Lecture 3.1

Update of Bulletin 17B to 17C
Expected Moments Algorithm, etc

Flood Frequency Analysis

Beth Faber, PhD, PE
Hydrologic Engineering Center

Goals

- Understand why we have Federal guidance on flood frequency analysis, and how its development has progressed
- Be aware of the differences between Bulletins 17B and 17C, and improvements Bulletin 17C can offer

Outline

- Why a Bulletin? History
- Bulletin 17B / 17C elements
  - Generalized Data description
  - Missing values, zero flows, low outliers (PILFs)
  - Historical information, high outliers
  - EMA Computation
  - Weighted skew
  - Confidence intervals

Why Flood Frequency Analysis Guidance?

- Late 1960s: many agencies were performing flood frequency analysis for many reasons
  - economic analysis, standards & permitting, project design
- National Flood Insurance Program (NFIP) defined "100-year" flow (and floodplain) as the regulatory threshold
  - a lot of **significance** to the 100-year flow estimate!
- Needed a consistent, objective, reproducible method to estimate that 100-year flow
  - *So any entity doing the analysis with the same data will get the same result (flood frequency curve, 1% exc. estimate)*
- There is often a trade-off between consistency for all sites and the best analysis for a given site…

Evolution of Guidance

1967 Bulletin 15 *brief testing*

fit LogPearson type III (**LP3**) to at-site gage data using Method of Moments

1976 Bulletin 17 *more extensive testing*

LP3, weighting with a regional skew coefficient, low-outlier censoring, use of historical information

1977 Bulletin 17A

Improved low-outlier identification and adjustment

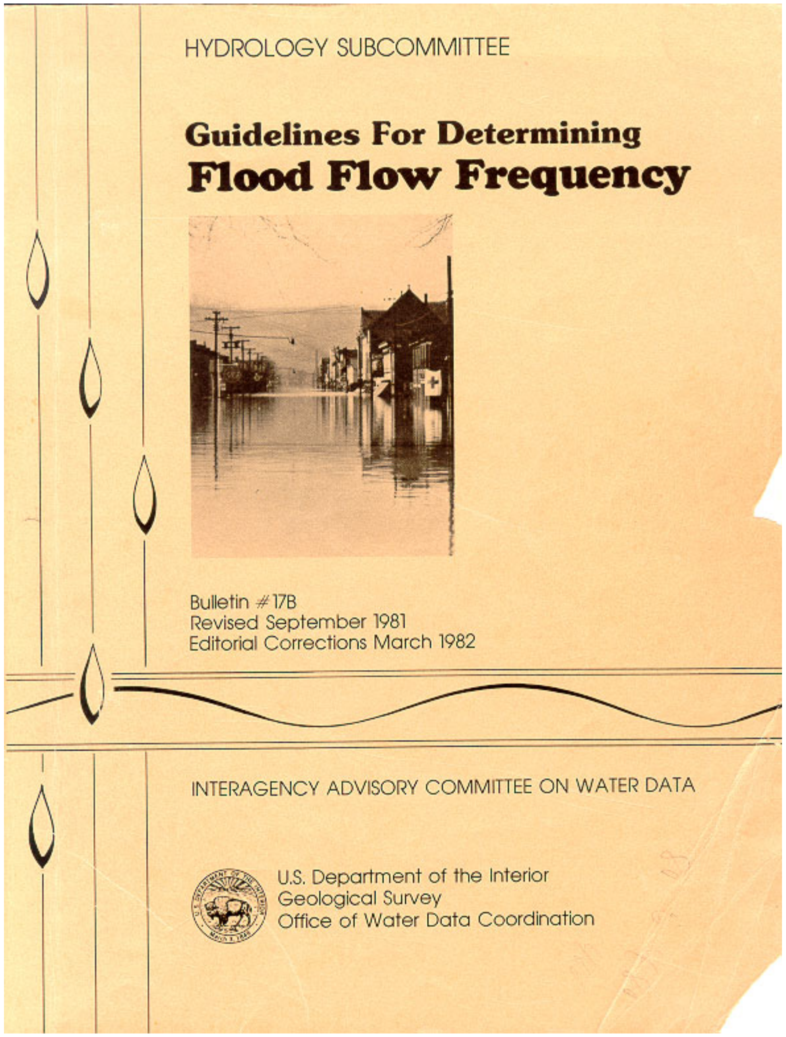
1982 Bulletin 17B

Improved use of historical information

1983 ***Water Resources Council abolished***

2018 Bulletin 17C

Fitting **LP3** by Expected Moments Algorithm (EMA), better confidence intervals, updated low outlier identification, improved regional skews



Bulletin 17B

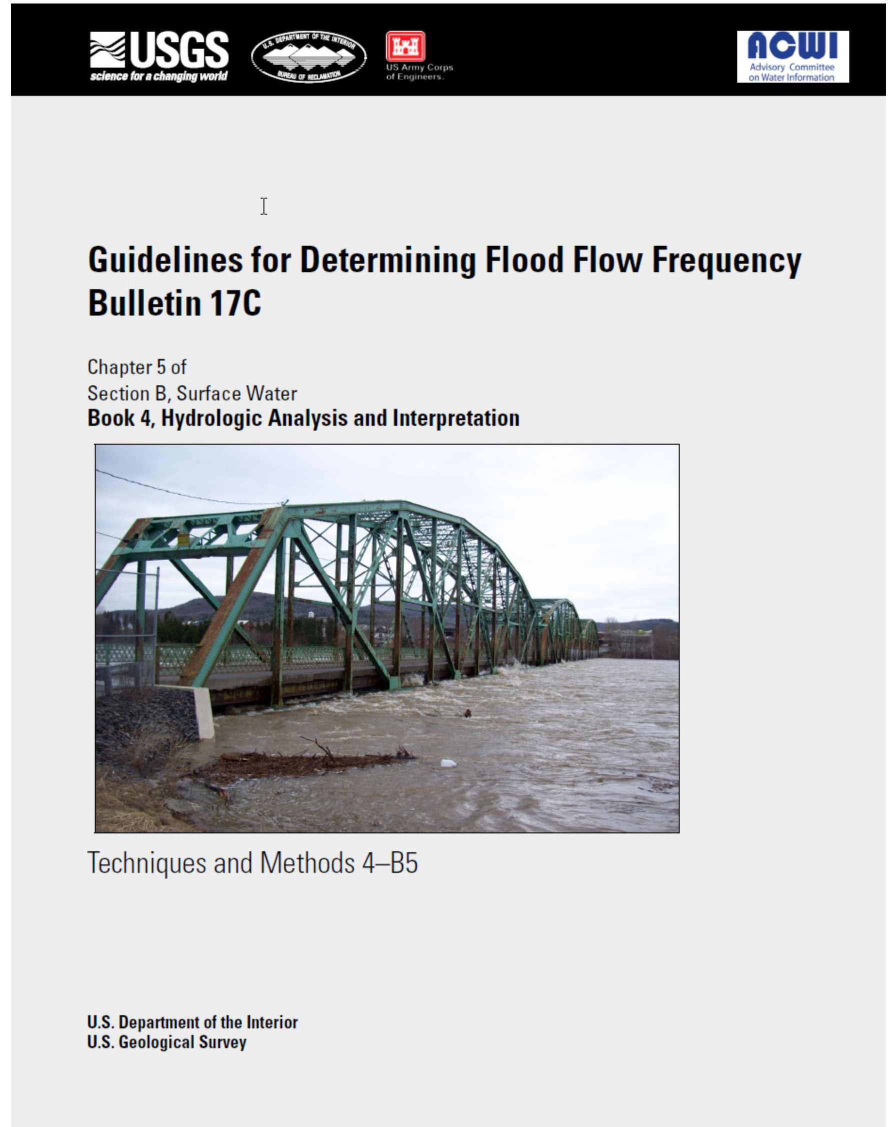
dated March 1982

←

Bulletin 17C

dated Feb 2018

→



Outline

- Why a Bulletin? History
- Bulletin 17B / 17C elements
  - Generalized Data description
  - Missing values, zero flows, low outliers (PILFs)
  - Historical information, high outliers
  - EMA Computation
  - Weighted skew
  - Confidence intervals

Bulletin 17B/C: Gaged Frequency Analysis

When a gage is present, and flows are unregulated, Bulletin 17B/C procedures must be applied

- Data set is each year's annual maximum streamflow
- Distribution fitted is **Log-Pearson III**, using MOM with $\log_{10}$flow
- Assumptions **IID**
  - annual peak flows are statistically independent **I-**
  - watershed characteristics are unchanged for the entire record (i.e., all floods from same probability distribution) **-ID**
  - sufficient record length to describe the "population" of floods

USGS's
National
Water
Information
System

http://waterdata.usgs.gov/usa/nwis/sw

USGS
science for a changing world

USGS Home
Contact USGS
Search USGS

National Water Information System: Web Interface

USGS Water Resources (Cooperator Access)

Data Category: Surface Water
Geographic Area: United States
GO

Peak Streamflow for the Nation

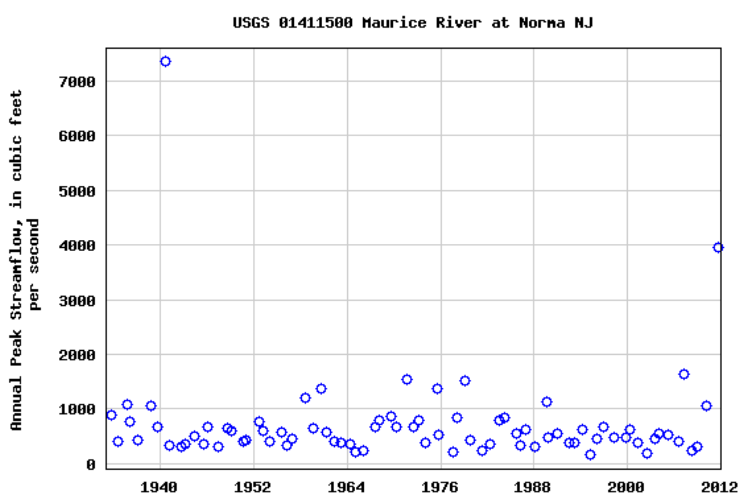
USGS 01411500 Maurice River at Norma NJ

Available data for this site: Surface-water: Peak streamflow GO

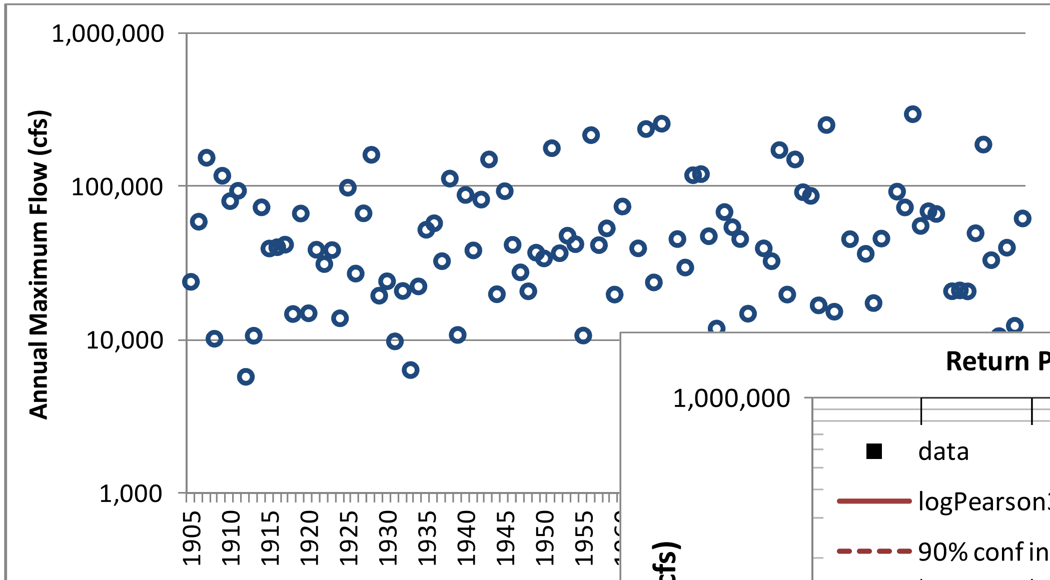
Salem County, New Jersey
Hydrologic Unit Code 02040206
Latitude 39°29'44", Longitude 75°04'37" NAD83
Drainage area 112 square miles
Gage datum 46.94 feet above NGVD29

Output formats

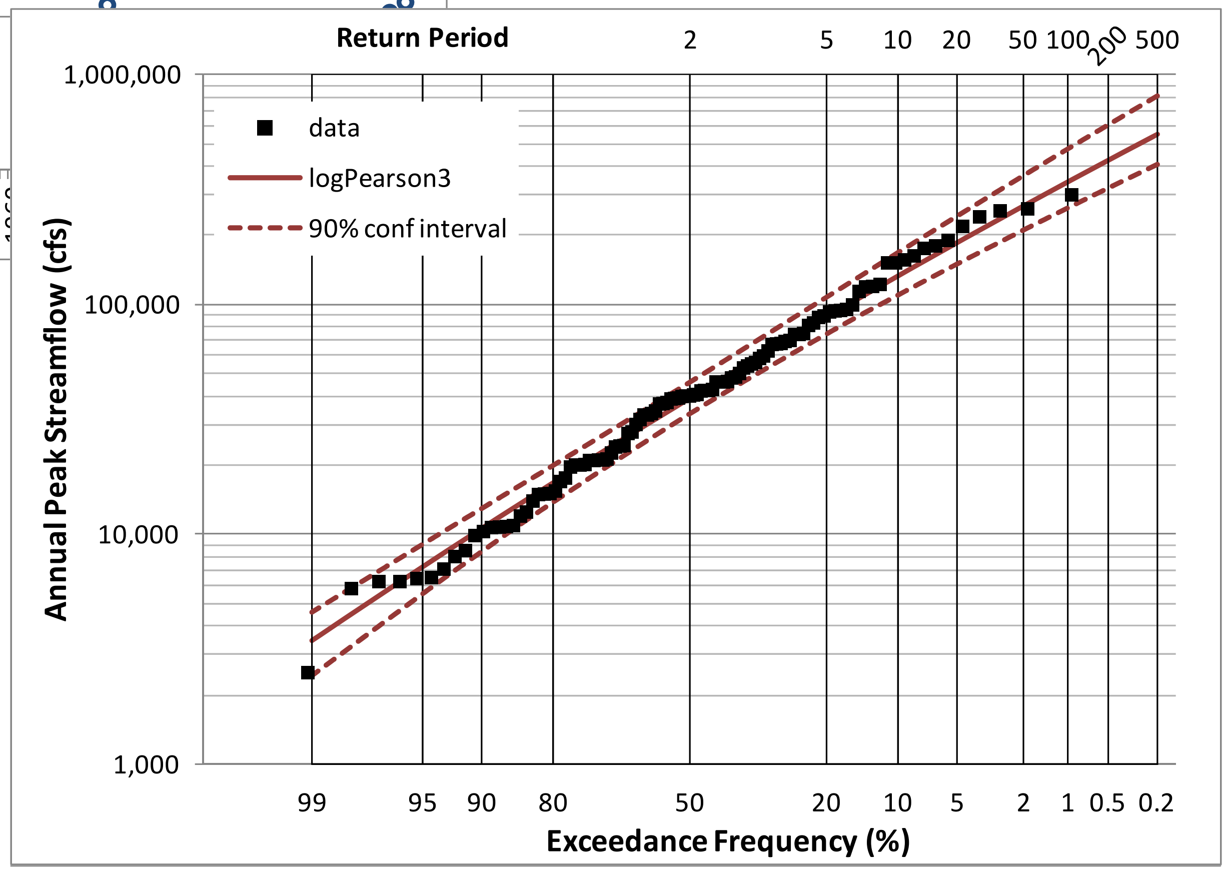
- Table
- Graph
- Tab-separated file
- peakfq (watstore) format
- Reselect output format



Download a presentation-quality graph



1,000,000
100,000
10,000
1,000
Annual Maximum Flow (cfs)
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955



Return Period
2
5
10
20
50
100
200
500
1,000,000
100,000
10,000
1,000
data
logPearson3
90% conf interval
Annual Peak Streamflow (cfs)
99
95
90
80
50
20
10
5
2
1
0.5
0.2
Exceedance Frequency (%)

Selection of Log Pearson Type III (LP3) for Annual Peak Flows

- Water Resources Council tests showed LP3 with regional info was best for estimating **100-year event**

  others: LogNormal, Gumbel, LogGumbel, 2-, 3-Gamma, but not GEV

  see Bulletin 17B, Appendix 14

- Greater flexibility than LogNormal – has a variable skew coefficient

  Normal has zero skew, logNormal has constant skew, P3 = "skewed Normal"

- **Recommendation**

  Bulletin 17B guidelines recommends use of Log-Pearson III distribution, with regional info, *unless some other distribution explains the observed data much better (must be well supported)*

"Future Work" (p. 27, B17B)

1. Selection of distribution and **fitting procedures**.
2. Mixed Distributions
3. **Low Outliers** (identification and treatment)
4. Use of **Historical data**
5. Correct **Confidence Intervals**
6. Incorporating precipitation frequency analysis.
7. Ungaged watersheds, limited gaging records.
8. Urbanization and regulation.

USGS science for a changing world   US Army Corps of Engineers   ACWI Advisory Committee on Water Information

Guidelines for Determining Flood Flow Frequency Bulletin 17C

Chapter 5 of
Section B, Surface Water
Book 4, Hydrologic Analysis and Interpretation

Techniques and Methods 4–B5

U.S. Department of the Interior
U.S. Geological Survey

Bulletin 17C

Finalized 2018

Differences in Bulletin 17C

- Uses the Expected Moments Algorithm (**EMA**) to estimate the LogPearson III distribution parameters still Method of Moments
- Allows a more general description of flow data as intervals, and requires definition of perception thresholds
- Has a more aggressive "low outlier" test
- Provides more accurate *(usually larger)* confidence intervals
- Improved regional skew method being used by USGS across US

➢ *If there are no low outliers and no historical information, the results of 17C are identical to 17B*

Computation of EMA parameter estimates

- EMA estimates product moments from sample data
  - *mean, standard deviation, skew coefficient*
- Due to the iterative nature of the EMA computation, it is not a "by hand" procedure like Bulletin 17B
- Computer code has been developed, and is present within USGS and USACE software (PeakFq, and HEC-SSP)
  - Bulletin 17C recommends use of this code

Outline

- Why a Bulletin? History
- Bulletin 17B / 17C elements
  - Generalized Data description
  - Missing values, zero flows, low outliers (PILFs)
  - Historical information, high outliers
  - EMA (Expected Moments Algorithm) computation
  - Weighted skew
  - Confidence intervals

EMA: Estimating Moments from Point Data

Mean (M), Standard Deviation (S), and Skew (G)

$$M \equiv \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$S \equiv \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i - M)^2}$$

$$G \equiv \frac{N}{S^3(N-1)(N-2)}\sum_{i=1}^{N}(X_i - M)^3$$

Method of Moments

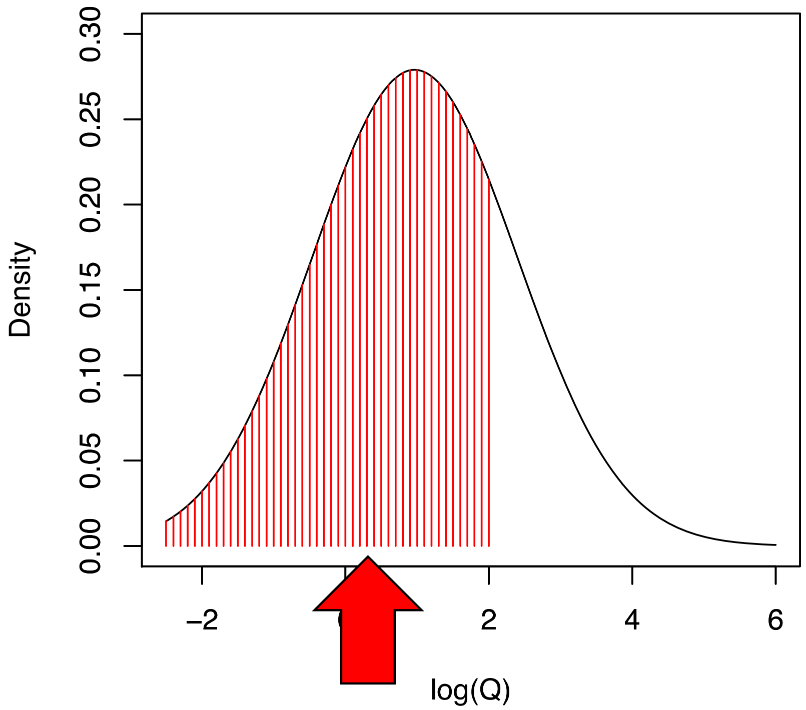
estimate population moments from sample moments

EMA: Estimating Moments from Interval Values

Moments are estimated with the assumption of **LP3**, so can use LP3 to define the likelihood of an interval

1. Given population moments
   – we can compute *expectation of mean*, *variance* and *skew* of interval observations

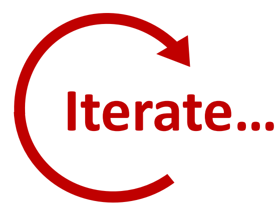
Computing Expected Moments



EMA: Estimating Moments from Interval Values

Moments are estimated with the assumption of **LP3**, so can use LP3 to define the likelihood of an interval

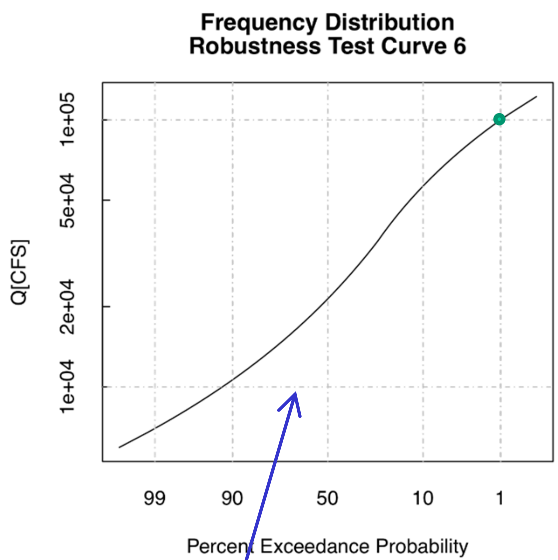
1. Given population moments
   – we can compute *expectation of mean*, *variance* and *skew* of interval observations
2. Given *expected mean, variance* and *skew* of interval observations,
   – we can estimate population moments



- Initial guess, and iterate…

Benefits of Expected Moments Algorithm (EMA)

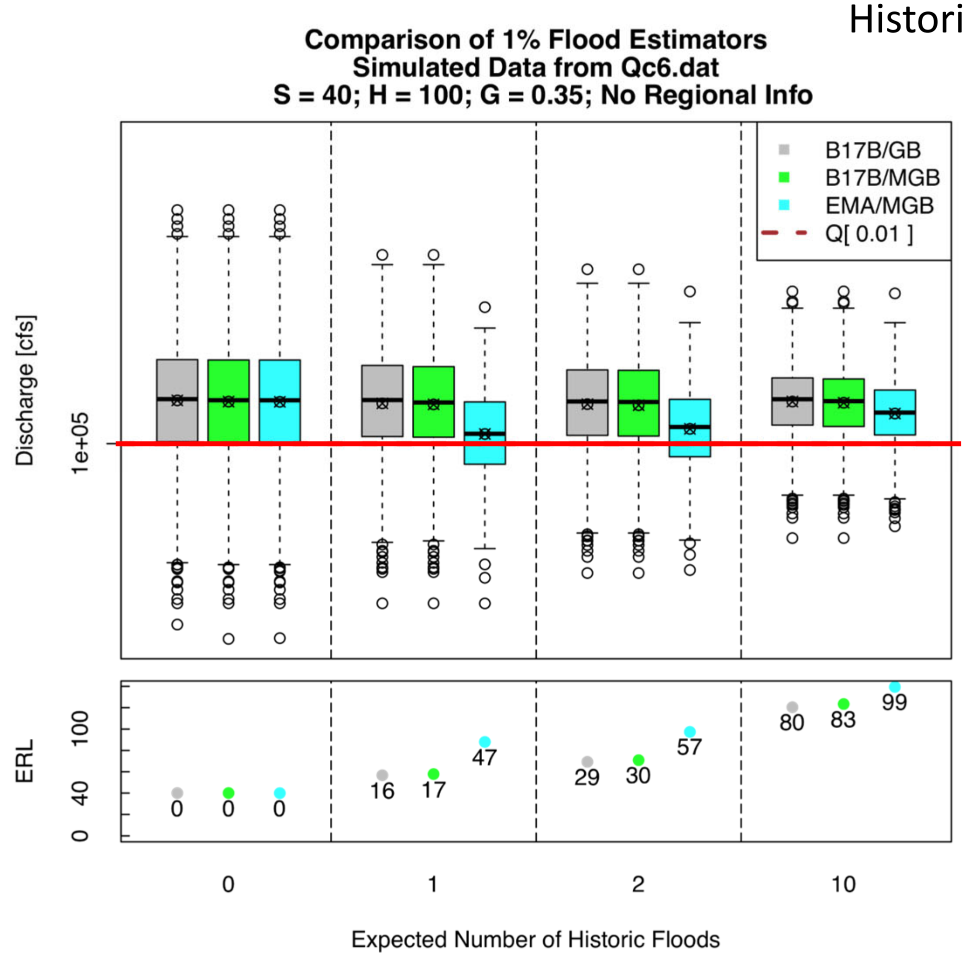
1. Consistent with Bulletin 17B *(LP3, MOM, etc)*
   - *Identical result if no historical info or low outliers*
2. Statistically efficient and robust



Not an LP3 distribution!

Estimating 1% event

Systematic = 40 years
Historical = 100 years



17B / GB
17B / MGB
17C / MGB

historical event threshold

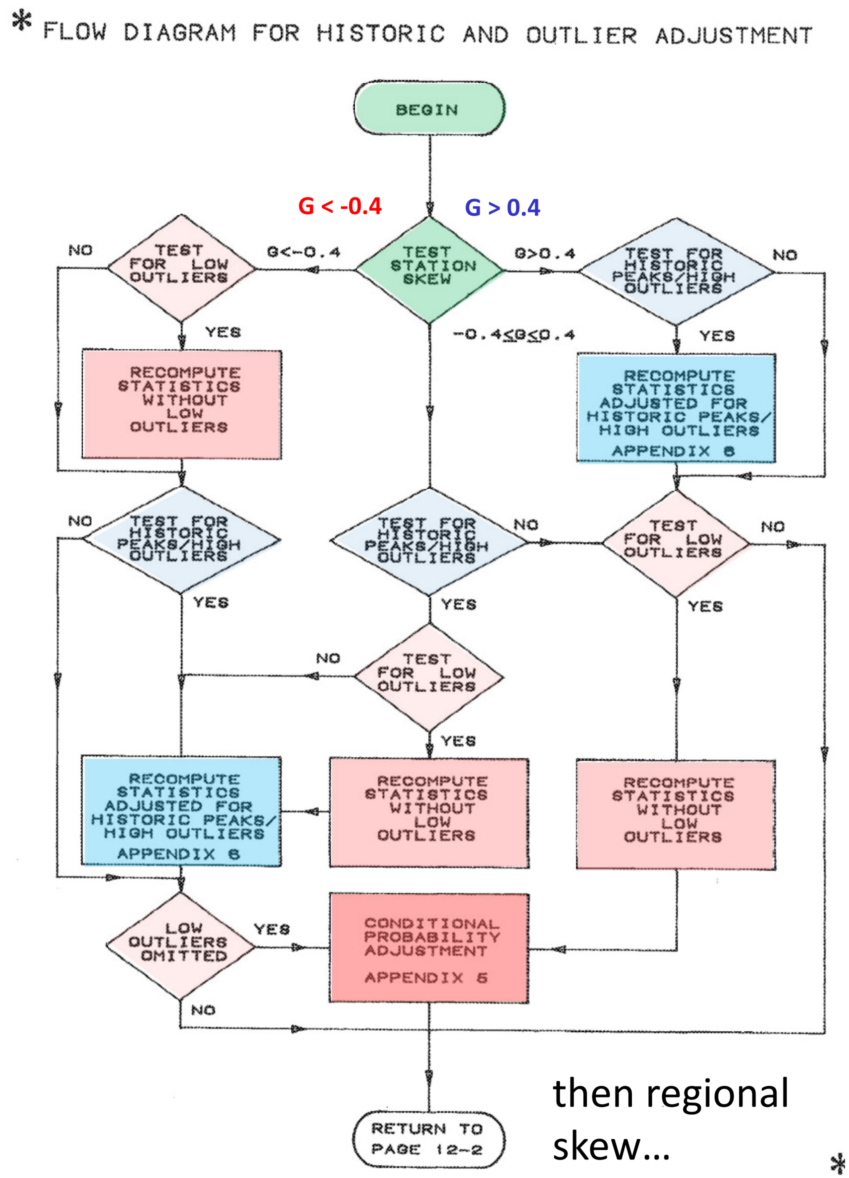
none 100-yr 50-yr 10-yr

Benefits of Expected Moments Algorithm (EMA)

1. Consistent with Bulletin 17B *(LP3, MOM, etc)*
   - *Identical result if no historical info or low outliers*
2. Statistically efficient and robust
3. Statistically tractable
4. <u>Simultaneous</u> consideration of low outliers, historical information and regional skew

Bulletin 17B: Complicated Flowcharts defining order of Computations

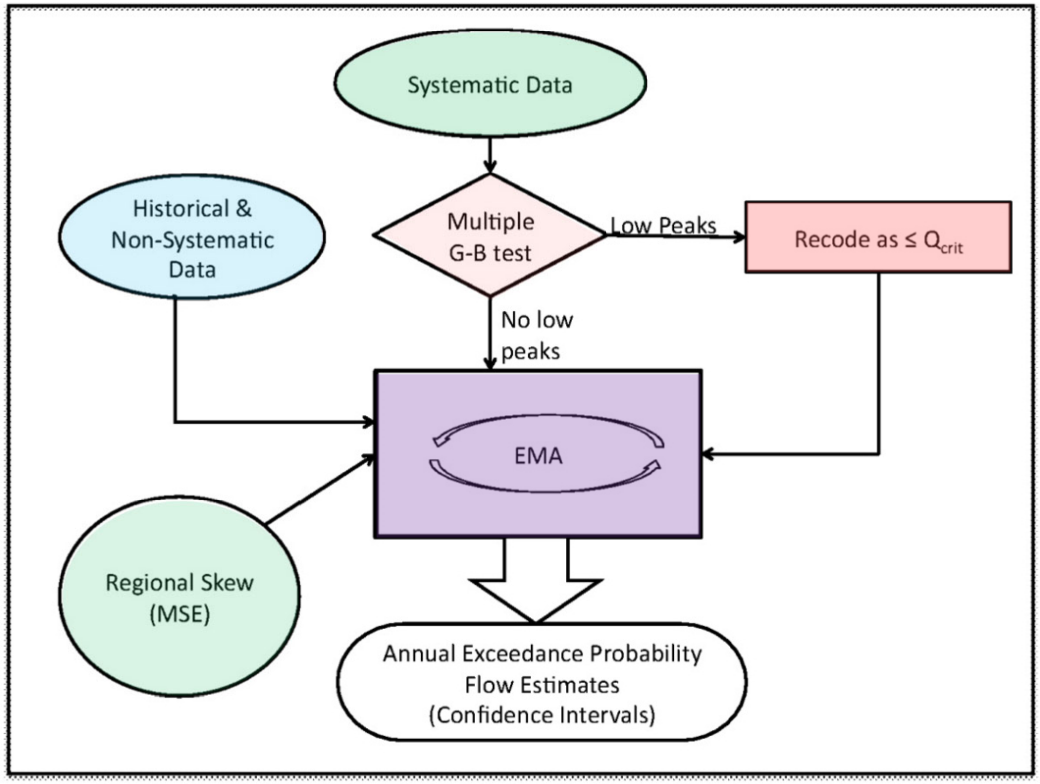
"Specialized" Procedures for Historic data and outlier adjustments



then regional skew…

*

EMA Flowchart for Bulletin 17C



Benefits of Expected Moments Algorithm (EMA)

1. Consistent with Bulletin 17B *(LP3, MOM, etc)*
   – *Identical result if no historical info or low outliers*
2. Statistically efficient and robust
3. Statistically tractable
4. <u>Simultaneous</u> consideration of low outliers, historical information and regional skew
5. More accurate confidence intervals

Outline

- Why a Bulletin? History
- Bulletin 17B / 17C elements
  - Generalized Data description
  - Missing values, zero flows, low outliers (PILFs)
  - Historical information, high outliers
  - EMA Computation
  - Weighted skew
  - Confidence intervals

Weighted Skew, B17B

- Skew is very difficult to estimate well from a limited sample.
  G_s = station skew
- Improve the estimate with a regional skew, developed using many gages in the region ("trade space for time")
  G_r = regional skew
- Can weight the skew based on the relative error in the skew estimates **$G_w = W_r G_r + W_s G_s$** *where* $(W_r + W_s) = 1$

MSE = mean squared error

$$G_w = \frac{{}^{1}/_{MSE_{G_r}}\, \mathbf{G_r} + {}^{1}/_{MSE_{G_s}}\, \mathbf{G_s}}{{}^{1}/_{MSE_{G_r}} + {}^{1}/_{MSE_{G_s}}}$$

$$G_w = \frac{MSE_{G_s}\mathbf{G_r} + MSE_{G_r}\mathbf{G_s}}{MSE_{G_r} + MSE_{G_s}}$$

Weighted Skew, B17C

- In the Expected Moments Algorithm, additional information (historical info, regional skew, etc) is brought in **simultaneously** with the systematic record, not after
- The skew weighting equation is not at the end, as in 17B, but part of the skew estimation in each iteration. ***Uses N in place of 1/MSE***.

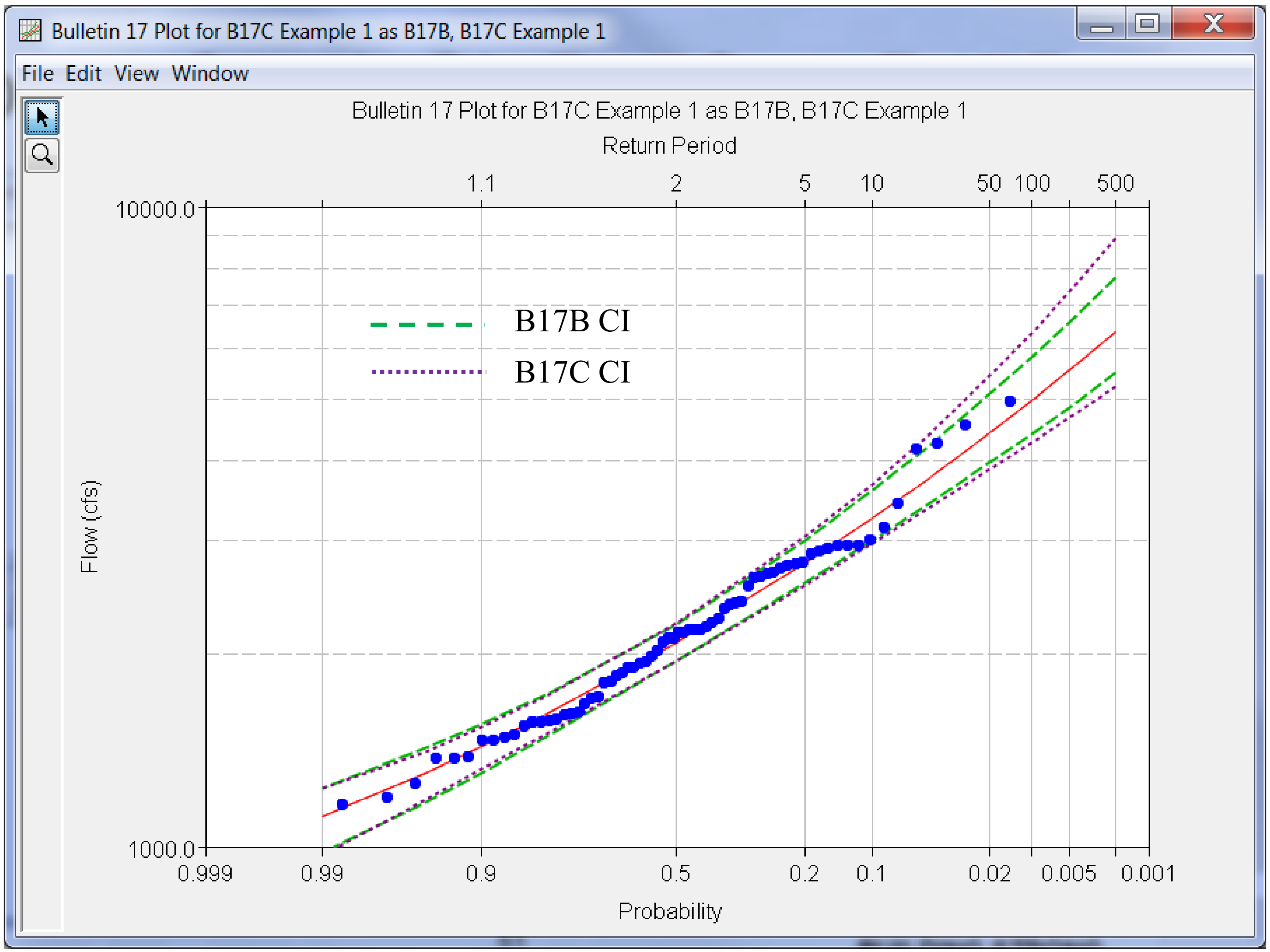
systematic censored regional

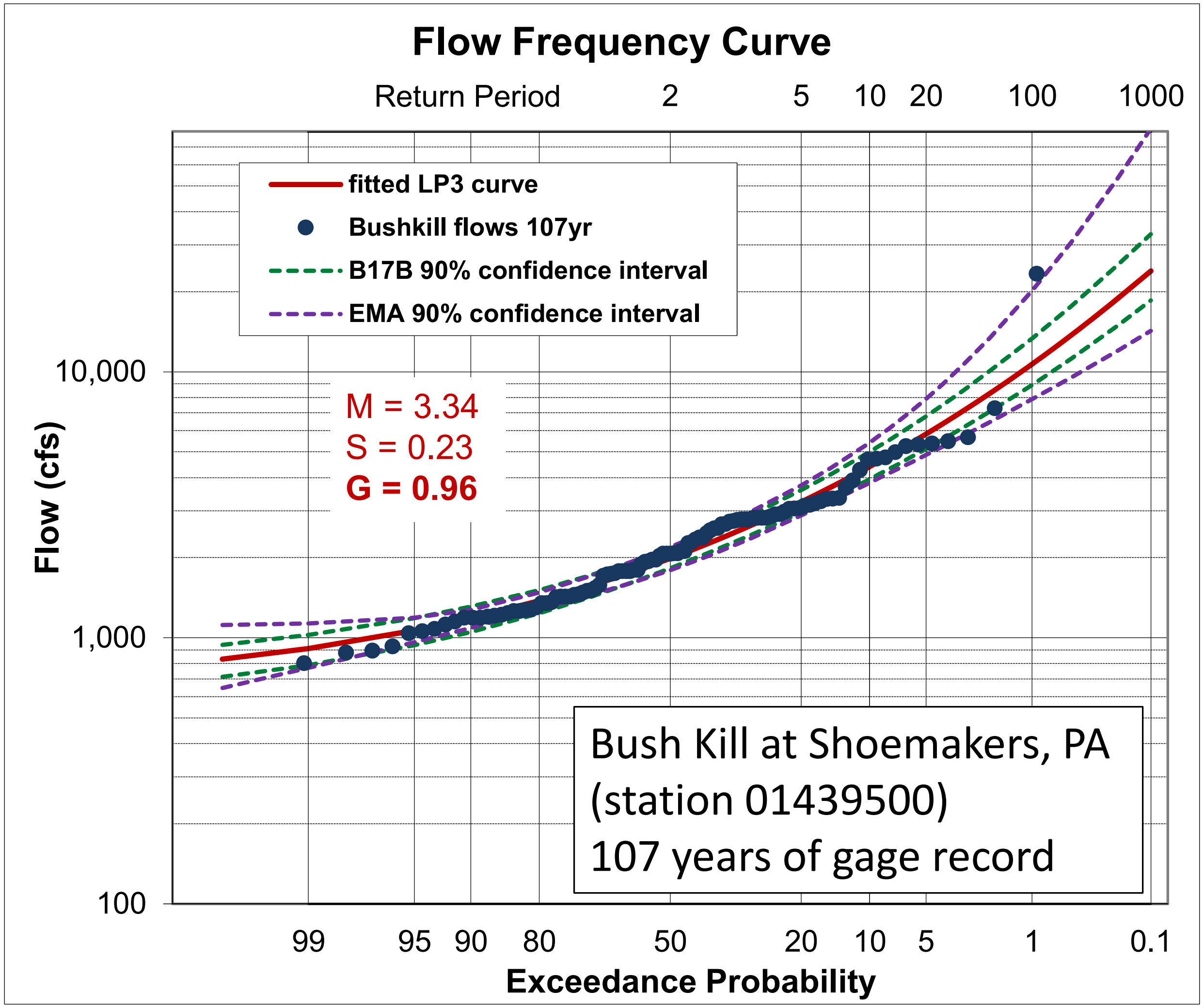
$$(7-10)\quad \hat{\gamma}_{j+1} = \frac{\sum_{i \in S}\left(\frac{X_i-\mu}{\sigma}\right)^3 + \sum_{i \in H} E\left[\left(\frac{X_i-\mu}{\sigma}\right)^3 \middle| T, \mu, \sigma, \hat{\gamma}_j\right] + N_R \hat{G}}{N_S + N_H + N_R}$$

More Correct Confidence Intervals

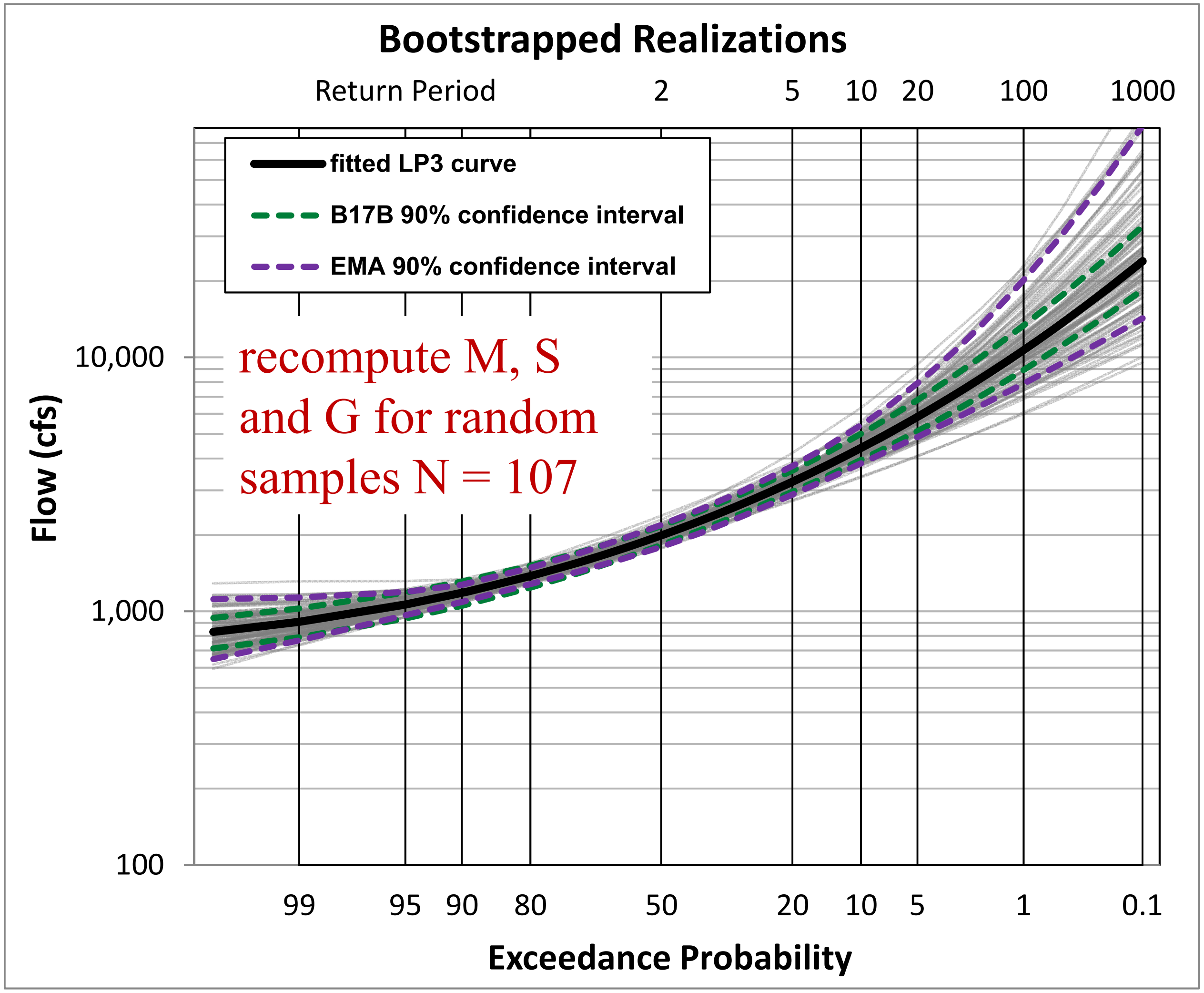
- We show confidence intervals around the estimated frequency curve to demonstrate our uncertainty
  - usually show 90% confidence interval
- The confidence interval formulas in B17B are based on log Normal – which has defined zero skew (can’t be changed)
  - intervals therefore <u>neglect</u> the uncertainty in the estimated <u>skew</u>
- EMA intervals **DO** include uncertainty in skew
- And also reflect <u>historical information</u> and <u>low outlier</u> adjustments

Moose Creek CI, B17C vs B17B

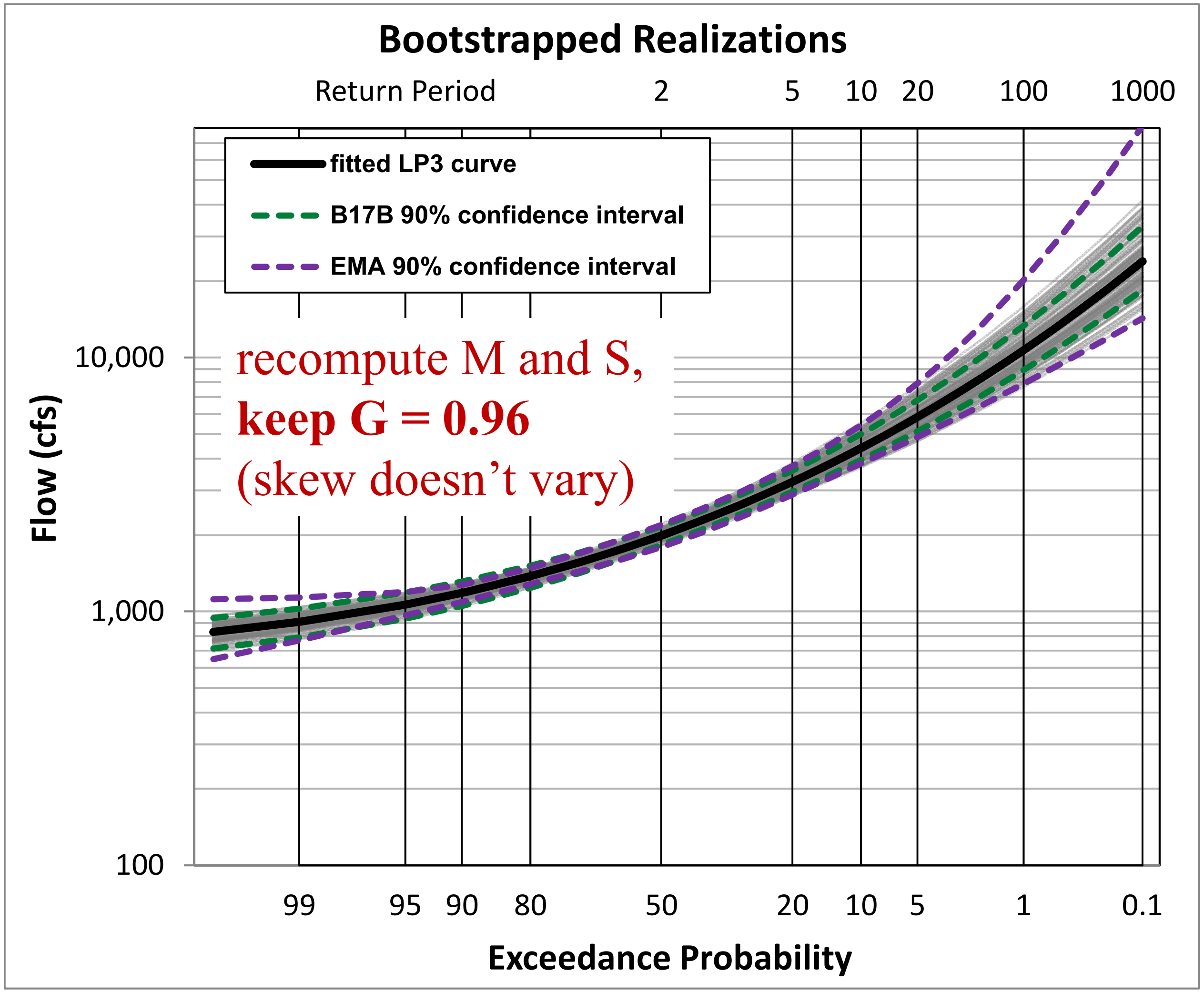




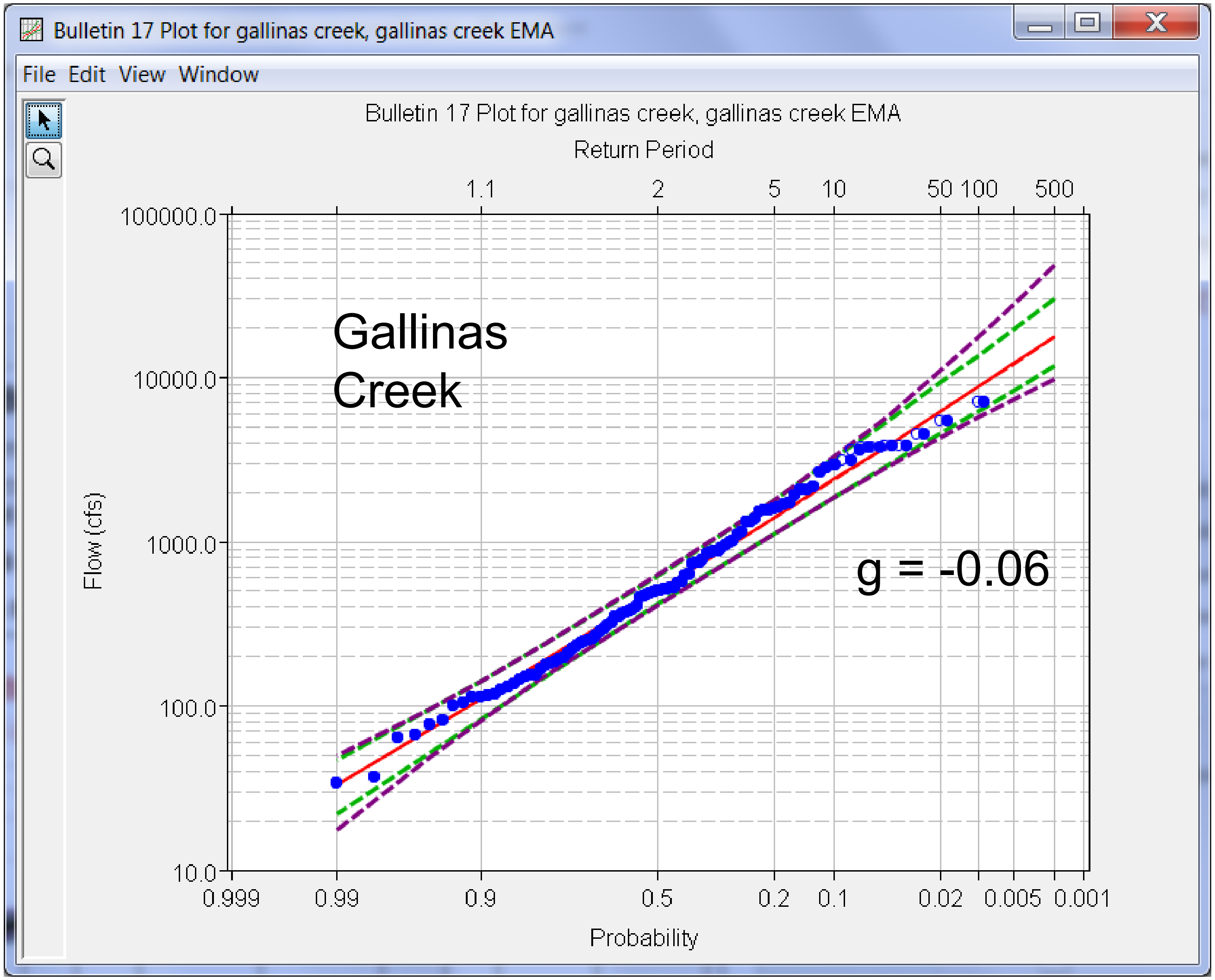
when estimated skew is far from zero, greater uncertainty and so greater difference in intervals



"parametric bootstrap"



"parametric bootstrap"



B17B and B17C confidence intervals more similar when estimated skew ≈ 0

Uncertainty Description

Bulletin 17B:

- The confidence intervals described in Bulletin 17B use the non-central t distribution
- They do not consider the uncertainty in the skew

Bulletin 17C:

- The confidence intervals in 17C do capture skew uncertainty, and so are more correct and generally wider than those from 17B

What 17B/C Does NOT Address

- Variables other than unregulated, instantaneous peak flow
  - max flow volume, channel stage, reservoir volume, precipitation, soil moisture, wave height, etc
- Flood frequency curves in un-gaged areas
- Regulated flows
- When the assumptions about **homogeneity**, reliability, randomness are challenged…
  - However, 17C recommends tests for trend and change
  - Allows for other methods, such as time-varying parameters

Outline

- Why a Bulletin? History
- Bulletin 17B / 17C elements
  - Generalized Data description
  - Missing values, zero flows, low outliers (PILFs)
  - Historical information, high outliers
  - EMA Computation
  - Weighted skew
  - Confidence intervals