Responses to Comment #48: Stationarity Issues – Comments on the draft report on Bulletin 17C – originally circulated by [redacted] on May 22, 2015

The following comments on Bulletin 17C were excerpted from [redacted] comments dated May 22, 2015. [redacted] is a member of the Hydrologic Frequency Analysis Work Group (HFAWG) and his May 22, 2015 comments were made on the April 20, 2015 draft of Bulletin 17C that was reviewed by the HFAWG. [redacted] comments were previously circulated among HFAWG members.

[redacted] **comment**: The way Bulletin 17B and the write-up of the 17C version approach stationarity are the crux of the problem. The approach in the existing guidance is rather pragmatic while the EMA procedure in the proposed 17C version requires a rigid and fairly extreme stationarity assumption. This puts EMA and thus the 17C changes in extreme conflict with recent results from climatology and meteorology.

Response: The stationarity assumptions in Bulletin 17B and Bulletin 17C are the same. Both approaches assume that the historic (or paleoflood) and systematic data reflect the same land use and climate. Admittedly, the historic peaks may have more uncertainty associated with them but the assumption in both methods is the data were collected under the same meteorological and hydrologic conditions.

Bulletin 17B and 17C have different approaches for incorporating historic and paleoflood data. The weighting procedure in Bulletin 17B does not give much weight to paleoflood data when the historic period is several hundred years. The EMA procedure in Bulletin 17C handles paleoflood data better than Bulletin 17B and we envision greater use of paleoflood data in the future.

[redacted] **comment**: Historical floods, high and low outliers, and zero flow cases all violate the homogeneity criterion of Bulletin 17B and represent failures of stationarity; Bulletin 17B admits this (by providing special procedures), but we will find that EMA (and thus draft Bulletin 17C) tries to use the same calculation even when its stationarity requirements are violated!

Response: Bulletin 17C has different special procedures for incorporating historic data and high outliers and censoring low floods than does Bulletin 17B. However, the special procedures in both methods assume that the moments of the logarithms of the annual flood peaks are not changing with time.

EMA uses a more general data specification such as a range of values rather than point values, and so is able to replace these data types with sensible ranges. Information on historical data and low floods (or zero flows) are still included but the procedure allows use of a range of values when the precise value is inappropriate. This is one of the strengths of the EMA procedure to allow for uncertainty in the historical or paleoflood data and the use of multiple censoring thresholds.

[redacted] **comment**: As a practical matter, the Bulletin 17B form of stationarity is limited to well-behaved gage records (excluding historical floods, outliers and zero flows as indicated above) which also happens

to be the set of cases over which B17B and EMA agree. This is consistent with the hypothesis that EMA ONLY WORKS WHEN THE ASSUMPTIONS THAT ARE CITED IN THE EMA LITERATURE ARE ACTUALLY TRUE.

Response: Bulletin 17B includes historic floods and outliers in the analysis. The reasons for the differences in the two methods are related to different procedures for incorporating the historic floods and high outliers and different procedures for censoring low outliers. Bulletin 17B and 17C agree when there are no historic floods, high outliers or low floods because the two procedures use exactly the same data. These differences and similarities in the methods are NOT related to the assumption of stationarity. Both Bulletin 17B and Bulletin 17C assume the logarithms of the annual peak flows are independent and identically distributed and follow a Pearson Type III distribution. This may not be true for the low floods but these floods will likely be censored in Bulletin 17C.

[redacted] **comment**: EMA advocates originally expected the algorithm to be robust with respect to outliers, but now admit the problems so obvious from early testing (per the following quote from John England's recently released "EMA/LP3 Development History") --

"**2009**

- Draft HFAWG testing report results (82 sites) by England, Cohn, and Steinberger shows low outlier issues with EMA"

What the EMA advocates still denied were the other "issues" where the "identical distribution" requirement was not true, e.g., the historical floods, high outliers, all the "interval" gimmickry (where the distribution was unknown, or at best unmeasurable) and those zero flows out west.

Response: The Bulletin 17C procedure now utilizes a different censoring technique for low floods than was used in 2009. The zero flows and very low floods will be censored using the new Multiple Grubbs-Beck technique. Testing on observed data at 82 gaging stations and simulated data demonstrated the new methodology to be superior to Bulletin 17B.

The use of intervals to describe flood events is a generalization of the data description that strengthens the ability of the EMA procedure to use information of varying levels of quality and also allows more explicit use of floods that are not in fact from the same distribution. This useful element of the EMA algorithm allows a standard computation to replace all the case-specific adjustments made in Bulletin 17B. For Bulletin 17C, all the adjustments are made together and the results do not depend on the order in which those adjustments were made which was a particular challenge for Bulletin 17B when both historical and low outliers were present in the same sample.

[redacted] **comment**: The only attempt to address the stationarity failures that compromise EMA was the MGB fix at the low end of the distribution. As we saw in the analysis distributed in June 2014, MGB as designed by the Cornell engineering department produces huge errors extending far beyond the stationarity problem with outliers, an inappropriate fix for zero flows, and does nothing to address the other stationarity failures (high outliers, historical floods, intervals, multiple thresholds, etc.)

Response: The existence of high outliers, historical floods, intervals and multiple thresholds are NOT stationarity failures. A completely stationary record might contain these types of data. Low outliers as well are not failures of stationarity, though maybe a failure of the identical distribution assumption due to different hydrologic processes, rather than a change in processes over time.

[redacted] **comment**: In fact, the stationarity requirements for EMA (from the 1997 paper) are NOT TRUE in each and every one of these cases. So B17B has carefully developed procedures to deal with these nonstationarities and EMA has replaced them with calculations that are designed to work only under strict assumptions of stationarity! In other words, the revision replaces proven procedures on new procedures that rely on a stationarity assumption that is patently FALSE.

Response: Bulletin 17B procedures do not deal with nonstationarity issues. There are no differences in Bulletin 17B and Bulletin 17C with respect to the stationarity assumption. The real issue of concern here may be the assumption that the logarithms of the annual peak flows are Pearson Type III distributed.

In both Bulletin 17B and Bulletin 17C, the moments of the Pearson Type III distribution are estimated for the total sample with no assumption these moments are changing with time. The EMA algorithm allows for more efficient use of historic and paleoflood data and tends to censor the low floods that may be from a different distribution.

The following comment on Potentially Influential Low Floods (PILFs) was circulated by Jerry Coffey in April 2015 but was submitted as part of Comment #48 during the public review period and is addressed below.

[redacted]**comment**: PILFs are not some magic bullet that increases robustness – just the opposite. Influential points (e.g., PILFs) are not an analytical alternative to outliers, they are either valid points in the tail of a distribution or outliers from the same locations. In other words, PILFs are either valid observations that substantially INCREASE THE EFFICIENCY of an estimator or outliers that represent contamination of the underlying distribution. As I demonstrated in July last year, the MGB procedure cannot distinguish between these two cases. It is throwing out the baby with the bath water – compromising extremely valuable data points that are NOT outliers. MGB is a perverse fix for the frailties of EMA that is not needed by the more robust existing procedure in Bulletin 17B.

Response: PILFs are low floods that have a very strong effect on the shape of the estimated frequency distribution at the high discharge end. The purpose of the MGB procedure is to eliminate the influence of low floods so that the small floods have little or no impact on the frequency estimates at high discharges. The PILFs are valid observations but their magnitude may depart significantly from the trend of the remaining data, the definition of an outlier in Bulletin 17B, or they may not be outliers in the context of the Bulletin 17B definition. The smallest observations in the data set do not convey meaningful or valid information about the magnitude of significant flooding but they do convey valid information about the frequency of significant flooding. Even though the magnitude of the PILFs are not used directly in the frequency computations, they fact they occurred and were censored are used in the

computations. The MGB test is actually more robust in censoring low floods than the GB test in Bulletin 17B that often needed to be subjectively increased to censor low floods.

The MGB test as currently used in Bulletin 17C was agreed to in a September 20, 2012 meeting of the HFAWG (http://acwi.gov/hydrology/Frequency/minutes/HFAWG_meeting_mins_September_20_2012.pdf). The recommended revisions to Bulletin 17B, that included the MGB test, were approved by a 12 to 1 vote at a June 12, 2013 meeting of the HFAWG (http://acwi.gov/hydrology/Frequency/minutes/HFAWG_mins_june12-13_2013_meeting_revised_071213.pdf). The MGB test was documented in a peer-reviewed journal paper in Water Resources Research in August 2013 (Cohn et al., 2013) and the performance evaluated in a peer-reviewed Water Resources Research paper in April 2016 (Lamontagne et al., 2016).

References

[redacted] 2013, A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series: Water Resources Research, Vol. 49, 5047-5058, August 2013.

[redacted] 2016, Robust flood frequency analysis: Performance of EMA with multiple Grubbs-Beck outlier tests: Water Resources Research, Vol. 52, 3068-3084, April 2016.