

Subcommittee on Hydrology Hydrologic Frequency Analysis Work Group

Bulletin 17-B Guidelines for Determining Flood Flow Frequency Frequently Asked Questions

INTRODUCTION

The Hydrologic Frequency Analysis Work Group is a work group of the Hydrology Subcommittee of the Advisory Committee on Water Information (ACWI). The Terms of Reference of this work group were approved by the Hydrology Subcommittee on October 12, 1999 and are available on the ACWI web page. The work group was formed to provide guidance on issues related to hydrologic frequency analysis and replaced the Bulletin 17B Work Group that had existed since 1989. The Hydrologic Frequency Analysis Work Group is open to individuals from public and private organizations. The current members of the work group are also given on the ACWI web page. The initial objectives of the work group are to

- Develop a set of frequently asked questions and answers on the use of Bulletin 17B guidelines,
- Prepare a position paper that provides guidance on determining the most appropriate methodology for flood frequency analysis for ungaged watersheds, and
- Prepare a position paper on methodologies for flood frequency analysis for gaged streams whose upstream flows are regulated by detention structures.

In response to the first objective above, the work group has prepared a list of frequently asked questions and answers that provide additional information relative to the implementation of the Bulletin 17B "Guidelines For Determining Flood Flow Frequency", dated March 1982 and developed by the Hydrology Subcommittee of the Advisory Committee on Water Data. These questions and answers supplement the guidelines given in Bulletin 17B and it is envisioned that these questions and answers will be modified or extended in the future as better information becomes available.

Any comments on these frequently asked questions and answers or any new questions and/or answers should be provided by email to Bill Kirby, a member of the Hydrologic Frequency Analysis Work Group, at wkirby@usgs.gov for review by the Work Group.

BULLETIN 17-B
GUIDELINES FOR DETERMINING FLOOD FLOW FREQUENCY
FREQUENTLY ASKED QUESTIONS

CONTENTS

==== 100-YEAR FLOOD =====
 ===== RECURRENCE INTERVAL =====
 ===== AVAILABILITY OF BULLETIN 17-B =====
 ===== LAKE STAGE FLOOD FREQ =====
 ===== SEASONAL FLOOD FREQ =====
 ===== SKEW COMPUTATIONS =====
 ===== MIXED POPULATION =====
 ===== FREQUENCY OF MINOR FLOODS =====
 ===== LOW OUTLIERS =====
 ===== HISTORICAL FLOODS AND HIGH OUTLIERS =====
 ===== OUTLIERS, GENERAL =====
 ===== LIMITS OF FREQUENCY CURVE EXTRAPOLATION =====

=====

==== 100-YEAR FLOOD =====

Q: What is the 100-year flood? Twice in the past 10 years, government officials have said that our river has had a 100-year flood? How can this be?

A: The 100-year flood is the stream flow rate, in cubic feet per second (cfs) or cubic meters per second (m³/s), or the water surface elevation, in feet or meters, that is exceeded by a flood peak in one year out of 100, on the long-run average, or, equivalently, exceeded with a probability of 1/100 (1 percent) in any one year. The 100-year terminology does not imply regular occurrence or that a given 100-year period will contain one and only one event. The 100-year flood also is called the 1-percent-chance flood, and this terminology calls attention to the fact that each year there is a chance that the 100-year flood will be exceeded. Flood occurrence is a random process, largely unpredictable over time spans longer than a few days or weeks. Thus, a rash of exceedances of the 100-year flood can occur in a short time by pure random chance and bad luck. In addition, the true 100-year flood is never known with certainty, but must be estimated from a small sample using uncertain assumptions about the flood-generating processes, with the result that the estimated 100-year flood may be lower (or higher) than the true value. Finally, some caution must be exercised in calculating and interpreting probabilities of events that have

already occurred. Consideration of pure random chance indicates that two exceedances of the 100-year flood could occur in 10 years for the same reason (and with about twice the likelihood) that three double-sixes could occur in 10 rolls of a pair of fair dice. The occurrence of either of these events might lead one to wonder whether the dice really were fair or whether the 100-year flood had been under-estimated. However, floods are continually occurring on thousands of streams around the country; although the probability is small that the 100-year flood would be exceeded twice in a particular 10 year period in any particular location, the fact that it happened during some 10-year period, somewhere in the country, is not surprising, and the fact that it happened on your river, rather than someone else's, should not be taken as an indication of anything wrong.

===== RECURRENCE INTERVAL =====

Q: What is a recurrence interval? My house was damaged by a flood last year, and I'm using my flood-insurance payment to make some improvements as well as repairs. A government report said that the recurrence interval was 100 years, and my friend who's doing the work says that it's safe to make the improvements, because another flood won't occur for 99 more years. Is that right?

A: A recurrence interval (also called a return period) is the expected (or average) length of time between occurrences of events of a specified type, such as floods that exceed a stated stage or discharge. The words "recurrence" and "period" do not imply regular predictable occurrence in time; the actual times between successive events may be either greater than or less than the average. Floods occur approximately as a sequence of independent random trials, with some probability of occurrence in each year. The average time between occurrences is equal to the reciprocal of the annual probability of occurrence: low-probability events are -- on average -- widely spaced in time. Mathematically, the waiting time to the next flood (from either a flood year or a non-flood year) has the same statistical distribution as the time between floods; thus, the recurrence interval is also the expected waiting time to the next occurrence of the event. For example, if the chance of an overflow of the stream banks is 50 percent (1/2) in any year, then the recurrence interval is 2 years. If the chance of overtopping a levee, say, is 1 percent (1/100) in any year, then the recurrence interval is 100 years. Although the average time between overtoppings and the expected waiting time to the next overtopping are 100 years, it nonetheless is possible that the levee is overtopped in two successive years: the probability is $0.01 \times 0.01 = 1/10,000$. However, if the levee has just been overtopped this year, the relevant probability issue is not the chance of two successive overtoppings, but rather the chance that the

levee will be overtopped again next year; that chance is 1/100. Likewise, if the levee has not been overtopped in the past 99 years, the relevant issue is not the law of averages but rather the chance that the levee will be overtopped next year; that chance is still 1/100. In both cases, the probability that there will be no overtoppings in the next N years is just the probability that the flood will be lower than the levee top in every one of the next N years: that probability is $(1-0.01)^{N}$ or 0.99 raised to the N-th power; the probability of one or more overtoppings in the next N years is $1 - .99^{N}$.

Years(N)	1	2	3	5	10	20	30	50	100	200
P{none}	.99	.98	.97	.95	.90	.82	.74	.61	.37	.13
P{>=1}	.01	.02	.03	.05	.10	.18	.26	.39	.63	.87

Thus, your chance of being caught by another flood before your friend finishes his work is pretty slim, but your chance of being flooded sometime in the next 20 years is a little higher than the chance of losing at Russian roulette. The chance of being flooded at some time during the term of a 30-year mortgage is a little higher than that of getting two heads on a row in coin tossing. The importance of duration of exposure in assessment of risk is clear.

Q: The river in my town had two big floods in one year about 15 years ago, one at Easter time and the other at Labor Day. I remember because the second one re-damaged my house within a month after I finished fixing it up after the first flood. So I was shocked when the government engineers doing a flood study in my town didn't show both of those floods, but showed only one flood for each year. How can a valid flood risk analysis be done if you ignore half of the floods?

A: The Bulletin-17-B guidelines recommend a procedure for estimating flood risk based on the maximum flood peak (the so-called "annual flood") in each year. This analysis yields estimates of the probability that the annual flood in any year will be greater than (or less than) any specified flow rate. The probability that the annual flood is less than some discharge x is called the annual non-exceedance probability for that discharge and often is denoted F(x); it is the probability that the year will be free of floods exceeding that level. The complementary probability 1-F(x) is called the annual exceedance probability; it is the probability that the annual flood will exceed level x in any year. If the annual maximum flood exceeds level x in some year, it is possible that one or more other floods, smaller than the annual maximum, might also exceed x. Thus the annual exceedance probability is the probability of one or more exceedances in the year. It is not the case that half the floods are ignored; all of the floods in a year have to be

examined in order to determine the maximum -- the annual flood. Although the Bulletin-17-B annual-flood methodology adequately answers the question it addresses -- differentiating between the likelihood of no flooding versus the likelihood of one or more floods in a year -- this information may not be sufficient to answer all questions that may be important for flood risk evaluation and economic analysis of flood damage. Information about the number of floods in a year or about the joint probability distribution of flood magnitudes and dates within the year may be needed for proper evaluation of flood damage to minor streamside facilities, especially damage due to interruption of use of low-volume road crossings, recreational facilities, and some industrial and commercial facilities. These questions are outside the scope of Bulletin 17-B; they may be addressed by so-called "partial-duration" or "peaks-above-a-base" frequency analysis, which considers all peaks that exceed a specified flood threshold (see Linsley, RK et al., 1982, Hydrology for Engineers, pp 359, 373-374). In more typical flood frequency applications such as flood control levees and 100-year flood-plain delineations, on the other hand, magnitudes of design floods are high, corresponding annual exceedance probabilities are low, probabilities of multiple occurrences in a year even lower, and the damages due to failure are catastrophic and long-lasting. In such cases, even if multiple floods do occur in one year, the damage due to the first is unlikely to be repaired completely before the second one occurs, and the total damage due to all floods in the year may be reasonably approximated by the damage due to the maximum flood acting alone. In these cases, then, the Bulletin-17-B annual-flood approach provides the flood-frequency information needed for engineering-economic planning and flood-risk evaluation.

===== AVAILABILITY OF BULLETIN 17-B =====

Question: Where can I obtain a copy of the publication, Guidelines for Determining Flood Flow Frequency, Bulletin 17B?

Answer: There are two sources. A copy of the original is available from the National Technical Information Service (NTIS), Springfield VA 22161, as report no. PB 86 157 278. As of August 2002, the price for a paper copy from NTIS was about \$60, including shipping, and could be ordered by calling (800) 553-6847.

A digital copy of Bulletin 17B is available in PDF format from FEMA's mitigation web site.

The URL is -- http://www.fema.gov/fhm/dl_flow.shtml. On that page,

click on the link to send an email to pdfarchive@floodmaps.net. A link to the PDF document will be delivered by a reply e-mail. The reply e-mail may take more than a day to arrive. The document size is approximately 28 Mbytes.

The Bulletin 17B PDF document also is available from the USGS web page at http://water.usgs.gov/osw/bulletin17b/bulletin_17B.html This page also has links to many of the Bulletin 17B references and to these Frequently Asked Questions.

Suggested format for citation of Bulletin 17-B:
U.S. Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency, Bulletin 17-B of the Hydrology Subcommittee: Reston, Virginia, U.S. Geological Survey, Office of Water Data Coordination, [183 p.]. [Available from National Technical Information Service, Springfield VA 22161 as report no. PB 86 157 278 or from FEMA on the World-Wide Web at http://www.fema.gov/mit/tsd/dl_flow.htm]

===== LAKE STAGE FLOOD FREQ =====

Question: Often lakes are major flood sources for communities. It is necessary to determine the 1% annual exceedance level (stage, elevation) for the lakes. Some lakes have more than 40 years of record for annual maximum elevation. Can the Bulletin 17B procedure be used to estimate stage frequency curves for these lakes?

Answer: Bulletin 17-B was written for riverine flood-flow frequency analysis, not for lake-level frequency analysis. Although the basic Bulletin-17-B methodology (method of moments, log-transformed data, Pearson Type III distribution) is not limited to flood flows, it has not been systematically tested and evaluated for application to lake levels. The Bulletin-17-B generalized skew map does not apply to lake levels, and it may be difficult or infeasible to develop generalized skews for lakes. In addition, lake levels do not have the natural zero value and the extreme variability and skewness of flood flows, so the use of log transforms of lake levels may not be necessary or beneficial. Thus, Bulletin 17-B should not be applied blindly or dogmatically to lake levels. Instead, graphical frequency analysis should be used. Consider the Bulletin-17-B guidance (page 2) on documentation of flood-frequency studies and on length of record (at least 10 years) needed for statistical analysis. Compute probability plotting positions using the Bulletin-17-B formula (Weibull formula, equation 11, page 27), and plot them versus untransformed lake levels on an arithmetic-normal probability grid. Use the shape of the plotted curve to decide what distribution to fit, or use a visually fitted manually-drawn curve. No distribution has been established for lake level frequency analysis, as the log-Pearson

III distribution for stream flow peaks. Extrapolate the lake-level frequency curve only with extreme caution: the form of the lake-level frequency curve is not known and lake levels may be much more sensitive to lake-shore topography than the peak flows are to flood plain topography. Do not extrapolate the level-frequency relationship if lake levels are under impacts of operation of structures or if other constraints exist. An additional complication is the effect of storage changes on the time-series characteristics of the lake level process. If the lake has an effective surface outlet, flood waters can drain out rapidly and the lake level returns quickly to its normal level. In this case, maximum lake levels in successive years will be substantially independent, so that traditional methods of estimation can be applied directly to the lake levels. If the lag-one serial correlation is less than about 0.4, then the impacts of assuming independence is not great. However, if the lake has no surface-water outlet (closed lake such as the Great Salt Lake) or if its outlet has much smaller capacity than the surface inflows to the lake, there is likely to be substantial serial dependence between lake levels in successive years. In this case a short record of the lake levels themselves is unlikely to show the true long-run distribution; instead, the lake-level CHANGES from year to year should be treated as independent random variables and used to define a distribution of annual lake-level changes. A hydrologic time-series model should be developed to represent the lake level in any year as the cumulative sum of all past level changes. The lake-level distribution then should be computed from mathematical analysis and/or simulation of the hydrologic time series model. Such a model recently was developed for Devils Lake, North Dakota (USGS Open-File Report 95-123, by G.J.Wiche and A.V.Vecchia). If the serial correlation exceeds about 0.4, then the time series modeling approach should be used.

==== SEASONAL FLOOD FREQ =====

Question: Can the Bulletin 17-B procedures be used to develop x-year flood peaks and volumes for a month or season or other part of a year? Such information may be needed for defining flood diversion requirements in a river during construction which is planned during the low flow season.

Answer: Bulletin 17-B was written for analysis of annual-maximum peak flows; flood volumes and partial-year maxima were not considered and the Bulletin-17-B procedures were not tested or evaluated for application to volumes or partial-year data. Nonetheless, the Bulletin-17-B procedure will provide a mathematical solution to a mathematical problem of fitting a distribution to data. The data can be limited by date. The largest flood peak or volume for the set range of dates can be calculated for each year (at least if a partial-duration flood

series or continuous daily flow record is available). The 17-B procedure can be applied to the derived data set. The fitted frequency curve should be scrutinized critically, since the adequacy of Bulletin-17-B procedures for partial-year data has not been established. Bulletin-17-B map skews cannot be used, only the computed station skew, unless special generalized skew studies are performed for the partial-year or volumetric data. The procedure can be applied using multiple gauge sites if some form of regional analysis is desired for analysis of an ungaged site. The result can be useful in defining the diversion flood requirement for that particular season.

However the result can be very misleading unless clearly identified with the season or months considered. The 100-year dry-season flood actually is exceeded -- during the dry season -- in one year out of 100, on average, and there is, on average, 100-year spacing between years containing dry seasons in which the 100-year dry-season flood is exceeded. Obviously, however, the 100-year dry-season flood may be much less than the 100-year flood based on the maximum flood in the full year, and years containing any exceedance (not just dry-season exceedances) of the 100-year dry-season flood may occur much more frequently than once in 100 years. Even if floods occur more or less uniformly throughout the year, restricting attention to a particular time period, say the month of March, will yield a March 100-year flood that is lower than the ordinary 100-year flood, for the same reason that the maximum of 12 throws of the dice is usually greater than the result of, say, the third throw.

One problem which often comes up, especially with peak flows, is that limiting the data by date may severely limit the amount of data available for analysis. If peaks above a base are available, then, even if there are no events in some years, the base discharge can be used to define a conditional-probability adjustment for those years. If only annual peaks are available, however, then it may be necessary to ignore the years in which the annual peak does not occur in the specified time period; the probability interpretation of the resulting frequency curve is very tricky -- the curve will give the probability that an annual peak occurring in March, say, will exceed discharge x , but leaves unanswered the real question, which is whether any peak in March will exceed x . Some other analysis to relate peak flow to daily flow or the use of some other gauge records may be advised. Most often the procedure is used with daily-mean flows which are much more abundant and more readily divided into seasons without fear of extreme data losses.

==== SKEW COMPUTATIONS ====

Question: What is the applicability of the skew map (Plate I) in Bulletin 17B? What is the maximum watershed drainage area for which it can be used?

Answer: The skew map in Bulletin 17B is provided as a convenience for users who do not wish to do the regional analysis needed to develop their own estimates of generalized skew. The use of the skew map should be consistent with the data used to develop it. The map was developed with data from watersheds smaller than 3,000 square miles and with essentially unregulated peak discharges (that differ from natural peak discharges by less than 15 percent). Periods when the annual peak discharge likely differed from natural flow by more than about 15 percent were excluded from the skew-map analysis. Thus, the maps should not be used with regulated flows. The results were presented as a map largely because it was not possible to find convincing relations between skew and basin characteristics or other peak flow statistics. Although basin storage is generally thought to be important, there is no widely accepted empirical or theoretical relation tying basin characteristics to logarithmic skew coefficients. Although the use of the skew map should be consistent with the data used to develop it, no strict upper limit on drainage area has been established for use of the map, and no procedures have been defined for using the map to determine generalized skew for basins larger than 3000 sq mi. Ideally, generalized skews for basins larger than 3000 sq mi should be determined by analysis of skews for nearby basins of similar size, especially when the map skew varies substantially across the basin. This is not always possible, however, and undoubtedly the Bulletin-17 skew map has been used for basins larger than 3000 sq mi. Although modest extrapolation may be acceptable, the Bulletin-17 map should not be extrapolated to drainage areas greater than about 2 or 3 times the size of the basins for which the map was developed, because the physiographic factors controlling the skew coefficients of large basins undoubtedly are different, at least in magnitude, from those effective in smaller basins.

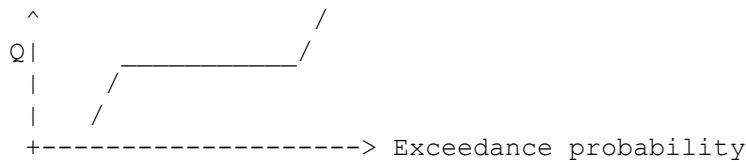
Question: How should the generalized skew be determined from Plate I in Bulletin 17B? That is, should the value from the centroid of the basin or from the station location be selected from the map in Bulletin 17B?

Answer: Using the Bulletin-17 skew map (Plate I), the generalized skew should be determined at the point of interest or outlet of the watershed rather than at the centroid of the watershed, to be consistent with how the map was prepared. The Bulletin-17 skew map was developed by plotting station skew values at the latitude and longitude of the gaging station and plotting iso-lines. No information about basin centroid locations was practically available. If some other skew map or relation is used, the gage or basin location should be specified in the same manner used in developing the map or relation.

 Question: Should Bulletin 17B be used for regulated watersheds (peak discharges differ by more than 15 percent from natural peaks), and if so, what skew value should be used?

Answer: Bulletin 17B can be used for regulated watersheds if the logarithms of the regulated peak discharges are reasonably consistent with a Pearson Type III distribution. A graphical comparison of the plotting positions to the computed frequency curve should be used to judge the reasonableness of using Bulletin 17B.

However, if the basin is regulated by a reservoir that is generally quite effective at reducing damaging flood peaks, then there may be a problem in assuming log-Pearson Type III shape. If the reservoir is quite effective, the upper middle range of flood magnitudes will be lowered relative to the unregulated condition, but the extreme upper tail, corresponding to overtopping of the reservoir, will be back up at the same magnitude as the unregulated flows. The resulting frequency curve will have the following general shape:



This is not an L-P-III shape. Moreover, the actual observed data are likely to show only the lower and middle (flat) segments, not the steep upper segment, so the fitted curve also will reflect only the lower flatter segments. Extrapolation of the fitted curve and observed data therefore are likely to grossly underestimate the potential magnitude of extreme rare floods.

To get an adequate representation of the upper steep segment of the frequency curve, one should try to increase the record length by routing historical floods through current reservoir storage or by developing and routing hypothetical floods (for a given frequency).

If Bulletin 17B is judged to be applicable, then station skew should be used in defining the final frequency curve. The skew map in Bulletin 17B should not be used in determining a weighted skew because the Bulletin 17B skew map is based on data for essentially unregulated watersheds. As always, the final determination of the skew coefficient and other frequency curve parameters should include consideration of the data, the flood-producing factors, the characteristics of the regulation rules, the characteristics of nearby similar regulated sites (if any), and the sensitivity of the results.

It is improper to use Bulletin 17B procedures on a non-homogeneous data set , i.e., one in which peak flows for certain years are regulated, and for other years not. The peak flows for the regulated period should be analyzed by themselves, or converted into unregulated flows before a frequency analysis is performed.

Question: What is the basis for giving greater weight to station skew when the generalized and station skew differ by more than 0.5 (page 15 Bulletin 17B)?

Answer: Bulletin 17B work group members felt that large deviations between generalized and station skew may indicate that the flood frequency characteristics of the watershed of interest are different from those used to develop the skew map. It is thought that station skew is determined by rainfall skew and basin storage and that there is considerable variability of response among different basins with similar observable characteristics, in addition to the random sampling variability in estimating skew from a short record. Therefore, it is considered more reasonable to give greater weight to station skew, after due consideration of the data and flood-producing characteristics of the basin. The difference of 0.5 was chosen based on engineering judgement and considering that the standard error of the Bulletin 17B skew map is 0.55.

===== MIXED POPULATION =====

Question: Floods in my study area are caused by hurricanes, by ice-affected flows, and by snowmelt, as well as by rainfall from thunderstorms and frontal storms. How do I determine whether mixed-population analysis is necessary or desirable?

Answer: Flood magnitudes are determined by many factors, in unpredictable combinations. It is conceptually useful to think of the various factors as "populations" and to think of each year's flood as being the result of random selection of a "population", followed by random drawing of a particular flood magnitude from the selected population. The resulting distribution of flood magnitudes is called a mixture distribution. However, the resulting mixture distribution in many cases is well-approximated by a log-Pearson Type III distribution (LPIII), and in such cases there is no benefit in going through the lengthy mixture calculation instead of using the LPIII tables to compute the distribution. In practice, one determines whether the distribution is well-approximated by the LPIII by comparing the fitted LPIII with the sample frequency curve defined by plotting observed flood magnitudes versus their empirical

probability plotting positions determined by eqn 10 (pg 26). If the fit is good, and if the flood record includes an adequate sampling of all relevant sources of flooding (all "populations"), then there is nothing to be gained by mixed-population analysis. Only if the sample frequency curve has sharp curvature (kinks), reverse curves, or other characteristics that prevent its being approximated by the LPIII, or if the available flood record omits important sources of flooding, is there any reason to perform a mixed-population analysis.

Question: Hurricanes are an important source of flooding in my study area, but the flood record at my study site is short and contains no hurricane events (or too few events to define the hurricane population). How do I perform a mixed-population analysis in this case?

Answer: First examine longer records from nearby sites to see if the hurricane population is inconsistent with an overall LPIII distribution. If mixed population analysis is indicated, then determine the distribution of hurricane-flood peaks for the study site. This might be done by determining historical hurricane-flood peaks at the study site using rainfall-runoff modeling, determination of peak-flow/drainage-area ratios, or other means. Plotting the hurricane-flood peaks versus Weibull probability-plotting positions ($n/(N+1)$) determines the CONDITIONAL distribution of hurricane-flood magnitudes. The observed record at the site defines the CONDITIONAL distribution of non-hurricane flood magnitudes. (Any hurricane events in the observed record first have to be moved to the hurricane record.) The two conditional distributions can be combined as follows to define the unconditional distribution of all floods at the site --

$$F(x) = P\{\text{annual flood } X < x\} =$$

$$= P\{X < x \mid X \text{ is hurricane flood}\} * P\{X \text{ is hurricane flood}\} +$$

$$P\{X < x \mid X \text{ is non-hurricane flood}\} * P\{X \text{ is non-hurricane flood}\}$$

wherein $P\{X \text{ is hurricane flood}\} = NH / LHP$
 $=$ number of hurricanes / length of historical
period analyzed for hurricanes

and $P\{X \text{ is non-hurricane flood}\} = 1 - NH/LHP$.

Note that ALL hurricane-caused annual floods during the period LHP must be included in the analysis to properly define the conditional distribution of hurricane floods.

==== FREQUENCY OF MINOR FLOODS ====

Question: I have to determine the 1.1-year flood for use in stream restoration analysis. The record contains several low outliers, and the computed frequency curve is not defined for the high exceedance probabilities that correspond to the 1.1-year flood. How do I proceed?

Answer: Bulletin 17 methodology is not designed for and should not be used to determine high-frequency low-recurrence-interval flood magnitudes or to determine risks due to occurrence of low-magnitude floods. This is the case whether or not there are low outliers, even if the computation does yield a value for the 1.1-year flood. Bulletin 17 is based on the annual-flood probability model, in which it is assumed that exactly one flood event occurs per year. This probability model usually is adequate as an approximation for risks due to large-magnitude low-frequency floods (which are unlikely to occur at all during any given year, and extremely unlikely to occur more than once during the year). The annual-flood model, however, does not adequately represent the occurrence of low-magnitude floods. In most streams, several low-magnitude flood events occur in most years. In most cases, if floods of this magnitude cause damages or other effects of concern, such as channel-forming activity, then each occurrence of such a flood will contribute to the cumulative effect for the year. Proper accounting of the risk requires consideration not only of the distribution of individual flood magnitudes but also of the distribution of the number of events that occur in a year. The annual-flood analysis does not furnish the necessary information about the likelihood of multiple flood occurrences per year. Generally speaking, the annual-flood model understates the total risk or total effect for the year because of undercounting of the number of minor floods in the year. Conversely, it overstates the recurrence interval of minor floods of a given magnitude, again because of failure to recognize the occurrence of multiple events per year. A different method of frequency analysis, called the "partial duration" method (because the "duration" of time associated with each flood event is only a "partial" year), or "peaks over threshold (POT)" method, which explicitly considers multiple events per year, is required. This methodology is described, for example, in Hydrology for Engineers (1982, by Linsley, Kohler, and Paulhus, McGraw-Hill, pages 359, 373-347) and consists of selecting all distinct well-separated flood peaks exceeding a given threshold magnitude, ranking them, estimating the recurrence intervals by the formula $T = (N+1)/m$ (where N is the record length, in years, and m is the rank of the peak), and plotting magnitude versus recurrence interval. The threshold is commonly set so that a long-run average of about 3 peaks per year will be recorded; thus recurrence intervals as low as about 1/3 year can be defined. Relating the recurrence-interval curve to probabilities of occurrence requires consideration of the frequency distribution of the number of above-threshold peaks per year, as summarized, for example, in the Handbook of Hydrology (1993, D.R. Maidment, editor, McGraw-Hill, page 18.37). (That having been said, if administrative

or regulatory requirements necessitate use of the 1.1-year annual flood, and if that value is not computed because of low outliers or zero flows, a value can be determined by graphical plotting of the low end of the frequency curve or by manual calculation using the "synthetic" statistics described in appendix 5 and printed by the computer as "Bulletin-17-B estimates". If more than 9 percent of the annual peaks equal zero, then the 1.1-year flood equals zero; if more than 1/3 of the peaks equal zero, then the 1.5-year flood equals zero.)

===== DATA QUALITY =====

Question: How important is data quality in the validity of Bulletin-17-B frequency results? What issues need to be checked?

Answer:

Data quality is obviously important to validity of the Bulletin-17-B frequency analysis, since the frequency analysis is basically nothing other than a standardized summary of the underlying flood data set. In a critical review of the flood data set, two broad sets of issues need to be considered: 1) relevance of the flood data set (and frequency analysis results) to estimation of future flood risk, and 2) accuracy of the data set as a representation of the flood events that actually occurred in the past.

In the first set of issues, factors such as flow regulation by dams, dam failures, stormwater management, effects of development (or reversion to undeveloped conditions) in the flood plain, stream channel improvement or restoration, and the effects of mining, forestry, agriculture, or reclamation from those activities, all have the potential to make all or part of the record unrepresentative of future flood risk. The significance of these factors and the nature of any adjustments that might be applied for estimation of future flood risk cannot be predicted in general and depends on the specific situation at each site; no simple guidelines can be given that could safely be followed blindly or dogmatically. The application of records of past floods (including frequency analysis results) to decision-making about the future is outside the scope of frequency analysis and belongs to the realm of engineering-economic decision making.

Regarding the accuracy of the data, it is helpful to consider the process for computing the flood record, the potential sources of error, and the steps taken to detect and correct errors. Most annual-peak flows are determined by sensing the stage or water level at the gage and reading the flow (discharge) from a stage-discharge rating curve. The rating curve is made by correlating direct measurements of discharge, made by current meters or similar devices, with concurrent measurements of stage. The accuracy of the annual peak flow value then depends on the accuracy of the stage reading and the accuracy of the stage-discharge relation. The accuracy of the stage-discharge relation, in turn, depends on the accuracy, number,

and flow magnitudes of the direct discharge measurements used to establish the relation. An important factor in promoting accuracy of records is a long-term organizational commitment and focus on production of records, along with a regularized process for checking and reviewing the data collection and computations, cross-checking the results against records at nearby streams, and annual publication of the records for public examination and use.

Issues of data accuracy are most likely to affect the top-magnitude floods in the data set. These events occur rarely, so there are fewer opportunities to define the stage-discharge rating for events in this range. In addition, these are the most destructive events, and are more likely to destroy or damage the gage, or impair the operation of the instrumentation. The uncertainty associated with historical peak discharges is usually greater than that associated with peaks that are part of the systematic record. The analyst should evaluate if the historical peak discharges are reliable enough to be used in the analysis. Issues that may be of concern include whether the historical sources provide sufficient substantive information to associate a stage or discharge with the historical event, whether the historical stage is referenced to the same gage datum as the stages used to develop the stage-discharge rating used to compute the discharge, and whether the stage-discharge rating adequately reflects the hydraulic conditions that existed in the channel and flood plain at the time of the historical event.

For historical data, attention must be paid to what is not in the data set as well as to the accuracy of the recorded historical peaks. As explained in more detail below, the Bulletin-17-B procedure for historical data involves defining a historical threshold discharge that separates the record into two classes of peaks which are given different weights in the computation. Bulletin 17-B requires that the threshold be set at a level high enough to ensure that it was not exceeded by any peaks that are not in the record. Any non-systematic peaks that are below the threshold are unusable statistically. Although the precise numerical value of the threshold is of little consequence, since it is not used for computation, setting the threshold to correctly identify the number and magnitudes of the peaks to be adjusted is critical to the accuracy of the historical adjustment. There is a tendency to casually assume that any peak that is outside of a period of systematic record is a true historical peak in the sense of Bulletin 17-B, and a tendency to improperly set the threshold at the level of the lowest such peak. Occasionally, records contain non-systematic peaks that are lower than many of the systematic peaks and contain few or no higher non-systematic peaks. In these cases, it is likely that higher peaks actually did occur outside the systematic record period but were not included in the

non-systematic record, thus violating the assumptions underlying the historical adjustment. Setting the threshold too low results in improper discounting of the high-magnitude peaks relative to the below-threshold peaks. The analyst should check that the number of peaks exceeding the threshold during the systematic record period is consistent with the number during the historical period, and should check that the threshold level is not so low that it could have been exceeded without anyone's taking note of it. Accuracy of the length of the historical period also is important because the value is used to compute the amount by which the above-threshold peaks are discounted; knowledge of local history is critical.

Question-- What is the relationship of the Federal Data Quality Act to flood data and flood-frequency analysis?

Answer -- The "Federal Data Quality Act" (officially known as Section 515 of Public Law 106-554, the Treasury and General Government Appropriations Act for Fiscal Year 2001) requires the Office of Management and Budget (OMB) and, through it, all Federal agencies to issue guidelines to ensure the "quality, objectivity, utility, and integrity" of information issued by the government. The agencies are required to develop procedures for reviewing and substantiating the quality (including objectivity, utility, and integrity) of information before it is released. The agencies also are required to establish administrative procedures by which persons affected by government-disseminated information can seek and obtain correction of information that does not conform to the quality guidelines. The general intent of the guidelines is that agencies should make their data-collection, data-analysis, and data-interpretation methods "transparent" by providing documentation of the methods; should ensure data and information quality by reviewing the methods used (including, as appropriate, consultation with experts and users); and should keep users informed about corrections and revisions. These guidelines and procedures apply to government-disseminated information in general, and thus apply to flood data and to the results of statistical flood frequency analysis. The guidelines and procedures apply not only to information produced internally by agencies themselves, but also to information supplied by outside sources.

===== **LOW OUTLIERS** =====

Q: What is a low outlier? How is it different from a zero flow? Why do we drop low outliers and zero flows? Aren't we overstating flood risk if we ignore flood peaks that are zero or near zero?

A: Outliers are observations that lie far out from the trend of

the rest of the data when plotted on a magnitude versus frequency graph. A smooth trend line, such as a statistical frequency function, does not fit data sets with outliers, and the fitted curve usually fails to fit the bulk of the data as well as the outlier. Low outliers are outliers at the low end of the data set, near zero, at least in comparison with the rest of the data. On a log-probability plot, the low outliers impart a strong downward curvature and a downward-drooping lower tail to the frequency curve. In comparison with the lower tail, the upper tail of the low-outlier-affected curve may appear relatively flat.

In the Bulletin-17-B context, low outliers differ from zero values in that computations with the logarithm of zero are impossible, whereas computations with logarithms of low outliers may be mathematically possible, but may overwhelm the computations of logarithmic moments (means, standard deviations, and skews) or distort the fit of the frequency curve to the data in the upper part of the data set, which are the data that represent significant flood or near-flood events. Since the zero and near-zero values in a flood data set are not the ones that convey valid or meaningful information about the magnitude of flooding, their numerical values are not used in the computation of the moments. However, in contrast to classical statistical treatments, where outliers are considered utterly spurious and are simply dropped from the data set, the Bulletin-17-B procedure recognizes that zero values and low outliers do convey valid and meaningful information about the frequency of flooding, and this information is used in Bulletin 17-B. Thus, the Bulletin 17-B procedure first uses the non-zero non-low-outlier data to define a conditional-probability curve which applies only to the non-zero non-low-outlier events; then the number of zeroes and low outliers is determined and used in the conditional-probability or "n-over-N" adjustment (equation 5-2) to adjust the probabilities from the conditional curve to properly reflect the frequency of occurrence of zeroes and low outliers.

Question: When should low flows that are not identified as low outliers using the 17B default procedure be censored as a result of the paragraph in Bulletin 17B on page 18 that reads, "If multiple values that have not been identified as outliers ...".

Answer: Bulletin-17-B detects low outliers by means of a statistical criterion (the Grubbs-Beck test) rather than by consideration of the influence of low-lying data points on the fit of the frequency curve. The test is based on the standardized distances, $(x_i - \bar{x})/\text{stdv}$, between the lowest observations and the mean of the data set. The test is easily defeated by occurrence of multiple low outliers, which exert a large distorting influence on the fitted frequency curve, but also increase the standard deviation, stdv , thereby making the standardized distance too small to trigger the Grubbs-Beck test. Therefore,

Bulletin 17-B (pg. 18) permits manually overriding the statistical criterion. Obviously, the intention is to allow as many low outliers to be designated as necessary to achieve a good fit to the part of the data set that contains the significant flood and near-flood events. Equally obviously, the intention is that the Grubbs-Beck result be used unless the resulting poor fit gives compelling justification for not doing so.

There is no universal method that can be followed blindly to achieve a good fit. The sensitivity analysis alluded to in Bulletin 17-B is based on the engineering-hydrologic-common-sense proposition that the smallest observations in the data set do not convey meaningful or valid information about the magnitude of significant flooding, although they do convey valid information about the frequency of significant flooding. Therefore, if the upper tail of the frequency curve is sensitive to the numerical values of the smallest observations, then that sensitivity is a spurious artifact based on the mathematical form of the assumed but in fact unknown flood distribution, and has no hydrologic validity. The sensitivity analysis determines whether the upper tail of the frequency curve is sensitive to the magnitude of the lowest values by iteratively treating them, one by one, as low outliers and plotting the estimated value of, say, the 100-year flood (or other percentage point or points characteristic of the upper part of the frequency curve) as a function of the number of low outliers. Frequently, the estimated 100-year flood will change noticeably and consistently, either increasing or decreasing, as the first few low outliers are identified, but then remain relatively constant, perhaps changing erratically, as additional data points are treated as low outliers. In such cases the identity of the spuriously influential data points -- the low outliers -- is clear, and the low-outlier threshold is set just above the magnitude of the highest spuriously influential data point. Note that the magnitude of the change resulting from low outlier treatment is not the deciding factor, but rather the change in the magnitude of change as additional points are treated as low outliers. In more complex cases, there may not be a clear demarcation of the low outliers, and the entire low end of the data set may be inconsistent with the fitting of the log-Pearson Type III distribution to the upper (hydrologically significant) part of the data. In such cases it may be necessary to rely on visual assessment of the fit of the upper part of the frequency curve, and Bulletin 17-B allows for this necessity.

Question: Does dropping multiple low outliers improve the estimate of the 100-year flood at the expense of distorting the estimates of the lower-recurrence-interval (10, 20, 50-year) floods?

Answer: No. The intent and the result of the low outlier adjustments are to improve the fit of the entire frequency curve

above the low-outlier threshold.

==== HISTORICAL FLOODS AND HIGH OUTLIERS =====

Question: What is the difference between a high outlier and a historical flood?

Answer:

High outliers and most historical floods both are exceptionally large floods. High outliers are exceptionally large floods that are contained in the systematic record, whereas historical floods were observed outside the period of systematic record. Systematic records are collected during periods of systematic stream gaging, usually continuous series of years, in which flood data are observed and recorded annually, regardless of the magnitudes of the floods. A nonsystematic record is collected and recorded sporadically, without definite criteria, usually in response to actual, perceived, or anticipated major flooding. The systematic record can be used directly in flood frequency analysis. The non-systematic record cannot be used unless additional information can be supplied to relate it to the population of all flood peaks. Bulletin 17-B requires that the non-systematic record be a complete record of all flood peaks that exceeded some threshold level during a definite historical time period.

A high outlier is an extraordinary flood that occurred during the period of systematic streamgaging. It is part of the systematic record and is treated just like the other systematic peaks in the preliminary steps of the Bulletin-17-B analysis. On a magnitude-vs-probability plot, the high outlier lies well above the fitted frequency curve, and the fitted frequency curve is steeper than the trend of the other plotted data points. Often historical information is available that indicates that the high outlier was the largest in a period longer than the period of systematic streamgaging. This historical information is used to adjust the frequency curve to take proper account of the extended time period associated with the high outlier. If no usable historical information is available, the high outlier is retained in the systematic record and used without adjustment.

A historical flood is a major flood that occurred outside of the period of systematic streamgaging. The stage or elevation of the historical flood is usually determined by high-water marks left by the flood and recorded by local residents, state departments of transportation (DOTs), railroad companies, local, state or Federal agencies. Stages of historical floods often are reported in local newspapers, diaries or Bibles of local residents, unpublished documents of state DOTs or railroad companies, and/or published

reports of local, state or Federal agencies. Because the historical event was not observed in accordance with definite statistical sampling criteria, and is not part of the systematic record, its relation to the underlying process of flood occurrence is uncertain. This is so regardless of the accuracy with which the stage and discharge might have been determined. For example, a historical flood that washed out a bridge might have been recorded although a larger flood that caused no damage might have gone unremarked. The historical flood cannot be used in flood frequency analysis unless additional information (historical threshold and historical period) is available to relate it to flood occurrence over a historical time period.

The computational procedures in Bulletin 17-B Appendix 6 are applied to both historical floods and high outliers.

Q: Why do we bother with historical floods and high outliers? Why don't we just use the systematic gage record?

A: We bother with historical floods and high outliers because systematic streamflow records usually are short and may be inconsistent with the longer-term flood history experienced and recorded non-systematically by the local community. Sometimes a short systematic record contains an extraordinary flood peak that stands head and shoulders above everything else in the systematic record and everything else experienced in the history of the local community. Conversely, the long-term community history may record one or more outstanding floods that are much larger than anything in the systematic record. In either case, the systematic record disagrees with long-term community experience. Such discrepancies must be resolved if the frequency analysis is to be a sound basis for planning. The Bulletin-17-B historical adjustment procedure provides a basis for reconciling these discrepancies.

Question: What is the high outlier threshold?

Answer: There are two different quantities that sometimes are called high outlier thresholds. One is the statistical high-outlier test criterion or threshold computed by the Grubbs-Beck test and described on page 17 of Bulletin 17-B. The other threshold, which B-17 does not clearly describe or distinguish from the statistical threshold, is the threshold that is used in (or implied by) the Bulletin-17-B historical adjustment procedure that actually is applied to the flood record. This threshold may be called the historical-adjustment threshold; it does not necessarily equal the statistical high outlier test criterion, and may be either higher or lower than the statistical test.

The statistical high-outlier test criterion is based

on the standardized distance, $(x_i - \bar{x})/stdv$, based on logarithms of peak flows, between the extreme top observation and the mean of the data set. The criterion or threshold value is the value that is unlikely (10 percent chance) to be exceeded by the LARGEST observation in a sample. If the largest observation actually is greater than the high-outlier threshold, that is an indication that the observations above the threshold are larger than would be expected for the given period of record, that they may be associated with a longer time period than the period of systematic record and that they may be distorting the fit of the frequency curve. The test itself is only a warning, not a definitive indication of anything wrong, and additional historical information must be supplied before any adjustment can be made. If appropriate historical information is available, the historical adjustment can be made even if the statistical test does not detect a high outlier.

The historical-adjustment threshold is specified either explicitly or implicitly in the course of defining the non-systematic historical record period and applying the Bulletin-17-B historical data adjustment. Bulletin 17-B does not describe or discuss this threshold clearly, but instead simply assumes that the Z highest peak discharges in the combined systematic and historical (non-systematic) record are known also to be the Z largest in a historical period longer than the systematic record period. If the threshold is stated explicitly, then Z is the number of peaks that exceed the threshold; if Z is given instead, then the threshold is implied to be somewhere between the Z and Z+1 ranked peaks. Although the number Z and the historic period length H are sufficient for computing the historical-data adjustment, an actual discharge threshold (or range of values) is needed to properly document the historical record, and should be reported along with the results of the historical-data adjustment. An actual threshold value or range is needed for two reasons. First, the number of events Z will become outdated, and will have to be updated, whenever a new peak occurs that exceeds the previous Z-th ranked peak. Second, and more important, one cannot legitimately claim, without any support, that the Z largest peaks in the record at hand are the largest in a longer period; it is necessary to demonstrate, or at least provide plausible support, that there actually were no other peaks that occurred but were not included in the record. In most cases, this support has to be based on the idea that if any peaks greater than some magnitude had occurred, people would have noticed and recorded them; that magnitude, the magnitude that would get almost everyone's attention and ensure that a record would be made, is the magnitude that should be determined and documented as the historical-adjustment threshold to support the historical-data adjustment.

It should be noted that the numerical value of the historical-adjustment threshold is not used in the computation of the historical data adjustment. The threshold value is used only

to separate the Z largest values from the remaining data points; the results of the adjustment are the same for threshold values anywhere between the Z and Z+1 ranked data points. This is important because the threshold cannot always be determined with much certainty or precision from the available historical information.

If there is not sufficient historical information available to determine a historical-adjustment threshold and length of historical period, then any historical (non-systematic) peaks are not usable for statistical analysis, because their relation to the underlying process of flood occurrence is unknown. Similarly, if the historical information is inadequate to adjust for the high outliers, then they should be retained as part of the systematic record and all peak discharges given equal weight in computing the moments (mean, standard deviation and skew).

 Question: How is the threshold determined in the historical adjustment procedure?

Answer: The historical-adjustment threshold discharge is chosen high enough such that all high outliers and historical floods included in the adjustment procedure are the only floods known to exceed the threshold in the historical period of H years. In other words, the record is known to be complete for all events exceeding the historical-adjustment threshold. There is no single procedure that can be followed blindly to determine the threshold from the historical information usually available. Determination of the threshold usually will be based on consideration of channel-bank and floodplain elevations, elevations of important structures, and the history of the neighboring community. Although the determination may involve elements of subjectivity and judgement, the choice of the historic threshold should be defensible given the available historical information. The historical-adjustment threshold often is less than the computed high-outlier threshold.

 Question: Must a peak discharge exceed the high outlier threshold to be included in the historical adjustment procedure?

Answer: No, the high outlier threshold (the statistical test criterion given by equation 7) is just used as guidance in determining whether a peak discharge is so large that it might require use of the Bulletin-17-B historical adjustment procedure. If the peak discharge exceeds the high outlier threshold, then the analyst should determine whether historical information is available that indicates the high outlier is the largest flood in a period longer than the systematic record. If there is useful historical information available, the

high outlier may be adjusted even though it does not exceed the computed threshold. Therefore, the historic threshold used in adjusting for high outliers (and historical floods) can be less than the computed high outlier threshold.

Question: What is the difference in the historical adjustment procedure for high outliers and historical floods?

Answer: There is no difference in the computation; the historical adjustment computation does not distinguish between historical peaks and high outliers. The same threshold is used for adjusting for both types of floods. The number of peaks above the threshold is denoted as Z, and no distinction is made between high outliers and historic peaks in determining the value of Z. Historically adjusted moments are computed using a weight of 1.0 for the Z above-threshold peaks and a weight $W = (H-Z)/(N+L)$ (H = total length of historical and systematic record; N+L = number of below-threshold systematic peaks, including low outliers and zeroes) for all systematic peaks below the threshold.

Question: Please clarify the definition of the variable "Z" in the historical adjustment procedure, appendix 6, and clarify the intended application of the procedure. Bulletin 17 seems to say that historical data should be used if possible, but appendix 6 seems to indicate that the historical peaks need to be the largest in the whole record.

Answer: Yes, Bulletin 17 does say that historical data should be used if possible. And yes, Appendix 6 does indicate that the historical adjustment is applied to the largest peaks in the whole record. The key words are "if possible." A couple of conditions must be met. First, historic peaks and high outliers, by their nature, are expected to be a biased (unrepresentative) sample of the population of all peaks. However, the Bulletin 17-B procedure assumes that the high outliers and historic peaks are an unbiased sample of the population of flood events that exceed the historical-adjustment threshold magnitude. If there is some question of that assumption, then the historical information is wholly or partly unusable. Any historical peaks that do not exceed the threshold cannot be used because their relation to the flood population is undefined. If potential high outliers do not exceed the threshold, they are used in the same way as any other ordinary systematic peaks. Second, the Bulletin-17-B historical adjustment procedure postulates the existence of exceptionally large floods (historic peaks and high outliers) in the data set. When such peaks are present, the systematic streamgaging record,

especially if short, may be inconsistent with the neighboring community's long-term knowledge of flood occurrence, and some reconciliation of the gage record with the community experience is required. The historical-adjustment procedure accomplishes this reconciliation by using the above-threshold (historic and high-outlier) peaks with unit weight and the below-threshold systematic peaks with the historical weight factor $(H-Z)/(N+L)$, in effect filling in the rest of the extended historical period with multiple copies of the below-threshold systematic record. Thus, Z represents the total number of peaks, systematic and historic, that exceed the threshold. It is quite possible and acceptable for Z to consist of, for example, 3 historical peaks that exceed the threshold, plus 3 systematic peaks that not only exceed the threshold but also exceed the 3 historical peaks; if there were several additional historical peaks that did not exceed the threshold, they would simply be ignored because their relation to the flood population would be indeterminate. It is also perfectly acceptable for Z to consist of one or more systematic peaks (high outliers) and no historical (non-systematic) peaks at all. However, if $Z = 0$ (no large peaks at all, only the knowledge that no peaks exceeded some threshold), the Bulletin 17-B historical adjustment has no effect -- the computation can be performed, but the frequency curve is unchanged.

--

Question: If all peaks that exceed the historical threshold are treated the same, why do we have all of this gobbledygook about "systematic peaks above the threshold," "high outliers," etc? Why don't we just call them "historical peaks" and be done with it?

Answer: All peaks that exceed the historical threshold are indeed treated the same IN THE HISTORICAL ADJUSTMENT COMPUTATION. However, in the preliminary analysis of the systematic record, the historical peaks are ignored whereas the high outliers are treated exactly like the other systematic peaks. Improper treatment of high outliers and historic peaks in the systematic-record analysis can adversely affect the final Bulletin-17-B frequency curve, primarily through incorrect skew coefficients and mis-identification of high and low outliers.

The systematic record is that portion of the record in which the annual peak is determined and documented for each year, regardless of the magnitude of the peak; if the peak is too small to measure, it nonetheless is recorded, but with a qualification code indicating that fact. Thus, the systematic record can be regarded as an unbiased random sample of the population of all floods, and the statistics of the sample can be taken as estimates of the corresponding characteristics of the population. The historical record consists of flood peaks that were observed outside of a systematic period of flood record. That is, the historical record is non-systematic. One might assume that such peaks were observed

and recorded because they were unusually large and noticeable (or were expected to be so), but there is no real guarantee of this unless additional historical evidence is available. Since one does not know how the historical sample is related to the flood population, one cannot use it for flood estimation unless additional historical information is provided. Bulletin 17's historical flood threshold and historical period together provide the information needed to make use of the historical peaks. If the threshold and historical period cannot be defined, or if the historical peaks are less than the threshold, then the historical peaks cannot be used.

===== OUTLIERS, GENERAL =====

QUESTION:

Why is there so much emphasis on low and high outliers?

RESPONSE:

Bulletin 17B does not explain this very well. It lumps a number of distinct problems and phenomena under the label "outlier," but does not give much explanation of how the Bulletin-17-B conception of outliers differs from the classical concepts developed in the literature on statistics and analysis of measurement data.

Bulletin 17-B defines outliers as data points that depart significantly from the trend of the remaining data when plotted as a frequency curve on magnitude-probability coordinates. By implication, outliers are data points that interfere with the fitting of simple trend curves to the data and, unless properly accounted for, are likely to cause simple fitted trend curves to grossly misrepresent the data. This definition is quite nebulous, furnishes little concrete guidance, and may be confusing to those unfamiliar with flood frequency analysis. However, flood data sets often do not conform to common statistical probability distributions and often contain observations that distort the fit of simple fitted frequency curves. Most flood data points are distributed within some range of moderate extent, but some values extend above the range by factors of 10 or more; thus statistical analysis usually is based on the logarithms of the flows. In arid environments, streams sometimes may be dry all year long, so that the annual maximum "flood" flow may be zero or, perhaps worse, a factor of 10 or more below the range of ordinary flows, so that computations with logarithms are impossible or prone to difficulties. Bulletin 17-B rather loosely gathers all of these issues, which generally may result from a range of causes, but often manifest themselves as poor fit of the fitted frequency curve, under the single and somewhat misleading term "outlier."

Classical statistical concepts of outliers involve ideas of rejection of spurious observations, such as surveying measurements of the azimuth of the North Church steeple rather than of the Blue Hill triangulation beacon. These ideas are not generally very relevant to frequency analysis of properly quality-assured published flood flow data, and Bulletin 17-B does not recommend or provide procedures for rejection of outliers.

The related and more modern notion of contamination of one measurement distribution by another is a special case of the concept of mixed populations mentioned briefly in another section of Bulletin 17-B. Different flood-generating hydrologic processes in one basin give rise to mixed populations of floods. The mixture may manifest itself as outliers, especially if the populations are quite different and one of them occurs relatively infrequently. In this case, the outlying observations could be called hydrologic outliers and could be treated either by the Bulletin-17-B outlier adjustments or by mixed-population analyses, as discussed in another one of these FAQs, with substantially similar results.

Another source of outliers is simply the purely chance occurrence of extraordinarily large observations in some samples. This kind of outlier is more common in flood distributions where one tail or the other generally is somewhat stretched out relative to the normal distribution, and is especially common in so-called "heavy-tailed" distributions such as the Pareto. These may be called "statistical" outliers, and are exemplified by the Bulletin-17-B concept of high outlier, in which it is postulated that the record length associated with the outlier is governed by historical (or paleoflood) evidence rather than simply by the systematic streamgaging record period.

Bulletin-17-B procedures do not involve "dropping" of outliers. High outliers are retained in the analysis as systematic peaks if usable historical information cannot be found. If historical information is available, the high outlier is properly discounted to a more appropriate time window. Low outliers are counted, not "dropped", and their frequency (which is the valid information that they contain relative to flood risk) is used to properly account for the occurrence of low outliers by means of the conditional probability adjustment of a preliminary conditional frequency curve based on the non-outlying observations.

The extensive discussion of outliers in Bulletin 17-B is necessary because of the prevalence of outlier-like effects in frequency analysis of flood data. Despite whatever faults might be found, the discussion in Bulletin 17-B is a generally worthy effort to provide the theoretical concepts and operational procedures needed to enable different analysts to produce reasonable and consistent fitted frequency curves in the variety of problems encountered in practice.

===== LIMITS OF FREQUENCY CURVE EXTRAPOLATION =====

Question: What are the limitations on flood frequency curve extrapolation?

We recently had a terrible flood in our town. I own a trailer park that was above the maximum flood level. I downloaded flood data and used the Bulletin-17-B flood frequency methodology to prove that the flood was a 5,320-year flood and that my trailer park was above the 6,000-year flood level. I wanted the Flood Agency to certify my trailer park, but they would say only that the flood was more than twice as big as the 100-year flood and that my park was outside the 100-year flood plain. Why doesn't the agency acknowledge the true rarity of this flood and the true safety of my property?

Answer: Extrapolation of flood frequency curves is limited primarily by the user's tolerance for uncertainty in the extrapolated results. The user of flood-frequency data needs to understand that these data carry substantial uncertainties with them, even if no extrapolation is involved. The user has to accept responsibility for using these results in such a way that errors in the results do not lead to catastrophic consequences to actions based on the results. Because of the vagaries of flood occurrence in time and space, any observed flood record is likely to give a more or less inaccurate representation of the true magnitude and frequency of flooding. This so-called random-sampling uncertainty is smallest near the middle of the flood distribution (the 2-year flood) and increases for larger less frequent flood magnitudes. This uncertainty is represented by the confidence limits in Bulletin 17-B; the limits are farther apart, representing greater uncertainty, in the tail of the distribution than in the center (near the 2-year flood). Random sampling uncertainty exists and is greater in the tail of the distribution even if extrapolation is not an issue and even if the mathematical form of the distribution is known. In practice, the record length or sample size usually is small (20-60 years) in relation to the annual exceedance probabilities or recurrence intervals of interest (100-500 years), so extrapolation is necessary for obtaining the needed information. Moreover, the mathematical formula that should be used for the extrapolation is not known with any confidence, and there is no agreed-upon procedure to assess or quantify the uncertainty in the extrapolation formula. As a result, the following rules generally are followed: 1) don't extrapolate if you don't have to; 2) if you do have to extrapolate, do so, but only as far as necessary; 3) seek additional information to provide independent corroboration of the extrapolated values (see Bulletin 17-B, pages 19-22); and 4) don't give too much credibility to or place too much reliance on the extrapolated values. For many types of engineering design and planning, there are authoritative design criteria that specify recurrence intervals or exceedance probabilities that must be used; in such situations, extrapolation to those levels is required, like it or not. Commonly used design recurrence intervals include 100 years, 500 years for design of scour protection for major bridges, and shorter intervals for less

important works. Bulletin 17-B shows recurrence intervals up to 500 years (annual exceedance probabilities down to 0.002) in the example problems; it may be assumed that there is a consensus that extrapolation out to that level, if necessary, is acceptable, even if not necessarily accurate or reliable.

In other cases, however, there may be no essential need to extrapolate. Estimation of the recurrence interval of an observed flood by long extrapolation of a frequency curve, for example, generally serves no useful purpose in terms of flood control or flood plain planning and management. (Think of it -- What difference does it make to the winner of a raffle -- or to the losers -- whether he had the one winning ticket in 100 or the one winning ticket in 1000?) If extrapolation is necessary, and, for that matter, even if it is not, prudence dictates that corroboration be sought, and that more corroboration be sought the longer the extrapolation. Thus, it

is always prudent to compare at-site Bulletin-17-B frequency curves with regional flood frequency relations and, if the extrapolation is longer, with flood records at comparable nearby sites and with regional rainfall and runoff relations (Bulletin 17-B, pages 19-22).

If long extrapolation is required, it probably is required because of a concern that exceedance of the design flow would cause catastrophic damage that must be avoided by setting an extremely high design flow; in such cases, if the extrapolated design flow is very uncertain, and if the uncertainty cannot be reduced by comparison with other regional flood information, it might be prudent to consider whether an alternative system design, having less catastrophic failure modes, might be preferable.
