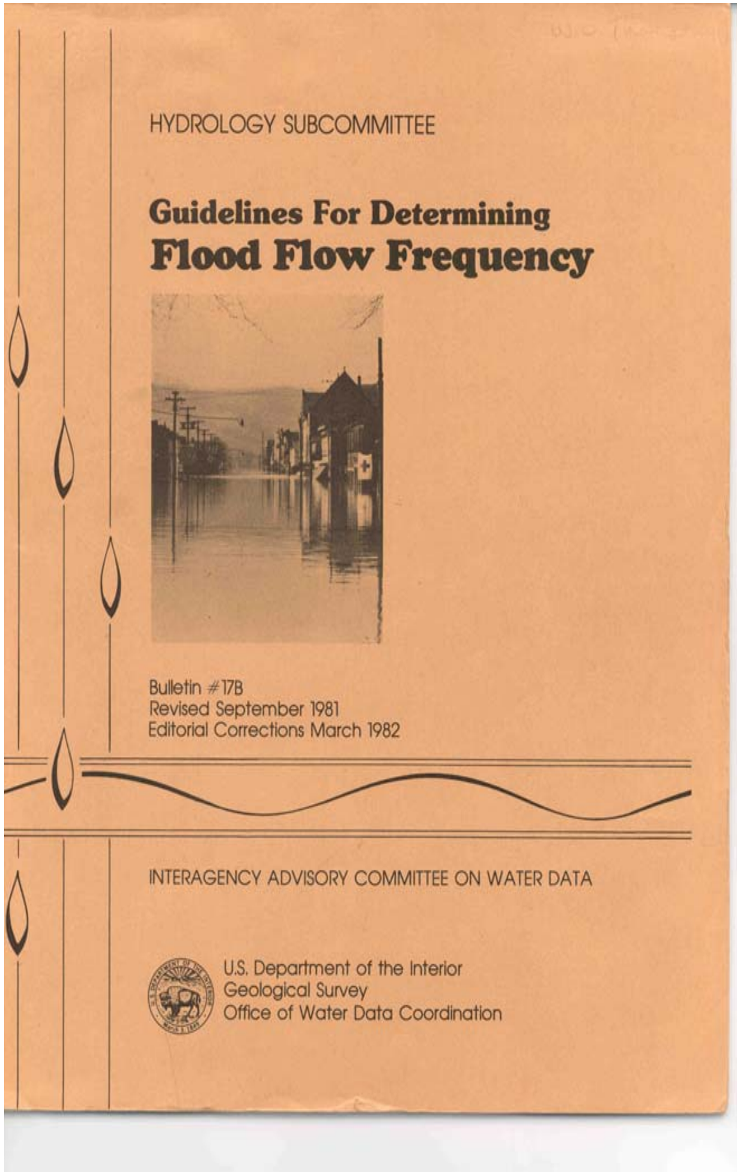
Washington, DC

Wilbert O. Thomas, Jr.

Michael Baker, Jr.

Existing Guidelines - Bulletin 17B

- **Bulletin 17B - Published in 1982, includes guidelines for:**
  - Fitting Pearson Type III distribution to logs of annual peak flows
  - Estimating generalized skew
  - Weighting generalized skew with station skew
  - Low- and high-outlier detection tests
  - Conditional probability adjustment for low outliers
  - Adjustments for historic flood data



Computing T-Year Events with the Pearson Type III Distribution

- **The T-year event is computed by the method of moments as:**

$$X_T = \overline{X} \pm KS$$

where X_T **= T-year low flow or T-year flood event.**

Computing T-Year Events with the Pearson Type III Distribution

$\overline{X}$ = logarithmic mean of annual values

S = logarithmic standard deviation of annual values

K = The Pearson Type III frequency factor (Appendix 3 - Bulletin 17B)

Annual Maximum Peak Discharges

N	06600500			FLOYD RIVER AT JAMES, IOWA		
H	06600500	423448	961836	1919		882.00
3	06600500		19350628	1460		15.20
3	06600500		19360310	40502		18.101
3	06600500		19370527	3570		17.20
3	06600500		19380915	2060		16.50
3	06600500		19390312	13002		16.101
3	06600500		19400605	1390		15.40
3	06600500		19410311	1720		16.20
3	06600500		19420604	6280		18.80
3	06600500		19430617	1360		15.20
3	06600500		19440513	7440		18.80
3	06600500		19450312	5320		18.40
3	06600500		19460301	1400		15.30

http://water.usgs.gov/nwis/sw

Baker

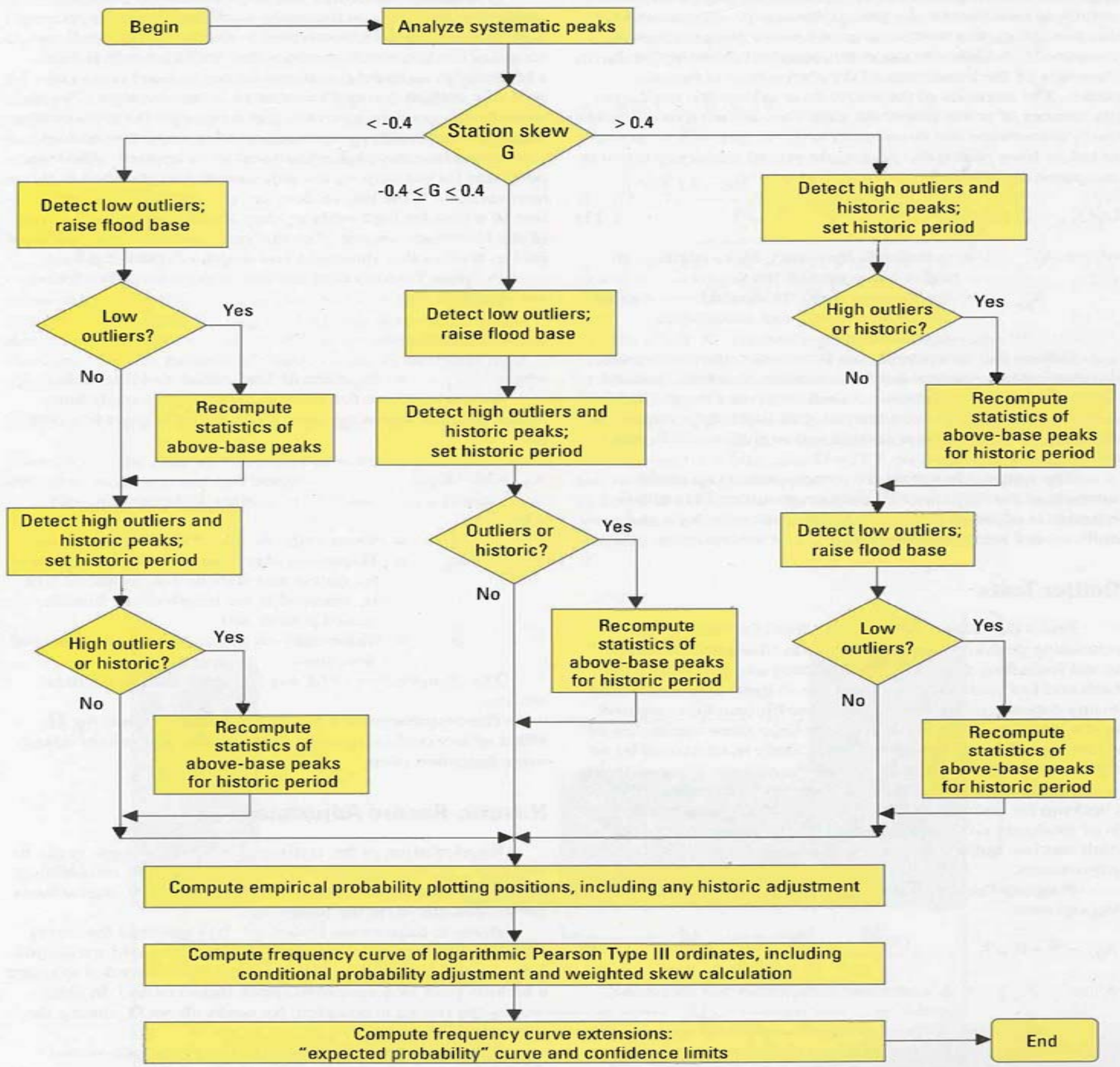


Figure 1. General flow chart for flood-frequency computations (modified from Interagency Advisory Committee on Water Data, 1982).

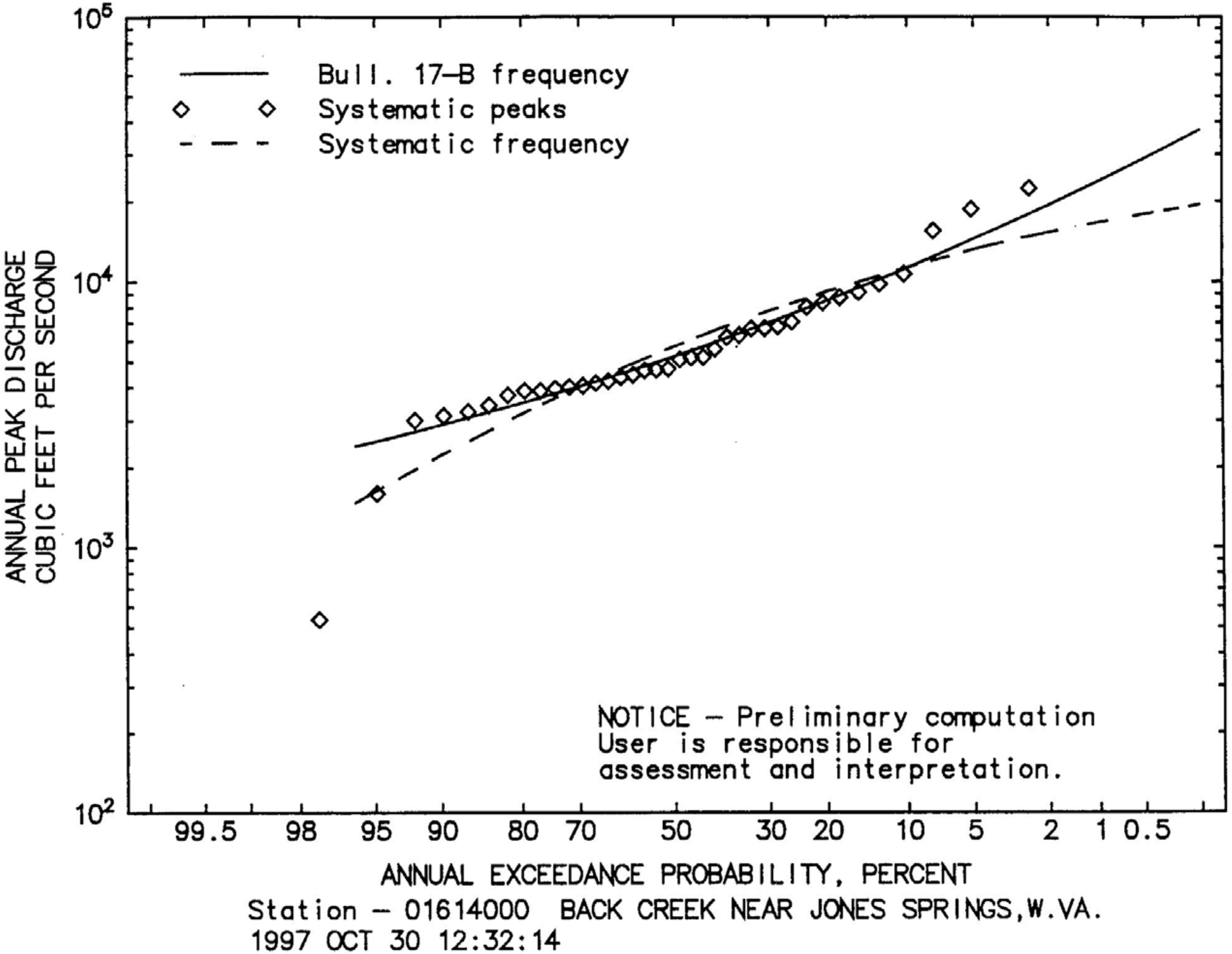
OUTLIERS AND USE OF HISTORIC FLOOD DATA

Definitions:

Outliers - Data points at (or near) the extremity of the frequency curve which departs from the trend of the data due to:

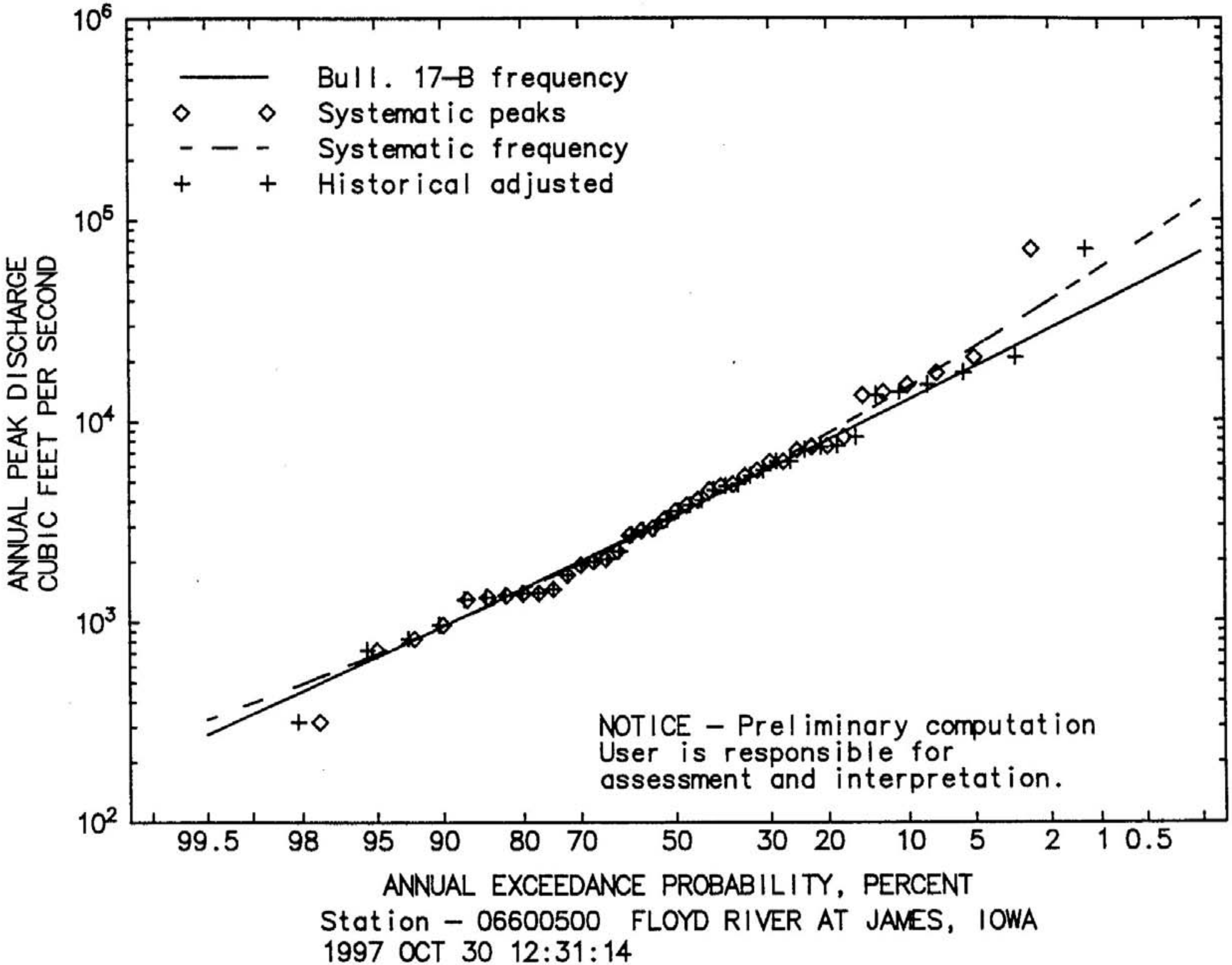
- **measurement error problems**
- **statistical sampling problems**

See examples of low and high outliers in the following two frequency plots.



Bull. 17-B frequency
Systematic peaks
Systematic frequency
ANNUAL PEAK DISCHARGE
CUBIC FEET PER SECOND
10^5
10^4
10^3
10^2
NOTICE — Preliminary computation
User is responsible for
assessment and interpretation.
99.5 98 95 90 80 70 50 30 20 10 5 2 1 0.5
ANNUAL EXCEEDANCE PROBABILITY, PERCENT
Station — 01614000 BACK CREEK NEAR JONES SPRINGS, W.VA.
1997 OCT 30 12:32:14

Baker

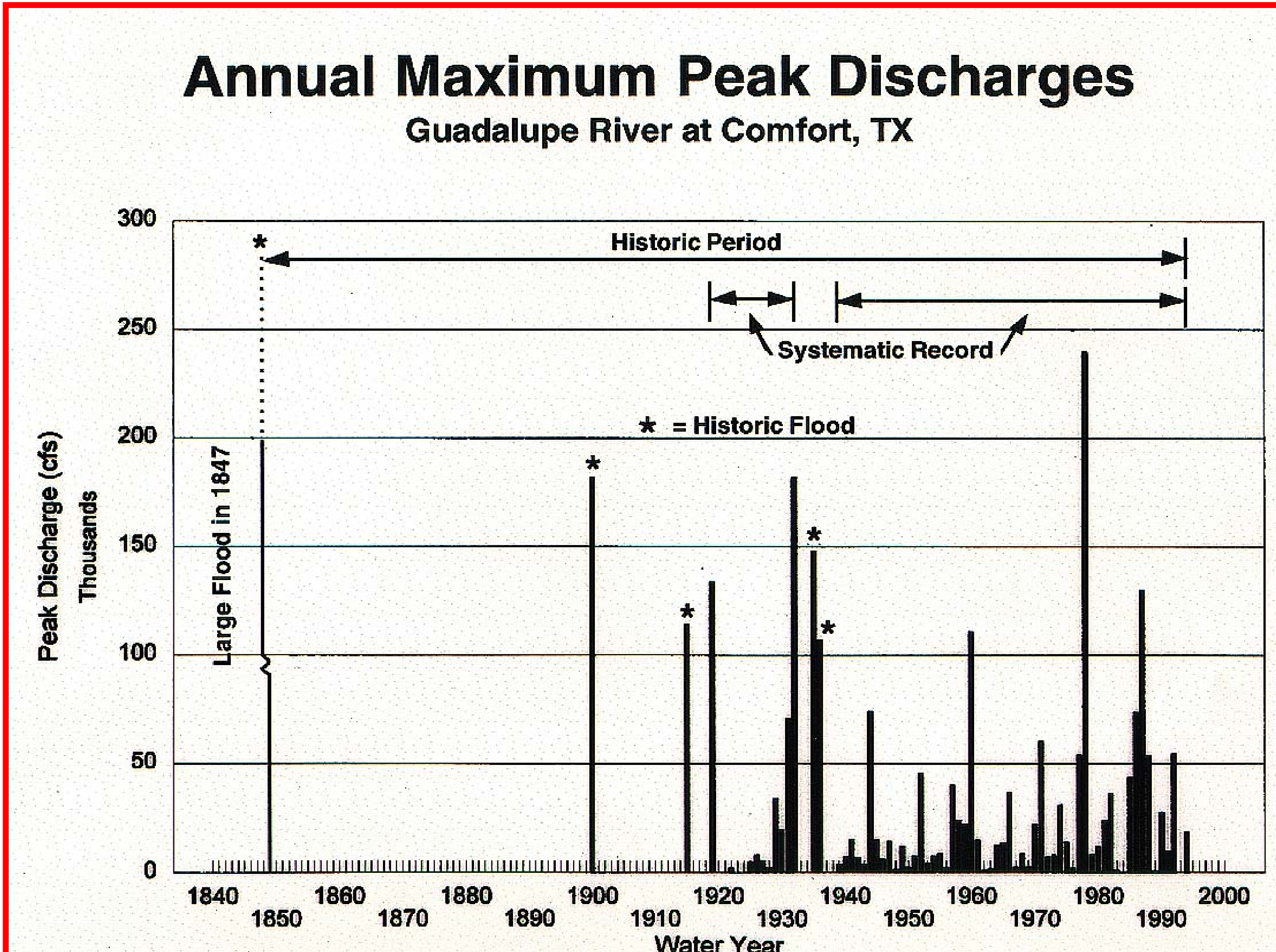


Bull. 17-B frequency
Systematic peaks
Systematic frequency
Historical adjusted
NOTICE – Preliminary computation
User is responsible for
assessment and interpretation.
ANNUAL PEAK DISCHARGE
CUBIC FEET PER SECOND
10^6
10^5
10^4
10^3
10^2
99.5
98
95
90
80
70
50
30
20
10
5
2
1
0.5
ANNUAL EXCEEDANCE PROBABILITY, PERCENT
Station – 06600500 FLOYD RIVER AT JAMES, IOWA
1997 OCT 30 12:31:14

OUTLIERS AND USE OF HISTORIC DATA

Historic Data - are large events that occurred outside of the systematic record (gaged period) for which magnitudes have been estimated.

Historic Information - knowledge gained from residents, newspaper accounts, reports, or other information, which provide guidance on the number of exceedances of a large event(s) within some specified historic period.


Annual Maximum Peak Discharges
Guadalupe River at Comfort, TX
Historic Period
Systematic Record
* = Historic Flood
Large Flood in 1847
Peak Discharge (cfs)
Thousands
300
250
200
150
100
50
0
1840
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
Water Year

HISTORIC ADJUSTMENT PROCEDURE

Compute statistics of systematic and historic data.

Compute the systematic record weight,

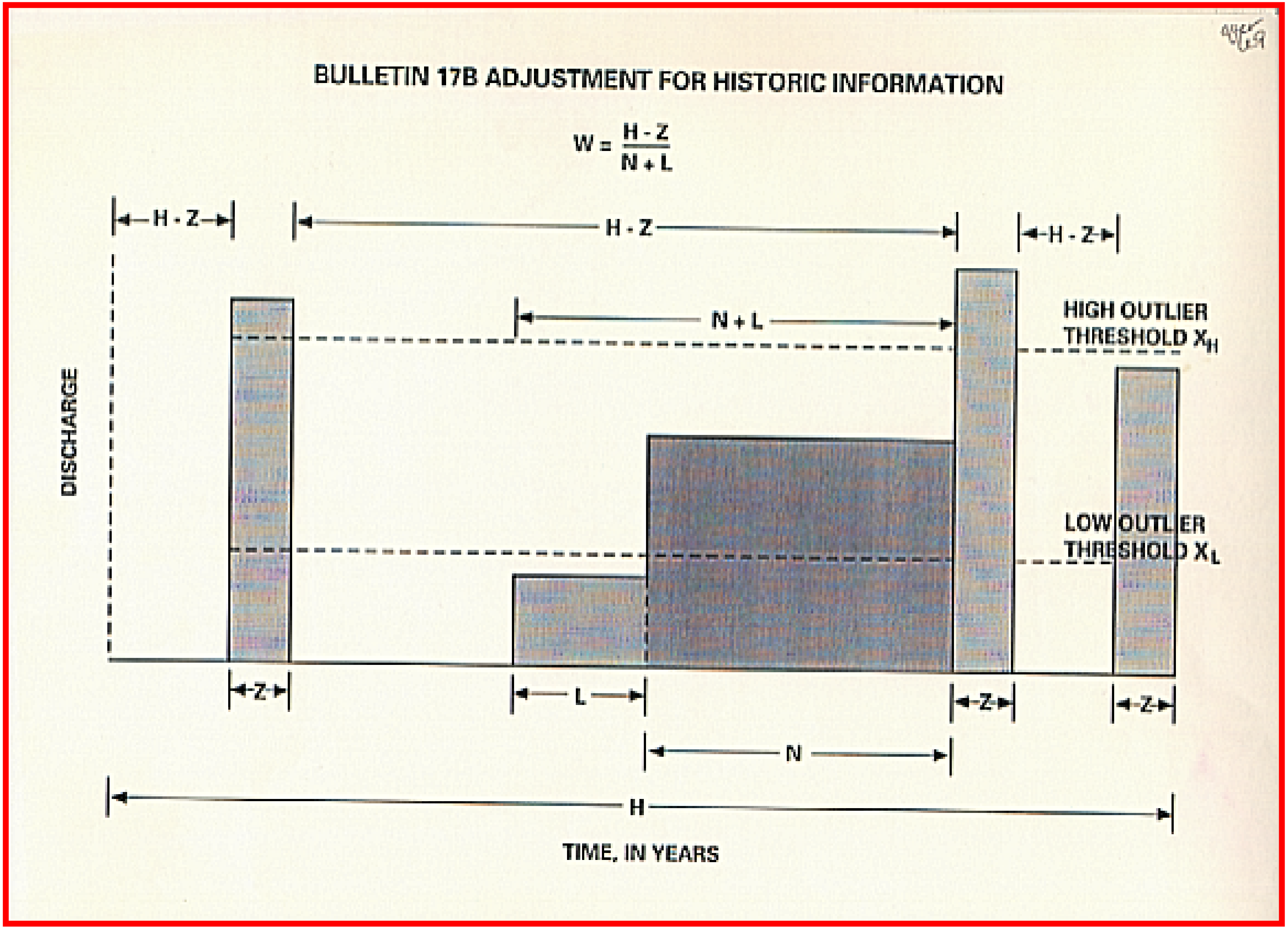
$W = (H - Z) / (N + L)$ - defined in following figure.

H = historic period, Z = number of historic/high outliers, N = systematic years of record, L = number of low outliers.

Historic peaks are given a weight of 1.

Compute the weighted mean, standard deviation and skew.

HISTORIC ADJUSTMENT PROCEDURE



DETECTION OF OUTLIERS

High outlier (X_H) threshold value is computed by:

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

Where:

X_H = high outlier threshold (log units)

$\overline{X}$ = logarithmic mean of systematic peaks excluding:

- zero flood events
- peaks below gage base
- outliers previously detected

DETECTION OF OUTLIERS

S **= logarithmic standard deviation**

K_N **=** K **value from Appendix 4 for sample size** N

Any peaks greater than threshold are considered high outliers. Historic information is needed to "adjust" the high outlier(s); if no available historic information, high outlier is left in systematic record.

*

Appendix 4

OUTLIER TEST K VALUES

10 PERCENT SIGNIFICANCE LEVEL K VALUES

The table below contains one sided 10 percent significance level K_N values for a normal distribution (38). Tests conducted to select the outlier detection procedures used in this report indicate these K_N values are applicable to log-Pearson Type III distributions over the tested range of skew values.

Sample size	K_N value	Sample size	K_N value	Sample size	K_N value	Sample size	K_N value
10	2.036	45	2.727	80	2.940	115	3.064
11	2.088	46	2.736	81	2.945	116	3.067
12	2.134	47	2.744	82	2.949	117	3.070
13	2.175	48	2.753	83	2.953	118	3.073
14	2.213	49	2.760	84	2.957	119	3.075
15	2.247	50	2.768	85	2.961	120	3.078
16	2.279	51	2.775	86	2.966	121	3.081
17	2.309	52	2.783	87	2.970	122	3.083
18	2.335	53	2.790	88	2.973	123	3.086
19	2.361	54	2.798	89	2.977	124	3.089
20	2.385	55	2.804	90	2.981	125	3.092
21	2.408	56	2.811	91	2.984	126	3.095
22	2.429	57	2.818	92	2.989	127	3.097
23	2.448	58	2.824	93	2.993	128	3.100
24	2.467	59	2.831	94	2.996	129	3.102
25	2.486	60	2.837	95	3.000	130	3.104
26	2.502	61	2.842	96	3.003	131	3.107
27	2.519	62	2.849	97	3.006	132	3.109
28	2.534	63	2.854	98	3,011	133	3.112
29	2.549	64	2.860	99	3.014	134	3.114
30	2.563	65	2.866	100	3.017	135	3.116
31	2.577	66	2.871	101	3.021	136	3.119
32	2.591	67	2.877	102	3.024	137	3.122
33	2.604	68	2.883	103	3.027	138	3.124
34	2.616	69	2.888	104	3.030	139	3.126
35	2.628	70	2.893	105	3.033	140	3.129
36	2.639	71	2.897	106	3.037	141	3.131
37	2.650	72	2.903	107	3.040	142	3.133
38	2.661	73	2.908	108	3.043	143	3.135
39	2.671	74	2.912	109	3.046	144	3.138

DETECTION OF OUTLIERS

Low outlier (X_L) threshold value is computed by:

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

Where:

X_L **= low outlier threshold (log units)**

$\overline{X}$ **= logarithmic mean of systematic peaks excluding:**

- **zero flood events**
- **peaks below gage base**
- **outliers previously detected**

DETECTION OF OUTLIERS

S = logarithmic standard deviation

K_N = K value from Appendix 4 for sample size N

Any peaks less than threshold are considered low outliers. When one or more low outliers are identified, they are censored (removed from the computations) and conditional probability adjustment made.

DETECTION OF OUTLIERS

If an adjustment for historic flood data has been made prior to the detection of low outliers, then use the following low outlier threshold equation:

$$\boxed{X_L = \tilde{M} - K_H \tilde{S}}$$

DETECTION OF OUTLIERS

Where:

X_L **= low outlier threshold (log units)**

$\tilde{M}$ **= historically adjusted mean (log units)**

$\tilde{S}$ **= historically adjusted standard deviation (log units)**

K_H **=** K **value from Appendix 4 for sample size** H

When one or more low outliers are identified, they are censored (removed from the computations) and conditional probability adjustment made.

BASIS FOR OUTLIER TESTS

The K_N values in Bulletin 17B are for a one-sided 10 percent significance level test (Grubbs and Beck, 1972).

- **Test for high and low outliers are made separately (one-sided)**
- **Hypothesis - there are no outliers**
- **10 percent chance of rejecting true hypothesis**

BASIS FOR OUTLIER TESTS

The test (Grubbs and Beck, 1972) was developed for detection of a single outlier from a normal distribution, but is used in Bulletin 17B for detection of multiple outliers for a Pearson Type III distribution.

Bulletin 17B Work Group evaluated various outlier tests using observed and simulated data (Thomas, 1985).

CONDITIONAL PROBABILITY ADJUSTMENT

Appropriate when annual peaks are less than the base of partial-record stations, when zero flows occur or when low outliers are identified in the record.

Basic steps in computations are:

1. Compute the frequency curve using the above base (or above low outlier criterion) peaks and station skew including detection of outliers and incorporation of historic information. This gives the conditional frequency curve with exceedance probabilities P_d **.**

CONDITIONAL PROBABILITY ADJUSTMENT

2. **Calculate the estimated probability $\tilde{P}$ that any annual peak will exceed the truncation level**

   where *N* is the number of peaks above the truncation level and *n* is the total number of years of record. The equation is (Jennings and Benson, 1969):

$$\boxed{\tilde{P} = \frac{N}{n}}$$

CONDITIONAL PROBABILITY ADJUSTMENT

If historic information has been included, then use the following formula:

$$\boxed{\tilde{P} = \frac{H - WL}{H}}$$

where H **is the historic record length,** L **the number of peaks truncated, and W the systematic record weight.**

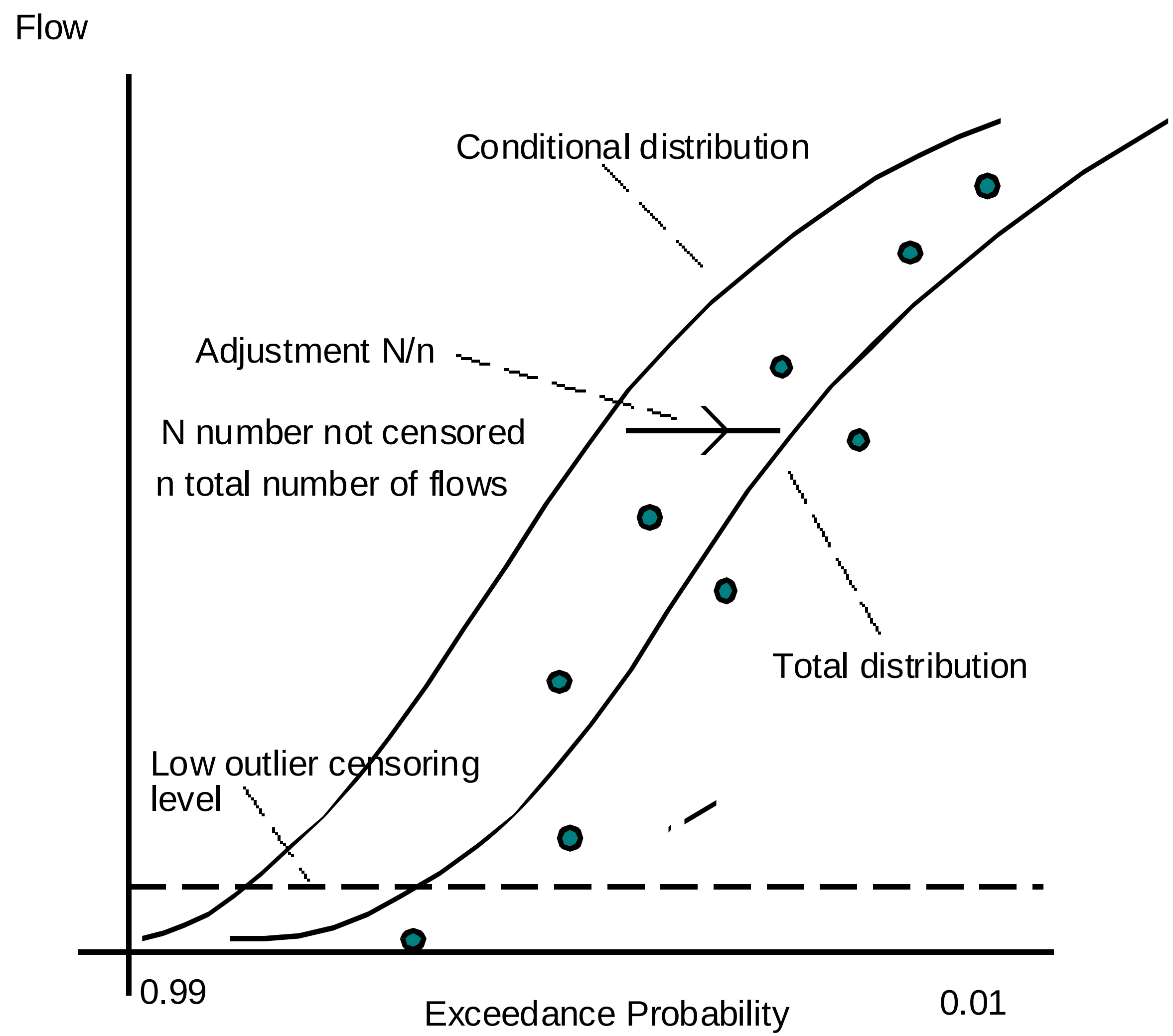
Baker

CONDITIONAL PROBABILITY ADJUSTMENT

3. Adjust the frequency curve computed in 1 above for those peaks below the truncation level. This gives the unconditional frequency curve with exceedance probabilities

$$\boxed{P = \tilde{P} * P_d}$$

4. Interpolate either graphically or mathematically using the unconditional frequency curve in 3 above to obtain the .01, .10 and .50 exceedance probability discharges.


Flow
Conditional distribution
Adjustment N/n
N number not censored
n total number of flows
Total distribution
Low outlier censoring
level
0.99
Exceedance Probability
0.01

CONDITIONAL PROBABILITY ADJUSTMENT

5. Compute the synthetic skew (G_S) standard deviation (S_S) and mean ($\bar{X}_s$) using the three exceedance probabilities from 4 above and the following equations:

$$G_S = -2.30 + 3.12 \{(\log (Q_{.01}/Q_{.10}) / \log (Q_{.10}/Q_{.50})\}$$

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

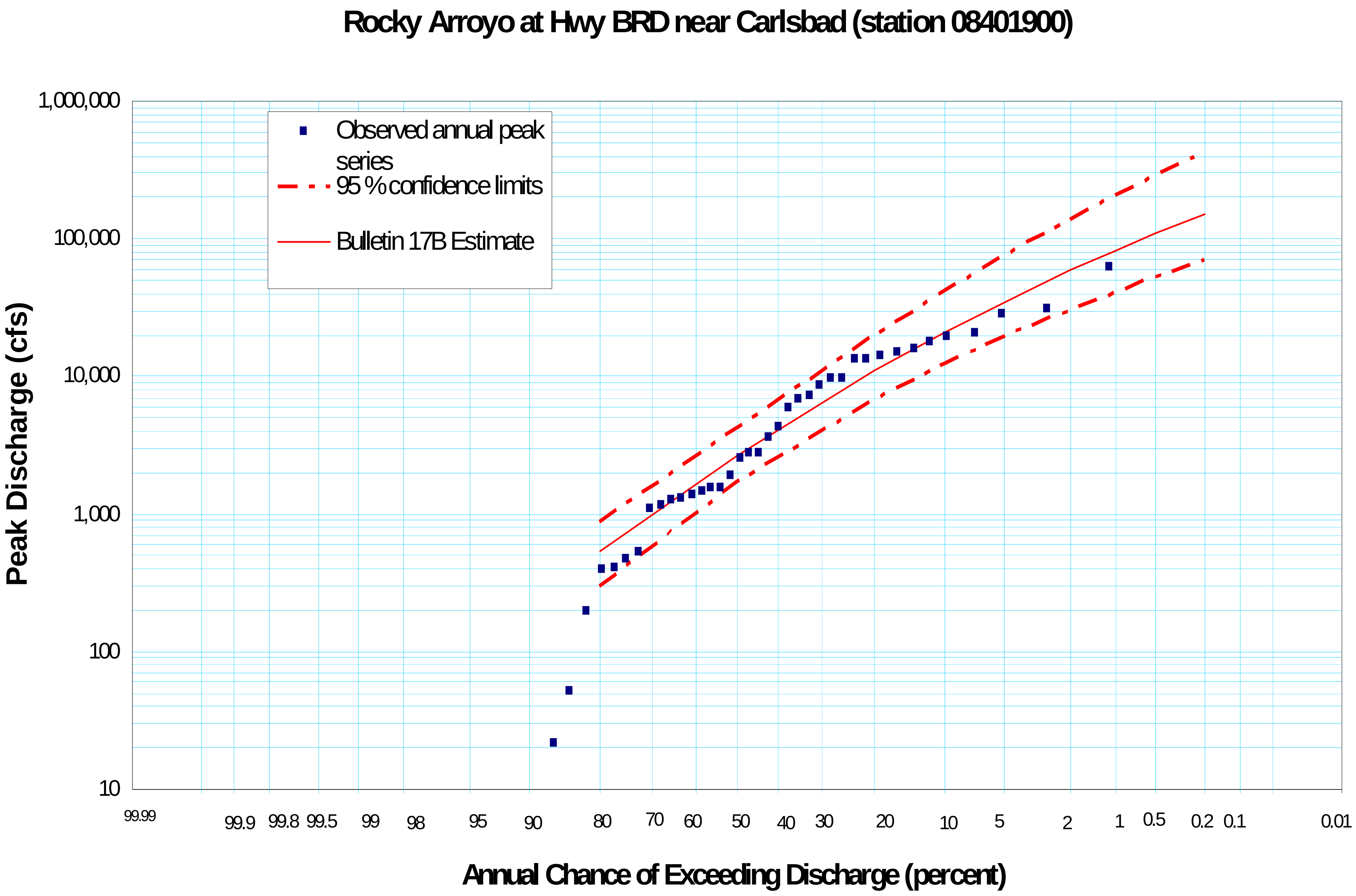
$$\bar{X}_s = \log (Q_{.50}) - K_{.50} (S_S)$$

CONDITIONAL PROBABILITY ADJUSTMENT

Where $Q_{.01}$, $Q_{.10}$ and $Q_{.50}$ are discharges with .01, .10, and .50 exceedance probabilities and $K_{.01}$, $K_{.10}$, and $K_{.50}$ are the corresponding Pearson Type III deviates.

6. **The final step is to compute the weighted skew using G_S computed in 5 above and the generalized skew G.**

Baker


Rocky Arroyo at Hwy BRD near Carlsbad (station 08401900)
Observed annual peak series
95 % confidence limits
Bulletin 17B Estimate
1,000,000
100,000
10,000
1,000
100
10
Peak Discharge (cfs)
99.99
99.9
99.8
99.5
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.5
0.2
0.1
0.01
Annual Chance of Exceeding Discharge (percent)

MOTIVATION FOR GENERALIZED SKEW

Computation of coefficient of skewness is needed in Bulletin 17B method of moments approach.

There is large uncertainty in computing coefficient of skewness for sample sizes commonly available in flood-frequency analysis.

Generalized skew is used to reduce uncertainty in estimating T-year events.

Baker

PLATE I IN BULLETIN 17B

- **Map based on 2,972 stations that had 25 or more years of record through 1973 and drainage areas less than 3,000 square miles.**
- **Only 144 low outliers were found based on the Bulletin 17 (not 17B) low-outlier test that was not very sensitive.**
- **Historic flood information was not used.**
- **The national map has a mean-square error (MSE) of 0.302. Regional values of the MSE would probably be more appropriate.**

Baker

WEIGHTING THE SKEW COEFFICIENT

The weighted skew coefficient is computed as follows:

$$\boxed{G_W = \frac{MSE_{\overline{G}}(G) + MSE_G\left(\overline{G}\right)}{MSE_{\overline{G}} + MSE_G}}$$

WEIGHTING THE SKEW COEFFICIENT

G_W = weighted skew coefficient

G = station skew

$\overline{G}$ = generalized skew

$MSE_{\overline{G}}$ = mean square error of generalized skew

MSE_G = mean square error of station skew.

The concept of weighting the station and generalized skew in proportion to their mean square errors was based on work by Tasker (1978).

Baker

WEIGHTING THE SKEW COEFFICIENT

The mean square error of the station skew (MSE_G) can be determined from Wallis, Matalas, and Slack (1974). MSE_G can be computed as:

$$MSE_G = \text{(Bias of skew coefficient)}^2 + \text{variance of skew coefficient}$$

The bias and variance of station skew coefficients for Pearson Type III random variables can be obtained from Wallis, Matalas, and Slack (1974).

Baker

WEIGHTING THE SKEW COEFFICIENT

The following equation (Wallis, Matalas, and Slack (1974) is used for computing MSE_G as a function of record length and skew

$$MSE_G = 10^{[A - B [LOG_{10} (N/10)]]}$$

Where

$$A = -0.33 + 0.08 |G| \text{ if } |G| \leq 0.90$$
$$-0.52 + 0.30 |G| \text{ if } |G| > 0.90$$
$$B = 0.94 - 0.26 |G| \text{ if } |G| \leq 1.50$$
$$0.55 \text{ if } > 1.50$$

Baker

REFERENCES

Beard, L.R., 1974, Flood flow frequency techniques: Center for Research in Water Resources, Technical Report CRWR-119, The University of Texas, Austin, Texas.

Benson, M.A., "Uniform Flood-Frequency Estimating Methods for Federal Agencies," *Water Resources Research,* Vol. 4, No. 5, Oct., 1968, pp. 891-908.

Grubbs, F.E., and Beck, G., "Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations," *Technometrics*, Vol. 14, No. 4, 1972, pp. 847-854.

Hardison, C.H., 1974, Generalized skew coefficient of annual floods in the United States: Water Resources Research, vol. 10, n0. 4, pp. 745-752.

Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency: Bulletin 17B of the Hydrology Subcommittee, U.S. Geological Survey, Office of Water Data Coordination, Reston, Virginia.

Jennings, M.E., and Benson, M.A., 1969, Frequency Curves for Annual Flood Series with Some Zero Flow Events or Incomplete Data: Water Resources Research, Vol. 5, No. 1, pp. 276-280.

National Academy of Sciences, 1978, Estimation of peak flows: prepared by the Task Force on Estimation of Peak Flows of the Panel on Flood Studies in Riverine Areas, Building Research Advisory Board, National Research Council, Washington, DC, 10p.

Baker

Natural Environment Research Council, 1975, Flood Studies Report, Volume I – Hydrological Studies, Water Resources Publications, Fort Collins, Colorado.

Tasker, G.D., 1978, Flood frequency analysis with a generalized skew coefficient: Water Resources Research, vol. 14, no.2, pp. 373-376.

Thomas, W.O., Jr., 1985, A uniform technique for flood frequency analysis: ASCE Journal of Water Resources Planning and Management, vol. 111, no. 3, pp. 321-327.

Wallis, J.R., Matalas, N.C., and Slack, J.R., 1974, Just a moment!: Water Resources Research, vol. 10, no. 2.