

Evaluation of Recommended Revisions to Bulletin 17B *

Timothy A. Cohn¹ and Nancy A. Barth² John F. England, Jr.³ Beth
Faber⁴ Robert R. Mason¹ Jerry R. Stedinger⁵

¹U.S. Geological Survey, Reston, VA 20192

²U.S. Geological Survey, Sacramento, CA, 95818

³Bureau of Reclamation, Denver, CO, 80225

⁴U.S. Army Corps of Engineers, Davis, CA, 95616

⁵Cornell University, Ithaca, NY, 14853

DRAFT 3.0: 3 May 2015

Edits by Julie E. Kiang, December 2015

compiled January 20, 2016

Unpublished Draft: Not for Citation

*The work in this report was conceived by the Hydrologic Frequency Analysis Work Group (HFAWG) under the Subcommittee on Hydrology (SOH), Advisory Committee on Water Data (ACWI). The HFAWG is currently chaired by Wilbert Thomas. A data subgroup and a testing subgroup were formed under HFAWG to examine potential changes to Bulletin 17B. The data subgroup consisted of Wilbert Thomas, Martin Becker, Don Woodward, Beth Faber, and Jerry Coffey. The HFAWG data subgroup provided a set of peak-flow data sets and suggested tests to perform in comparing Bulletin 17B and EMA.

Contents

1	Introduction	12
2	Literature Sources: The History of Flooding and Flood Risk Estimation	16
3	Metrics for Evaluating Flood-Frequency Estimators	17
4	Estimation	20
4.1	Parameter Estimation	20
4.2	Identification of Outliers and Potentially Influential Low Floods	21
4.3	The MGB Test Algorithm for Identifying PILFs	23
4.4	Would Identification of Fewer PILFs Yield Better Fits?	25
4.5	Treatment of Floods Identified as Outliers: CPA and EMA	29
4.6	Estimation Methods Considered in this Report	32
4.7	Software	34
5	Comparisons of Methods	35
5.1	Understanding the Graphical Presentation of Results	35
5.2	Studies with LP3 Distribution	39
5.3	Studies with LP3 Distribution and Regional Skew	41
5.4	Robustness Studies	45
5.4.1	Robustness with Respect to Pearson Type 3 Population	45
5.4.2	Robustness with Respect to Mixed Population Constructed from Two LP3 Distributions	46
5.4.3	Robustness with Respect to Population Constructed from Two LP3 Distributions	46
5.4.4	Robustness with Respect to Population Constructed from Two GEV Distributions	49
6	Examples Based on Real Data at Selected Test Sites	53
6.1	82 Example Test Sites	53
6.2	Sites with Systematic Gage Data and No Low Outliers or Historical Information	54
6.3	Sites with Historical Information	57
6.4	Sites with Systematic Gage Data and Low Outliers	58
6.5	Sites with Low Outliers, Historical and/or High Outliers	67
6.6	Resampling Studies	73
6.7	Summary	73

7	Conclusions	81
A	Appendix: Characteristics of 82 Test Sites	83
B	Appendix: Graphical Comparisons Between EMA and B17B at the 82 Test Sites	96
B.1	Systematic Data Sites	96
B.2	Sites with Historical Information	123
B.3	Sites with Low Outliers	143
B.4	Sites with a Combination of Data Types	164
B.5	Studies with LP3 Distribution and Regional Skew	182
B.6	Additional Studies with Specific Frequency Curves	186

List of Figures

1	One-Day Peak Floods from 1932-2008 Measured at Sacramento River at Shasta Dam (Cohn et al., 2013, Figure 3) . .	31
2	Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a Log-Pearson Type 3 distribution with skew $\gamma = 0.0$	40
3	Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a Log-Pearson Type 3 distribution with skew $\gamma = -0.5$	42
4	Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a Log-Pearson Type 3 distribution with skew $\gamma = 0.5$	43
5	Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from robustness test curve 3.	47
6	Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a mixed population based on robustness test curve 4.	48
7	Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from robustness test curve 5.	50
8	Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from robustness test curve 6.	52
9	Relative Percent Difference (RPD) for B17B/GB and EMA/MGB estimators for 10%, 1%, and 0.2% exceedance probabilities. Includes 26 sites without historical flood information where no low outliers were identified by Grubbs-Beck tests.	55

10	Relative Percent Difference (RPD) for B17B/GB and EMA/MGB estimators for 10%, 1%, and 0.2% exceedance probabilities. Figure represents 19 sites with historical information where no low outliers were identified by Grubbs-Beck tests.	59
11	Pryor Creek near Billings, MT (06216500)	60
12	Relative Percent Difference (RPD) for B17B/GB, B17B/MGB, and EMA/MGB estimators for 10%, 1%, and 0.2% exceedance probabilities. Figure represents 20 sites where low outliers were identified by Grubbs-Beck(GB) or generalized Grubbs-Beck (MGB), and no historical information.	62
13	Number of floods identified as low outliers or below gage-base using the standard Grubbs-Beck (GB) and generalized Grubbs-Beck (MGB) tests. Figure represents 20 sites without historical information.	63
14	Percent low outliers identified using the generalized Grubbs-Beck (MGB) test. Figure represents 20 sites without historical information.	64
15	Orestimba Creek near Newman, CA (11274500), fit after application of the multiple Grubbs-Beck (MGB) test for low outliers.	66
16	Santa Cruz River near Lochiel, AZ (09480000)	68
17	Relative Percent Difference (RPD) for B17B/GB, B17B/MGB, and EMA/MGB estimators for 10%, 1%, and 0.2% exceedance probabilities. Represents 17 sites with historical flood information where low outliers were identified by the Grubbs-Beck (GB) or multiple Grubbs-Beck (MGB) test.	69
18	Arroyo Mocho near Livermore, CA (11176000)	70
19	Ten Mile Creek near Rimini, MT (06062500)	72
20	Results based on resampled data, $N_S = 40$ and $N_H = 100$, drawn from observed discharges at “Historical” category site 03011020.	74
21	Results based on resampled data, $N_S = 40$ and $N_H = 100$, drawn from observed discharges at “Low Outlier” category site 11152000.	75
22	Results based on resampled data, $N_S = 40$ and $N_H = 100$, drawn from observed discharges at “Gage Only” category site 14048000.	76
23	Results based on resampled data, $N_S = 40$ and $N_H = 100$, drawn from observed discharges at “Low Outlier” category site 14321000.	77

24	Results based on resampled data drawn from observed discharges at “Combination” category site 13185000.	78
25	Site 01076500 with Systematic Data Only	97
26	Site 01439500 with Systematic Data Only	98
27	Site 01555500 with Systematic Data Only	99
28	Site 01635500 with Systematic Data Only	100
29	Site 02037500 with Systematic Data Only	101
30	Site 02256500 with Systematic Data Only	102
31	Site 03183500 with Systematic Data Only	103
32	Site 05586500 with Systematic Data Only	104
33	Site 06406000 with Systematic Data Only	105
34	Site 06710500 with Systematic Data Only	106
35	Site 07208500 with Systematic Data Only	107
36	Site 07382000 with Systematic Data Only	108
37	Site 08380500 with Systematic Data Only	109
38	Site 08387000 with Systematic Data Only	110
39	Site 10128500 with Systematic Data Only	111
40	Site 11266500 with Systematic Data Only	112
41	Site 12134500 with Systematic Data Only	113
42	Site 12414500 with Systematic Data Only	114
43	Site 12437950 with Systematic Data Only	115
44	Site 12451000 with Systematic Data Only	116
45	Site 14021000 with Systematic Data Only	117
46	Site 14048000 with Systematic Data Only	118
47	Site 14137000 with Systematic Data Only	119
48	Site 15072000 with Systematic Data Only	120
49	Site 16518000 with Systematic Data Only	121
50	Site 16587000 with Systematic Data Only	122
51	Site 01350000 with Systematic and Historical Data	124
52	Site 01562000 with Systematic and Historical Data	125
53	Site 01636500 with Systematic and Historical Data	126
54	Site 02138500 with Systematic and Historical Data	127
55	Site 03011020 with Systematic and Historical Data	128
56	Site 03051000 with Systematic and Historical Data	129
57	Site 03159500 with Systematic and Historical Data	130
58	Site 03550000 with Systematic and Historical Data	131
59	Site 03558000 with Systematic and Historical Data	132
60	Site 03606500 with Systematic and Historical Data	133
61	Site 04293500 with Systematic and Historical Data	134
62	Site 06216500 with Systematic and Historical Data	135

63	Site 06600500 with Systematic and Historical Data	136
64	Site 06898000 with Systematic and Historical Data	137
65	Site 07067000 with Systematic and Historical Data	138
66	Site 08167000 with Systematic and Historical Data	139
67	Site 08378500 with Systematic and Historical Data	140
68	Site 09482500 with Systematic and Historical Data	141
69	Site 12413000 with Systematic and Historical Data	142
70	Site 01668000 with Low Outliers; no historical information . .	144
71	Site 03345500 with Low Outliers; no historical information . .	145
72	Site 05572000 with Low Outliers; no historical information . .	146
73	Site 06176500 with Low Outliers; no historical information . .	147
74	Site 07203000 with Low Outliers; no historical information . .	148
75	Site 08133500 with Low Outliers; no historical information . .	149
76	Site 08150000 with Low Outliers; no historical information . .	150
77	Site 08189500 with Low Outliers; no historical information . .	151
78	Site 09241000 with Low Outliers; no historical information . .	152
79	Site 09480000 with Low Outliers; no historical information . .	153
80	Site 10234500 with Low Outliers; no historical information . .	154
81	Site 11028500 with Low Outliers; no historical information . .	155
82	Site 11152000 with Low Outliers; no historical information . .	156
83	Site 11274500 with Low Outliers; no historical information . .	157
84	Site 11383500 with Low Outliers; no historical information . .	158
85	Site 12307500 with Low Outliers; no historical information . .	159
86	Site 13302500 with Low Outliers; no historical information . .	160
87	Site 13343660 with Low Outliers; no historical information . .	161
88	Site 14321000 with Low Outliers; no historical information . .	162
89	Site 16068000 with Low Outliers; no historical information . .	163
90	Site 03289500 with a Combination of Low Outliers, Historical and/or High Outliers	165
91	Site 05270500 with a Combination of Low Outliers, Historical and/or High Outliers	166
92	Site 05291000 with a Combination of Low Outliers, Historical and/or High Outliers	167
93	Site 05464500 with a Combination of Low Outliers, Historical and/or High Outliers	168
94	Site 06062500 with a Combination of Low Outliers, Historical and/or High Outliers	169
95	Site 06897000 with a Combination of Low Outliers, Historical and/or High Outliers	170

96	Site 06933500 with a Combination of Low Outliers, Historical and/or High Outliers	171
97	Site 07138600 with a Combination of Low Outliers, Historical and/or High Outliers	172
98	Site 08164000 with a Combination of Low Outliers, Historical and/or High Outliers	173
99	Site 08171000 with a Combination of Low Outliers, Historical and/or High Outliers	174
100	Site 09361500 with a Combination of Low Outliers, Historical and/or High Outliers	175
101	Site 09471000 with a Combination of Low Outliers, Historical and/or High Outliers	176
102	Site 11176000 with a Combination of Low Outliers, Historical and/or High Outliers	177
103	Site 11464500 with a Combination of Low Outliers, Historical and/or High Outliers	178
104	Site 11522500 with a Combination of Low Outliers, Historical and/or High Outliers	179
105	Site 12039500 with a Combination of Low Outliers, Historical and/or High Outliers	180
106	Site 13185000 with a Combination of Low Outliers, Historical and/or High Outliers	181
107	Results are based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a Log-Pearson Type 3 distribution with skew $\gamma = 0.0$ Regional skew is assumed to be 0.0 with $MSE = 0.15$	183
108	Results are based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a Log-Pearson Type 3 distribution with skew $\gamma = -0.5$. Regional skew is assumed to be -0.5 with $MSE = 0.15$	184
109	Results are based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a Log-Pearson Type 3 distribution with skew $\gamma = 0.05$. Regional skew is assumed to be 0.5 with $MSE = 0.15$	185
110	Results are based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from robustness test curve 1	187
111	Results are based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from robustness test curve 2	188

List of Tables

1	Impact of restricting search for PILFs to bottom quartile of sample at the 13 sites where more than 25% of the observations had been identified as PILFs using the standard Multiple Grubbs-Beck test. The change in the 1%-exceedance event appears in the right-most column, and ranged from $-20\%(0.20)$ to $+58\%(0.58)$, with a median increase of $+6\%$ and average increase of $+9\%$. Here $\hat{Q}_{0.01,[50]}$, $N_{[50]}$ refer to the estimated 1%-exceedance flood with the 50% limit and the corresponding number of identified PILFs; $\hat{Q}_{0.01,[25]}$ and $N_{[25]}$ refer to the corresponding statistics for the 25% limit.	27
2	Characteristics of at-site streamflow records used in comparing Bulletin 17B and EMA estimators	84

Executive Summary

For the past 35 years, Bulletin 17B has guided flood-frequency analyses in the United States. During this period much has been learned about both hydrology and statistical methods. In keeping with the tradition of periodically updating the Bulletin 17B Guidelines in light of advances in our understanding and methods, the Hydrologic Frequency Analysis Work Group (HFAWG) was charged by the Subcommittee on Hydrology (SOH) of the Advisory Committee on Water Information (ACWI) to consider possible updates to Bulletin 17B.

The purpose of this report is to consider the statistical performance of possible revisions to Bulletin 17B procedures. Of particular interest are procedures designed to accommodate more general forms of flood information. The concern is how the proposed procedures would affect the precision, accuracy and robustness of flood-frequency estimates.

The investigations reported here focus on techniques for:

- Incorporating information related to historical flooding that occurred outside the period of systematic streamgaging;
- Identification of potentially influential low floods (PILFs); and

The proposed changes, which mostly involve generalizing Bulletin 17B's method-of-moments procedures by using the Expected Moments Algorithm (EMA), are relatively modest, at least in the sense that they would not affect the main features of Bulletin 17B. The proposed methods include:

- Continued use of the log-Pearson Type III (LP3) distribution;
- Continued use of the Method-of-Moments fitting method applied to the logarithms of annual-peak-flow data; and

- Identification of low outliers and PILFs using a generalized Grubbs-Beck-type criterion that is sensitive to multiple potentially influential low flow (“PILF”; such low floods were previously denoted “low outliers”).

The hydrological literature already provides extensive support for the theory behind the proposed changes. The remaining question is practical: How well do the proposed methods perform under typical and realistic conditions and specifically with difficult records occasionally encountered in practice?

In order to answer these questions, the HFAWG commissioned the work reported here. Four major sets of results are provided:

- Monte Carlo simulations of fitting procedures employing data drawn from simulated LP3 populations;
- Monte Carlo simulations of fitting procedures employing data drawn from non-LP3 populations that were selected to reflect likely deviations of flood series, based on the experience of HFAWG members, from LP3 distributions;
- A direct comparison of the fitted LP3 distributions for 82 real “test sites” identified by an independent Data Group as both “typical” and “challenging” for flood frequency estimation; and
- Simulations of fitting procedures using records obtained by re-sampling with replacement from the longest of the 82 test-site records.

Collectively these studies provide a reasonably comprehensive, valid and robust assessment of the properties of the Bulletin 17B methods and proposed alternatives.

The experiments and analysis indicate that the flood quantile estimators proposed as a revision of Bulletin 17B:

- Perform generally as well as, and in some cases much better than, Bulletin 17B estimators in terms of the Mean Square Error (MSE) of flood quantiles estimates;
- Allow for incorporation and efficient statistical treatment of broader classes of flood-frequency data and information, including historical information, binomial data and interval data; and
- Generally confirm studies and the theoretical findings reported in the hydrological literature that would support use of updated estimation procedures that have been developed since Bulletin 17B was published.

1 Introduction

Flooding is the costliest natural hazard facing the United States in terms of loss of both life and property (Mileti, 1999; ACWI, 2011). While policy makers and planners have an array of tools to reduce flood losses through structural and non-structural mitigation, doing so in practice requires a quantitative, uniform and consistent approach for estimating flood risk (Tasker and Thomas, 1978; Thomas, 1985; Griffis and Stedinger, 2007). Recognizing this, the federal government has developed standard guidelines for performing flood-frequency analyses, and published these guidelines in a document known as “Bulletin 17B” (IACWD, 1982). The methods defined in Bulletin 17B inform literally millions of decisions about land use and construction, emergency response and recovery, and countless other governmental and private-sector activities in the United States. The practical value of federal guidelines is not in dispute.

Nonetheless Bulletin 17B (IACWD, 1982, “B17B”) has its limitations. The document itself recognized that additional work was needed to improve some of its procedures (IACWD, 1982, p. 27, “Future Studies”). In light of additional research conducted to address these concerns, both researchers and practicing hydrologists have recently called for updates to the B17B Guidelines. The purpose of this report is to evaluate some proposed modifications to the flood-frequency methods specified in Bulletin 17B.

In November of 2005, the HFAWG proposed the following (ACWI, 2011, minutes from February 2005 HFAWG meeting):

Based on recently completed research, the HFAWG proposes

to investigate the following possible improvements in Bulletin 17B:

1. Evaluate and compare the performance of the Expected Moments Algorithm (EMA) (Cohn et al., 1997) to the weighted-moments approach of Bulletin 17B (Appendix 6) for analyzing data sets with historic information and paleoflood data.
 - Apply EMA and Bulletin 17B to gaging station data that include low and high outliers and historic data and those that do not. Develop criteria for determining if EMA provides more accurate and consistent flood estimates.
 - Review and evaluate the published literature for comparisons of EMA to conventional Bulletin 17B procedures.
 - Recommend improved plotting position formula when historic data are available.
2. Evaluate and compare the performance of EMA to the conditional probability adjustment of Bulletin 17B for analyzing data sets with low outliers and zero flows.
 - Apply EMA and Bulletin 17B to gaging station data the include low and high outliers and historic data and those that do not (same data set as noted above). Develop criteria for determining if EMA provides more

accurate and consistent flood estimates.

3. Describe improved procedures for estimating generalized/regional skew.
 - Evaluate revisions needed in Bulletin 17B to describe improved procedures for estimating generalized/regional skew based on recently completed research.
4. Describe improved procedures for defining confidence limits.
 - Evaluate revisions needed in Bulletin 17B to describe new procedures for defining confidence limits that include the uncertainty in the skew coefficient.
 - Describe confidence limit procedures for EMA (if adopted).

This report presents results of analyses that focus on items 1 and 2. Items 3 and 4 are addressed in previous literature. Advances in regional skew analyses are describes in a series of USGS reports Lamontagne et al. (2012); Veilleux and Stedinger (2010); Veilleux et al. (2012); Confidence interval issues have been discussed in (Cohn et al., 2001; Cohn, in prep, 2015) (also see Cohn et al. (2013)) and earlier in Chowdhury and Stedinger (1991). More recent improvements in computational algorithms for computing confidence intervals based on EMA flood frequency analyses are the topic of a publication in press (Cohn 2015). Note that none of these proposed changes addresses issues that arise when fitting flood frequency curves to non-stationary conditions that may arise due to land use change, changes to water management practices, or climate variability or change. This is left

to future work.

The report is organized in the following manner. Chapter 2 provides a brief review of literature on the topic addressed herein and, through references, a more general review of the larger field of flood frequency analysis. Chapter 3 discusses criteria for judging estimator performance including statistical, practical and operational concerns. Chapter 4 reviews the statistical methods that are considered in this report, primarily the existing Bulletin 17B recommendations and the proposed Expected Moment Algorithm (EMA). Chapter 5 presents a set of Monte Carlo results based on synthetic samples from a variety of assumed and alternative distributional families, as well as resampling at long-record sites. Chapter 6 reviews the performance of the Bulletin 17B and EMA procedures at 82 real sites that were selected to represent conditions that exist at sites throughout the United States. Chapter 7 presents conclusions, a summary of the main findings of the research as reported in the earlier chapters. Two appendices are included that provide meta-data on the real-world sites employed in this study and additional figures illustrating estimator performance.

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

Bulletin 17B (IACWD, 1982) defines the current method for conducting peak-flow frequency analyses in the U.S. It descends from a series of studies and guidelines beginning with Bulletins 13 (ICOWR, 1966), 15 (USWRC, 1967), 17 (USWRC, 1976), and 17A (USWRC, 1977). Each of these includes a list of relevant citations.

The history of the Bulletins and a discussion of the methods recommended in the current Bulletin 17B appears in Thomas (1985), Stedinger et al. (1993) and Griffis and Stedinger (2007). There is also a vast literature on methods for flood-frequency estimation, and an extensive list of citations can be found in Stedinger et al. (1993), Griffis and Stedinger (2007), and Dawdy et al. (2012). American Institutes for Research (AIR, 2001) provides a comprehensive chronology of the history of flood risk estimation as well as a comprehensive bibliography of the flood-frequency literature prior to 2000. Stedinger et al. (1993) provides a bibliography of statistical techniques employed in flood frequency analyses.

3 Metrics for Evaluating Flood-Frequency Estimators

In practice, the benefits of more accurate flood information depend on many factors that are site-specific. In particular, benefits depend on the physical, geological, social, and other characteristics of the site. However, in trying to assess the performance of flood-risk estimation techniques, one typically simplifies the problem by considering a small number of generally accepted criteria that are believed to characterize adequately an estimator's performance.

The criteria employed in this study can be divided into several groups:

- *Operational*

1. *Ease of Application* Methods should be relatively easy to implement;
2. *Applicability to Available Data* Methods should be able to make efficient use of the data types available for flood investigations;
3. *Uniformity of Methods* Where possible, standardized methods and software should be used to ensure that different people performing the same analysis will obtain the same risk estimates. This will also ensure that frequency estimates are fully reproducible.

- *Statistical*

1. *Bias and consistency* On average, risk estimates should approx-

imately equal the true risk; as more information becomes available, risk estimates should converge to the true risk;

2. *Efficiency* Estimators should extract as much *information* from the *data* as possible so as to minimize the mean square error of estimated statistics;
3. *Quantified uncertainty* Estimates should be accompanied by quantitative assessments of their uncertainty;

- *Political, Legal and Institutional Criteria* Standard methods should be consistent, uniform, and easily explained. They should satisfy legal requirements, and should serve institutional requirements of the federal government and its National Flood Insurance Program, among others.

A set of metrics are employed in this report to quantify the differences among different parameter estimation procedures.

BIAS The expected difference between the estimate and true value of the parameter of interest.

MSE Mean Square Error (MSE) is defined as the expected squared difference between the estimate and true value of the parameter of interest. This is sometimes expressed as the sum of the variance plus the bias squared.

ERL The efficiency of estimators that employ historical flood information is quantified in terms of Effective Record Length (ERL), the amount of equivalent systematic record that, by itself, would provide the preci-

sion achieved with the combination of both systematic and historical information.

AG A more general measure of estimator efficiency is the Average Gain (AG), which expresses the benefit of each year of historical information in terms of the information in a year of systematic data.

RPD The Relative Percent Difference (RPD) statistic is used to quantify differences among estimators.

Taken as a whole, these metrics provide the quantitative foundation for characterizing the performance of Bulletin 17B flood frequency methods and proposed alternatives.

4 Estimation

For most of this study, flood quantile estimates were computed using two parameter estimation methods (B17B and EMA) and two PILF (Potentially Influential Low Floods) identification procedures (Grubbs-Beck (GB) and Multiple Grubbs-Beck (MGB)). Section 4.2 discusses the motivation of the GB procedure in Bulletin 17B and of a proposed replacement, MGB. The characteristics of the MGB method are discussed in section 4.3. Section 4.5 provides a discussion of the conditional probability adjustment (CPA) used by Bulletin 17B and the proposed EMA approach to analyze records containing zero flows, PILFs and other events described as being less than a specified threshold. Section 4.6 discusses three specific combinations of parameter estimators and PILF identification procedures considered in this report. Section 4.7 identifies the software used in the studies.

4.1 Parameter Estimation

Bulletin 17B (IACWD, 1982) recommends use of the method of moments to estimate the first three sample moments, although a number of additional procedures are employed for dealing with special situations. All of the procedures are documented in the Bulletin, and additional information on the procedures is found in Thomas (1985), Stedinger et al. (1993) and Griffis and Stedinger (2007).

The Expected Moments Algorithm (EMA) is a straightforward generalization of the method of moments that is designed to accommodate both the point data envisioned in Bulletin 17B (IACWD, 1982) as well as various

forms of non-standard interval data. The details of EMA are discussed in Cohn et al. (1997), Cohn et al. (2001) and Griffis et al. (2004b). At most sites across the United States, where only systematic annual peak flow (APF) data are currently available, the B17B and EMA frequency estimates are essentially identical. However, where non-standard data are present, the two estimates will differ. In some cases the differences can be substantial (see Section 6).

4.2 Identification of Outliers and Potentially Influential Low Floods

Among the proposed improvements to Bulletin 17B is a new multiple Grubbs-Beck (MGB) low-outlier test. MGB is a natural generalization of the Bulletin 17B Grubbs-Beck test, and addresses item 3 in the Bulletin 17B list of needed future work (IACWD, 1982): “The treatment of outliers both as to identification and computational procedures.”

PILF identification and treatment are important issues in flood frequency analysis guidelines because unusually small observations, which often result from physical processes that are of little relevance to the estimation of large floods, can nevertheless have a large influence on statistical estimates of extreme flood quantiles. In arid regions, for example, channel losses often result in annual flood peaks that are zero. As a result, an LP3 distribution cannot fit the entire flood record. Moreover, samples from an LP3 distribution with substantial negative log-space skew ($\gamma \leq -0.5$) typically contain so-called “low outliers,” and, using log-space moments, those unusually small values have undue influence and can result in poor estimates

of the large flood quantiles of interest in flood risk management. Unusually small values are therefore a real concern because it is imperative that the estimators defined in a Bulletin 17C, in addition to being efficient, possess the characteristic of *robustness* – meaning that they perform reasonably well even when underlying assumptions are violated.

Robustness is sometimes achieved in statistical analyses through explicit adaptation: The analyst looks for problems and then addresses the critical issues. In frequency analyses, for example, one often uses a probability plot to examine if sample data are consistent with a fitted curve (Stedinger et al., 1993). The authors of Bulletin 17B dealt with the robustness issue by explicitly identifying low outliers that “depart significantly from the trend of the remaining data.” They note that “The retention, modification and deletion of these outliers can significantly affect the statistical parameters computed from the data.” Thus the authors of Bulletin 17B explicitly recognized that we do not want small flood values distorting the model that describes the distribution of large floods (the “trend of the remaining data”). They explicitly note that failure to address this issue would “significantly affect the [computed] statistical parameters.”

Barnett and Lewis (1994, pp. 7-9) discuss this idea of using outlier tests to identify unusual observations that otherwise might have undue influence in an analysis. Low-outlier identification tests provide an objective and standard method for identifying outliers, ensuring that different hydrologists analyzing the same data will arrive at the same conclusions.

The purpose of using the Bulletin 17B GB test, and the MGB extension considered here, is to identify *Potentially Influential Low Floods*, or PILFs.

Although the low flows that are unusually different from the remaining data have previously been termed "low outliers", a more precise term is Potentially Influential Low Floods (PILFs). This report uses PILFs most often when describing the MGB extension, but, in keeping with previous terminology, uses "low outliers" when describing comparisons and test results. The defining characteristic of PILFs is that they potentially have a large influence on the upper tail of the fitted frequency curve. For example, when data sets are highly negatively skewed, the smallest observations can be very influential in determining the estimated skewness coefficient, and as a result the estimated value of the 1%-exceedance probability (100-year flood). In order to provide a robust and objective procedure, Bulletin 17B employs the Grubbs-Beck (GB) test to identify low outliers (PILFs).

4.3 The MGB Test Algorithm for Identifying PILFs

The Bulletin 17B Grubbs-Beck (GB) test provides an objective method for identifying values that should be treated as "low outliers," in the terminology employed in that report. However, the Bulletin 17B GB test "low-outlier" threshold is based on the assumption that only the smallest observation in the sample might be a low outlier. As a result, even though multiple low outliers in flood data are common, the GB test rarely identifies more than a single low outlier (Lamontagne et al., 2013). To provide an objective criterion for multiple low outlier identification, a multiple Grubbs-Beck test (MGB) was developed that employs the actual distribution of the i^{th} smallest observation in a sample of n independent LP3 variates with zero skew (Cohn et al., 2013).

Cohn et al. (2013) provides the probability $p(i;n)$ that the i^{th} smallest observation in a normal sample of size n might be as small or smaller than the value observed. If $p(i;n)$ is small enough, then the i^{th} smallest observation is considered unusual, and all values equal to and less than the i^{th} smallest observation are re-coded as censored values. Rosner (1975) explains the advantages of this generalization of the Grubbs-Beck statistic in the context of an analogous two-sided multiple outlier test. The specialized test adopted here considers only outliers in the lower tail. Spencer and McCuen (1996) and McCuen (2002) also discuss outlier tests based upon a generalized Grubbs-Beck statistic.

The procedure for actually identifying Potentially Influential Low Floods (PILFs) has two steps:

1. Starting at the median and sweeping *outward* towards the smallest observation, each observation is tested and is identified as a PILF if $p(i;n) \leq 0.5\%$. If the k^{th} observation is identified as a PILF, the outward sweep stops and the k^{th} and all smaller observations are also identified as PILFs.
2. Then, as with the current Grubbs-Beck procedure in Bulletin 17B, an *inward* sweep starts at the $(k + 1)^{st}$ smallest observation, where the observation is identified as a PILF if $p(k+1;n) \leq 10\%$. This is repeated, sweeping towards the median, until an observation m fails to be identified as a PILF by the inward sweep, at which point the inward sweep stops.

The number of PILFs identified by the procedure is then the larger of k and

$m - 1$.

Bulletin 17B also used a 10% significance test but employed only a single *inward* sweep. A second critical difference is that the new inward sweep uses the $p(i;n)$ function which correctly describes whether the i^{th} largest observation is unusual in a sample of size n .

The MGB's first outward sweep seeks to determine if there is some break in the lower half of the data that would suggest the sample is best treated as if it had a number of PILFs. The second sweep using a less severe significance level, $p(i;n) \leq 10\%$, mimics Bulletin 17B's willingness to identify one or more of the smallest observations as PILFs so that the analysis is more robust.

If a record has, for example, 5 zero flows, then the smallest non-zero flow is considered to be the 6^{th} smallest observation in the record. This correctly reflects the fact that the flood record included 5 smaller values. The GB test in Bulletin 17B includes no mechanism for correcting its threshold when testing the smallest non-zero flood value in a record containing one or more zeros, or below-threshold discharges at sites with crest-stage gages. This is particularly problematic because sites with zero flows are likely to include one or more small or near-zero flood values which should be identified as PILFs. The MGB test solves this problem.

4.4 Would Identification of Fewer PILFs Yield Better Fits?

The MGBT was defined to search for PILFs – *potentially* influential low floods – in the bottom 50% of the sample. Citing various reasons, some practicing hydrologists have expressed concern about declaring half of a

sample to be “low outliers” or PILFs. In part, this reflects concern about the number of “low outliers” but the focus should be on achieving robustness in estimating large floods, not to utilize as many point observations as possible. In any case, one can determine the consequence of limiting the number of identified PILFs to, for example, one quarter of the sample rather than half. This question is addressed here through consideration of the 82 test sites. One immediate observation is that the question is relevant to only a handful of cases; the MGB test identifies fewer than 25% PILFs at 69 of the 82 test sites. The 25% constraint would not be binding at these sites, and the results at these sites are not affected by limiting the number of PILFs to 25%.

Table 1 reports the impact the 25% limit at the remaining 13 test sites. At these sites, limiting the number of PILFs to 25% of the sample will affect estimates of the 1% exceedance event, which appears in the right-most column. The changes ranged from -20% to $+58\%$ with a median difference of $+6\%$ and average increase of $+9\%$.

On average, use of the lower quartile (25% limit) resulted in a modest increase in the 100-year flood estimate. The question then becomes essentially subjective: Are the fitted frequency curves “better” with the 25% limit or the 50% limit? In the three cases where the absolute value of the change exceeded 20%, the 100-year flood increased by 22%, 30% and 50%. If the MGBT has been censoring too heavily, we would expect to see no consistent increase or decrease from censoring less. To see an average increase of 9%, with such consistently positive increases when large differences occur, suggests that the 25% constraint leads to admitting small peaks that unduly

Table 1: Impact of restricting search for PILFs to bottom quartile of sample at the 13 sites where more than 25% of the observations had been identified as PILFs using the standard Multiple Grubbs-Beck test. The change in the 1%-exceedance event appears in the right-most column, and ranged from $-20\%(0.20)$ to $+58\%(0.58)$, with a median increase of $+6\%$ and average increase of $+9\%$. Here $\hat{Q}_{0.01,[50]}$, $N_{[50]}$ refer to the estimated 1%-exceedance flood with the 50% limit and the corresponding number of identified PILFs; $\hat{Q}_{0.01,[25]}$ and $N_{[25]}$ refer to the corresponding statistics for the 25% limit.

Site	USGS ID	$\hat{Q}_{0.01,[50]}$	$N_{[50]}$	$\hat{Q}_{0.01,[25]}$	$N_{[25]}$	$(\frac{Q_{0.01,[25]} - Q_{0.01,[50]}}{Q_{0.01,[50]}})$
17	03345500	49700	34	54100	1	0.09
23	05291000	16000	29	12800	20	-0.20
37	07138600	1110	12	983	5	-0.12
41	08133500	26800	23	34900	0	0.30
45	08171000	113000	27	110000	19	-0.02
50	09241000	4810	22	4550	17	-0.05
56	10234500	1110	45	1360	0	0.22
58	11152000	30800	46	32100	18	0.04
59	11176000	2440	19	3860	1	0.58
61	11274500	13600	29	14500	19	0.06
63	11464500	21600	15	25400	5	0.18
65	12039500	50400	47	56500	0	0.12
73	13302500	18000	42	18500	16	0.03

affect the 1%-exceedance flood estimate.

Appendix B includes showing flood magnitude plotted against Annual Exceedance Probability (AEP) for these three cases. Visual examination of the plots suggests that the algorithm that started its search at the median made reasonable decisions, as can be seen in the figures (in Appendix B):

08133500 MGBT [50] identifies 23 PILFS while MGBT[25] identifies none. Note the sharp break in the data just above the lower quartile that MGBT[50] catches and MGBT[25] would not (Figure 75).

11176000 Note a weak break in the data just above the lower quartile. Using the MGBT[50] identifies this discontinuity and results in a fit that is much more consistent with the 8 largest observations. B17B/GB with one outlier seems to overestimate the 1%-exceedance flood when compared with the data (Figure 102).

10234500 Note kink visible just below the median which MGBT[50] identifies, with the result that the MGBT recommends censoring almost half the data. Given the shape of the data, that seems reasonable. Which approach provided the best fit? With the more extreme censoring, less weight is placed on the smallest observations and the B17B/MGB and EMA/MGB fits are much more consistent with the 7 largest observations in the sample. In short, the MGB as defined seems to achieve the goal of matching the trend in the largest observations (Figure 80).

In 10 of 13 cases the differences were small. In three cases the differences were greater than 20% and *all* were positive. In at least two of the three

cases, better fits were obtained with the more aggressive censoring, and it is arguable that this is the case in all three. Thus MGBT[50] is arguably a good choice for identifying *potentially* influential low floods.

4.5 Treatment of Floods Identified as Outliers: CPA and EMA

The logarithm of zero is minus infinity, so zero flows cannot be employed in a standard method-of-moments fit using the logarithms of the flood magnitudes. As discussed here, the two estimation methods, B17B and EMA, deal with zeros and low outliers in different ways.

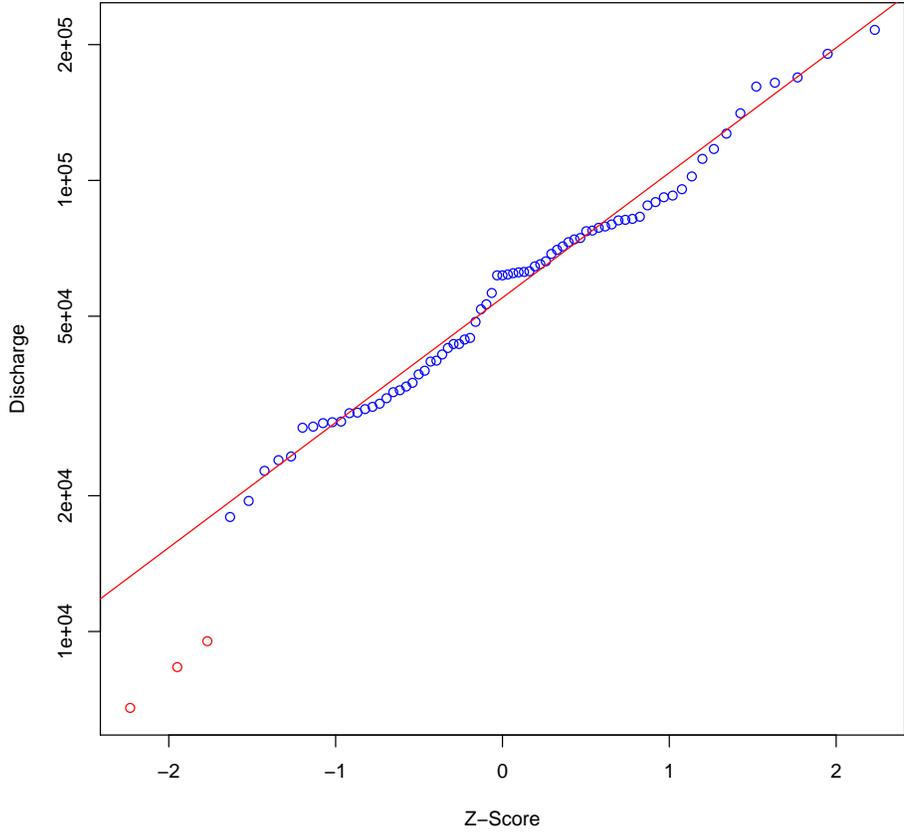
To deal with such low outliers, Bulletin 17B employs the Conditional Probability Adjustment (CPA) developed by Jennings and Benson (1969). If the r smallest floods in a record of length n are identified as outliers less than some threshold q_T , then the CPA fits a LP3 distribution to the retained $(n - r)$ large values. Then for any large flood value $q > q_T$, its non-exceedance probability is estimated as r/n plus $(1 - r/n)$ times the probability it would not be exceeded given the distribution fit to the $(n - r)$ retained observations. Then for convenience and computation of confidence intervals, the 50, 90, and 99 percentiles of that “probability-adjusted” distribution are computed and used to compute the three parameters of a “synthetic” LP3 distribution that is used to represent flood risk at the site. This procedure is relatively robust, and has been shown to work well with zeros and low outliers Griffis et al. (2004b).

EMA uses ranges and thresholds to describe historical flood information, floods measured with limited precision, and PILFs. In particular, PILFs

can be represented as being less than the smallest retained observation when the LP3 distribution parameters are estimated. This allows EMA to address a large range of circumstances, while appropriately representing all of the available information. Censoring several of the smallest observations in a sample results in almost no decrease of precision in flood quantile estimates with EMA or with CPA (Griffis et al., 2004b). However, EMA can be much more sensitive than CPA to PILFs that are not afforded special treatment. Suppose a record had three low outliers, and we identified two. Then using CPA, the value of the smallest two observations are completely ignored, and the frequency-adjustment (CPA) corrects for omitting two observation in the curve fitting step. Thus 66.7% of the low outliers have been dropped, greatly reducing their impact. With EMA, if there are 3 low outliers and two are identified, EMA represents the two smallest outliers as being less than the third. This can make things worse.

Consider, for example, the Sacramento River 1-day flood record discussed in Cohn et al. (2013, Figure 3) (reproduced here as Figure 1), which has three small values that are each at least a factor of two smaller than the other observations. The GBT identifies one low outlier in this case; MGBT correctly identifies all 3, which can be seen immediately to be the most reasonable decision. Because EMA in some sense takes the data seriously, one needs to identify all of the PILFs to reliably avoid potential problems. GB frequently fails to identify obvious outliers, other than the smallest flow. Given the well understood sensitivity of the EMA method to low outliers, no EMA/GB combination is considered in this report. EMA does not perform reliably when paired with GB. As will be shown, the MGB outlier criterion

Figure 1: One-Day Peak Floods from 1932-2008 Measured at Sacramento River at Shasta Dam (Cohn et al., 2013, Figure 3)



identifies obvious outliers and PILFs in the lower half of troublesome data sets. As a result, as will be shown here, both B17B and EMA when used with the MGB outlier identification procedure are reliable flood-quantile estimators in terms of their ability to avoid problems caused by unusually small peaks.

4.6 Estimation Methods Considered in this Report

Bulletin 17B (IACWD, 1982) procedures include a Grubbs-Beck (GB) outlier identification procedure, a weighted moment computation if a historical record period is included with a perception threshold that was exceeded at least once, use of a conditional probability adjustment (CPA) if a record contains zero flows, low outliers, or peaks below gage-base, and finally computation of a weighted skewness coefficient ((Griffis and Stedinger, 2007, figure 2); (IACWD, 1982, pages 12-2 through 12-3)). All four of those steps, as needed, are components of the flood frequency method denoted here as B17B.

The proposal under consideration is to replace those procedures with the multiple Grubbs-Beck (MGB) test for identification of low outliers (PILFs) (Cohn et al., 2013), and the Expected Moments Algorithm (EMA) for the estimation of the parameters of an LP3 distribution simultaneously considering zeros and low outliers as censored data, one or more periods with different perception thresholds for historical floods, and a regional skewness coefficient with specified mean square error (Cohn et al., 1997; Griffis et al., 2004b). An advantage of the EMA approach is that all of the data is treated consistently in a single parameter estimation step, rather than as

a sequence of procedures that first address historical information, then zeros and low outliers, and finally regional skew information (or the opposite ordering of low outlier and historical adjustments when the skew coefficient is less than -0.4). Uncertainty analysis of the B17B sequence of steps is more difficult than that of the unified EMA procedure, and EMA, because it allows simultaneous use of all of the data, is more efficient. For example, an informative regional skew should be used to inform the interpretation of historical and interval information, rather than being introduced as a separate and independent final adjustment of just the skewness coefficient as recommended by Bulletin 17B. The EMA algorithm also allows for different thresholds to describe historical information from different periods, and the use of interval estimates of flood peaks to represent measurement uncertainty, which is often appropriate when representing the magnitude of historical flood information.

The studies that follow consider three estimation combinations:

1. EMA (Cohn et al., 1997) with a generalized multiple Grubbs-Beck test (MGB) (Cohn et al., 2013) for detecting multiple potentially influential low outliers in a flood series (EMA/MGB);
2. B17B (IACWD, 1982) with the standard Grubbs-Beck (GB) method for identifying low outliers followed by the conditional probability adjustment (CPA) to address zeros, floods identified as low outliers, and peaks below gage-base (B17B/GB); and
3. B17B where the new MGB was employed to identify multiple potentially influential low floods followed by the conditional probability ad-

justment (CPA) (IACWD, 1982) to address zeros, floods identified as low outliers, and peaks below gage-base (denoted B17B/MGB).

4.7 Software

The USGS PEAKFQ version 7.0 program was used to create the comparisons between B17B and EMA estimates. PeakfqSA version 0.995, a development version of PEAKFQ written by Tim Cohn, was used to check the EMA results generated in PEAKFQ 7.0.

The Monte Carlo and resampling results reported in section 5 are based on the underlying code employed in PEAKFQ, with analysis routines prepared by Tim Cohn employing a front-end “driver” routine developed by Jery Stedinger called “monte.f”

Results based on the Bulletin 17B Guidelines are referred to as “B17B” or PEAKFQ/B17B. Results based on the proposed EMA methods are denoted “EMA” or PeakfqSA. Additional letters are appended, yielding names of the form “B17B/GB,” “B17B/MGB,” and “EMA/MGB,” to specify explicitly which low-outlier test procedure was employed.

5 Comparisons of Methods

Hundreds of distinct Monte Carlo and resampling experiments were conducted, each involving 10000 replicate samples, in order to determine how well the B17B/GB, B17B/MGB, and EMA/MGB estimators satisfy the criteria listed in section 3. These Monte Carlo experiments are described in this chapter.

5.1 Understanding the Graphical Presentation of Results

In order to simplify the discussion, comparisons between estimators for cases with different historical information are presented concisely using a single graphical format. Each figure considers the performance of the estimators when data are drawn from a specific population. Figure 2, for example, corresponds to an LP3 distribution with a log-skew $\gamma = 0.0$; Later graphs consider estimator performance when data are drawn from other populations.

Each graphic has two parts. The top panel contains 12 boxplots (Tukey, 1977), divided into four groups of three, showing the distribution of three estimators for $Q_{1\%}$, the 1%-exceedance event (i.e., the "100-year flood"). The central colored rectangle spans the 25th to 75th percentiles of the estimators' distributions. The central line indicates the median. Additional information about the distribution of the estimations is shown by the whiskers and hinges. The whisker length is 1.5 times the interquartile range. Observations beyond the whiskers are plotted as individual circles. For those cases where the population mean is known, an additional symbol is present on

each box, a circle with a cross, which indicates the location of the mean.

The three estimators are defined below and in IACWD (1982), Cohn et al. (1997), Griffis et al. (2004a) and Cohn et al. (2013). Although flood quantiles other than the 1% exceedance were investigated, the results were found to be insensitive to which quantile was considered. Because $Q_{1\%}$ is the flood of interest for many federal activities, this case is reported.

Each of the four groups corresponds to $N_S = 40$ years of systematic data. The first group involves no historical information ($E[H] = 0$). The other 3 have a $N_H = 100$ year historical period during which different perception thresholds are set so that, in expectation, 1, 2 or 10, historic floods would have been recorded. The red dashed line indicates the true value of $Q_{1\%}$, where it is known; in some later plots, where $Q_{1\%}$ is estimated by interpolation between real observations, a purple line is used to indicate an estimate of $Q_{1\%}$ because the true value of $Q_{1\%}$ is unknown.

The colored dots appearing in the lower panel in each graph indicates the effective record length (ERL) of each estimator. ERL is defined in this case, where $N_S = 40$ and $N_H = 100$, as the ratio:

$$\text{ERL} \equiv N_S \left\{ \frac{\text{MSE}[\log(\hat{Q}_{1\%}(N_S = 40, N_H = 0))]}{\text{MSE}[\log(\hat{Q}_{1\%}(N_S = 40, N_H = 100))]} \right\} \quad (1)$$

where N_S is the length of the systematic record, N_H is the length of the historical period, and

$$\begin{aligned} \text{MSE}[\log(\hat{Q}_{1\%}(N_S, N_H))] &\equiv (\text{Bias}[\log(\hat{Q}_{1\%}(N_S, N_H))])^2 + \\ &\quad \text{Var}[\log(\hat{Q}_{1\%}(N_S, N_H))] \end{aligned} \quad (2)$$

ERL expresses, in a rough way, how well an estimator exploits historical information by expressing its effect in terms of an equivalent number of years of exclusively systematic data. Note that the estimator variance, denoted Var in equation 2, is simply the variance of the 10000 replicate estimates of $\hat{Q}_{1\%}$. The Bias, however, is not always so easily defined. Where the flood distribution is known, the Bias is defined in the conventional manner as the average difference between $\hat{Q}_{1\%}$ and the true mean. When dealing with real data (e.g. in section 6), however, the true value $Q_{1\%}$ is not known. In these cases, for want of a better assumption, Bias is assumed to be zero for all estimators. Thus, the ERL is computed using only the estimated variance. The bottom left-most box contains three numbers in lavender. These are, respectively, the percentage of observations identified as low outliers (PILFs) by the Grubbs-Beck test with Bulletin 17B (B17B/GB), the multiple Grubbs-Beck test with Bulletin 17B (B17B/MGB), and the multiple Grubbs-Beck test with EMA (B17B/MGB). The number of low outliers with B17B/MGB and EMA/MGB are in fact always identical because the tests do not depend on the fitting procedure.

The triplets of black numbers in the three boxes to the right indicate the average gain (AG), in percent, associated with each year of historical flood information, for B17B/GB, B17B/MGB and EMA/MGB, respectively. The AG quantifies the relative value of an additional year of historical information to an additional year of systematic gage record, and is defined as:

$$AG \equiv 100 \left(\frac{ERL - N_S}{N_H} \right) \quad (3)$$

AG is the percentage increase of *ERL* for each year in the historical period. Where the expected number of historic floods is high – on the right side of the graph – the AG is typically also high. Where no historic floods can be expected – the group on the left – the average gain is zero. ¹. The three colored circles in the bottom box represent indicate the ERL corresponding to each estimator and censoring threshold.

¹Note: AG results are based on 10000 replicate samples, and have a corresponding uncertainty of about 1%

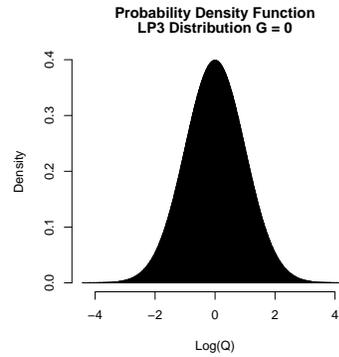
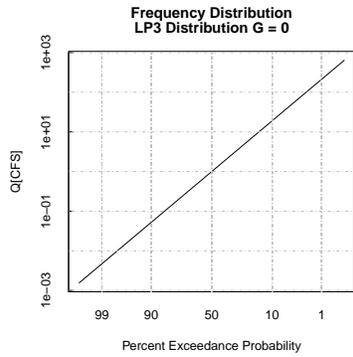
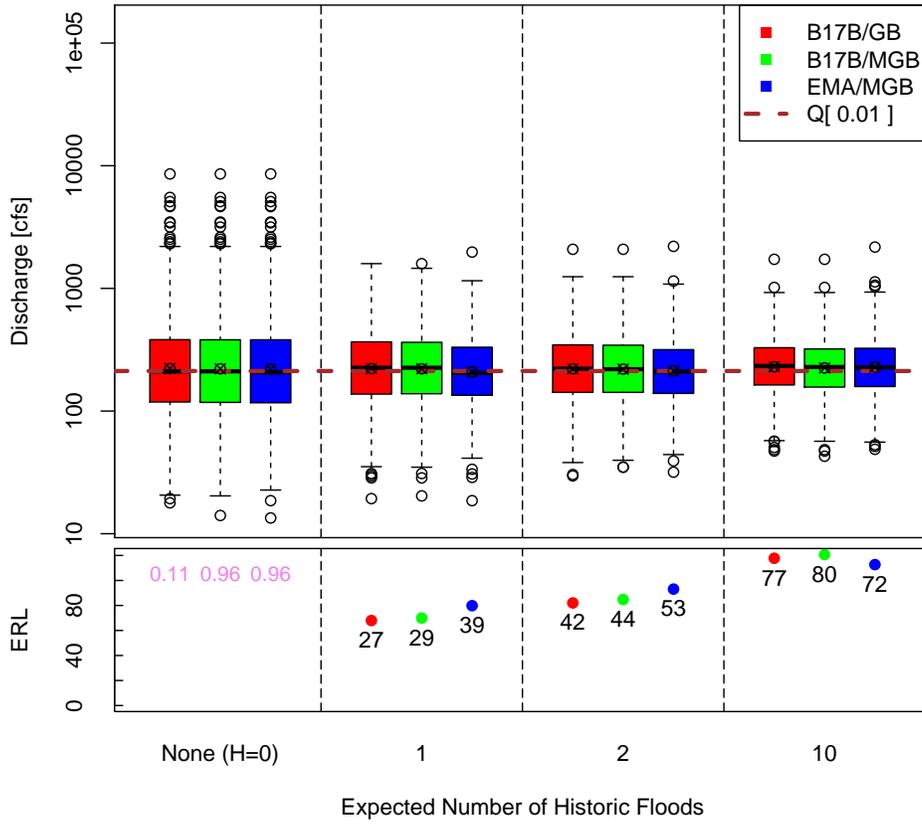
5.2 Studies with LP3 Distribution

Figures 2 through 4 correspond to the case where we are fitting the LP3 distribution to LP3 data assuming we have no regional information. In these cases we expect good results for all the estimators because we are applying the correct model for the population from which the data are drawn. The distribution of the Monte Carlo population, depicted as a frequency plot and as a probability density function of the logarithm of Q , appears below the main plot. Cases with regional skew information are addressed in Section 5.3.

Figure 2, depicts the case when the population skew, γ , is zero and low outliers are rare. The figure shows that the 3 estimators are identical when there is no historical information (the left-most three boxplots). If historical information is present with a perception threshold at approximately the 1% exceedance flood level (Stedinger and Cohn, 1986), the EMA/MGB (AG=39) method performs substantially better than B17B/GB (AG=27) or B17B/MGB (AG=29). The same conclusion applies when the threshold is at the 2% exceedance level, where the average gains are 42, 44, 53 for B17B/GB, B17B/MGB, and EMA/MGB, respectively. When the threshold is at the 10% exceedance level, which corresponds to a rare situation, all of the estimators perform extremely well, with average gains of 77, 80, 72.

Figure 3 depicts the case when the population skew is $\gamma = -0.5$. In this case many low outliers are to be expected. Figure 3 reveals several interesting properties of the estimators. First, the center of the boxplots are substantially above the hashed line (the true 1% exceedance level), indi-

Figure 2: Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a Log-Pearson Type 3 distribution with skew $\gamma = 0.0$.



cating that all of the estimators are biased upward when only systematic data are employed. This phenomenon is actually well known (Kirby, 1974; Stedinger et al., 1993): The method-of-moments estimator for the skew coefficient is biased toward zero, and thus method-of-moments quantile estimators are biased upwards for populations with negative skews and downwards for populations with positive skews (see figure 4). When historical information is present, the EMA/MGB method performs substantially better than B17B/GB or B17B/MGB.

Figure 4 depicts the case when the population skew is $\gamma = 0.5$ and no low outliers are to be expected. This case reveals the opposite bias seen in figure 3. However, in this case all of the estimators make good use of the historical information because, for positively skewed populations, the smaller values in the dataset have little impact on the sample moments and therefore do not influence the results regardless of which estimator is employed. In figures 2 and 4 with $H = 10$ expected floods, both B17B estimators do a little better than EMA, though visually the boxplots are indistinguishable. In all other cases EMA did essentially as well as the B17B estimators.

5.3 Studies with LP3 Distribution and Regional Skew

Figures 107 - 109 show the same cases as figures 2 - 4 except that regional skew information has been added. Synthetic regional skews are modeled and generated as normal variate with a mean of the observed skew and a variance of 0.15 – a typical value consistent with Bayesian/GLS skew maps (Lamontagne et al., 2012; Parrett et al., 2011; Gotvald et al., 2006). The addition of regional skew information improves all of the estimators,

Figure 3: Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a Log-Pearson Type 3 distribution with skew $\gamma = -0.5$.

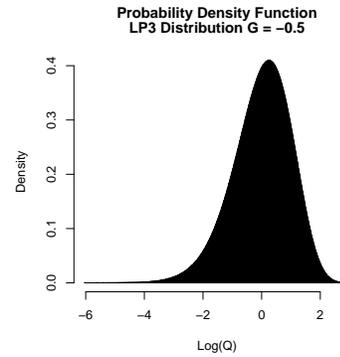
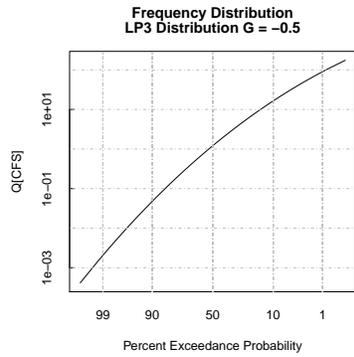
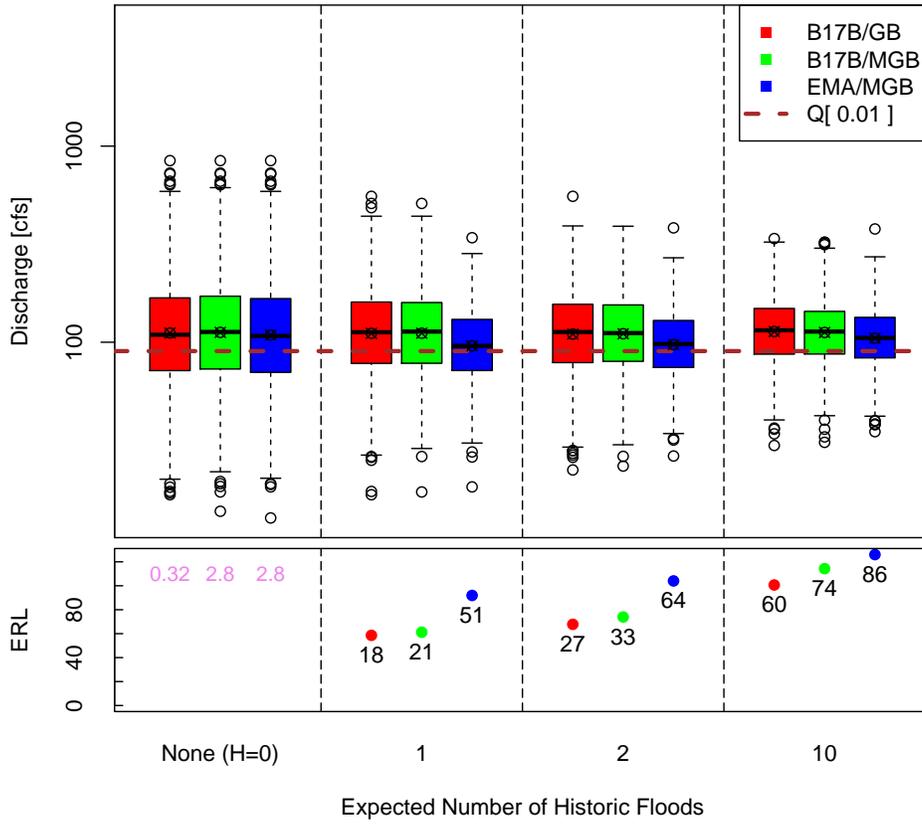
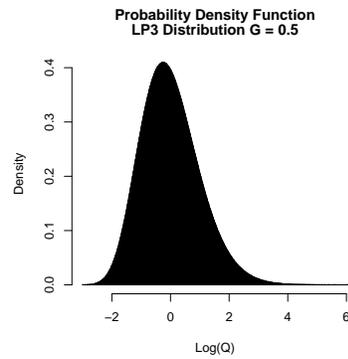
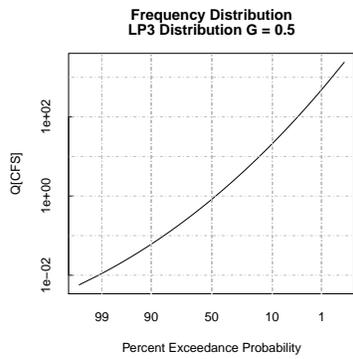
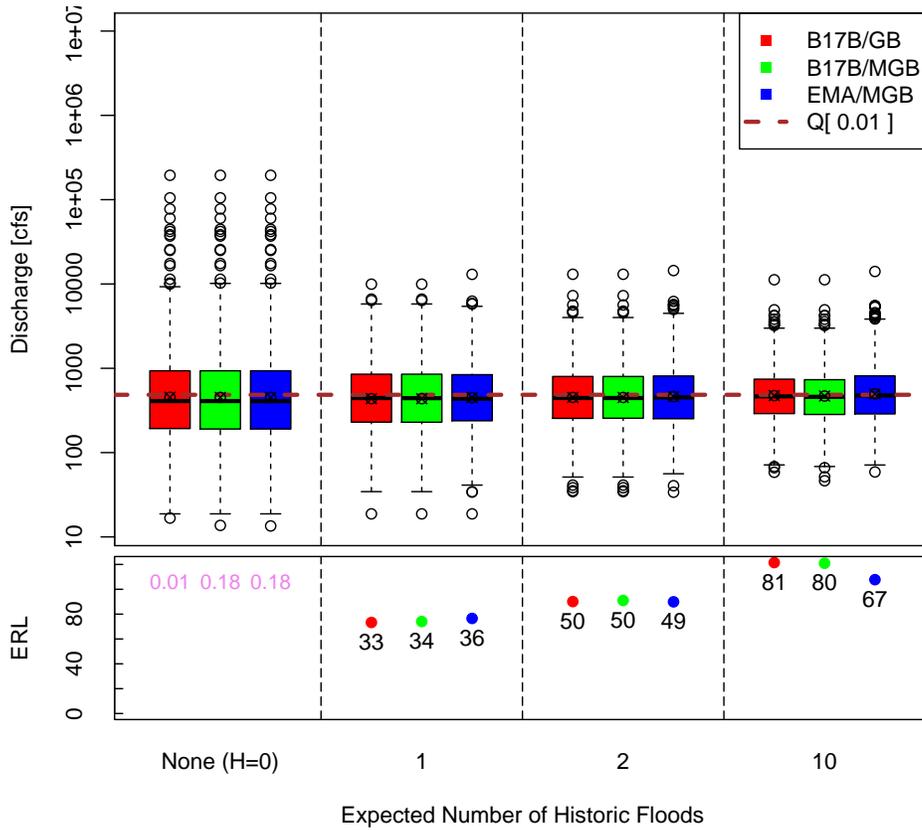


Figure 4: Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a Log-Pearson Type 3 distribution with skew $\gamma = 0.5$.



including EMA/MGB and B17B/GB. The impact is essentially the same for all of the estimators and most pronounced for the negative skew example. In conclusion, while accurate regional skew is a valuable addition to frequency analyses, it does not need to be considered as an important factor in determining the relative performance of the three estimators. For the cases with historical information where 1 or 2 historical floods are expected, again EMA provides more precise quantile estimators than the B17B estimators. For the case where 10 historical floods are expected and $\gamma = 0.5$, there is a virtual tie; when $\gamma = 0.5$ and $\gamma = 0$ (figures 107-108), B17B/MGB does a little better than EMA, while B17B/GB does slightly better only when $\gamma = 0$ (figure 107). However, it is noteworthy that the value of historical information, when carefully employed with EMA (see figure 2), ranges from an average gain of 39% when 1 historical flood is expected, to 72% when 10 historical floods are expected.

5.4 Robustness Studies

To test their robustness to data with different distributions, the three estimators, B17B/GB, B17B/MGB, and EMA/MGB, were applied to 6 test curves. These are referred to as “robustness test curves 1-6.” The first two of these test curves are based on LP3 data with different skew values, and are similar to those discussed in section 5.2. Figures 110, 111 in the appendix also depict these two test curves. The remaining 4 test curves are used to test the performance of the three estimators, B17B/GB, B17B/MGB, and EMA/MGB when applied to non-LP3 data.

There are an infinite number of distribution “curves” that could be used to test robustness. While these curves represent only a tiny portion of the non-LP3 universe, they were chosen because it is believed they reflect at least some of the non-LP3 populations that have been observed in practice. For example, mixed populations may arise due to the existence of multiple peak flow generating processes in a watershed.

5.4.1 Robustness with Respect to Pearson Type 3 Population

Figure 5 depicts the performance of the estimators when data are drawn from a Pearson Type 3 (P3) population, not the LP3 (robustness test curve 3). As can be seen in the figure, the P3 has a substantially different shape than the LP3. All of the estimators are biased when fitting this population, as is expected because it is not the assumed population. However, the EMA/MGB estimator does perform slightly better than the two other estimators when historical information is present. In the absence of historical

information, differences arise only due to outliers.

5.4.2 Robustness with Respect to Mixed Population Constructed from Two LP3 Distributions

Figure 6 depicts the performance of the estimators when data are drawn from robustness test curve 4, a mixed population created by choosing the maximum of observations drawn from two different LP3 distributions, the first with parameters

$$\{M, S^2, G\} = \{4.1212, 0.29^2, 1.00\}$$

and the second with parameters

$$\{M, S^2, G\} = \{4.0900, 0.13^2, 0.15\}$$

Without historical information, all three estimators are nearly identical. When historical information is present, the estimators are no longer identical with EMA/MGB performing slightly better.

5.4.3 Robustness with Respect to Population Constructed from Two LP3 Distributions

Figure 7 depicts the performance of the estimators when data are drawn from robustness test curve 5, a constructed population based on two LP3 distributions, the first with parameters

$$\{M, S^2, G\} = \{4.3438, 0.41^2, -1.00\}$$

Figure 5: Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from robustness test curve 3.

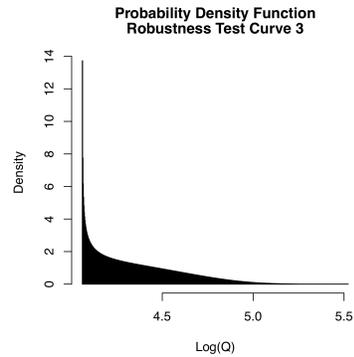
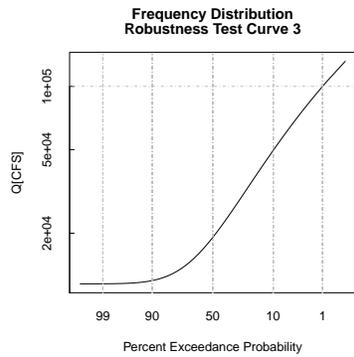
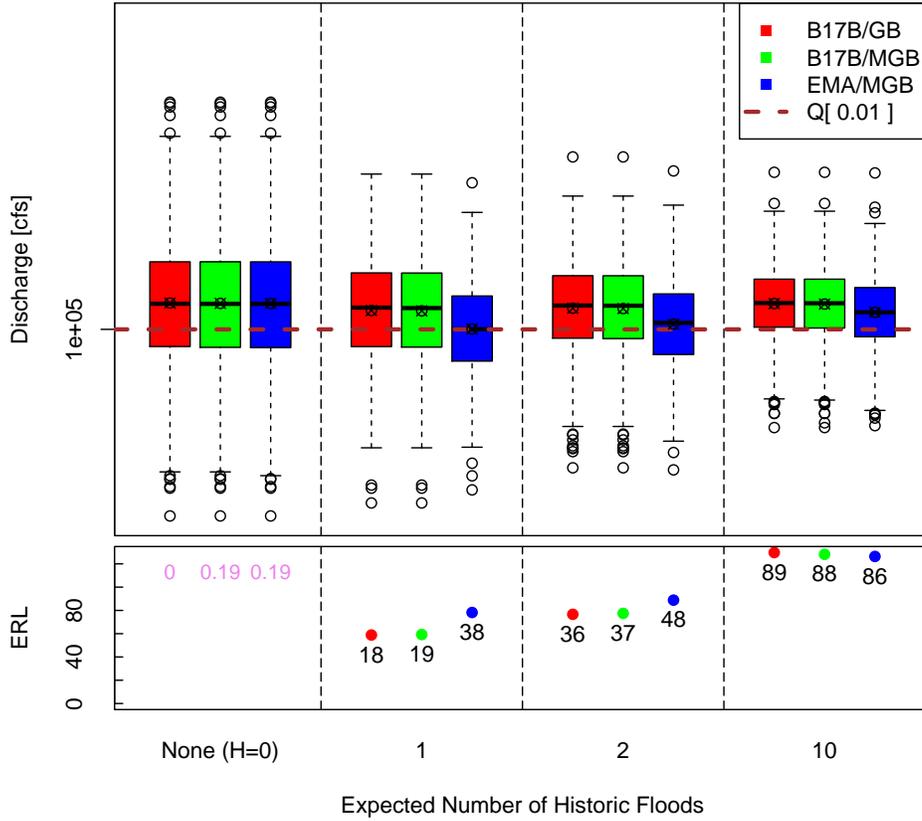
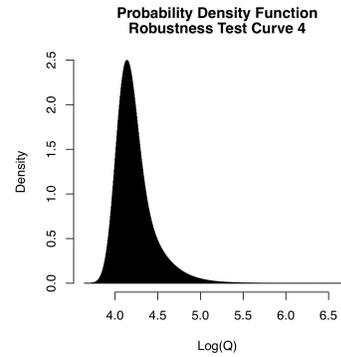
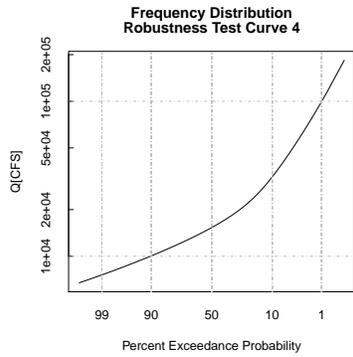
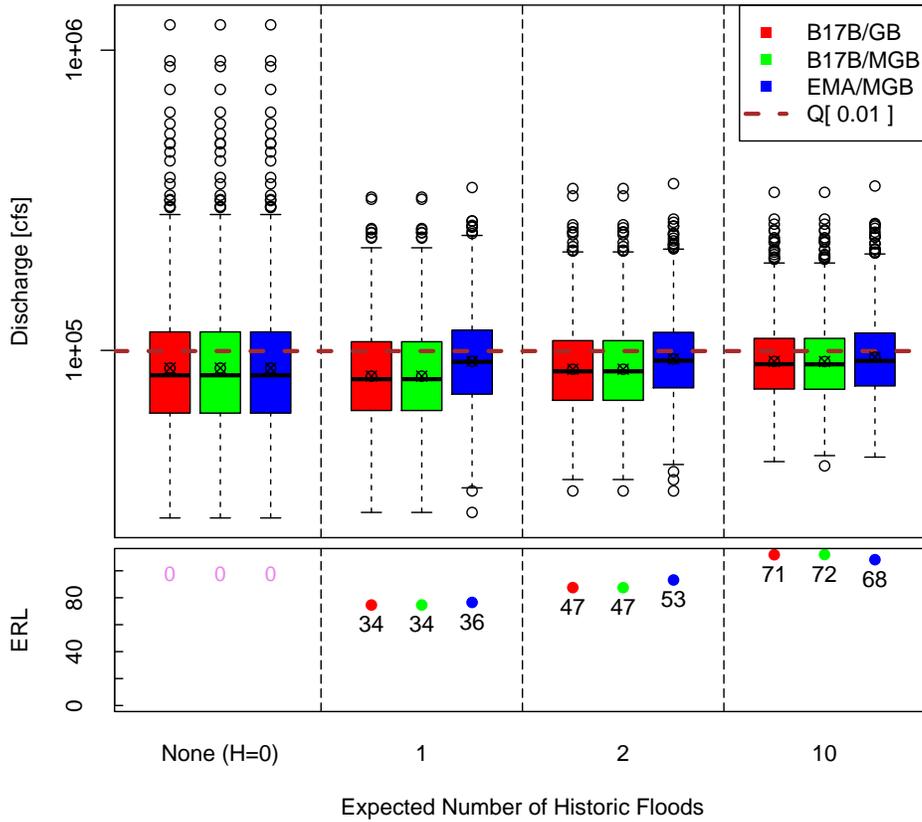


Figure 6: Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a mixed population based on robustness test curve 4.



and the second with parameters

$$\{M, S^2, G\} = \{4.3936, 0.50^2, -0.20\}$$

The lower half of the distribution function employs the first parameters, and the upper half is based on the second parameters. The two parent distributions are spliced together at their shared median. Fitting data from this population again tests the robustness of the estimators with a misspecified population, but this time when low outliers are present. Without historical information, EMA/MGB and B17B/MGB are nearly identical, and both perform better than B17B/GB. When historical information is present, EMA/MGB performs much better than the other estimators.

5.4.4 Robustness with Respect to Population Constructed from Two GEV Distributions

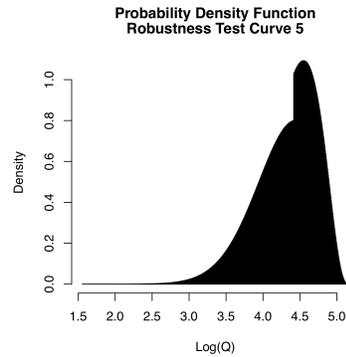
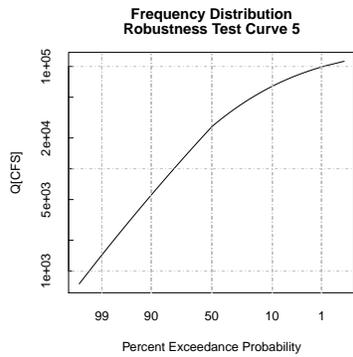
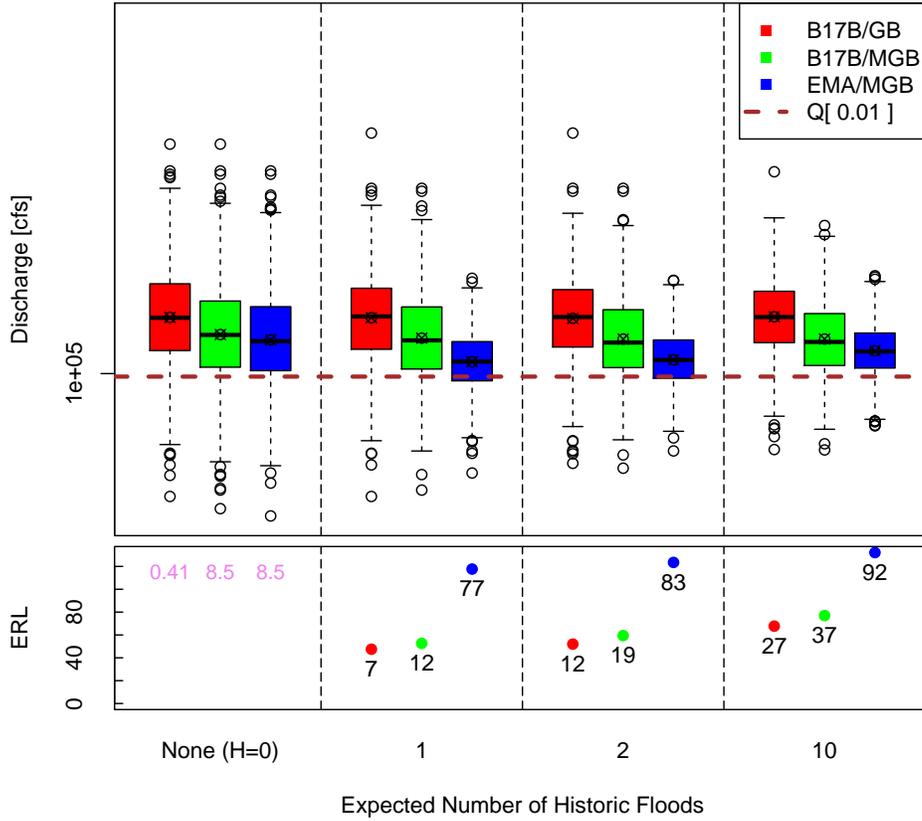
Figure 8 depicts the performance of the estimators when data are drawn from robustness test curve 6, a constructed populations based on two Generalized Extreme Value (GEV) distributions, the first with location, scale, and shape parameters

$$\{\kappa, \alpha, \xi\} = \{0.08, 24326, 6378\}$$

and the second with parameters

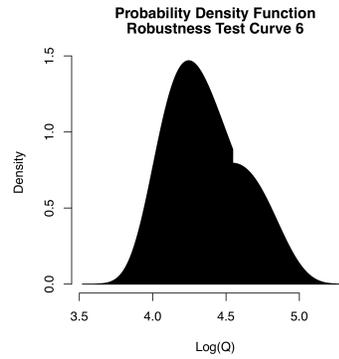
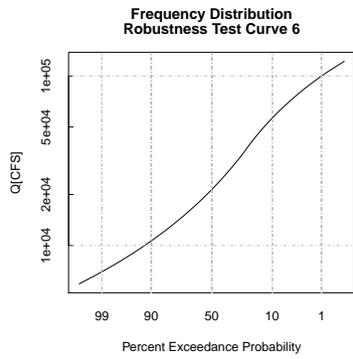
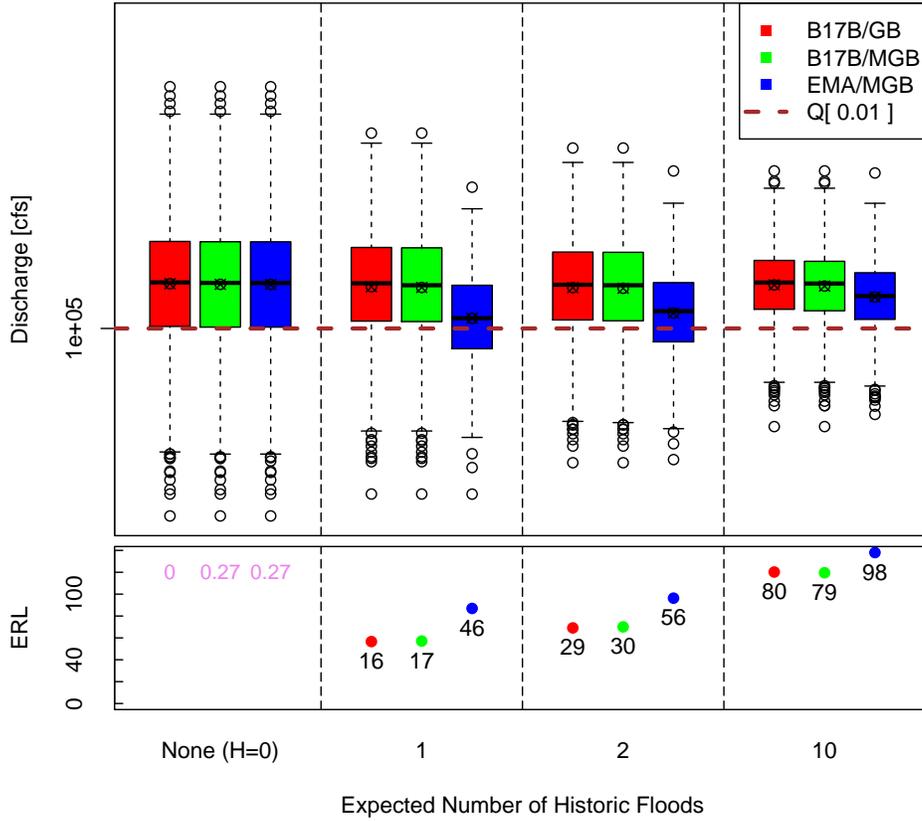
$$\{\kappa, \alpha, \xi\} = \{-0.55, 10000, 17330\}$$

Figure 7: Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from robustness test curve 5.



The lower three quarters of the distribution function employs the second parameters, and the upper quarter is based on the first parameters. The distributions have the same upper quartile, which is where they join. As was seen in figure 7 and many other examples, when historical information is present, EMA/MGB performs much better than the other estimators.

Figure 8: Monte Carlo results based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from robustness test curve 6.



6 Examples Based on Real Data at Selected Test Sites

6.1 82 Example Test Sites

Eighty-two USGS streamflow-gaging (streamgages) stations were selected by a subset of the HFAWG named the “Data Group.” The types of data found at these sites are organized into four categories in this report:

1. Systematic Gage Data, no historical or low outlier data (26 sites);
2. Historical Data, possibly including high outliers (19 sites);
3. Low Outliers; no historical information (20 sites);
4. Low Outliers, Historical and/or High Outliers (17 sites).

These eighty-two sites include all of the sites used as examples in Bulletin 17B (B17B) were also included in this study. Clearly 82 sites is a limited sample of the many thousands of streamgage records throughout the Nation. However, the set is believed to cover the range of situations, and particularly the most difficult situations, that arise in practice.

The respective estimated frequency curves for the sites are presented graphically in appendix B. Because we do not know the true frequency curve, judgments about the various estimates are necessarily subjective; graphs provide a convenient way to visualize the differences. The magnitude of observed differences, however, can be summarized in terms of a statistic, the

relative percent difference (RPD), defined as:

$$RPD \equiv 100 \left(\frac{\hat{Q}_p^{EMA/MGB} - \hat{Q}_p^{B17B/*}}{\hat{Q}_p^{B17B/*}} \right) \quad (4)$$

where p corresponds to the quantile of interest and $B17B/*$ refers to $B17B/GB$ or $B17B/MGB$, respectively, depending on which estimators are being compared. Although the RPD does not tell us which estimator is better, it does quantify the magnitude of observed differences between the estimators.

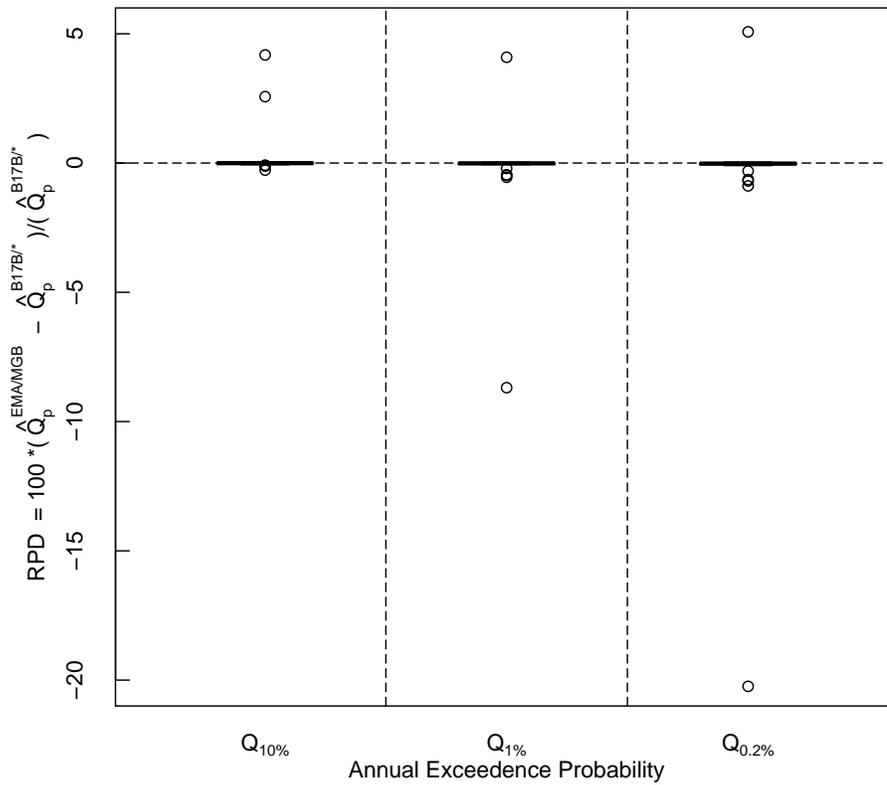
6.2 Sites with Systematic Gage Data and No Low Outliers or Historical Information

The first category, “Gage Only” data, included 26 sites. The sites have systematic data with no historical information, no below gage-base floods, and no low outliers identified by either using the GB or MGB tests. Almost all RPDs were zero, as expected. In this case all of the estimators are, in theory, identical². As can be seen in figure 9, however, there were two sites where, in fact, the RPD was not zero.

At site 02037500, James River at Richmond (figure 29), the RPD ranged between 3 to 5% for the 3 quantiles. This was because the 1937 peak discharge was recorded with a qualification code indicating the discharge was greater than the reported 152,000 [cfs] value. PEAKFQ/B17B, the USGS software that implements Bulletin 17B, by default omitted the 1937 peak. A user-supplied point discharge value of 152,000 [cfs] was instead used to

²Apparent differences between PEAKFQ/B17B and PeakfqSA results of less than 1% in RPD occur because PEAKFQ/B17B rounds quantile estimates to between 2 and 3 significant digits; PeakfqSA does no rounding. The estimated moments, which are not rounded, are identical in these cases

Figure 9: Relative Percent Difference (RPD) for B17B/GB and EMA/MGB estimators for 10%, 1%, and 0.2% exceedance probabilities. Includes 26 sites without historical flood information where no low outliers were identified by Grubbs-Beck tests.



characterize the 1937 peak. This is a “known problem” with B17B, but it bears repeating: PEAKFQ/B17B is not well adapted to incorporating non-standard discharge values; in some cases the way it handles them is to ignore them altogether. As a result, the PEAKFQ/B17B estimates, which reflect B17B methods, do not properly employ the data.

EMA, the alternative method, does accommodate interval data. Thus the 1937 peak could be correctly described as inside an interval from 152,000 to infinity. This was used in the EMA analysis to capture the additional information associated with the 1937 peak. Because EMA employs an interval range greater than 152,000[cfs] for the large 1937 peak, EMA estimated higher values for the upper portion of the frequency curve.

Site 05586500, Hurricane Creek near Roodhouse (figure 32), had a similar problem. It showed a RPD range from 4 to -20%. The sixth lowest recorded discharge had a qualification code indicating the recorded value was less than 70 [cfs]. PEAKFQ/B17B employed a gage base at 70 [cfs], consequently omitting five additional recorded point discharge values below 70 [cfs]. No low outliers were identified by either the GB or MGB test. Thus PEAKFQ/B17B unnecessarily truncated a portion of the left hand tail due to the qualification code of one observation. A user-supplied interval discharge range from 0 to 70 [cfs] was properly set in EMA for that single water year, and the five recorded point discharges below 70 [cfs] were included in the flood frequency analysis. Because EMA used all recorded discharges and the 0 to 70 [cfs] censored data, EMA’s estimates were lower particularly at the upper end of the curve.

In both cases where the estimates differed, it was because EMA can accom-

modate data properly that PEAKFQ/B17B cannot accommodate.

6.3 Sites with Historical Information

The testing for the “Historical Data” category included 19 sites, some of which included high outliers. The historical data sites illustrate a fundamental difference between EMA and B17B (Stedinger and Cohn, 1986; England et al., 2003).

EMA estimates, and therefore the RPDs, are sensitive to the historic threshold and historical period employed. All attempts were made to manually enter the same values into both software programs. However, some adjustments were made to accommodate PEAKFQ/B17B’s inability to use interval discharge ranges. Many sites had one to three recorded historic peaks that exceeded the historic threshold. A few sites had recorded gage heights at or near the record value with missing discharges. Interval discharge values were set in EMA to accommodate these observations; because PEAKFQ/B17B has no corresponding capability, point discharge values were estimated by relating log-space discharge to log-space gage height for those years and these were entered into PEAKFQ/B17B with the use of similar historic thresholds. Additionally, as a default, PEAKFQ/B17B sets an historic threshold at the lowest recorded historic value for a user-specified historic period. Thus all missing years of information in the historic record are effectively set to the lowest historic threshold. If a systematic record is missing any discharges (a broken systematic record) in a historical period, the missing data is set to the same historic threshold. This is not the case if there is missing data in a purely systematic record. PEAKFQ/B17B will assume no information is

known about those missing years of systematic record.

Overall, the majority of sites with historical information showed a positive RPD in the flood estimates (figure 10). The interquartile distances ranged from approximately 0 to 10% for $\hat{Q}_{1\%}$. Three sites had higher RPD for $\hat{Q}_{1\%}$, ranging from 11 to 37%. EMA and B17B provide substantially different estimates for these three sites, all of which include historical information and high outliers. Site 06216500, Pryor Creek near Billings, had one large flood that was the largest in an extended historical period of 99 years. Figure 11 shows the frequency plots for both B17B and EMA. Based on visual inspection of the frequency curves, EMA seems to provide better fit to the data, most clearly to the high outlier in the right hand tail.

6.4 Sites with Systematic Gage Data and Low Outliers

The testing for the “Low Outliers” category utilized 20 sites with low outliers identified using either the standard Grubbs-Beck (GB) test recommended by B17B or the multiple Grubbs-Beck (MGB) test (Cohn et al., 2013). These sites did not include historic data. Two sets of comparisons were made between the results of EMA/MGB with:

1. B17B using the GB test followed by the conditional probability adjustment (CPA) (B17B/GB), and
2. B17B using the MGB identified low outlier threshold followed by the CPA (B17B/MGB).

Current PEAKFQ/B17B software does not include a MGB test option, so, where needed, the MGB threshold was computed independently and then

Figure 10: Relative Percent Difference (RPD) for B17B/GB and EMA/MGB estimators for 10%, 1%, and 0.2% exceedance probabilities. Figure represents 19 sites with historical information where no low outliers were identified by Grubbs-Beck tests.

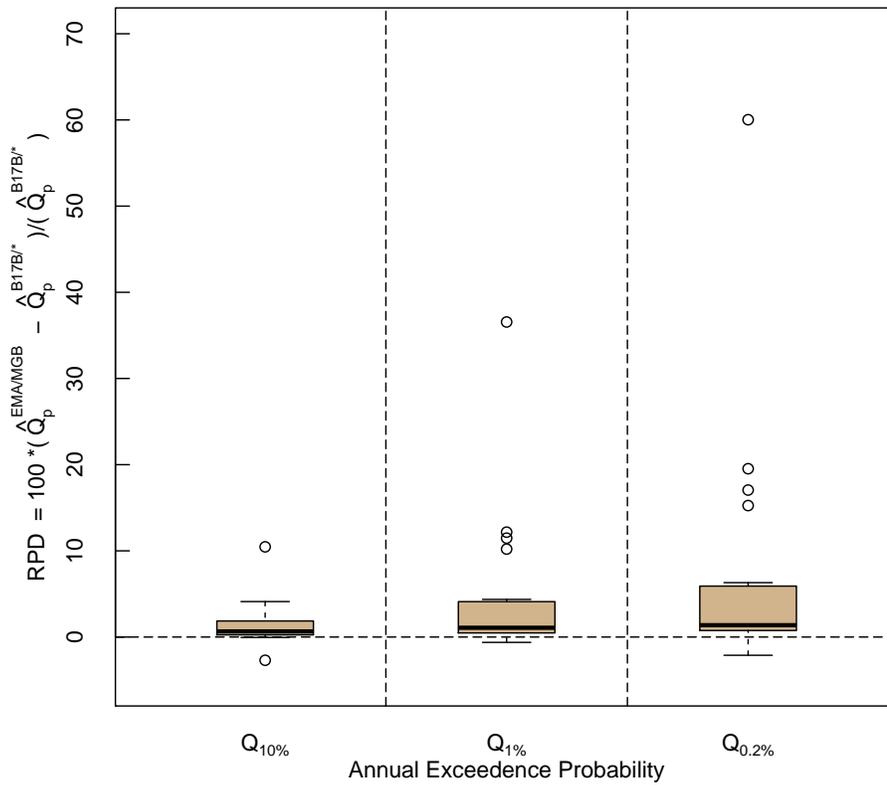
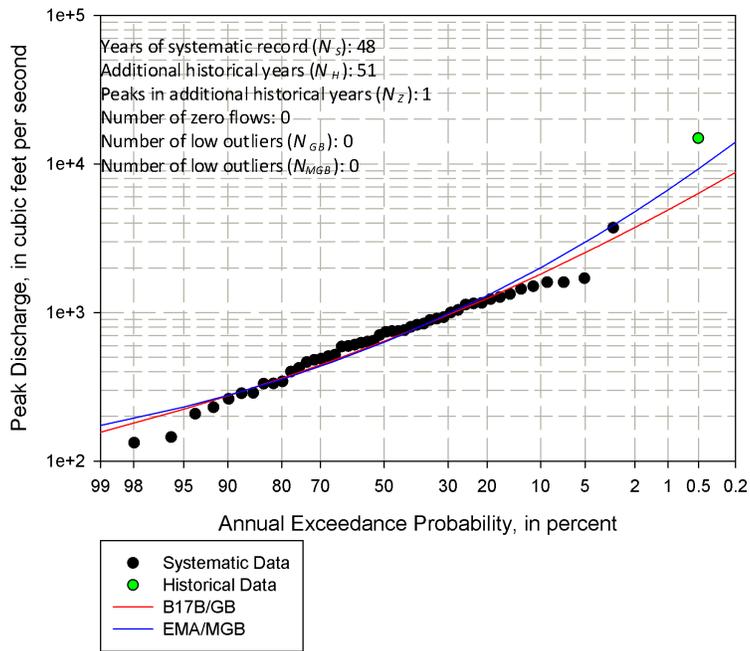


Figure 11: Pryor Creek near Billings, MT (06216500)

Pryor Creek near Billings, MT
(Station 06216500)



entered into PEAKFQ as a user-supplied low outlier threshold.

The systematic flood series for site 08133500, North Concho River at Sterling City, TX (Figure 75), included three peaks with a qualification code indicating the discharge was less than the reported value of 300 [cfs]. As discussed in the systematic “Gage Only” section (6.2), PEAKFQ/B17B arbitrarily set a gage base for the entire record, in this case omitting an additional 17 systematic point discharges less than 300 [cfs] that did not have remark codes. In EMA, the 3 peaks were recoded as between 0 and 300 [cfs]. However, the MGB test identified a low outlier threshold of 634 [cfs]. Thus EMA/MGB identified 23 low outliers and B17B/GB (with a default-set gage base of 300 [cfs]) identified 20 low outliers. The result is that the estimated flood quantiles do not differ by very much.

Flood estimate comparisons between EMA/MGB and B17B/GB for (figure 12) shows the median RPD is essentially zero. However, the RPD has substantial variability, indicating that the B17B/GB and EMA/MGB estimators behave differently when low outliers are present. This increased variability in RPD is attributed to the very different number of low outliers identified in the flood series and the methods used to handle low outliers in the frequency analysis, i.e., EMA’s low outlier censoring versus B17B’s CPA. Of the 20 sites in this low outlier category, the GB test found only 0 to 2 low outliers per site above gage base, while the MGB test found 1 to 46 low outliers (figure 13) At some sites, the MGB identified nearly 50% of the recorded floods (figure 14) as “low outliers.” This is the upper bound on the percentage of peaks that MGB will test and designate as low outliers.

Figure 12: Relative Percent Difference (RPD) for B17B/GB, B17B/MGB, and EMA/MGB estimators for 10%, 1%, and 0.2% exceedance probabilities. Figure represents 20 sites where low outliers were identified by Grubbs-Beck(GB) or generalized Grubbs-Beck (MGB), and no historical information.

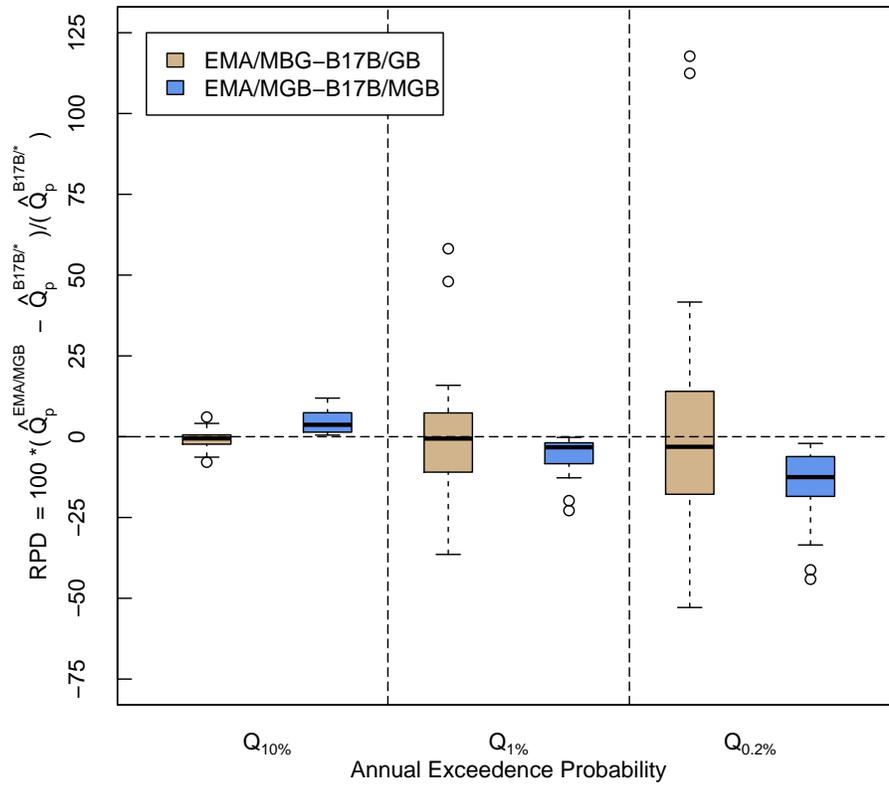


Figure 13: Number of floods identified as low outliers or below gage-base using the standard Grubbs-Beck (GB) and generalized Grubbs-Beck (MGB) tests. Figure represents 20 sites without historical information.

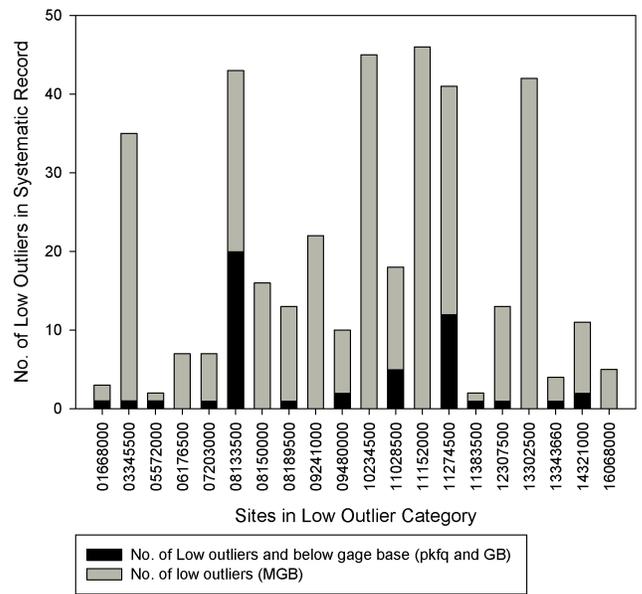
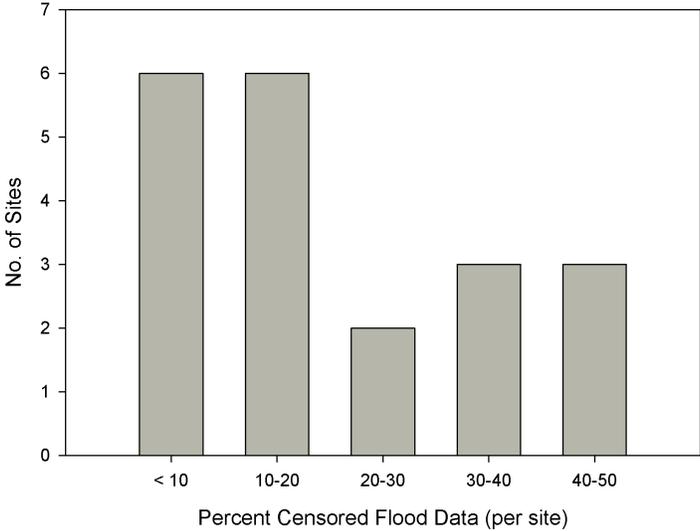


Figure 14: Percent low outliers identified using the generalized Grubbs-Beck (MGB) test. Figure represents 20 sites without historical information.

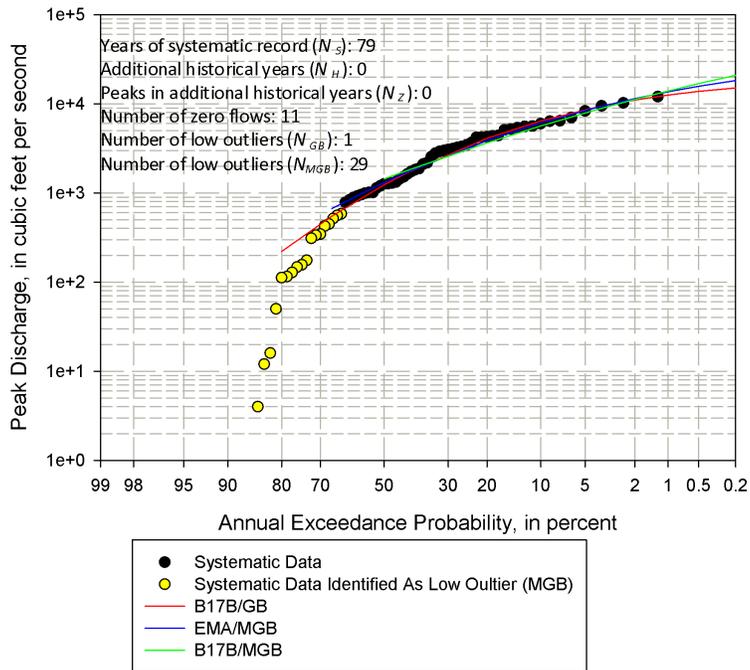


The observed RPDs were both positive and negative, a result of complex interactions between different low-outlier tests and different fitting procedures. However, use of the EMA/MGB method seems to produce a better fit to the upper portion of the frequency curve than is obtained with B17B/GB. Orestimba Creek near Newman (site 11274500, figure 15), which is also presented in B17B, represents an extreme case of low outliers – a flood series of 79 years including 11 zero floods. The LP3 distribution cannot describe the full range of observed flood flows at this site because the support for the LP3 distribution vanishes for $Q \leq 0$ (Cohn et al., 2013). PEAKFQ/B17B treats all zero flows as below gage base, while EMA regards them as ordinary low outliers.

The Grubbs-Beck (GB) test yields a low-outlier threshold of 10.8 [cfs] and identification of a single additional low outlier. The MGB test, in contrast, yields a low outlier threshold of 782 [cfs], identifying 29 peaks, or 37% of the data, as low outliers. The RPD between EMA/MGB and B17B/GB ranges from 9% to 21% for the 1%, and 0.2% exceedance probability estimates, respectively (figure 15).

It is interesting to note that B17B/MGB, while close to EMA/MGB in this case, provides a different and much poorer fit to the data, as can be seen in figure 15. The B17B/MGB curve lies above the data for exceedance probabilities between 40-50%, and below the data for exceedance probabilities between 10-40%. Above the 10% level, the B17B/MGB curve rises more sharply than the EMA/MGB curve, which appears to match the concave downward trend in the data.

Figure 15: Orestimba Creek near Newman, CA (11274500), fit after application of the multiple Grubbs-Beck (MGB) test for low outliers.



The pattern seen at Orestimba is also observed at other sites with multiple low outliers. For example, at Santa Cruz River near Lochiel, AZ (09480000, figure 16) one sees that EMA/MGB, by treating the influential small peaks as low outliers, results in a reasonably close fit in the right-hand tail. On the other hand, the B17B/GB estimator generates a frequency curve that is greatly exceeded by two of the sample values.

6.5 Sites with Low Outliers, Historical and/or High Outliers

Seventeen of the 82 sites included the combination “Low Outlier, Historical and/or High Outlier.” This category contains sites whose flood series have low outliers with

1. High outliers in a systematic record, or
2. High outliers in a historical period.

Nothing fundamentally new appeared in these cases. The RPD between EMA/MGB and B17B/GB are similar to those found in both the “Low Outlier” and “Historical” categories. The median RPD for the $\hat{Q}_{10\%}$ estimates remained near zero while the median RPD were slightly positive for the $\hat{Q}_{1\%}$ and $\hat{Q}_{0.2\%}$ estimates (figure 17). About a third of the sites in this category showed a RPD greater than 15% for the 0.2% estimates and three sites were less than 18%. The largest RPD between EMA/MGB and B17B/GB was found at site 11176000, Arroyo Mocho near Livermore, CA. The RPDs for the estimates were between -48 and -68%, for the $\hat{Q}_{1\%}$ and $\hat{Q}_{0.2\%}$, respectively. EMA/MGB found 19 low outliers in the systematic record (figure 18) while B17B/GB found only one low outlier above gage

Figure 16: Santa Cruz River near Lochiel, AZ (09480000)

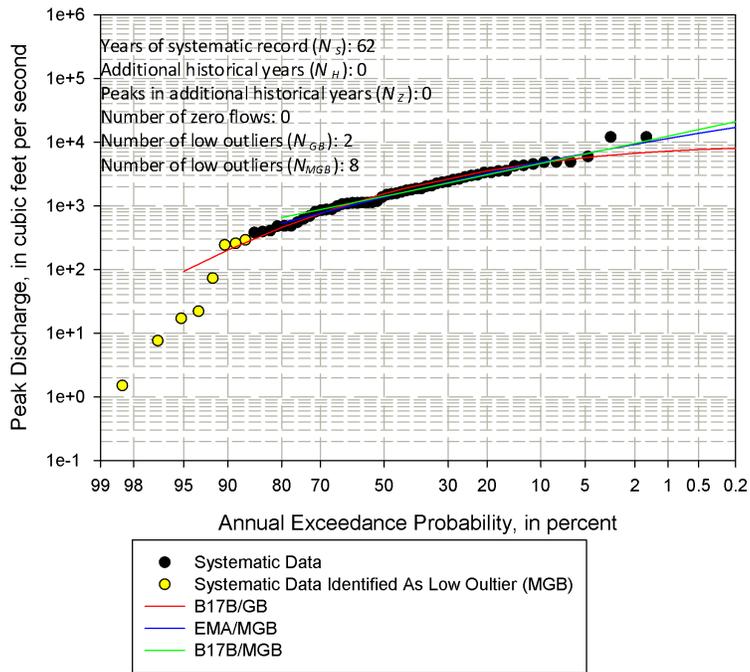


Figure 17: Relative Percent Difference (RPD) for B17B/GB, B17B/MGB, and EMA/MGB estimators for 10%, 1%, and 0.2% exceedance probabilities. Represents 17 sites with historical flood information where low outliers were identified by the Grubbs-Beck (GB) or multiple Grubbs-Beck (MGB) test.

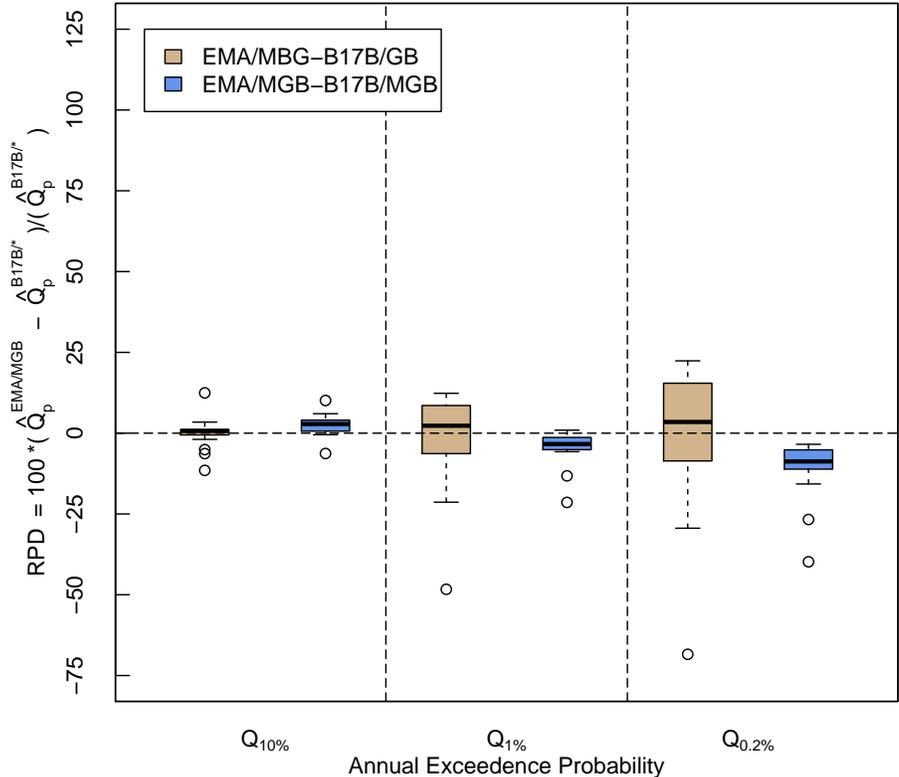
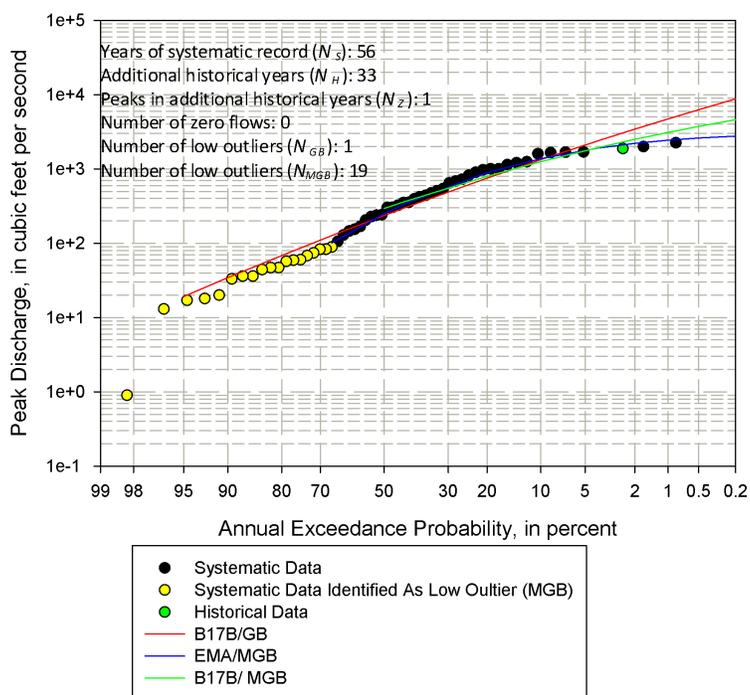


Figure 18: Arroyo Mocho near Livermore, CA (11176000)

Arroyo Mocho near Livermore, CA
(Station 11176000)



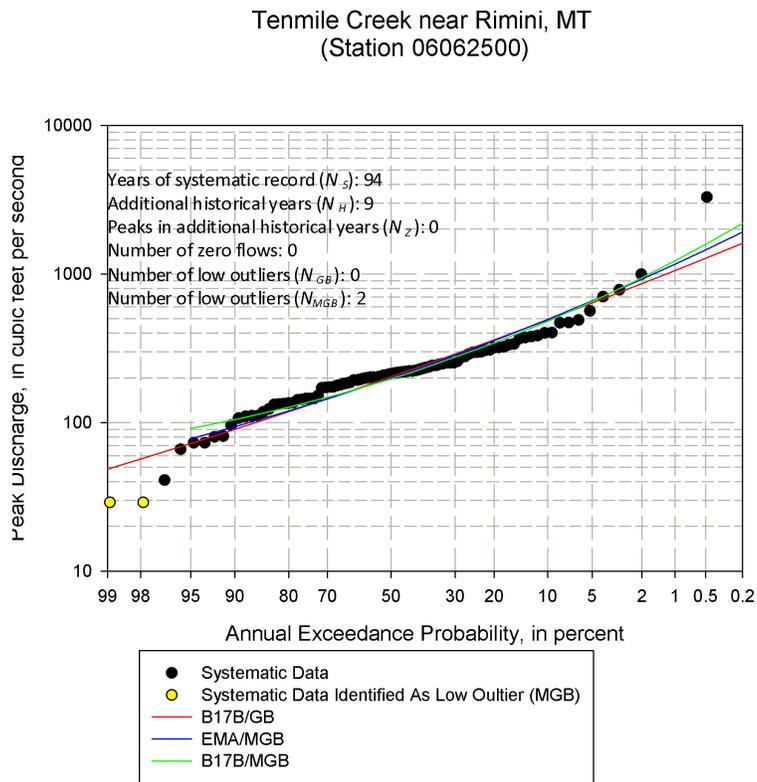
base. Additionally, one historic peak was recorded in the flood series. By censoring multiple low outliers, EMA/MGB more accurately fit the right-hand tail.

A large positive RPD between EMA/MGB and B17B/GB was found at site 06062500, Tenmile Creek near Rimini, MT. A 10 and 19% RPD difference was found for the $\hat{Q}_{1\%}$ and the $\hat{Q}_{0.2\%}$ estimates, respectively. EMA/MGB found 2 low outliers, while B17B/GB found none (figure 19). Here EMA seems to provide a fit that is more consistent with the trends exhibited by the largest 7 observations.

Figures 18 and 19 illustrate an important point: Both B17B/GB and EMA/MGB often provide a good fit to the data they employ. However, B17B/GB can be highly influenced by PILFs with the result being a poor fit at the high end of the distribution. EMA/MGB, on the other hand, avoids this problem by identifying and recoding PILFs so that their exact magnitudes do not distort the fit in the right-hand tail.

The RPD between EMA/MGB and B17B/MGB when a flood series has both low and high outlier data is similar to that found when only low outliers are present. The median value and interquartile range for the estimates in the “combination category” was slightly positive while the median values and interquartile ranges for the estimates in the “low outlier” category were slightly negative. As expected, the variance in the RPD was minimized for all estimates when the same low outlier threshold was used. The higher estimate from B17B/MGB is illustrated by figure 18 (11176000, Arroyo Mocho near Livermore, CA) where a -21% RPD difference for the $\hat{Q}_{1\%}$ was found. The MGB low outlier threshold of 106 [cfs] was used in B17B/MGB. The

Figure 19: Ten Mile Creek near Rimini, MT (06062500)



MGB test identified almost 33% of the peaks as low outliers. Using the same MGB test, based on visual inspection, EMA/MGB and B17B/MGB both show an improved fit versus B17B/GB.

6.6 Resampling Studies

Figures 20 - 24 depict results from applying the three estimators, B17B/GB, B17B/MGB, and EMA/MGB, to resampled data from five of the longest-record (> 100 observations) sites among the 82 “test” sites considered in section 6. In these cases, we do not know what the true value of the 1% exceedance event is, so the figures instead employ an interpolated value based on the two largest observations in the dataset. However, this is an unreliable estimator; as noted in section 6.7, many of the “test” sites were selected specifically because they contained high outliers. In reviewing figures 20 - 24, it is likely best to use one’s judgement about the *reasonableness* of the results, possibly referring back to the test-site results, rather than trying to conjure up a strict quantitative assessment. However, it is noteworthy that, when historical information is present, the EMA/MGB estimator generally outperforms the other estimators in terms of ERL, with site 03011020 in figure 20 being an exception.

6.7 Summary

82 streamflow-gaging stations were chosen as a representative sample of long-term sites whose flood series include a variety of situations and problems that are believed to be found throughout the U.S. The flood data was divided into four categories:

Figure 20: Results based on resampled data, $N_S = 40$ and $N_H = 100$, drawn from observed discharges at “Historical” category site 03011020.

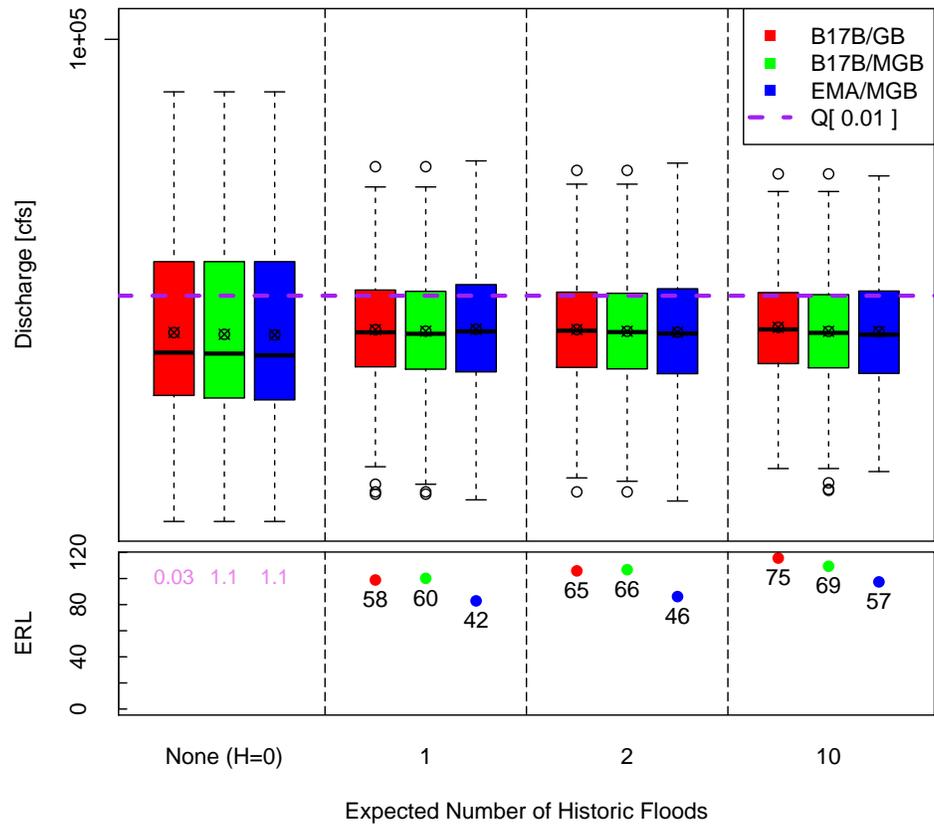


Figure 21: Results based on resampled data, $N_S = 40$ and $N_H = 100$, drawn from observed discharges at “Low Outlier” category site 11152000.

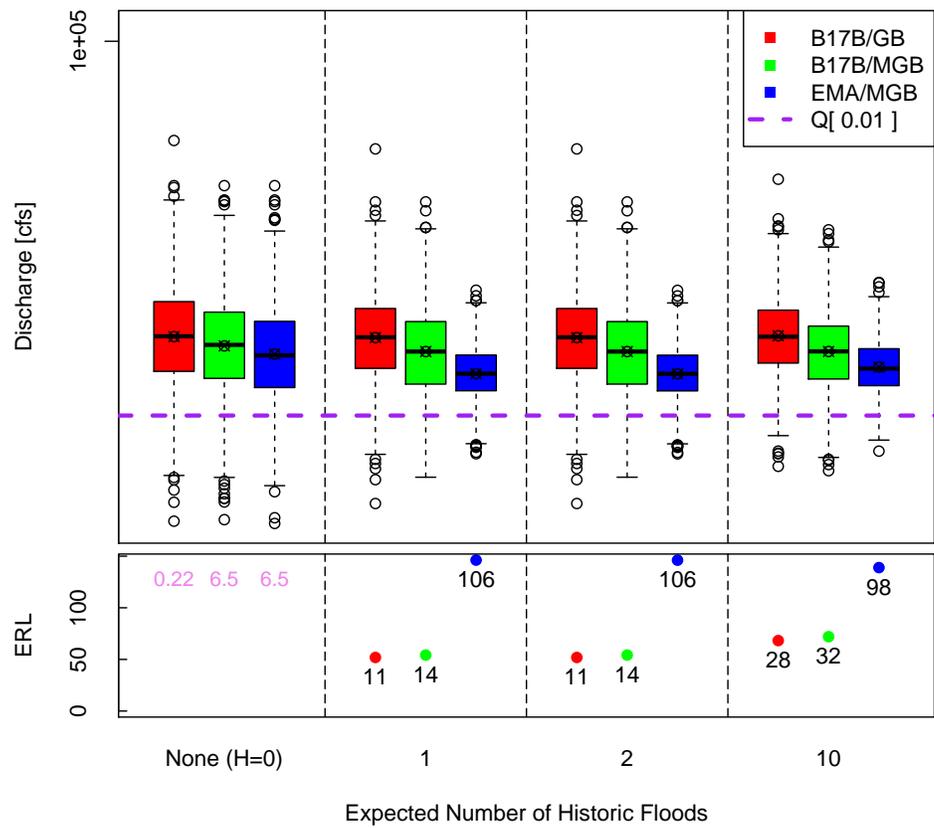


Figure 22: Results based on resampled data, $N_S = 40$ and $N_H = 100$, drawn from observed discharges at “Gage Only” category site 14048000.

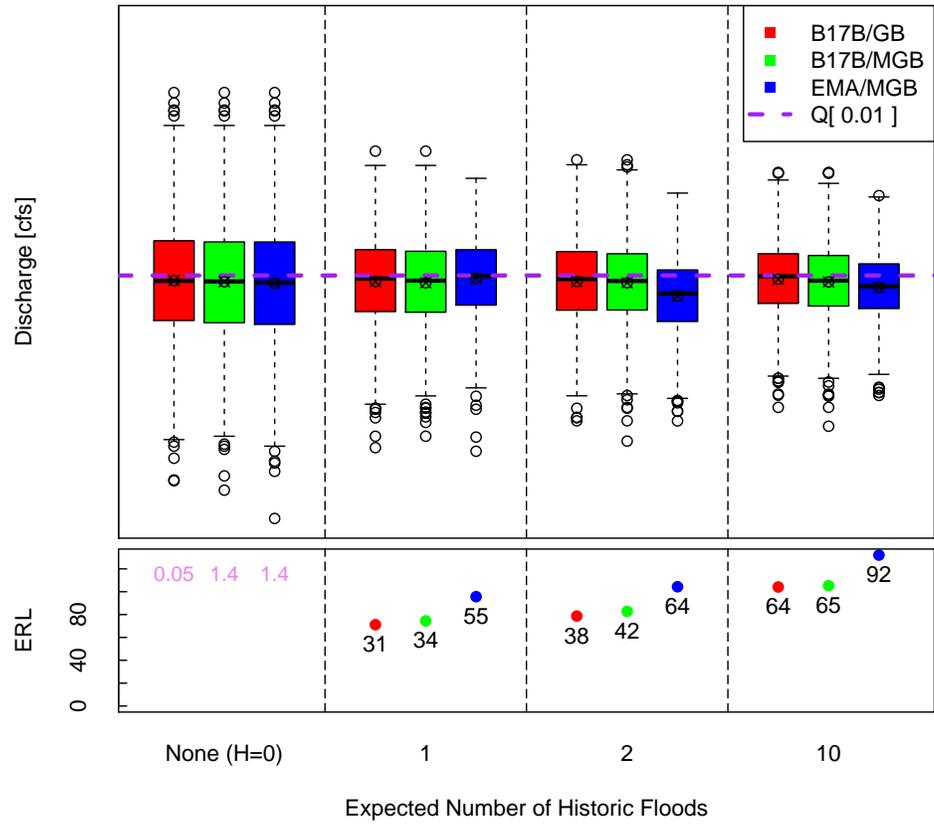


Figure 23: Results based on resampled data, $N_S = 40$ and $N_H = 100$, drawn from observed discharges at “Low Outlier” category site 14321000.

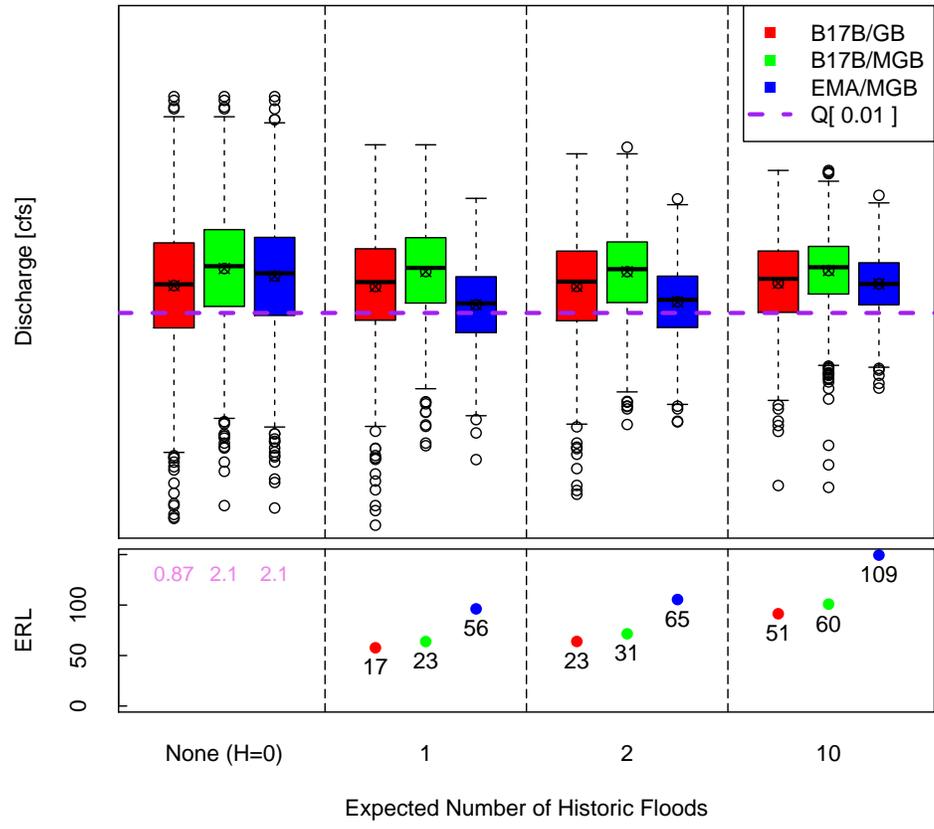
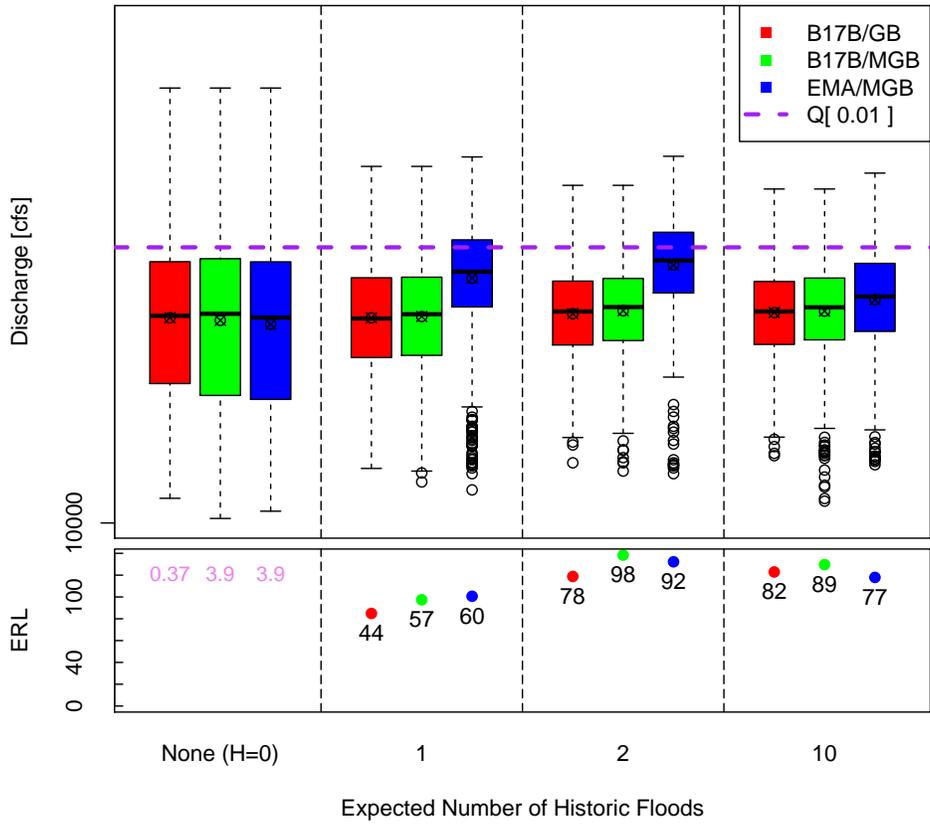


Figure 24: Results based on resampled data drawn from observed discharges at “Combination” category site 13185000.



1. Systematic Gage Data, no low outliers or historical information,
2. Historical and/or High outliers,
3. Low Outliers, and
4. Low Outliers, with Historical and/or High Outliers.

The performance of EMA/MGB, B17B/GB and B17B/MGB methods were compared visually for each of the sites, and a relative percent difference (RPD) statistic was calculated corresponding to each of three exceedance probabilities to compare the EMA and B17B estimates.

When only systematic flood data were present, identical flood estimates (RPD equal to zero) were obtained except for those cases where B17B cannot accommodate non-standard data correctly. When historical and/or high outliers were present, observed RPDs were positive more than half the time. This is at least in part due to the 82 sites that were selected for testing, many of which included historical “high outliers.” EMA tends to attach more “weight” to historical flood information than does B17B (Stedinger and Cohn, 1987), so where the historical period includes an unusually high peak, the EMA/MGB estimate will tend to be higher than the B17B/GB estimate.

Flood series that contain multiple low outliers exhibited a range of results primarily due to the very different number of low outliers identified by the GB and MGB tests and the methods used to handle low outliers in the frequency analysis, i.e., EMA’s censoring versus B17B’s CPA. When the same low outlier threshold was used for the EMA and B17B’s CPA fitting

procedure, the RPDs were usually smaller and negative. For those sites with low outliers with historical, and/or high outlier data, similar RPD were found to those in the historical and low outlier categories. This combination group included both low outliers and high outliers. The RPD were notably more positive for these estimates.

EMA/MGB generally identified more low outliers, when low outliers were present. In these cases, based on subjective assessment EMA/MGB always provided a closer fit to the largest peaks in the dataset.

Additional studies were conducted that involved resampling data from sites with the longest records. Although the true value of the 1% exceedance event is not known for these sites, the resampling experiments confirmed that the EMA/MGB estimator performed reasonably well in all cases and generally provided higher average gains than the alternative estimators.

7 Conclusions

The study reported here was designed to determine how well proposed changes to Bulletin 17B would perform in practice compared to the current recommendations in Bulletin 17B. In particular, the investigations focused on techniques for:

- Incorporating information related to historical flooding that occurred outside the period of systematic streamgaging;
- Addressing the identification and treatment of PILFs (zero flows and low outliers);

In order to answer these questions, Monte Carlo studies were conducted by resampling from real records, or drawing random samples from specified LP3 and non-LP3 distributions. In addition, problematic datasets from across the country were selected to serve as test cases for comparing the estimators. To summarize, results are presented for:

- Monte Carlo simulations employing data drawn from specific LP3 populations;
- Monte Carlo simulations employing data drawn from non-LP3 populations that were selected to reflect likely deviations, based on the experience of the Data Group, from the hypothesized LP3 distribution;
- A direct subjective comparison of results at 82 real “test sites” identified by an independent Data Group as both “typical” and “challenging” for flood frequency estimation;

- Resampling with replacement of the data at the 82 sites.

It is believed that, taken together, these studies provide a reasonably comprehensive, valid and robust assessment of the properties of the Bulletin 17B procedures and proposed extensions and improvements.

The results demonstrate that the proposed alternative method, denoted EMA/MGB:

- Generally performs at least as well as, and in some cases much better than, Bulletin 17B procedures in terms of the Mean Square Error (MSE) of flood quantile estimates;
- Allows for incorporation of more general types of flood-frequency information, thereby voiding some annoying problems that arise when applying Bulletin 17B in practice with datasets containing non-standard flood data (exceedances of thresholds, and “less-than” values).

In summary, the results here generally confirmed other studies published in the hydrological literature that have found that EMA generally provides improved flood frequency estimates.

A Appendix: Characteristics of 82 Test Sites

Table 2: Characteristics of at-site streamflow records used in comparing Bulletin 17B and EMA estimators.

DA	drainage area in square miles;	$N_{<gd}$	number of point discharge values below gage base;
N_{Pk}	number of peaks;	Q_{GB}	Grubbs-Beck critical value (cfs)–low-outlier criterion;
N_S	number of systematic peaks;	N_{GB}	number of low outliers identified using Grubbs-Beck test;
N_0	number of zero peaks;	N_{MGB}	number of low outliers identified using multiple Grubbs-Beck test
N_H	length of historical period;		
N_Z	number of historic peaks; n, Jr.,		
Q_{gd}	gage base discharge (cfs);		

Site	USGS ID	Station Name	DA	N_{Pk}	N_S	N_0	N_H	N_Z	Q_{gd}	$N_{<gd}$	Q_{GB}	N_{GB}	N_{MGB}
1	01076500	Pemigewasset River at Plymouth, NH	622	106	106	0	106	0	0	0	6316.7	0	0
2	01350000	Schoharie Creek at Prattsville, NY	237	100	99	0	106	1	0	0	2064	0	0

DRAFT

Site	USGS ID	Station Name	DA	N_{Pk}	N_S	N_0	N_H	N_Z	Q_{gd}	$N_{<gd}$	Q_{GB}	N_{GB}	N_{MGB}
3	01439500	Bush Kill at Shoemakers, PA	117	102	102	0	102	0	0	0	427.5	0	0
4	01555500	East Mahantango Creek near Dalmatia, PA	162	81	81	0	81	0	0	0	784.9	0	0
5	01562000	Raystown Branch Juniata River at Saxton, PA	756	99	98	0	122	1	0	0	2977	0	0
6	01635500	Passage Creek near Buck- ton, VA	86.5	78	78	0	78	0	0	0	267.5	0	0
7	01636500	Shenandoah River at Mil- lville, WV	3041	97	95	0	141	2	0	0	4178.7	0	0
8	01668000	Rappahannock River near Fredericksburg, VA	1595	100	100	0	103	0	0	0	5209.8	1	2
9	02037500	James River near Rich- mond, VA	6753	76	76	0	76	0	0	0	14965.1	0	0

85

DRAFT

Site	USGS ID	Station Name	DA	N_{Pk}	N_S	N_0	N_H	N_Z	Q_{gd}	$N_{<gd}$	Q_{GB}	N_{GB}	N_{MGB}
10	02138500	Linville River near Nebo, NC	66.7	89	88	0	95	1	0	0	461.7	0	0
11	02256500	Fisheating Creek at Palm- dale, FL	311	79	79	0	79	0	0	0	227.9	0	0
12	03011020	Allegheny River at Sala- manca, NY	1608	107	107	0	146	0	0	0	8502.9	0	0
13	03051000	Tygart Valley River at Bel- ington, WV	406	104	103	0	123	1	0	0	3992.9	0	0
14	03159500	Hocking River at Athens, OH	943	78	77	0	137	1	0	0	3655.4	0	0
15	03183500	Greenbrier River at Alder- son, WV	1364	115	115	0	115	0	0	0	10190.8	0	0
16	03289500	Elkhorn Creek near Frank- fort, KY	473	72	70	0	94	2	0	0	3050.9	1	6

Site	USGS ID	Station Name	DA	N_{Pk}	N_S	N_0	N_H	N_Z	Q_{gd}	$N_{<gd}$	Q_{GB}	N_{GB}	N_{MGB}
17	03345500	Embarras River at Ste. Marie, IL	1516	99	99	0	101	0	0	0	1698.2	1	34
18	03550000	Valley River at Tomotla, NC	104	101	100	0	113	1	0	0	862.1	0	0
19	03558000	Toccoa River near Dial, GA	177	85	84	0	157	1	0	0	1022.6	0	0
20	03606500	Big Sandy River at Bruce- ton, TN	205	70	67	0	114	3	0	0	704.3	0	0
21	04293500	Missisquoi River near East Berkshire, VT	479	92	91	0	180	1	0	0	4328.8	0	0
22	05270500	Sauk River near St. Cloud, MN	1030	75	75	0	101	0	0	0	152.1	1	1
23	05291000	Whetstone River near Big Stone City, SD	398	83	83	0	101	0	0	0	23.4	0	29
24	05464500	Cedar River at Cedar Rapids, IA	6510	109	108	0	311	1	0	0	3187.3	0	5

Site	USGS ID	Station Name	DA	N_{Pk}	N_S	N_0	N_H	N_Z	Q_{gd}	$N_{<gd}$	Q_{GB}	N_{GB}	N_{MGB}
25	05572000	Sangamon River at Monti- cello, IL	550	101	101	0	103	0	0	0	755.6	1	1
26	05586500	Hurricane Creek near Roodhouse, IL	2.3	45	44	0	45	1	70	6	38	0	0
27	06062500	Tenmile Creek near Rimini, MT	30.9	94	94	0	103	0	0	0	28.1	0	2
28	06176500	Wolf Creek near Wolf Point, MT	251	37	37	0	84	0	0	0	0.5	0	7
29	06216500	Pryor Creek near Billings, MT	440	49	48	0	99	1	0	0	101.6	0	0
30	06406000	Battle Cr at Hermosa, SD	178	61	61	0	61	0	0	0	1.8	0	0
31	06600500	Floyd River at James, IA	886	76	76	0	119	0	0	0	282.2	0	0
32	06710500	Bear Creek at Morrison, CO	164	98	98	0	120	0	0	0	19.3	0	0

Site	USGS ID	Station Name	DA	N_{Pk}	N_S	N_0	N_H	N_Z	Q_{gd}	$N_{<gd}$	Q_{GB}	N_{GB}	N_{MGB}
33	06897000	East Fork Big Creek near Bethany, MO	95	53	52	0	77	1	0	0	376.2	1	6
34	06898000	Thompson River at Davis City, IA	701	80	79	0	126	1	0	0	1171.9	0	0
35	06933500	Gasconade River Jerome, MO	2840	92	88	0	114	4	4320	1	4369.7	0	18
36	07067000	Current River at Van Bu- ren, MO	1667	99	98	0	107	1	0	0	2536.9	0	0
37	07138600	White Woman C Tr near Selkirk, KS	38	39	39	4	54	0	0	0	2.5	0	12
38	07203000	Vermejo River near Daw- son, NN	301	76	76	0	81	0	0	0	59.7	1	6
39	07208500	Rayado Creek near Cimar- ron, NN	65	86	86	0	96	0	0	0	7	0	0

Site	USGS ID	Station Name	DA	N_{Pk}	N_S	N_0	N_H	N_Z	Q_{gd}	$N_{<gd}$	Q_{GB}	N_{GB}	N_{MGB}
40	07382000	Bayou Cocodrie near Clearwater, LA	240	72	72	0	88	0	0	0	272.6	0	0
41	08133500	N Concho Rv at Sterling City, TX	588	65	65	0	68	0	300	20	130.7	0	23
42	08150000	Llano Rv near Junction, TX	1854	91	91	0	95	0	0	0	34.4	0	16
43	08164000	Lavaca Rv near Edna, TX	817	73	72	0	75	1	0	0	734.7	1	1
44	08167000	Guadalupe Rv at Comfort, TX	839	76	72	0	163	4	0	0	104.9	0	0
45	08171000	Blanco Rv at Wimberley, TX	355	87	86	0	142	1	0	0	57.6	1	27
46	08189500	Mission Rv at Refugio, TX	690	71	71	0	71	0	0	0	65.9	1	12
47	08378500	Pecos River near Pecos, NM	189	87	85	0	90	2	0	0	70.3	0	0

DRAFT

Site	USGS ID	Station Name	DA	N_{Pk}	N_S	N_0	N_H	N_Z	Q_{gd}	$N_{<gd}$	Q_{GB}	N_{GB}	N_{MGB}
48	08380500	Gallinas Creek near Mon-tezuma, NM	84	93	93	0	95	0	0	0	16.2	0	0
49	08387000	Rio Ruidoso at Hollywood, NM	56	120	56	57	0	0	0	0	19.1	0	0
50	09241000	Elk River at Clark, CO	216	78	78	0	93	0	0	0	1025.5	0	22
51	09361500	Animas River at Durango, CO	692	102	102	0	113	0	0	0	1374.4	1	2
52	09471000	San Pedro River at Charleston, AZ	1234	95	95	0	105	0	0	0	490.9	1	3
53	09480000	Santa Cruz River near Lochiel, AZ	82.2	62	62	0	62	0	0	0	10.1	2	8
54	09482500	Santa Cruz River at Tucson, AZ	2222	94	93	0	119	1	0	0	559.6	0	0
55	10128500	Weber River near Oakley, UT	162	105	105	0	106	0	0	0	579.8	0	0

91

DRAFT

DRAFT

Site	USGS ID	Station Name	DA	N_{Pk}	N_S	N_0	N_H	N_Z	Q_{gd}	$N_{<gd}$	Q_{GB}	N_{GB}	N_{MGB}
56	10234500	Beaver River near Beaver, UT	91	97	97	0	97	0	0	0	40.4	0	45
57	11028500	Santa Maria C near Ra- mona, CA	57.6	71	71	11	97	0	0	0	0	1	13
58	11152000	Arroyo Seco near Soledad, CA	244	105	105	0	105	0	0	0	447.9	0	46
59	11176000	Arroyo Mochó near Liver- more, CA	38.2	57	56	1	89	1	0	0	2.4	1	19
60	11266500	Merced R A Pohono Bridge nr Yosemite, CA	321	94	94	0	94	0	0	0	819.6	0	0
61	11274500	Orestimba C near New- man, CA	134	79	79	11	79	0	0	0	10.8	1	29
62	11383500	Deer C near Vina, CA	208	94	94	0	99	0	0	0	521.4	1	1
63	11464500	Dry C near Cloverdale, CA	87.8	40	39	0	43	1	0	0	693.5	1	15

92

DRAFT

Site	USGS ID	Station Name	DA	N_{Pk}	N_S	N_0	N_H	N_Z	Q_{gd}	$N_{<gd}$	Q_{GB}	N_{GB}	N_{MGB}
64	11522500	Salmon R A Somes Bar, CA	751	86	86	0	99	0	0	0	1923.5	1	1
65	12039500	Quinault River at Quinault Lake, WA	264	97	96	0	101	1	0	0	9162.2	0	47
66	12134500	Skykomish River near Gold Bar, WA	535	82	82	0	82	0	0	0	2227.9	1	0
67	12307500	Moyie River at Eileen, ID	755	53	53	0	53	0	0	0	6220.4	0	12
68	12413000	Nf Coeur D Alene River at Enaville, ID	895	74	71	0	99	3	0	0	3271.3	0	0
69	12414500	St Joe River at Calder, ID	1030	92	92	0	100	0	0	0	4310.5	0	0
70	12437950	East Fork Foster Creek Trib near Bridgeport, WA	4.75	21	21	0	21	0	0	0	0.8	0	0
71	12451000	Stehekin River at Stehekin, WA	321	89	89	0	100	0	0	0	3396.8	0	0

DRAFT

Site	USGS ID	Station Name	DA	N_{Pk}	N_S	N_0	N_H	N_Z	Q_{gd}	$N_{<gd}$	Q_{GB}	N_{GB}	N_{MGB}
72	13185000	Boise River near Twin Springs, ID	830	102	100	0	140	2	0	0	1840.6	1	25
73	13302500	Salmon River at Salmon, ID	3760	96	96	0	99	0	0	0	2039.2	0	42
74	13343660	Smith Gulch Tributary near Pataha, WA	1.85	20	20	0	20	0	0	0	0.7	1	3
75	14021000	Umatilla River at Pendleton, OR	637	57	57	0	86	0	0	0	1066.7	0	0
76	14048000	John Day River at Mcdonald Ferry, OR	7580	105	105	0	117	0	0	0	2549.2	0	0
77	14137000	Sandy River near Marmot, OR	263	99	99	0	99	0	0	0	2894.4	0	0
78	14321000	Umpqua River near Elkton, OR	3683	104	104	0	105	0	0	0	18022.5	2	9
79	15072000	Fish C near Ketchikan, AK	32.1	91	91	0	95	0	0	0	1352.2	0	0

94

DRAFT

Site	USGS ID	Station Name	DA	N_{Pk}	N_S	N_0	N_H	N_Z	Q_{gd}	$N_{<gd}$	Q_{GB}	N_{GB}	N_{MGB}
80	16068000	Eb Of Nf Wailua River near Lihue, Kauai, HI	6.27	95	95	0	95	0	0	0	418.8	0	5
81	16518000	West Wailuaiki Stream near Keanae, Maui, HI	3.66	90	90	0	96	0	0	0	926.2	0	0
82	16587000	Honopou Stream near Huelo, Maui, HI	0.64	98	98	0	100	0	0	0	36.7	0	0

B Appendix: Graphical Comparisons Between EMA and B17B at the 82 Test Sites

B.1 Systematic Data Sites

Figure 25: Site 01076500 with Systematic Data Only

Pemigewasset River at Plymouth, NH
(Station 01076500)

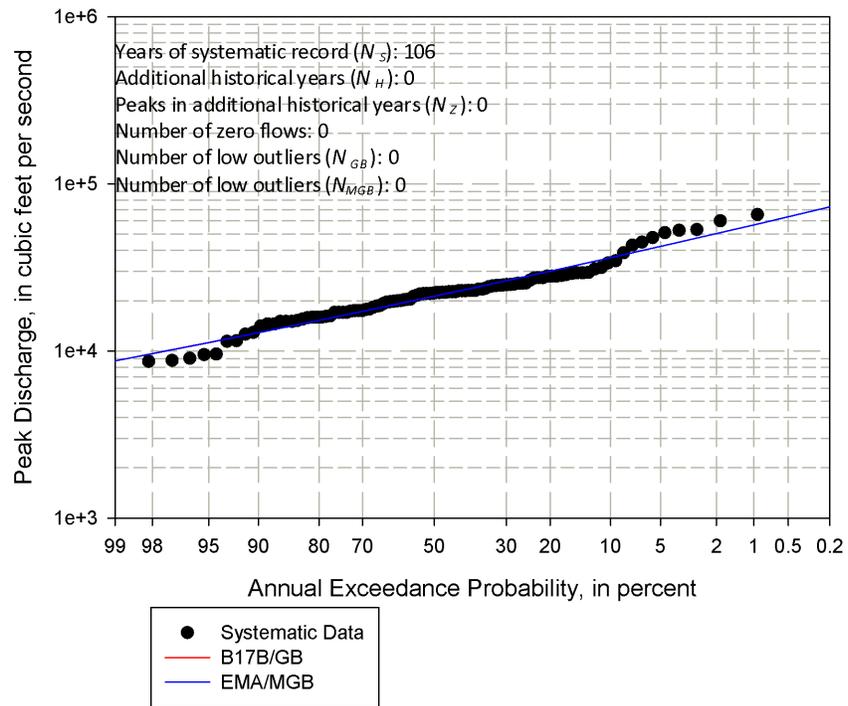


Figure 26: Site 01439500 with Systematic Data Only

Bush Kill at Shoemakers, PA
(Station 01439500)

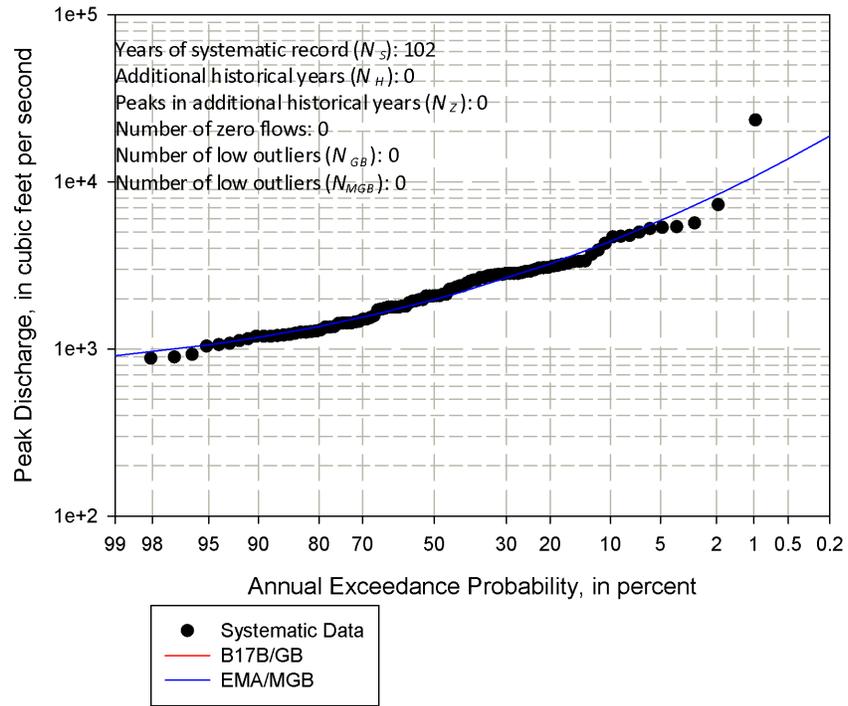


Figure 27: Site 01555500 with Systematic Data Only

East Mahantango Creek near Dalmatia, PA
(Station 01555500)

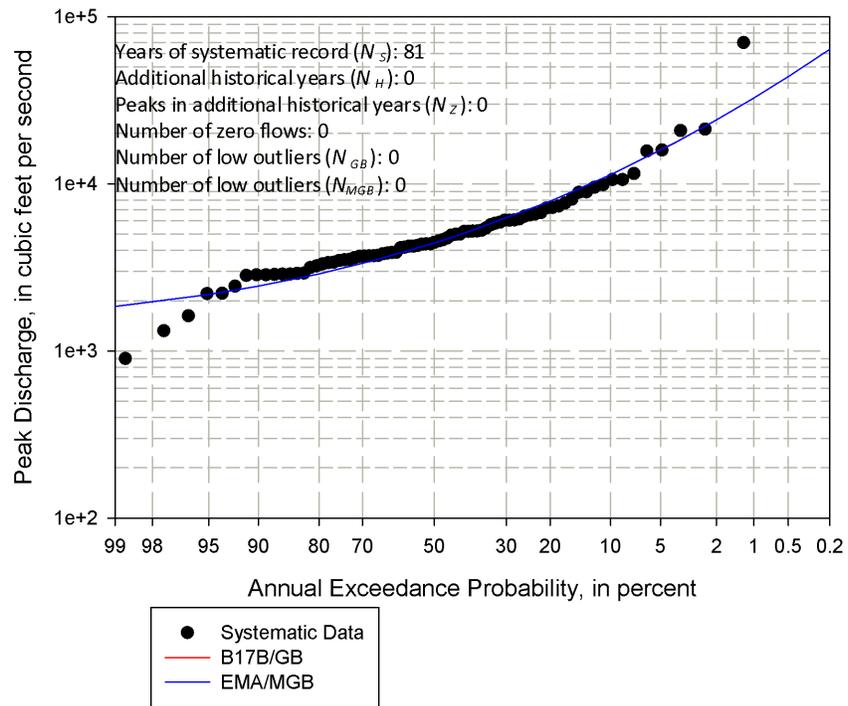


Figure 28: Site 01635500 with Systematic Data Only

Passage Creek near Buckton, VA
(Station 01635500)

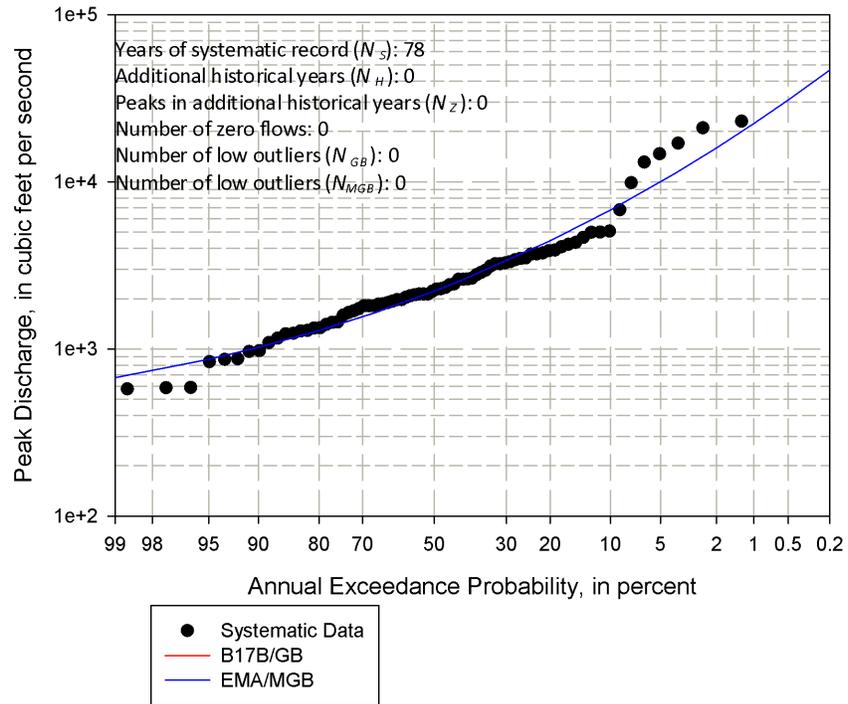


Figure 29: Site 02037500 with Systematic Data Only

James River near Richmond, VA
(Station 02037500)

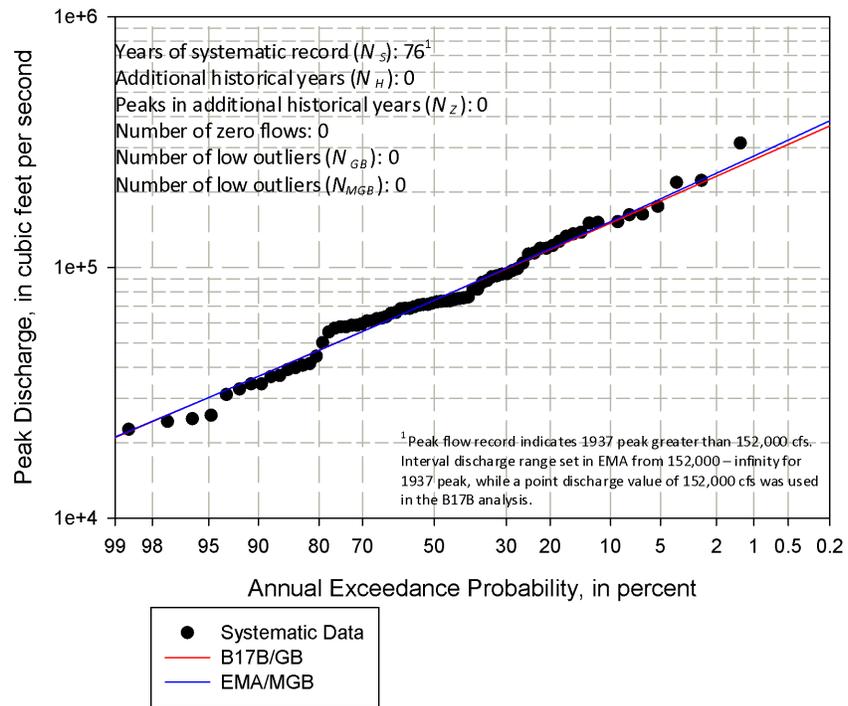


Figure 30: Site 02256500 with Systematic Data Only

Fisheating Creek at Palmdale, FL
(Station 02256500)

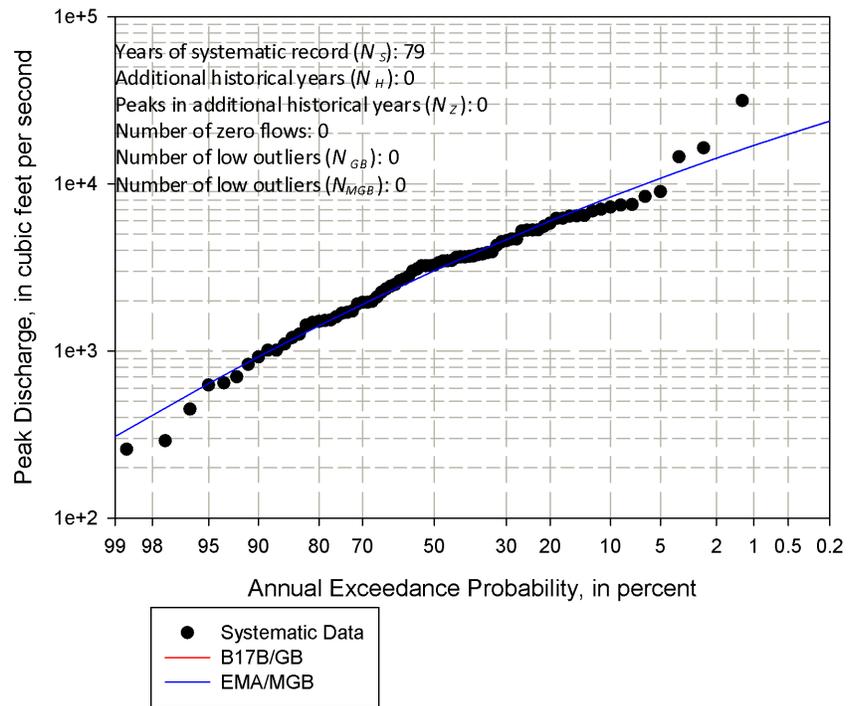


Figure 31: Site 03183500 with Systematic Data Only

Greenbrier River at Alderson, WV
(Station 03183500)

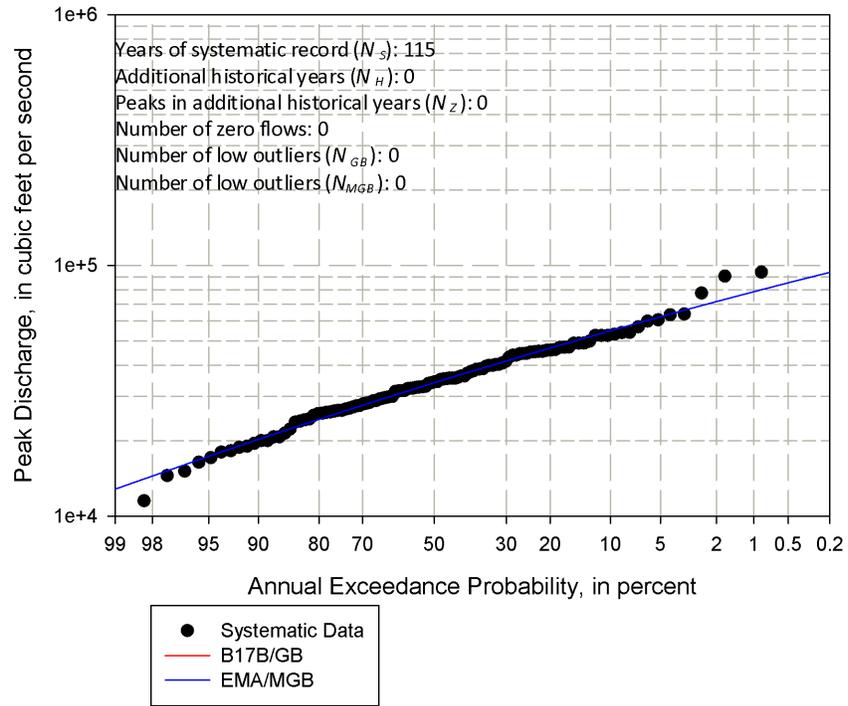


Figure 32: Site 05586500 with Systematic Data Only

Hurricane Creek near Roodhouse, IL
(Station 05586500)

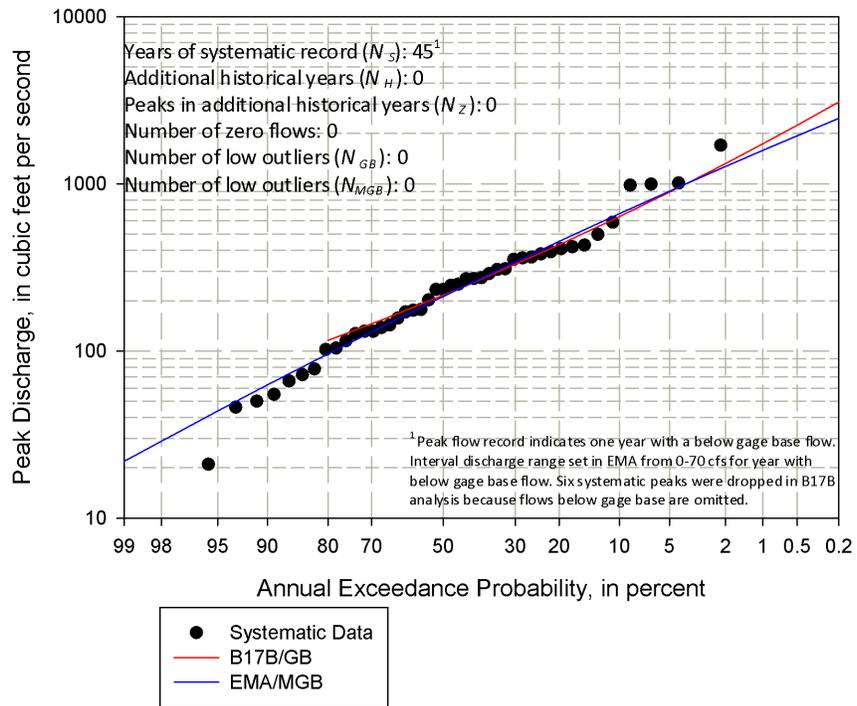


Figure 33: Site 06406000 with Systematic Data Only

Battle Creek at Hermosa, SD
(Station 06406000)

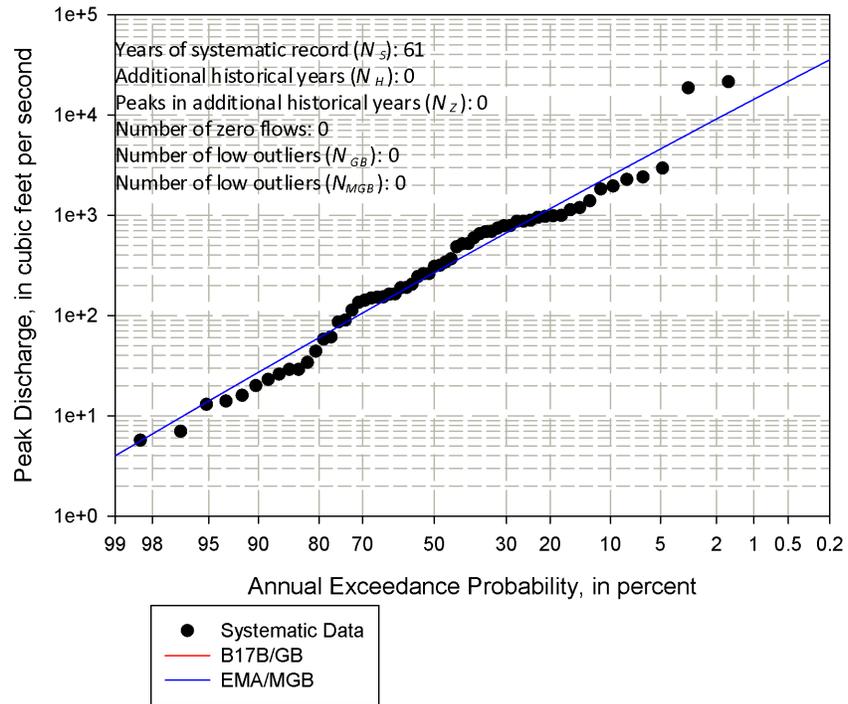


Figure 34: Site 06710500 with Systematic Data Only

Bear Creek at Morrison, CO
(Station 06710500)

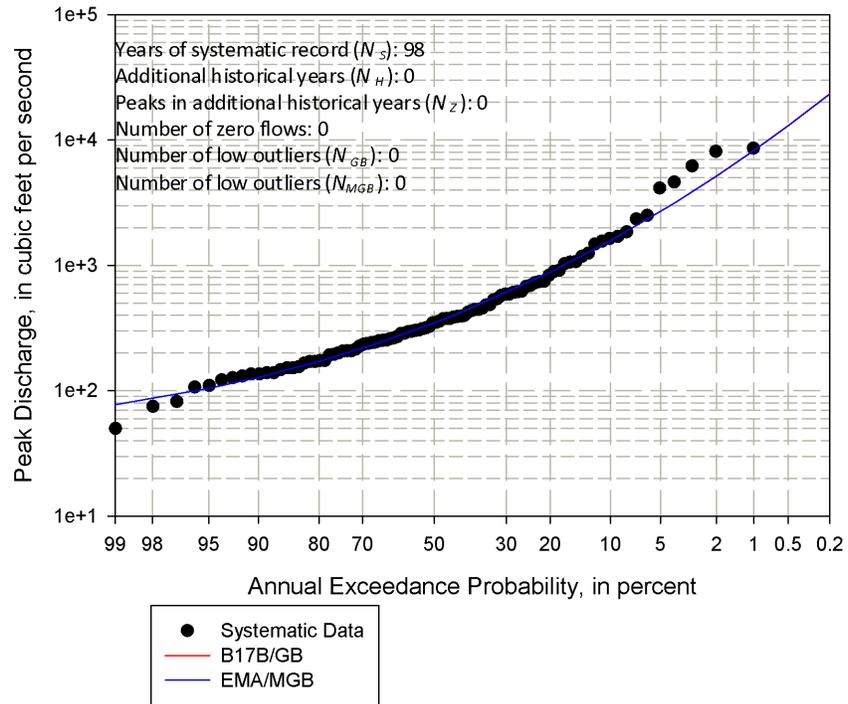


Figure 35: Site 07208500 with Systematic Data Only

Rayado Creek near Cimarron, NM
(Station 07208500)

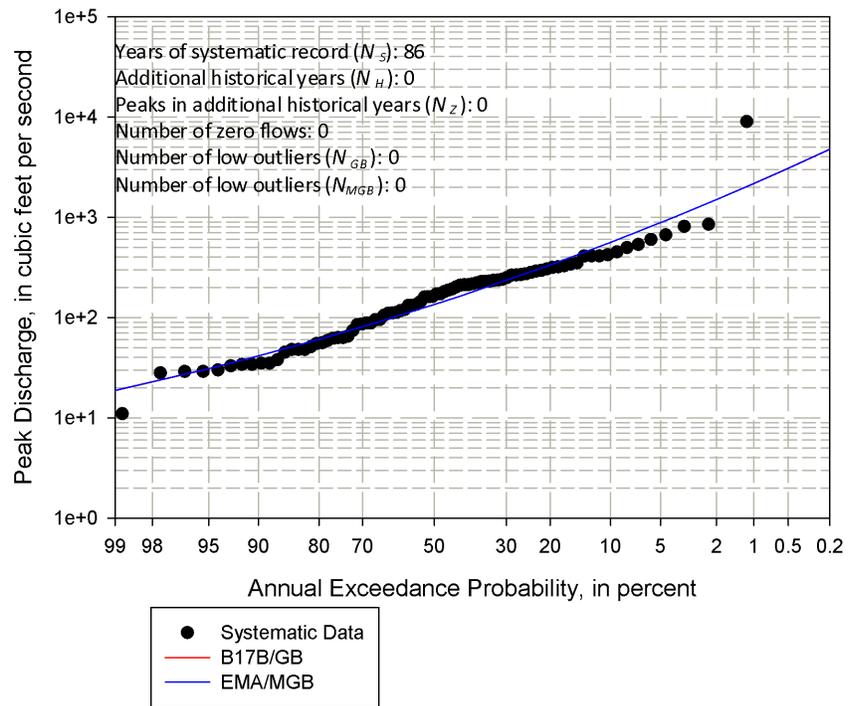


Figure 36: Site 07382000 with Systematic Data Only

Bayou Cocodrie near Clearwater, LA
(Station 07382000)

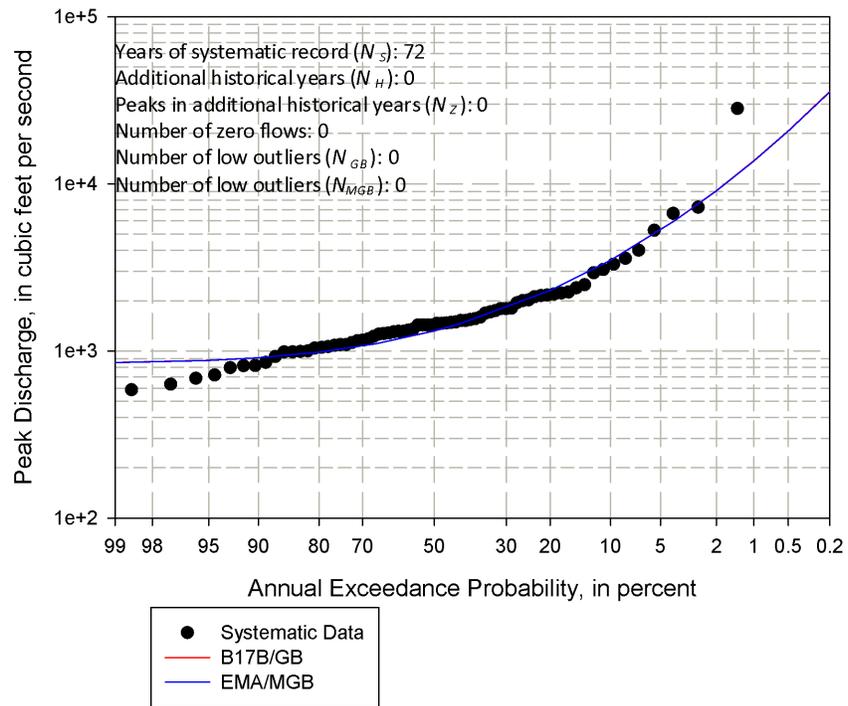


Figure 37: Site 08380500 with Systematic Data Only

Gallinas Creek near Montezuma, NM
(Station 08380500)

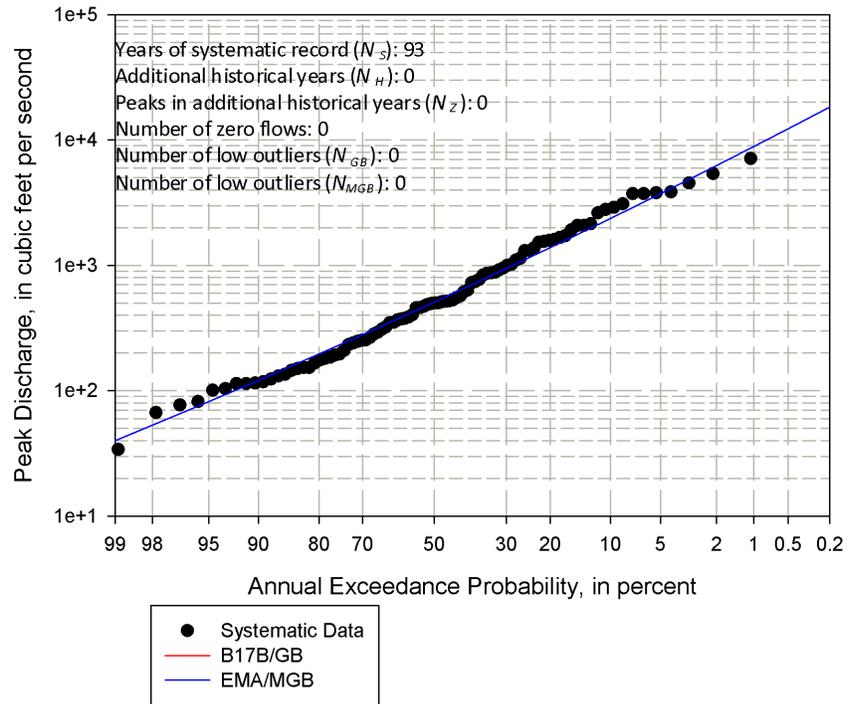


Figure 38: Site 08387000 with Systematic Data Only

Rio Ruidoso at Hollywood, NM
(Station 08387000)

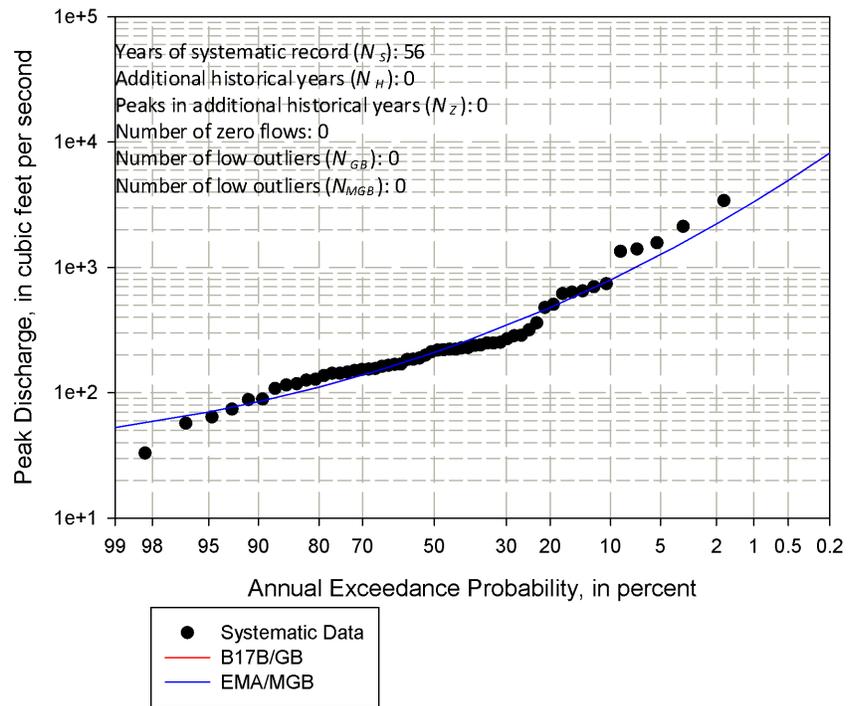


Figure 39: Site 10128500 with Systematic Data Only

Weber River near Oakley, UT
(Station 10128500)

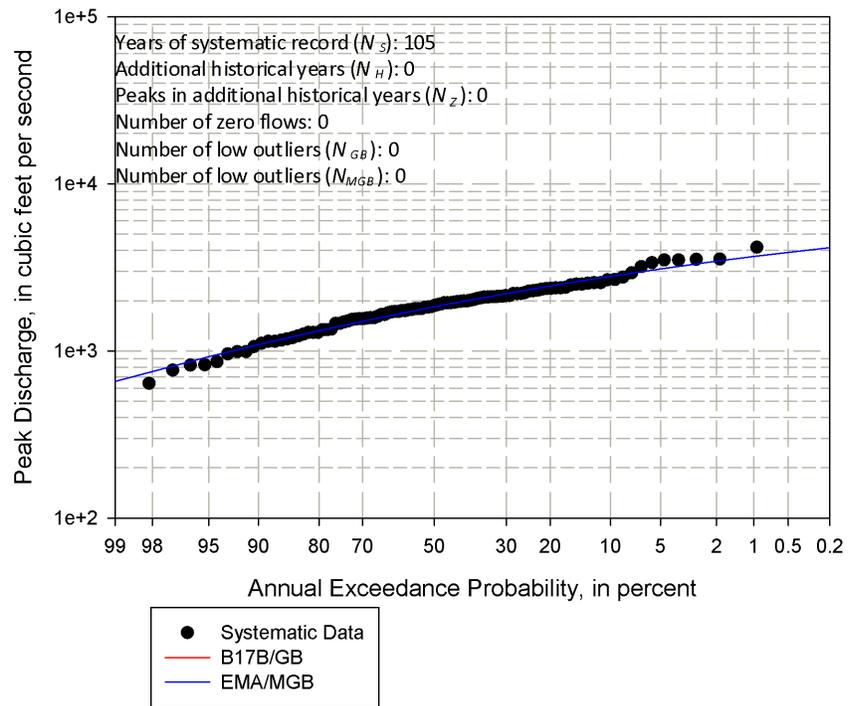


Figure 40: Site 11266500 with Systematic Data Only

Merced River at Pohono Bridge near Yosemite, CA
(Station 11266500)

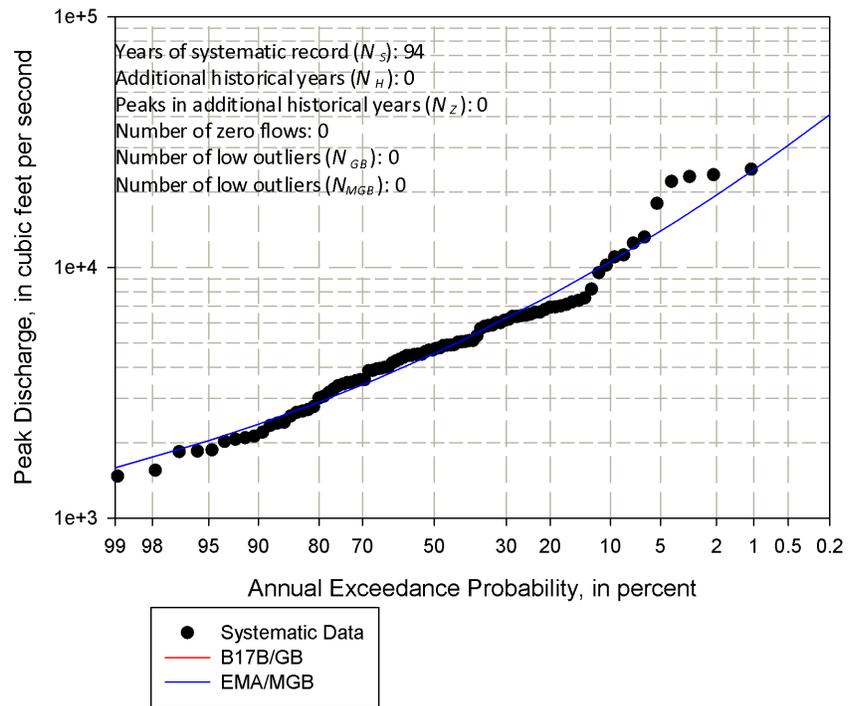


Figure 41: Site 12134500 with Systematic Data Only

Skykomish River near Gold Bar, WA
(Station 12134500)

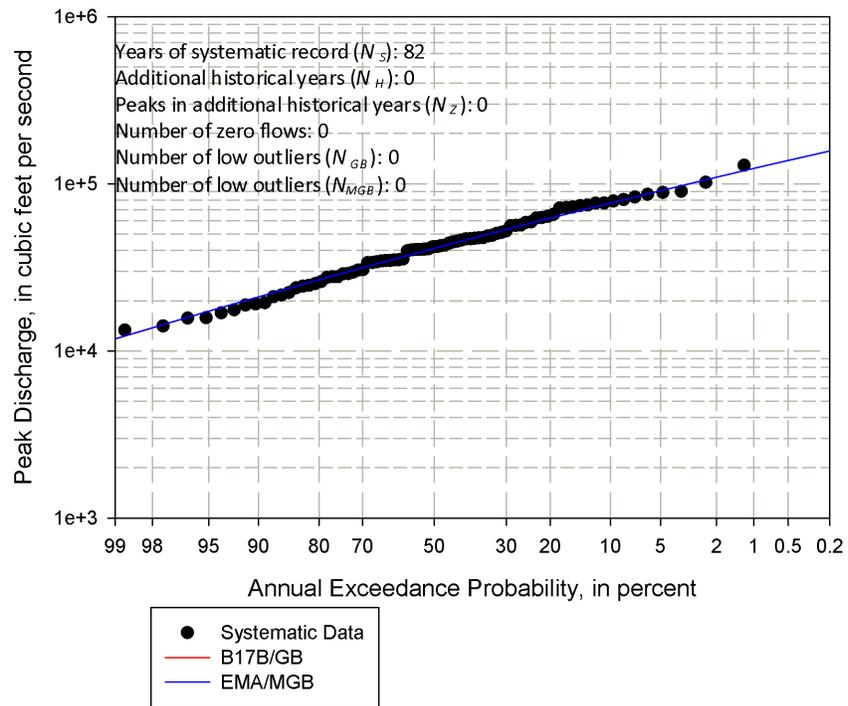


Figure 42: Site 12414500 with Systematic Data Only

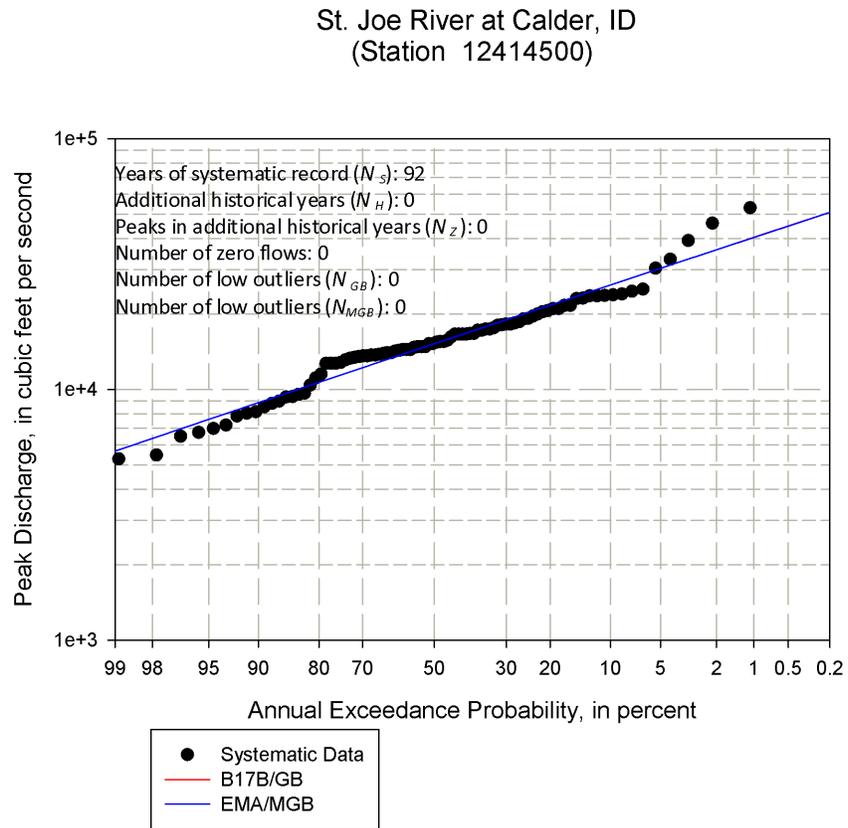


Figure 43: Site 12437950 with Systematic Data Only

East Fork Foster Creek Trib near Bridgeport, WA
(Station 12437950)

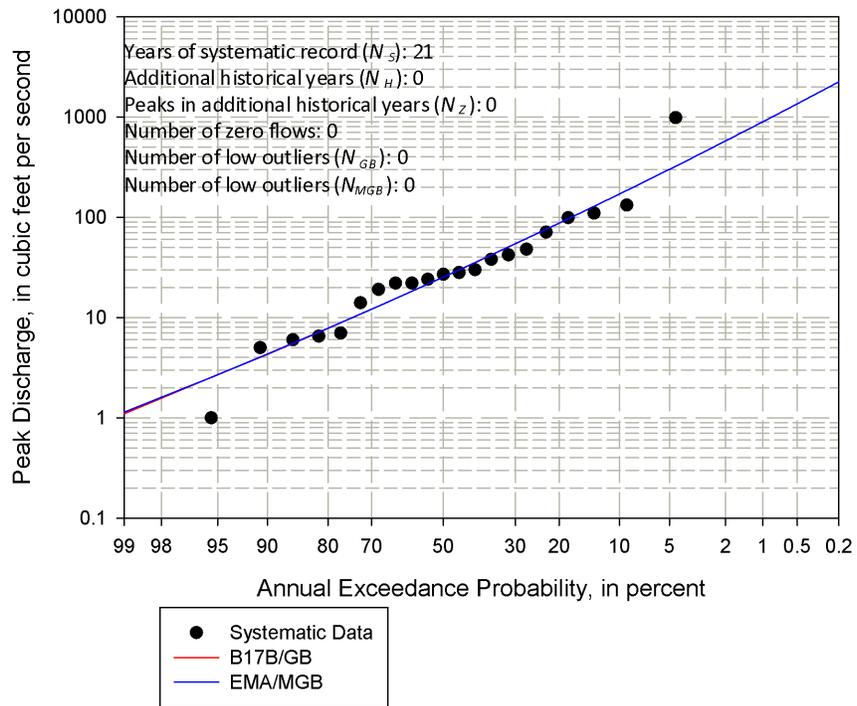


Figure 44: Site 12451000 with Systematic Data Only

Stehekin River at Stehekin, WA
(Station 12451000)

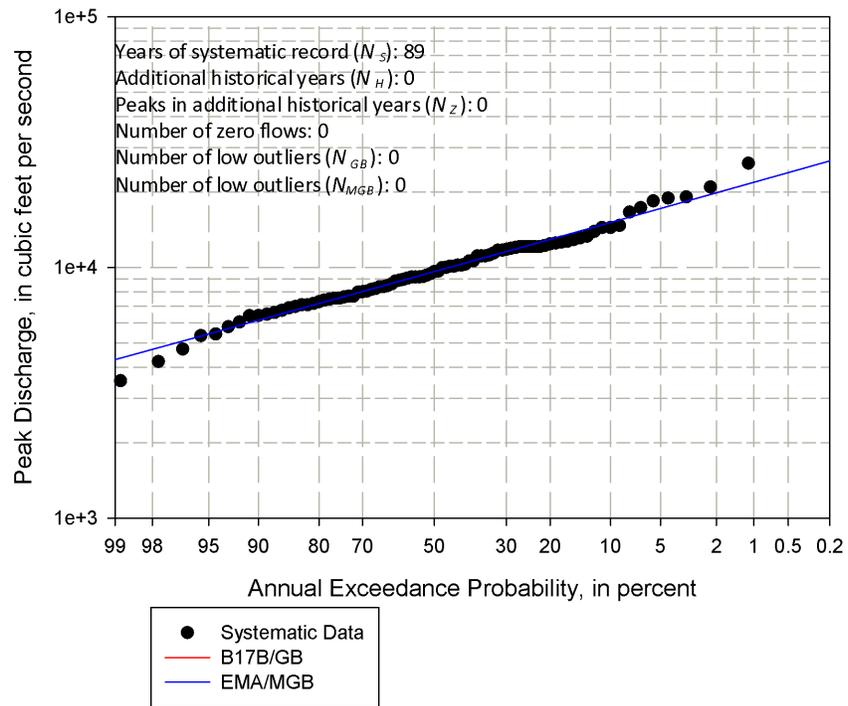


Figure 45: Site 14021000 with Systematic Data Only

Umatilla River at Pendleton, OR
(Station 14021000)

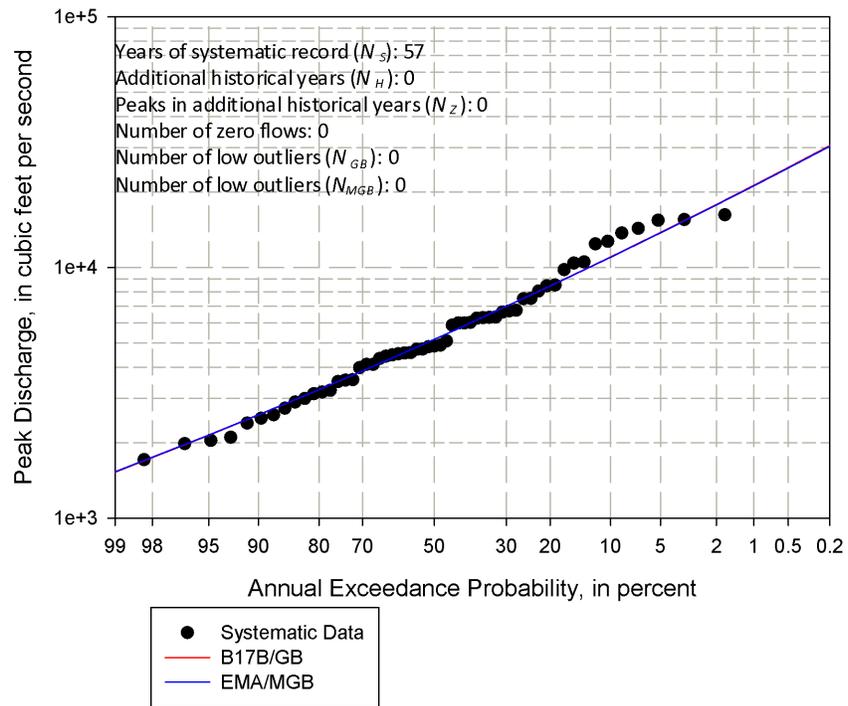


Figure 46: Site 14048000 with Systematic Data Only

John Day River at McDonald Ferry, OR
(Station 14048000)

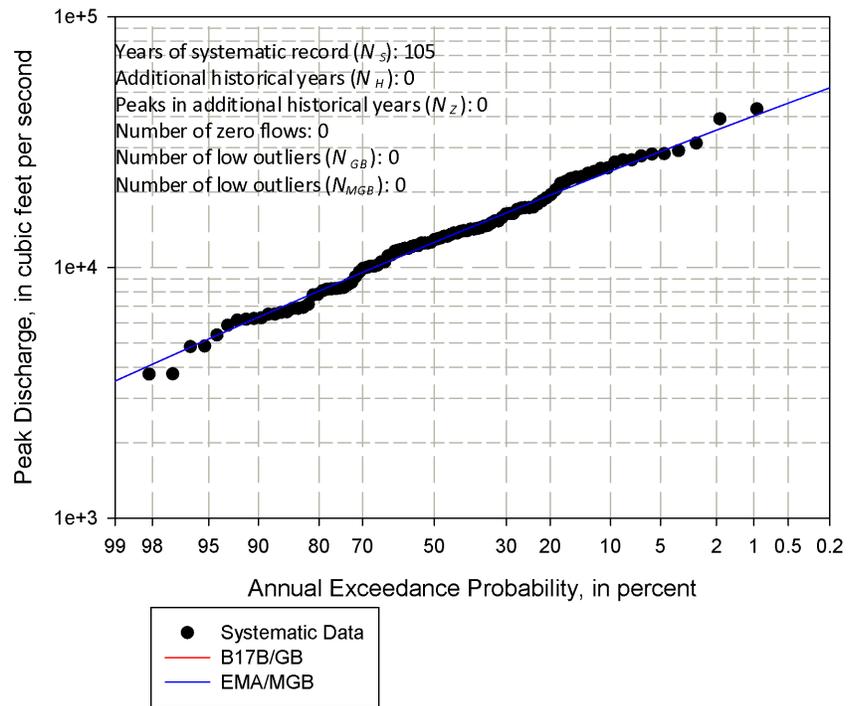


Figure 47: Site 14137000 with Systematic Data Only

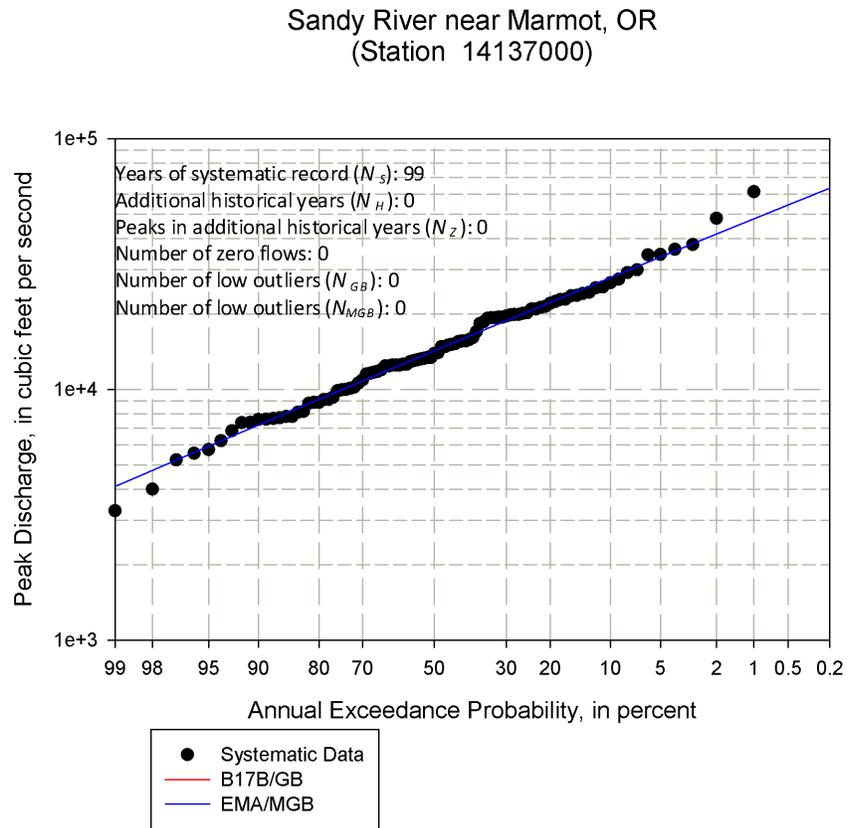


Figure 48: Site 15072000 with Systematic Data Only

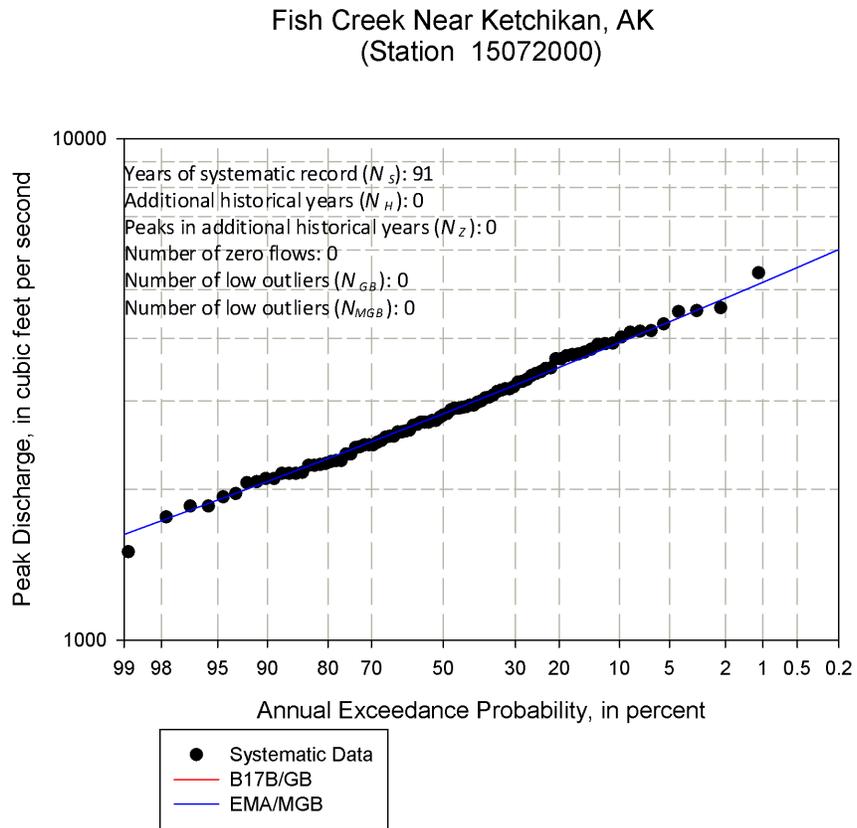


Figure 49: Site 16518000 with Systematic Data Only

West Wailuaiki Stream near Keanae, HI
(Station 16518000)

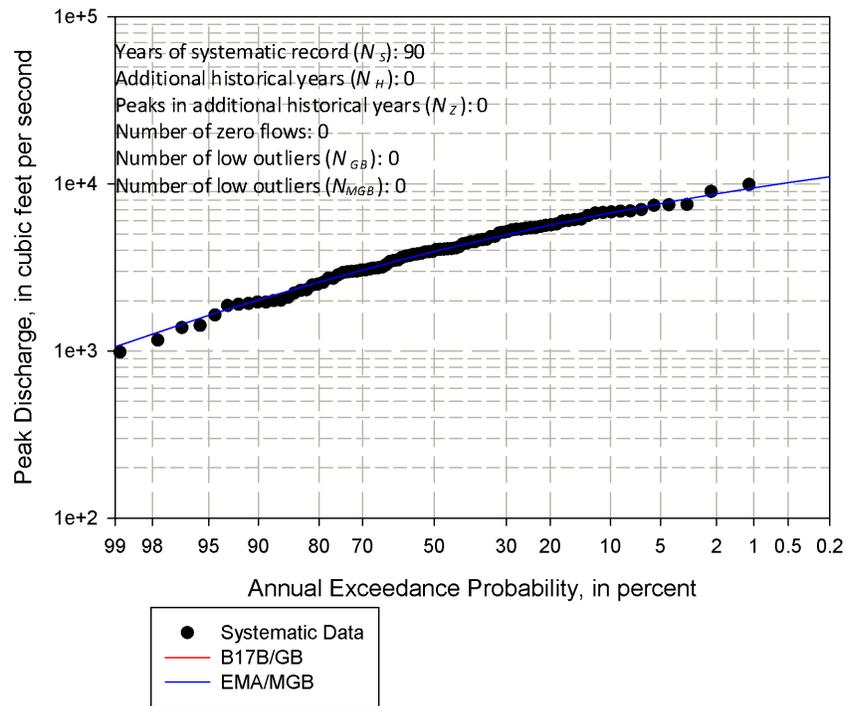
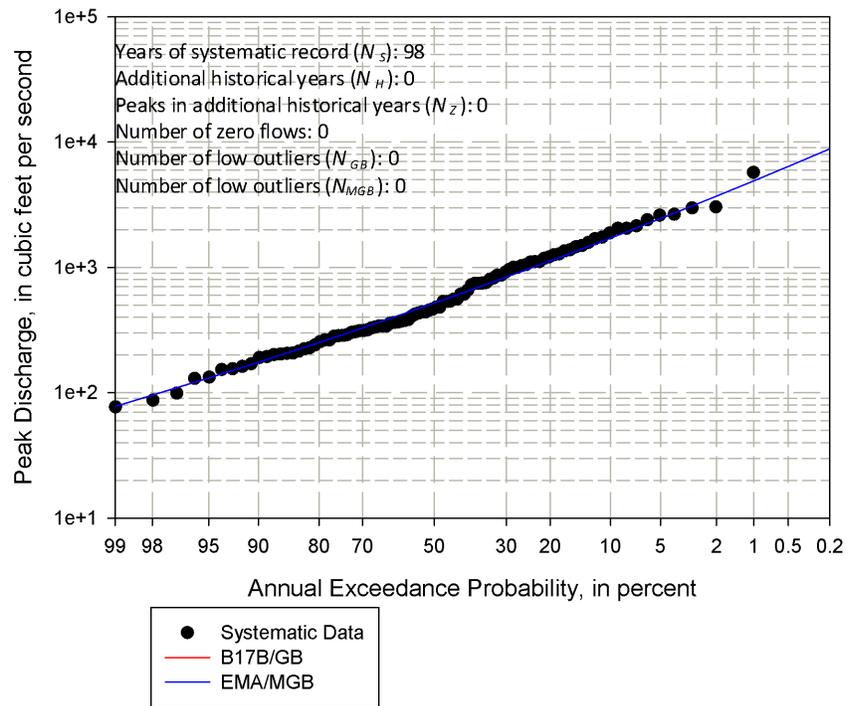


Figure 50: Site 16587000 with Systematic Data Only

Honopou Stream near Huelo Maui, HI
(Station 16587000)



B.2 Sites with Historical Information

Figure 51: Site 01350000 with Systematic and Historical Data

Schoharie Creek at Prattsville, NY
(Station 01350000)

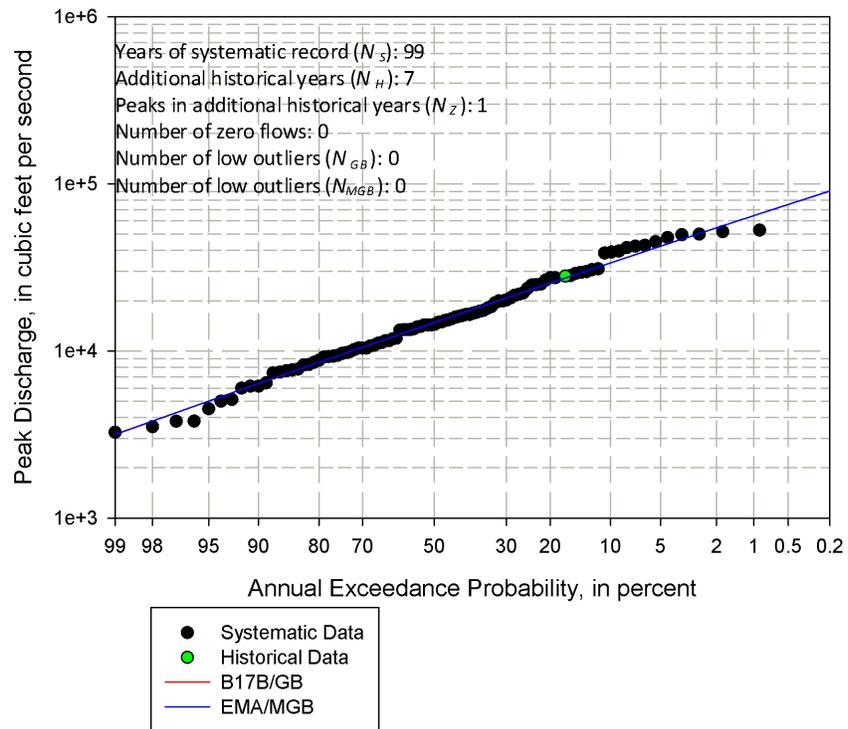


Figure 52: Site 01562000 with Systematic and Historical Data

Raystown Branch Juniata River at Saxton, PA
(Station 01562000)

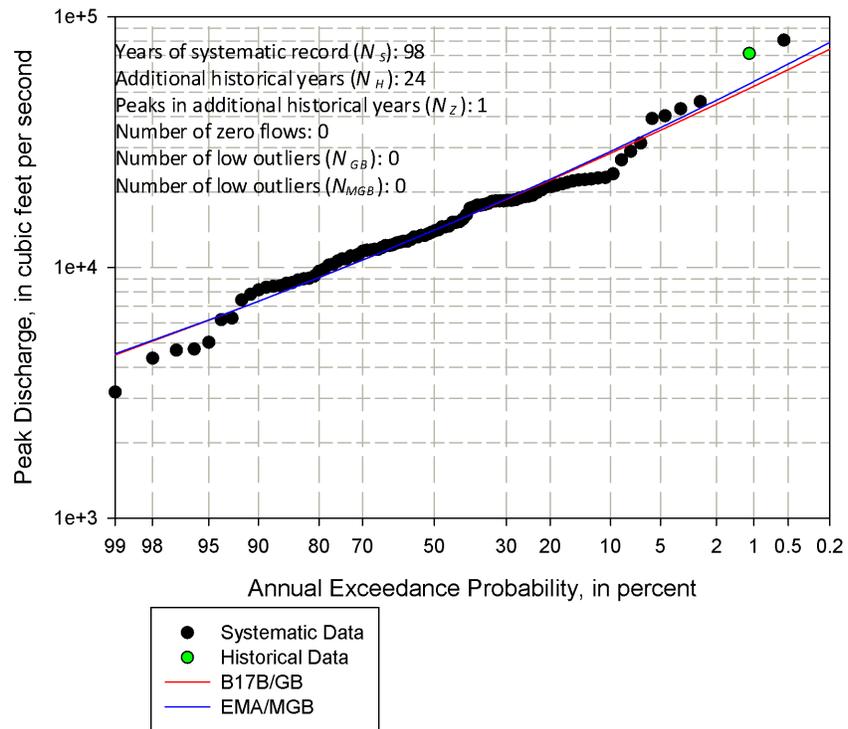


Figure 53: Site 01636500 with Systematic and Historical Data

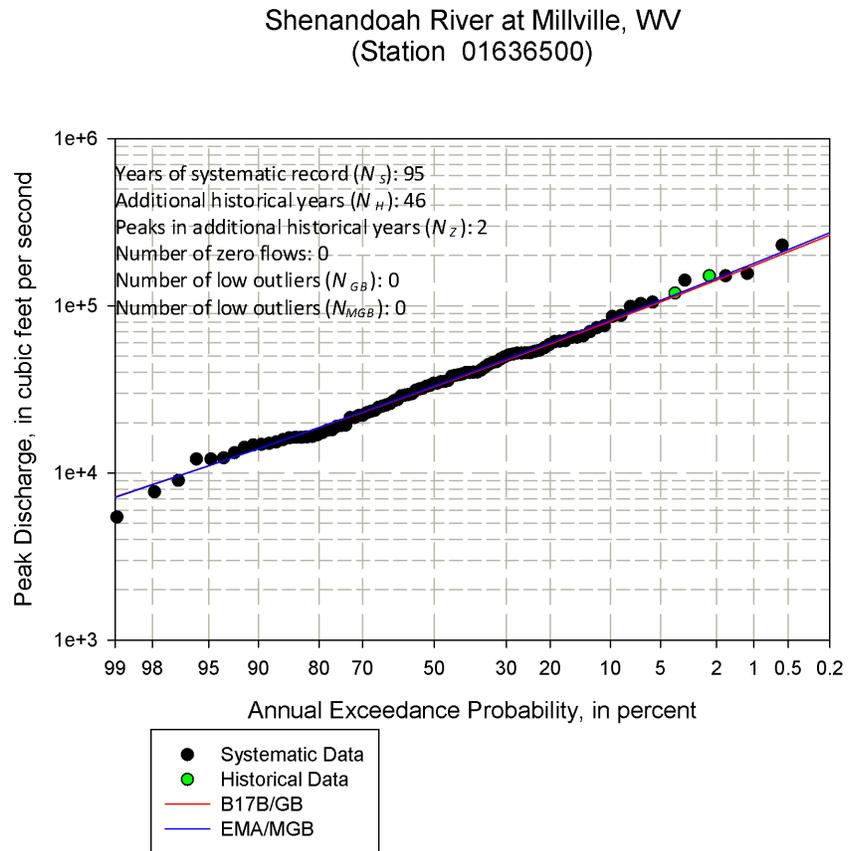


Figure 54: Site 02138500 with Systematic and Historical Data

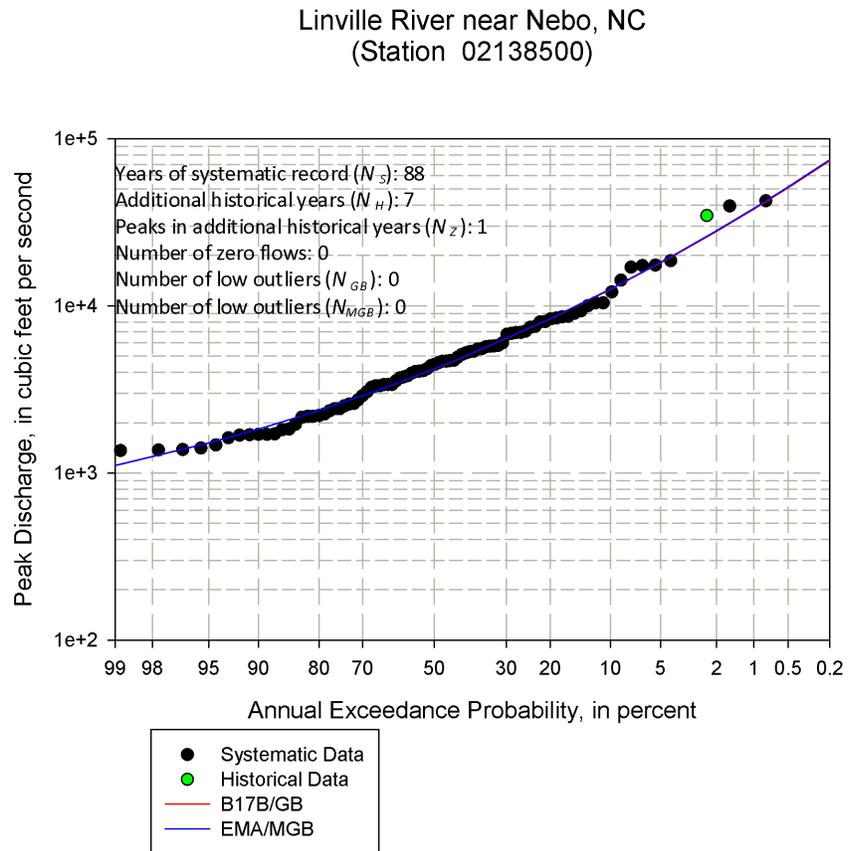


Figure 55: Site 03011020 with Systematic and Historical Data

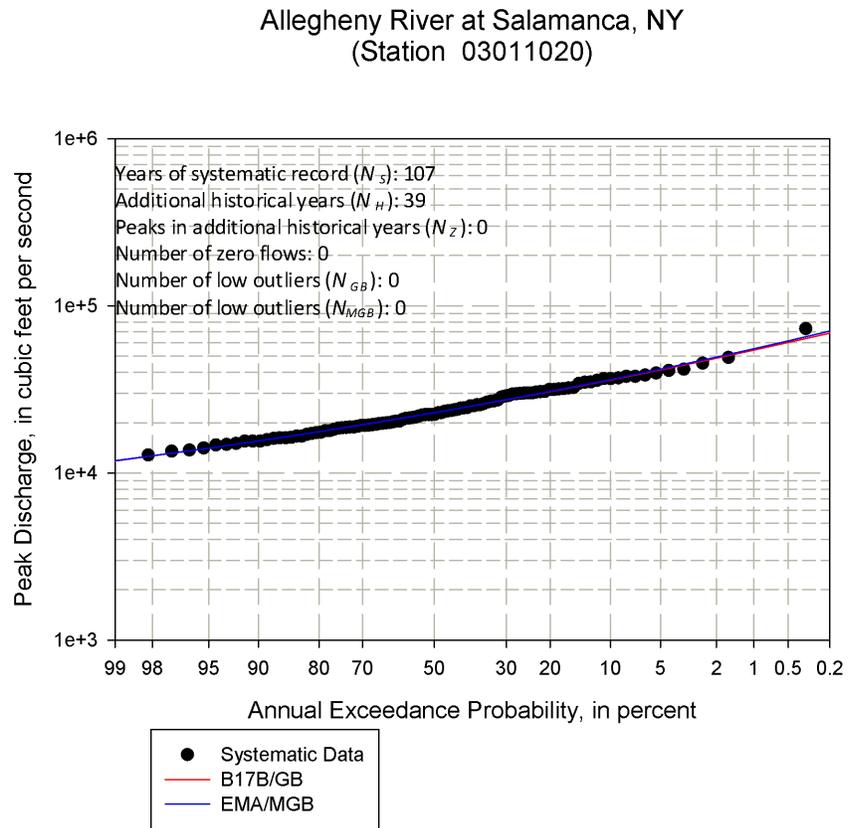


Figure 56: Site 03051000 with Systematic and Historical Data

Tygart Valley River at Belington, WV
(Station 03051000)

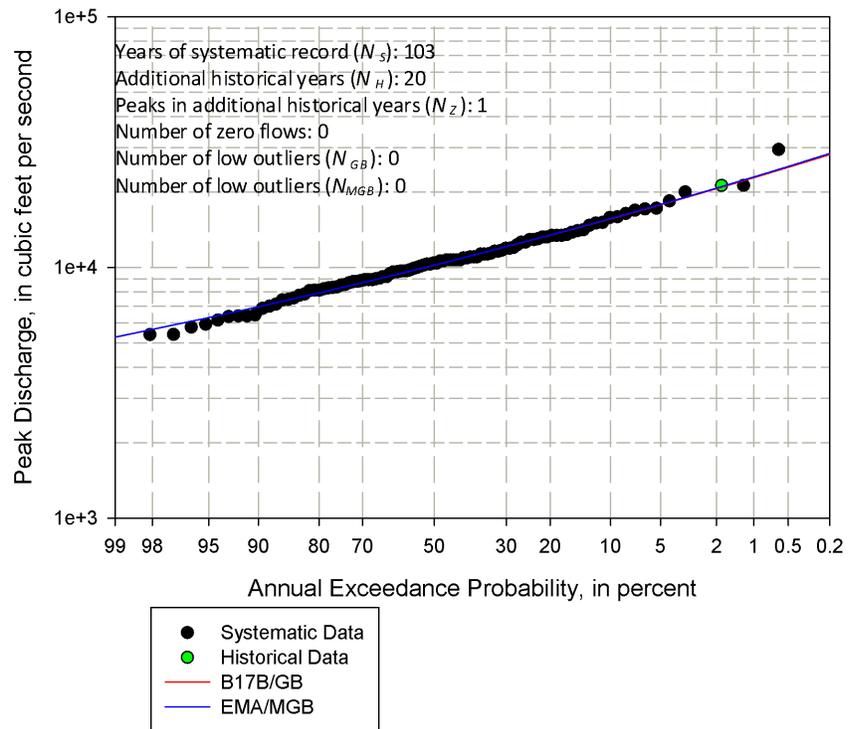


Figure 57: Site 03159500 with Systematic and Historical Data

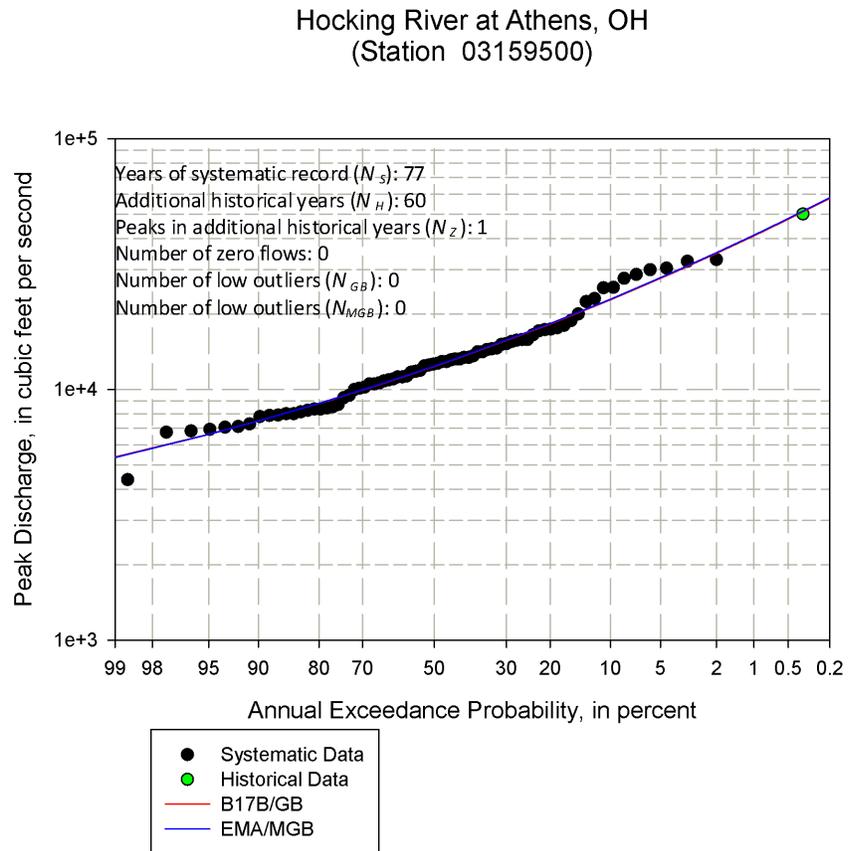


Figure 58: Site 03550000 with Systematic and Historical Data

Valley River at Tomotla, NC
(Station 03550000)

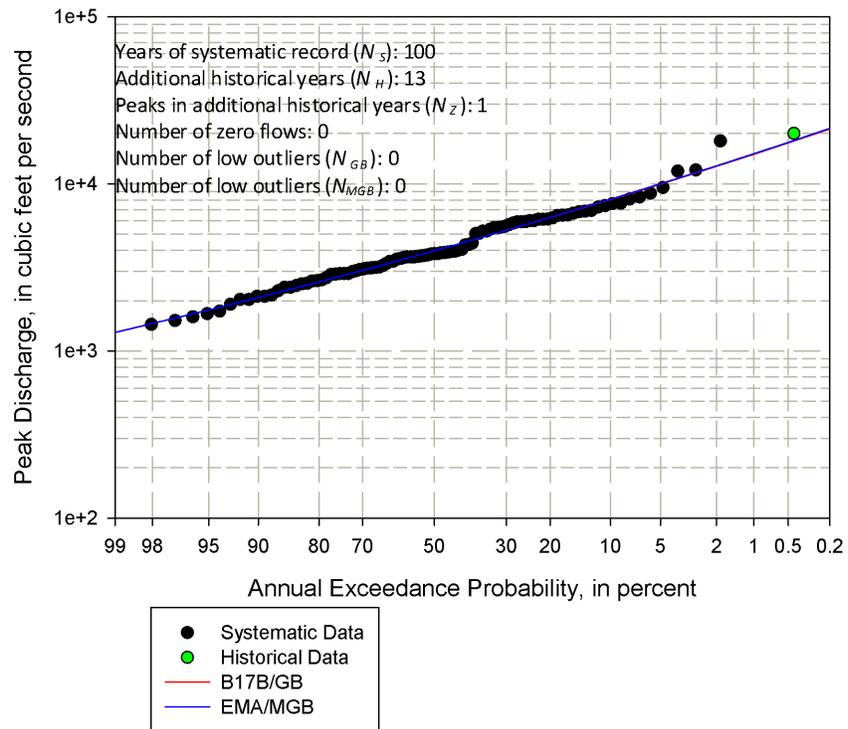


Figure 59: Site 03558000 with Systematic and Historical Data

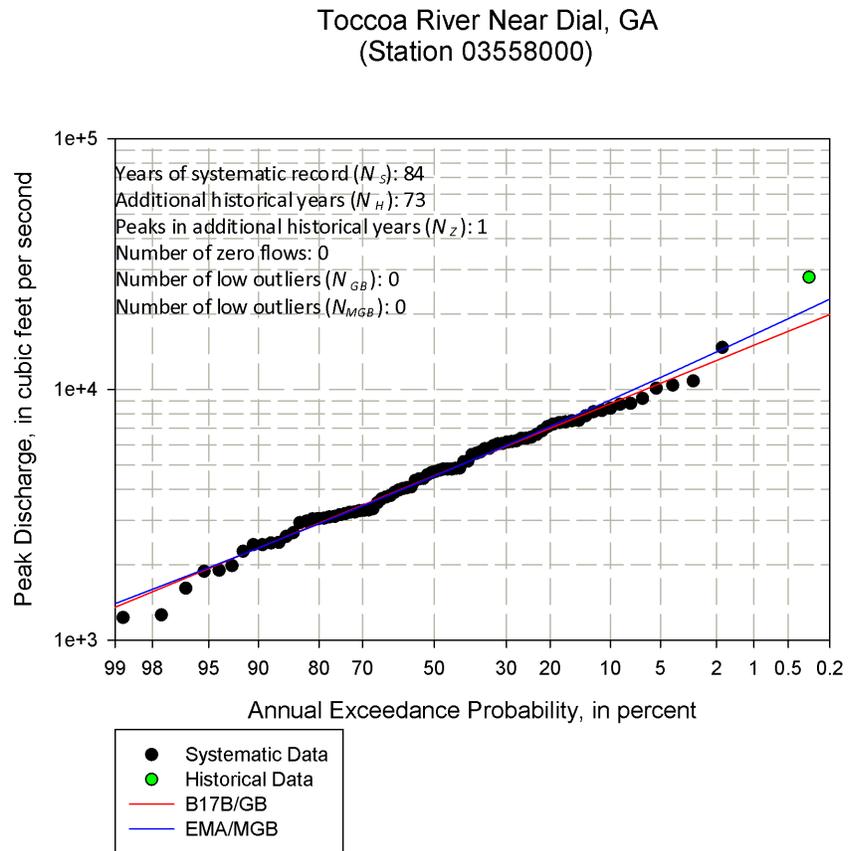


Figure 60: Site 03606500 with Systematic and Historical Data

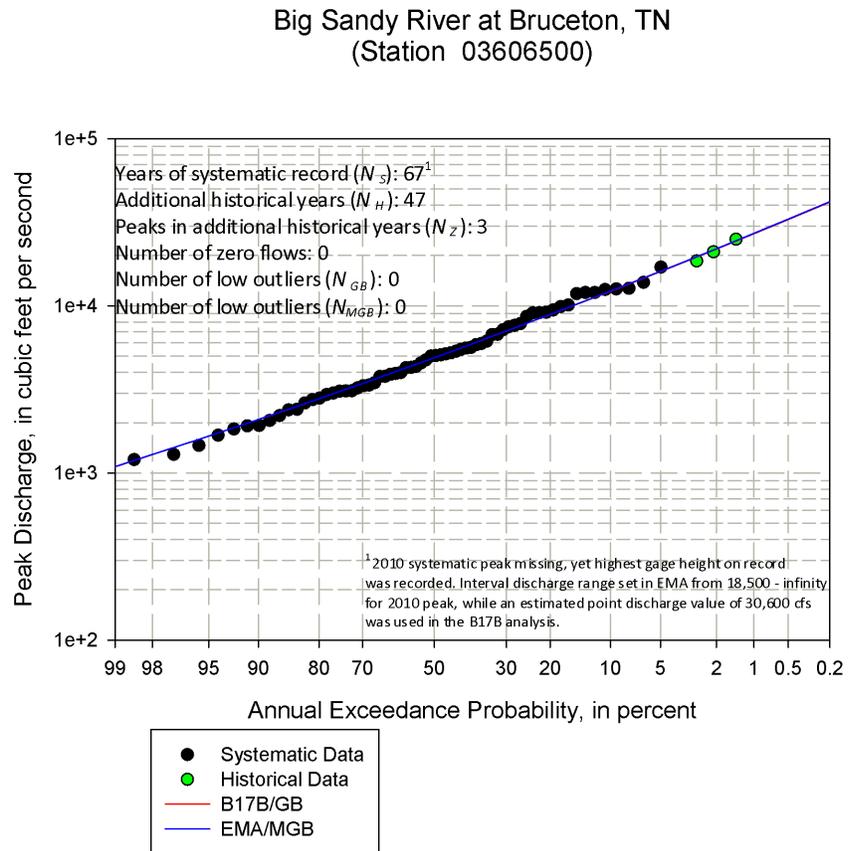


Figure 61: Site 04293500 with Systematic and Historical Data

Missisquoi River Near East Berkshire, VT
(Station 04293500)

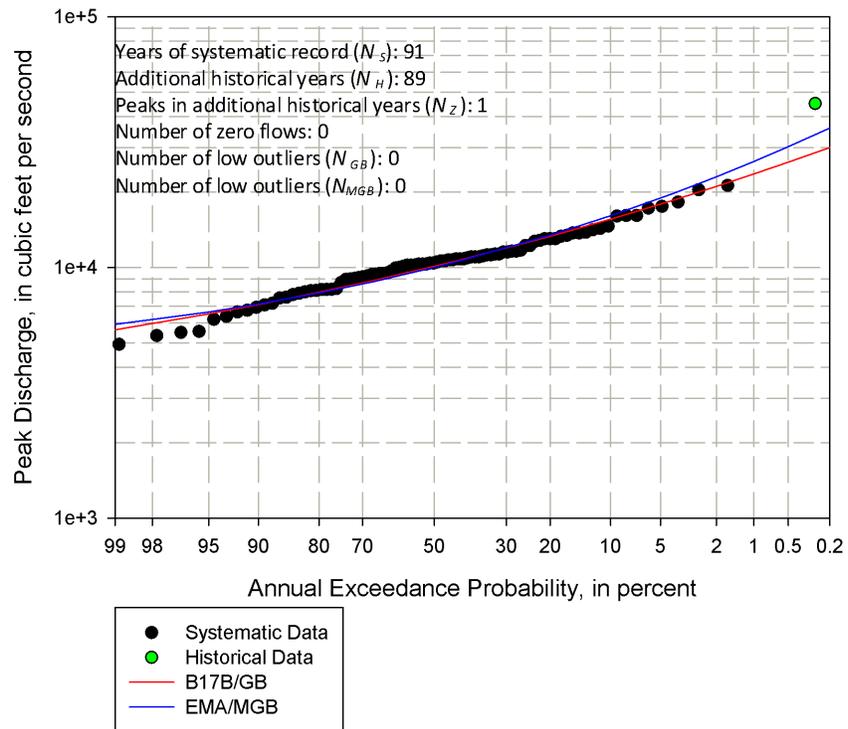


Figure 62: Site 06216500 with Systematic and Historical Data

Pryor Creek near Billings, MT
(Station 06216500)

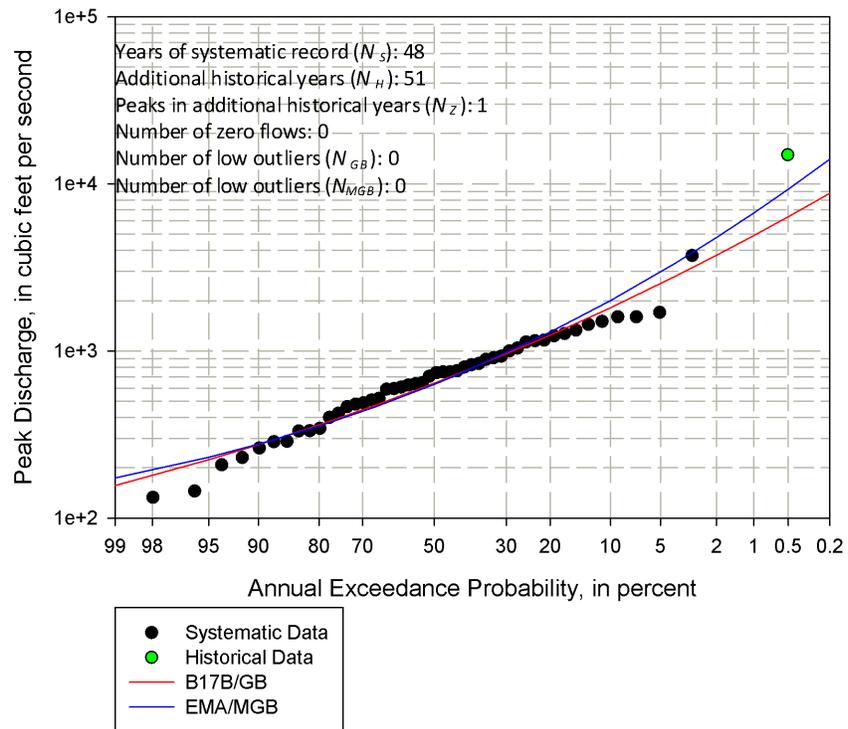


Figure 63: Site 06600500 with Systematic and Historical Data

Floyd River at James, IA
(Station 06600500)

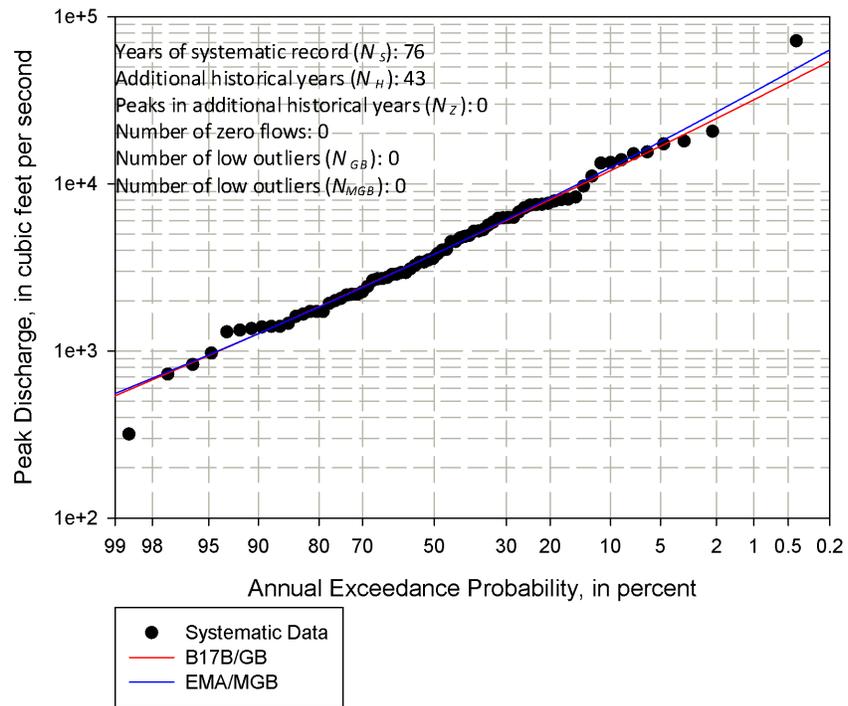


Figure 64: Site 06898000 with Systematic and Historical Data

Thompson River at Davis City, IA
(Station 06898000)

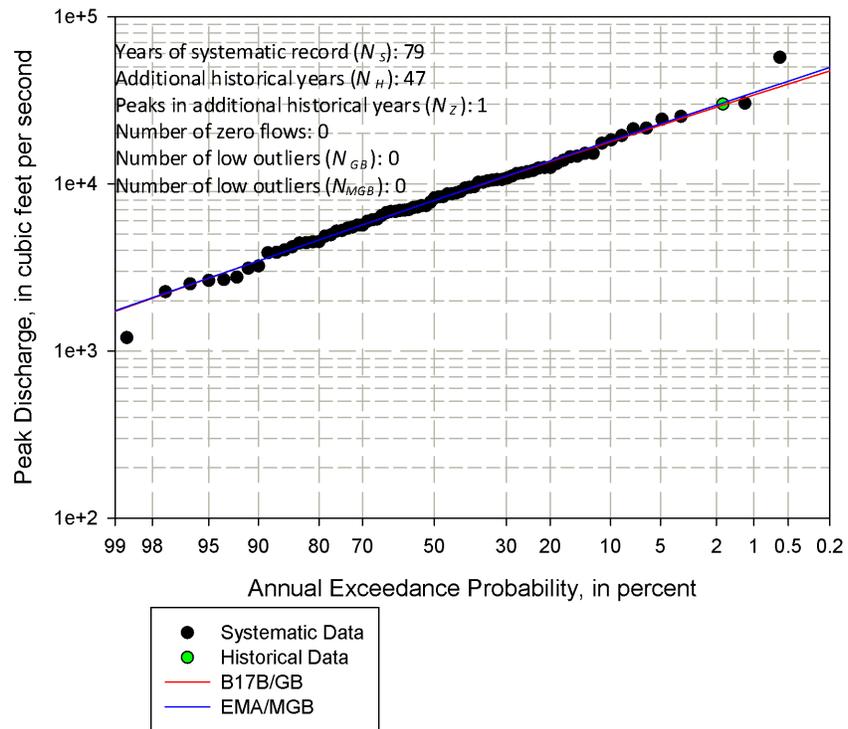


Figure 65: Site 07067000 with Systematic and Historical Data

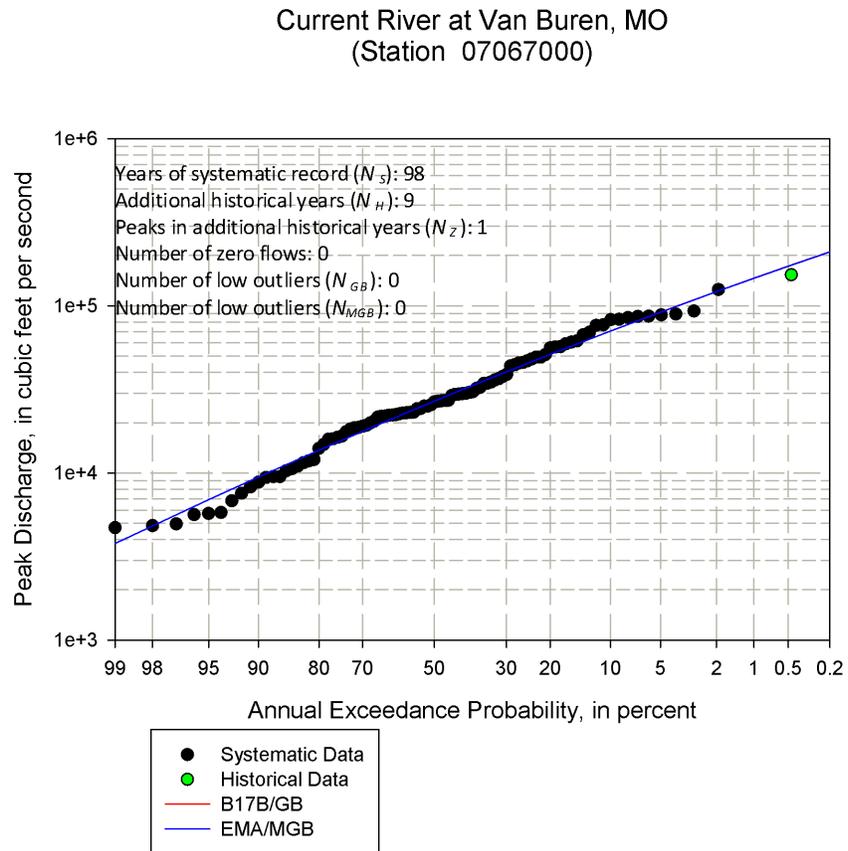


Figure 66: Site 08167000 with Systematic and Historical Data

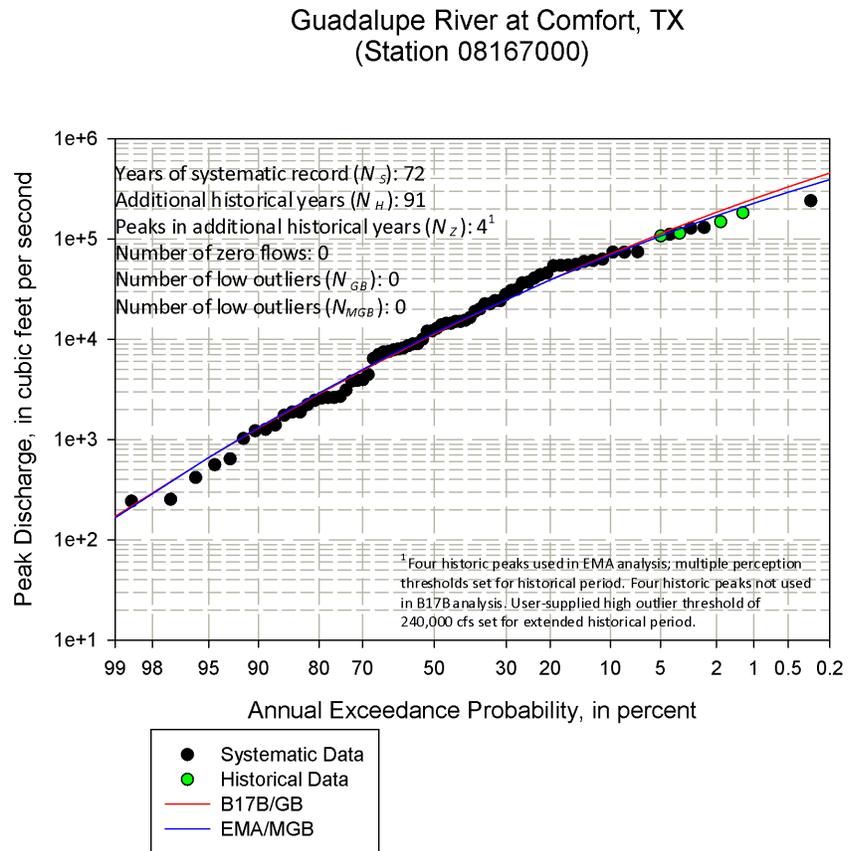


Figure 67: Site 08378500 with Systematic and Historical Data

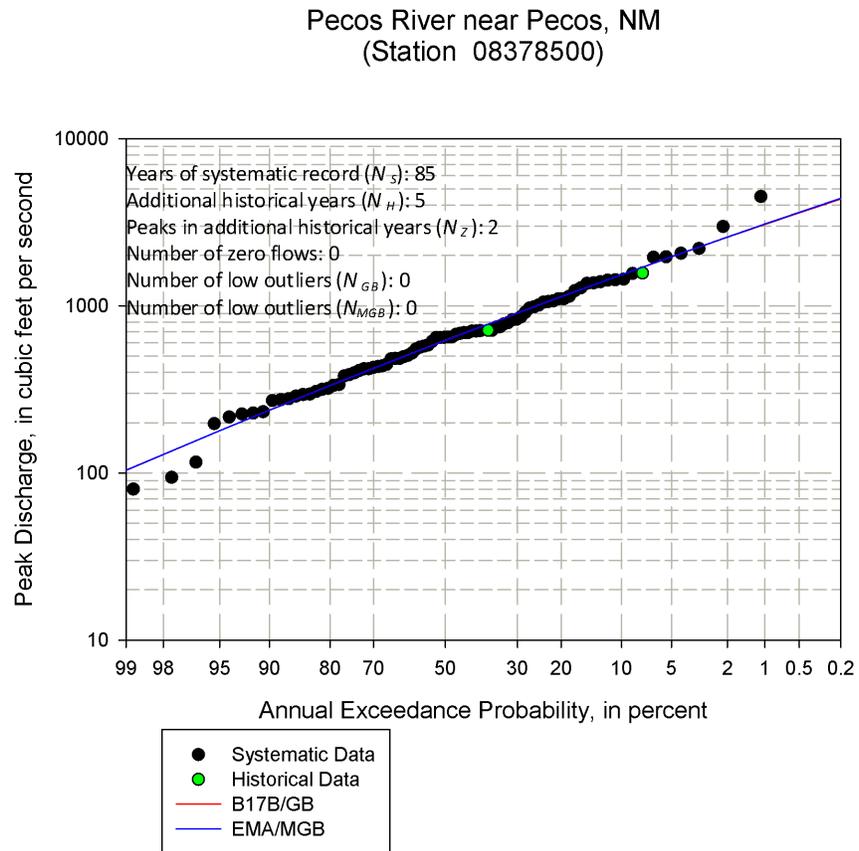


Figure 68: Site 09482500 with Systematic and Historical Data

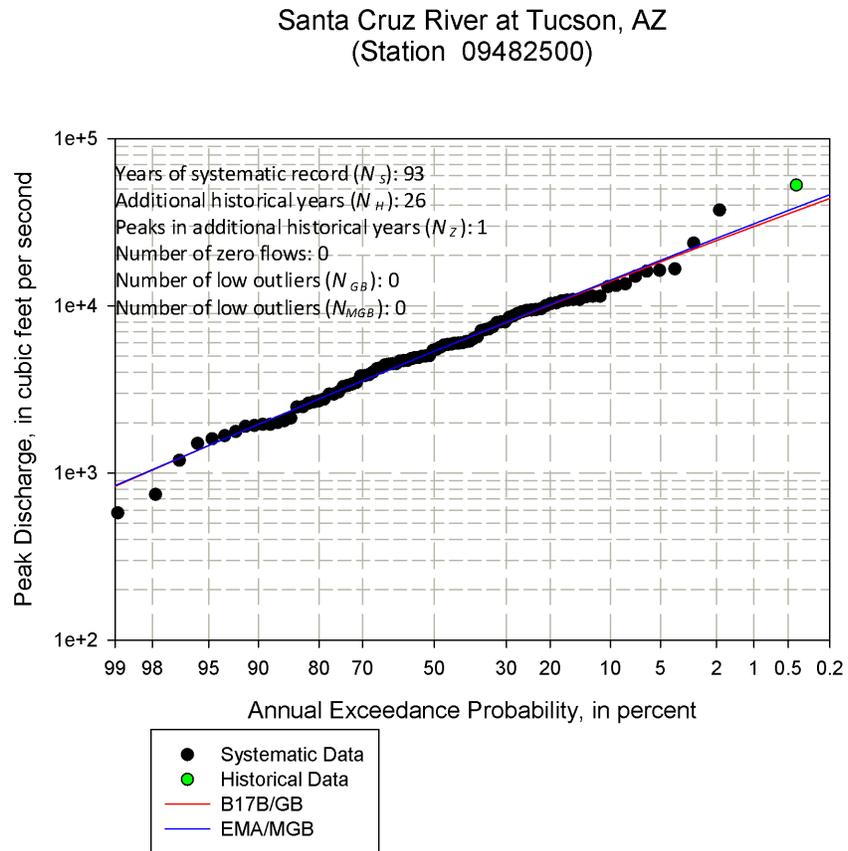
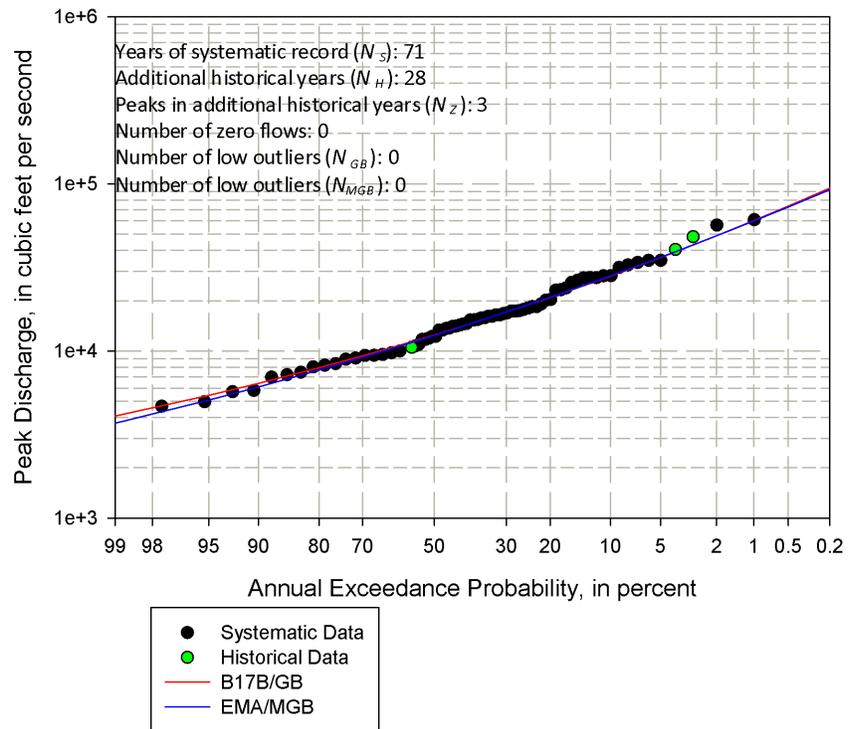


Figure 69: Site 12413000 with Systematic and Historical Data

NF Coeur D Alene River at Enaville, ID
(Station 12413000)



B.3 Sites with Low Outliers

Figure 70: Site 01668000 with Low Outliers; no historical information

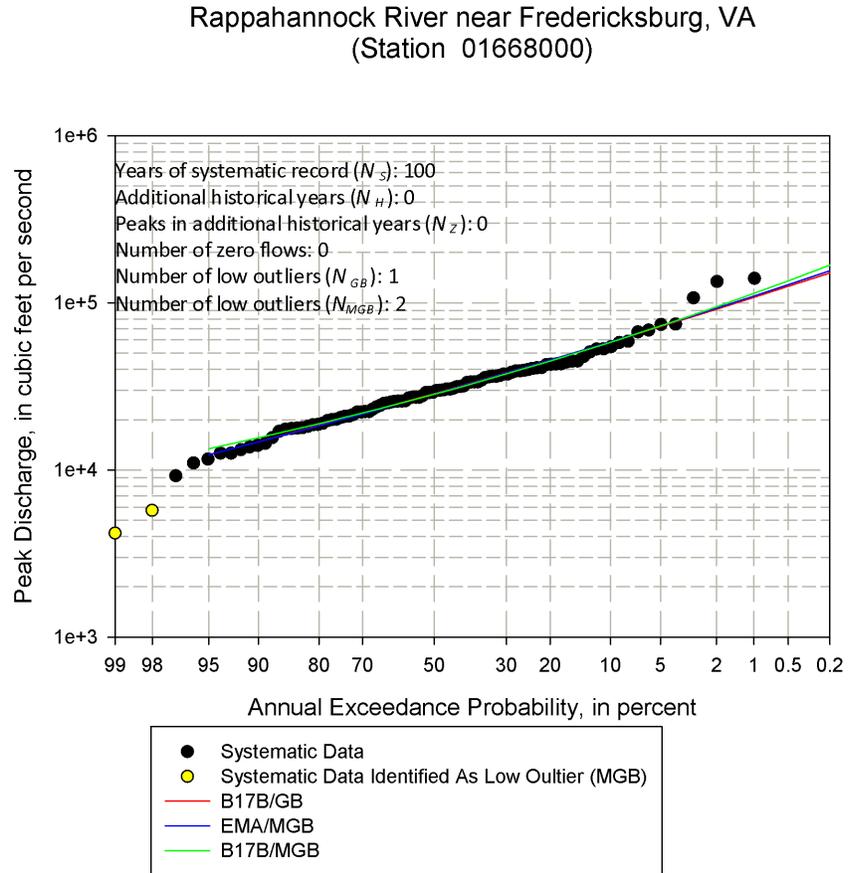


Figure 71: Site 03345500 with Low Outliers; no historical information

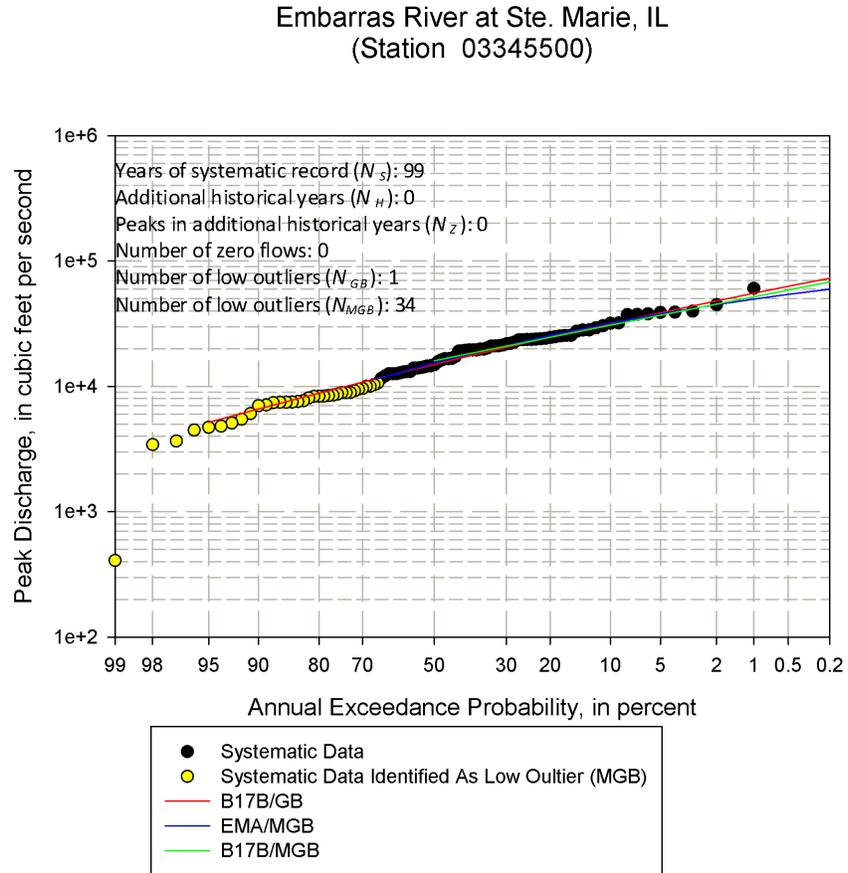


Figure 72: Site 05572000 with Low Outliers; no historical information

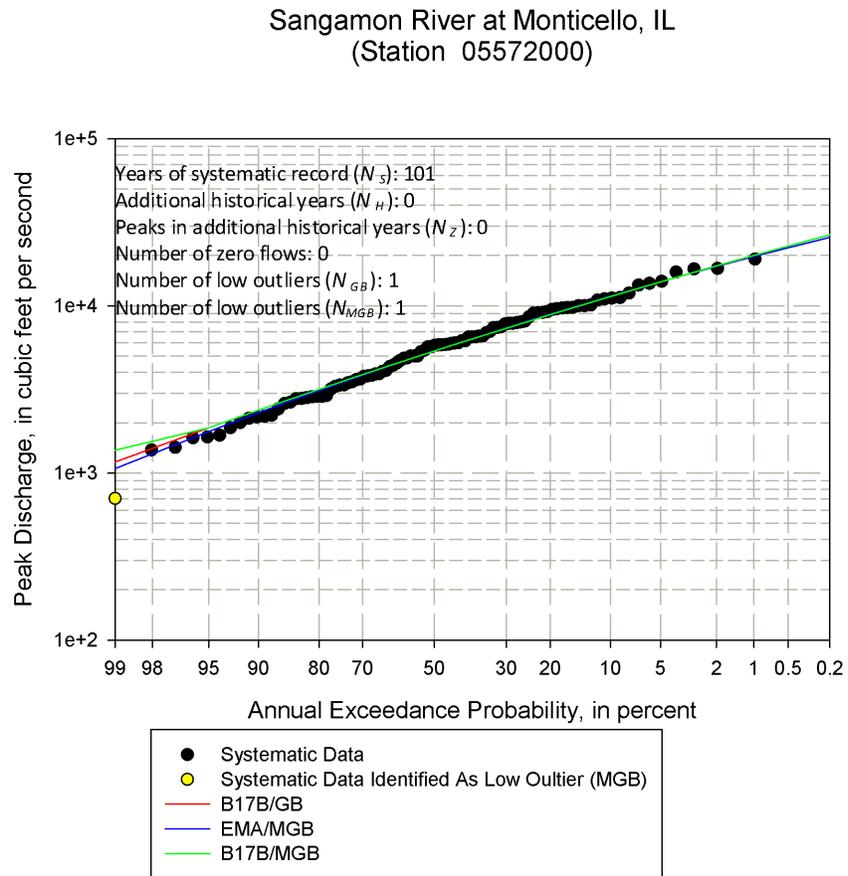


Figure 73: Site 06176500 with Low Outliers; no historical information

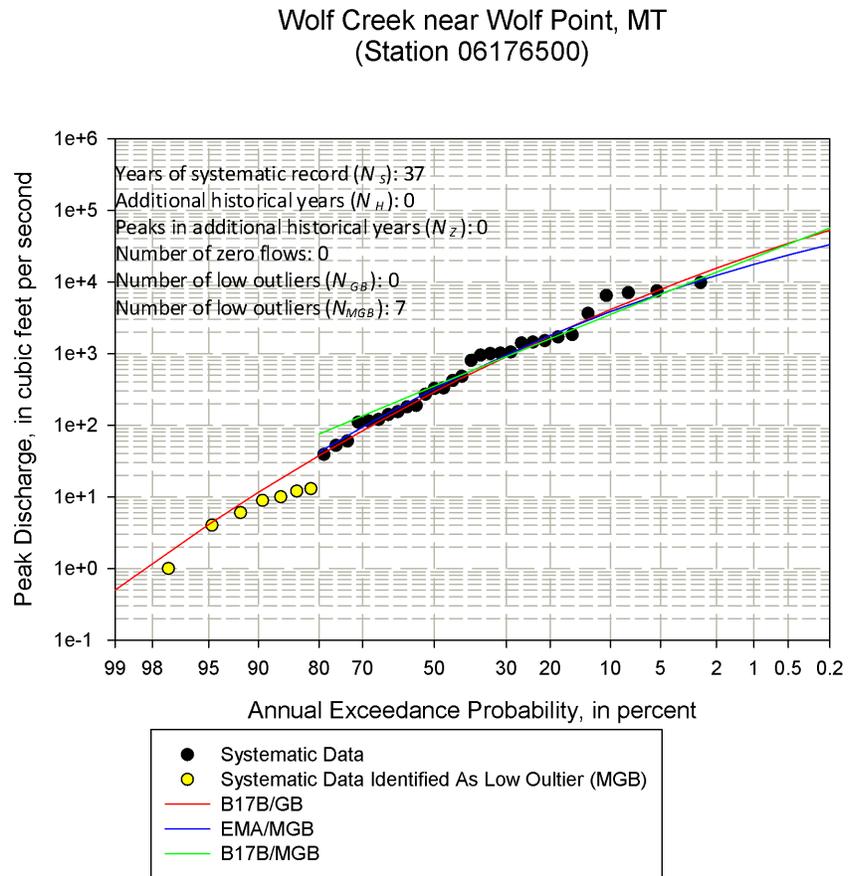


Figure 74: Site 07203000 with Low Outliers; no historical information

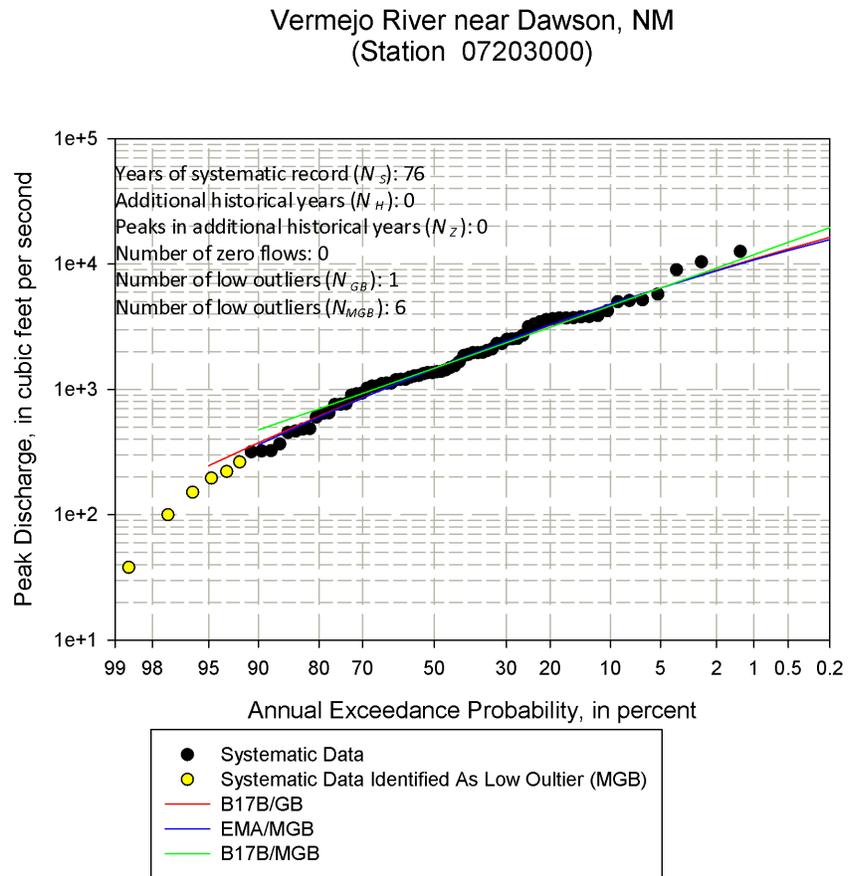


Figure 75: Site 08133500 with Low Outliers; no historical information

North Concho River at Sterling City, TX
(Station 08133500)

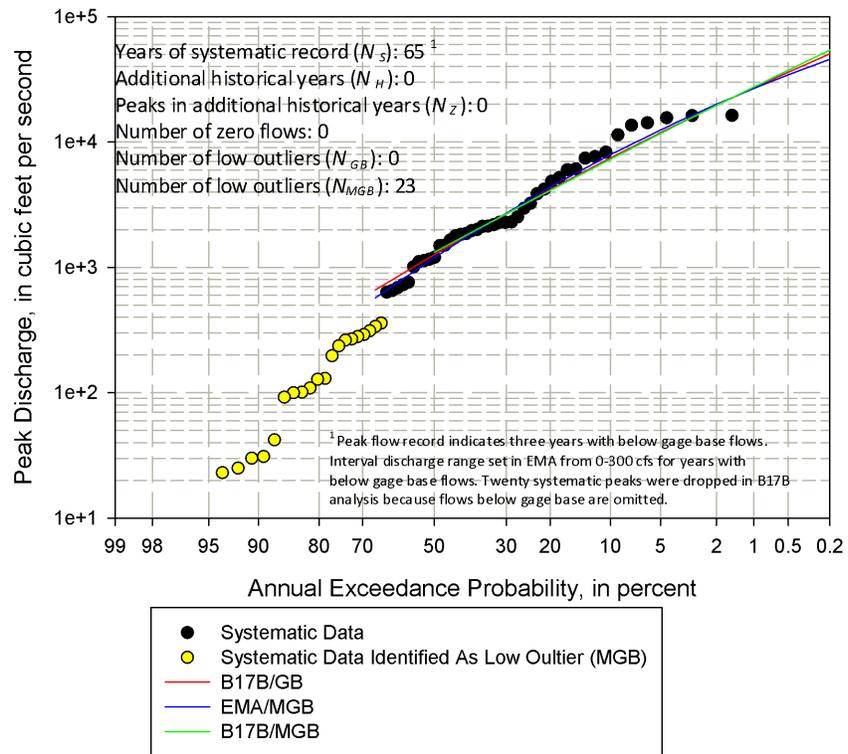


Figure 76: Site 08150000 with Low Outliers; no historical information

Llano River near Junction, TX
(Station 08150000)

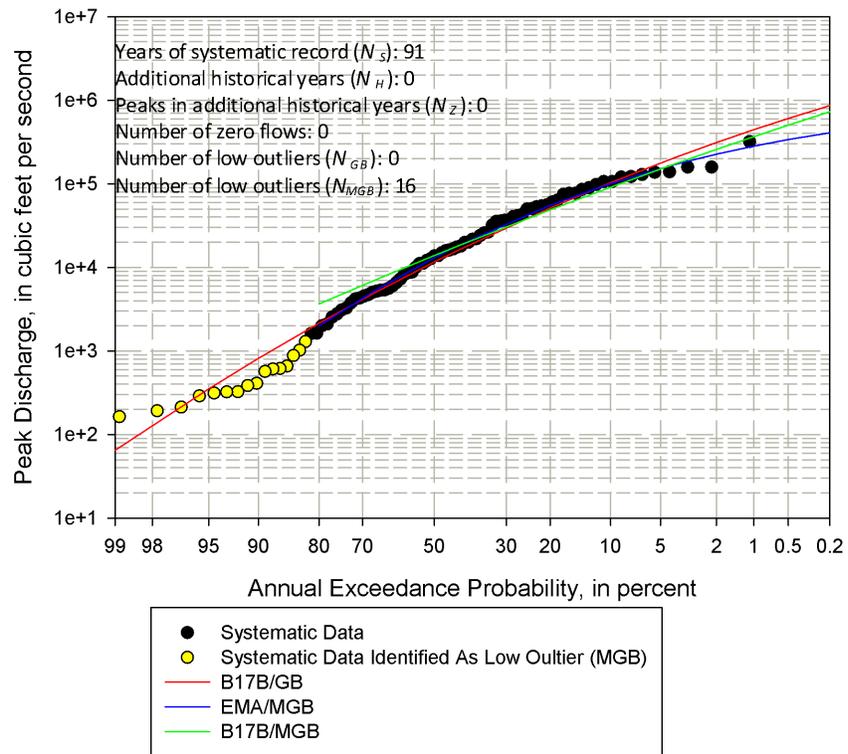


Figure 77: Site 08189500 with Low Outliers; no historical information

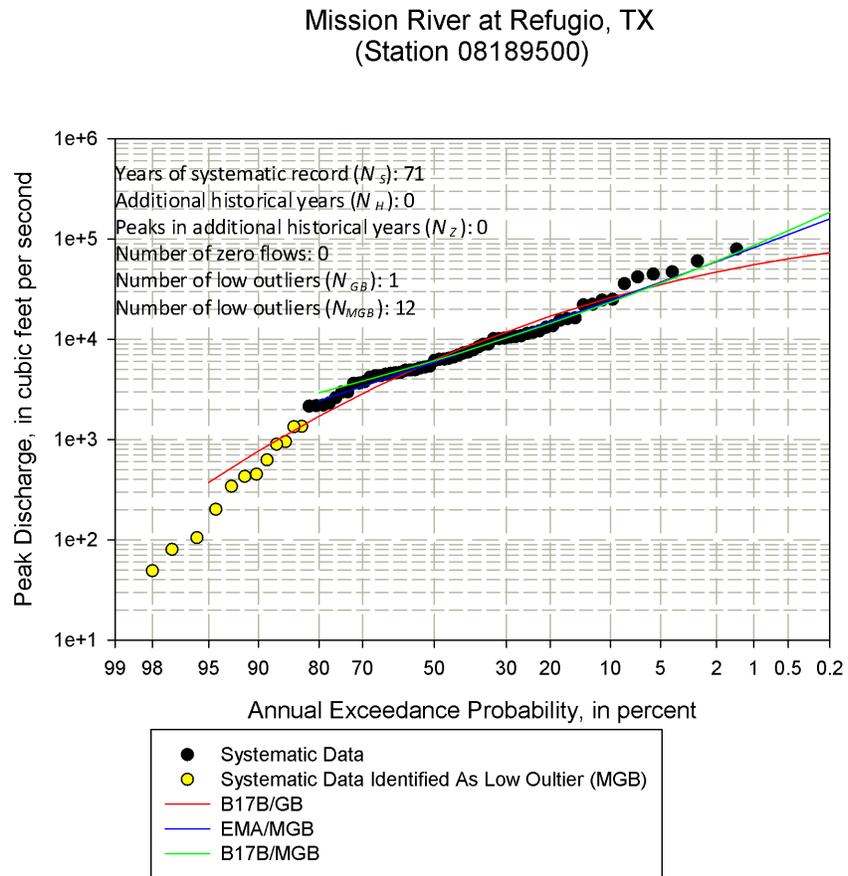


Figure 78: Site 09241000 with Low Outliers; no historical information

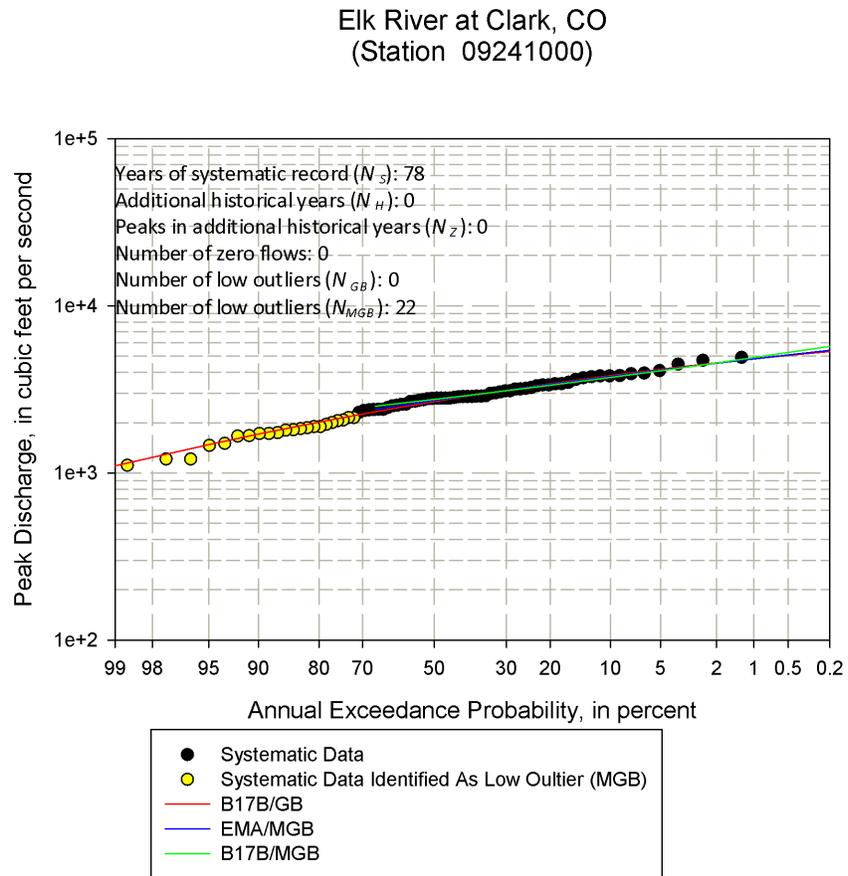


Figure 79: Site 09480000 with Low Outliers; no historical information

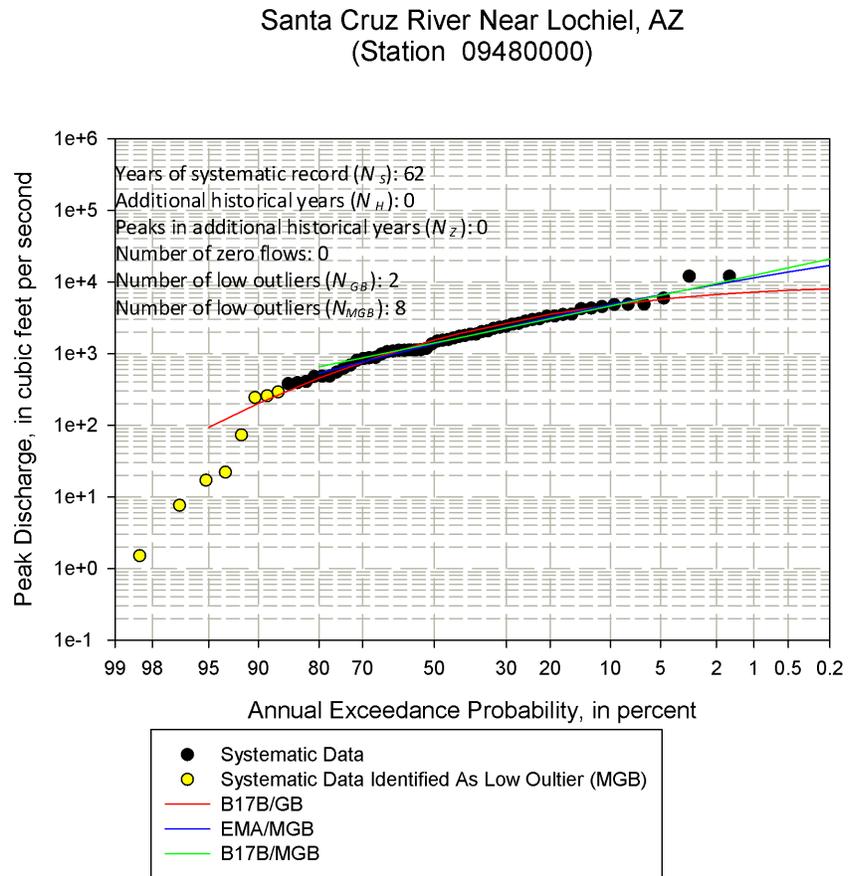


Figure 80: Site 10234500 with Low Outliers; no historical information

Beaver River near Beaver, UT
(Station 10234500)

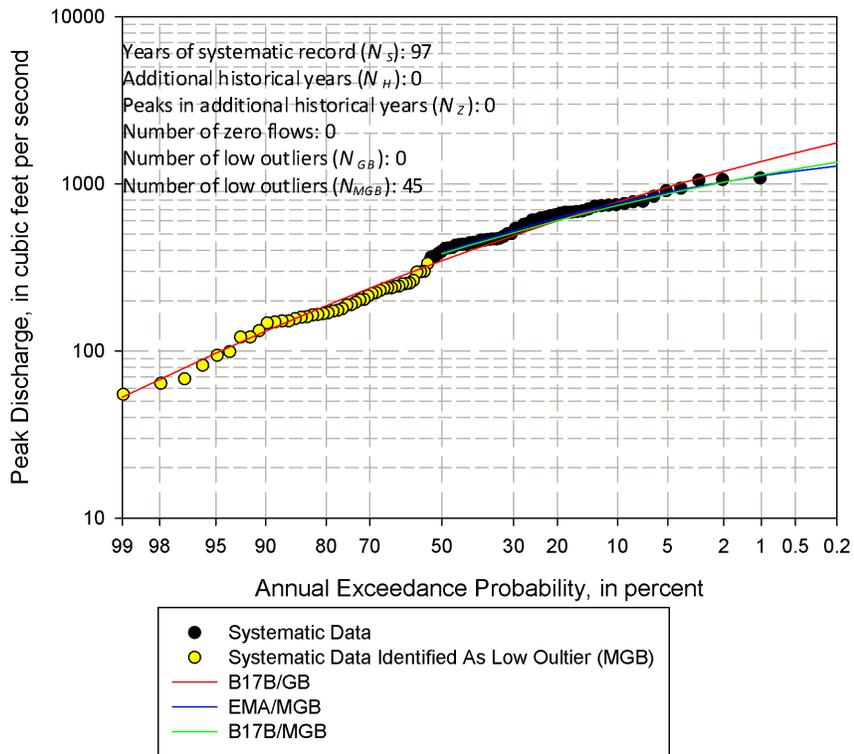


Figure 81: Site 11028500 with Low Outliers; no historical information

Santa Maria Creek Near Ramona, CA
(Station 11028500)

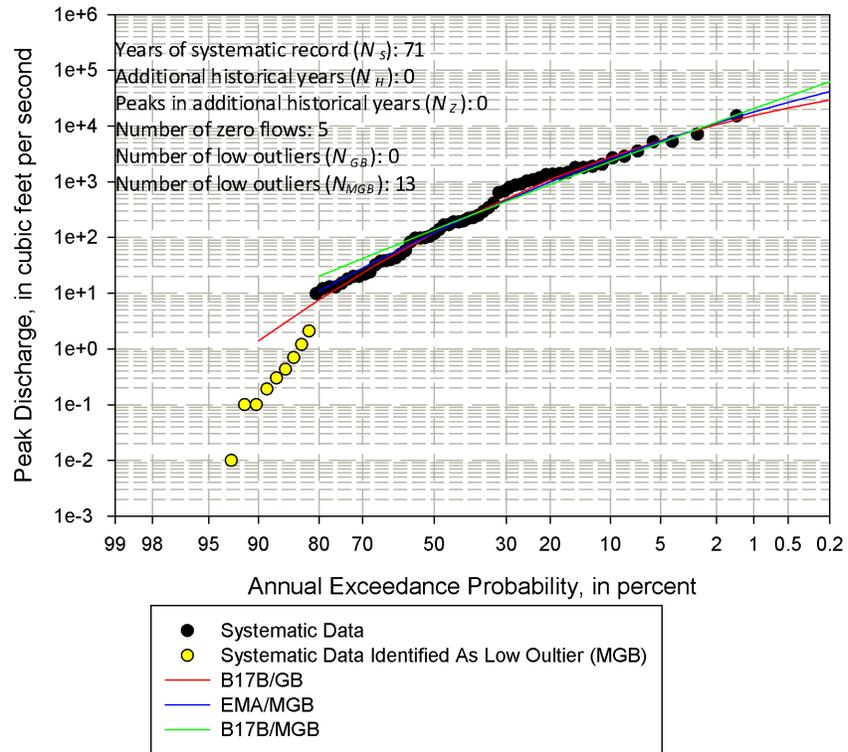


Figure 82: Site 11152000 with Low Outliers; no historical information

Arroyo Seco near Soledad, CA
(Station 11152000)

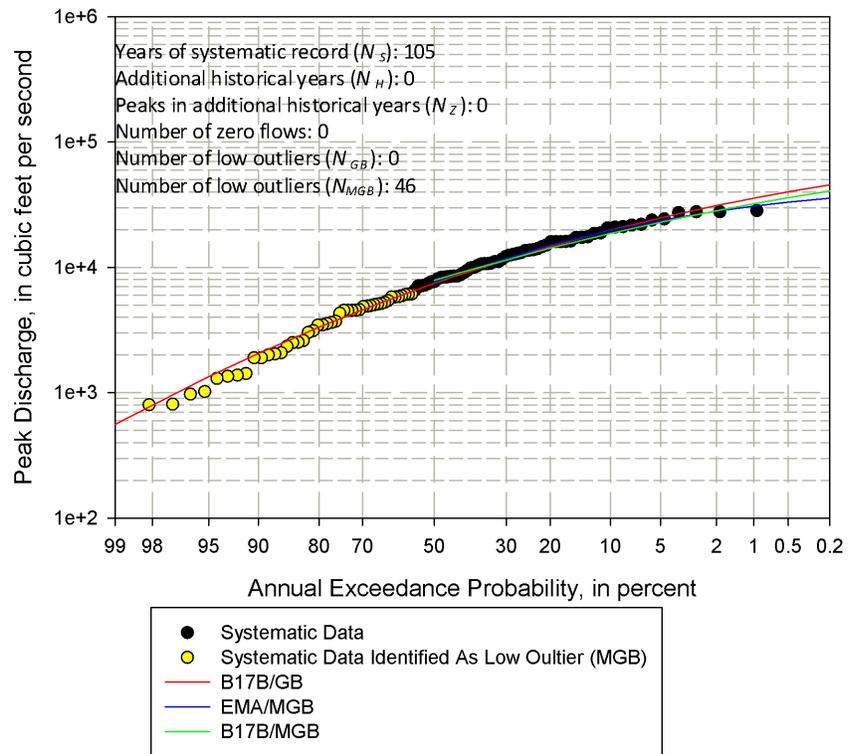


Figure 83: Site 11274500 with Low Outliers; no historical information

Orestimba Creek near Newman, CA
(Station 11274500)

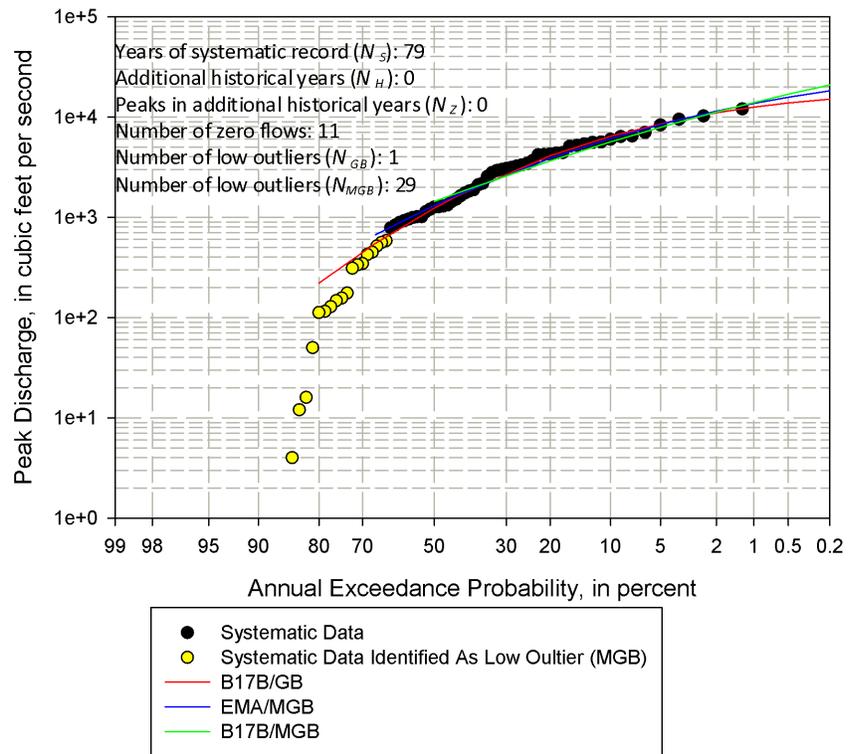


Figure 84: Site 11383500 with Low Outliers; no historical information

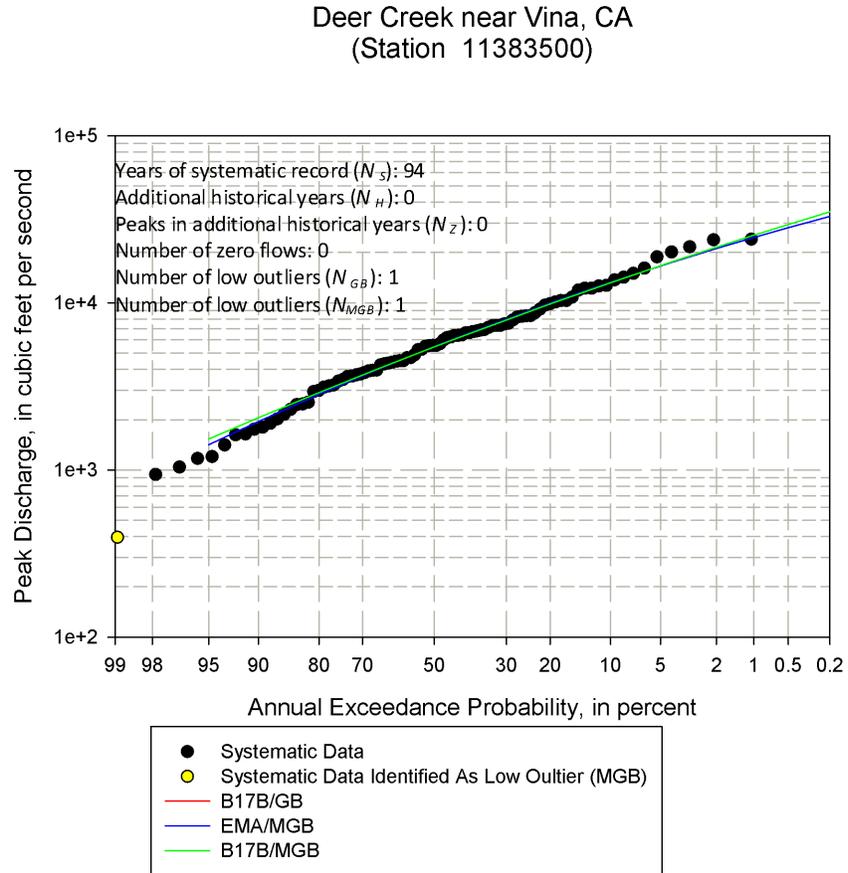


Figure 85: Site 12307500 with Low Outliers; no historical information

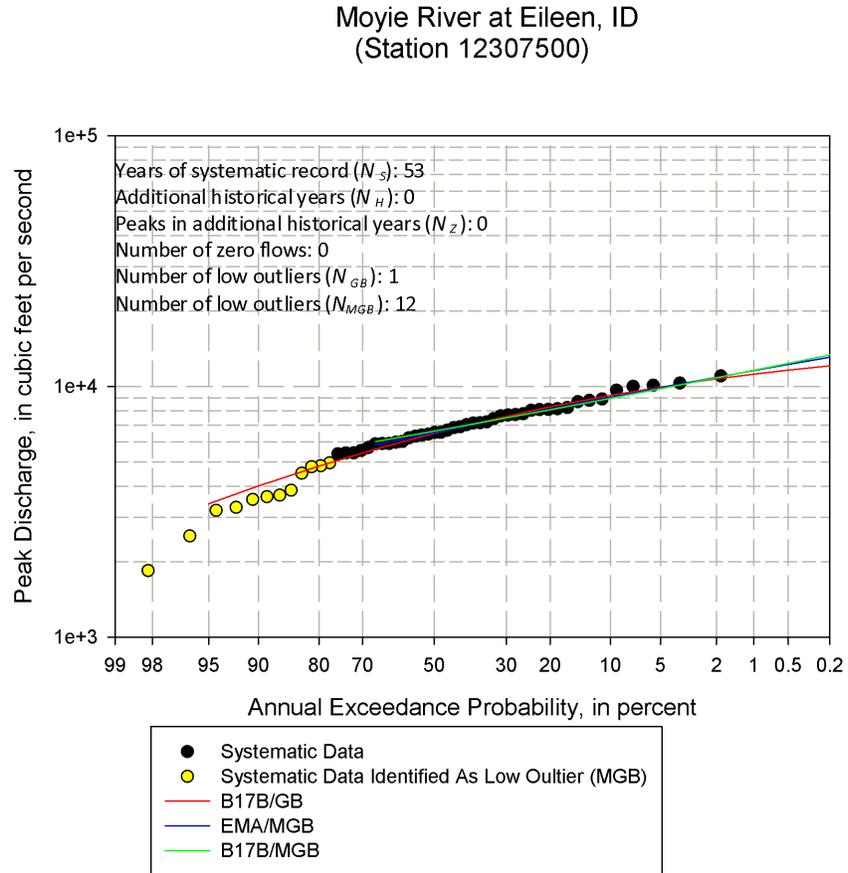


Figure 86: Site 13302500 with Low Outliers; no historical information

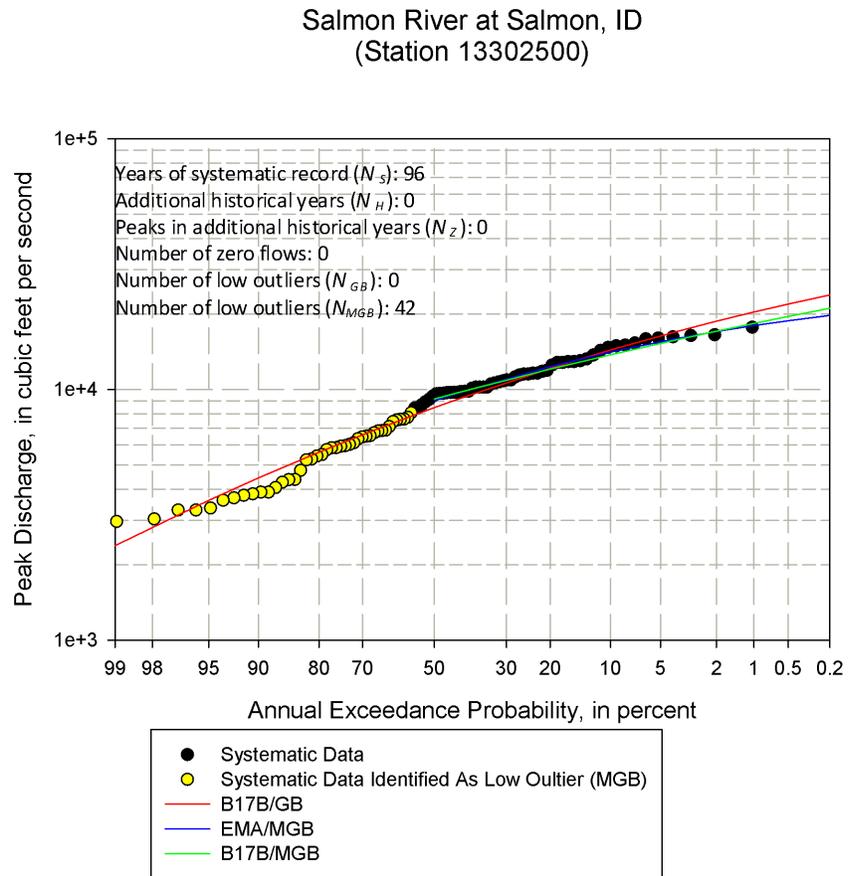


Figure 87: Site 13343660 with Low Outliers; no historical information

Smith Gulch Tributary near Pataha, WA
(Station 13343660)

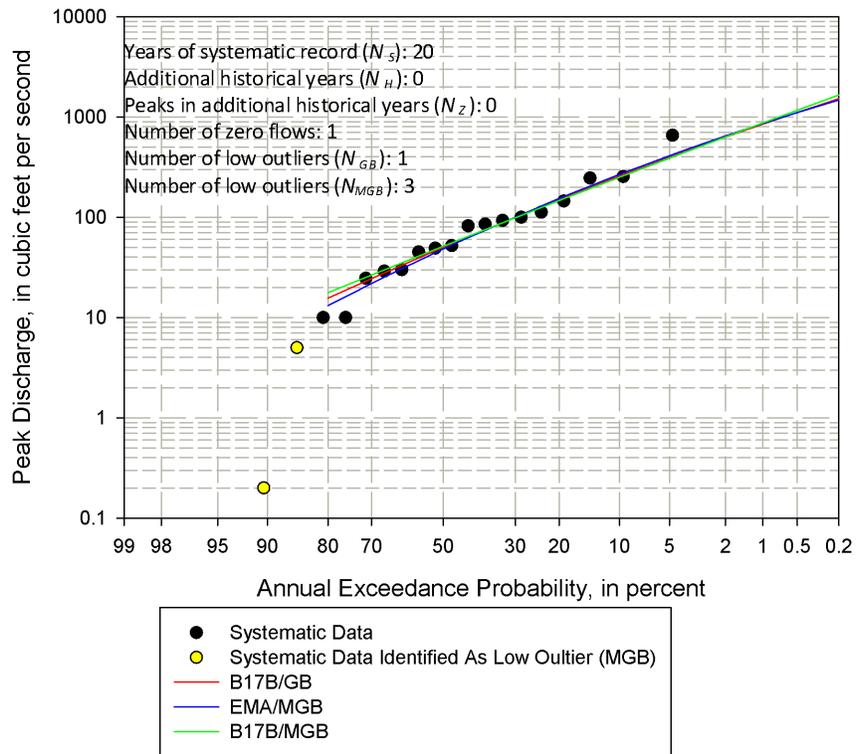


Figure 88: Site 14321000 with Low Outliers; no historical information

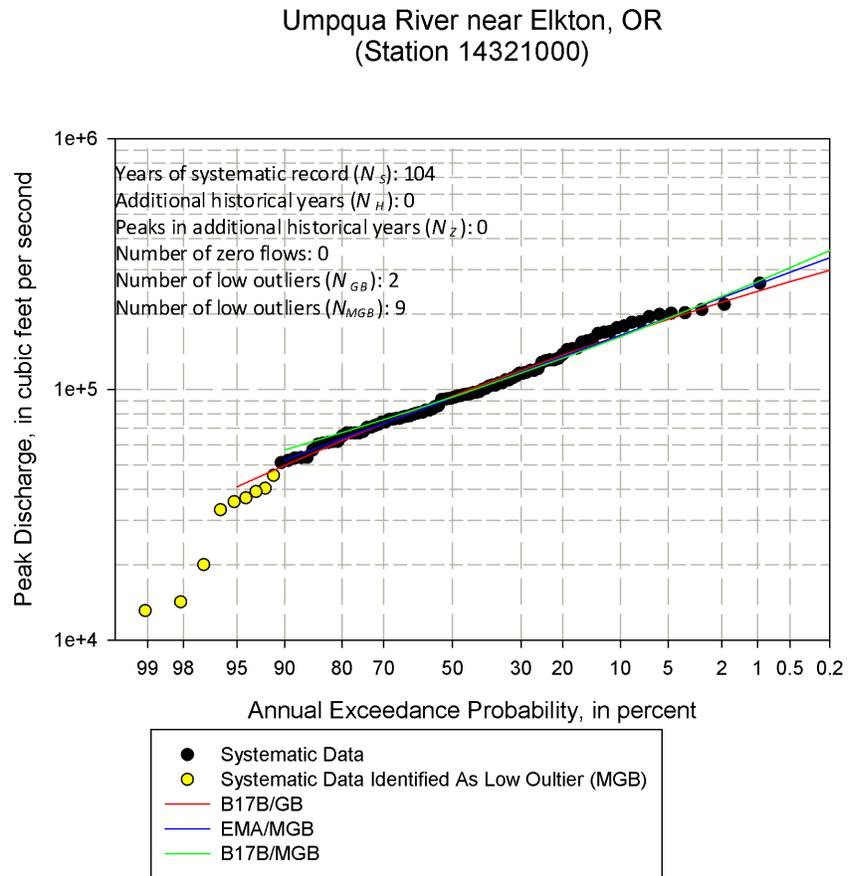
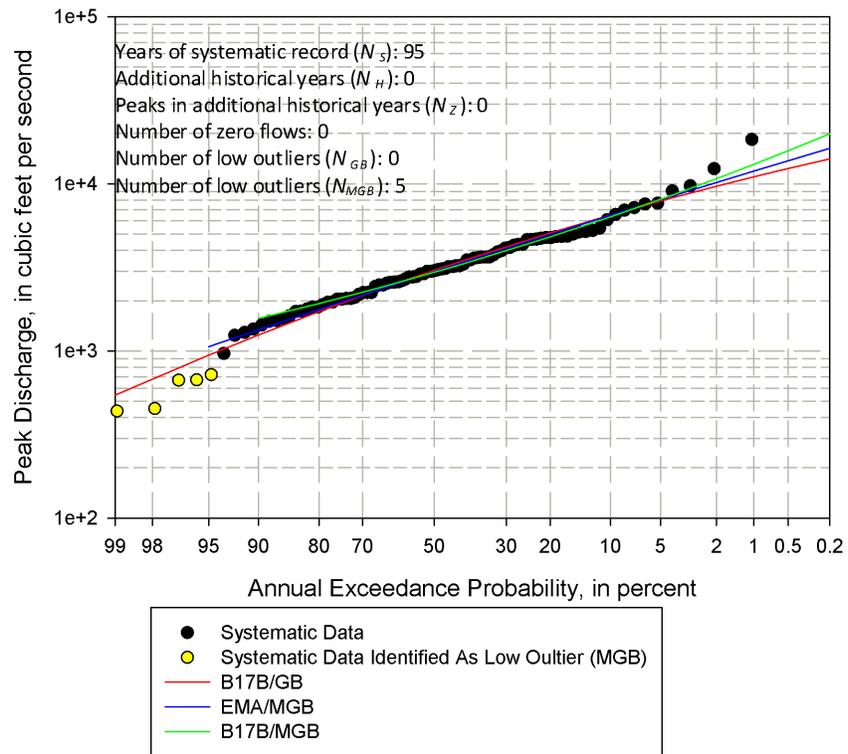


Figure 89: Site 16068000 with Low Outliers; no historical information

EB of NF Wailua River near Lihue, HI
(Station 16068000)



B.4 Sites with a Combination of Data Types

Figure 90: Site 03289500 with a Combination of Low Outliers, Historical and/or High Outliers

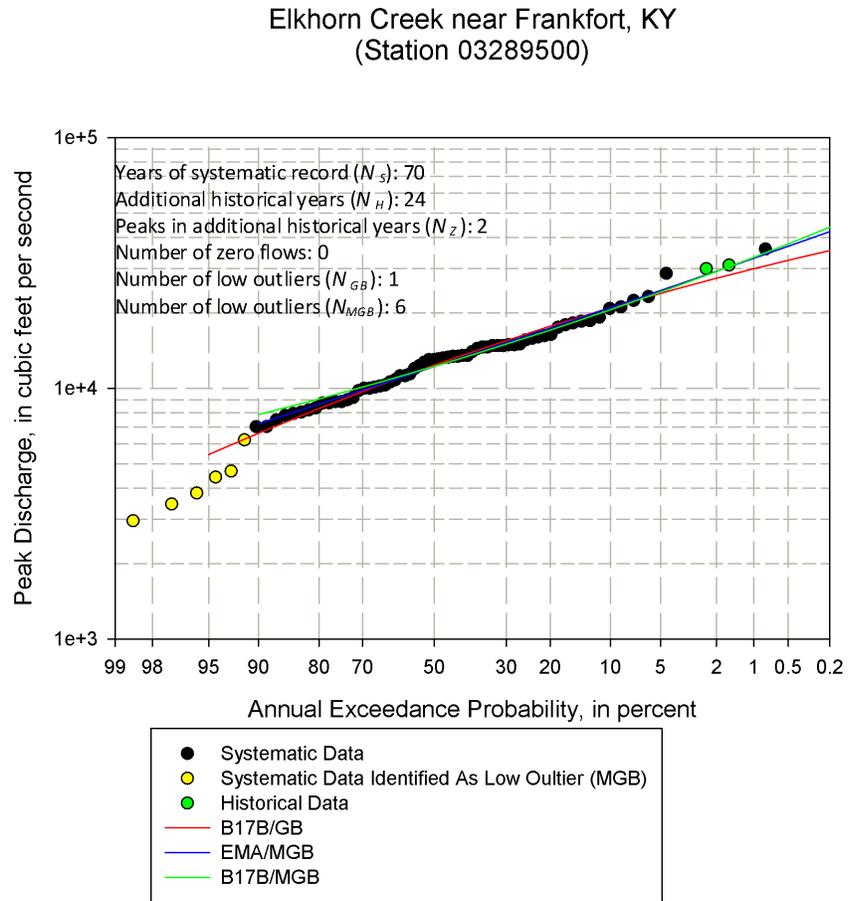


Figure 91: Site 05270500 with a Combination of Low Outliers, Historical and/or High Outliers

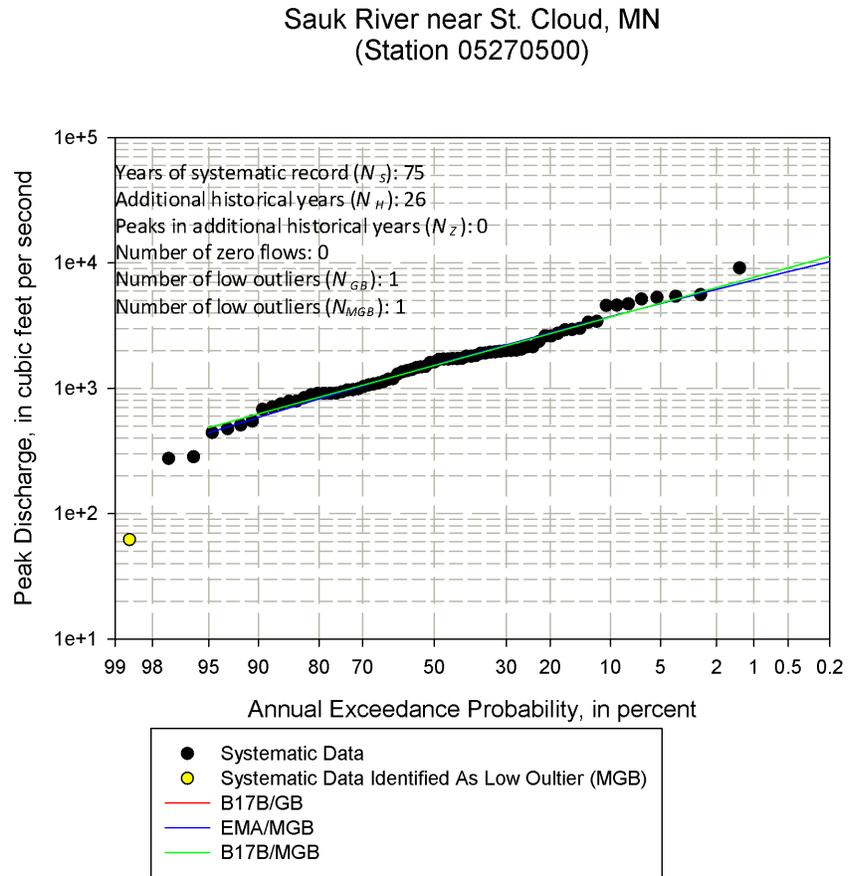


Figure 92: Site 05291000 with a Combination of Low Outliers, Historical and/or High Outliers

Whetstone River near Big Stone City, SD
(Station 05291000)

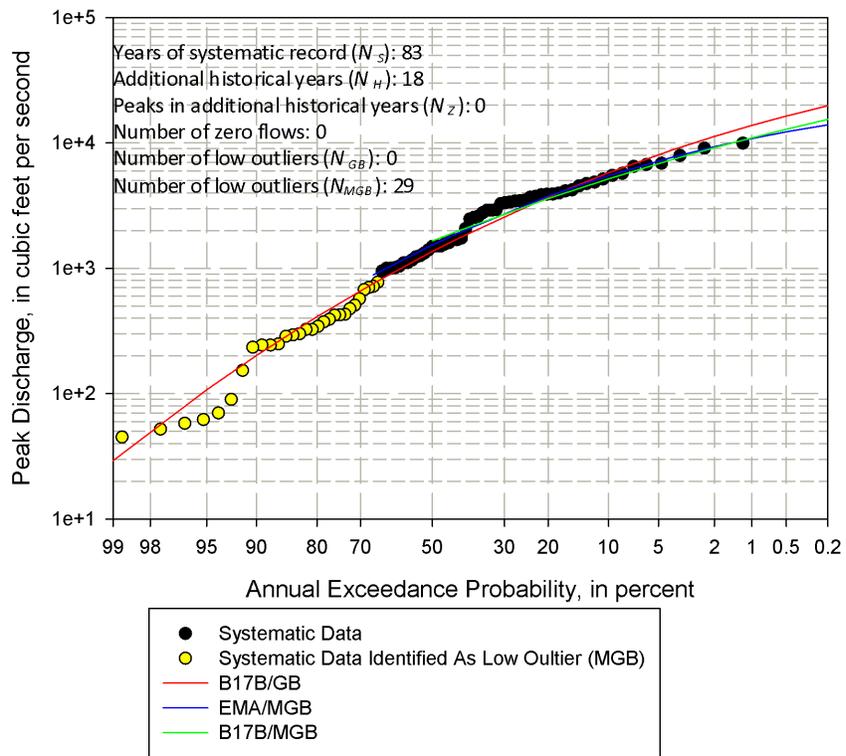


Figure 93: Site 05464500 with a Combination of Low Outliers, Historical and/or High Outliers

Cedar River at Cedar Rapids, IA
(Station 05464500)

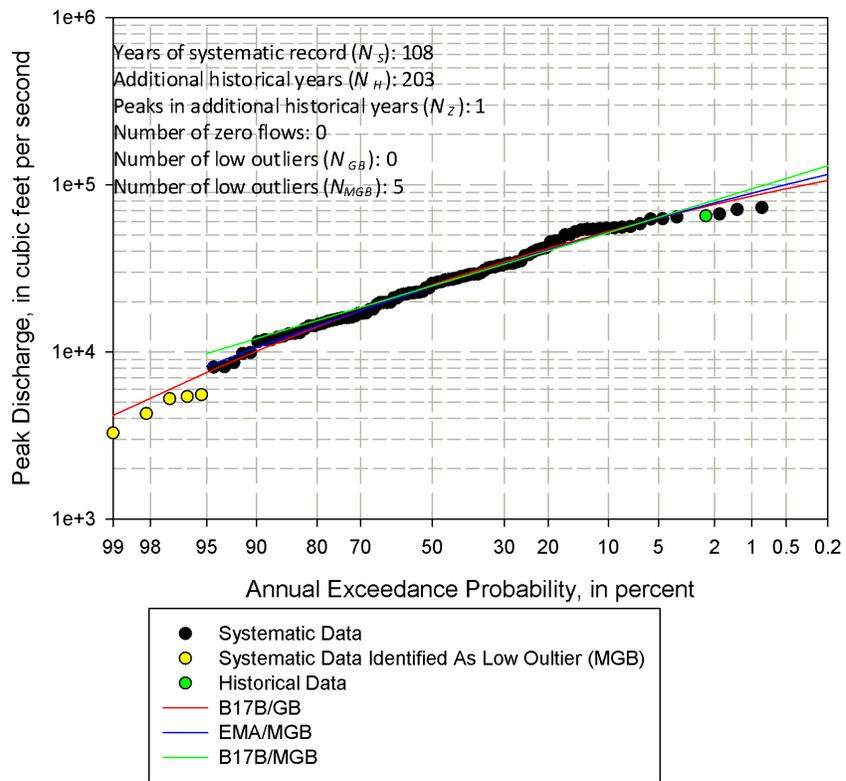


Figure 94: Site 06062500 with a Combination of Low Outliers, Historical and/or High Outliers

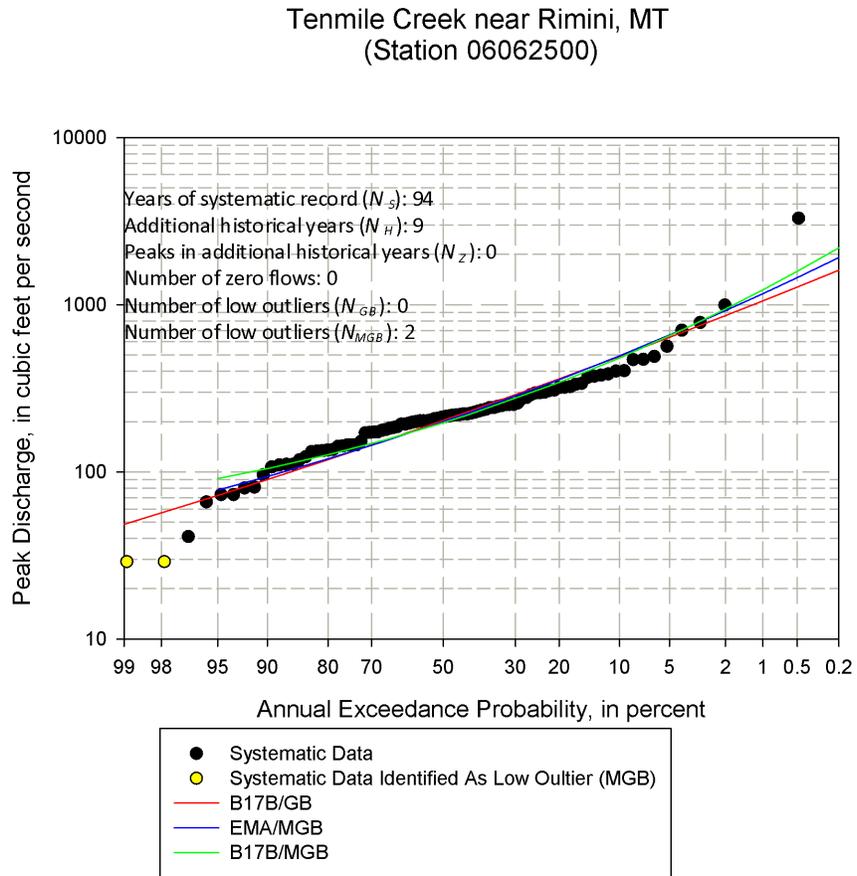


Figure 95: Site 06897000 with a Combination of Low Outliers, Historical and/or High Outliers

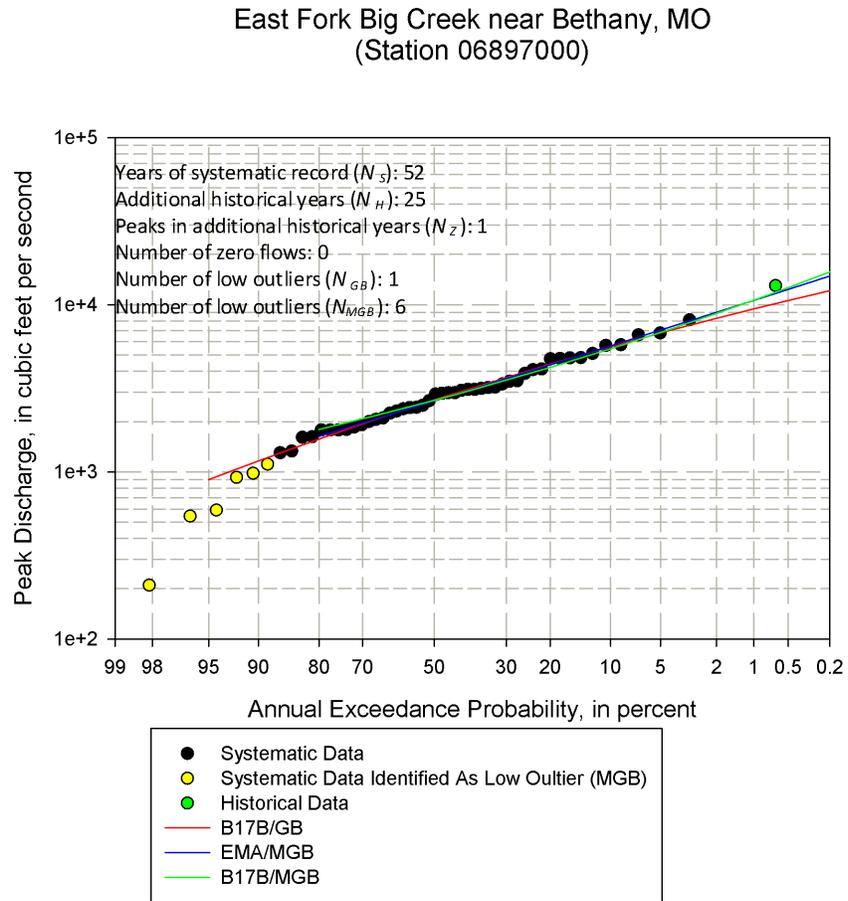


Figure 96: Site 06933500 with a Combination of Low Outliers, Historical and/or High Outliers

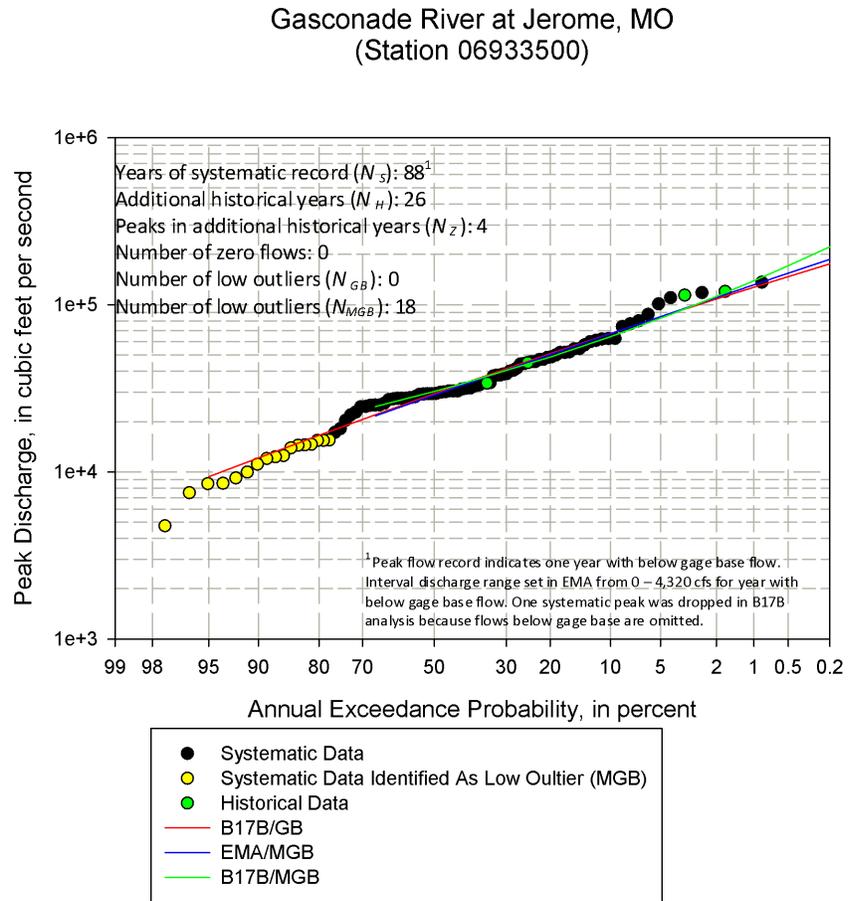


Figure 97: Site 07138600 with a Combination of Low Outliers, Historical and/or High Outliers

White Woman Creek Tributary near Selkirk, KS
(Station 07138600)

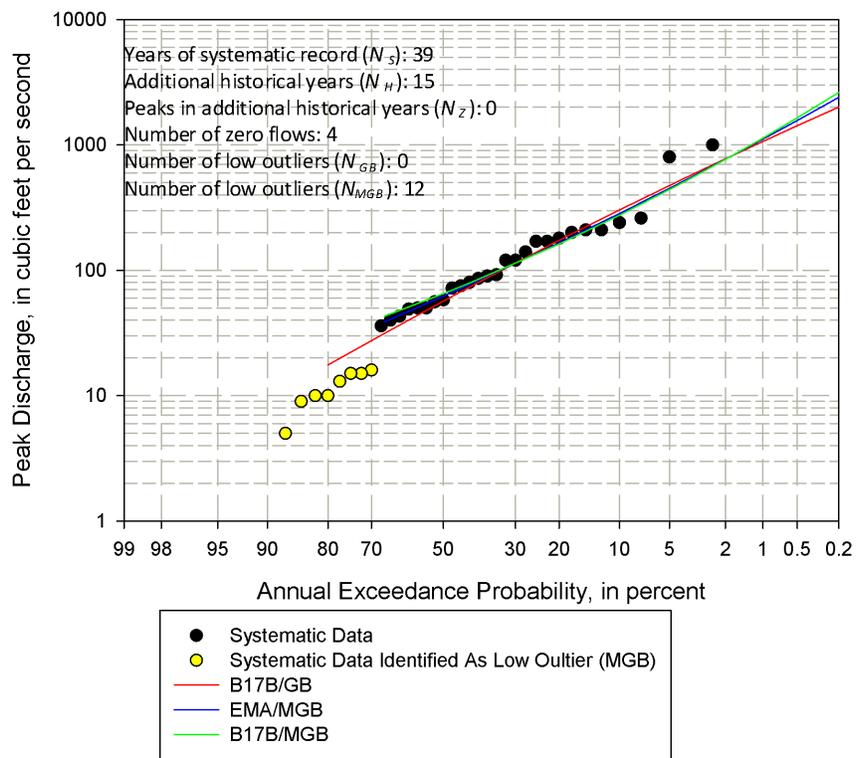


Figure 98: Site 08164000 with a Combination of Low Outliers, Historical and/or High Outliers

Lavaca River near Edna, TX
(Station 08164000)

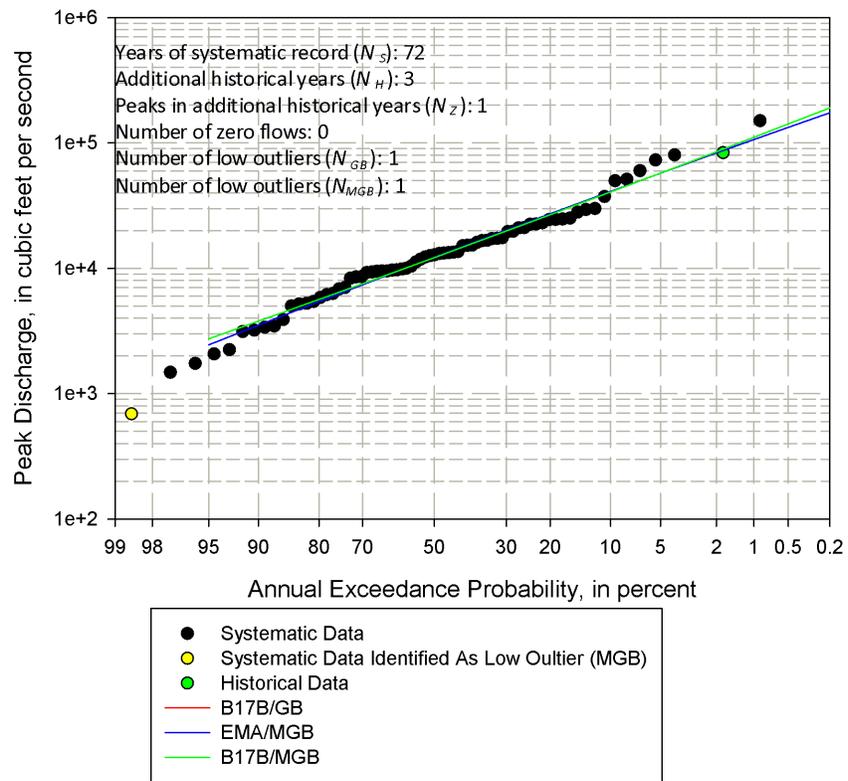


Figure 99: Site 08171000 with a Combination of Low Outliers, Historical and/or High Outliers

Blanco River at Wimberley, TX
(Station 08171000)

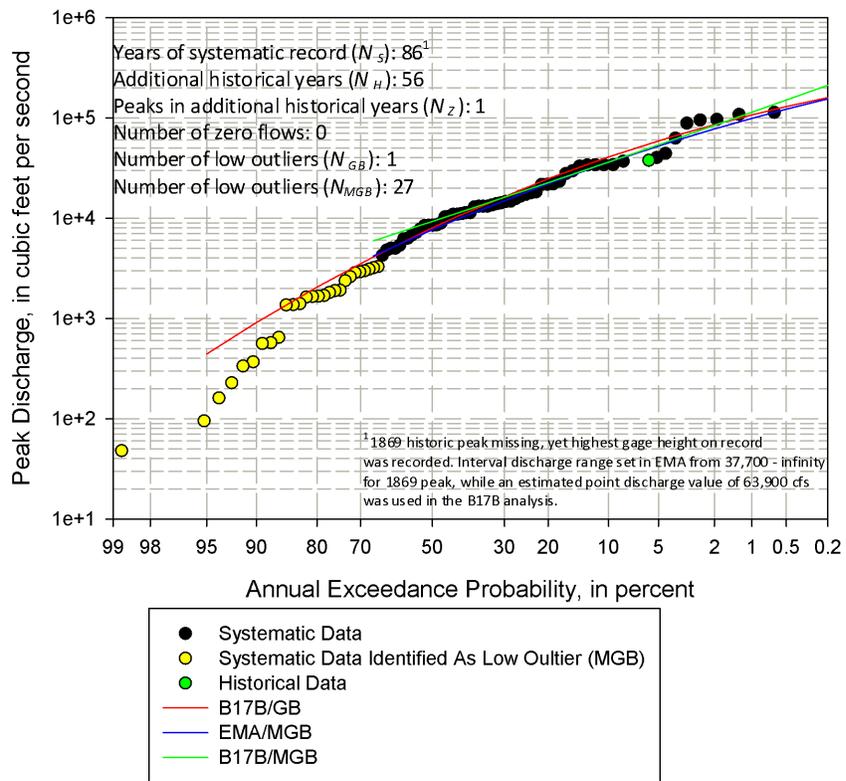


Figure 100: Site 09361500 with a Combination of Low Outliers, Historical and/or High Outliers

Animas River at Durango, CO
(Station 09361500)

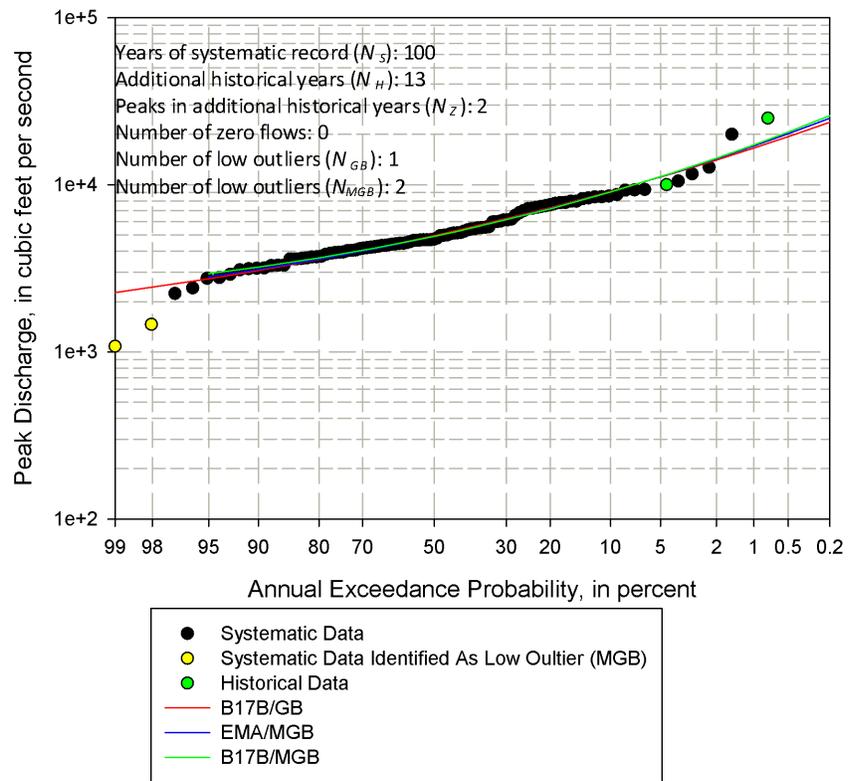


Figure 101: Site 09471000 with a Combination of Low Outliers, Historical and/or High Outliers

San Pedro River at Charleston, AZ
(Station 09471000)

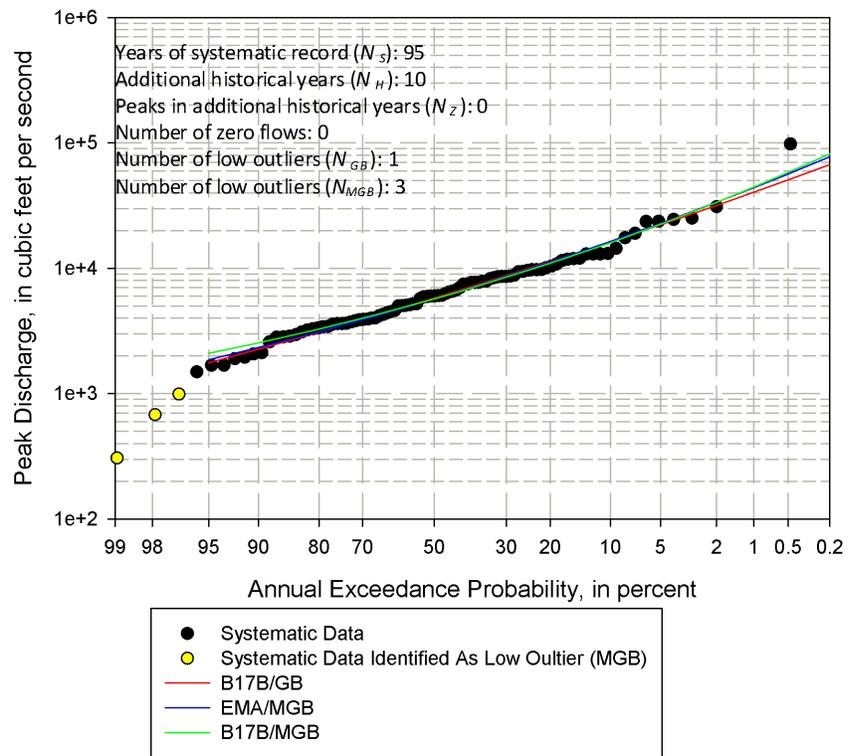


Figure 102: Site 11176000 with a Combination of Low Outliers, Historical and/or High Outliers

Arroyo Mocho near Livermore, CA
(Station 11176000)

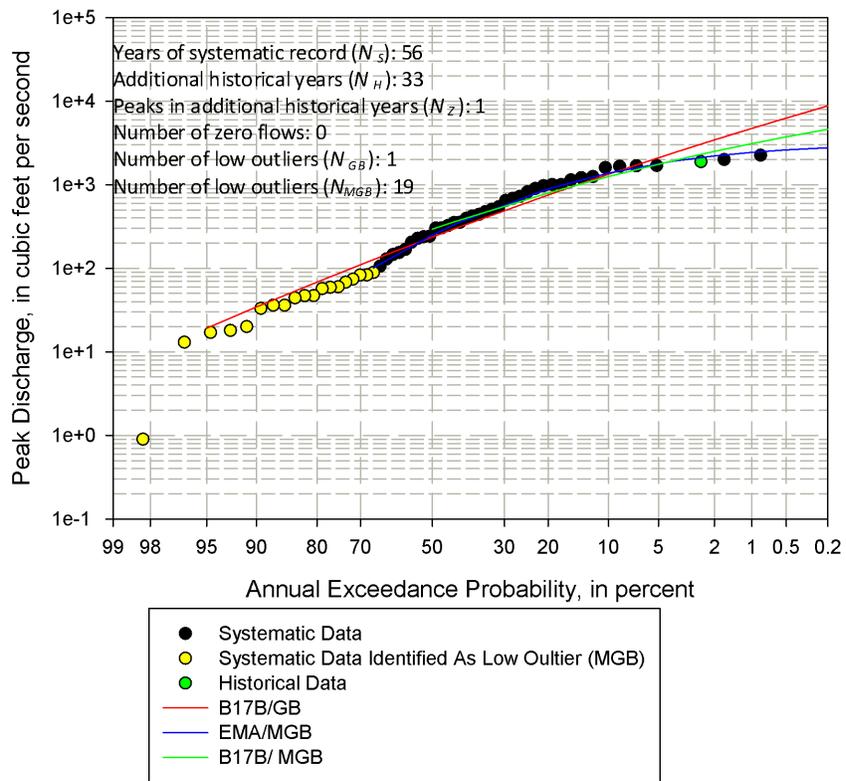


Figure 103: Site 11464500 with a Combination of Low Outliers, Historical and/or High Outliers

Dry Creek near Cloverdale, CA
(Station 11464500)

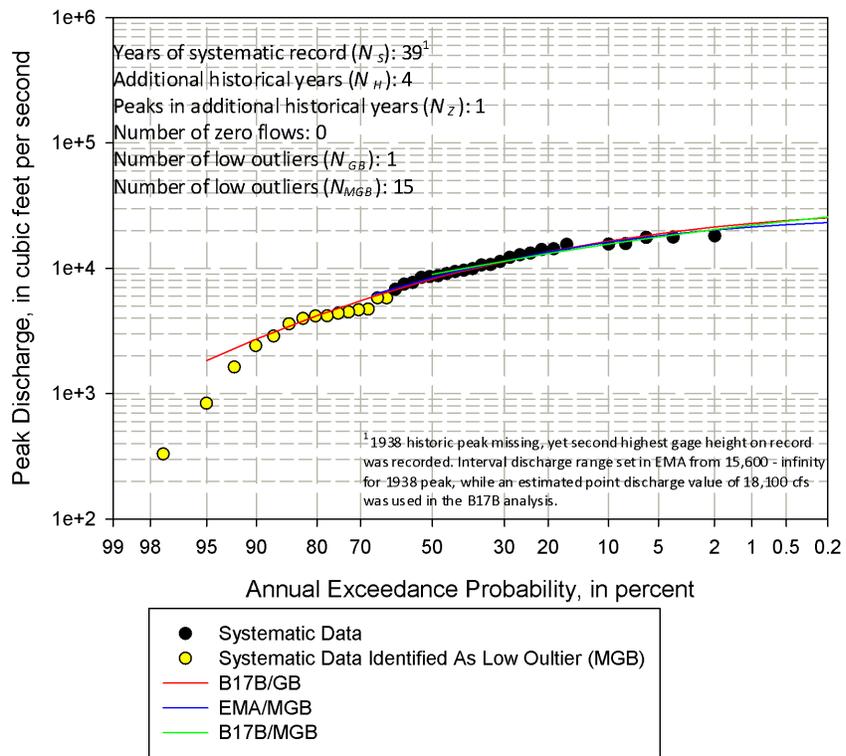


Figure 104: Site 11522500 with a Combination of Low Outliers, Historical and/or High Outliers

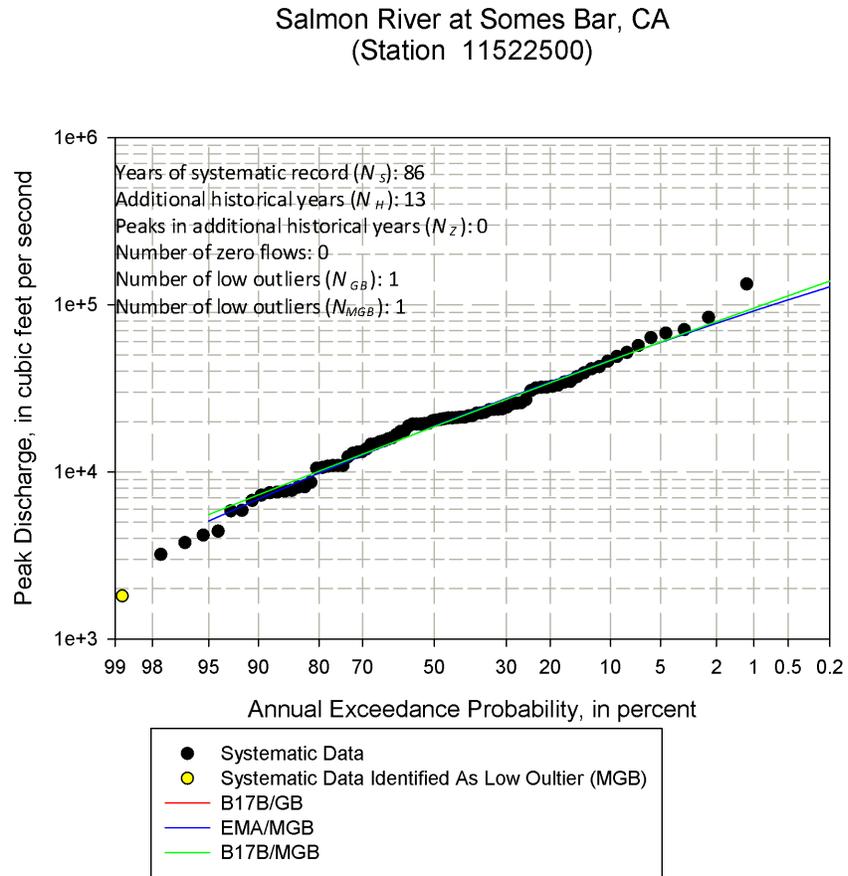


Figure 105: Site 12039500 with a Combination of Low Outliers, Historical and/or High Outliers

Quinault River at Quinault Lake, WA
(Station 12039500)

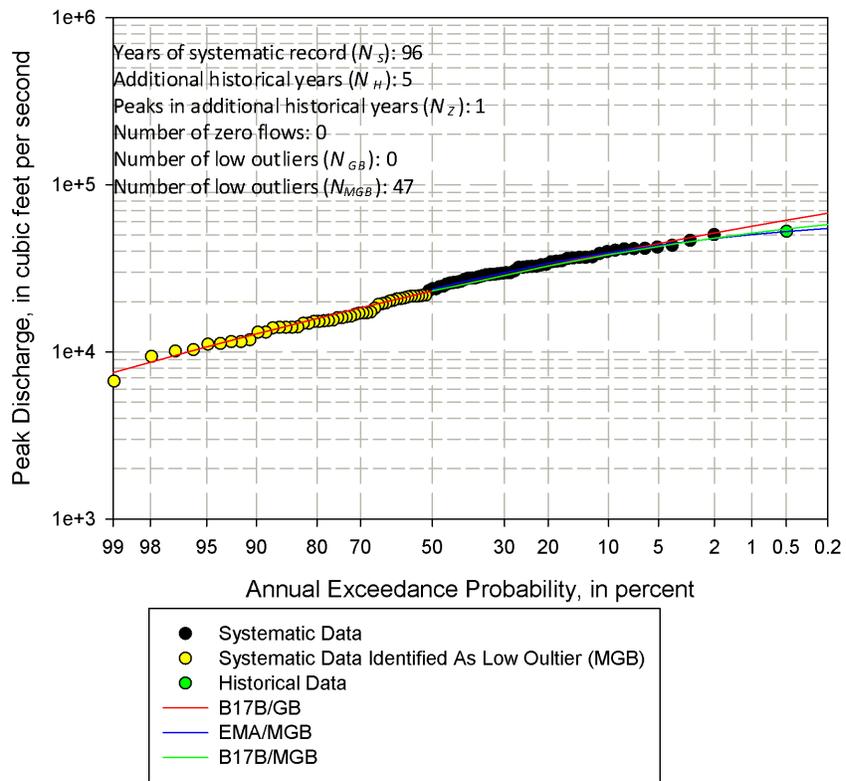
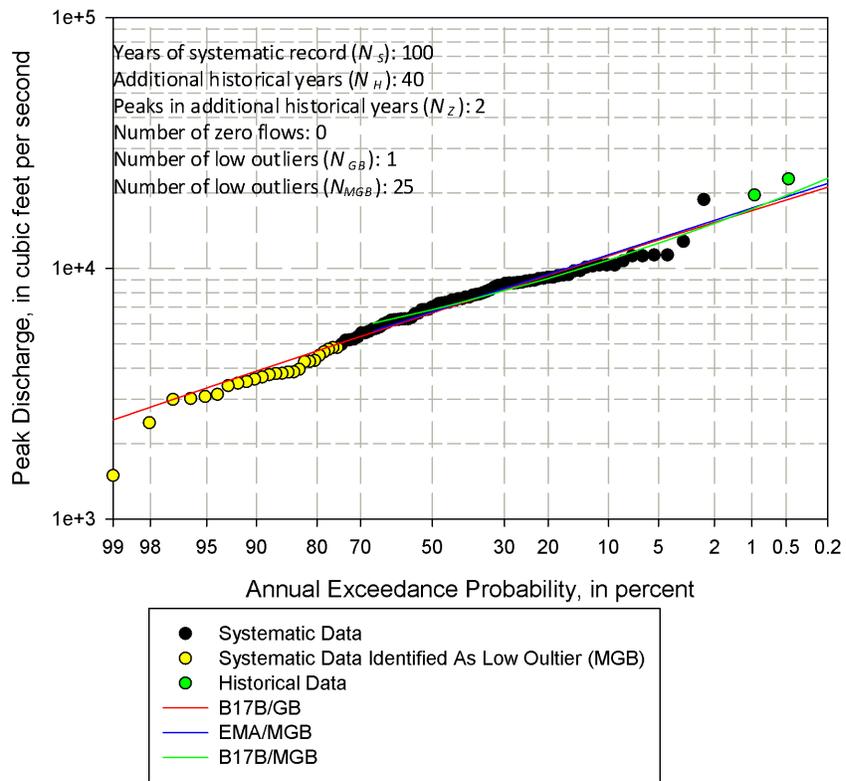


Figure 106: Site 13185000 with a Combination of Low Outliers, Historical and/or High Outliers

Boise River near Twin Springs, ID
(Station 13185000)



B.5 Studies with LP3 Distribution and Regional Skew

Figures 107 - 109 show the same cases as Figures 2 - 4 except that regional skew information has been added with a MSE of 0.15 – a typical value consistent with Bayesian/GLS skew maps Lamontagne et al. (2012); Parrett et al. (2011); Gotvald et al. (2006). As expected, all of the estimators perform better with regional information. Aside from that, however, there is little difference between the corresponding figures.

Figure 107: Results are based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a Log-Pearson Type 3 distribution with skew $\gamma = 0.0$. Regional skew is assumed to be 0.0 with $MSE = 0.15$.

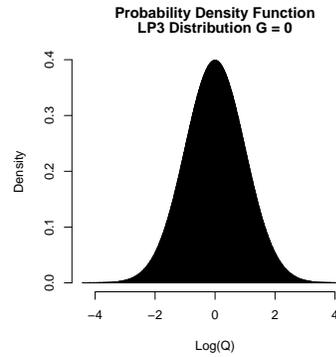
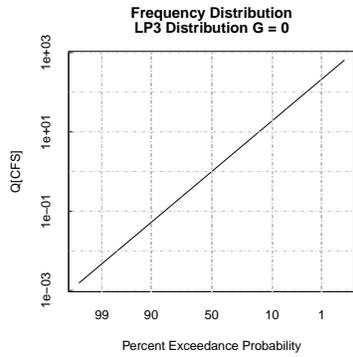
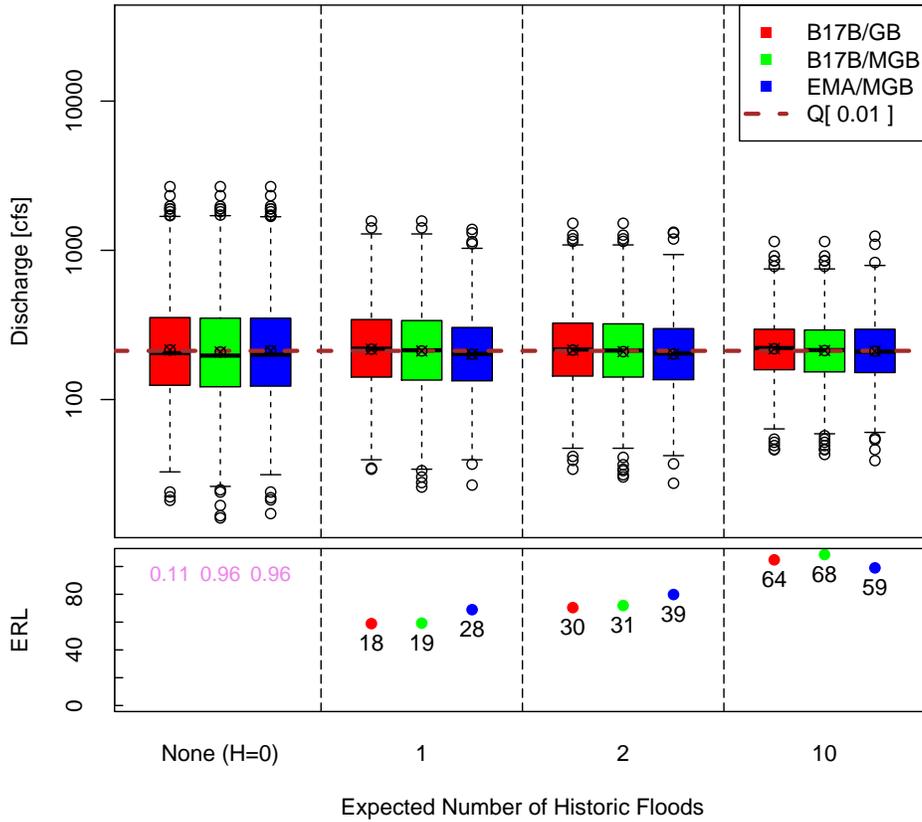


Figure 108: Results are based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a Log-Pearson Type 3 distribution with skew $\gamma = -0.5$. Regional skew is assumed to be -0.5 with $MSE = 0.15$.

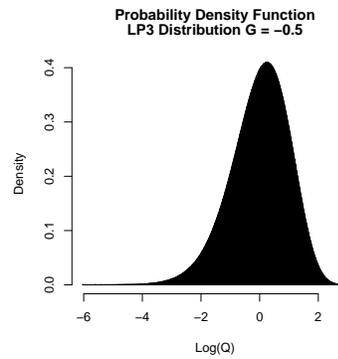
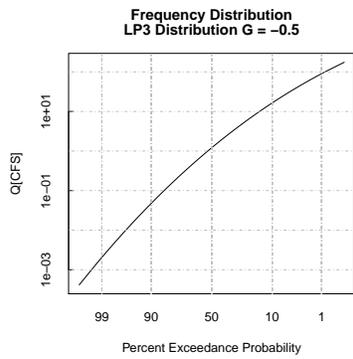
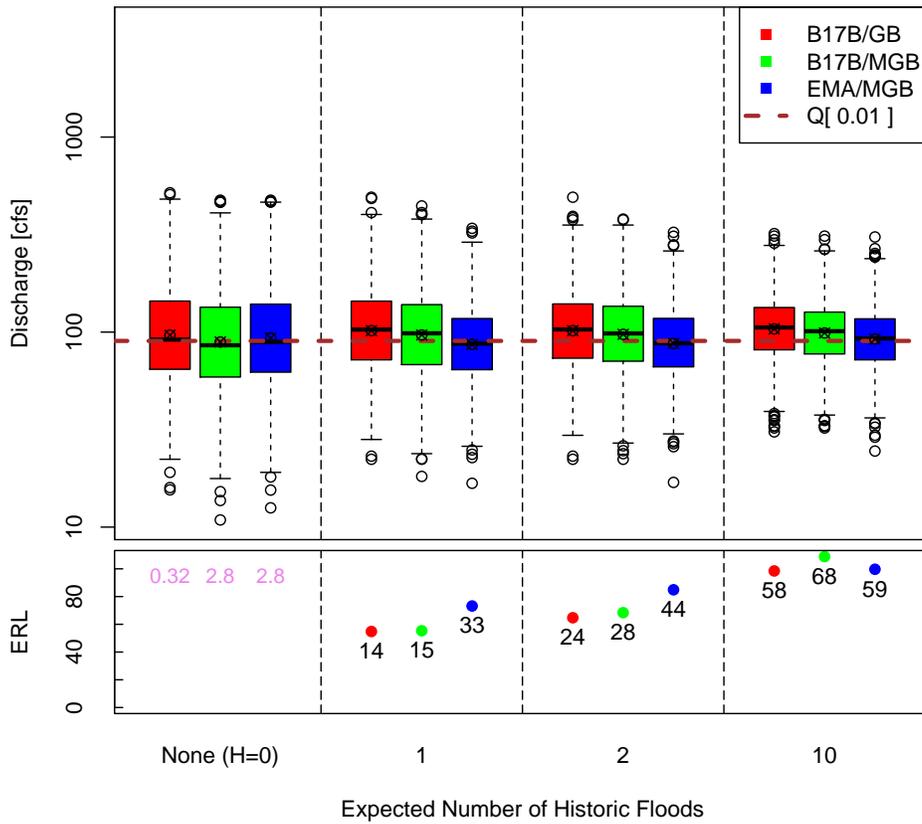
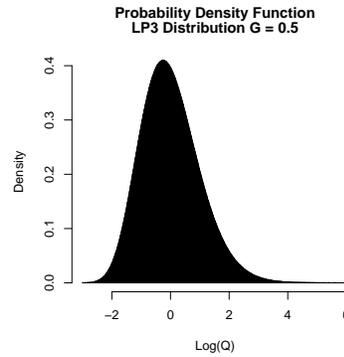
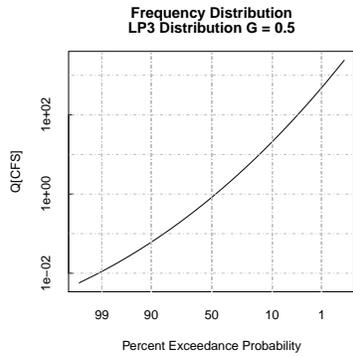
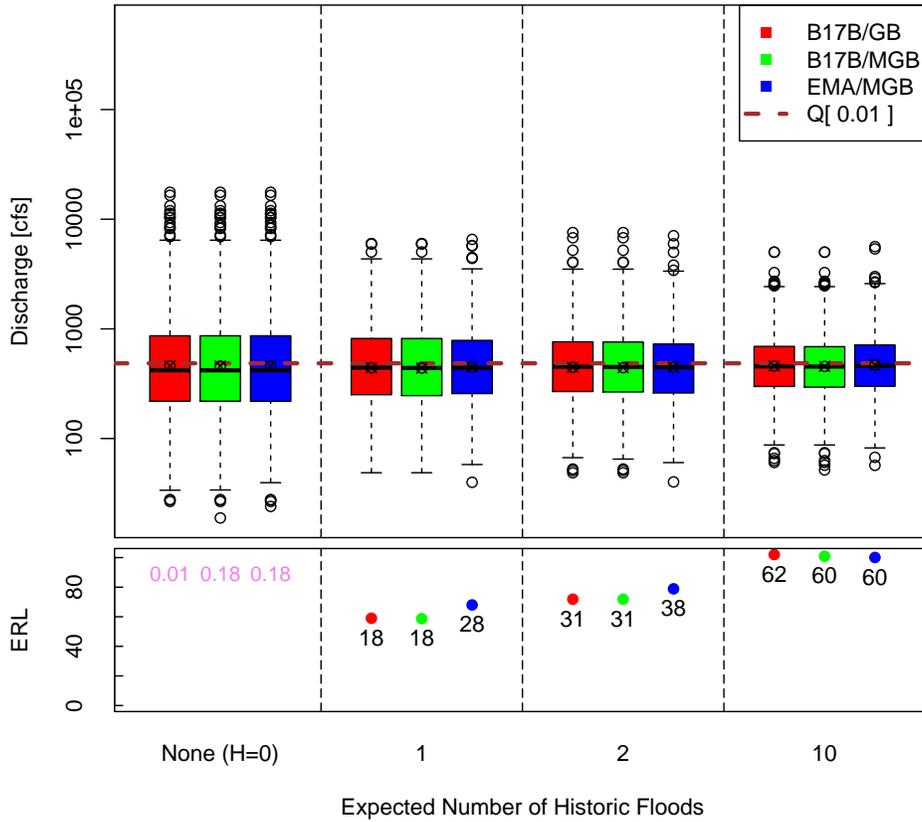


Figure 109: Results are based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from a Log-Pearson Type 3 distribution with skew $\gamma = 0.05$. Regional skew is assumed to be 0.5 with $MSE = 0.15$.



B.6 Additional Studies with Specific Frequency Curves

Figures 5 - 8 and 110 - 111 apply the three estimators, EMA/MGB, B17B/GB and B17B/MGB, to data drawn from specific populations selected to test the estimators' performance. The Cases are labeled "robustness test curve 1-6" in recognition of their origins. Figures 5 - 8 are presented in the main body of this report. Figures 110 and 111 are presented here in the appendix because they do not actually deal with robustness but rather with specific LP3 populations. In fact, because the estimators are all invariant with respect to location and scale, these two cases duplicate cases already considered in the report. Figure 110 depicts essentially the same case as figure 4 with a population skew of $\gamma = 0.5$. Figure 111 depicts essentially the same case as figure 3 with a population skew of $\gamma = -0.5$.

Figure 110: Results are based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from robustness test curve 1

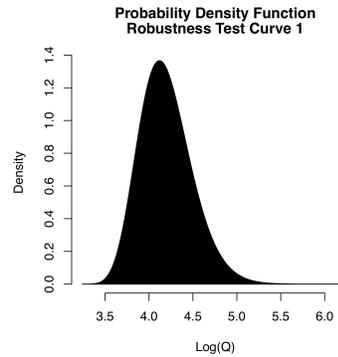
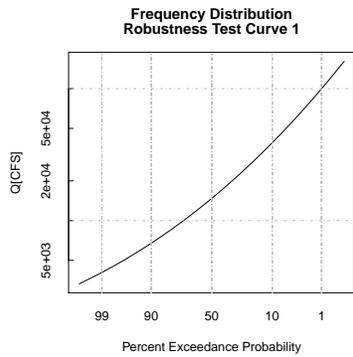
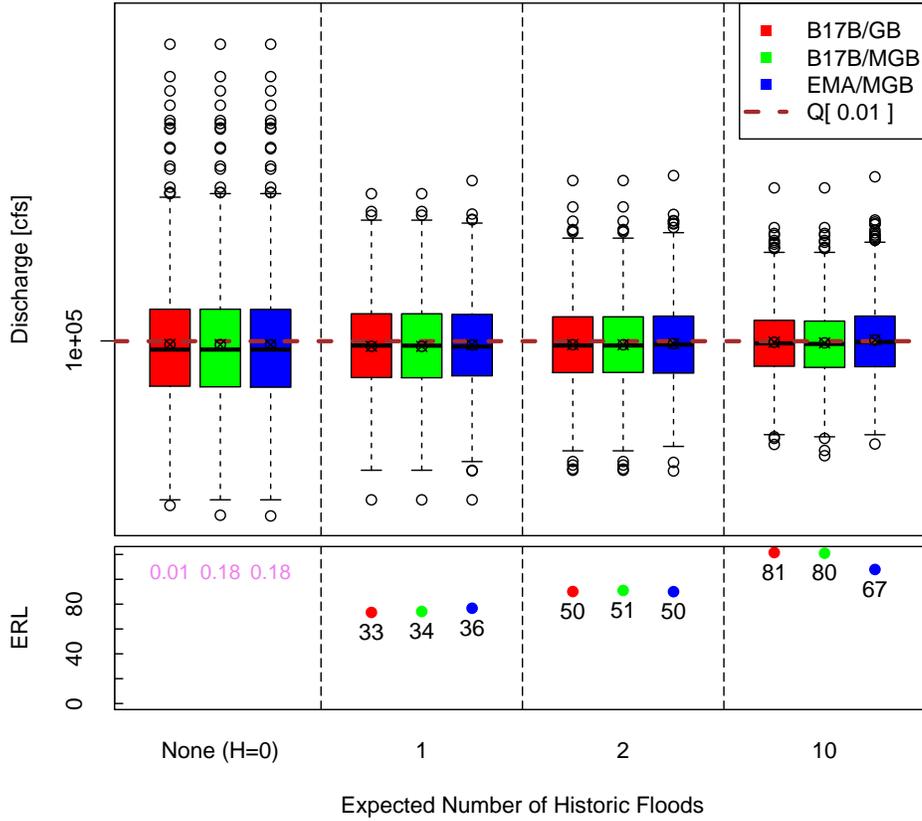
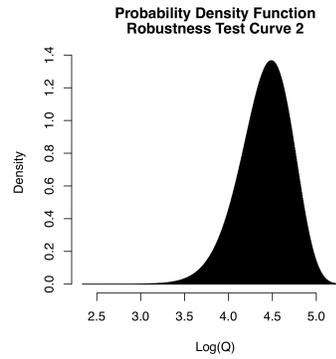
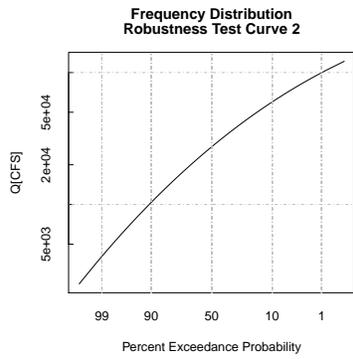
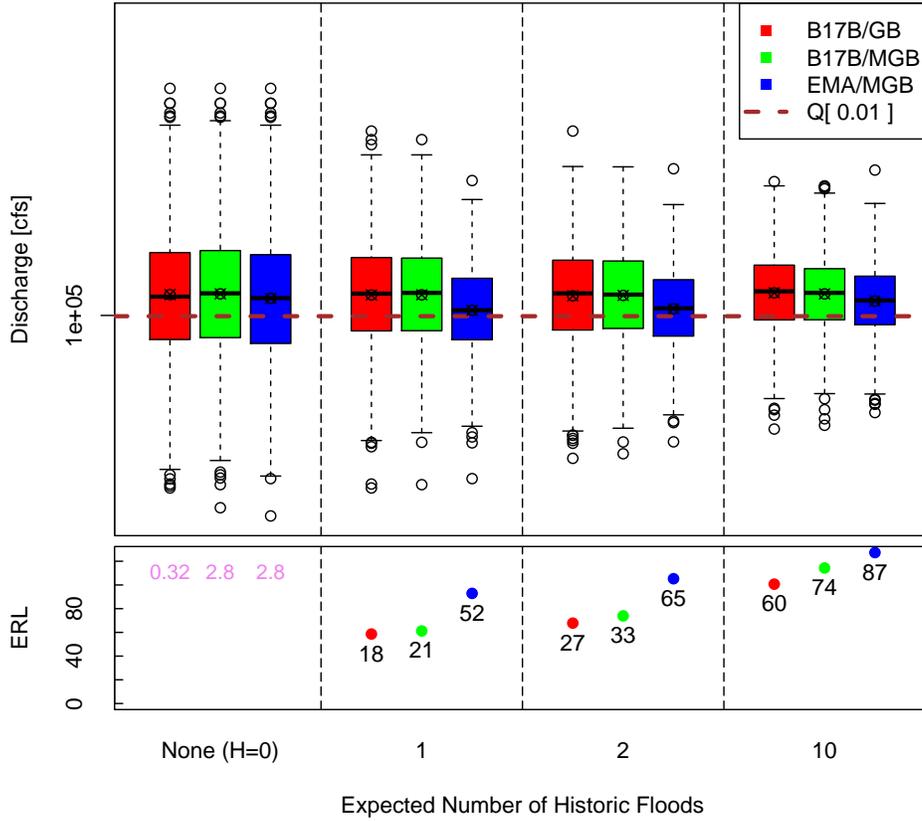


Figure 111: Results are based on 10000 replicate samples of size $N_S = 40$ and $N_H = 100$ drawn from robustness test curve 2



References

- ACWI. Advisory committee on water information, subcommittee on hydrology, hydrologic frequency analysis work group, 2011. URL <http://acwi.gov/hydrology/Frequency/>.
- AIR. A chronology of events affecting the National Flood Insurance Program. Report, American Institutes for Research, 2001.
- V. Barnett and T. Lewis. *Outliers in statistical data*. Wiley New York, 1994.
- J. U. Chowdhury and J. R. Stedinger. Confidence interval for design floods with estimated skew coefficient. *Journal of Hydraulic Engineering*, 117: 811–831, 1991.
- T. A. Cohn. Computation of confidence intervals for ema-based flood quantile estimates. *to be submitted to Water Resources Research*, page 29, in prep, 2015.
- T. A. Cohn, W. M. Lane, and W. G. Baier. An algorithm for computing moments-based flood quantile estimates when historical flood information is available. *Water Resources Research*, 33(9):2089–2096, 1997.
- T. A. Cohn, W. M. Lane, and J. R. Stedinger. Confidence intervals for Expected Moments Algorithm flood quantile estimates. *Water Resources Research*, 37(6):1695–1706, 2001.
- TA Cohn, JF England, Jr., CE Berenbrock, RR Mason, JR Stedinger, and JR Lamontagne. A generalized grubbs-beck test statistic for detecting

multiple potentially influential low outliers in flood series. *Water Resources Research*, 49(8):5047–5058, 2013.

David R Dawdy, Veronica W Griffis, and Vijay K Gupta. Regional flood-frequency analysis: How we got here and where we are going. *Journal of Hydrologic Engineering*, 17(9):953–959, 2012.

JF England, Jr., R. D. Jarrett, and J. D. Salas. Data-based comparisons of moments estimators using historical and paleoflood data. *Journal of Hydrology*, 278(1):172–196, 2003.

Anthony J Gotvald, Toby D Feaster, and J Curtis Weaver. Magnitude and frequency of rural floods in the southeastern united states, 2006: Volume 1, georgia. Technical report, U. S. Geological Survey, 2006.

V.W. Griffis and J.R. Stedinger. Evolution of Flood Frequency Analysis with Bulletin 17. *Journal of Hydrologic Engineering*, 12:283, 2007.

VW Griffis, JR Stedinger, and TA Cohn. Log Pearson type 3 quantile estimators with regional skew information and low outlier adjustments. *Water Resources Research*, 40(7), 2004a.

VW Griffis, JR Stedinger, and TA Cohn. Log Pearson type 3 quantile estimators with regional skew information and low outlier adjustments. *Water Resources Research*, 40(7):W07503, 2004b.

IACWD. Guidelines for determining flood flow frequency, Bulletin 17-B. Technical report, Interagency Committee on Water Data, Hydrology Subcommittee, 1982.

ICOWR. Methods of flow frequency analysis, Bulletin No. 13. Technical report, Interagency Committee on Water Resources, Subcommittee on Hydrology, Washington, D.C., 1966.

M. E. Jennings and M. A. Benson. Frequency curves for annual flood series with some zero events or incomplete data. *Water Resources Research*, 5 (1):276–280, 1969.

W. Kirby. Algebraic boundedness of sample statistics. *Water Resources Research*, 10(2):220–222, 1974.

Jonathan R Lamontagne, Jery R Stedinger, Charles Berenbrock, Andrea G Veilleux, Justin C Ferris, and Donna L Knifong. Development of regional skews for selected flood durations for the central valley region, california, based on data through water year 2008. Technical report, USGS, 2012. URL <http://pubs.usgs.gov/sir/2012/5130/>.

Jonathan R Lamontagne, Jery R Stedinger, Timothy A Cohn, and Nancy A Barth. Robust national flood frequency guidelines: What is an outlier? In *Showcasing the Future*, pages 2454–2466. ASCE, 2013. doi: 10.1061/9780784412947.242.

R.H. McCuen. *Modeling hydrologic change: statistical methods*. CRC, 2002.

D.S. Mileti. *Disasters by design*. Joseph Henry Press Washington, DC, 1999.

Charles Parrett, Andrea Veilleux, JR Stedinger, NA Barth, Donna L Knifong, and JC Ferris. Regional skew for california, and flood frequency for

selected sites in the sacramento–san joaquin river basin, based on data through water year 2006. Technical report, U. S. Geological Survey, 2011.

B. Rosner. On the detection of many outliers. *Technometrics*, 17(2):221–227, 1975.

C.S. Spencer and R.H. McCuen. Detection of outliers in Pearson type III data. *Journal of Hydrologic Engineering*, 1:2, 1996.

J. R. Stedinger and T. A. Cohn. Flood frequency analysis with historical and paleoflood information. *Water Resources Research*, 22(5):785–793, 1986.

J. R. Stedinger and T. A. Cohn. Historical flood-frequency data: Its value and use. In V. P. Singh, editor, *Regional Flood Frequency Analysis*, pages 273–286. D. Reidel, 1987.

J. R. Stedinger, Richard M. Vogel, and Efi Foufoula-Georgiou. *Frequency Analysis of Extreme Events*, chapter 18, page 99. McGraw Hill, Inc., 1993.

G. D. Tasker and W. O. Thomas. Flood frequency analysis with pre-record information. *ASCE Journal of the Hydraulics Division (HY2)*, 104(2): 249–259, 1978.

WO Thomas, Jr. A uniform technique for flood frequency analysis. *Journal of Water Resources Planning and Management*, 111:321, 1985.

John W Tukey. Exploratory data analysis. *Reading, Ma*, 231, 1977.

USWRC. *A uniform technique for determining flood flow frequencies, Bulletin No. 15*. U.S. Water Resources Council, Subcommittee on Hydrology, Washington, D.C., 1967.

USWRC. *Guidelines for Determining Flood Flow Frequency, Bulletin No. 17*. U.S. Water Resources Council, Subcommittee on Hydrology, Washington, D.C., 1976.

USWRC. *Guidelines for Determining Flood Flow Frequency, Bulletin No. 17A*. U.S. Water Resources Council, Subcommittee on Hydrology, Washington, D.C., 1977.

AG Veilleux, JR Stedinger, and DA Eash. Bayesian wls/gls regression for regional skewness analysis for regions with large crest stage gage networks. In *World Environmental and Water Resources Congress 2012: Crossing Boundaries*, pages 2253–2263. ASCE, 2012.

Andrea G Veilleux and Jery R Stedinger. Bayesian gls analysis of california regional skew. In *World Environmental and Water Resources Congress 2010 :Challenges of Change*, pages 2422–2431. ASCE, 2010.