Hydrologic Frequency Analysis Work Group (HFAWG) Meeting
June 12-13, 2013
Michael Baker, Jr., 3601 Eisenhower Ave., Alexandria, VA

The Hydrologic Frequency Analysis Work Group (HFAWG) met at the office of Michael Baker, Jr., 3601 Eisenhower Ave., Alexandria, Virginia on Wednesday, June 12, and Thursday, June 13. The meeting on June 12 was the official meeting for the entire HFAWG and the meeting on June 13 was for the Testing Group to plan future action items. The agenda for the June 12, 2013 meeting is given in **Attachment 1**. Sixteen people attended the June 12 meeting in person and nine people attended by conference call and live meeting. The attendees at the June 12 meeting are given in **Attachment 2**.

Investigations and Progress Since September 2012

After introductions were made and the agenda was approved, Will Thomas gave a powerpoint presentation titled "EMA-Bulletin 17B Investigations and Progress since September 2012". The presentation covered the following major topics:

- Why the work group was updating Bulletin 17B – to deal with interval data and multiple high and low thresholds, to make more efficient use of historical data, to deal more effectively with zero flows, peaks below a gage base and low peaks, to develop more accurate confidence limits, and to improve the estimation of generalized skew
- The recommended changes in Bulletin 17B were listed and noted that there would be a vote later in the meeting on sending these recommendations to the Subcommittee on Hydrology (SOH)
- The voting process was explained and the HFAWG members eligible to vote were identified (http://acwi.gov/hydrology/Frequency/HFAmembership.html)
- The September 2012 meeting primarily discussed the Multiple Grubbs-Beck (MGB) test for identifying low peaks and the adoption of a slightly revised MGB test
- The action items from the September 2012 meeting were discussed and noted as being completed – rerun 82-station sample with revised MGB test, rerun the Monte Carlo simulations with the revised MGB test, update the March 2012 Testing Report with the new results, and release version 7.0 of the USGS PeakFQ software
- Discussed two comparisons of EMA and Bulletin 17B for gaging stations in Texas with low peaks where the 1-percent chance discharges differed significantly

The two gaging stations in Texas were:

- Llano River near Junction, Texas (08150000), 91 years of systematic record, where the EMA 1-percent chance discharge is 36.4 percent less than the Bulletin 17B estimate.
- Mission River at Refugio, Texas (08189500), 71 years of systematic record, where the EMA 1-percent chance discharge is 48.0 percent higher than the Bulletin 17B estimate.

The point of providing results for these two stations was to illustrate that sometimes EMA provides a higher estimate and sometimes Bulletin 17B provides a higher estimate. Will provided 1-percent chance discharges for the two stations when censoring different numbers of low peaks with each method and also provided the standard deviation and skew for each analysis.

The observations or conclusions were:

- When the skew is slightly negative and the standard deviation is significantly inflated by inclusion of the low peaks, Bulletin 17B will tend to provide higher estimates than EMA if the low peaks are not censored. When the censoring of low peaks has more impact on the standard deviation than on the skew, Bulletin 17B gives higher estimates than EMA.
- When the skew is very negative because of inclusion of low peaks, Bulletin 17B will tend to provide lower estimates than EMA. When the censoring of low peaks has more impact on the skew than on the standard deviation, Bulletin 17B gives lower estimates.
- There is no systematic trend for EMA to be higher or lower than Bulletin 17B when censoring low peaks. For the 21 stations where the 1-percent chance discharge differed by 9 percent or more, EMA gave higher 1-percent chance estimates at 11 stations and Bulletin 17B gave higher 1-percent chance estimates at 10 stations.
- The tendency for EMA or Bulletin 17B to give higher or lower estimates is dependent on the impact on the standard deviation and skew when the low peaks are censored. Will noted that this tendency had been observed in many Bulletin 17B analyses.

Recommended Revisions to Bulletin 17B

John England gave a powerpoint presentation titled "Recommended Revisions to Bulletin 17B and Plans for Bulletin 17C". John discussed each of the seven recommended changes in Bulletin 17B and the reasons why these changes were needed. These recommendations are listed below:

1. Historical Information, Low Outliers, Interval Data and Zero Flows. Replace the Historical Weighting Procedure and the Conditional Probability Adjustment (CPA) with an Expected Moments Algorithm (EMA) analysis when such special procedures are needed.

2. Low Outlier Identification. Generalize the simple Grubbs-Beck outlier test recommended in Bulletin 17B with the new Multiple Grubbs-Beck test (Cohn et al., 2013a,b; Lamontagne et al., 2013) for the identification of potentially influential low flows.

3. Confidence Intervals. Replace the confidence interval formulas in Bulletin 17B which neglect the uncertainty in the estimated coefficient of skewness with a computation based on an EMA analysis, that includes skewness uncertainty and reflects historical information and low outlier adjustments based on Cohn et al. (2001) and subsequent numerical improvements for large skews.

4. Derivation of Regional Skew. Revise statements in Bulletin 17B on the derivation of a regional skewness estimator and its precision to reflect recent advances in regional statistical analyses

5. Plotting Positions. Replace the single threshold historical plotting position with the multiple-threshold plotting positions suggested by Hirsch and Stedinger (1987).

6. Climate Change. Replace the outdated statements in Bulletin 17B on "Climate Trends" with a revised statement reflecting the current understanding of climate variability and climate change.

7. Expected Probability. Remove the discussion of Expected Probability from Bulletin 17B. The method is no longer used.

These recommendations are discussed in detail in the memorandum to the Subcommittee on Hydrology (SOH) from the HFAWG that is given as **Attachment 3**. John explained that the HFAWG was going to vote later in the meeting on forwarding these recommended changes to the SOH. Since the recommended changes are discussed in some detail in Attachment 3, no further details are provided here.

Improved Confidence Intervals with EMA

Tim Cohn gave a powerpoint presentation titled "Improved Confidence Intervals with EMA". Tim explained that the existing Bulletin 17B confidence limits did recognize the uncertainty in skew (the third moment), and ignored historical information and regional skew information. Tim noted that the EMA confidence limits described in the Cohn et al. (2001) paper in Water Resources Research were asymptotically correct and correctly recognized historical and regional information. However, those confidence limits were "kinky" and were too wide in samples with skew far from zero (see Figure 1 below). The skew of the data in Figure 1 is 1.93 for the 88 peaks.

Tim has corrected the problems shown in Figure 1 by transforming to uncorrelated central moments and using Guassian quadrature instead of numerical derivatives to estimate the confidence limits. An example of the revised confidence limits for the same station is given in Figure 2 below.

Tim showed the results of Monte Carlo simulations that illustrated the revised confidence limits cover the data the correct percentage of the time. The revised confidence limits are in the USGS PeakFQ Version 7.0 and Tim's standalone program PeakfqSA Version 0.995.

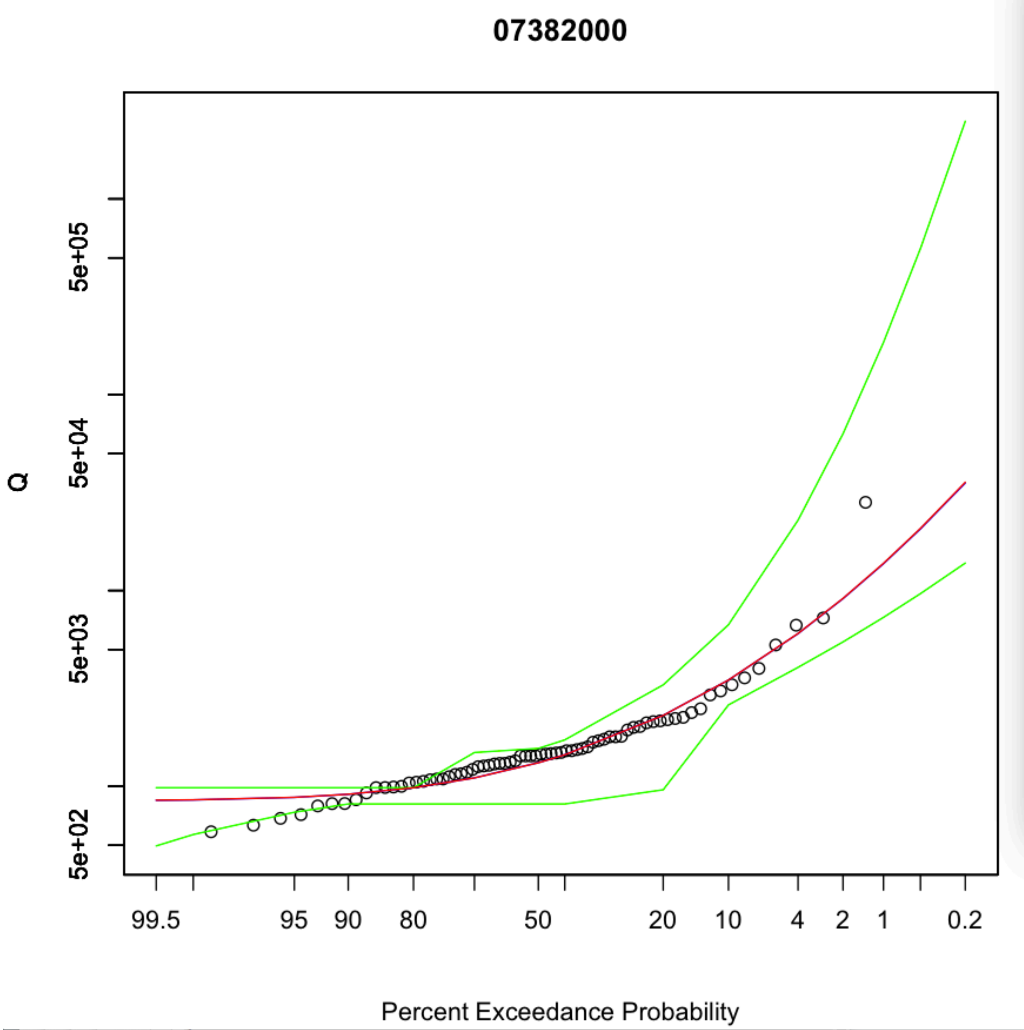

Figure 1. Example of “kinky” confidence limits for station 07382000 that are too wide for large skew values (based on procedures in Cohn et al., 2001).

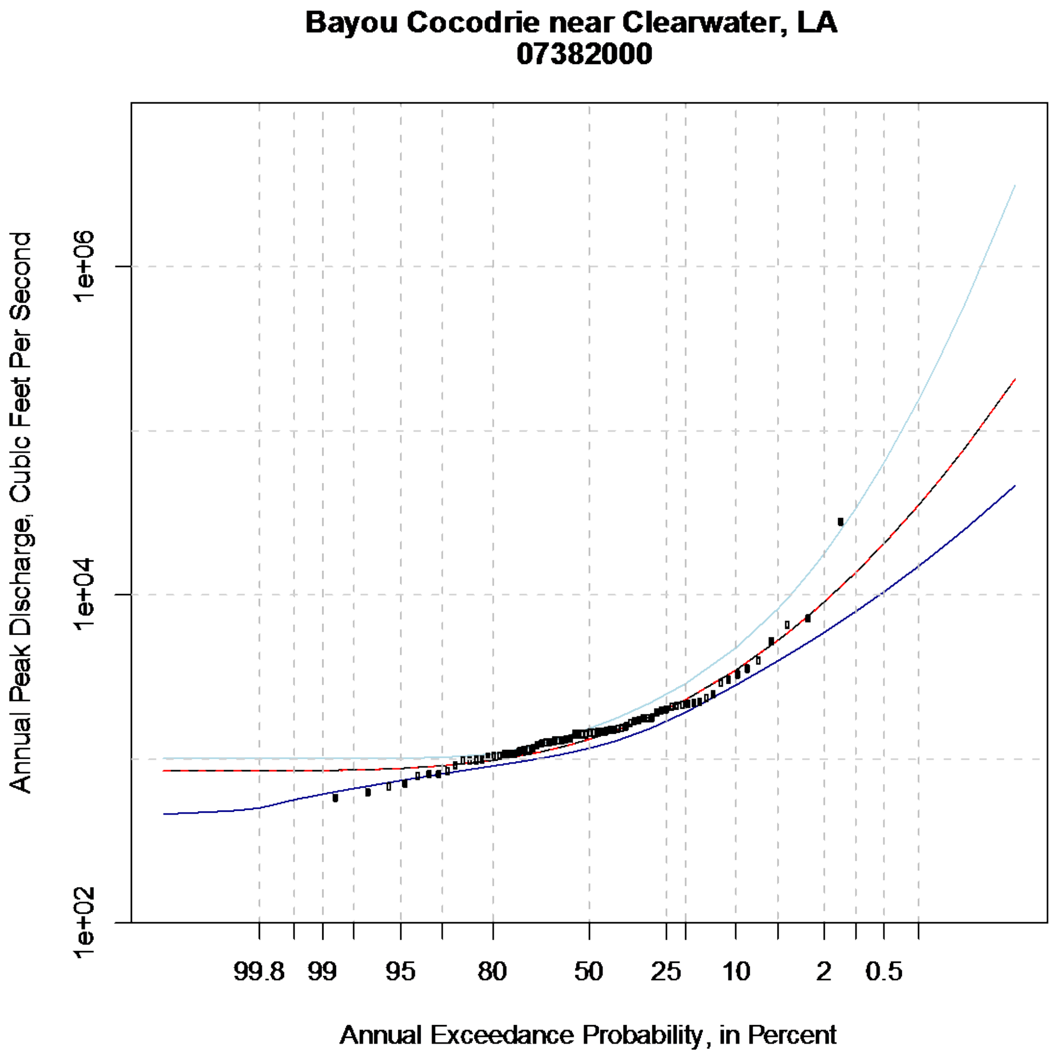

Figure 2. Example of revised confidence limits for EMA for station 07382000 as of 2013.

Multiple Grubbs-Beck Test and Potentially Influential Low Flows

Jery Stedinger gave a powerpoint presentation titled "Robust National Flood Frequency Guidelines: What is an Outlier?" Jery noted that the definition of a low outlier has changed from the Bulletin 17B definition "Data points which depart significantly from the trend of the data". Jery noted that an outlier could be a zero flow or near zero flow or a Potentially Influential Low Flow (PILFs). The new Multiple Grubbs-Beck (MGB) test is used to identify PILFs that have too much influence or leverage on the upper end of the frequency curve. Jery explained how the MGB test identifies low peaks in an outward and inward sweep. The details of how the MGB test works is explained in the ASCE EWRI paper by Lamontagne et al. (2013) titled "Robust National Flood Frequency Guidelines: What is an Outlier?" that was previously distributed to HFAWG members. Monte Carlo simulations were used to demonstrate that the MGB test is robust in identifying PILFs and demonstrate there is no loss in efficiency when censoring low peaks from samples with negative skew or zero skew. For samples with positive skew, PILFs are seldom identified. The paper by Lamontagne et al. (2013) and Jery's presentation provides additional justification for the use of the new MGB test.

Testing Report – Responses to Comments and Action Items

The revised Testing Report was sent out to HFAWG members on April 19, 2013 with a request to members to provide review comments on the report. Review comments and questions on the report and responses to those comments were provided to HFAWG members on June 10 just prior to this meeting and are provided in **Attachment 4**. There was a limited discussion of the review comments during the meeting. In addition, NRCS requested that Professor Rick McCuen of the University of Maryland provide comments on the April 19, 2013 Testing Report. Dr. McCuen's comments and responses from the Testing Group are provided in **Attachment 5**. These comments were also discussed briefly during the meeting.

During the discussion of the review comments on the Testing Report and in the earlier presentations, a number of issues (referred to as parking lot issues) came up and are documented as follows:

1. There is a need for a document that describes the systematic and historic data sources, collection, use and interpretation of the data and threshold input to EMA.
2. There is a need for a national skew map or national skew estimation procedure.
3. Recommendation 6 on climate change needs additional discussion and possible revision before inclusion in Bulletin 17C.
4. Jerry Coffey asked the question why the confidence interval in Figure 2 did not include all the points. The response was that the confidence interval is for the Pearson Type III frequency curve and not the data points. Bill Merkel commented that the data points are based on plotting positions and a change in plotting positions would cause the data points to change. Jerry commented that the confidence intervals reflect the errors in the centroid and slope of the curve and the fact that some very high flows fall within the bounds around the EMA curve reflects the larger error in the EMA slope.
5. With the identification of several PILFs at some stations, Zhida Song-James asked whether the new Bulletin 17C procedures would be applicable for estimating the more frequent floods (environmental flows) like the 1.5- and 2-year floods. The response was that Bulletin 17C would only be applicable to the larger floods like the 10-year floods and greater. The documentation in Bulletin 17C needs to be clear that the new techniques are applicable for flood flows like the 10-year flood and greater. The lower environmental flows should be based on the partial duration series, not the annual maximum series.
6. Martin Becker asked about the definition of flood risk management. Based on the discussion that followed, this term relates to reducing flood risk through the use of levees, dams and floodplain management. The suggestion was to refer to publications of USACE and FEMA on this subject. The comment was made that EMA is in the best interest of flood risk management because it provides a more defensible estimate of flood risk.

7. There is a need for documentation for high outliers. Someone asked the question about the purpose of the high outlier test and it is to identify those flood peaks that are unusually large for a given sample size. For these large peaks, the analyst is encouraged to find historical information. If no historical information can be found, the flood peak should be retained in the sample assuming the data are accurate.
8. Jerry Coffey suggested that the censoring of low peaks in the MGB test should be capped at a percentile less than the median in order to reduce the number of PILFs identified. Don Woodward suggested that the 25th percentile might be a good place to start. Tim Cohn commented that he had originally tried the 25th percentile as a starting point for the downward sweep but later changed it to the median. Tim indicated he never came up with a good quantitative criterion for judging what starting point was "better".
9. There was a suggestion that more evaluation of the "best results" for 82 stations was needed and more interpretation of the Monte Carlo sampling.
10. There is a need to provide written responses to all of Rick McCuen's questions and comments on the Testing Report. Attachment 5 is a more complete response to Rick's comments.
11. Before the Testing Report is finalized, there is a need for more dialogue between reviewers, additional peer review and distribution of comments to all reviewers. Robert Mason commented that the Testing Report will receive additional review independent of the HFAWG as it goes through the USGS report approval process.
12. The minutes of this meeting should reflect the comments and responses on the Testing Report. Attachments 4 and 5 document responses to review comments received to date.
13. Jerry Coffey handed out a one-page document where he assessed the quality of the B17B fit to the frequency curve. Jerry divided the 82 stations into three groups where the B17B fit to the empirical data is good, where the B17B fit is less good, and where the B17B fit had problems. Jerry's qualitative assessment is given as **Attachment 6**.

Discussion and Vote on Recommended Revisions to Bulletin 17B

The next part of the meeting was related to a discussion and vote on the recommended revisions to Bulletin 17B that are given in Attachment 3.

Jery Stedinger made a motion to adopt the June 12, 2013 memorandum (with any revisions made during this meeting) as the HFAWG report to the Subcommittee on Hydrology with the following attachments:

- Link to the latest version of the Testing Report,
- Minutes of the June 12, 2013 meeting, and
- Questions/responses on the Testing Report as received by the June 12, 2013 meeting.

Don Woodward seconded the motion and the Chair asked for a vote. There are 16 voting members of the HFAWG excluding the Chair as documented on the HFAWG web site at http://acwi.gov/hydrology/Frequency/HFAmembership.html. Thirteen of the 16 voting members were present and voted. The vote was 12 to 1 in favor of adopting the recommended changes in Bulletin 17B and sending the memorandum in Attachment 3 to the Subcommittee on Hydrology. The results of the voting are given in **Attachment 7**.

The memorandum in Attachment 3 and the attachments listed above will be forwarded to the Chair of the Subcommittee on Hydrology (SOH) prior to the next meeting. At this time, the date of the next SOH meeting has not been determined.

Testing Group Meeting on June 13, 2013

The Testing Group met on June 13, 2013 to discuss revisions to the Testing Report and make plans for drafting Bulletin 17C. The title of the Testing Report was revised to "Evaluation of Recommended Revisions of Bulletin 17B". The content of the report was reorganized as shown in **Attachment 8**.

The vision for Bulletin 17C is that it will be a relatively short document with appendices posted on the web. This is consistent with Bulletin 17B where most of the information is contained in the appendices.

Will Thomas
Chair, HFAWG
July 3, 2013 – revised July 12, 2013

Attachment 1. Agenda for June 12, 2013 meeting.

Subcommittee on Hydrology of the Advisory Committee on Water Information
Hydrologic Frequency Analysis Work Group (HFAWG) Meeting
June 12, 2013
Michael Baker, Jr., Inc., 3601 Eisenhower Ave., Alexandria, VA

Detailed Agenda

Meeting Activities:

1. Discuss improvements and accomplishments since the September 2012 HFAWG meeting:
   a. software that implements all recommended techniques – USGS PeakFQ 7.0 (beta);
   b. improved version of Multiple Grubbs-Beck test;
   c. revised confidence intervals.
2. Discuss the April 19, 2013 Testing Report
3. Vote on a motion to approve recommended revisions to Bulletin 17B and plans for Bulletin 17C.

Time	***Topic, Presentation and Relevant Documents***	***Presenters/ Discussion Leads***
10:00 am - 10:15 am	Introductions, Member Agencies, Call to Order, Voting Procedures, Motion to Approve Agenda	Will Thomas, HFAWG Chair
10:15 am - 10:45 am	EMA-Bulletin 17B investigations and progress since September 2012 • *Presentation* (slides) – Will Thomas	Will Thomas, HFAWG Chair
10:45 am – 11:00 am	Recommended Revisions to Bulletin 17B and Presentation of Motion to Revise Bulletin 17B • *Recommendations Memorandum to SOH* (handout) – John England • *Presentation of Recommendations and Motion* (slides) – John England	Will Thomas, John England, Tim Cohn, Beth Faber
11:00 am – 11:20 am	Improved Confidence Intervals with EMA • *Presentation* (slides) – Tim Cohn	Tim Cohn
11:20 am – 11:50 am	Multiple Grubbs-Beck Test and Potentially-Influential Low Flows • *Presentation* (slides) – Jery Stedinger	Jery Stedinger, Tim Cohn, Jonathan Lamontagne
11:50 am – 1:30 pm	Testing Report – Responses to Comments and Action Items • *Responses to comments* (sent via email) *Working Lunch (provided)*	Group Discussion
1:30 pm- 3:00 pm	Discussion and Vote on Recommended Revisions to Bulletin 17B and Plans for Bulletin 17C	HFAWG Members
3:00 pm – 4:00 pm	Work Groups and Assignments for Bulletin 17C and future work	Will Thomas, HFAWG

Attachment 2. Attendees at the June 12, 2013 meeting of the Hydrologic Frequency Analysis Work Group (HFAWG).

Name	Company/Agency	Telephone
Jerry Coffey	Retired OMB	703-944-4845
Robert Mason	USGS	703-648-5305
John England	USBR	303-445-2541
Tim Cohn	USGS	703-395-0204
Will Thomas	Michael Baker/ASFPM	703-334-4935
Zhida Song-James	Michael Baker/ASFPM	703-317-6257
Beth Faber	HEC-USACE	530-756-1104 x335
Victor Hom	NWS	301-713-0006 x173
Martin Becker	Self	404-876-3900
Jery Stedinger	Cornell University	607-257-8016
Ken Eng	USGS	703-648-5843
Sanja Perica	NWS/OHD	301-713-1669
William Merkel	NRCS	301-504-3936
Don Woodward	Global Ecosystems Center	301-216-5948
Thomas Nicholson	NRC	301-251-7498
Mohammed Haque	NRC	301-415-4041
By phone/live meeting		
Mike Eiffe	TVA	
Nancy Barth	USGS	
Mark Bandurraga	Ventura County	
Bruce Rindahl	Ventura County	
Andrea Vellieux	USGS	
Jonathan Lamontagne	Cornell University	
Claudia Hoeft	NRCS	
John Onderdonk	NRC	
Ben Pratt	SRBC	

Attachment 3. Recommendations Memorandum to the Subcommittee on Hydrology

Subcommittee on Hydrology, Advisory Committee on Water Information
Hydrologic Frequency Analysis Work Group (HFAWG)

MEMORANDUM

TO: Subcommittee on Hydrology

FROM: Hydrologic Frequency Analysis Work Group

SUBJECT: Recommended Revisions to Bulletin 17B

DATE: June 12, 2013

It has been 30 years since the last revision of Bulletin 17B in March of 1982. At that time, it was recognized that continued investigation and improvements of flood frequency techniques were needed. In fact, Bulletin 17B (pages 27-28) included a list areas where additional research was recommended by the Work Group in 1982.

Basis for Recommendations

The Hydrologic Frequency Analysis Work Group (HFAWG) has completed a series of studies to develop and justify proposed revisions to Bulletin 17B in the following four main areas:

1. Historical information and the weighted-moments approach;
2. Low outlier detection and treatment;
3. Procedures for estimating generalized/regional skew; and
4. Procedures for estimating confidence limits.

These studies were based on the published literature and the testing plan presented to the Subcommittee on Hydrology in 2006 (HFAWG, 2006) and summarized by Will Thomas (HFAWG, 2007). The January 2006 proposal observes that “the possible changes are considered significant improvements and would warrant the publication of a new Bulletin 17C”.

The results of the testing plan appear in Cohn et al. (2013a), with additional technical details in Cohn et al. (2013b), England and Cohn (2012), and Lamontagne et al. (2013) and references therein. The key change is the adoption of the Expected Moments Algorithm (EMA) framework for the analysis of data sets containing zeros, outliers, interval flow estimates, multiple thresholds, or historical and/or paleoflood information as the appropriate generalization of the method-of-moments to address such situations.

The simple and clear tests described in Cohn et al. (2013a) demonstrate that the Expected Moments Algorithm (EMA) with the log-Pearson Type III distribution appears to always perform as well, and in many cases with extreme censoring and/or historical information, does much better than the special algorithms in Bulletin 17B for dealing with such cases. Moreover, EMA is able to make use of a wider range of data types reflecting interval estimates and multiple *threshold for historical information and low outliers, common data occurrences that B17B does not address.*

Recommended Revisions to Bulletin 17B

While retaining the basic structure of Bulletin 17B that uses the method of moments in log-space with the log-Pearson Type III distribution and weighted skew coefficient, the HFAWG recommends the adoption of several corrections and extensions to those procedures. Most of these revisions follow from the list of needed research included in Bulletin 17A and 17B. All of the changes are in the spirit of the procedures currently recommended in Bulletin 17B.

We recommend that a new Bulletin 17C be issued with the following revisions.

1. Historical Information, Low Outliers, Interval Data and Zero flows. Replace the Historical Weighting Procedure and the Conditional Probability Adjustment (CPA) with an Expected Moments Algorithm (EMA) analysis when such special procedures are needed.

2. Low Outlier Identification. Generalize the simple Grubbs-Beck outlier test recommended in Bulletin 17B with the new Multiple Grubbs-Beck test (Cohn et al., 2013a,b; Lamontagne et al., 2013) for the identification of potentially influential low flows.

3. Confidence Intervals. Replace the confidence interval formulas in Bulletin 17B which neglect the uncertainty in the estimated coefficient of skewness with a computation based on an EMA analysis, that includes skewness uncertainty and reflects historical information and low outlier adjustments based on Cohn et al. (2001) and subsequent numerical improvements for large skews.

4. Derivation of Regional Skew. Revise statements in Bulletin 17B on the derivation of a regional skewness estimator and its precision to reflect recent advances in regional statistical analyses

5. Plotting Positions. Replace the single threshold historical plotting position with the multiple-threshold plotting positions suggested by Hirsch and Stedinger (1987).

6. Climate Change. Replace the outdated statements in Bulletin 17B on "Climate Trends" with a revised statement reflecting the current understanding of climate variability and climate change.

7. Expected Probability. Remove the discussion of Expected Probability from Bulletin 17B. The method is no longer used.

Further details on the each of these seven recommendations are provided in a subsequent section.

The U.S. Army Corps of Engineers, Bureau of Reclamation, and U.S. Geological Survey have invested substantial resources in conducting technical studies and developing these recommendations. See Olsen (2011) and the March 2012 testing group memorandum (HFAWG, 2012).

Perspective on Recommended Bulletin 17B Revisions

The recent studies by Cohn et al. (2013a), Cohn et al. (2013b), England and Cohn (2012), and previously published studies on issues related to Bulletin 17B that are listed in Cohn et al. (2013a), Stedinger and England (2005), and Stedinger and Griffis (2008), provide the technical basis for making recommendations to revise some aspects of Bulletin 17B. In addition to the recent data testing and simulation studies described in Cohn et al. (2013a), there have been many additional studies conducted by HFAWG members, collaborators, and others since 1995. Beyond such studies, EMA has been used in several regional frequency studies (California, Iowa, Arizona) conducted by the USGS, and in site-specific investigations conducted by the USBR.

The use of EMA has three advantages.

1. EMA is the reasonable extension of the Bulletin 17B LP-III method of moments approach to deal in a consistent statistical framework with ALL of the sources of information likely to be available. For simple cases with only a systematic record and a regional skew, the EMA algorithm reverts to the method of moments recommended in Bulletin 17B.

2. EMA deals with interval and multiple threshold data that CPA and the Historical Weighted-Moments procedures in Bulletin 17B do not. This new capability allows use of an expanded data set as well as describing better what is actually known. Crest-stage gages and sites with historical information yield observations that are best described by intervals, where thresholds often change over time.

3. Adoption of EMA will provide confidence intervals (CIs) that include skew uncertainty, which is neglected in Bulletin 17B, and will reflect the information provided by observations described by different intervals, for which Bulletin 17B has no option.

Plans for Bulletin 17C

The Hydrologic Frequency Analysis Work Group (HFAWG) commenced discussions and outlined studies on potential improvements to Bulletin 17B at the November 2005 HFAWG meeting. Stedinger and England (2005) summarized the critical papers, results available at that time, and listed proposed changes to Bulletin 17C. After the January 2006 HFAWG presentation to SOH, we commenced development of testing plans (HFAWG, 2007) and outreach.

Stedinger and Griffis (2008) provide further discussion and perspectives on the flood frequency literature and investigations for improving Bulletin 17B. That editorial in the *ASCE Journal of Hydrologic Engineering,* along with presentations at many professional meetings, including ASCE, ASFPM, US Flood Management Association, and the Federal Interagency Hydrologic Modeling Conference (England and Cohn, 2007, 2008; Stedinger et al. 2008; Thomas et al., 2008, 2010; Lamontagne et al. 2013), and Federal agency discussions (Olsen, 2011), ensured that the profession was aware of our efforts and the likely outcome.

We have developed a plan and recommendations for retaining, updating, replacing and deleting each of the sections and appendices within Bulletin 17B. The recommended revisions and status of particular sections within Bulletin 17B are summarized in two tables that are presented below We recommend four activities to prepare and implement a Bulletin 17C.

1. Develop the Bulletin 17C Document
    - writing team consists of HFAWG members, with USGS, USACE, and Reclamation as major contributors;
    - develop outline of document;
    - prepare draft for review;
    - define review, comment and approval process (in conjunction with SOH, ACWI, and public comments);
    - prepare final version;
    - define future revision process.

2. Develop Supporting Materials to Bulletin 17C
    - web site for FAQ, references, software links (also used as outreach);
    - prepare various conference papers, journal articles and related reports;
    - webinars to SOH and wider hydrology and engineering community on the technical background, improvements and materials, such as EMA, MGB, etc.

3. Develop Software for Bulletin 17C
    - individual agency software packages are under development (USGS, USACE, Reclamation);
    - provide application examples with software.
    - NOTE: USGS has completed and released beta software with examples for testing, and has held several webinars in 2012 and 2013.

4. Conduct Outreach and Training on Bulletin 17C
    - present Bulletin 17C update plan at professional meetings;
    - develop training on Bulletin 17C materials;
    - provide software demonstrations and training with Bulletin 17C and specific software packages.

Details on Recommendations for Bulletin 17B Revisions

The recommended changes and status of particular sections within Bulletin 17B are summarized in Table 1 (Bulletin 17B Main Report) and Table 2 (Bulletin 17B Appendix below). Refer to pages 9-25 in IACWD (1982) for additional details.

Table 1: Bulletin 17B Main Report features, condition, and recommendation

Bulletin 17B Assumption or Step	***Procedure***	***Condition***	***Recommendation***
IV.	**Data Assumptions**		
IV.A.	Climatic Trends	OK	Update
IV.B.	Randomness of Events	OK	Update
IV.C.	Watershed Changes	OK	Keep and review
IV.D.	Mixed Populations	OK	Keep
IV.E.	Reliability of Flow Estimates	OK	Update/supplement
V.	**Determination of Flood Frequency Curve**		
V.A.	Series Selection - Annual floods	OK	Keep and Clarify
V.B.	**Statistical Treatment**		
V.B.1	The Distribution – LP3	OK	Keep
V.B.2.	Fitting the Distribution (Method of Moments)	OK	Generalize with EMA
V.B.3.	Estimating Generalized Skew	Narrow	Replace
V.B.4.	Weighting the Skew Coefficient	OK	Generalize
V.B.5.	Broken Record	OK	Generalize with EMA
V.B.6.	Incomplete Record/Crest-stage gages (CPA)	Limited	Replace with EMA
V.B.7.	Zero flood years (CPA)	Limited	Replace with EMA
V.B.8.	Mixed-population	OK	Keep and review
V.B.9.	Outliers (Grubbs-Beck test)	Limited	Generalize with MGB
V.B.10.	Historic flood data	Limited	Replace with EMA
V.C.	**Refinements to Frequency Curve**		
V.C.1.	Comparisons with Similar Watersheds	OK	Keep and review
V.C.2.	Flood Estimates from Precipitation	OK	Keep and review
VI.	**Reliability Application**		
VI.A.	Confidence Limits	Limited	Replace with EMA
VI.B.	Risk	OK	Keep
VI.C.	Expected Probability	Unnecessary	Delete
VII.	**Potpourri**		
VII.A.	Non-conforming special situations	OK	Amend
VII.B.	Plotting Positions	Limited	Generalize
VII.C.	Future Studies	OK	Update

Table 2: Bulletin 17B Appendix features, condition, and recommendation

Bulletin 17B Appendix	***Procedure***	***Condition***	***Recommendation***
1	References	Outdated	Update
2	Glossary and Notation	Outdated	Update
3	Table of K Values	Keep	Review and revise to electronic format
4	Outlier Test K Values	Unneeded	Delete
5	Conditional Probability Adjustment	Limited	Delete
6	Historic Data	Limited	Delete
7	Two-Station Comparison	OK	Keep
8	Weighted Independent Estimates	OK	Keep
9	Confidence Limits	Incorrect	Replace
10	Risk	Awkward	Keep and review
11	Expected Probability	Unnecessary	Delete
12	Flow Diagrams and Example Problems	Outdated	Update
13	Computer Program	Outdated	Update
14	"Flood Flow Frequency Techniques" report summary	Limited	Delete
New	Additional Resources For Special Situations (Provide links to FAQ, other websites, etc.)	---	New
New	Trend Tests	---	New
New	Multiple-Threshold Plotting Positions	---	New
New	Expected Moments Algorithm (EMA)	---	New

Pertinent information on seven of the recommendations follows. Additional details are provided in Cohn et al. (2013a), other reports presented to HFAWG, and references therein.

1. Historical Information, Low Outliers, Interval Data and Zero flows. Flood records can contain zeros and low outliers, potentially influential low flows, crest-stage observations with different thresholds, historical information with one or more thresholds, and perhaps quantification of uncertainty in the measurement of some large events, all of which can be efficiently, accurately, and consistently represented as interval data. B17B does not provide for interval data, thus an extension is needed to the Bulletin 17B weighted moments (WM) and conditional probability adjustment (CPA) procedures. The Expected Moments Algorithm (EMA) has been developed as a collaborative effort with the USGS, Reclamation, and Cornell University (Lane, 1995; Cohn et al., 1997; England et al., 2003; Griffis et al., 2004, Cohn et al. 2012). The Expected Moments Algorithm (EMA) provides a single statistically-consistent framework for estimating the parameters of the LP3 distribution with the wide range of data types experienced in practice (England and Cohn, 2012), and for estimating the uncertainty in estimated model parameters, flood quantiles and related parameters. We recommend that EMA be adopted as the appropriate extension of the current Bulletin 17B method-of-moments approach.

2. Low Outlier Identification. Bulletin 17B employs the Grubbs-Beck test for low outliers. That test is intended to determine if the smallest observation in a sample of size *n* is inconsistent with a normal distribution using a type I error of 10%. Experience has shown that flood frequency studies in the Western United States, and other arid areas, need to be sensitive to more than one outlier per sample. Furthermore, some records have one or more zero flows, and for such records the Grubbs-Beck test is not an appropriate test to employ to check if the smallest positive (non-zero) observation is an outlier at the 10% level. Thus we recommend that the Grubbs-Beck test in Bulletin 17B be generalized to the Multiple Grubbs-Beck (MGB) test described in Cohn et al. (2013b) and Lamontagne et al. (2013) to evaluate if one or more observations should be considered to be low outliers because they are potentially-influential observations that could negatively affect the fitted LP3 distribution. Furthermore we recommend that the Multiple Grubbs-Beck test be employed in the two step algorithm described in the Testing Report (Cohn et al., 2013a). Examples in Cohn et al. (2013a), Cohn et al. (2013b), and Lamontagne et al. (2013) illustrate the advantages of this MGB procedure.

3. Confidence Intervals. Currently the formulas in Bulletin 17B for generating confidence intervals assume the weighted skewness coefficient is without error. This misrepresents the uncertainty in flood quantile estimators and the errors that result are well documented (Chowdhury and Stedinger, 1991). As a result, the Bulletin 17B confidence intervals fail to cover quantiles with the specified frequency (Cohn et al., 2001). To eliminate that limitation, and to be consistent with the use of EMA for parameter and flood quantile estimation, we recommend that flood studies use the EMA procedures for describing the uncertainty in estimated parameters and quantiles. These EMA confidence interval procedures are initially described in Cohn et al. (2001) and have been revised to include gaussian quadrature numerical solution techniques that provide improved estimates for large skew coefficients. While one would hope that we could correct the confidence intervals now utilized in B17B, the ad-hoc nature of the adjustment procedures (historical, low outliers, zero flows, weighted skew) now employed by B17B simply does not lend itself to an accurate and uniform confidence interval estimation procedure.

4. Derivation of Regional Skew. The weighting of a regional skewness estimator with the at-site skewness estimator remains an important and innovative component of the Federal guidelines in Bulletin 17B. The current weighting formula is an important part of the 1982 revision. However the value of that step depends on use of the best available regional information and the appropriate description of the precision of that information, which enters into the calculation of the weighted skewness estimator. Bayesian GLS procedures extract far more regional information for skew estimation than is presented in B17B Plate 1, and provide a relatively unbiased estimate of their precision. Moreover, we have an additional 30 years of data with which to develop regional skew estimators. We recommend that the section on "Estimating Generalized Skew" be revised to reflect GLS Bayesian procedures described in USGS SIR 2010-5260, USGS SIR 2009-5043, USGS SIR 2009-5158, and USGS SIR 2009-5156.

5. Plotting Positions. Bulletin 17B recommends plotting procedures applicable for a single threshold. However modern applications employing multi-threshold crest stage data, or multiple threshold historical information, require a multiple-threshold plotting position strategy. Thus we recommend that the probability plotting position formula with a single threshold be replaced by the corresponding multiple-threshold plotting positions recommended by Hirsch and Stedinger (1987) and the Handbook of Hydrology (Stedinger et al., 1993, p. 18.42. A plotting parameter *a* = 0, corresponding to a Weibull formula, is recommended as a default value, consistent with current practice. Other plotting parameters, including 0.40 (Cunnane), 0.44 (Gringorten), and 0.50 (Hazen) could also be considered. There would be no change for complete data sets that lack zeros, censored data, low outliers, or historical information.

6. Climate Change. The current statement in Bulletin 17B about "Climate Trends" is inaccurate based on our current understanding of climate variability and climate change, and should be re-written. (See IV. Data Assumptions, A. Climate Changes.). We recommend it be replaced with the following text:

"There is much concern about changes in flood risk associated with climate variability and long-term climate change. Time invariance was assumed in the development of this guide. In those situations where there is sufficient scientific evidence to facilitate quantification of the impact of climate variability or change in flood risk, this knowledge should be incorporated in flood frequency analysis by employing time-varying parameters or other appropriate techniques. All such methods employed need to be thoroughly documented and justified."

7. Expected Probability. Bulletin 17B contains a discussion of expected probability on page 24 with an expanded explanation including Tables and formula in Appendix 11. Given that results were only available for the case of normal data with zero skew, and that the one agency that used that adjustment no longer does so, we recommend that the discussion of Expected Probability in the Bulletin be omitted from Bulletin 17C.

References

Cohn, T.A., W.L. Lane, and W.G. Baier (1997) An algorithm for computing moments-based flood quantile estimates when historical information is available, Water Resour. Res., 33(9), 2089–2096.

Cohn, T.A., W.L. Lane, and J.R. Stedinger (2001) Confidence Intervals for EMA Flood Quantile Estimates, *Water Resources Research*, *37*(6), 1695-1706.

Cohn, T.A., Barth, N.A., England, J.F. Jr., Faber, B., Mason, R.R. and Stedinger, J.R. (2013a) Evaluation of Recommended Revision of Bulletin 17B. Draft USGS Open File Report, July 2013.

Cohn, T.A., England, J.F. Jr., Berenbrock, C., Mason, R.R. Stedinger, J.R., and Lamontagne, J.R. (2013b) A Generalized Grubbs-Beck Test for Detecting Multiple Potentially Influential Low Outliers in Flood Series, revised manuscript submitted to Water Resources Research, June, 41 p. ftp://ftp.usbr.gov/jengland/HFAWG/EMA_LowOutliers.final.pdf

England, J.F. Jr. and Cohn, T.A. (2007) Scientific and Practical Considerations Related to Revising Bulletin 17B: The Case for Improved Treatment of Historical Information and Low Outliers, American Society of Civil Engineers, EWRI World Water & Environmental Resources Congress, May 15-19, 2007, Tampa, FL, 9 p.

England, J.F. Jr. and Cohn, T.A. (2008) Bulletin 17B Flood Frequency Revisions: Practical Software and Test Comparison Results, American Society of Civil Engineers, EWRI World Water & Environmental Resources Congress, May 13-16, 2008, Honolulu, HI, 11 p.

England, J.F. Jr. and Cohn, T.A. (2012) Diverse Extreme Flood Data for Flood Frequency and Case Studies with the Expected Moments Algorithm. Draft Report Submitted to the Hydrologic Frequency Analysis Work Group, March, 41 p. ftp://ftp.usbr.gov/jengland/HFAWG/EMA_Multiple_Censoring_Examples.pdf

England, J.F., Jr., J.D. Salas, and R.D. Jarrett (2003) Comparisons of two moments-based estimators that utilize historical and paleoflood data for the log Pearson type III distribution, Water Resour. Res., 39(9), 1243, doi:10.1029/2002WR001791.

Griffis, V.W., J.R. Stedinger, and T.A. Cohn (2004) LP3 Quantile Estimators with Regional Skew Information and Low Outlier Adjustments, Water Resources Research, 40, W07503, doi:1029/2003WR002697.

Hirsch, R.M. and J.R. Stedinger (1987) Plotting Positions for Historical Floods and their Precision, Water Resources Research, 23(4), 715-727.

Hydrologic Frequency Analysis Work Group (HFAWG) (2006) Flood Frequency Research Needs. Memorandum Submitted to Subcommittee on Hydrology, January 16, 3 p. http://acwi.gov/hydrology/Frequency/minutes/HFAWG_for_revising_Bulletin17B_4.pdf

Hydrologic Frequency Analysis Work Group (HFAWG) (2007) Summary of Gaging Station Data and Testing Methods for the EMA and Bulletin 17B Comparisons. Prepared by Will Thomas, Hydrologic Frequency Analysis Work Group, Subcommittee on Hydrology, revised August 8, 4 p. http://acwi.gov/hydrology/Frequency/minutes/Testing_Methods_for_the_EMA_and_Bulletin_17B_080707.pdf

Hydrologic Frequency Analysis Work Group (HFAWG) Testing Group (2012) Recommended Revisions to Bulletin 17B. Prepared by Testing Group, Hydrologic Frequency Analysis Work Group, Subcommittee on Hydrology, March 26, 8 p. ftp://ftp.usbr.gov/jengland/HFAWG/Testing-Recommendations-to-HFAWG-revised-26March2012.doc

Interagency Committee on Water Data (IACWD) (1982) Guidelines for determining flood flow frequency: Bulletin 17-B. Hydrology Subcommittee, March 1982 (revised and corrected), 28 p. and appendices.
http://water.usgs.gov/osw/bulletin17b/dl_flow.pdf

Lamontagne, J.R., J.R. Stedinger, T.A. Cohn, and N. Barth (2013) Robust National Flood Frequency Guidelines: What is an Outlier?, inWorld Environmental & Water Resources Conference, Cincinnati, OH, edited by C.L. Patterson, S.D. Struck, and D.J. Murray, ASCE EWRI.

Lane, W.L. (1995) Method of moments approach to historical data. Informal handout, Bulletin 17B Working Group, Hydrology Subcommittee, Interagency Advisory Committee on Water Data.

Olsen, J.R. (2011) Meeting on Flood Frequency Analysis and Bulletin 17B revisions. Memorandum for Record, U.S. Army Corps of Engineers, Institute for Water Resources, Alexandria, VA, 2 p. ftp://ftp.usbr.gov/jengland/HFAWG/Olsen-USACE-Bulletin17C-MFR-10Jan11-Mtg%20Final.pdf

Stedinger, J.R. and England, J.F. Jr. (2005) Key Documents for Discussion of Evolution of Bulletin 17B and Proposed Changes. Submitted to Hydrologic Frequency Analysis Work Group, Subcommittee on Hydrology, November 3, 9 p. ftp://ftp.usbr.gov/jengland/HFAWG/Revision-Bulletin17B_11-05.doc

Stedinger, J.R. and Griffis, V.W. (2008) Flood frequency analysis in the United States: Time to update. Journal of Hydrologic Engineering, 13(4):199–204.

Stedinger, J.R., R.M. Vogel and E. Foufoula-Georgiou (1993) Frequency Analysis of Extreme Events, Chapter 18, Handbook of Hydrology, D. Maidment (ed.), McGraw-Hill, Inc., New York.

Stedinger, J.R., W.O. Thomas, T.A. Cohn (2008) Update to Bulletin 17B: Flood Frequency Guidelines in the U.S., Floodplain Management Ass., 2008 Ann. Conference, Paradise Point Resort San Diego, CA, Sept 2-5.

Thomas, W., J. England, T. Cohn, and B. Koper (2008) Recent Developments in Flood Frequency Analysis Including Plans to Update Bulletin 17B. Assoc. State of Flood Plain Managers, Annual Conference, Reno, NV, May 18-23.

Thomas, W., England, J., Cohn, T. and Steinberger, N. (2010) Recent Developments in Flood Frequency Analysis Including Plans to Update Bulletin 17b. Proceedings, Fourth Federal Interagency Hydrologic Modeling Conference, Las Vegas, NV, Jun 28-Jul 1, 2010, 4 p.

Attachment 4.

Questions and Comments on the April 19, 2013 Testing Report *Updating Bulletin B17B for the 21st Century* by Cohn, Barth, England, Faber, Mason and Stedinger

6/9/2013

Question 1: Since the report on numbered page 8 states that the purpose of the report is to consider the advantages and disadvantages of possible changes to Bulletin 17B, what are the disadvantages of EMA and EMA MGB; and, where are they stated in the testing report?

Response: We will restate the purpose of the report to be "to document the relative performance of EMA and Bulletin 17B on observed and simulated data." As noted below, the cited promise is more than the report provides. The report includes the results of the tests requested, and some additional Monte Carlo results and explanations. The powerpoint presentations provided to the HFAWG at the March 19, 2012 and September 20, 2013 meetings discuss the advantages and disadvantages of the two estimation procedures as well the HFAWG summary report of recommendations to the SOH. See also editorial by Stedinger and Griffis (2008), and proceedings paper by England and Cohn (2007)

Stedinger, J.R., and Griffis V. W., Flood Frequency Analysis in the United States: Time to Update (editorial), *J. of Hydrologic Engineering, 13*(4), 199-204, April 2008.

England, J.F. Jr. and Cohn, T.A. (2007) Scientific and Practical Considerations Related to Revising Bulletin 17B: The Case for Improved Treatment of Historical Information and Low Outliers, American Society of Civil Engineers, EWRI World Water & Environmental Resources Congress, May 15-19, 2007, Tampa, FL, 9 p.

ACTION: The HFAWG will develop a summary of recommendations that better articulates the pros and cons of the two methods and their potential adoption. The Testing Report (Cohn. etal., 2013) will focus on the results of the requested tests. Edit Testing Report to clarify its purpose.

Question 2: Please cite any published, scientific literature endorsing EMA and MGB and/or procedures for identifying PILP's by authors other than the authors of the testing report?

Response: Active researchers such as Jose Salas, Richard Vogel, Veronica Griffis, and Charles Kroll have written papers on EMA, but all have published with members of the HFAWG. The Bureau of Reclamation reviewed and endorsed EMA prior to MGB being developed. The motivation for adopting EMA is given the Reclamation's "Guidelines for Evaluating Hydrologic Hazards" (ftp://ftp.usbr.gov/jengland/Dam_Safety/Hydrologic__Hazard_Guidelines_final.pdf). Authors of these guidelines included persons not involved in developing the Testing Report

The MGB test was developed as part of research conducted by the HFAWG so references to MGB are for papers authored by HFAWG members. The MGB test is constructed following the reasoning in Rosner (1975, 1983) but the test is being applied differently. The Cohn et al. (2013) paper on the MGB test,that was submitted to the Water Resources Research journal, has a literature review of all pertinent research on low outliers.

ACTION: None. Relevant citations have been provided.

Question 3: For each of the following stations, which is the most correct 100-year flow determination when comparing the EMA MGB and Bulletin 17B results; and, what is the basis for determining most correct in each case: 06176500, 08150000, 08189500, 09480000, 10234500, 11028500, 11152000, 11176000, 05291000, and 06216500?

Response: The data group requested to see the methods applied to real data. And that has been done. Sometimes one can see that for a particular record one estimator does poorly. But because we do not know the right answer, generally it is difficult or impossible to make a valid scientific claim that one is most correct. A possible way to judge whether one method is more reasonable when applied to observed data is to evaluate the data and utilize engineering (hydrologic) judgment.

The comparison of B17B and EMA based on actual, resampled data was not without value. It helped us detect an unexpected outcome. Such was the case in the early pre-MGB comparisons when we discovered that EMA often performed more poorly than B17B when low outliers were involved because it (unlike B17B which uses CPA) was attempting to incorporate information on the influential low peaks. That comparison, in part, led us to realize the pathology of over fitting low peaks at the expense of high peaks and to develop the MGB.

But the fairer test for evaluating statistical estimators is by looking at their average error in repeated sampling when the population from which the samples is drawn is known. Such analyses have been included. See section 5 of the report. The same approach was followed when looking into the challenges posed by non-LP3 and mixed populations studies see section 5.4 of the report.

ACTION: There is no universal way to determine which is "most correct" when the true distribution is not known. In real world applications, an engineer or hydrologist must make a decision based on the available data and engineering (hydrologic) judgment.

Question 4: What statistical criterion was originally used and which criterion is presently being used to determine the best technique? I cannot find a reference to this in the report.

Response: Chapter 3 entitled "Metrics for Evaluating Flood-Frequency Estimators" provides the statistical criteria employed in the evaluation. The equations for the Effective Record Length (ERL) and Average Gain (AG) are given on pages 49-50 of the report.

Question 5: Is the present system broken or are we trying to update a system that is 30 years old?

Response: We are doing both. We are basically trying to improve a methodology that is 30 years old by more accurately incorporating historical information, detecting low peaks that are influential, improving the computation of confidence limits and plotting positions, and providing a description of a better way to estimate generalized skew and its precision.

ACTION: Response has been provided. HFAWG recommendation should be clearer on this point.

Question 6: Is the current system of determining confidence limits really broken?

Response: The confidence limits in Bulletin 17B were always approximate, but the best they knew how to do in 1976. The new approach more accurately includes historic information, uncertainty in the skew coefficient, and the uncertainty in censored and interval data. The following reference points out some deficiencies in the Bulletin 17B confidence limits:

Chowdhury, J. U., and J.R. Stedinger, Confidence Intervals for Design Floods with Estimated Skew Coefficient, *Journal of Hydraulic Engineering, 117*(7), 811-831, 1991.

ACTION: Adopt new and more accurate confidence interval computation.

Question 7: Is the proposed system of determining flood frequencies dependent on the EMA MGB technique totally?

Response: Is the question are we going to use EMA/MGB for all applications or only some applications; that is, use Bulletin 17B for some applications and EMA/MGB for others? The plan is to use EMA/MGB for all applications (historic data, low peaks, systematic record only). As noted in the presentations and the Testing Report, if there are no historic data, low peaks or interval data, then EMA is the same as the log-space method of moments estimator in Bulletin 17B. EMA is an effective extension of the moments estimator in Bulletin 17B that addresses interval flood descriptions.

ACTION: None. Clarification provided.

Comment 8: I would like to see an independent review of the proposed techniques by someone not involved in its formulation.

Response: The HFAWG members are providing the independent review. That is the purpose of the HFAWG. The Subcommittee on Hydrology (SOH) and the ACWI will also provide independent reviews on Bulletin 17C. Lastly, there will be a public comment period on Bulletin 17C through the Federal Register. A paper on the MGB Test was submitted to Water Resources Research for review in December 2012. There was a recent presentation at the May 2013 EWRI World Congress discussing the new MBGT and its impact on B17B and EMA performance.

John England and Timothy Cohn published a number of papers on EMA with coauthors not on the HFAWG. A series of articles by Griffis and Stedinger has appeared in the Journal of Hydrologic Engineering, and Water Resources Research. Clearly publication in those professional journals required independent reviews conducted by the journal editors and anonymous reviewers.

ACTION: Independent review to be provided by the HFAWG, SOH, ACWI and public comment.

Comment 9: The conclusions have changed little but the methods and results have changed more. In the two latest meetings we have had graphic illustrations of the multi-year process of fine tuning the EMA to perform better on the challenge data set. In September we saw this cut-and-try approach in all of its gory details. Over the years we have made similar claims for calculations that we now find deficient (see page 27 of the current report).
Response: We have gotten better at addressing a much wider set of problems cases. That is progress, and it is a good thing. This is what scientific research looks like, and what one would expect to see in a work group. Moreover, we hope to continue to improve in the future as we can better couple the physical characteristics and location of watershed with the flood risk across the floodplain, as well as addressing possible climate change impacts.

Question 9: Why haven't we documented the process we all saw in September, i.e., fine tuning the estimator to the test data?
Response: Please see the report submitted and discussed at the September 2012 meeting, and the Proceeding paper from the 2013 EWRI Congress. It appears that the Cohn et al. (2013) MGB paper is about to be accepted by Water Resources Research. Does the question mean that we should document the path we took to get where we are? What is the point of writing that history at this time?

Lamontagne, J. R., J. R. Stedinger, T. A. Cohn, and N. A. Barth, Robust National Flood Frequency Guidelines: What is an Outlier? World Environmental and Water Resources Congress 2013: Showcasing the Future, American Society of Civil Engineers, pp, 2454-66, Cincinnati, Ohio, May 19-23, 2013.

ACTION: Will continue to develop documentation of new procedures.

Comment 10: Throughout section B, we see numerous sites where B17/GB does not perform well, and EMA/MGB does just as badly.

Question 10: Why isn't this acknowledged and explained in the report? ("does as well or better" is SPIN, not analysis)

Response: Please be more specific as to the cases where you feel the situation has been misrepresented. The statement, “does as well or better" seems like a very appropriate conclusion.

ACTION: Will continue to attempt to refine the discussion included in the Testing Report.

Comment 11: In parts of section B, there are some instances where differences are driven by historical extrapolations beyond p = .01.

Response: What are “historical extrapolations”? Are you referring to extrapolations of the frequency curve beyond p = 0.01 or the use of historic data that exceeds the 0.01 flood?

Question 11: Since EMA assumes stationarity beyond 100 years and this is probably not true (more on this later), for estimates with historical components, should we consider the B17/GB estimate to be likely to be more accurate?
Response: B17B involves fitting the LP3 with method-of-moments. EMA is used to fit the LP3 with essentially method-of-moments. The only difference is that EMA properly accounts for any historical data or censored data, where as B17B utilizes ad-hoc approaches. There is no difference between the two approaches concerning stationarity assumptions. In either case, we would expect the same treatment as we would apply to records affected by landuse change and dam-release influenced records. The analyst needs to access the data for trends, and to the extent that he can, ensure that those trends are either removed or incorporated into the analysis (a topic that we would envision being taken up with a effort to develop a follow-on B17D).

ACTION: None. Issue explained.

Comment 12: In section B.3, there are numerous cases where points labeled as "outliers" by MGB that do not appear to be outliers at all.
Response: The issue is whether observations are PILFs. How have you determined that the points in questions should not be treated as PILFs? And if we treat a few too many points as PILFs, does it cause any harm?

This was a topic that was discussed at great length in the September 2012 HFAWG meeting. See the powerpoint presented by Stedinger at that time for an overview of the important points. Also see the 2013 EWRI conference proceedings paper, and the new Cohn et al. (2013) MBGT paper being reviewed by Water Resources Research.

Question 12: Why isn't this disconnect discussed and explained?
Response: The philosophy related to detecting PILFs is discussed in other documents besides the Testing Report. See report given to the HFAWG at the September 2012 meeting Stedinger presentation at that time, the minutes of the September 2012 meeting that reflect why we are identifying PILPs rather than low outliers in the Bulletin 17B definition, and the paper below:

Lamontagne, J. R., J. R. Stedinger, T. A. Cohn, and N. A. Barth, Robust National Flood Frequency Guidelines: What is an Outlier? World Environmental and Water Resources Congress 2013: Showcasing the Future, American Society of Civil Engineers, pp, 2454-66, Cincinnati, Ohio, May 19-23, 2013.

ACTION: Will continue to develop documentation of new procedures.

Comment 13: Elsewhere in section B.3, some deviations on the low end of the curve strongly suggest a mixture of two or more patterns influencing the empirical curve. In these cases the differences between the estimated curves appear to be driven more by MGB censoring than by B17 versus EMA.

Question 13: Why isn't the likelihood of mixtures (and the implications beyond MGB censoring) discussed and explored?

Response: Almost all flood records are mixture. This is not a new concern. See discussion in Bulletin 17B. The important point is that MBG helps ensure that those distributions that affect the low end of the flood-frequency distribution, do not overly influence the high end.

ACTION: When this task has been completed, HFAWG can proceed to address other issues.

Comment 14: In sections B.5 and B.6 everything past the first column depends on long-period stationarity, an assumption which is almost surely false. The same problem afflicts the ERL and AG measures.

Question 14: Why is this not discussed and explained?

Response: This is beyond the Testing Report, and is a subject of discussion in the literature.

Stedinger, Jery R., and Veronica W. Griffis, Getting From Here to Where? Flood Frequency Analysis and Climate, J. Amer. Water. Resour. Assoc., 47(3), June 2011, Pages: 506–513, DOI: 10.1111/j.1752-1688.2011.00545.x, 1 June 2011.

Comment 15: Executive Summary, page 8 – The purpose of the report is to document the relative performance of EMA and Bulletin 17B on observed and simulated data rather than to "consider the advantages and disadvantages of possible changes to Bulletin 17B procedures".

Response: Yes, the purpose of the Testing Report will be revised.

Comment 16: Introduction, page 13 – Suggest briefly listing the advances in regional skew analyses that have been described in USGS reports. Also suggest briefly listing the improvements in the computation of confidence intervals. There is no need to describe these advancements in any detail but a listing of improvements will inform the reader what improvements are planned.
Response: This is beyond the scope of the testing report. Instead see:

Veilleux, A.G., and J. R. Stedinger, Appendix. Regional Skewness Regression, In Eash, D.A., Barnes, K.K., and Veilleux, A.G., 2013, Methods for estimating annual exceedance-probability discharges for streams in Iowa, based on data through water year 2010: U.S. Geological Survey Scientific Investigations Report 2013-5086, 63 p. with appendix., http://pubs.usgs.gov/sir/2013/5086/.

Lamontagne, J.R., Stedinger, J.R., Berenbrock, Charles, Veilleux, A.G., Ferris, J.C., and Knifong, D.L., 2012, Development of regional skews for selected flood durations for the Central Valley Region, California, based on data through water year 2008: U.S. Geological Survey Scientific Investigations Report 2012–5130, 60 p., URL - http://pubs.usgs.gov/sir/2012/5130/, 2012.

Veilleux, A.G., J. R. Stedinger, and D. A. Eash, Bayesian WLS/GLS Regression for Regional Skewness Analysis for Crest Stage Gage Networks, Paper 227, Crossing Boundaries, Proceedings World

Environmental and Water Resources Congress, ed. Eric D. Loucks, Amer. Society of Civil Engineering, Albuquerque, New Mexico, pp. 2253-63, May 20-24, 2012.

Veilleux, A.G., J. R. Stedinger, J.R. Lamontagne, Bayesian WLS/GLS Regression for Regional Skewness Analysis for Regions with Large Cross-Correlations among Flood Flows, *World Environmental and Water Resources Congress 2011: Bearing Knowledge for Sustainability*, R. E. Beighley and M. W. Killgore (eds.), Amer. Soc. of Civil Engin., Palm Springs, CA, Paper #324, DOI: 10.1061/41173(414)324, May 22 – 26, 2011.

Comment 17: Section 4.1 The Sites and Data, page 17 – The range of record lengths or mean or median of the record lengths should be provided to illustrate that most stations had more than 75 years of record. The authors should indicate that data through the 2010 water year were used, if available, and that the attributes of the stations are given in Section A. Note that Skykomish River near Gold Bar, WA (12134500) shows that Ngb = 1 in Section A but it should be Ngb = 0. That is, the Grubbs-Beck test did not identify one low peak for this station.

Response: Appreciate editorial suggestion.

Comment 18: Section 4.3 Methods, page 18 – This section should indicate that the three combinations of parameter estimators and outlier identification procedures include Bulletin 17B with the MGB test.

Response: Editorial change will be made.

Comment 19: Section 4.3.2, page 20 – The statement is made that the GB test rarely identifies more than a single outlier. Note that the Grubbs-Beck test identified 2 outliers for the 2 of the 37 stations that had identified outliers by either test.

Response: Recall that many of the 82 records considered in the report were selected because they were unusual, to highly unusual. So this set of unusual records contained more than the average number of unusual records.

Comment 20: Section 4.4, page 24 – The description of CPA is correct but it would be clearer to describe the procedures for estimating the exceedance probability (as given in Appendix 5 of Bulletin 17B), not the non-exceedance probability.

Response: Appreciate editorial suggestion. We will review the presentation.

Comment 21: Figure 3, page 31 – Figure 3 is the first use of box plots and it would be helpful to the reader to give a brief description of box plots (median, interquartile range, whiskers, outlying observations) just prior to Figure 3. The readers will then have a better understanding of the information given in the box plot figures.

Response: Appreciate editorial suggestion.

Comment 22: Figure 11, page 42 – There is no vertical scale on this figure and no legend describing the methods.

Response: Appreciate editorial suggestion.

Comment 23: Figures 23-25 do not have vertical scales.
Response: Appreciate editorial suggestion.

Attachment 5.

Responses to Professor Rick McCuen's Comments on the April 19, 2013 Testing Report
By HFAWG Testing Group
7/3/2013

The memo addresses comments provided on Professor McCuen. The memo first addresses his summary page, followed by responses to notes he made in the margins of the report.

Comments on

UPDATING BULLETIN B17B FOR THE 21ST CENTURY

GENERAL COMMENT: The analyses themselves are likely completed as intended and under the assumptions made. However, serious concern should be given to the objectivity of many of the observations/conclusions. Primary concerns are as follows:

Prof. McCuen is correct that the report for the most part describes results of tests that were requested by the HFAWG. We will return to explore what we think Prof. McCuen means as "objectivity" when we respond to #2 below.

1. Hydrologic reasoning for many of the conclusions is missing. Most of the conclusions depend entirely on statistical assessment. For example, for some data sets a large percentage of the observations are censored even though a hydrologic explanation for the supposed outliers is not given. It would be helpful to know if the censored values were from a different type of storm event.

Several issues here:

(i) At the September 2012 HFAWG meeting, a vocabulary was developed to explain how we proposed to extend the Grubbs-Beck outlier test to more aggressively identify what we called PILFs – Potentially Influential Low Flows. PILFs are low values that POTENTIALLY will be highly influential, so we give them special treatment so as to let the larger values in the record dominate in the fitting of a flood curve. We are not claiming that PILFs are true rogue values, outliers, or badly measured flows. Our concern is that they have too much *leverage* and thus potentially too much *influence*. From a statistical point of view.

(ii) Yes, it would be wonderful to employ more hydrologic insight in the analysis, and to know if the low values reflected different hydrologic processes. Indeed that should be done by those conducting individual investigations. However, the focus here was the development and testing of objective and consistent statistical procedures for identifying PILFs, and the analysis of records with censored observations. This was not explained clearly in the April 2013 Testing Report, but appears in other documents.

2. Most assessments are not objective, just subjectively evuluated even though the report stresses the importance of consistency. Objective decision criteria are lacking. Decisions appear to be subjectively made. Just a few examples:
   a. "performs substantially better" (page 51)
   b. "performs slightly better" (56)
   c. "perform better" (58)

It seems that quantitative results, such as a bias or mean square error, are not cited with such comments.

These comments focus on text in Chapter 5, which reports Monte Carlo studies using specified parametric frequency curves, and resampling of real records (a form of Monte Carlo analysis). The variance estimators was computed, then converted to an Effective Record Length, and Average Gains (AG). The average gain values are reported for all cases – so there are quantitative metrics that reflect the variance. The bias did not effect the computation.

The analysis will be improved by computing AG based upon the MSE. Then the quantitative performance metric would reflect both bias and variance. However, AG had been defined in terms of variance, because for the resampling studies we do not know what the true value of the quantiles are, and thus cannot determine bias and MSE. In the next revision of the report we should use AG_{MSE} for the Monte Carlo studies, and AG_{var} for the resampling studies when the true T-year flood is not known.

3. The analyses seem to place an over emphasis on historical data. Two of the three criteria (page 16) stress historic data. I would suspect that historic data are less accurate than the flows in the systematic record. The criteria on page 16 need to be reassessed and replaced with more useful statistics.

The motivation for the development of EMA was to make better use of historical information. Research has shown that EMA makes better use of historical information than B17B, and there is not much difference with the quality of the fit reflecting low outliers. So indeed the report stresses historical information. Still cases with and without low outliers are included in the analyses.

In terms of the criteria, ERL and AG were introduced in chapter 3 (page 16), but defined in section 5.1; these two sections will be combined so that one can see what quantitative criteria were used. Yes, historical information can be less accurate. However, large floods during the period of systematic record also have larger measurement errors because they may be beyond the rating curve.

```
4.  The MGB test assumes normality.  If flow data are assumed
to be from LP3 populations, wouldn't an outlier test that is
based on LP3 values be more appropriate and lead to more
accurate assessment of outliers?
```

First, the MGB test does assume normality, which is equivalent to the LP3 distribution with zero skew. So zero skew & LP3 is the null hypothesis of the outlier tests. Second, Spencer and McCuen (1996) is an important paper that provides statistical tools one could use to do what is proposed here. However, the intent is to identify PILFs which occur in samples with a negative skew; negative skew indicates that in log space the smallest observations have become more important in moment computations than the largest observations. Thus we adopt zero-skew LP3 as our concern, and seek to categorize the smallest statistically-significant observations in such records as PILFs. See Cohn et al. (2013) and Lamontagne et al. (2013) for an elaboration of this idea. We are not proposing a test of the skew, or a test for real outliers; rather we seek to identify PILFs so as to make the estimation procedure more robust.

Cohn, T.A., J. F. England, R. R. Mason, J. R. Stedinger, and J. R. Lamontagne. A generalized Grubbs-beck test for detecting multiple potentially influential low outliers in flood series. Under review *Water Resour. Res.,* 2013.

Lamontagne, J. R., J. R. Stedinger, T. A. Cohn, and N. A. Barth, Robust National Flood Frequency Guidelines: What is an Outlier? C.L. Patterson, S.D. Struck, and D.J. Murray (eds.), J. World Environmental and Water Resources Congress 2013: Showcasing the Future , American Society of Civil Engineers, pp, 2454-66, Cincinnati, Ohio, May 19-23, 2013.

Spencer, C.S. and R.H. McCuen, Detection of outliers in Pearson type III data, J*. of Hydrologic Engr., 1*(1), 2-10, 1996.

```
5.  There seems to be an over reliance on box plots as a means
of drawing conclusions.  It is important to keep in mind that
the quantiles of a box plot based on sample data are values of
a random variable and thus have confidence intervals.  For the
sample sizes used, the intervals can be quite wide, which makes
decisions based on box plots subject to concern in the same way
other statistics are.
```

Box plots were a nice way to display the distributions of the different estimators so their entire distribution (including extreme outliers) could be assessed visually, rather than reporting just a variance or MSE. Prof. McCuen is correct that the precision of the box plots should be a concern. These are paired tests in that all three estimators considered the same set of “random” records. Thus the box plots show relatively precisely the relative performance of the 3 estimators across THOSE samples. The Effective Record lengths and Average Gains are also reported. In the final version of the report. larger sample sizes will also be employed.

6. Some of the assumptions made in the analyses need better justification. For example, the use of infinity and zero on pag 29 do not seem realistic. A flow of infinity is unlikely and a flow of zero would suggest a dry river--no baseflow. While zero flows are likely in some areas, are they realistic on Hurricane Creek, IL?

Prof. McCuen is of course correct. If the gauge record shows a flow was less than 100 cfs, and thus was not recorded, one can certainly argue that while zero is a correct lower bound, with some effort perhaps it could have been established that 30-100 cfs is the range in which the flow occurred. Will use of 0-100 instead of 30-100 make a difference? In most cases it will not; it will not make a difference if the fitted model assigned a probability of almost zero t the range 0-30 cfs. Similarly, if physical evidence and records show a flood exceeded 10,000 cfs, then with some effort one might be able to document and defend an upper bound of perhaps 40,000 cfs. But if it is really difficult to resolve an upper bound, use of 10,000-to-infinity cfs is not incorrect. In most cases where the upper bound is a value with an extremely small exceedance probability, the differences will not matter. To say that a flood did not occur in 10,000-to-infinity as opposed to a flow range with a large upper bound that has an exceedingly small exceedance probability does not provide any significant difference in information and it does not affect the parameters of the fitted curve.

Suppose Z were the standardized logarithms of a flood series, and Z is standard normal. If a hydrologist learned that in a historical year Z was less than+5, there is no information in that record. A standard normal random variable is essentially always less than 5 standard deviations. But the basic point is that hydrologists should do their best to determine appropriate interval estimates for historical and unusual low floods.

SUMMARY STATEMENT: The report seems to be premature. Considera more work is needed. A better set of objective criteria, both hydrologic and statistical, are needed. The three on page 16 an box plots are not definitive. The subjective assessments need t be supported with objective results.

With the materials Prof. McCuen had, and the incomplete status of the April 2013 Draft of the Testing Report, the conclusion that the report is "premature" is reasonable. The April 2013 Draft will be improved. We provided Rick with the latest versions of Cohn et al. (2013) and Lamontagne et al. (2013) to provide more background on the MGB Test. Additionally, the excel spreadsheet comparing estimates for the 82 stations and minutes of the March 19, 2012 and September 19, 2012 HFAWG meetings were provided to Rick for background information.

REMARKS IN TEXT MARGINS

p. 14 Suggest discuss availability of data and its accuracy – this is in other documents. But report could cite reference. Accuracy is very site and event specific.

p. 15 Asks for equations for bias and variance, and how will tell if significant.
Equations were in another section; report to be reorganized.

p. 16a Will reorganize the report to have the equations with the text definitions that appear here.

p. 16b See question 3.

p. 17 Asks if analysis assumes that historical data has the same accuracy as large peaks within the systematic record. Analysis uses historical records as recorded in USGS files. Accuracy of peak measurements depends on many factors.

p. 20 Asks for clarification of assumptions of MGB test. Asks if it was normal? While development of the MGB test will be left to other documents, it would make sense to provide more information in this report. In fact in section 4.3.3 on pages 20-21 the text is clear that the MGB test considers samples from a normal sample. See discussion of question #4.

p. 21 Asks if normality assumption is important – yes it is. See discussion of question #4.

p. 27 Recommends use a sensitivity analysis if you have non-standard data. Yes, of course.

p. 29 Asks if infinity and 0 are realistic: see question #6.

p. 29 Asks to explain how important the differences in the representation of the data were. Hope to add such a discussion in final version of report.

p. 30 Asks, how big were differences, and but would it make a difference in design? Design issue is a good question that the engineers would need to consider.

p. 32a Remark noted. See remark on page 33.

p. 32b See question 6, and response.

p. 33 – Figure 4 – comment does not need a reply. However, EMA employs historical information more effectively, and as a result matches the largest systematic flood and the large historical flood better. As a result does not match the smaller peaks below the 1/0.90 = 1.11 year flow quite as well as Bulletin 17B estimator. A 3-paraemter distribution cannot do everything, and the large floods described by the almost 50 years of historical data are the phenomena of interest.

p. 34 Asks for more precision in describing differences between differences – will try.

p. 35 McCuen notes observation was subjective.

p. 37a Yes 111% and 113% are essentially identically. McCuen wanted precision, so here it is.

p. 37b Issue of a hydrologic reason for decisions is addressed in question 1. Outlier test is a statistical exercise. Should be augmented by hydrologic review.

p. 38 Formula for RPD got lost and will be restored.

p. 46 Need to restore definition of RPD.

p. 46 Asks which is correct? Monte Carlo with LP3 data shows that EMA is overall more accurate with historial data. Generally cannot really tell which is better if only have a single sample. However, sentence in question should be revised.

p. 47 Did not test significance of bias comment.

p. 47b Definition and intent of outliers discussed in question #1.

p. 48a Question marks will be resolved.

p. 48b Box plots are visual. ERL and AG reveal value of historical information.

p. 49 Did not study distribution of ERL criterion. ERL is very important because in an understandable metric it quantifies the value of historical information, which was a key issue in the analysis. ERL is a very fair comparison of the methods in the Monte Carlo analysis. It does not always correspond to the real gain if the variation of the MSE is not exactly linear with sample size – but that is not a big deal.

p. 50 Asks how often historical information is available. Presentation included a graphic by Robert Mason showing the frequency occurrence of historical information across the country.

p. 50 Asks for more precision. Asks about the importance of the difference.

p. 51 Asks that criterion be clarified. Report reorganization should help.

p. 53 Requests appropriate clarification of several statements. We should deal with these when report is edited.

pp. 56, 58, 61, 63, 69 Request clarification of statements in text. We should deal with these when report is edited.

Attachment 6. An assessment of the quality of the B17B fit by Jerry Coffey.

8 Months Plus a Few Hours -- Another Perspective June 12, 2013

We can divide the 82 test sites into three groups based on the quality of the B17 fit --

In Group I, the B-17 fit to the empirical data is good. Since the fit at these 38 sites is good we cannot reasonably expect EMA to offer improvement.

In Group II, the B17 fit is less good (some modest room for improvement but small changes only). Since both curves were close to the data, slight advantages were determined by comparing the slope of the fitted curve to the nominal slope of points on the right side of the curve. The B17 curves had a slight advantage in six sites, while EMA showed slight advantages in four sites. None of the fits were poor and the results were pretty much a wash in these 11 sites.

In Group III, B17 fits had problems -- there was serious room for improvement. Large deviations of data points from fitted curves in the 33 sites left substantial concerns about the results of the B17 analysis. EMA appears to have improved the fit at one site (page 164 of the report), in the other 32 challenging sites EMA failed to make any improvement (a failure rate of 97%). In these cases where substantial improvements over the B17 results would have been very useful, EMA gave us virtually nothing.

Note: The actual groupings are shown below, numbers refer to the pages in the report where the site graphs appear. A poignant commentary on the theory argued for MGB appears in the results on pages 139 and 143. Wholesale censoring of low-end points (22 and 46 "outliers" respectively) produces results nearly identical to plain vanilla B17.

GROUP I -- B17 fit good -- 83, 95, 96, 97, 99, 102, 103, 104, 106, 107, 108, 110, 112, 113, 114, 115, 116, 117, 119, 125, 126, 129, 132, 133, 135, 139, 142, 143, 144, 145, 146, 149, 153, 157, 158, 161, 165, 167. Total of 38 sites

GROUP II -- B17 fit less good (but not bad) -- 105, [90, 134, 138, 140, 152, 155 -- advantage B17], [137, 141, 147, 154 -- advantage EMA with MGB "outlier" counts of 16, 45, 42, and 29, respectively]. Total of 11 sites.

GROUP III -- B17 fit not good -- 84, 85, 86, 87, 88, 89, 91, 92, 93, 94, 98, 100, 101, 111, 118, 120, 121, 122, 123, 127, 128, 131, 136, 148, 150, 156, 159, 160, 162, 163, 164, 166, 168. Total of 33 sites.

Attachment 7. Summary of Voting on the Recommended Revisions to Bulletin 17B as given in Attachment 3.

No.	Name	Organization	Vote on recommended changes in Bulletin 17B
1	Siamak Esfandiary	Federal Emergency Management Agency	Not present
2	Don Woodward	Global Ecosystems Center	Yes
3	Martin Becker		Yes
4	Will Thomas	Chair	
5	Zhida Song-James	Michael Baker, Jr.	Yes
6	Tim Cohn	U.S. Geological Survey	Yes
7	Mike Eiffe	Tennessee Valley Authority	Yes
8	Beth Faber	U.S. Army Corps of Engineers	Yes
9	John England	Bureau of Reclamation	Yes
10	Jerry Coffey	Mathematical Statistician	No
11	Rocky Durrans	University of Alabama	Not present
12	Joe Krolak	Federal Highway Administration	Not present
13	William Merkel	Natural Resources Conservation Service	Yes
14	Sanja Perica	National Weather Service	Yes
15	Jery Stedinger	Cornell University	Yes
16	Thomas Nicholson	Nuclear Regulatory Commission	Yes
17	John Onderdonk (for Samuel Lin)	Federal Energy Regulatory Commission	Yes
		Number of members voting	13
		Yes votes	12
		No votes	1
		Abstain votes	0

Attachment 8. Reorganized Table of Contents for the Testing Report

Evaluation of Recommended Revisions to Bulletin 17B

Testing Group

CONTENTS

1 Introduction

2 Literature Sources: The History of Flooding and Flood Risk Estimation

3 Metrics and Estimators
- 3.1 Operational and Statistical Metrics [old section 3]
- 3.2 Identification of outliers and potentially influential low flows [old section 4.3.1]
- 3.3 MGB Test for Identifying PILFs [old section 4.3.2]
- 3.4 Quantile Estimators Explored in study [old section 4.3.3]

4 Comparisons Based on Monte Carlo (old section 5)
- 4.1 Understanding the graphics (old section 5.1)
- 4.2 Studies with LP3 (was 5.2}
- 4.3 Studies of LP3 distribution with regional skew (old section 5.3]
- 4.4 Robustness studies (old section 5.4)
  - 4.4.1 Robustness with respect to Pearson Type 3 Population (old section 5.4.1)
  - 4.4.2 Robustness with respect to Population Constructed from Two LP3 Distributions (old section 5.4.2)
  - 4.4.3 Robustness with respect to Population Constructed from Two GEV Distributions (old section 5.4.3)
- 4.5 Resampling Studies (old section 5.5)

5 Comparisons of Estimator Based on Real Data at Selected Test Sites (old section 4)
- 5.1 The Sites and Data
- 5.2 Software
- 5.3 Sites with Systematic Gage Data and No Low Outliers or Historical Information (old section 4.5)
- 5.4 Sites with Historical Information (old section 4.6)
- 5.5 Sites with Systematic Gage Data and Low Outliers (old section 4.7)
- 5.6 Sites with Low Outlier, Historical and/or High Outliers (old section 4.8)
- 5.7 Summary (old section 4.9)

6 Conclusions