Limitations on Extrapolating Flood Frequency Distributions

Beth Faber, Ph.D., P.E., Research Hydraulic Engineer
Hydrologic Engineering Center
U.S. Army Corps of Engineers
609 Second Street
Davis, CA 95616

Tel: (530) 756-1104
Email: Beth.A.Faber@hec01.usace.army.mil
Web: www.hec.usace.army.mil

BIOGRAPHICAL SKETCH

Beth Faber received her Bachelor of Science degree in Civil Engineering from the University of Colorado at Boulder in May 1991. From June 1991 to July 1995, she was employed by Denver Water in Denver, Colorado, performing analysis and operation of the city's reservoir system. During that period she also completed a Master's degree in Civil Engineering at the University of Colorado at Boulder.

Between 1995 and 2000, she completed the doctoral program at Cornell University, interspersed with brief periods of employment with Pacific Gas & Electric in San Francisco and Denver Water. She is currently employed at the US Army Corps of Engineers' Hydrologic Engineering Center in Davis, California.

Extrapolating Frequency Curves, or Not

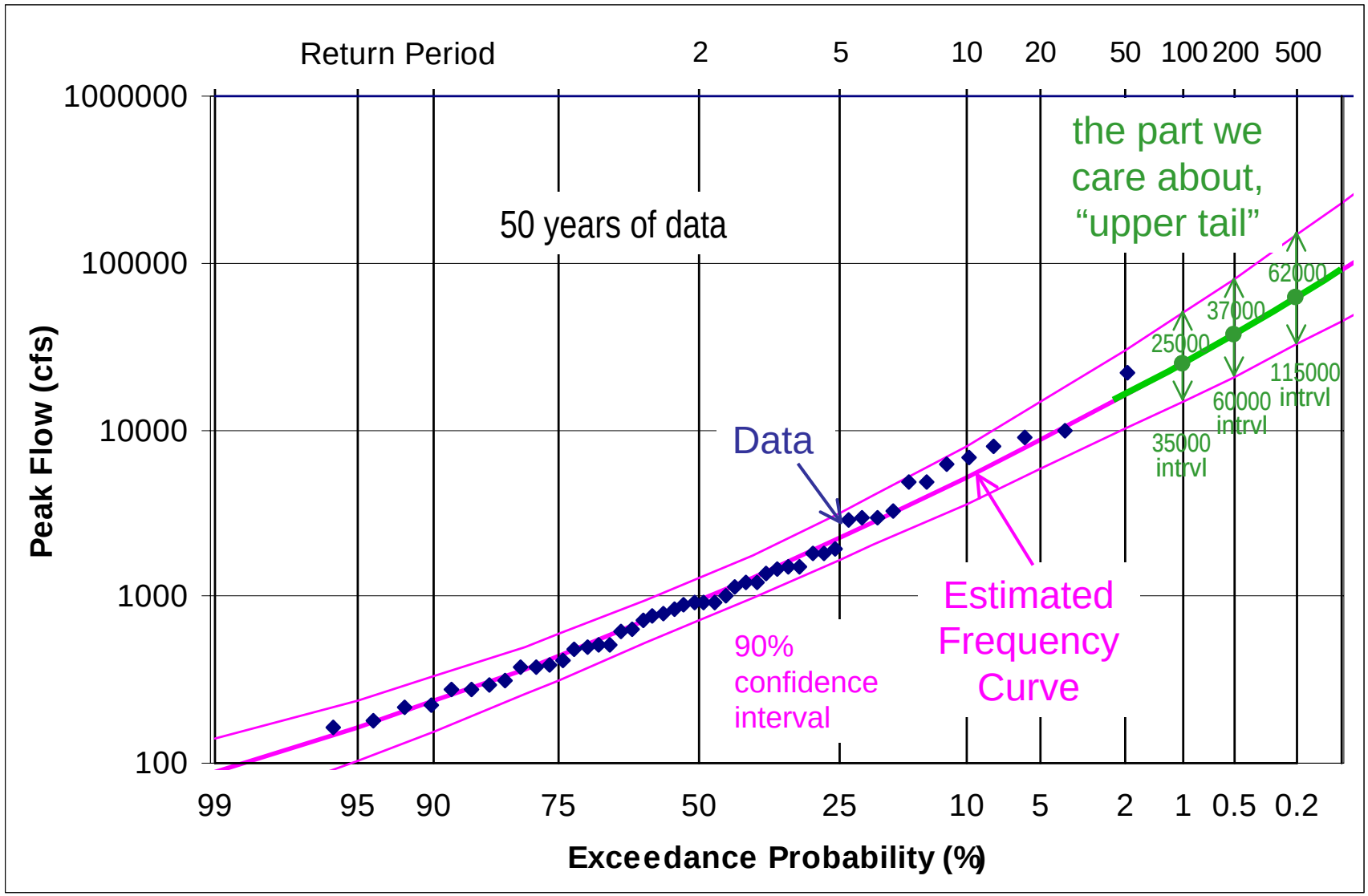
Beth Faber, Ph.D., P.E.
Hydrologic Engineering Center
U.S. Army Corps of Engineers

April 13, 2007

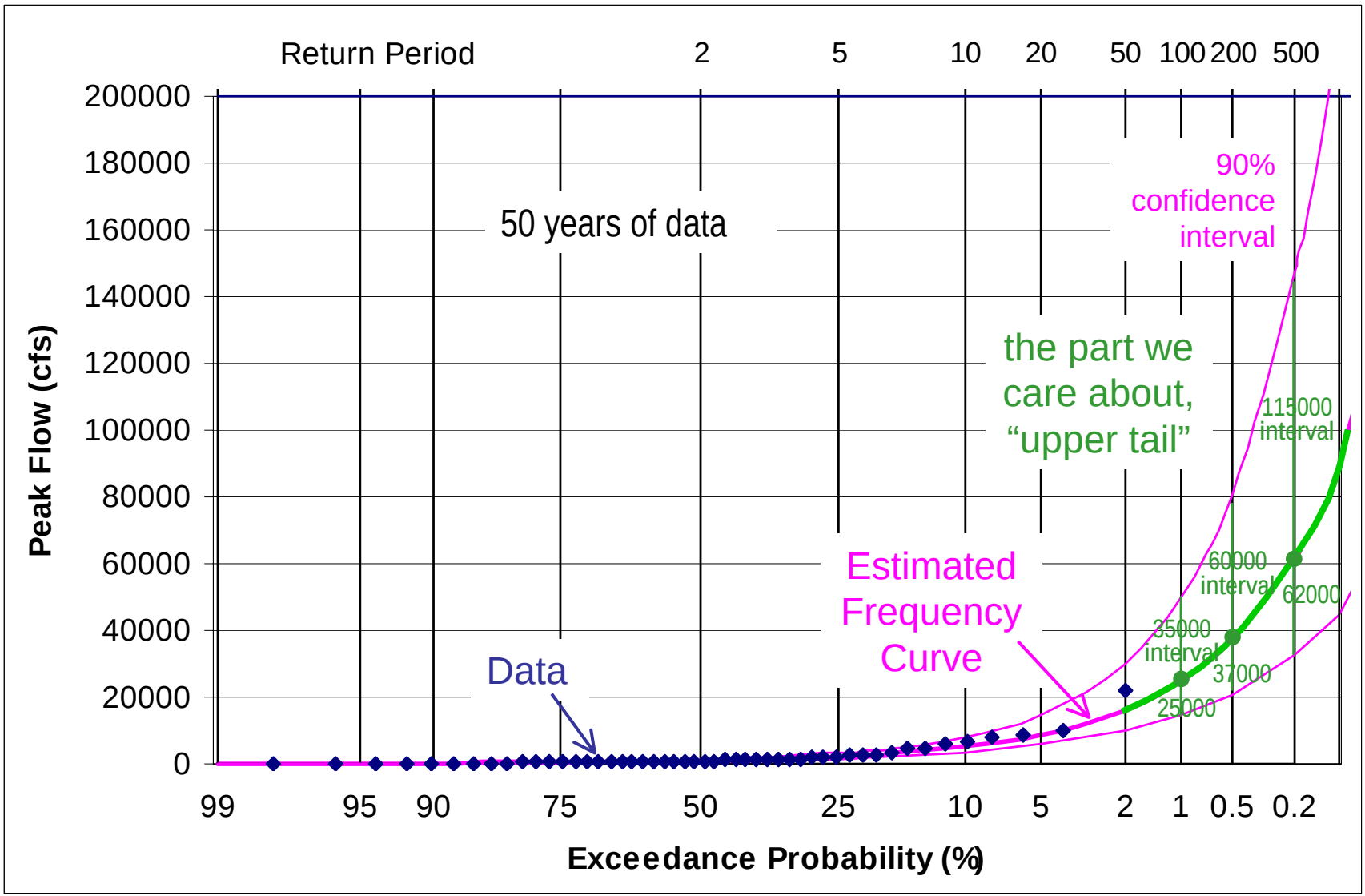
Purpose

- To understand the level of uncertainty in estimation of a frequency curve
- To understand the difficulty in estimating the upper end or upper tail of a frequency curve – “extrapolation”
- To consider the options in extending a frequency curve

Extrapolating?



Extrapolating?



Factors Contributing to Uncertainty

Measurement Error

- How well are large flows measured?

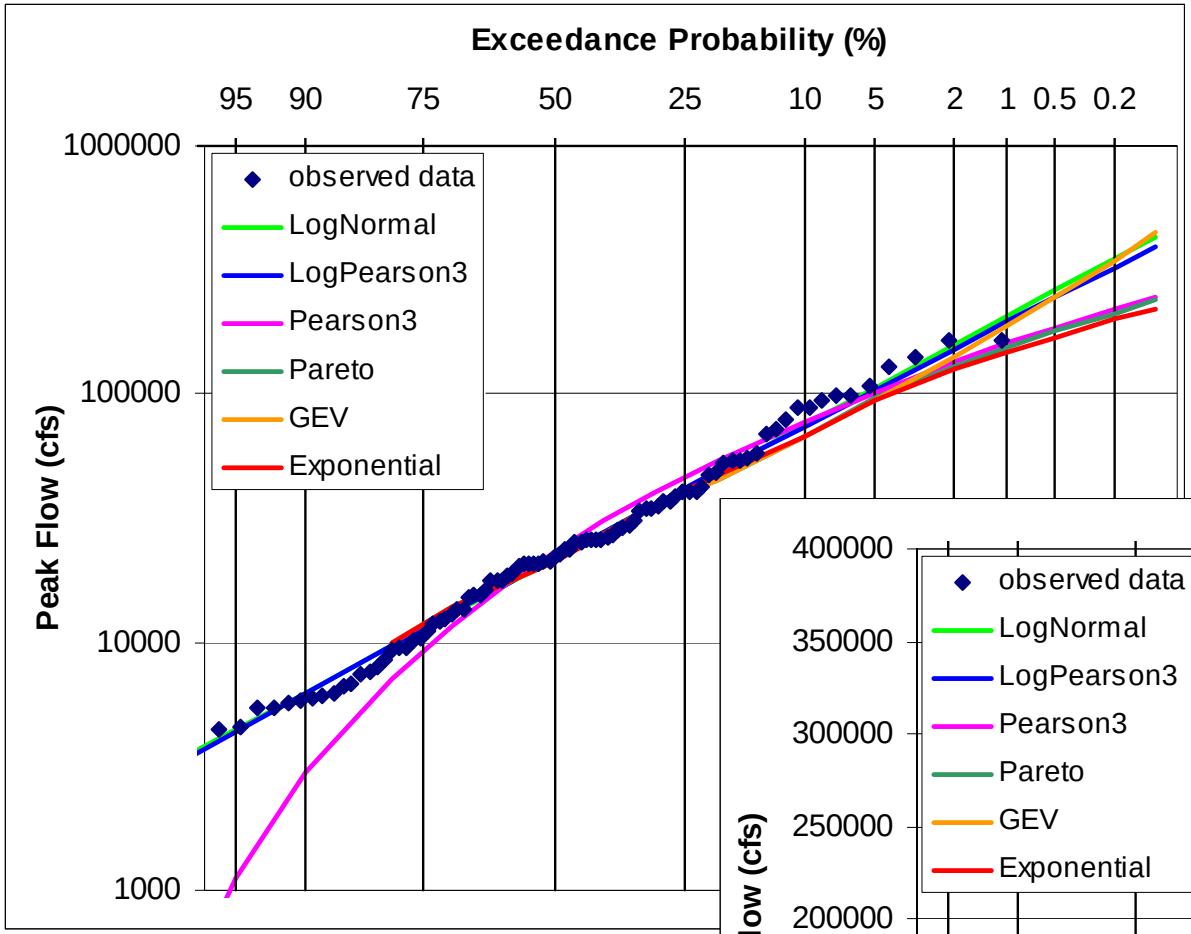
Model Error

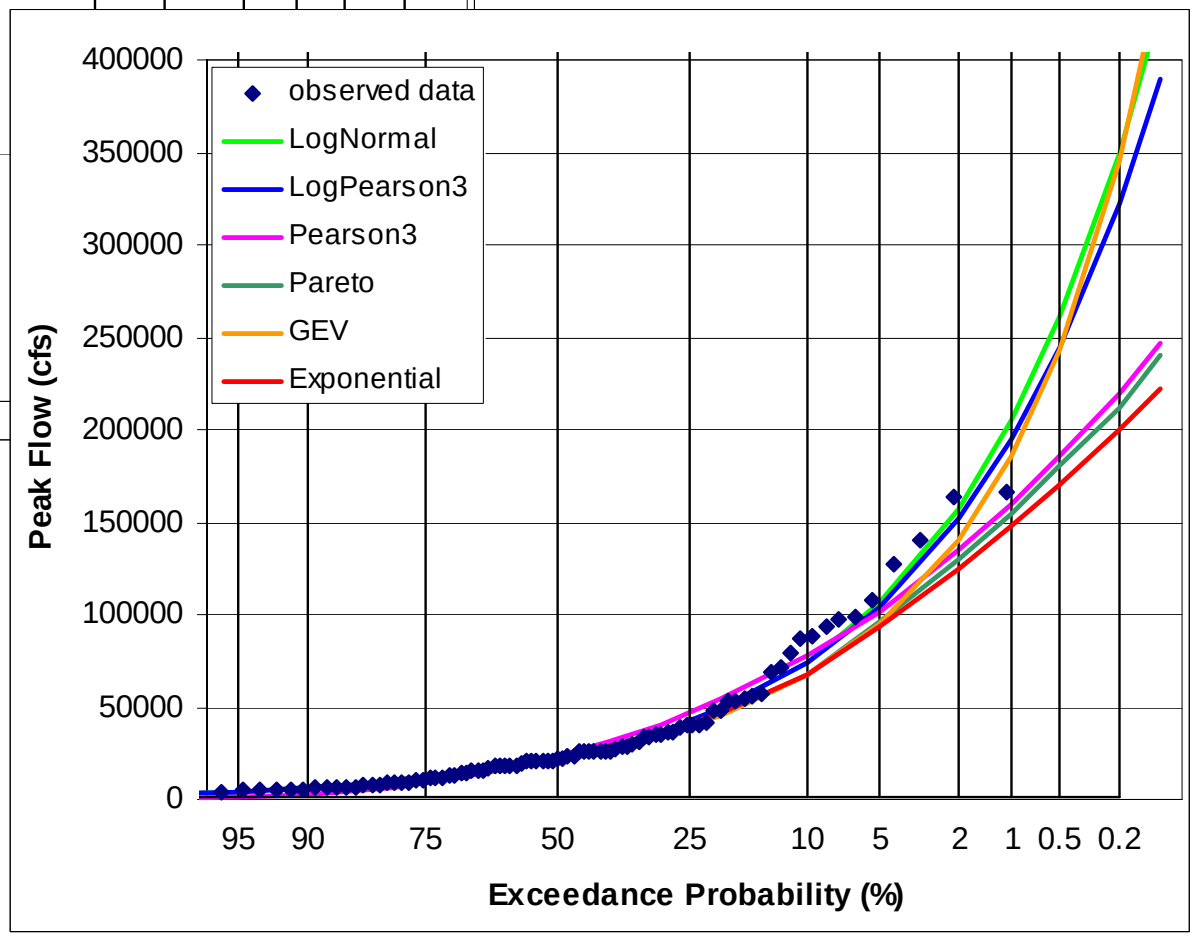
- Does Log-Pearson III distribution describe flow frequency well? GEV? Some other distribution?
- What method is used to estimate parameters?

Sampling error

- Error due to limited sample size
  - sample is not representative *I'll focus on this…*

Several Probability Distributions fit to the same data





Estimating a Frequency Curve

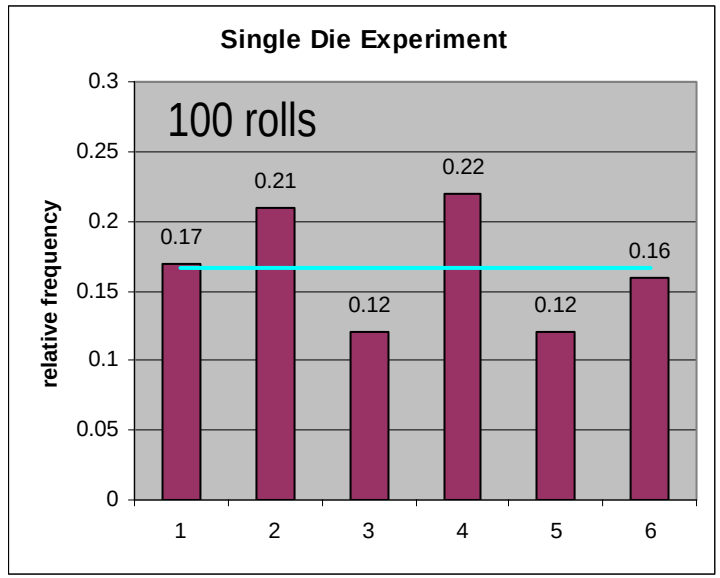
- *Question:* When we have 100 years of record, how *well* can we estimate a frequency curve and the 100-year event?
- *Answer:* Let's do some statistical experimentation to find out…

Statistical Experiments

- How do you test an estimation procedure?
  - *create limited samples of a known situation, so we can compare the sample estimates to the "right answer"*
  - start with a presumed "reality" – a probability distribution
  - draw a random sample from that "reality"
  - estimate "reality" from the sample (pretending we didn't really know), and compare
  - repeat…
- A simple example…

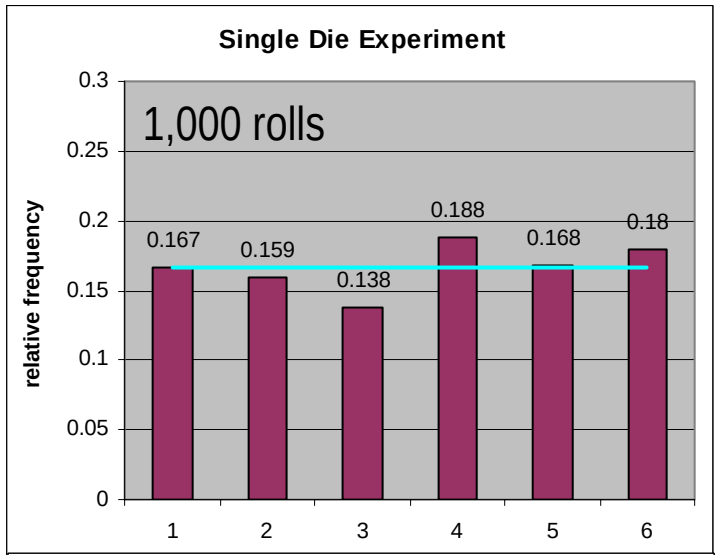
Statistical Experiments

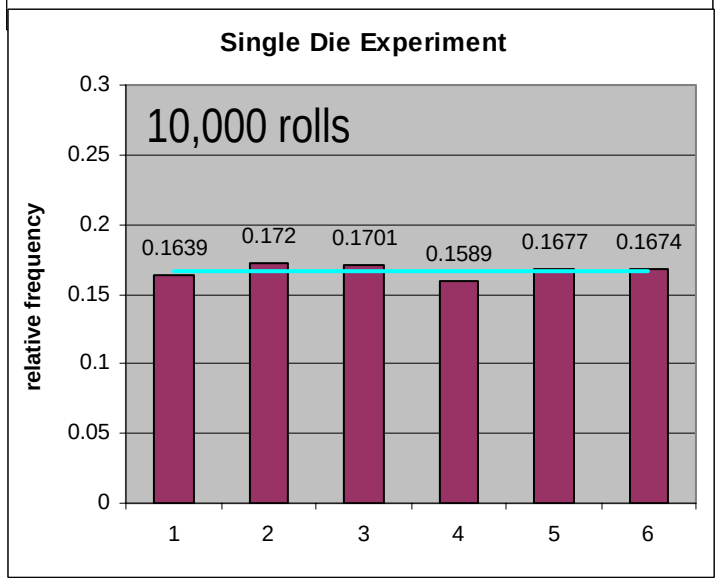
- Testing a 6-sided die for fairness. How many rolls?
- Presume a fair die, so each side has known prob = 1/6 = 0.167
- "roll" die 100 times, estimate probability of each side by relative frequency
  - estimatorA = relative frequency of each side in 100 rolls
  - Is it a good estimator?



Statistical Experiments

- How could we improve the estimator?
- In this case, we could add more data…
  - estimatorB = relative frequency with 1,000 rolls
  - estimatorC = relative frequency with 10,000 rolls
- Usually, can't add more data. But can determine the success of the estimator with a given sample size.



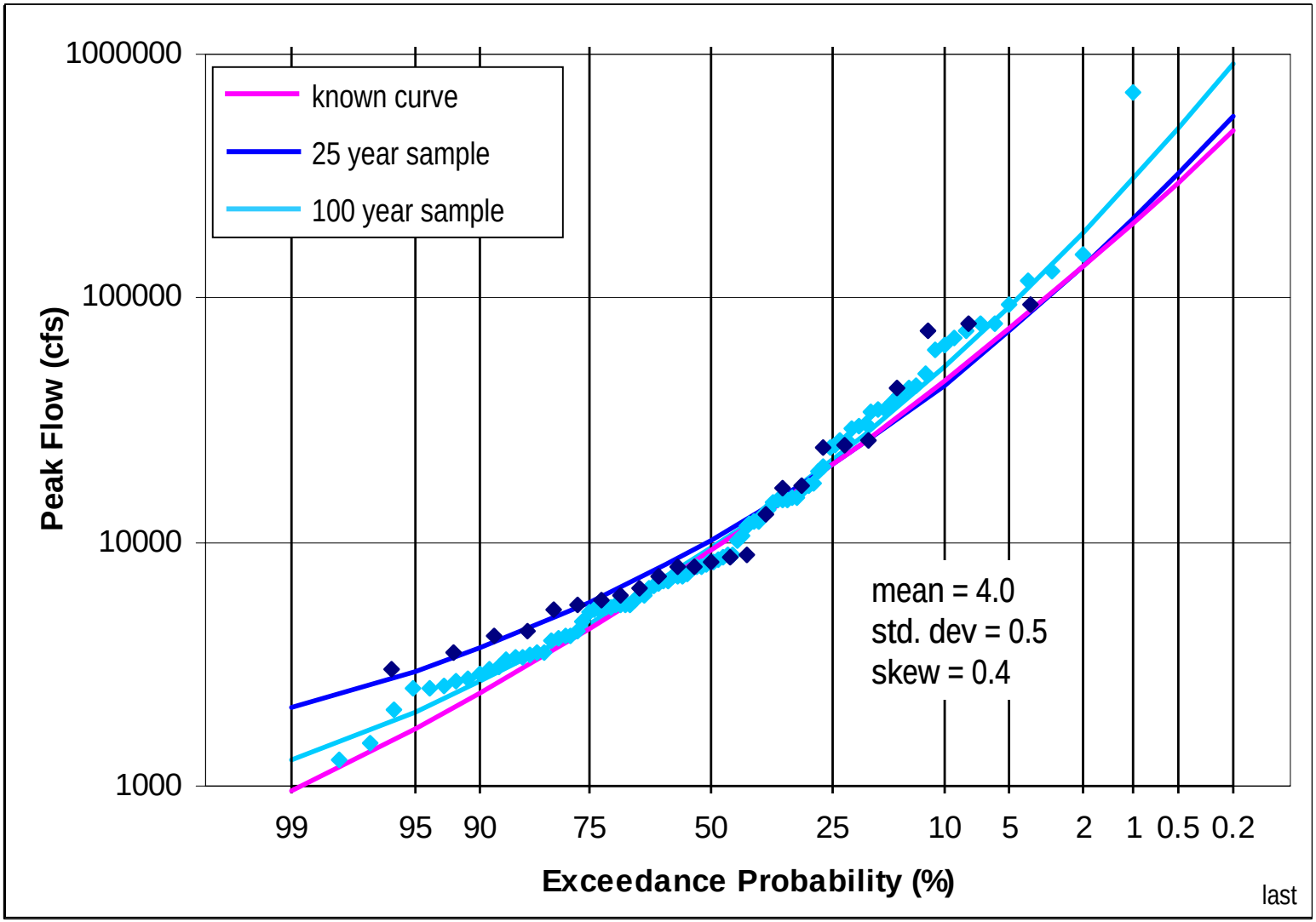


Estimating a Frequency Curve

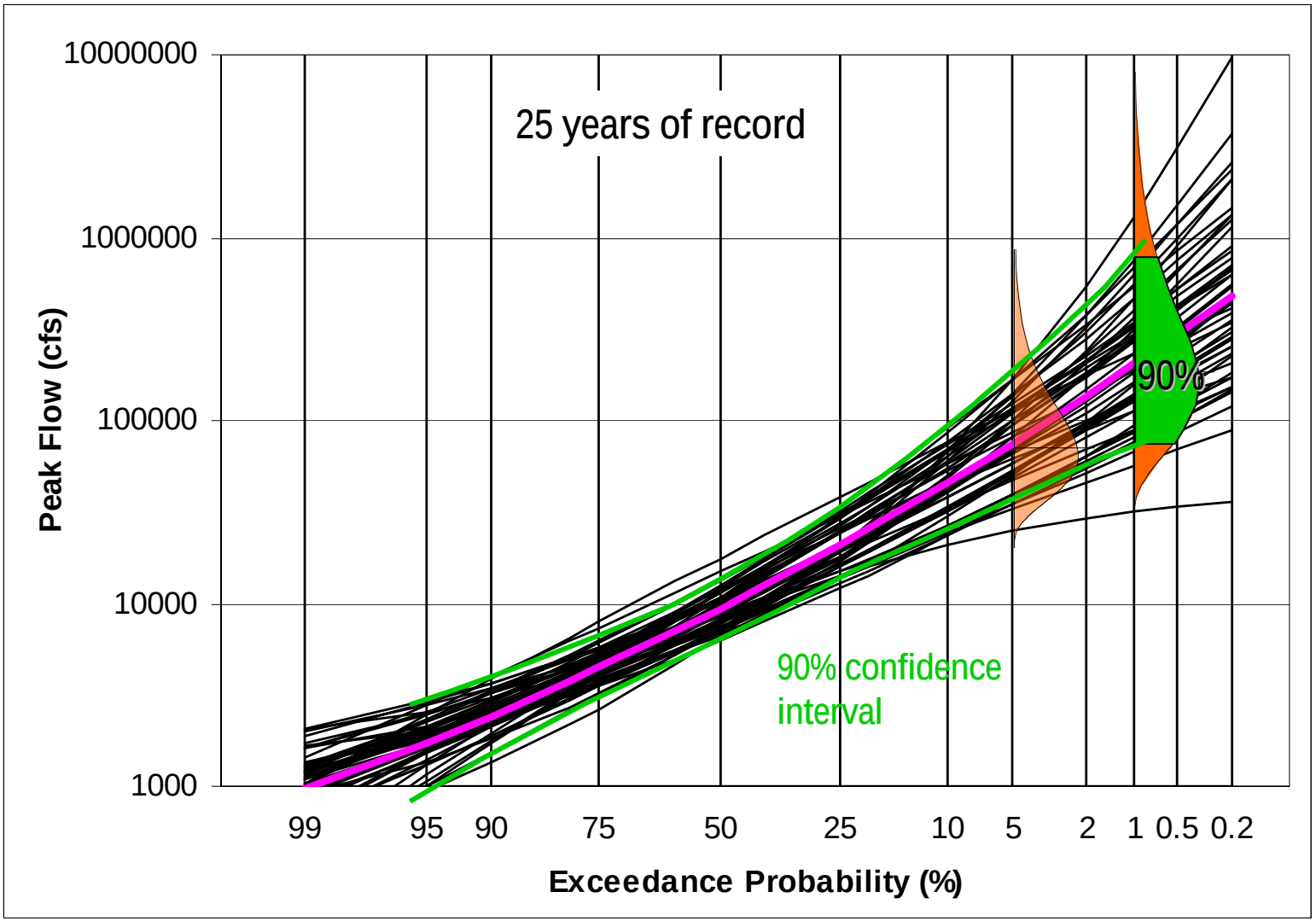
- *Question*: When we have 100 years of record how well can we estimate a frequency relationship?
- We'll do a statistical experiment in which we
  - propose a known probability relationship
  - draw samples from the known relationship to create 100 year records
  - estimate the frequency curve from those samples
  - compare the estimates to the known curve
- *NOTE: This experiment will show us "sampling error" but not model error*

100 year sample of “known” curve

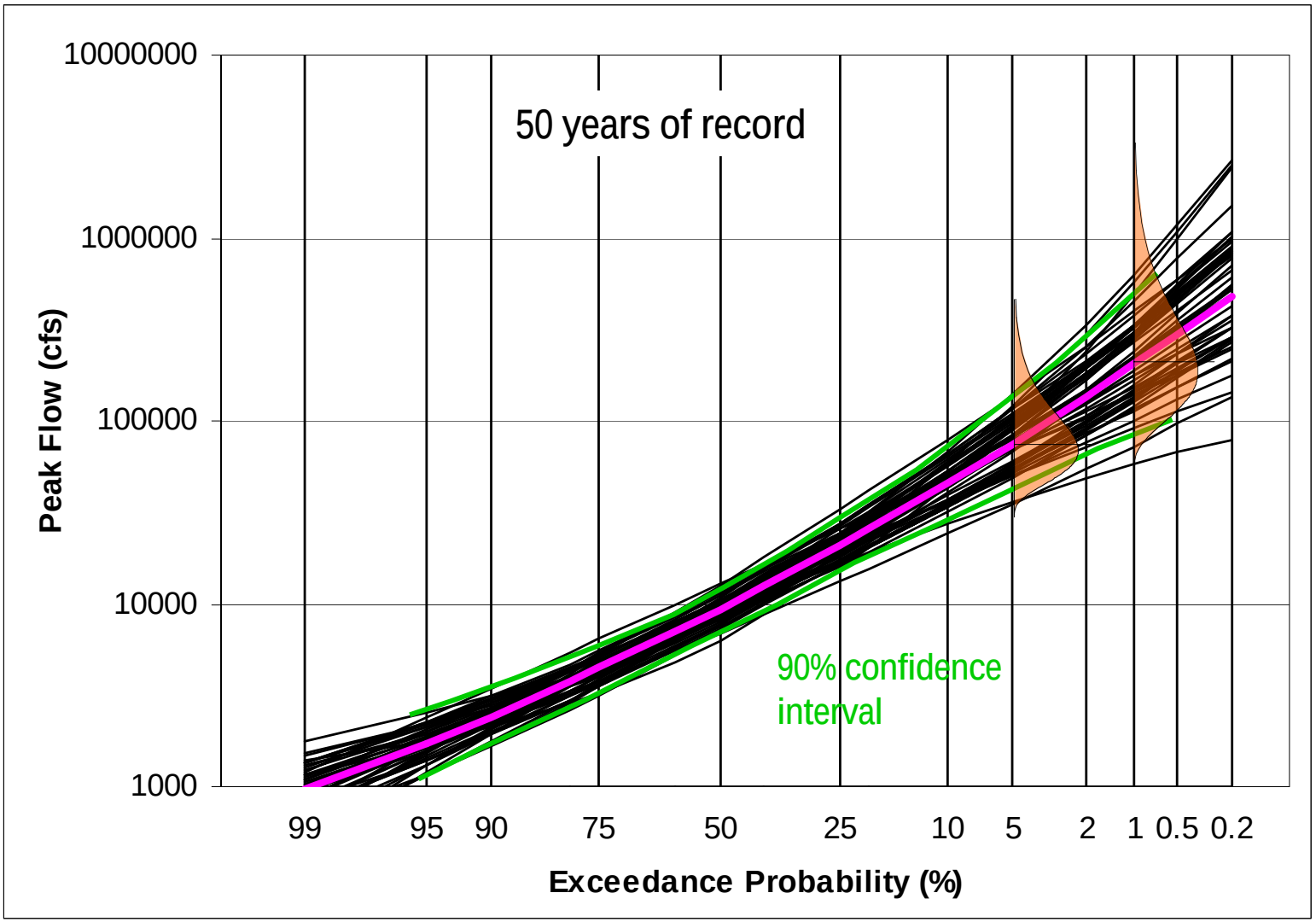
25 year sample of “known” curve



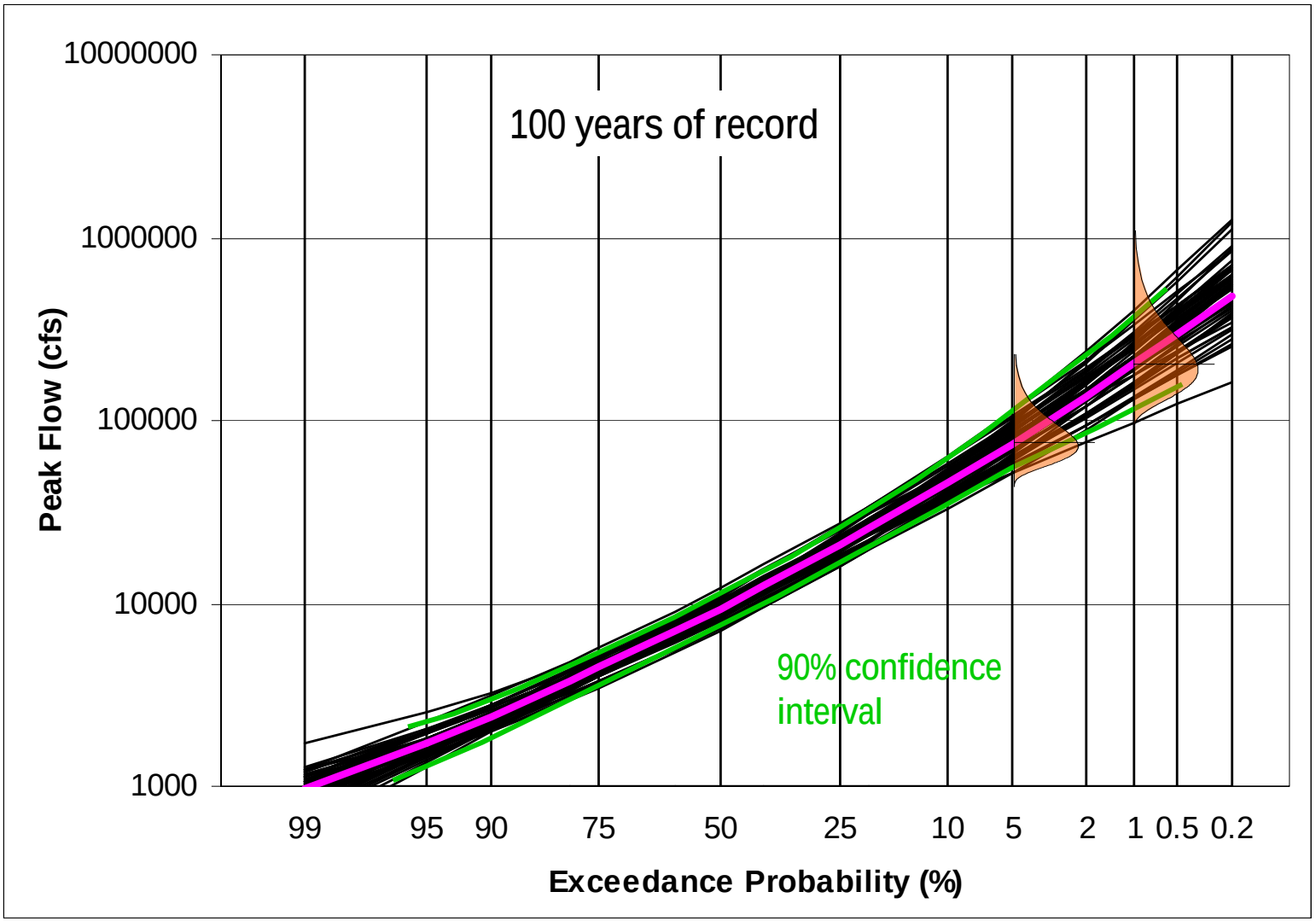
What does that tell us?



What does that tell us?



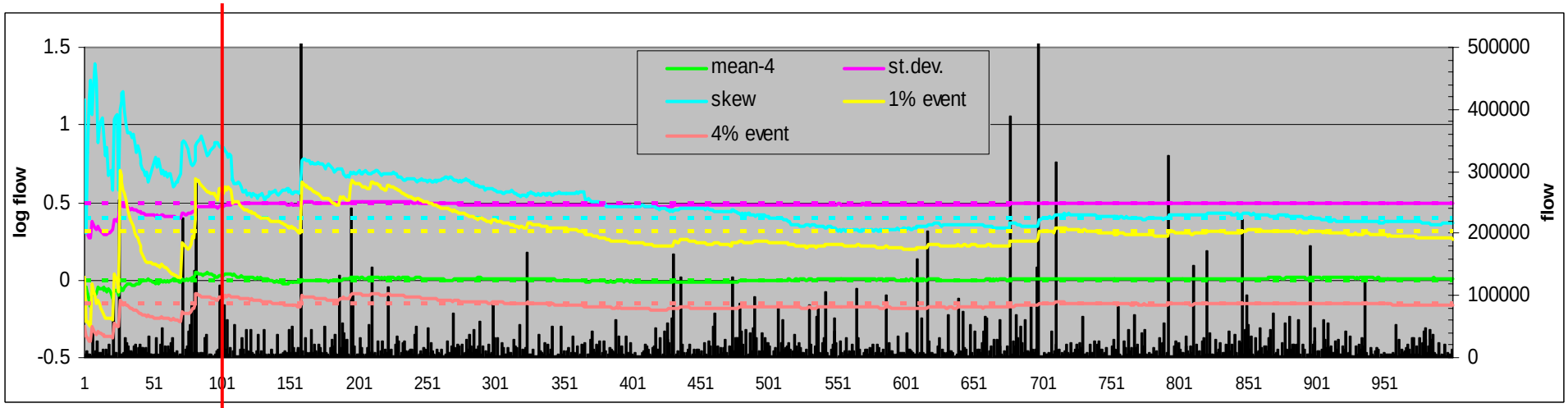
What does that tell us?

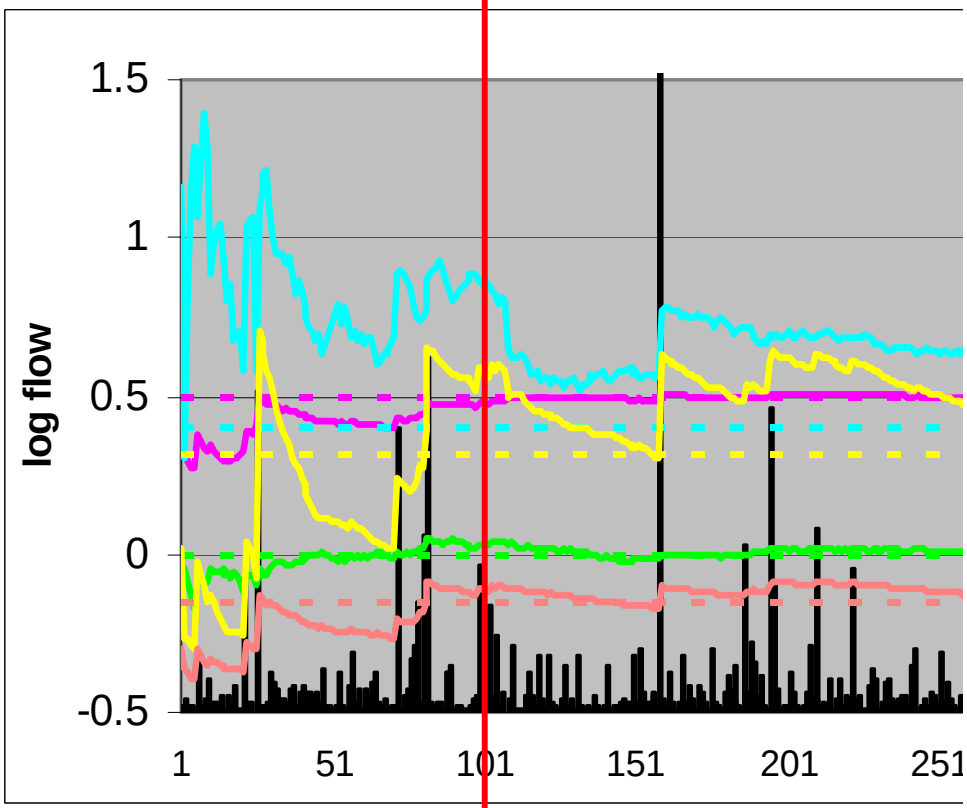


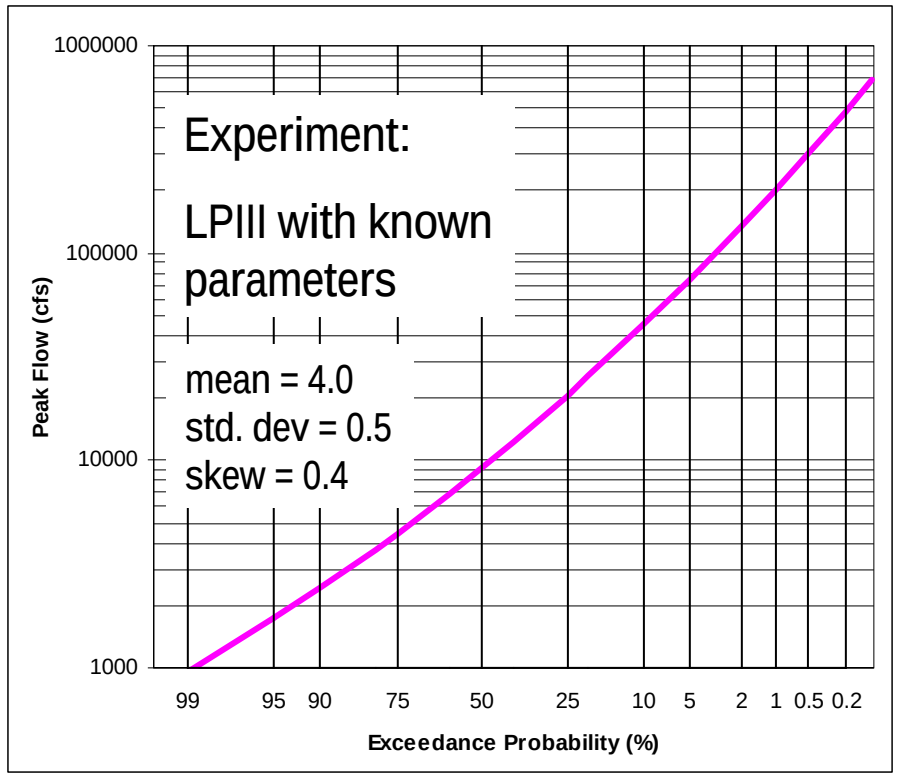
How many years does it take?

- How many years of observed record is enough to estimate the frequency curve to the events of interest?
- A simple answer would be that “we need 100 years of data to know the 100-year (1% chance) event.”
- …it’s both easier and harder than that…

Experiment: 1000 years of data, sampled from LPIII with known parameters

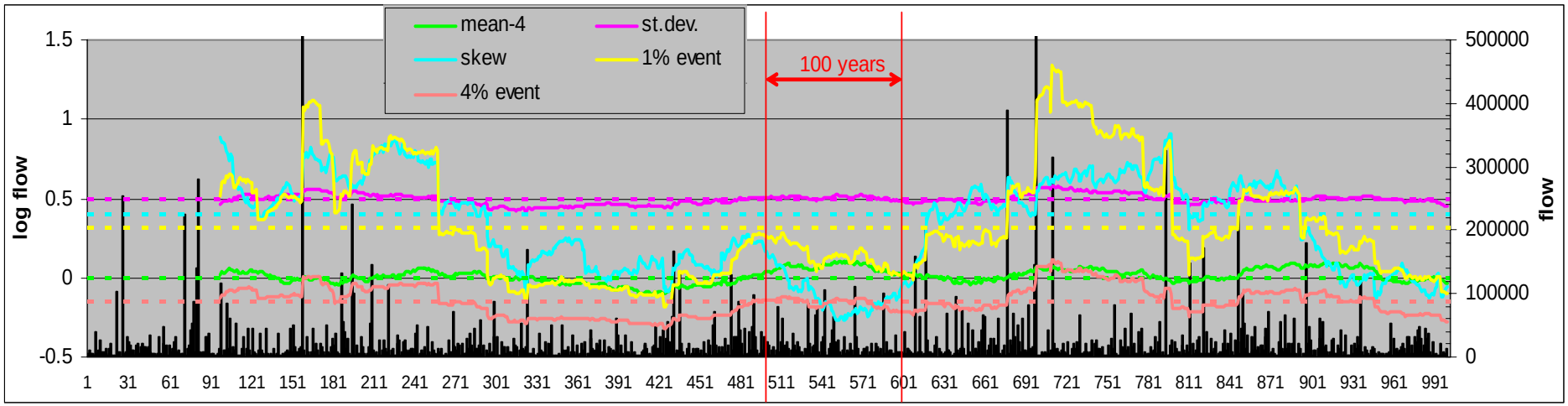


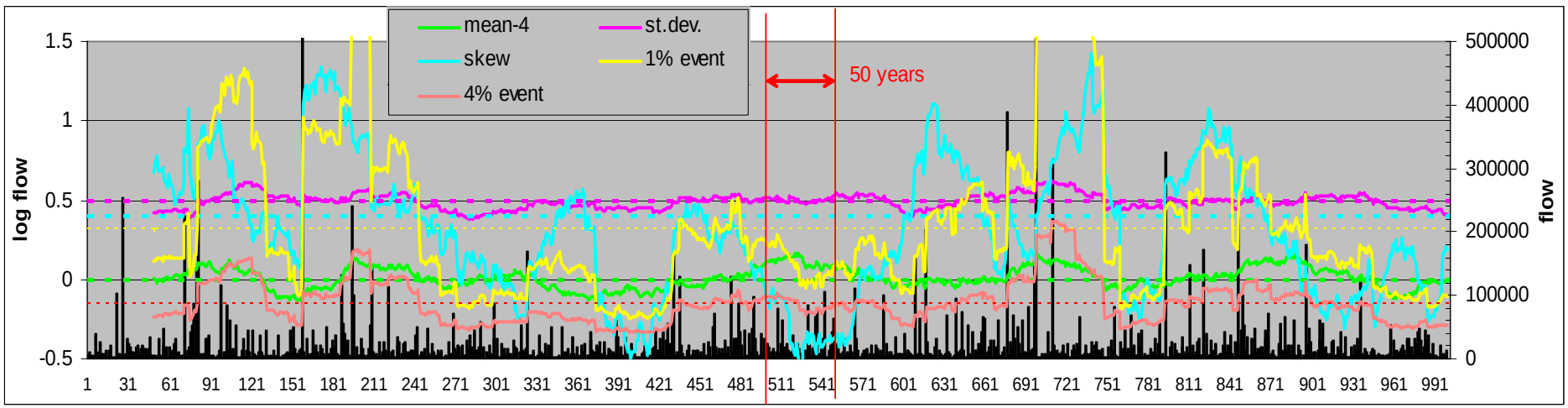




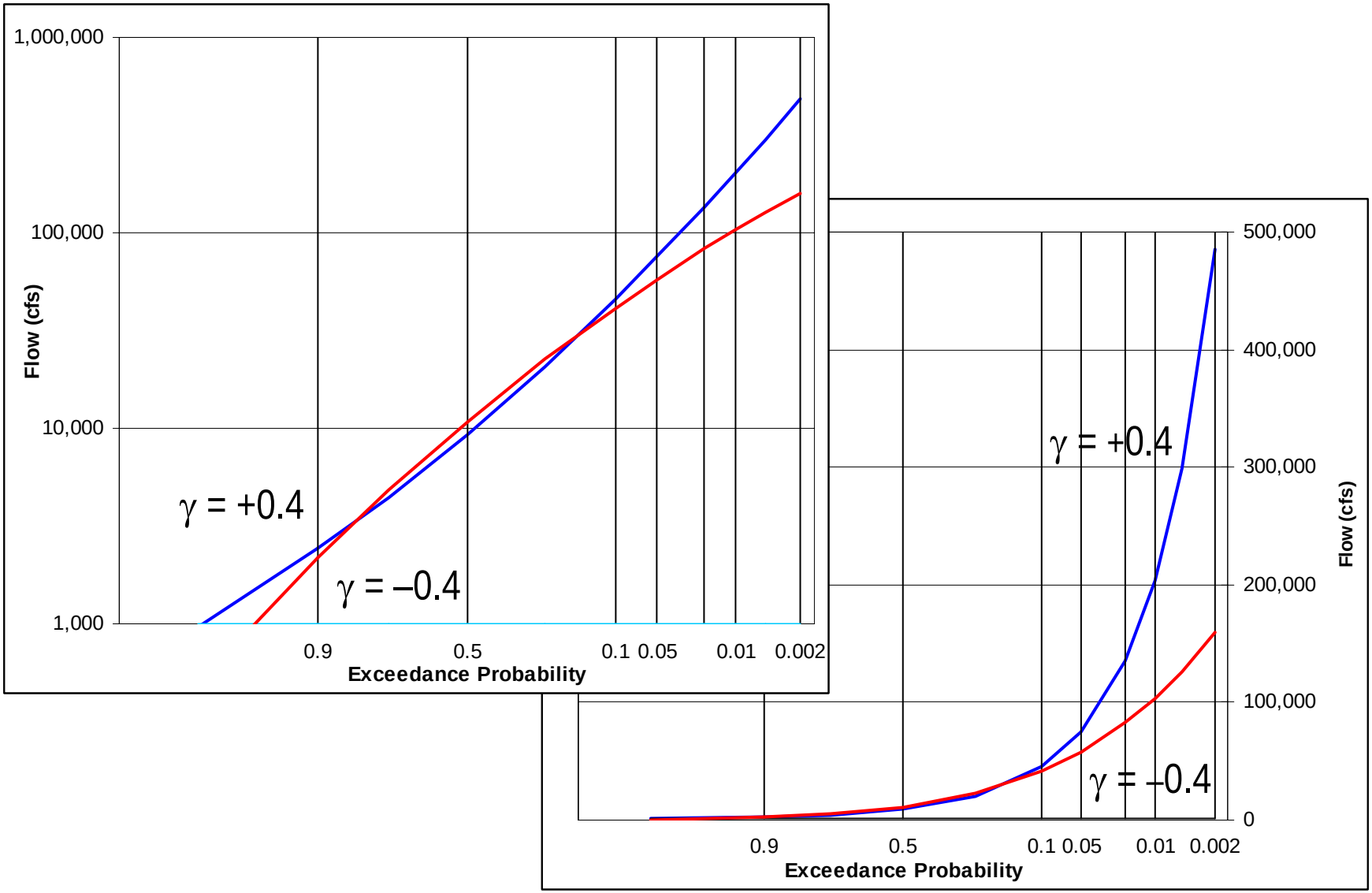
Let's look at different sample sizes…

Same data, sliding 100 year window and 50 year window



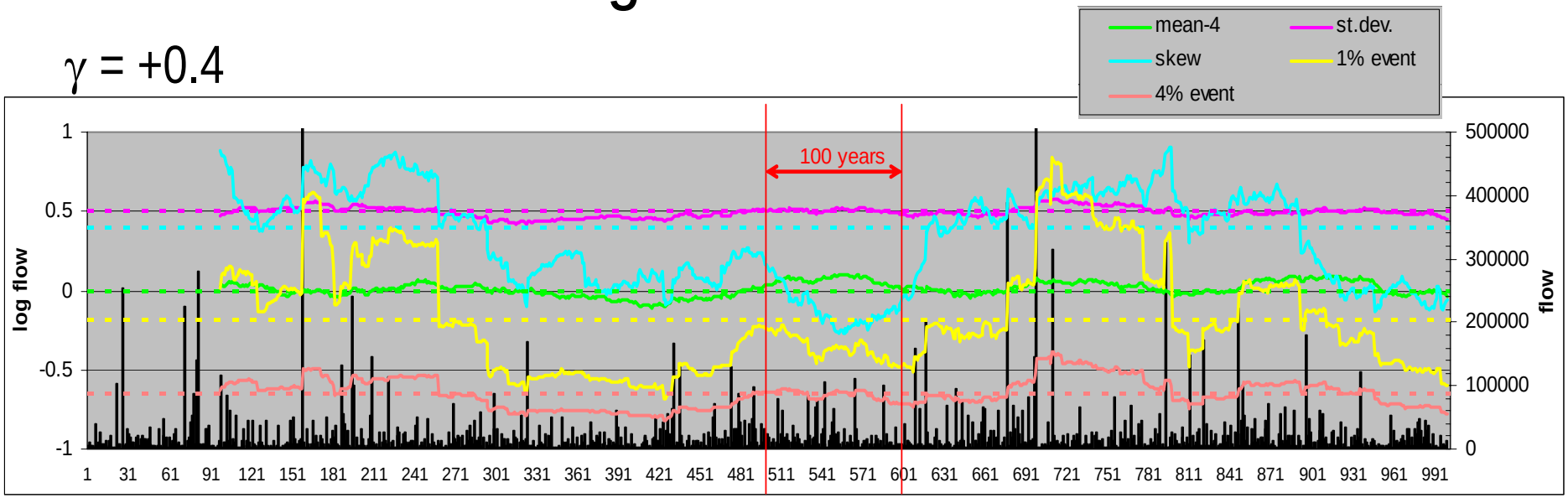


What about a more negative skew?

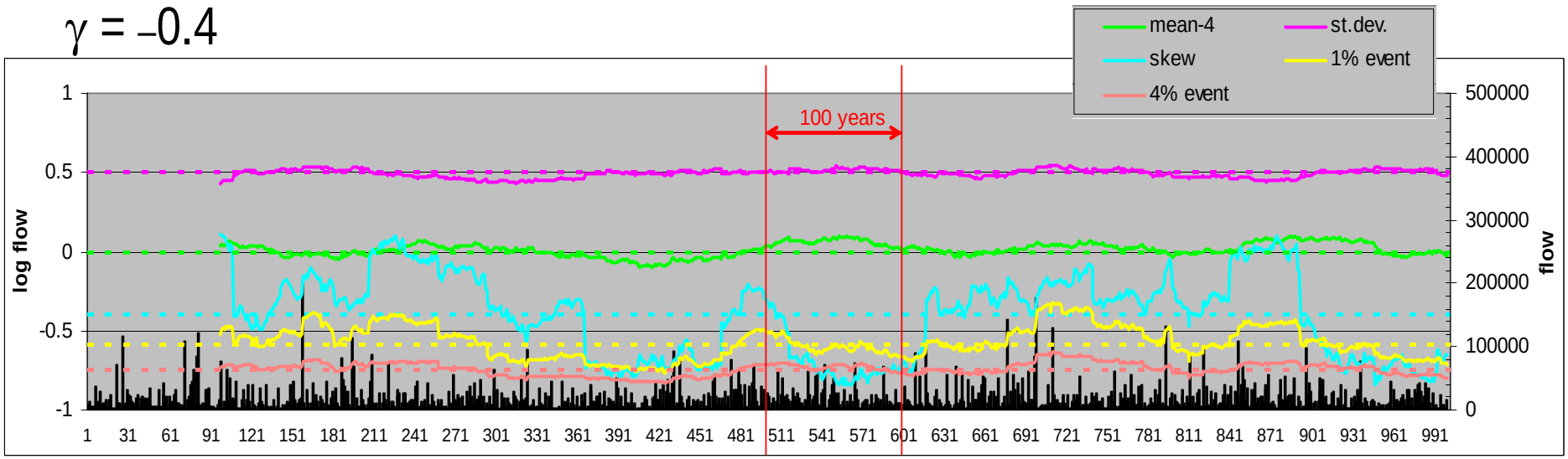


What about a more negative skew?

γ = +0.4



γ = –0.4



When skew is more negative, change has less influence on the 100-yr event

What should we do?

There are several methods for either improving the frequency curve estimate, or extending the curve

improved parameter estimates (data and method)

1. Historical flood or Paleoflood data
2. Regionalization (pooling data)
3. Baysian inference

assumption that precip-frequency analysis is more reliable

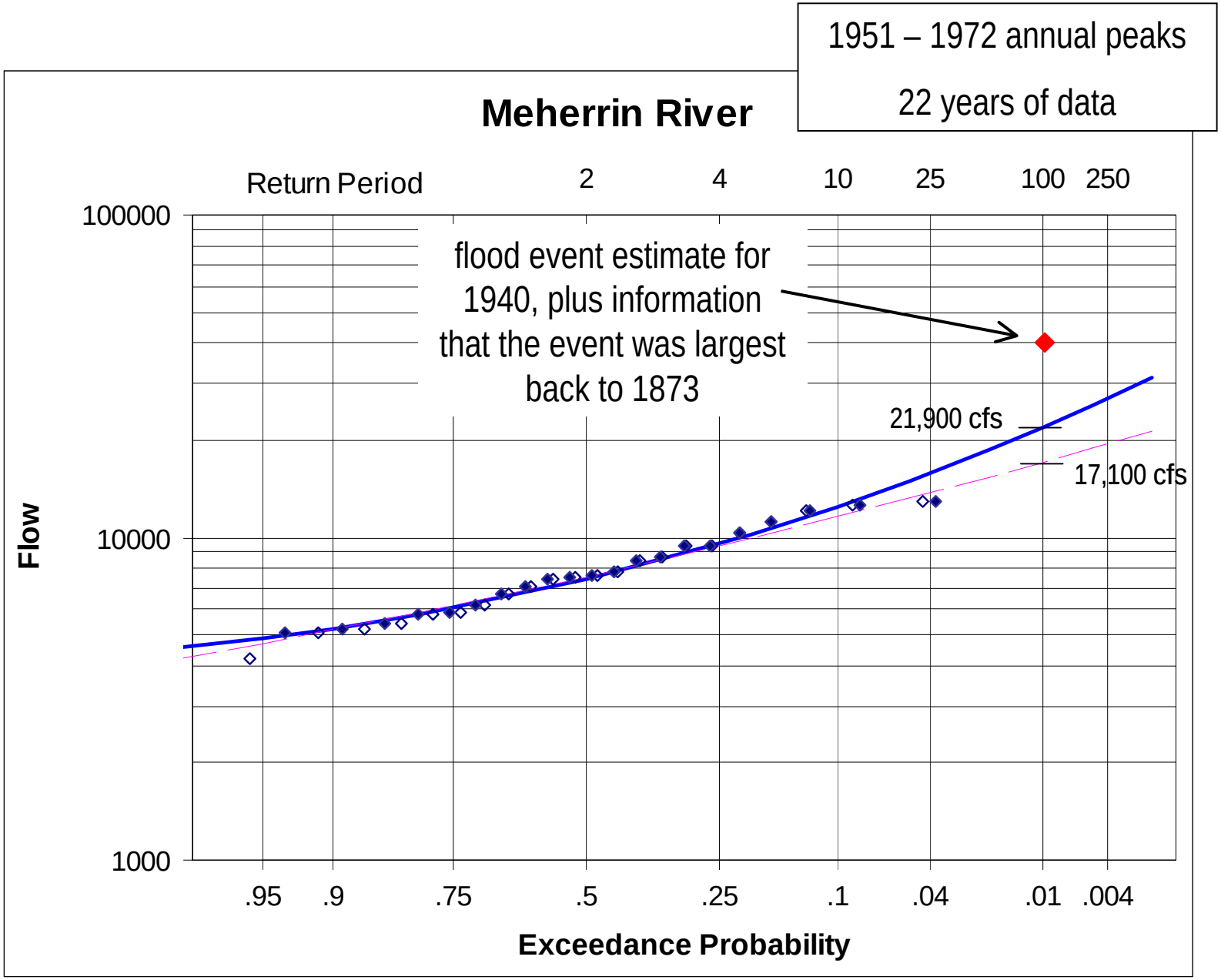
4. Hydrologic modeling of frequency-based flood events
5. Stochastic Event Flood Model (SEFM)
6. GRADEX method

new physical model describing scaling

7. Multi-fractal approach

Historical flood or Paleoflood data

- Often, historical records of large floods are available for a time that precedes the streamflow gage
  - A newspaper story from 1877 describes a flood that reached the steps of the town library
- Physical evidence sometimes exists to demonstrate a large flood occurred in the past
  - slack water deposits far from the channel that can be dated to 300 years ago
- Adding estimates of these large events to the record extends the effective years of record, and improves the estimate of the entire frequency curve
  - Bulletin 17B Method, Expected Moments Analysis


1951 – 1972 annual peaks
22 years of data
Meherrin River
Return Period
2
4
10
25
100
250
100000
10000
1000
Flow
flood event estimate for 1940, plus information that the event was largest back to 1873
21,900 cfs
17,100 cfs
.95
.9
.75
.5
.25
.1
.04
.01
.004
Exceedance Probability

Regionalization

- “Trading space for time”
- When pool data on flood flow or rainfall from many independent sites, the result is similar to experiencing a longer record
  - have eliminated much of the sampling error because have observed many more events.
- Based on assuming consistent meteorology across a region
- For flood flow analysis, regionalize dimensionless parameters such as skew, or develop relationships between basin parameters and flow quantiles

Baysian Analysis

- In a Baysian analysis, an assumption is made of a “prior” distribution, based on expert knowledge of the process and data at other locations.
- Gage data is used to update the prior distribution into a “posterior” distribution.
- This technique has been explored for extreme rainfall probabilities

reference:

Coles, Stuart G, and Jonathan A Tawn, “A Bayesian Analysis of Extreme Rainfall Data,” Applied Statistics, 1996, 45, No. 4, pp 463-478

Hydrologic Modeling

Computing points on the upper end of the flow frequency curve with a rainfall/runoff model of the river basin

- Calibrate a rainfall/runoff model to the frequency curve estimated from gage data
- Develop precipitation events that have a specified frequency (NWS data, regional precip-frequency study)
  - these are based on precipitation frequency, which can use more available data than flow frequency
- Model the runoff from the frequency-based precipitation to determine flow of the same frequency

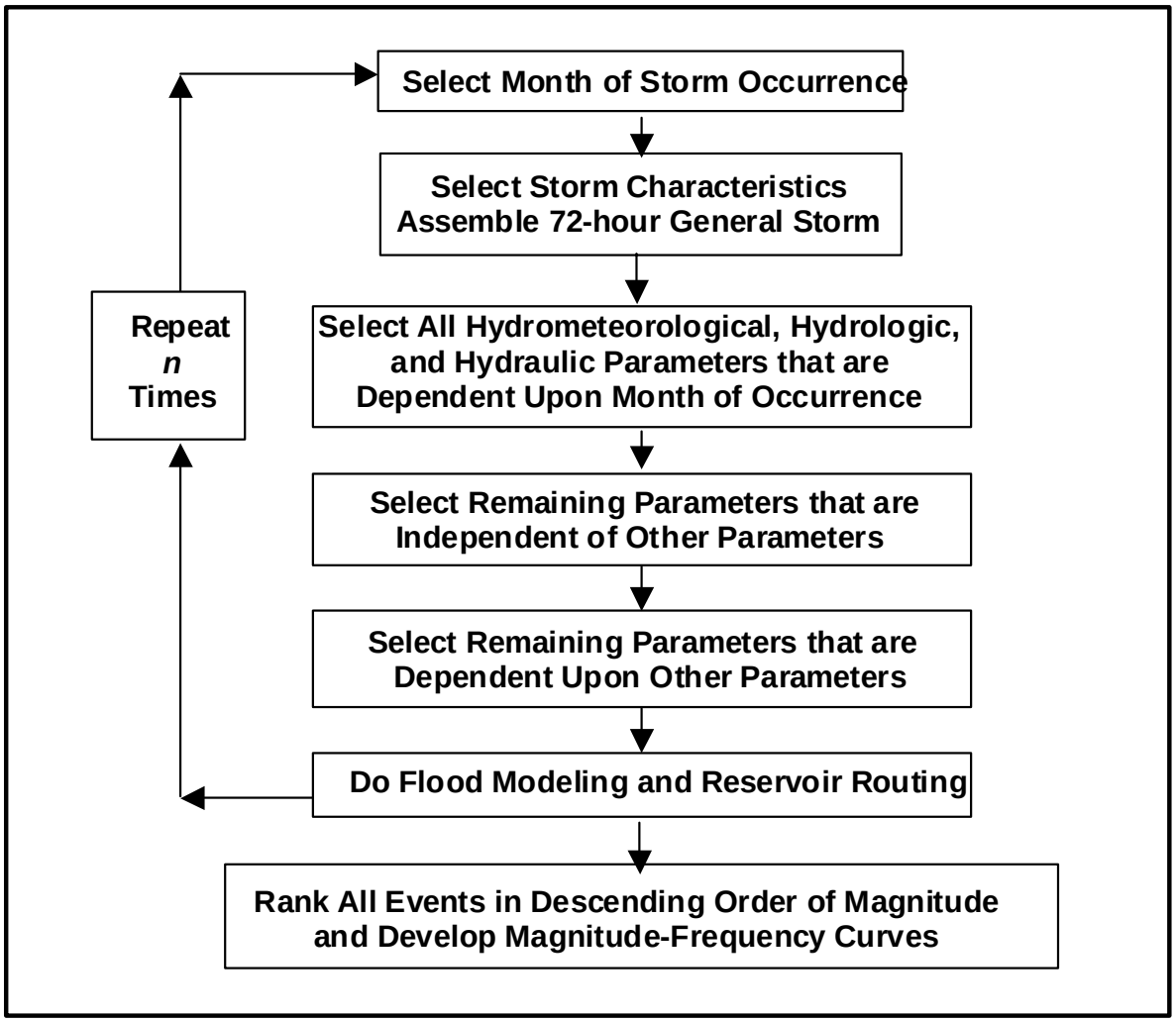
Stochastic Flood Event Model

- Depends on the precipitation-frequency relationship and a rainfall/runoff model of the basin
- Randomly sample the precipitation-frequency curve thousands of times, and model the runoff from each to generate thousands of flood flows
  - samples the uncertainty in model parameters as well as precipitation and flow
- Estimate a frequency curve with the thousands of flows, rather than just 50 or 100 actual events.

reference:
MGS Engineering Consultants, Inc

Stochastic Simulation

Simulation Procedure

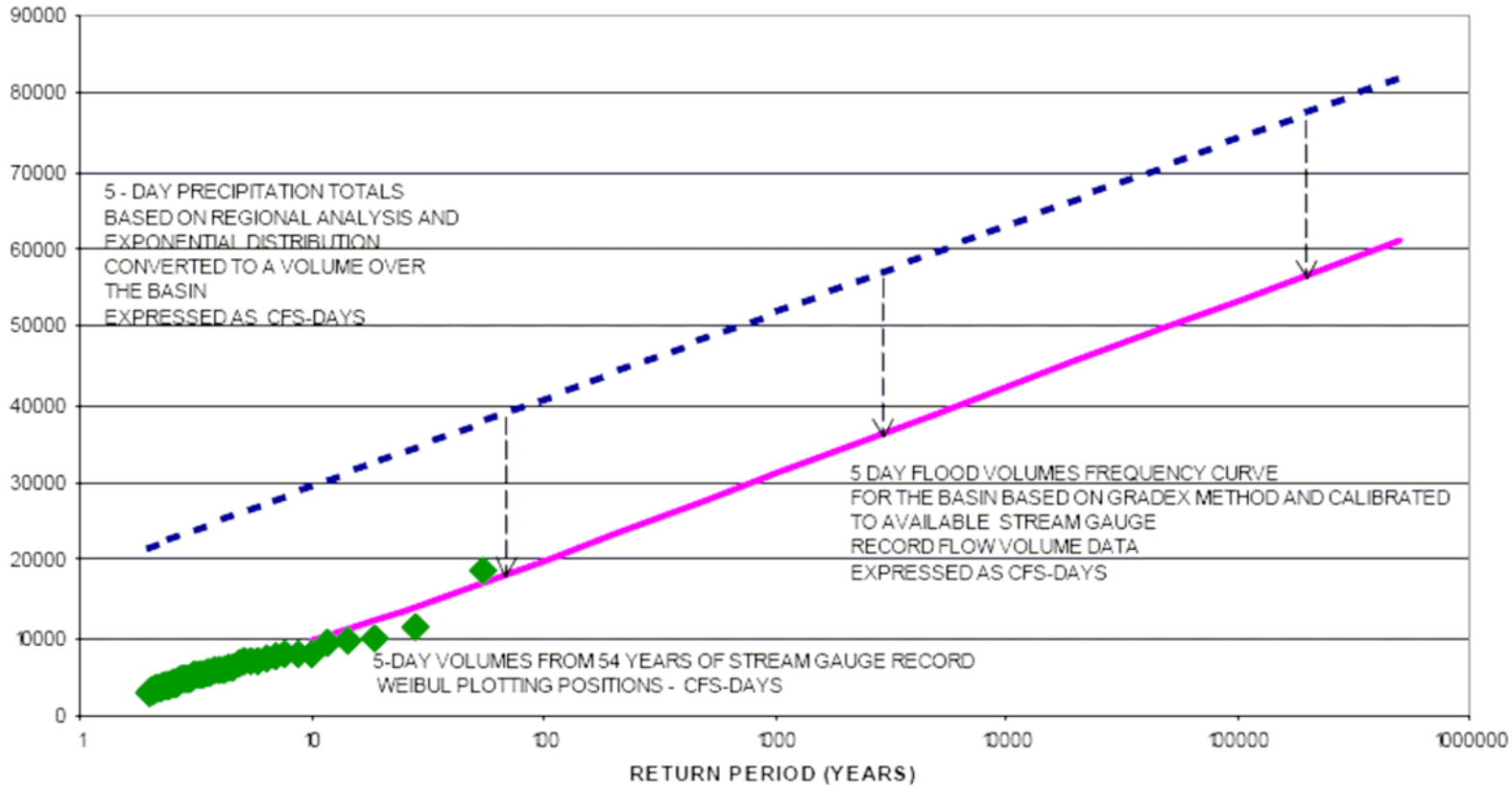


100,000 simulations per production run

MGS Engineering Consultants, Inc

GRADEX Method

Assumption is that slope of upper end of the flood frequency curve follows the slope of the precipitation frequency curve



Multi-fractal Flood Frequency Analysis

- Based on observing that physical processes exhibit scaling behavior in space and time
  - daily flow relationships are similar to monthly, which are similar to annual and to extreme
  - the intention is that frequencies of extreme events can be estimated with a short series of daily data

Reference:

CEA Technologies Project #T052700-0212, PI: Daniel Schertzer