Comparing Statistical Approaches to Estimating Floods

Johannes (Joe) DeVries, Ph.D., P.E., P.H., D.WRE
David Ford Consulting Engineers
2015 J St., Suite 200
Sacramento, CA 95814

Tel: (916) 447-8779
E-mail: jjdevries@ford-consulting.com
Web: www.ford-consulting.com

BIOGRAPHICAL SKETCH

Joe DeVries is an internationally recognized expert in hydrology and hydraulic engineering and has over 40 years of experience. He spent nearly two decades at the University of California, Davis as a research engineer and lecturer. He worked as a hydraulic engineer/hydrologist for the California Department of Water Resources and for various consulting engineering firms. He spent 4 years as a hydraulic engineer at the Hydrologic Engineering Center. He taught in the Departments of Civil and Environment Engineering and Land Air and Water Resources at U.C. Davis. He was Associate Director of the Water Resources Center of the University of California for five years

He has served as a consultant to the Corps of Engineers, National Weather Service, government agencies in California, Argentina, India, and Taiwan, the United Nations, the World Bank, USAID, and to several engineering firms. For past five years he has been employed by David Ford Consulting Engineers in Sacramento.

Dr. DeVries has a PhD in Hydraulic Engineering from the University of California, Davis; MS and BS degrees in Civil Engineering from the University of Michigan, Ann Arbor; and a BS in Engineering and Science from Calvin College in Grand Rapids, Michigan.

He is a registered civil engineer in California (1963) and a registered hydrologist with the American Institute of Hydrology (1986). He is a Life Member of the American Society of Civil Engineers (2001) and was inducted as a Diplomate of the American Academy of Water Resources Engineers (October 2005).

ABSTRACT

Results from computing the flood frequency on the American River using 3-day discharge data using different methods is shown. The analyses were made only to provide comparisons between the various methods and not make any recommendations as to the most appropriate flood flow frequency method to use. Extrapolations are made for the 100-year, 200-year, and 500-year floods.

Comparing Statistical Approaches to Estimating Floods

Joe DeVries, Ph.D., P.E., P.H., D.WRE
David Ford Consulting Engineers

Michael Anderson, Ph.D., P.E.
State Climatologist, California Department of Water Resources

Presentation Outline

- Introduction
- Graphical approach
- Bulletin 17B
- Mixed populations
- A stochastic model
  (MGS report to HEC, 2005)
- Conclusions

Introduction

- Driving Questions:
  - How do we define flood risk?
  - How big was that flood?
- Frequency Based Description Issues:
  - What to sample
    (annual peak vs. partial duration series)
  - How long to sample
    (Is 50 years enough data?)

Looking at stream flow statistics

- Stream flow record is a sample in time from the "population" of all possible annual maximum stream flows
- Population – all stream flows over an indefinitely long time period
- For a good representation of the statistics we need:
  - Accurate records
  - Representative records
  - Sufficiently long record
  - Stationarity of record

Required length of record for flood estimation

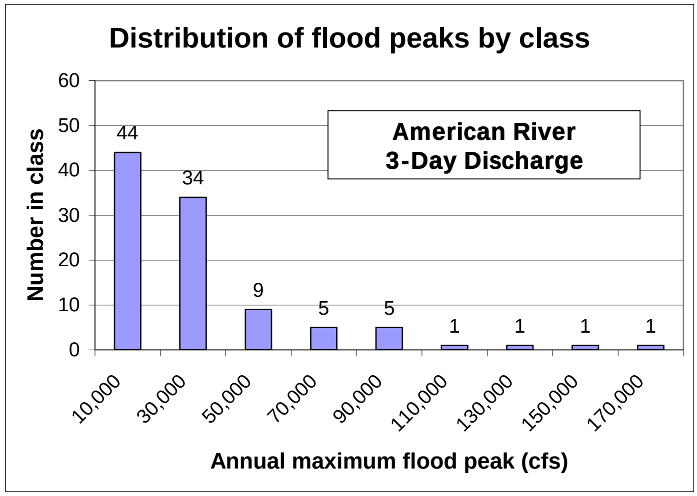
Average recurrence interval (years)	10% error	25% error
10	90	18
50	110	39
100	115	48

Length of record (years) required to estimate flood of various recurrence intervals with 95% confidence. (Gordon, McMahon, and Findlayson, 1992, *Stream Hydrology*, John Wiley & Sons, New York)

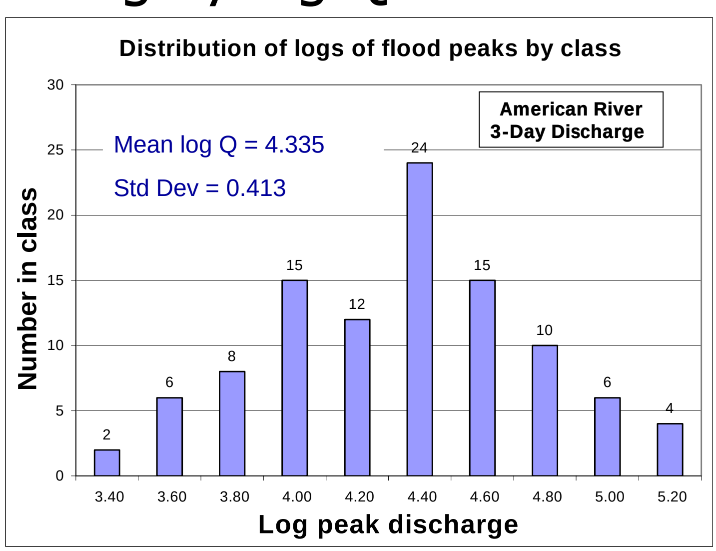
Example using American River data

- Graphical methods
- Bulletin 17B Computations
  - entire record
  - before Folsom
  - after Folsom
- Should we use Mixed Population Analysis?
- Stochastic Model Results

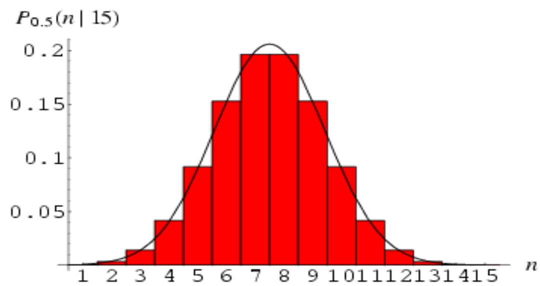
Ranking of floods by class



Ranking by log Q



Example: Normal distribution as limiting case of a discrete distribution



The normal distribution is the limiting case of a discrete binomial distribution $P_p(n \mid N)$ as the sample size N becomes large, in which case $P_p(n \mid N)$ is normal with mean and variance

$$\mu = N p \quad (4)$$
$$\sigma^2 = N p q. \quad (5)$$

Plotting positions

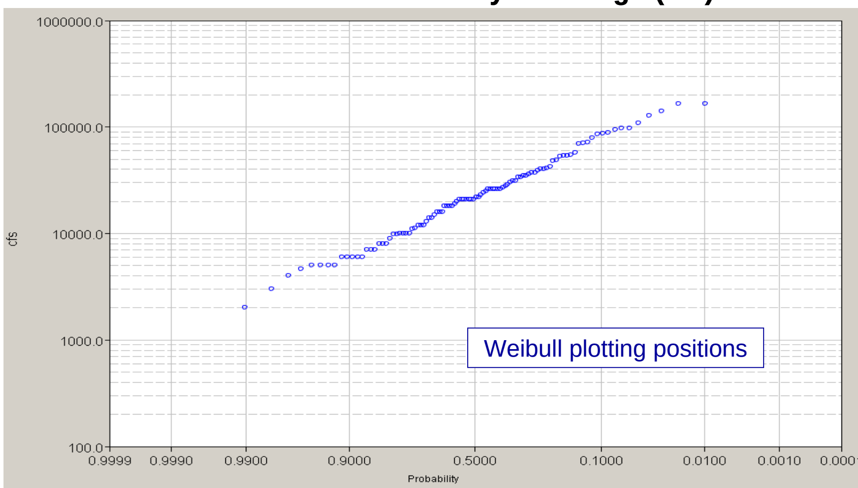
- Different plotting position formulas tend to give similar values in middle of data, but give different values at ends
- Weibull – most commonly applied
- GEV - general extreme value distribution from Hosking
- Cunnane – fits many distributions. Cunnane found some formulas were biased & overestimate highest flood peaks & recommended the value of a = 0.4.

Name	Probability of Exceedance, P	Avg. Recurrence interval, T
Weibull	m/(N + 1)	(N + 1)/n
GEV	(m – 0.35)/N	N/(m – 0.35)
Cunnane	(m-a)/[(N+1)-2a]	[(N+1)-2a]/(m-a)

N = number of years of record, m = rank of the event, a = a constant

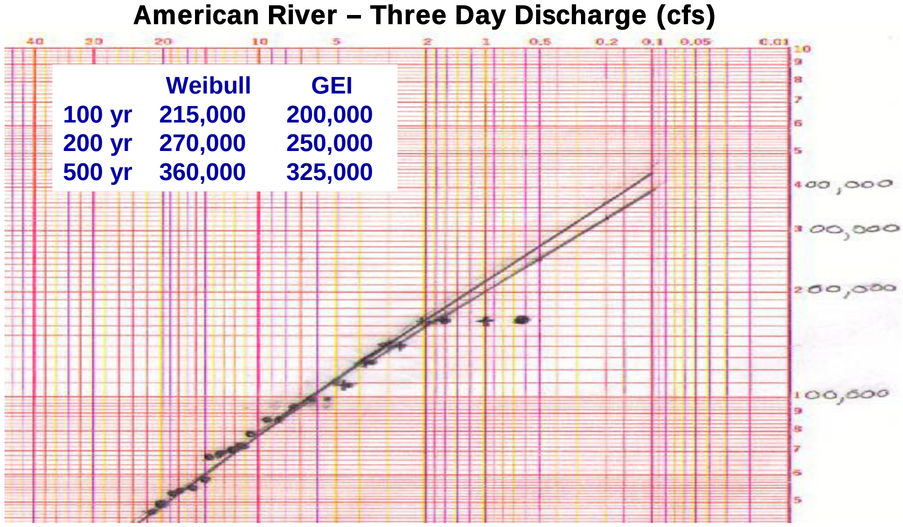
Graphical frequency analysis

American River – Three Day Discharge (cfs)



Graphical frequency analysis

American River – Three Day Discharge (cfs)



	Weibull	GEI
100 yr	215,000	200,000
200 yr	270,000	250,000
500 yr	360,000	325,000

Problem with graphical analysis

- Method can work well for defining frequent floods
- It is recommended that for larger recurrence intervals a theoretical probability distribution should be fitted to obtain "more consistent, objective estimates." (Gordon, *et al.* 1992)

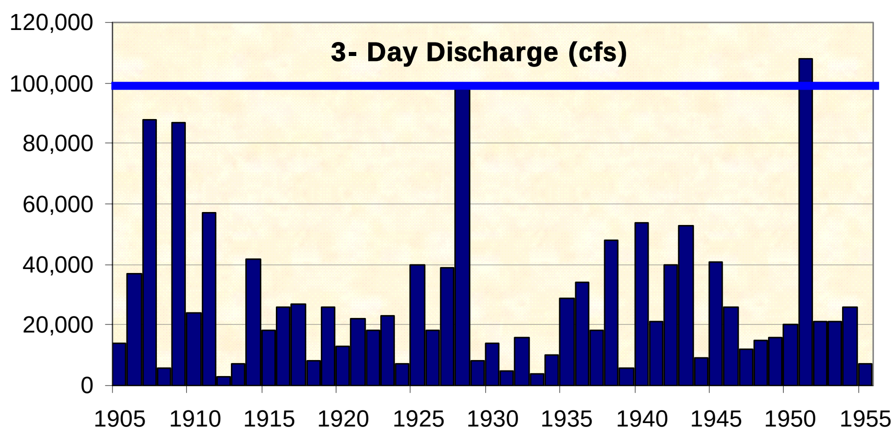
Bulletin 17B

- Flood frequency guidelines published by Water Resources Council
- Suggests Log Pearson Type III (LP III) distribution
- Minimum of 10 years of data needed, 30 preferable
- Three parameters:
  - Mean
  - Standard deviation
  - Skew factor

American River Example

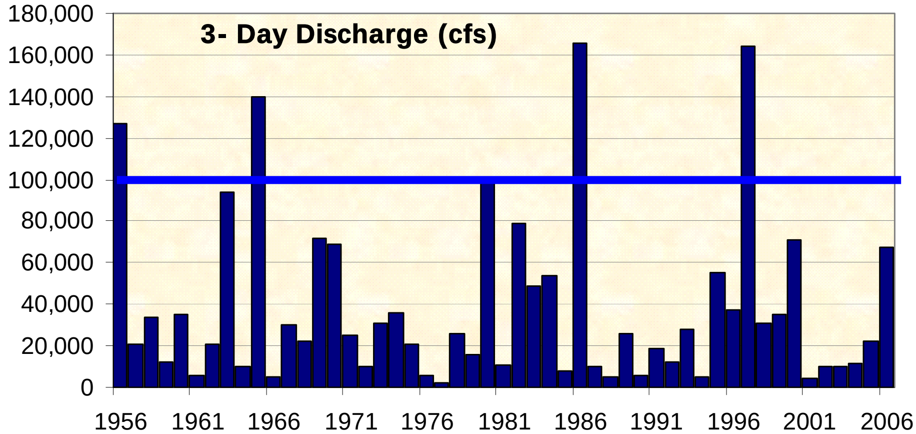
- Annual peak flow record from USGS
- 102 years of record
- Record split at 1955 (pre/post Folsom Dam)
- Following Bulletin 17B guidelines 1%, 0.5%, and 0.2% exceedence (100, 200, and 500-year return period) flows were computed

50-year record American River flows 1905-1955



First 50 years of flow data at Fair Oaks and encompasses data set used to size Folsom Dam. Note only 2 peaks come near or pass 100,000 cfs.

50-year record, part 2
American River flows 1956-2006



Second 50 years after Folsom Dam built. Note 4 peaks over 120,000 cfs & 2 others close to 100,000 cfs. Larger & more peaks than 1st 50 years. Size & number of flood peaks have distinct impact on flood frequency estimates.

Bulletin 17B Results: 3-Day Discharge

American River	Total Record	Pre - 1955	Post - 1955
Mean	4.34	4.31	4.37
Std. Dev.	0.41	0.37	0.46
Skew	-.010	-.109	-.035
100-yr Q (cfs)	196,000	133,000	262,000
200-yr Q (cfs)	248,000	161,500	338,000
500-yr Q (cfs)	330,000	202,500	460,000

Note 100-year flow for post-1955 sample is greater than 500-year flow for pre-1955 sample and greater than 200-year flow for total record.

Mixed populations

- Floods may be caused by different types of storms
- It may be appropriate to apply a “mixed-population analysis” if the storm characteristics are different
- Examples are hurricane floods vs. general storm floods, ice-affected flooding vs. warm weather floods, snow melt, etc.

How about the American River?

- Should we consider mixed-population analysis for the American River?
- More than one type of storm produces floods on this watershed
- Can we consider this “mixture distribution” as being well approximated by LP III distribution?

Approximation by LP III

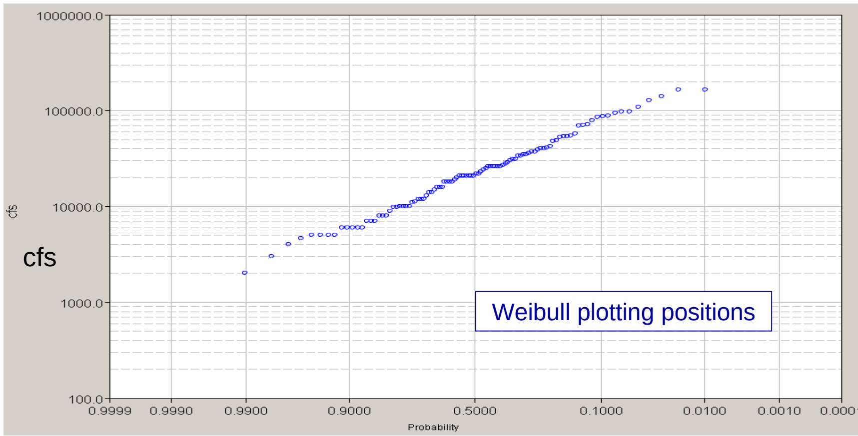
- We compared the fitted LPIII with the frequency curve defined by plotting observed flood magnitudes versus empirical probability plotting positions.
- For the American River the resulting mixture distribution appears to be well-approximated by a log-Pearson Type III distribution (LPIII).

FAQ on Mixed Populations

If the fitted LP III is compared with the frequency curve and there is good agreement between them, then "… there is no benefit in going through the lengthy mixture calculation."

"In practice, one determines whether the distribution is well-approximated by the LP III by comparing the fitted LPIII with the sample frequency curve defined by plotting observed flood magnitudes vs. empirical probability plotting positions."

American River frequency curve: 3-Day Discharge



Only if the sample frequency curve has sharp curvature (kinks), reverse curves, or other characteristics that prevent its being approximated by the LPIII, or if the available flood record omits important sources of flooding, is there any reason to perform a mixed-population analysis. (http://acwi.gov/hydrology/Frequency/B17bFAQ.html Mixed Populations)

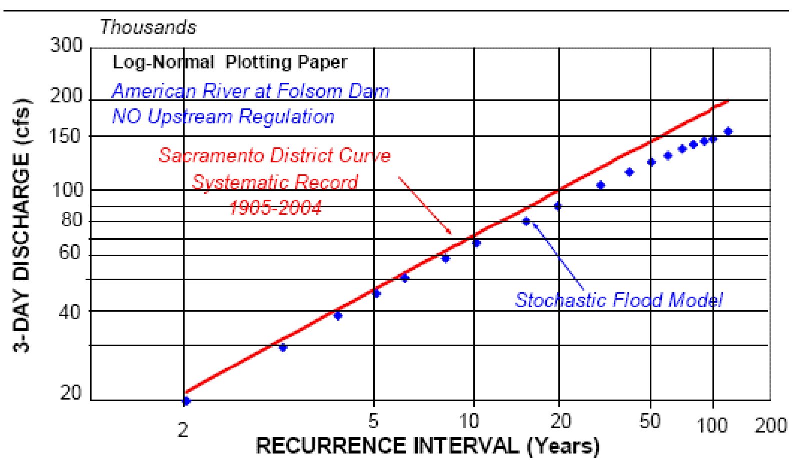
MGS Stochastic Model

- HEC-1 model of American River Watershed
- 12 parameters treated as random variables
  - Seasonality of storm occurrence
  - 72-hour storm magnitude
  - Temporal and spatial distribution of storm
  - Temperature temporal pattern
  - Sea level temperature
  - Freezing level
  - Antecedent precipitation
  - Antecedent snowpack
  - Antecedent soil moisture
  - Upstream reservoir storage
  - Initial stream flow
  - Folsom Lake Storage

MGS Stochastic Model - continued

- Model calibrated to match 100-year flow from systematic record
- 24 candidate storms from 1950-2002 period used as prototype storms for stochastic storm resampling
- 75,000 simulations run to generate magnitude-frequency relationships
- Full report on model including 13 appendices available on HEC's website at www.hec.usace.army.mil/publications/pub_download.html

MGS Stochastic Model



MGS Stochastic Model Results

American River – Three Day Discharge

Return Period	Flood Peak Discharge
100 year	150,000 cfs
200 year	180,000 cfs
500 year	223,000 cfs

Summary of Results (in cfs)

American River – Three Day Discharge

Return Period	Bulletin 17B (first 50 yrs)	Bulletin 17B (last 50 yrs)	Bulletin 17B (100 yrs)	Graphical Weibull	Graphical GEI	Stochastic Model
100 Year	133,000	262,000	196,000	215,000	200,000	150,000
200 Year	161,500	338,000	248,000	270,000	250,000	180,000
500 Year	202,500	460,000	330,000	360,000	325,000	223,000

Estimated Standard Project Flood Peak Discharge approx. 205,000 cfs

What we looked at:

- Graphical approach
- Bulletin 17B
  - Including split record
- Evaluation of using mixed-population analysis
- A stochastic model

Conclusions

- All methods have the common problem of extrapolation to infrequent flood events
- Using only the last 50 yrs of record for the American River in estimating infrequent flood events provides the most conservative estimate
- Stochastic modeling gave lower flood estimates than frequency-based analyses

Disclaimer

- Our analyses were made only to provide comparisons between the various methods
- The numerical values that we determined may or may not be the correct values for the estimated flood magnitudes
- We are not making a recommendation as to the most appropriate flood flow frequency method to use