Asquith, W.H., Slade, R.M. Jr., and Judd, L.J., 1995, Analysis of low-outlier thresholds for log Pearson Type III peak-streamflow frequency analysis in Texas, *in* Texas Water'95 component conference of American Society of Civil Engineers first international conference, San Antonio, Tex., 1995, Proceedings: San Antonio, Tex., American Society of Civil Engineers, p. 379-84.

Analysis of Low-Outlier Thresholds for Log Pearson Type III Peak-Streamflow Frequency Analysis in Texas

By William H. Asquith[1], Raymond M. Slade[2], Jr., *and* Linda J. Judd[3]

ABSTRACT

An evaluation of values for low-outlier thresholds pertinent to Log Pearson Type III peak-streamflow frequency analysis was done for 262 streamflow-gaging stations in Texas with at least 20 years of data from natural drainage basins. A natural basin contains less than 10 percent impervious cover and has less than 10 percent of its drainage area controlled by reservoirs. Two-hundred seven of the 262 stations have a distribution of annual peak discharges that require the selection of a low-outlier threshold in order to improve the fit of the Log Pearson Type III distribution curve. About 87 percent of the 207 stations require a low-outlier threshold higher than that estimated by the Interagency Advisory Committee on Water Data. Therefore, about 13 percent of the stations require a low-outlier threshold lower than or equal to that estimated by the Committee. Multiple linear regression analysis was used to estimate low-outlier threshold values, based on selected statistics of the annual peak discharges. These statistics are the mean, standard deviation, and skew of the logarithms of the annual peak discharges. The resulting regression equation is presented, with discussion of its use and limitations.

INTRODUCTION

Low outliers represent small values for annual peak streamflow—those values that depart from the low end of the fitted peak-streamflow frequency relation. The presence of low outliers can cause the Log Pearson Type III distribution (LPIII) curve to improperly fit the peak discharges, especially those for the larger recurrence intervals. Consequently, low outliers must be identified by a low-outlier threshold in order to improve the peak-streamflow frequency relation. As shown in table 1, the selection of a low-outlier threshold can substantially affect the LPIII frequency curve. Therefore, careful consideration should be given to select the most appropriate value for the threshold. The purpose of this report is to identify, for Texas stations, a procedure for determining appropriate values for low-outlier thresholds.

The Interagency Advisory Committee (IAC) on Water Data (1982, p. 17–19 and Appendix 5) recommends that a statistical test be applied to identify low-outlier threshold values. In a recent peak-streamflow frequency study done by Thomas and others (1994) for the southwestern United States, it was determined that the IAC's threshold-selection procedure was not appropriate for identifying low-outlier thresholds in the southwestern United States (S.D. Waltemeyer, U.S. Geological Survey, written comm., 1995). Therefore, Thomas and others developed a procedure for identifying thresholds. About 48 percent of the 1,059 streamflow-gaging stations analyzed by Thomas and others for peak-streamflow frequency required low-outlier thresholds that identified one or more annual peak discharges.

[1] Civil Engineer, U.S. Geological Survey, Austin, Texas.

[2] Hydrologist, U.S. Geological Survey, Austin, Texas.

[3] Environmental Engineer, U.S. Geological Survey, Austin Texas.

Table 1. Effect of various values of low-outlier threshold on computed values for the 100-year peak discharge, West Nueces River near Bracketville, Texas

Systematic record: 47 years

Historical record: 113 years

Historical peak: 550,000 cubic feet per second

Type of skew calculation: station—that produced by systematic and historical record only

Low-outlier threshold (cubic feet per second)	Number of low outliers removed	Computed 100-year peak discharge (cubic feet per second)
0.0	2	1,480,000
1.0	6	647,000
10.0	9	262,000
12.0	11	251,000
52.0	12	255,000
100	13	308,000

SELECTION OF LOW-OUTLIER THRESHOLDS IN TEXAS

Typically, thresholds other than those provided by the IAC procedure are determined primarily by visually fitting the LPIII curve to the annual peak discharges (Thomas and others, 1994). Various thresholds are used, and the resulting peak-streamflow frequency curves produced by the LPIII are visually evaluated. The appropriate threshold is that associated with the curve best fitting the data. Visually fitting the LPIII curve without regard to the threshold provided by the IAC was the primary procedure used to identify the threshold values for 262 selected long-term Texas stations—those with at least 20 years of annual peak discharges from natural basins. A natural basin is a basin for which the annual peak discharges are not affected by reservoirs, regulations, diversion, urbanization, or other anthropogenic activities. In addition to visual fitting of the LPIII curve, comparisons of the peak-streamflow frequency for nearby stations were made in order to further justify the associated thresholds. These comparisons were based on the relation between peak-streamflow frequency for nearby stations and selected basin characteristics such as contributing basin drainage area, main channel slope, main channel length, and a basin shape factor (length squared divided by basin drainage area). Additionally, those stations with skew values greater than 1.0 or less than -1.0 were reconsidered for further threshold adjustments.

Of the 262 stations, 207 or about 79 percent require a threshold value that exceeded one or more annual peak discharges; thus no threshold was required for 55 stations. About 87 percent of the 207 stations require a low-outlier threshold larger than that estimated by the IAC procedure. Therefore, about 13 percent of the stations require a low-outlier threshold less than or equal to that estimated by the IAC procedure. The threshold values for the 207 stations are referred to as observed low-outlier thresholds.

In addition to visual fitting, low-outlier thresholds selected by Thomas and others (1994) were limited to a recurrence interval of less than 1.4 years. For the Texas stations, only 6 of the 207 stations have thresholds exceeding the peak discharge associated with a recurrence interval of 1.4 years; none of the thresholds exceeded the 2-year peak discharge.

A comparison was made of the 100-year peak discharges computed as a result of using the observed threshold procedure and the IAC threshold procedure. The results of the comparison for the longest-record stations—those with at least 50 years of annual peak discharges—and the shortest-record stations—those with 20 to 23 years of annual peak discharges—are presented in table 2. These longest- and shortest-record stations are subsets of the 262 stations used in this report and were chosen because they represent the two extremes of the available record length for the stations. The mean standard errors of the percent changes indicate substantial differences in 100-year peak discharges for many stations, regardless of the period of record.

Table 2. Comparison of the mean standard error of the percent changes in 100-year peak discharges computed using low-outlier thresholds identified by the observed threshold procedure and the Interagency Advisory Committee threshold procedure

[Percent change is defined as the 100-year peak discharge from the observed threshold, minus that from the Interagency Advisory Committee threshold, divided by the 100-year peak discharge from the observed threshold. Mean standard error of the percent changes is defined as the square root of the sum of the square of the percent changes divided by the number of stations minus one.]

	Mean standard error of the percent change (in percent)
Longest-record stations: (51 stations)	12.8
Shortest-record stations: (32 stations)	24.5

REGRESSION EQUATION

Multiple Linear Regression

Values for the observed threshold were used as the dependent variable and compared by regression analysis to values for hydrologic and statistical independent variables. Hydrologic variables consist of values for contributing drainage area, main channel slope, main channel length, and a basin shape factor. The hydrologic variables were determined to be insignificant in the analysis. However, three statistical variables based on the annual peak discharges—those annual peak discharges associated with the systematic record only—were found to be significant; these are the mean, standard deviation, and skew coefficient of the annual peak discharges. The values for the dependent variables were transformed to their common logarithms ($\log_{10}$) to improve the linearity of the relations. A stepwise regression procedure was used with the F-value of each successive independent variable greater than 1.5 (alpha = 0.05). All variables in the regression equation are presented in $\log_{10}$ units.

The equation is:

$$\log_{10}(LOT) = a(M) + b(SD) + c(SKEW) + d, \qquad (1)$$

where LOT = low-outlier threshold (cubic feet per second),
M = mean of the logarithms of the systematic annual peak discharges,
SD = standard deviation of the logarithms of the systematic annual peak discharges,
SKEW = skew coefficient of the logarithms of the systematic annual peak discharges,
a = 1.09,
b = -0.584,
c = 0.140, and
d = -0.799.

The correlation coefficient squared (R-squared) for the equation is 0.75, indicating that 75 percent of the variance in the values for the dependent variable is explained by the three independent variables. A comparison of observed low-outlier thresholds and thresholds estimated by regression equation is shown in figure 1. The simple correlation coefficients between the dependent and independent variables are as follows:

Mean of the logarithms of the systematic annual peak discharges	0.84
Standard deviation of the logarithms of the systematic annual peak discharges	-0.52
Skew coefficient of the logarithms of the systematic annual peak discharges	0.23

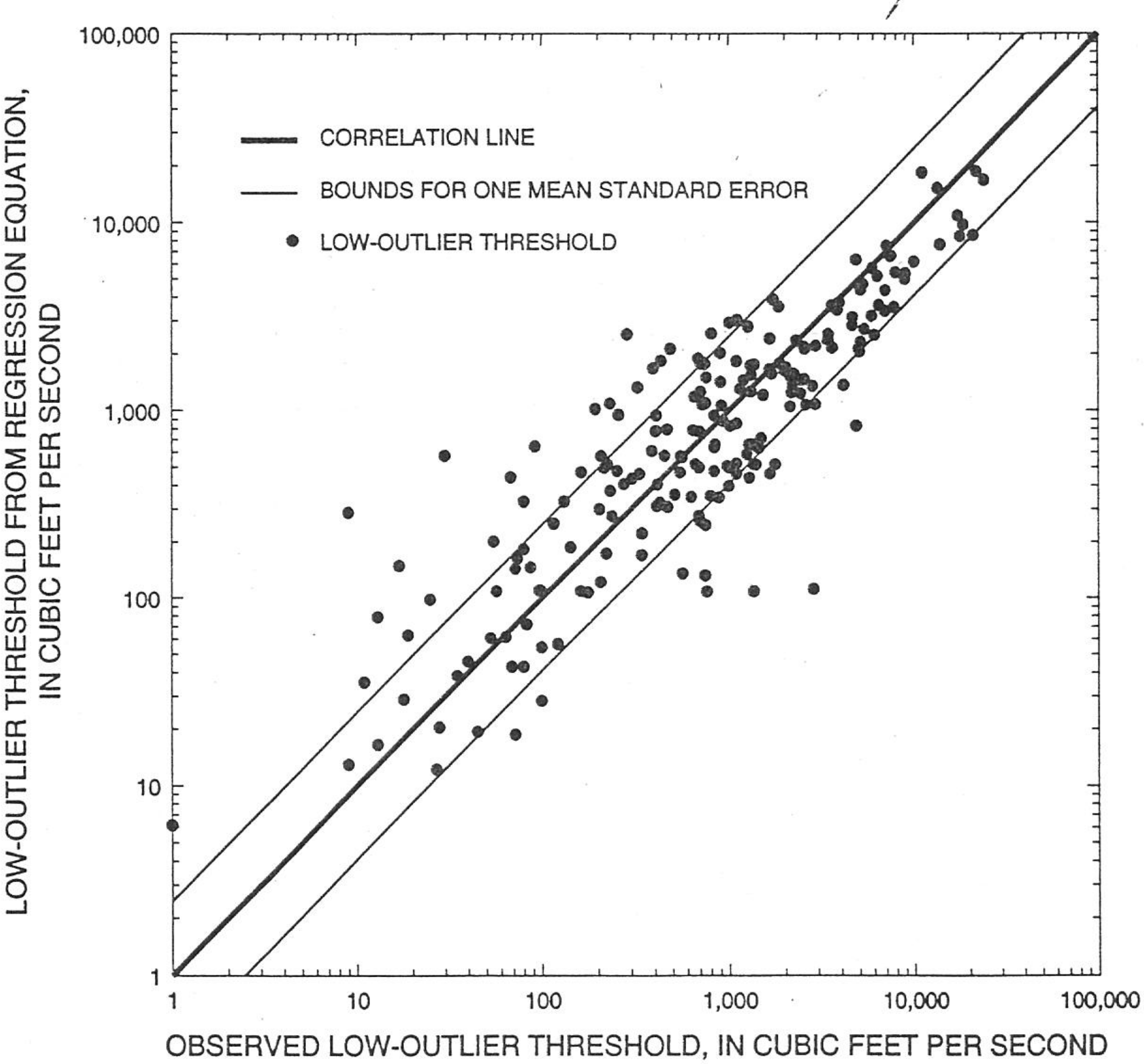

Figure 1. Comparison of observed low-outlier thresholds and thresholds estimated by regression equation for long-term stations in Texas.

Analysis of the residuals of the regression equation was made but is not presented here. The results indicate that the residuals are not correlated with either the length of record or any of the three independent variables, and thus these factors probably do not bias the equation.

The mean standard error of the equation is 0.39 $\log_{10}$ units or about 111 percent; this value is substantially smaller than the standard deviation of the dependent variable, which is 0.78 $\log_{10}$ units or greater than 240 percent. Values for the independent variables were computed by the peak-streamflow frequency analysis; however, potential errors exist in the values for the dependent variable because of the procedure used for selecting those values. Low-outlier threshold values that are between values for sequentially ordered annual peak discharges would not affect the resulting frequency curve. For instance, increasing the value for a specific threshold would not change the resulting frequency curve until that value exceeded the value for an annual peak discharge not excluded by the original threshold value. For example, the frequency curve for the station presented in table 1 would not change for threshold values exceeding 12 but not exceeding 52 cubic feet per second.

Evaluation of Regression Equation

A comparison was made of the 100-year peak discharges computed as a result of using the threshold estimated from equation 1 and that computed from the observed threshold. The results of the comparison, made for the longest-record stations—those with at least 50 years of annual peak discharges—and the shortest-record stations—those with 20 to 23 years of annual peak discharges—are presented in table 3. The mean standard errors of the percent changes are lower than those presented in table 2, which indicates that the regression equation is a more reliable estimator of observed thresholds than is the IAC procedure.

A peak-streamflow frequency analysis was done using thresholds estimated from equation 1 for the 55 stations for which thresholds were not required. The mean standard error of the percent change for the 100-year peak discharge is 10.8 percent. This value indicates that equation 1 does not substantially bias the computed peak discharges for those stations for which thresholds are not required.

Table 3. Comparison of the mean standard error of the percent changes in 100-year peak discharges computed using low-outlier thresholds identified by the observed threshold procedure and the regression equation

[Percent change is defined as the 100-year peak discharge from the observed threshold, minus that from the regression equation threshold, divided by the 100-year peak discharge from the observed threshold. Mean standard error of the percent changes is defined as the square root of the sum of the square of the percent changes divided by the number of stations minus one.]

	Mean standard error of the percent change (in percent)
Longest-record stations: (51 stations)	9.04
Shortest-record stations: (32 stations)	7.73

Limitations

The regression equation can be used to estimate the low-outlier threshold for peak-streamflow frequency analysis for stations in Texas. Equation 1 estimates the appropriate threshold for a station; however that value should not be used without visual verification of the fit of the LPIII curve to the annual peak discharges. The independent variables for the stations used should be within the range of those for which the equation is based. These ranges, by variable, are:

Log_{10} systematic mean	$1.900 < \text{mean} < 4.842$
Log_{10} systematic standard deviation	$0.125 < \text{standard deviation} < 1.814$
Log_{10} systematic skew coefficient	$-2.714 < \text{skew coefficient} < 0.698$

Additionally, the equation was developed from stations with at least 20 years of data; therefore, the applicability of the equation for stations with less than 20 years of data is questionable, and its effects on peak-streamflow frequency analysis should be evaluated. The majority of the contributing drainage areas for the stations used range from 0.2 to 35,441 square miles; therefore, the equation might not be applicable for drainage areas outside this range.

REFERENCES

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

Thomas, B.E., Hjalmarson, H.W., and Waltemeyer, S.D., 1994, Methods of estimating magnitude and frequency of floods in the Southwestern United States: U.S. Geological Survey Open-File Report 93-419, 211 p.