Lecture 4.1

Graphical Frequency Analysis & Plotting Positions

Beth Faber, PhD, PE

May 2022

Flood Frequency Analysis

Goals

- Know how a graphical frequency curve differs from an analytical frequency curve
- Understand plotting positions, what they estimate, and how they differ from each other
  - and that plotting positions are the basis for a graphical frequency curve
- Become familiar with the methods of developing graphical frequency curves for regulated flow and stage

Graphical Frequency Analysis

1. Used for estimating an **empirical** frequency curve with no assumed distribution

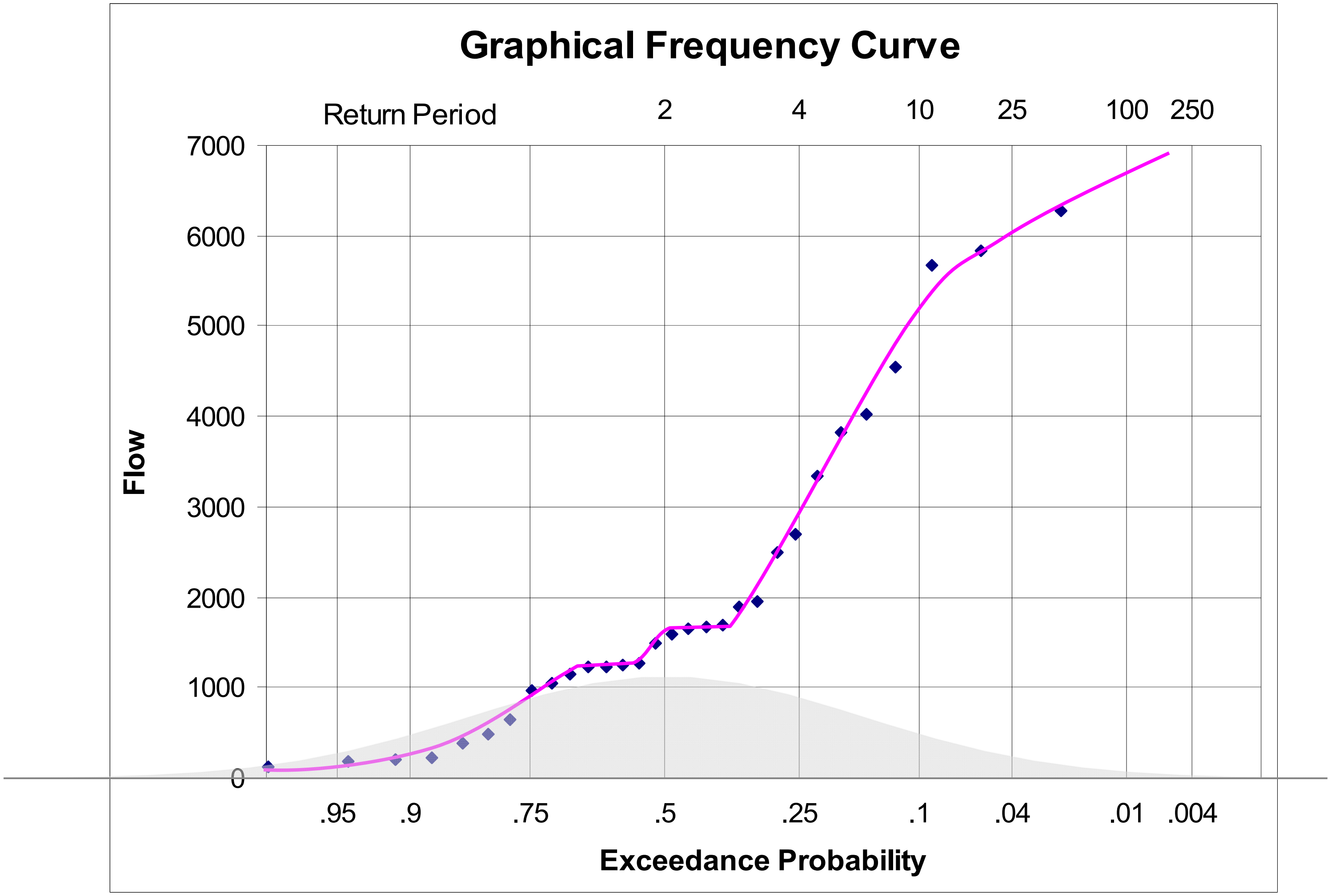
2. Used for regulated flows, channel or pool stages, partial duration series

Outline

- Graphical Frequency Analysis
- Plotting Positions
  - Statistical estimation theory
  - Sampling Distributions to define uncertainty
- Non-analytic (Graphical) Frequency Curves
  - Regulated Flow Frequency Curves
  - Stage Frequency Curves
- Partial Duration Series *(peaks over threshold)*
  - as graphical, rather than analytical

What Is Graphical Frequency Analysis?

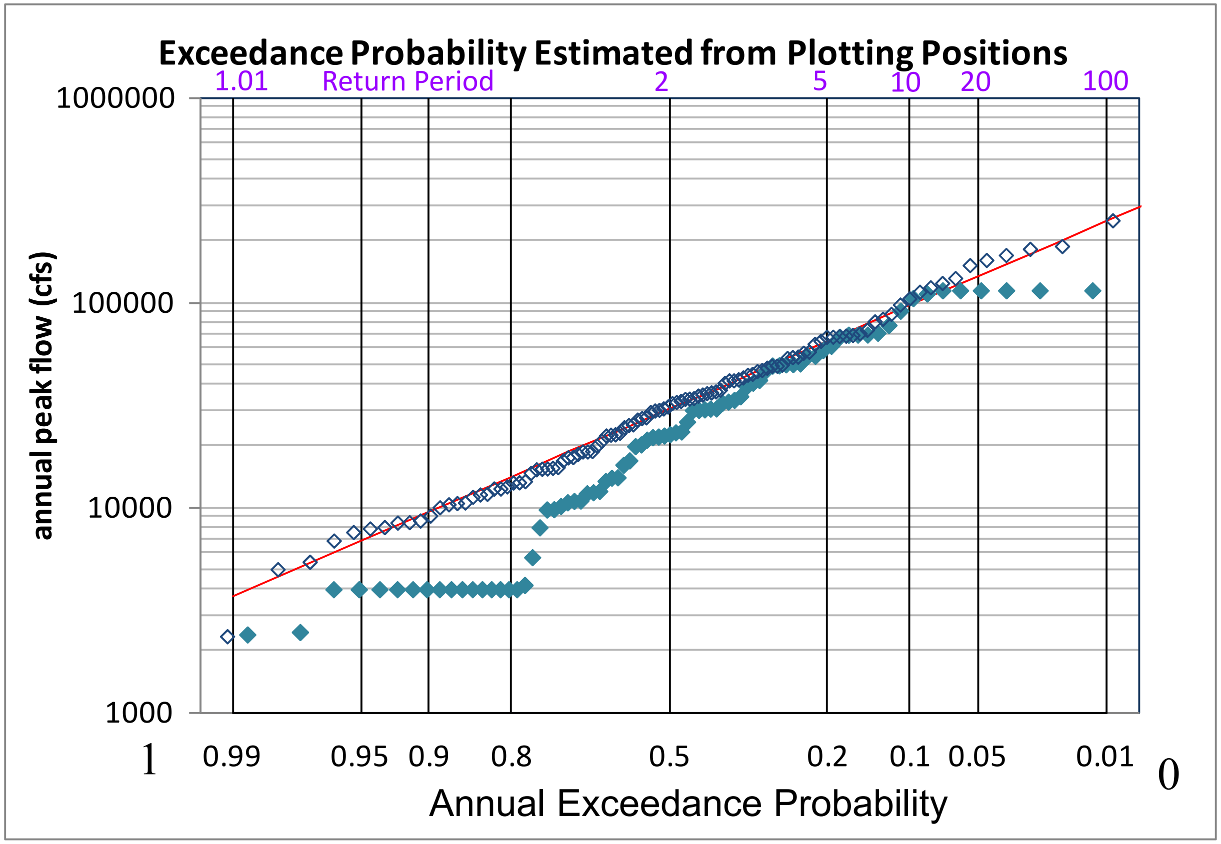
- Graphical analysis involves fitting a curve "by hand" to observed data plotted on probability paper *(i.e., on a Normal probability axis)*
- The *horizontal* location at which the data is plotted is based on a selected plotting position — an empirical estimate of exceedance probability


Graphical Frequency Curve
Return Period
2
4
10
25
100
250
7000
6000
5000
4000
3000
2000
1000
0
Flow
.95
.9
.75
.5
.25
.1
.04
.01
.004
Exceedance Probability

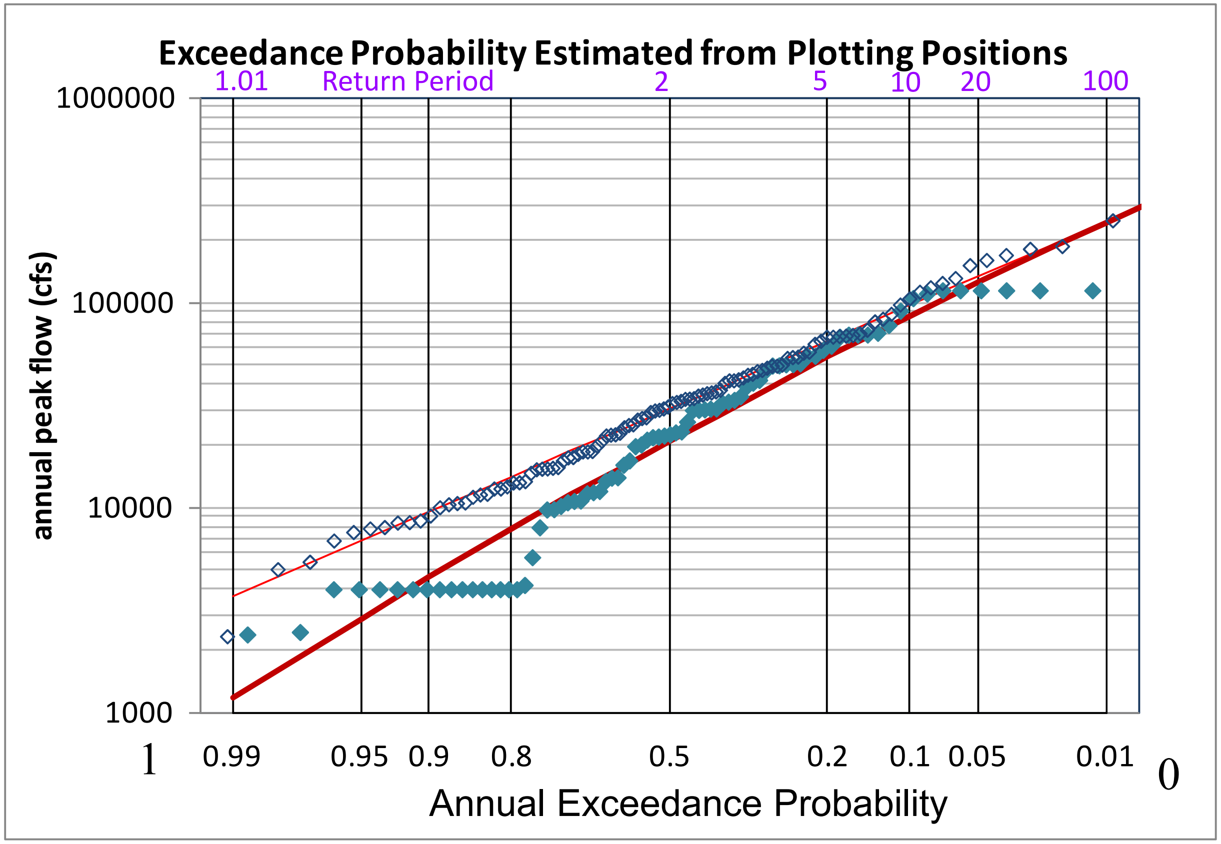
Why Use Graphical Analysis?

1. Plotting the data is a useful visual check on analytical frequency curves (e.g., frequency curves based on LPIII)
2. Sometimes, plotted frequencies of stage and regulated flow data cannot be adequately described by an analytical distribution such as Normal, or LPIII
   - Under these circumstances, we use a graphical frequency curves – also called non-analytic

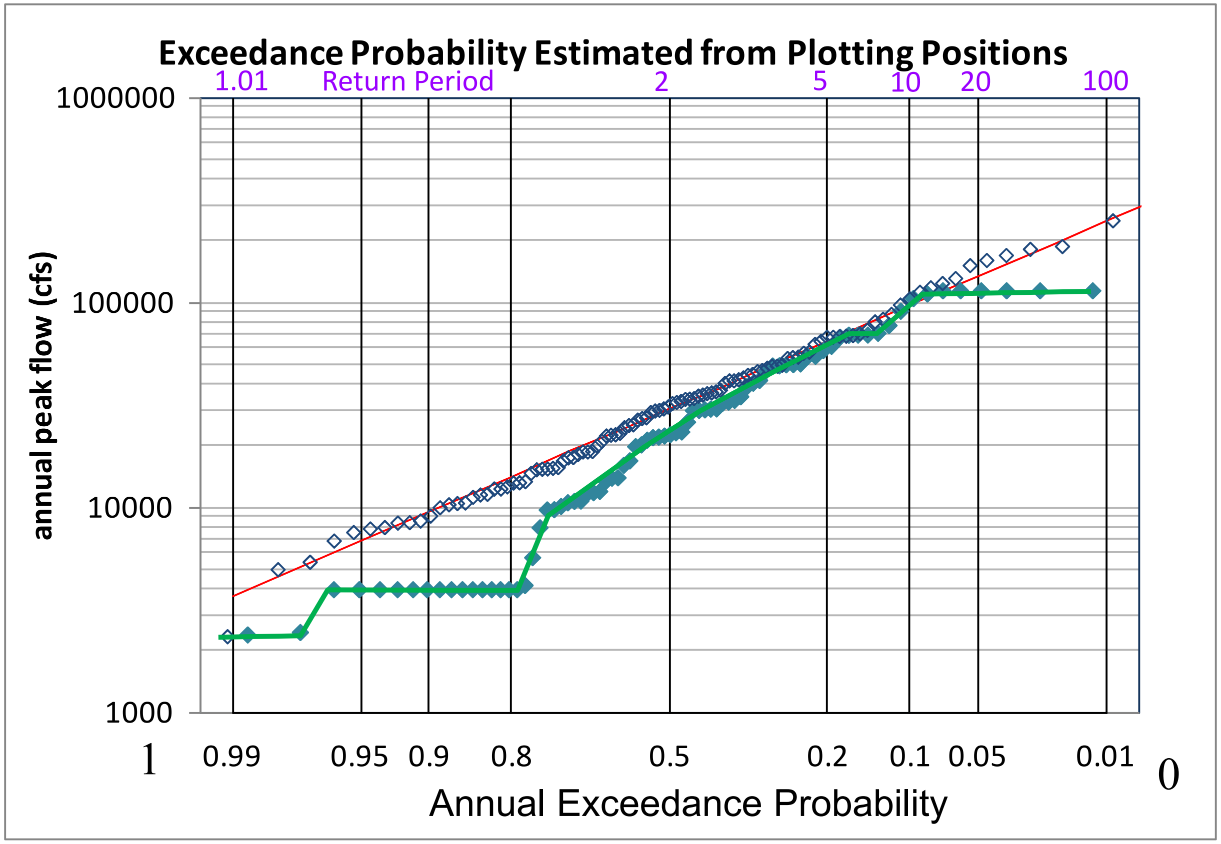
Plotted Annual Maximum Regulated Flows



Does an Analytical Curve Make Sense?



Graphical Curve is Better for this



Outline

- Graphical Frequency Analysis
- Plotting Positions
  - Statistical estimation theory
  - Sampling Distributions to define uncertainty
- Non-analytic (Graphical) Frequency Curves
  - Regulated Flow Frequency Curves
  - Stage Frequency Curves
- Partial Duration Series *(peaks over threshold)*
  - as graphical, rather than analytical

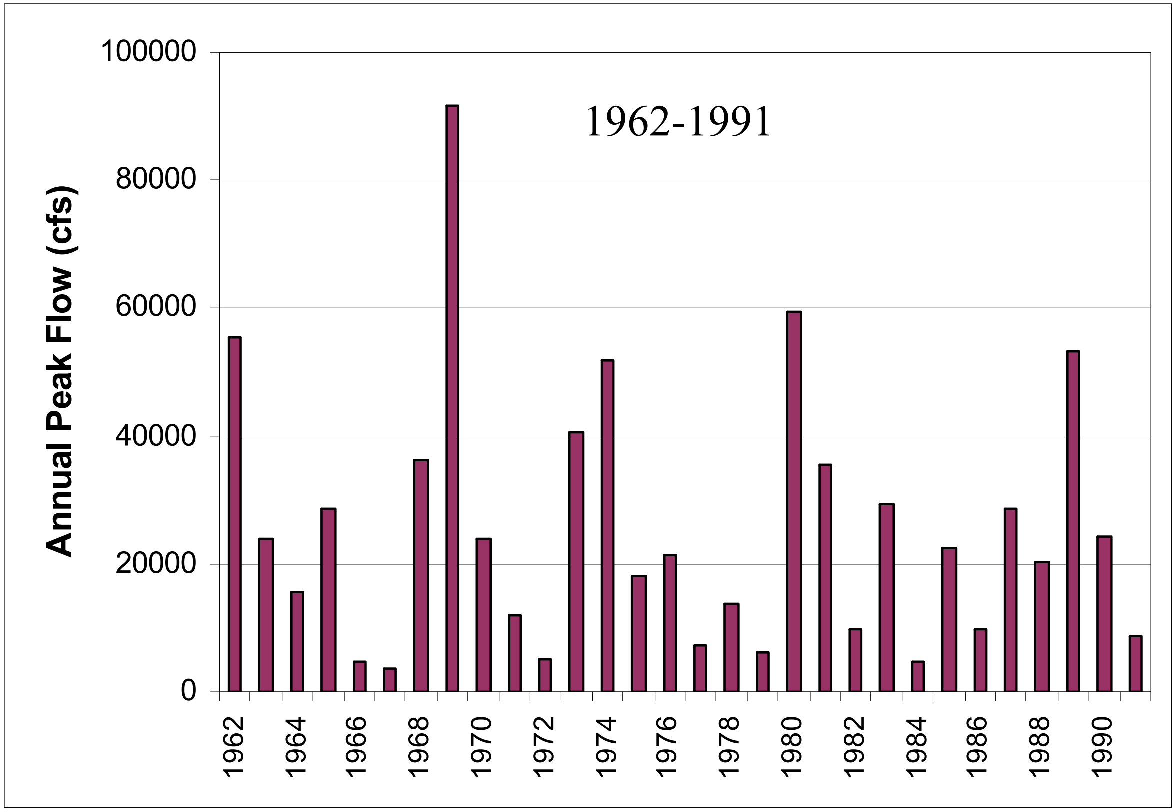
Plotting Position – basic idea

- A plotting position is an estimate of the exceedance probability for some observation of a random variable, such as flow or stage
- For a period of record, plot each observation on a probability axis by estimating its exceedance probability by its relative frequency

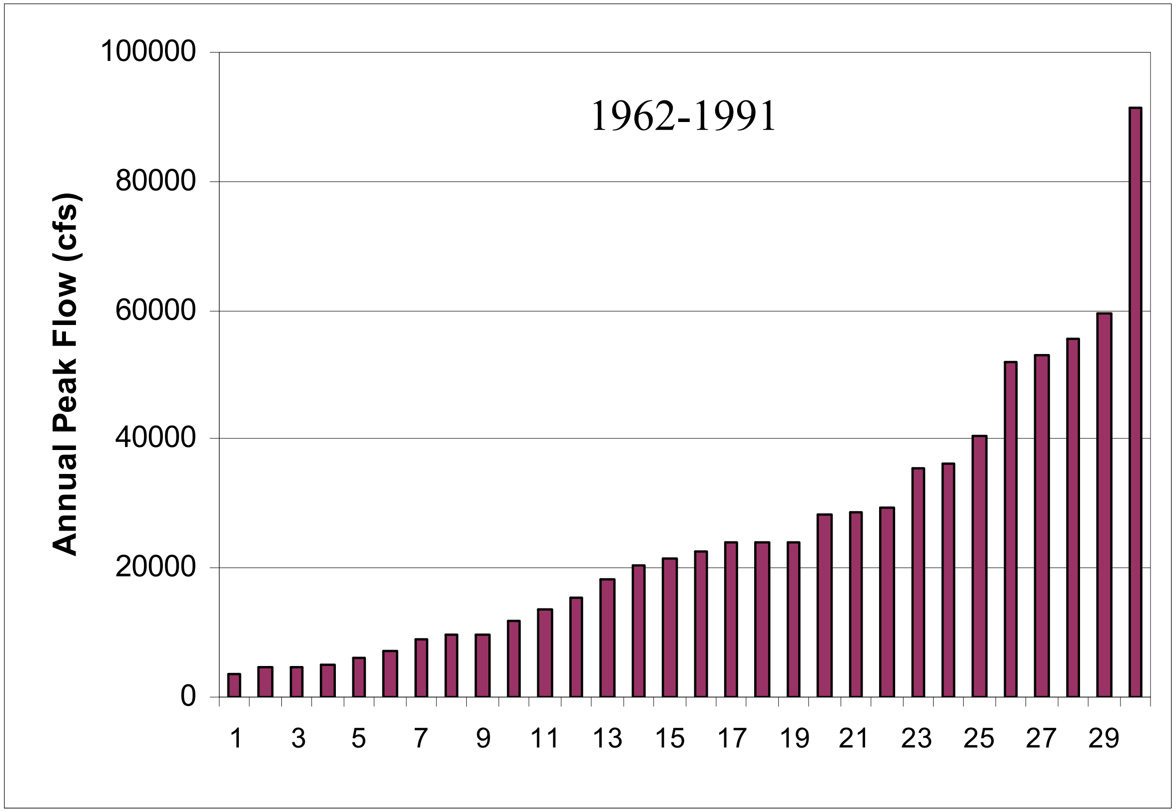
$$\text{Estimated Probability} = \frac{\text{\# of Occurrences}}{\text{\# of Trials}} \quad \rightarrow \text{\# years}$$

$$= \text{Relative Frequency}$$

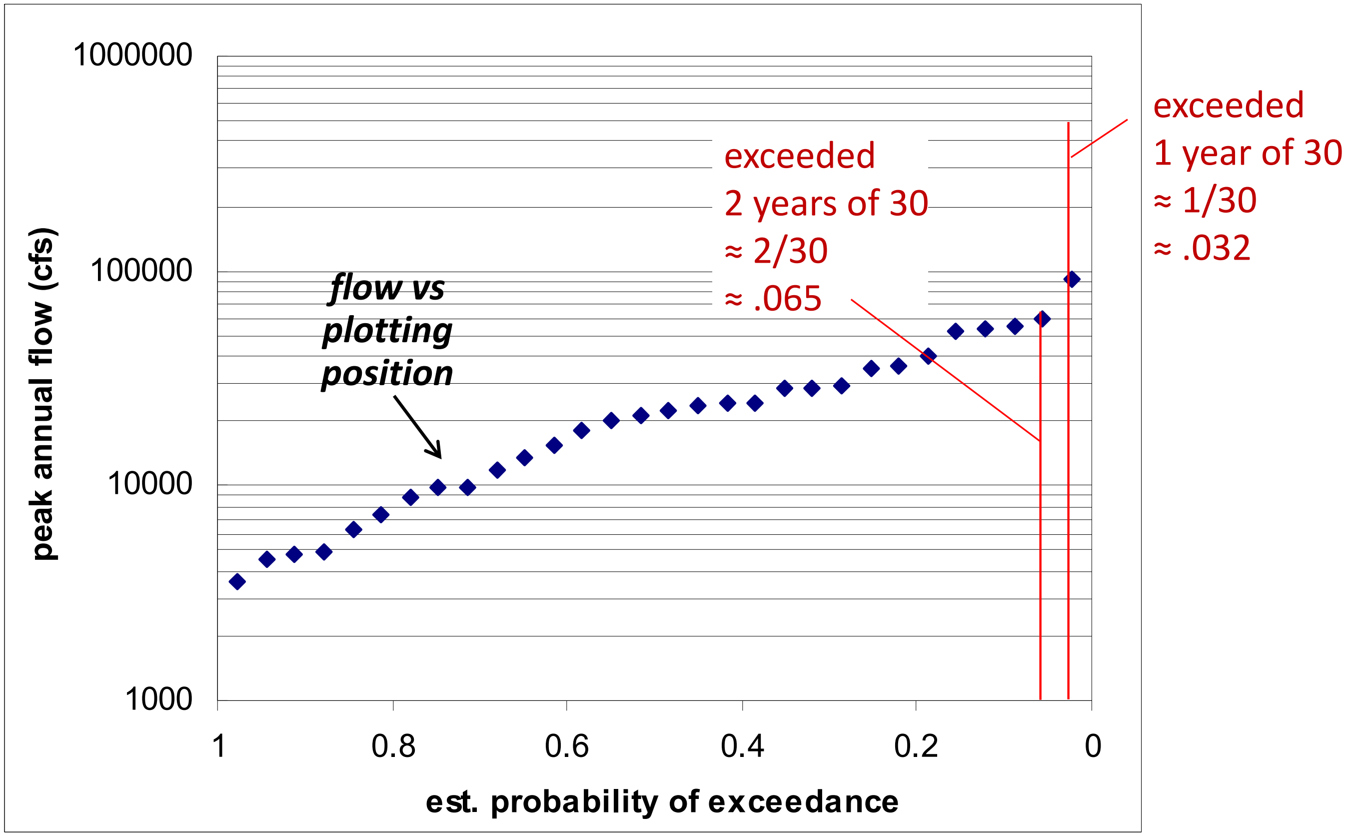
a 30-year record of annual peak flows



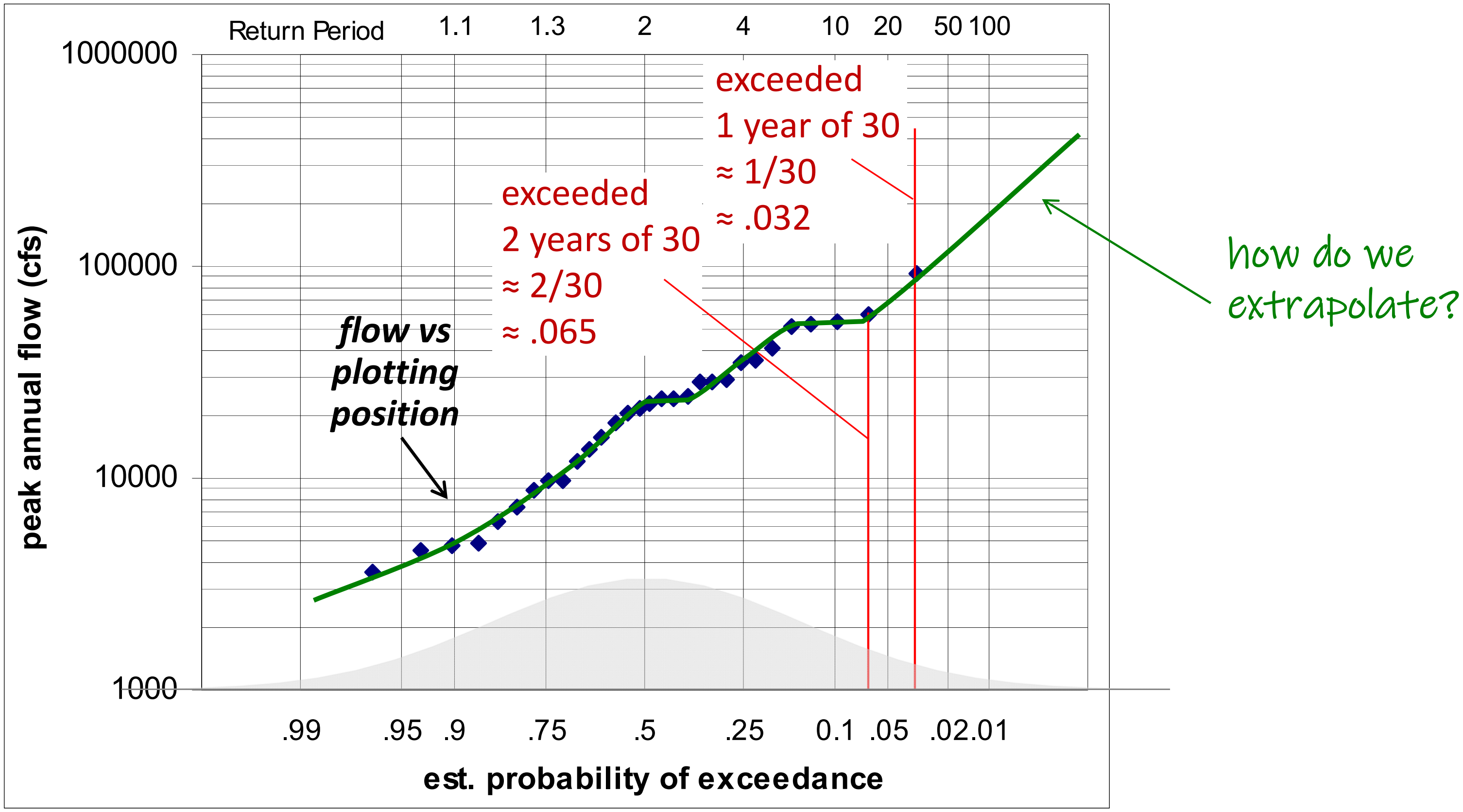
Rank the annual events…



Estimate Exceedance Probability by Relative Frequency



Use log(flow) and frequency scale



Remember the simple plotting position?

- Rank the N observed flows from largest to smallest

The top ranked event has rank	m = 1	the rank is the number of times the observation has been equaled or exceeded...
The second largest event has rank	m = 2	
...		
The smallest event has rank	m = N	

- Perform a class interval analysis (histogram) with exactly one observed value per interval

incremental probability of each	1/N
exceedance probability of each	m/N
simple Plotting Position	m/N

A Problem with simple PP = m/N

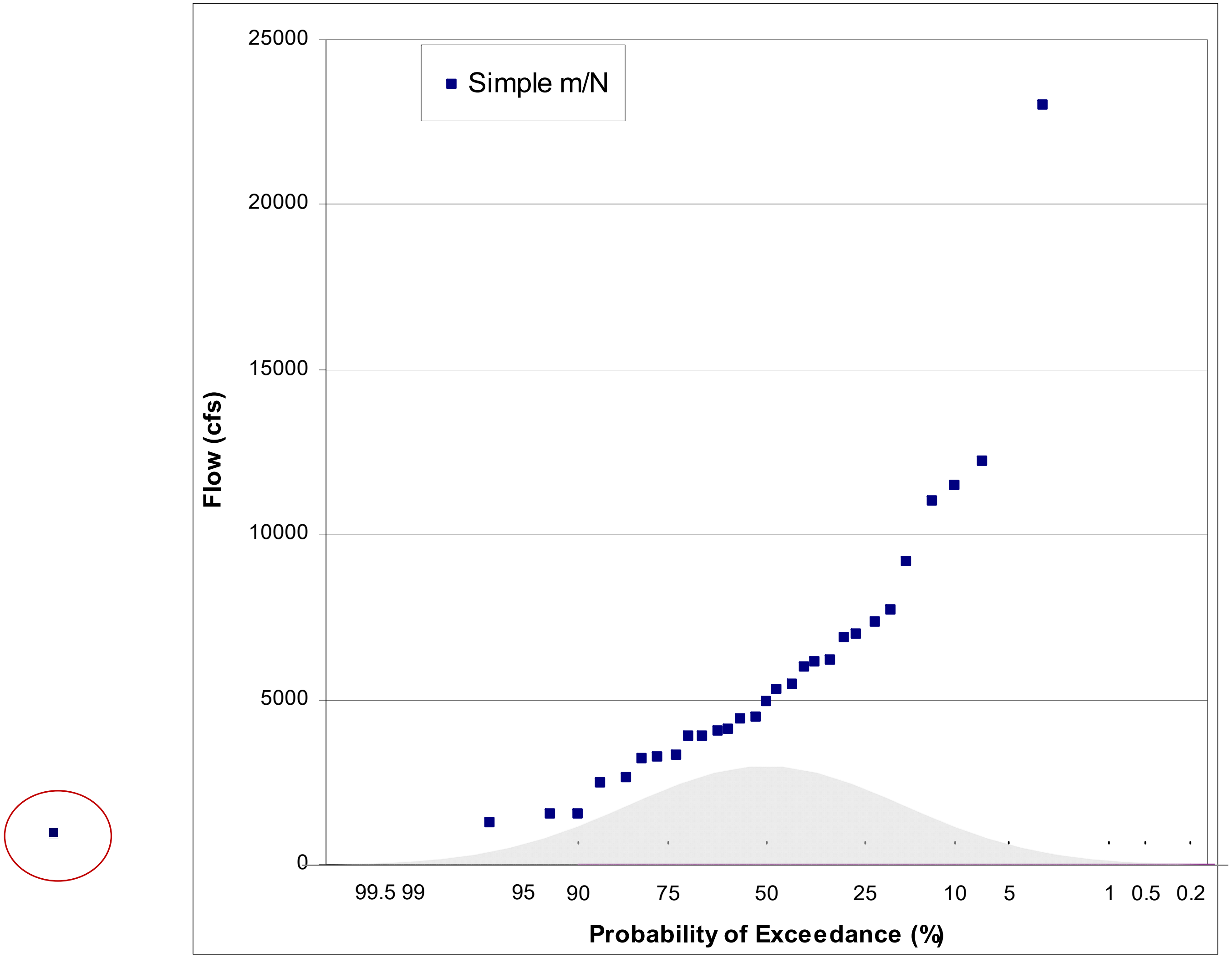
- Probability of exceeding the smallest value is N/N = 1.0
  - this is unreasonable
  - it also implies a problem with the exceedance probability estimates for other values
- Plotting position m/N is therefore statistically biased

EXAMPLE:

MILL CREEK, LOS MOLINOS, CALIFORNIA

PLOTTING POSITION m/N

water year	date	flow (cfs)	ranked flow (cfs)	rank m	plotting position m/N
1929	3feb	1520	23000	1	0.03
1930	15dec	6000	12200	2	0.07
1931	23jan	1500	11500	3	0.10
1932	24dec	5440	11000	4	0.13
1933	16mar	1080	9180	5	0.17
1934	29dec	2630	7710	6	0.20
1935	4jan	4010	7320	7	0.23
1936	21feb	4380	6970	8	0.27
1937	14feb	3310	6880	9	0.30
1938	11dec	23000	6180	10	0.33
1939	8mar	1260	6140	11	0.37
1940	28feb	11400	6000	12	0.40
1941	10feb	12200	5440	13	0.43
1942	6feb	11000	5280	14	0.47
1943	8mar	6970	4910	15	0.50
1944	4mar	3220	4430	16	0.53
1945	5feb	3230	4380	17	0.57
1946	21dec	6180	4070	18	0.60
1947	12feb	4070	4010	19	0.63
1948	23mar	7320	3870	20	0.67
1949	11mar	3870	3870	21	0.70
1950	4feb	4430	3310	22	0.73
1951	16nov	3870	3230	23	0.77
1952	26dec	5280	3220	24	0.80
1953	9jan	7710	2630	25	0.83
1954	17jan	4910	2480	26	0.87
1955	11nov	2480	1520	27	0.90
1956	22dec	9180	1500	28	0.93
1957	24feb	6140	1260	29	0.97
1958	24feb	6880	1080	30	1.00



Simple m/N
Flow (cfs)
25000
20000
15000
10000
5000
0
99.5 99
95
90
75
50
25
10
5
1
0.5
0.2
Probability of Exceedance (%)

Choosing a Plotting Position

How Should a Plotting Position Be Chosen?

Plotting position is an estimator of population exceedance probability

- an estimator is generally a function of the data
- two plotting position estimators seen so far are the Median and Weibull plotting positions

Other estimators discussed so far have been for the MEAN, standard deviation and skew

Remember that an estimator for the mean was given as:

$$\hat{\mu} = \overline{X} = \sum_{i=1}^{i=N} \frac{X_i}{N}$$

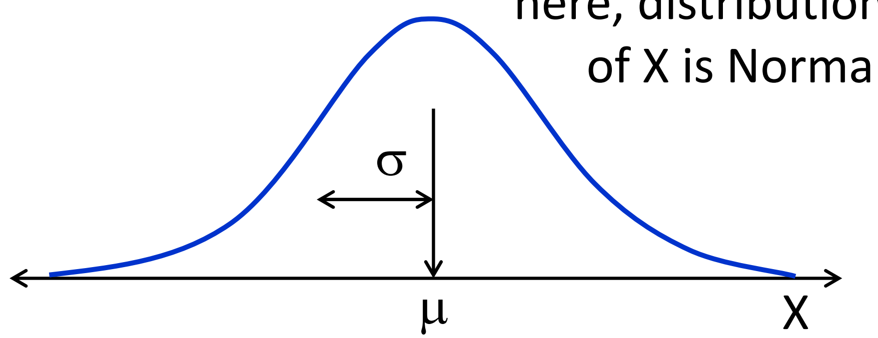
Creating a Sampling Distribution

Consider the sample mean, an unbiased estimator of the pop. mean

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{i=N} X_i$$

Let's examine properties of estimating the mean from 10 data points

- There exists an actual "population" probability distribution of X
- The distribution of X has a mean μ and standard deviation σ
- As an experiment, we'll generate 10-member random samples of flow X from the distribution of X above, 100 times

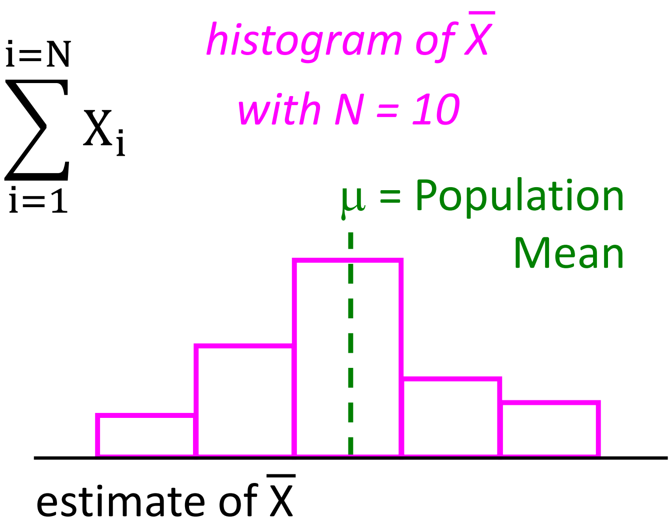


Creating a Sampling Distribution

- Generate 100 different 10-member samples of flow X
- Compute an estimate of the mean, $\overline{X}$, from each sample of 10

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{i=N} X_i$$

- A frequency analysis of the 100 estimates of $\overline{X}$ *(one from each 10-member sample)* would result in this <u>histogram</u>
- the average of 100 $\overline{X}$ would be close to the population value of the mean, μ
- the histogram of $\overline{X}$ could be approximated by a <u>Normal distribution</u>

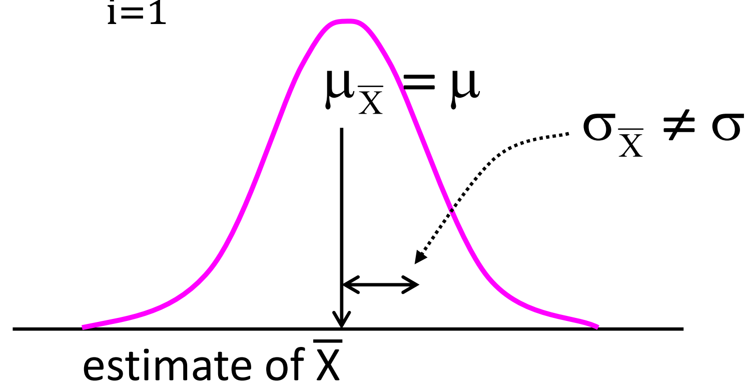


Sampling Distribution for the Mean

- CENTRAL LIMIT THEOREM: Since the estimator adds identical RVs, as N becomes very large, the distribution of $\overline{X}$ is Normal, with mean equal to the population value.
- The pink distribution to the right is the sampling distribution for the sample mean, $\overline{X}$

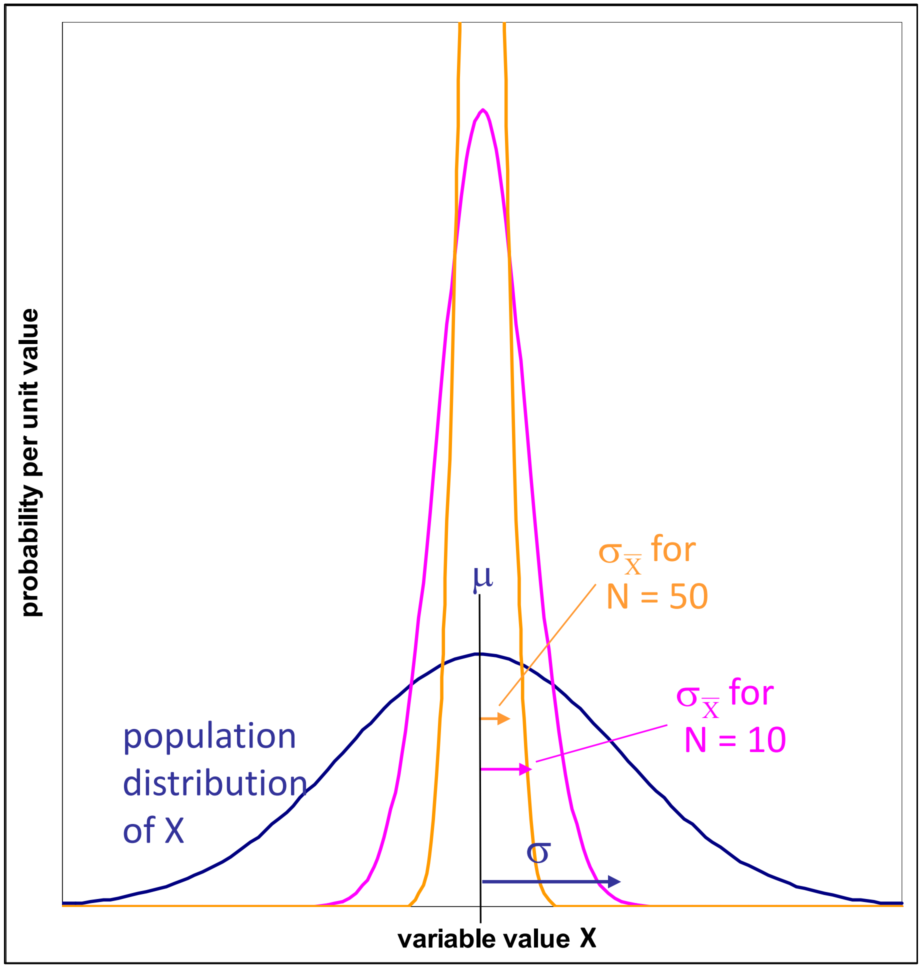
$$\overline{X} = \frac{1}{N}\sum_{i=1}^{i=N} X_i$$

distribution of sample means, $\overline{X}$

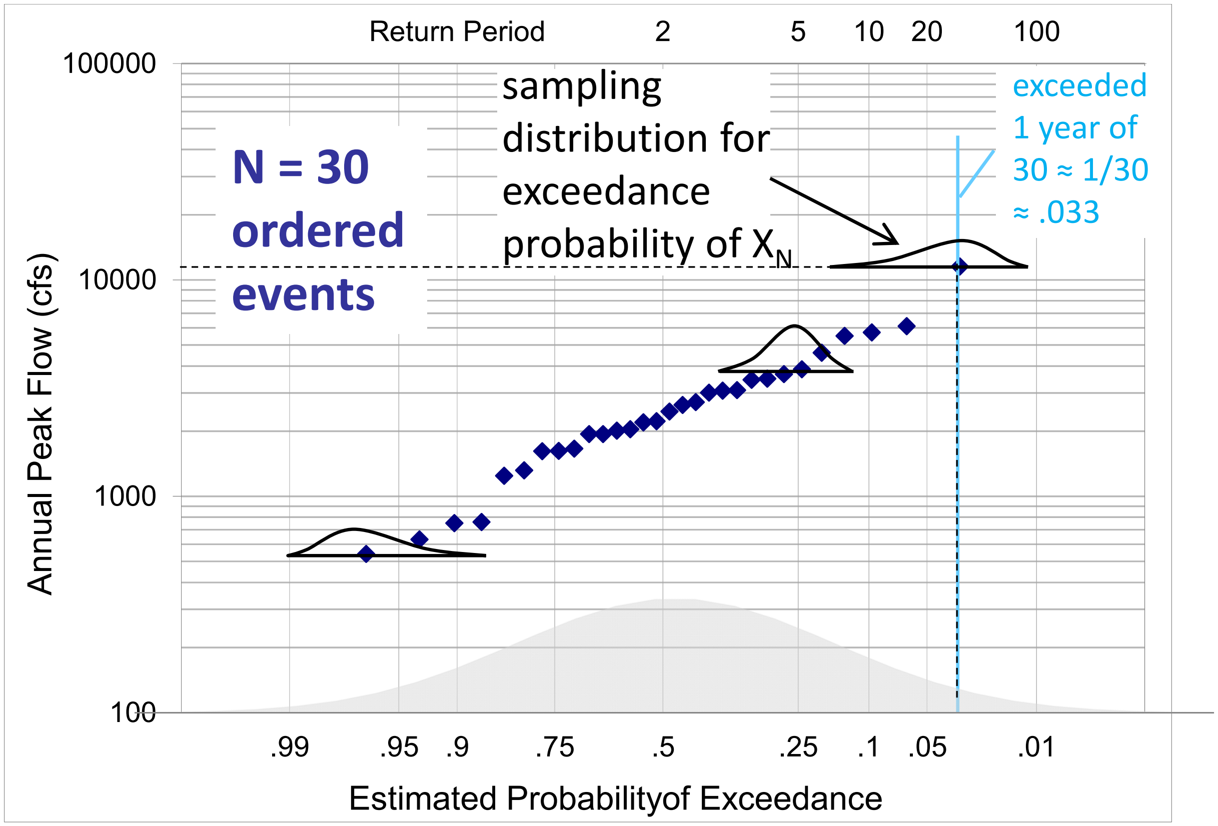


The standard deviation of the distribution-of-sample-means is equal to $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}$

Sampling Distribution for the Mean



Sampling Distribution for Plotting Position



The sampling distribution for the estimate of exceedance probability **NOT symmetrical**

Not Normal because Central Limit Theorem does not apply. *RVs are not summed*

What are we uncertain about?

- The **estimated** exceedance probability of the largest event in the record is based on the record length
  - i.e., the largest event in N=30 years is assigned a likelihood near 1/30, or 0.033 annual likelihood
- ***In what time period is it actually the largest? Longer than N? Shorter?***
- Can do a **sampling experiment** to see what the actual exceedance probability might be for the largest event in 30 years….
  - Sample 30 years from U[0,1] as the AEP of each event
  - Choose the smallest AEP = *largest event* from each sample
  - Repeat for 1,000 samples

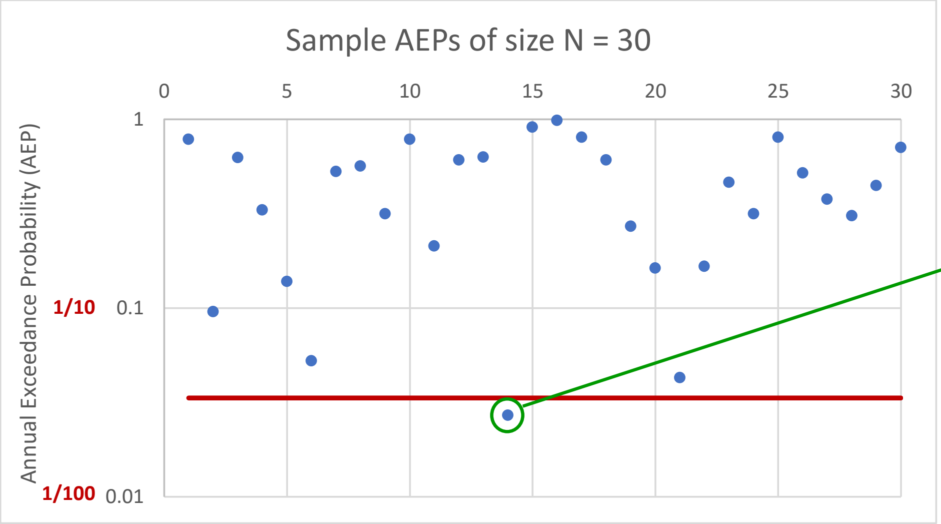
Source: Dave Margo, USACE / RMC

What's the real exceedance prob?

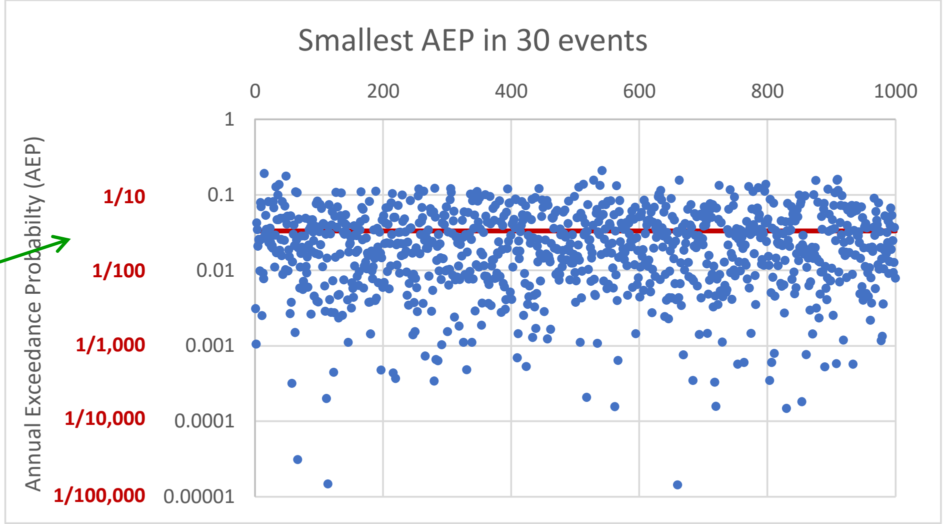
- Sample 30 "years" as 30 AEPs ~ U[0,1]
- Note the smallest AEP
- Repeat for 1,000 samples

Source:
Dave Margo,
USACE / RMC

1 sample of 30 events as AEPs



1000 samples size N=30, <u>smallest</u> AEP



Weibull Plotting Position

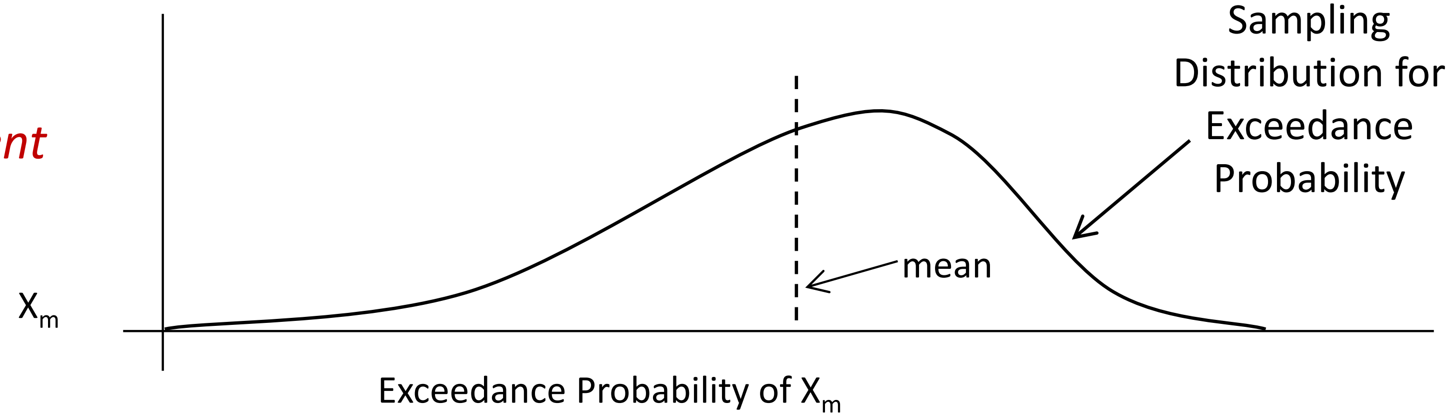
The Weibull plotting position is the mean of the sampling distribution at X_m, which results in an unbiased estimate of the exceedance probability

$$PP_m = \frac{m}{N+1}$$

Return Period of largest event = N + 1

For N=30, $PP_1 = 0.032$

Sampling Experiment
Smallest AEP of 30: average of 1000 samples = 0.031



Median Plotting Position

The Median plotting position is the median value of the sampling distribution for estimating the population probability of X_m

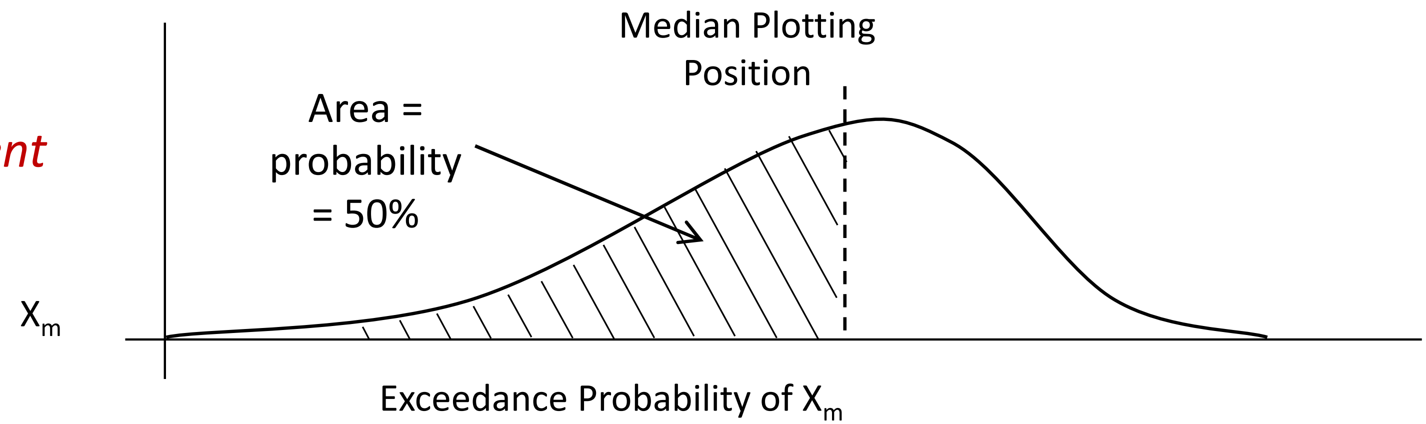
An approximation for N < 100 is:

$$PP_m = \frac{m - 0.3}{N + 0.4}$$

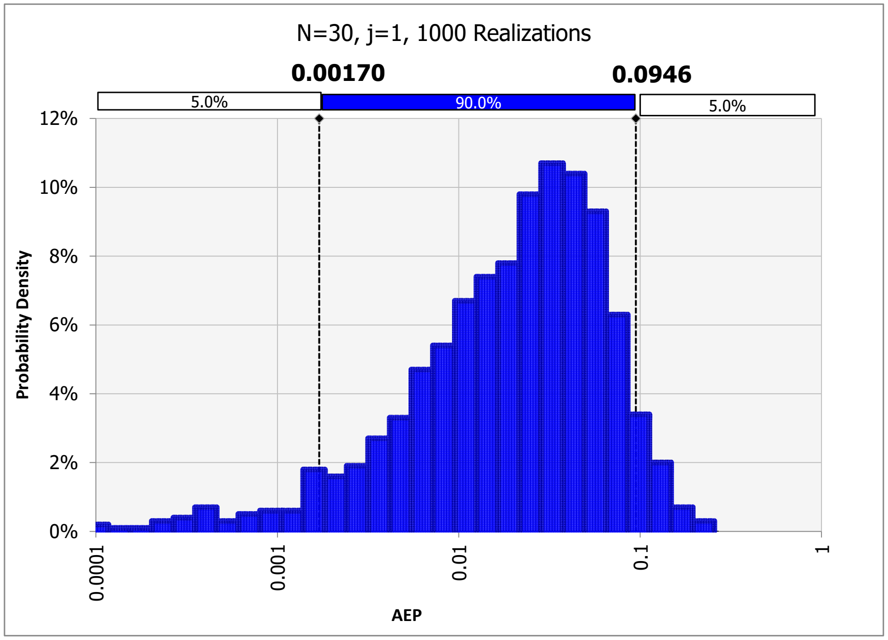
Return Period of largest event = 1.5N + 1

For N=30,
$PP_1 = 0.023$

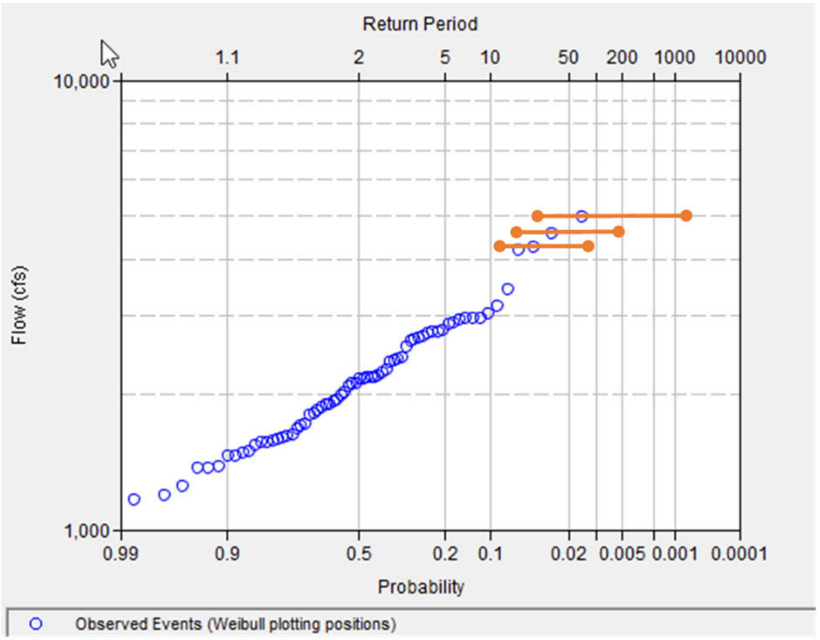
Sampling Experiment
Smallest AEP of 30: median of 1000 samples = 0.023



Plotting Position Uncertainty



j=1 (largest event) has almost 2 orders of magnitude uncertainty



Plot with Weibull PP
Depiction of plotting position uncertainty for the largest three events

Source:
Dave Margo, USACE / RMC

Plotting Position Choice

Recommend using Weibull for graphical frequency curves when economic analysis involved

- Calculation of damages requires an expected or mean probability, so use Weibull to estimate graphical a curve
- The Weibull plotting position gives an expected value

Recommend using Median when comparing plotting positions to compute frequency curve such as LP3

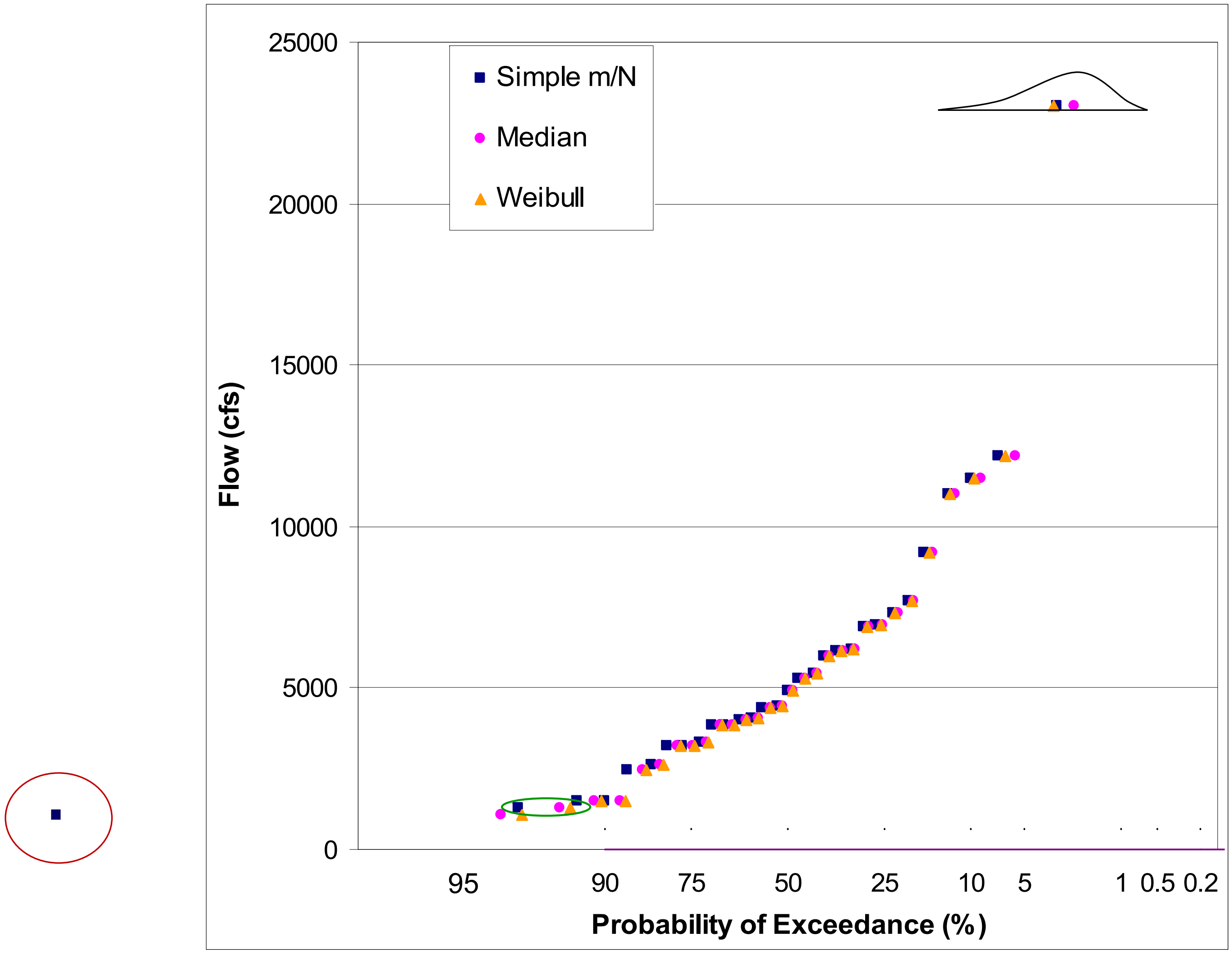
- Because comparing to a median estimate of LP3

EXAMPLE:

MILL CREEK, LOS MOLINOS, CALIFORNIA

PLOTTING POSITION Comparison

water year	date	flow (cfs)	ranked flow (cfs)	rank m	plotting position m/N	Median	Weibull
1929	3feb	1520	23000	1	0.03	0.023	0.032
1930	15dec	6000	12200	2	0.07	0.056	0.065
1931	23jan	1500	11500	3	0.10	0.089	0.097
1932	24dec	5440	11000	4	0.13	0.122	0.129
1933	16mar	1080	9180	5	0.17	0.155	0.161
1934	29dec	2630	7710	6	0.20	0.188	0.194
1935	4jan	4010	7320	7	0.23	0.220	0.226
1936	21feb	4380	6970	8	0.27	0.253	0.258
1937	14feb	3310	6880	9	0.30	0.286	0.290
1938	11dec	23000	6180	10	0.33	0.319	0.323
1939	8mar	1260	6140	11	0.37	0.352	0.355
1940	28feb	11400	6000	12	0.40	0.385	0.387
1941	10feb	12200	5440	13	0.43	0.418	0.419
1942	6feb	11000	5280	14	0.47	0.451	0.452
1943	8mar	6970	4910	15	0.50	0.484	0.484
1944	4mar	3220	4430	16	0.53	0.516	0.516
1945	5feb	3230	4380	17	0.57	0.549	0.548
1946	21dec	6180	4070	18	0.60	0.582	0.581
1947	12feb	4070	4010	19	0.63	0.615	0.613
1948	23mar	7320	3870	20	0.67	0.648	0.645
1949	11mar	3870	3870	21	0.70	0.681	0.677
1950	4feb	4430	3310	22	0.73	0.714	0.710
1951	16nov	3870	3230	23	0.77	0.747	0.742
1952	26dec	5280	3220	24	0.80	0.780	0.774
1953	9jan	7710	2630	25	0.83	0.813	0.806
1954	17jan	4910	2480	26	0.87	0.845	0.839
1955	11nov	2480	1520	27	0.90	0.878	0.871
1956	22dec	9180	1500	28	0.93	0.911	0.903
1957	24feb	6140	1260	29	0.97	0.944	0.935
1958	24feb	6880	1080	30	1.00	0.977	0.968



25000
20000
15000
10000
5000
0
Simple m/N
Median
Weibull
Flow (cfs)
95
90
75
50
25
10
5
1
0.5
0.2
Probability of Exceedance (%)

Outline

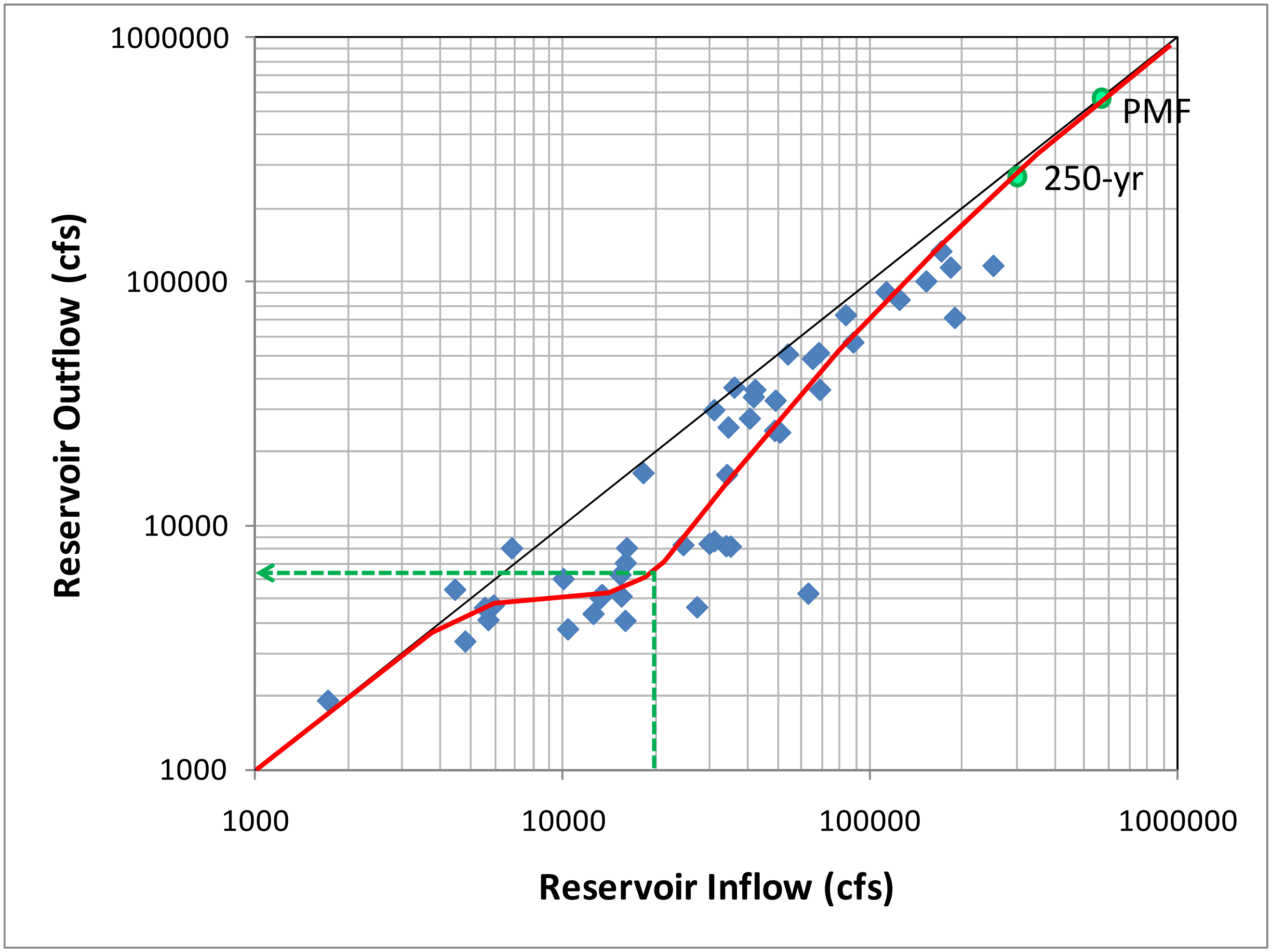
- Graphical Frequency Analysis
- Plotting Positions
  - Statistical estimation theory
  - Sampling Distributions to define uncertainty
- Non-analytic (Graphical) Frequency Curves
  - Regulated Flow Frequency Curves
  - Stage Frequency Curves
- Partial Duration Series *(peaks over threshold)*
  - as graphical, rather than analytical

Regulated Frequency Curves

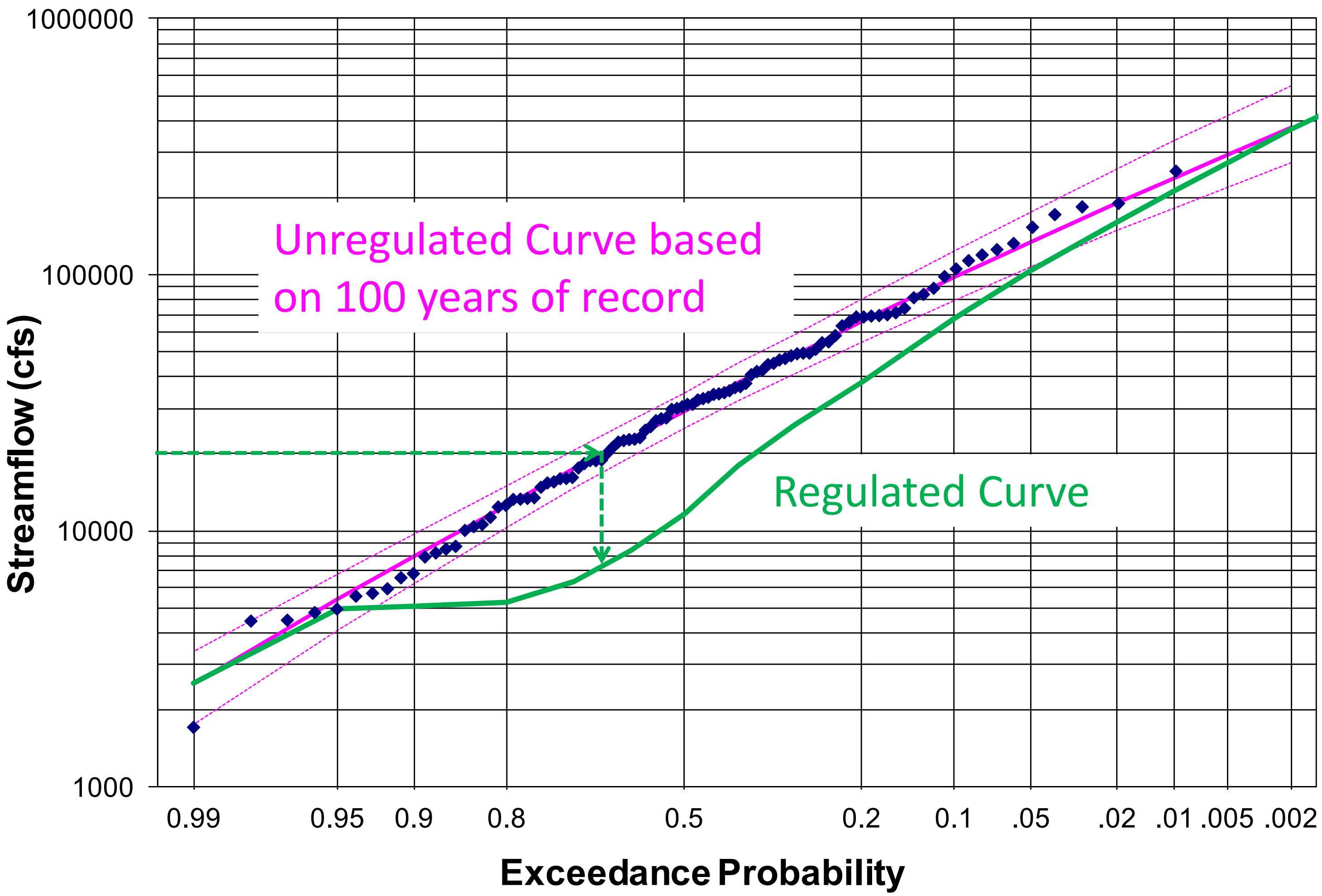
Two methods to produce a graphical regulated flow frequency curve:

1. Develop unregulated flow frequency curve with LP3 fit. Combine with reservoir regulation relationship to produce a regulated flow frequency curve
    - Note the added uncertainty in the inflow/outflow relationship
2. Estimate a regulated frequency curve with regulated flow data and graphical frequency analysis

Historical Reservoir Regulation Data



might instead create this relationship with a reservoir simulation model



1000000
100000
10000
1000
Streamflow (cfs)
0.99
0.95
0.9
0.8
0.5
0.2
0.1
.05
.02
.01
.005
.002
Exceedance Probability
Unregulated Curve based on 100 years of record
Regulated Curve

Regulated Frequency Curves

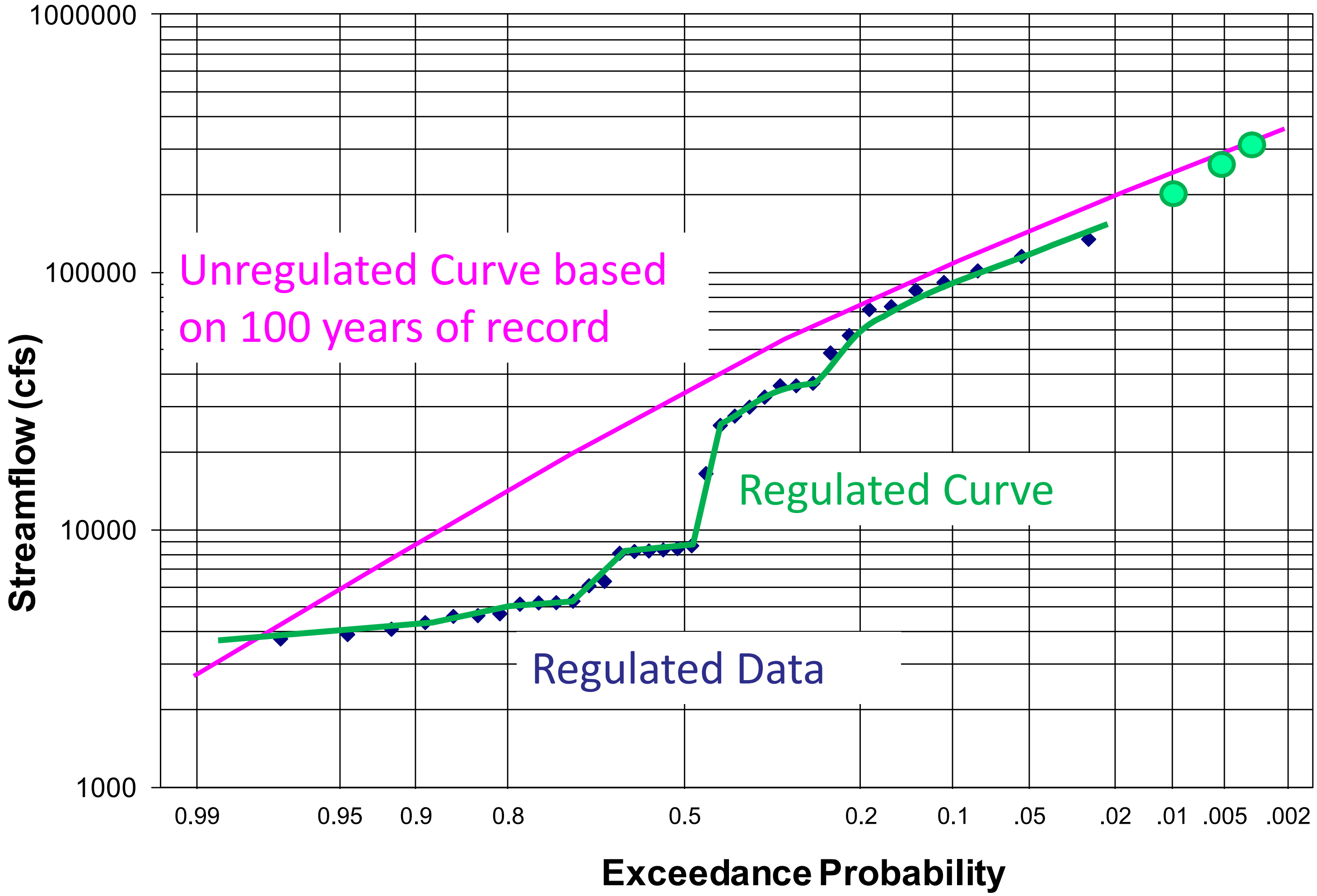
Second method: graphical fit to annual peak regulated flow data

Source of regulated flow data:

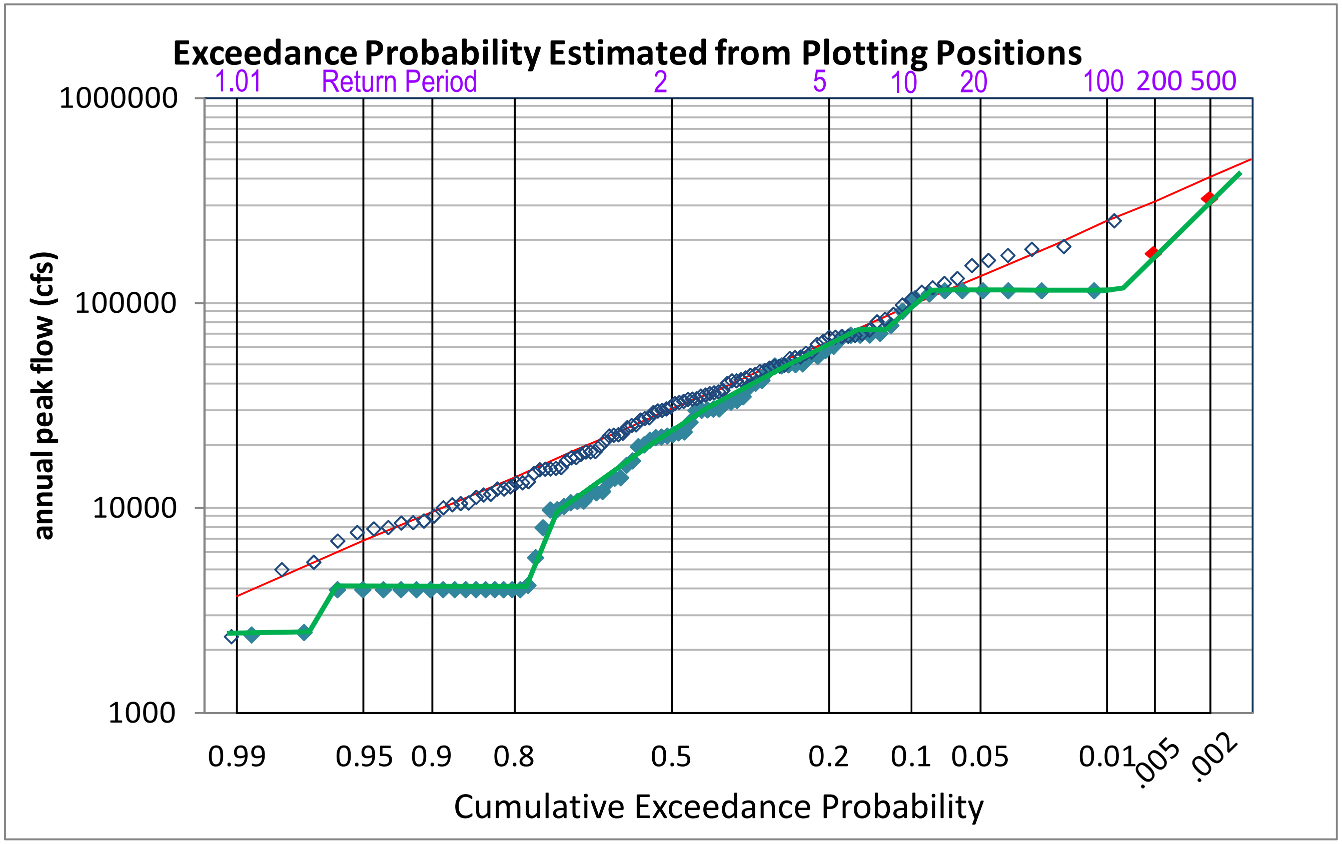
1. Observed data downstream of dam
2. Route inflows through dam with reservoir model
   - Use period of record inflows
   - Estimate inflows from continuous simulation model

Hypothetical events are needed to estimate upper end of frequency curve

- – define hypothetical inflow flood events (specified exc prob)
  - exceedance probability of outflow assumed same as inflows
  - starting water surface elevation is a critical consideration


Unregulated Curve based on 100 years of record
Regulated Curve
Regulated Data
Streamflow (cfs)
1000000
100000
10000
1000
0.99
0.95
0.9
0.8
0.5
0.2
0.1
.05
.02
.01
.005
.002
Exceedance Probability

Graphical Curve is Better for this



41

Additional Uncertainties in Regulation

- Often, there is incidental flood control provided by reservoir water supply pool
  - Reservoir starts with more flood space
  - Non-flood reservoirs upstream might reduce peak
- There are uncertainties in flood operation
  - Real-time flood forecasts are sometimes used, and are contain a great deal of uncertainty
  - Releases might be curtailed or diminished due to unforeseen contingencies, not using full capacities.
  - Might reduce release to account for runoff below reservoir

Outline

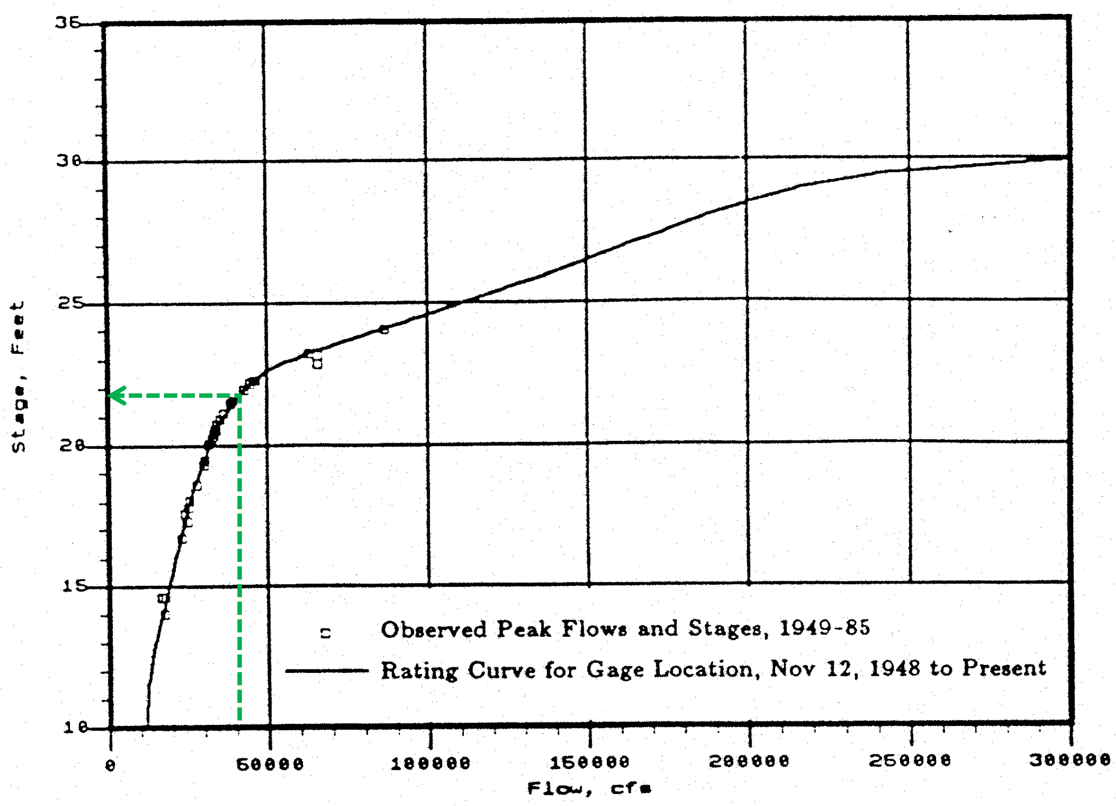
- Graphical Frequency Analysis
- Plotting Positions
  - Statistical estimation theory
  - Sampling Distributions to define uncertainty
- Non-analytic (Graphical) Frequency Curves
  - Regulated Flow Frequency Curves
  - Stage Frequency Curves
- Partial Duration Series *(peaks over threshold)*

Stage-Frequency Curves

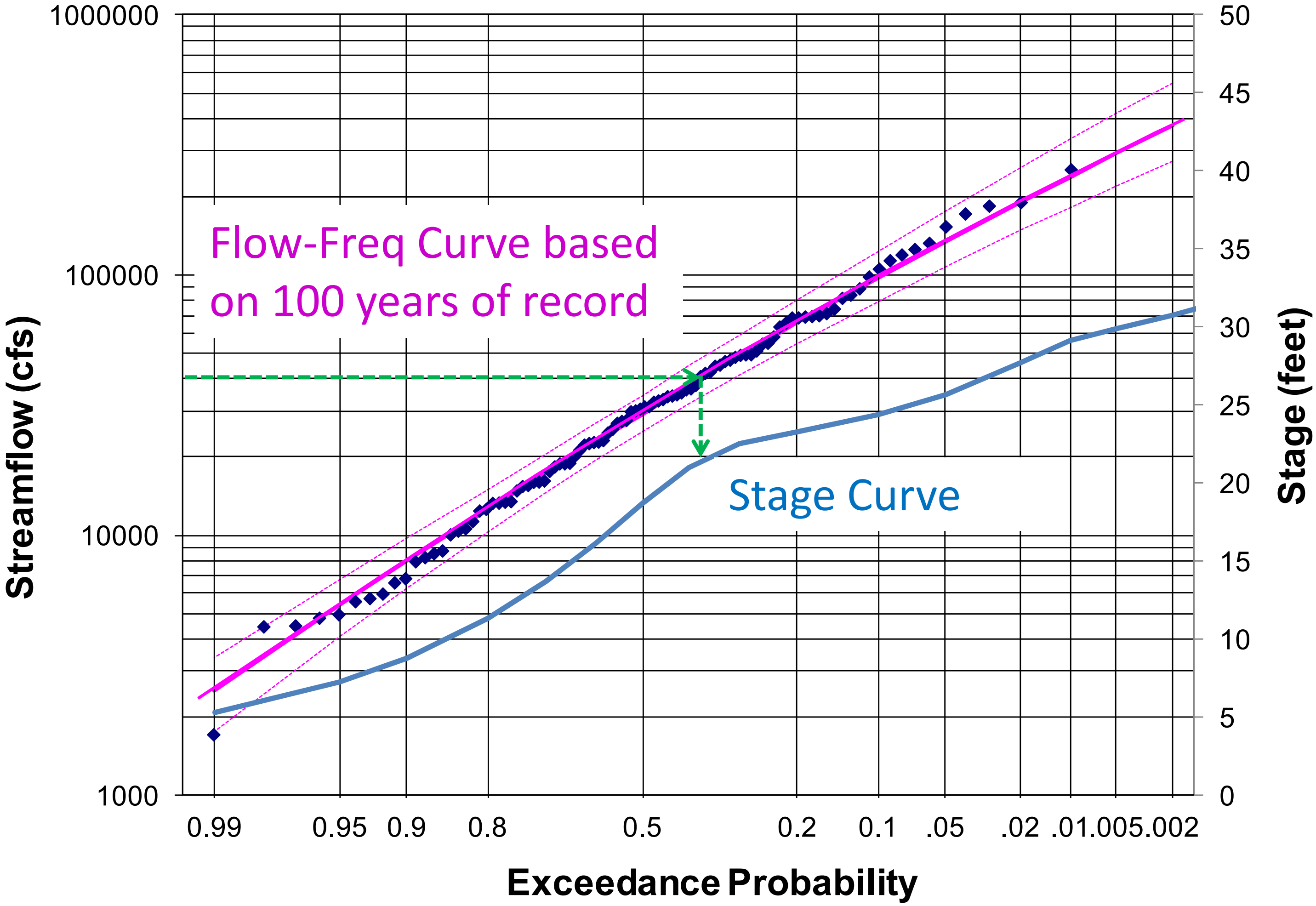
Two methods to produce a stage-frequency curve:

1. Develop a flow frequency curve. Combine with CURRENT flow-stage rating curve to produce stage-frequency curve
   - requires a one-to-one relationship between flow and stage, which might not be feasible when complex water systems involved (backwater, tidal)
   - better if flow is not regulated, and can use analytical fit
2. Estimate stage-frequency curve with stage data and graphical frequency analysis

Example Stage/Flow Rating Curve



Resulting Stage-Frequency Curve



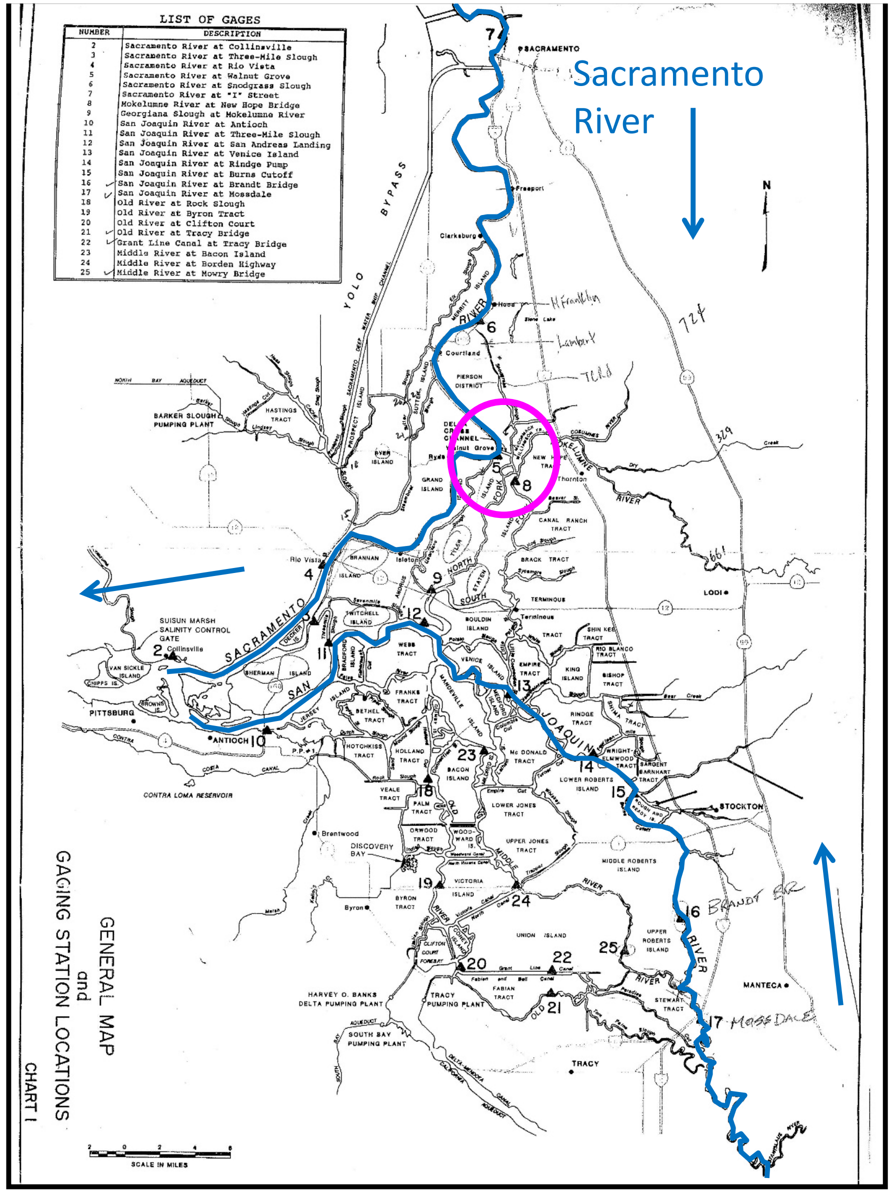
Stage Frequency Analysis

2nd Method, graphical fit to annual peak stage data:

- Use graphical frequency analysis, because analytical curve would smooth out relevant irregularities
- Pay attention to relevant stages: bottom, bankfull, etc
- A linear rather than log-scale is more reasonable
- Source of stage data:
  1. Observed gaged record (USGS peak flow file)
  2. Continuous simulation of hydraulic model, based on surrounding longer record gages

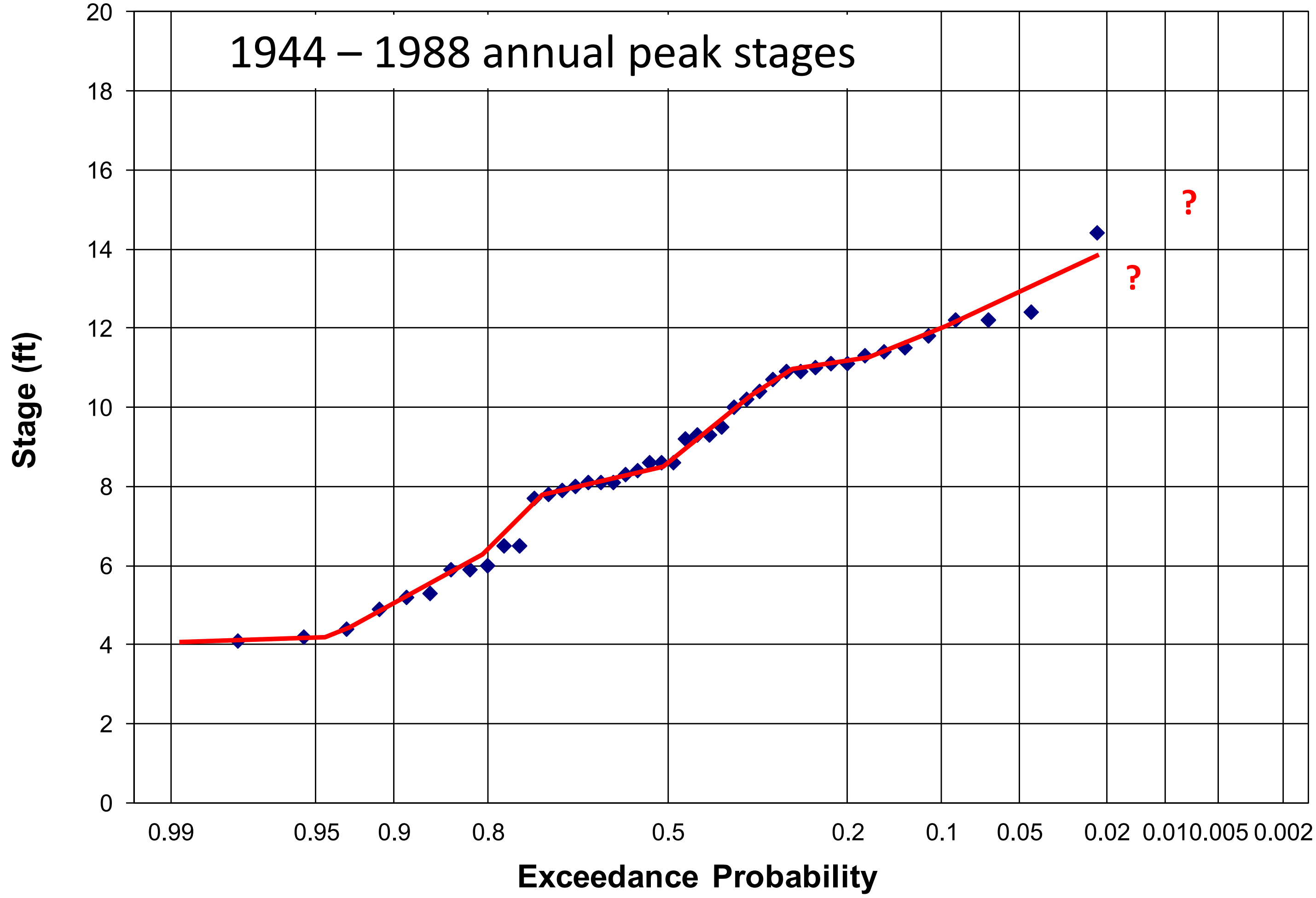
Sometimes, it is not possible to define a one-to-one relationship between flow and stage

Sacramento-San Joaquin Delta, confluence of Sacramento and San Joaquin Rivers

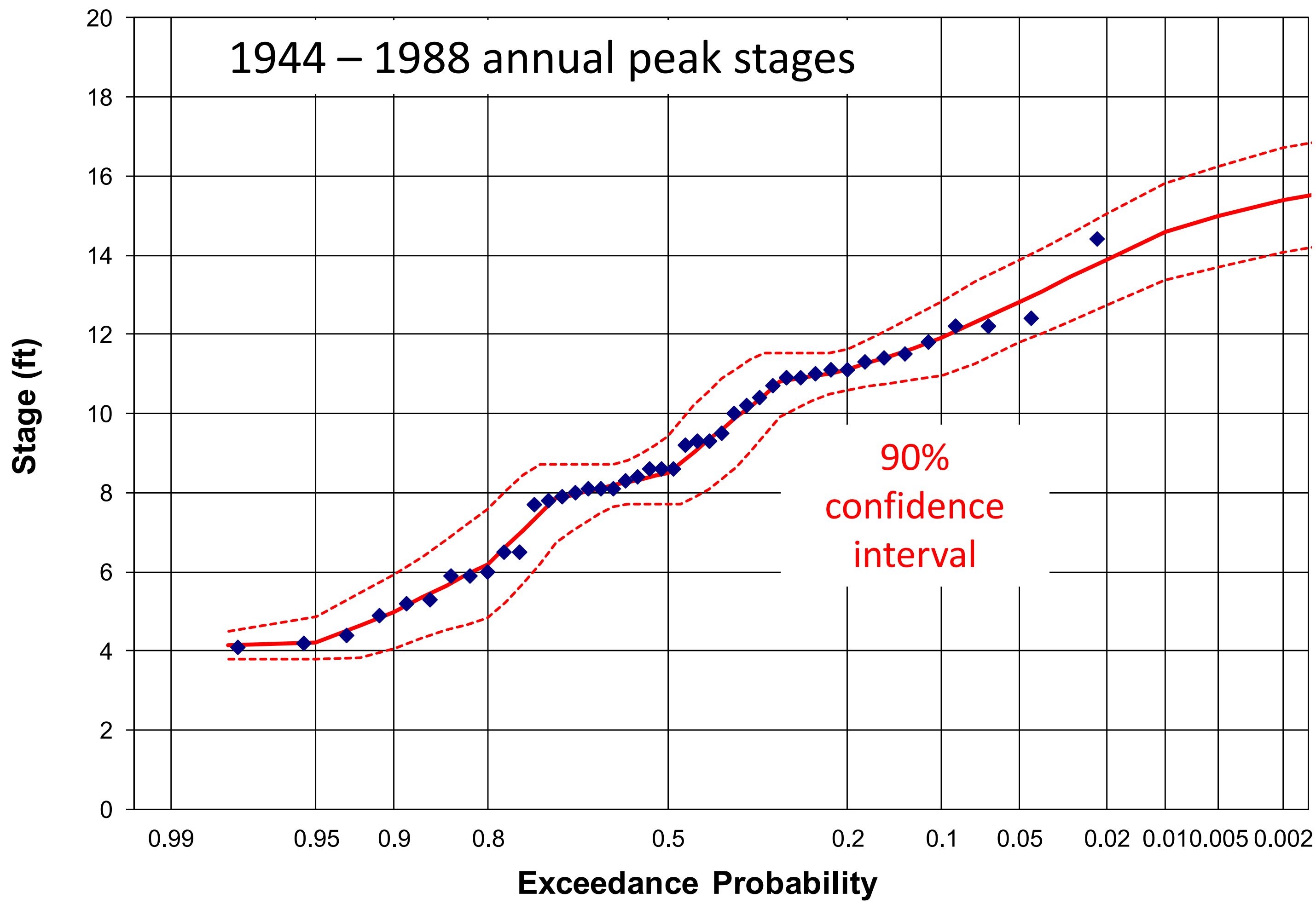


San Joaquin River

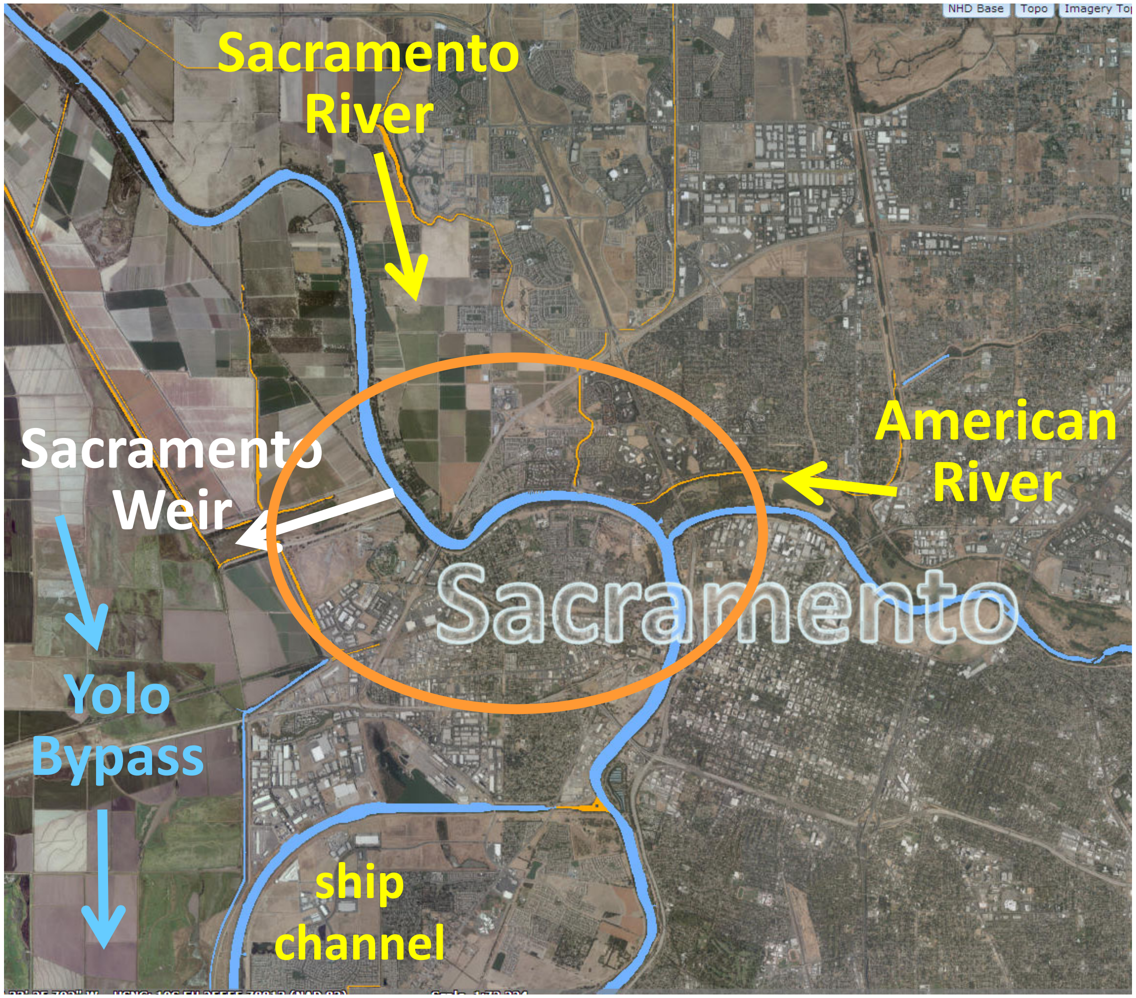
Sacramento River at Walnut Creek, Stage-Frequency

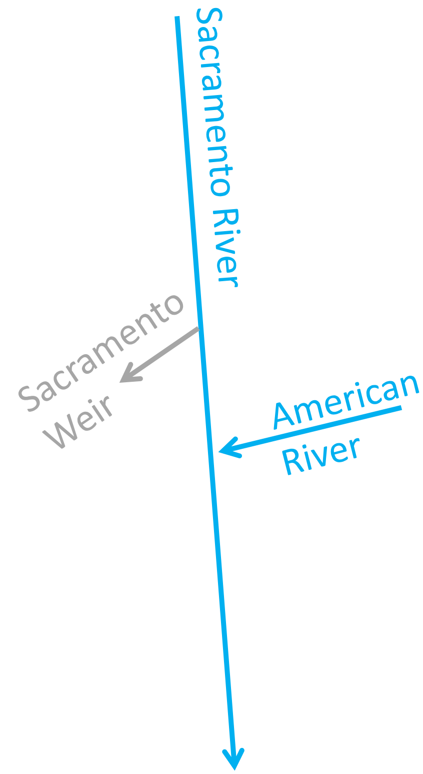


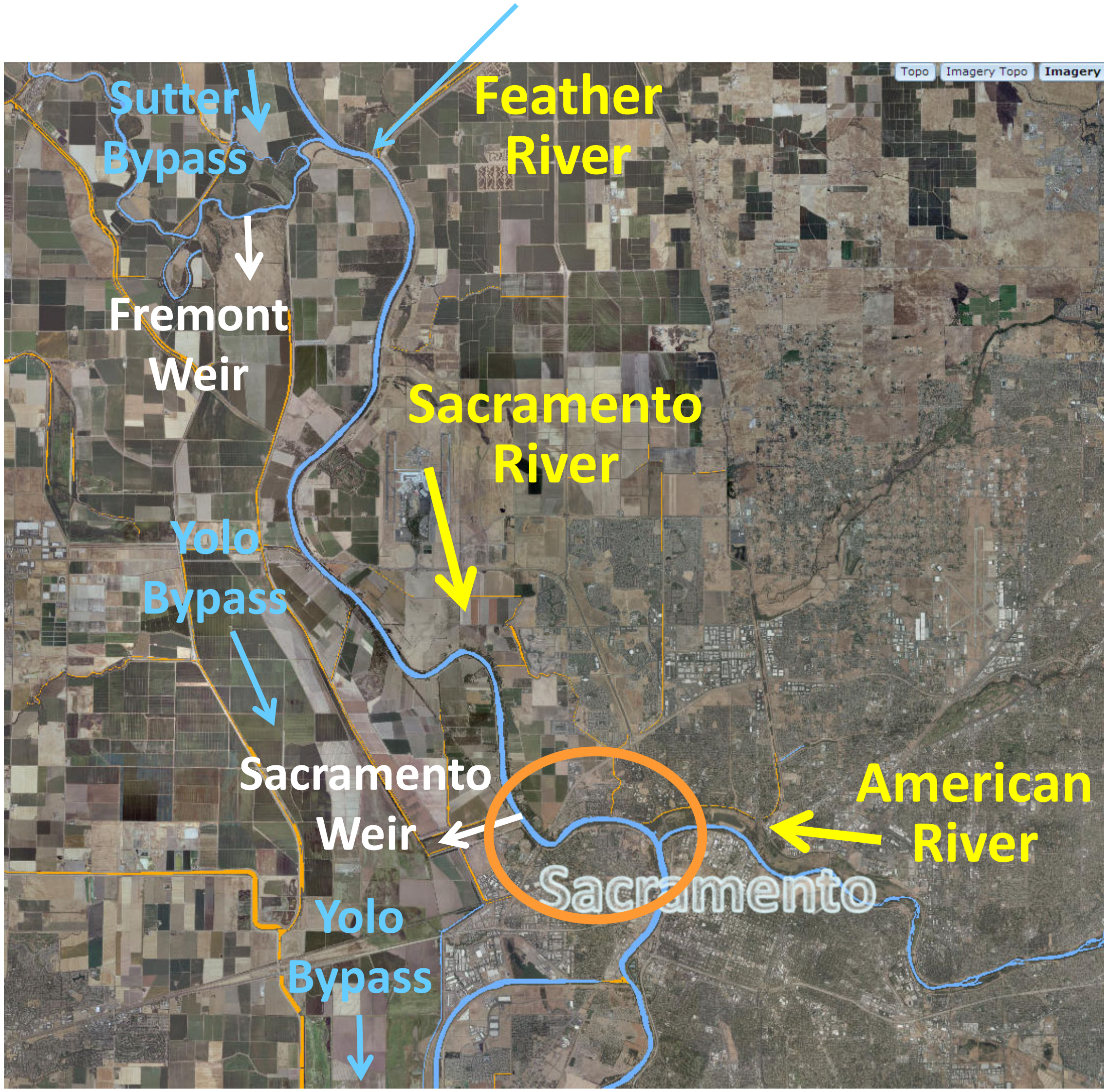
Sacramento River at Walnut Creek, Stage-Frequency

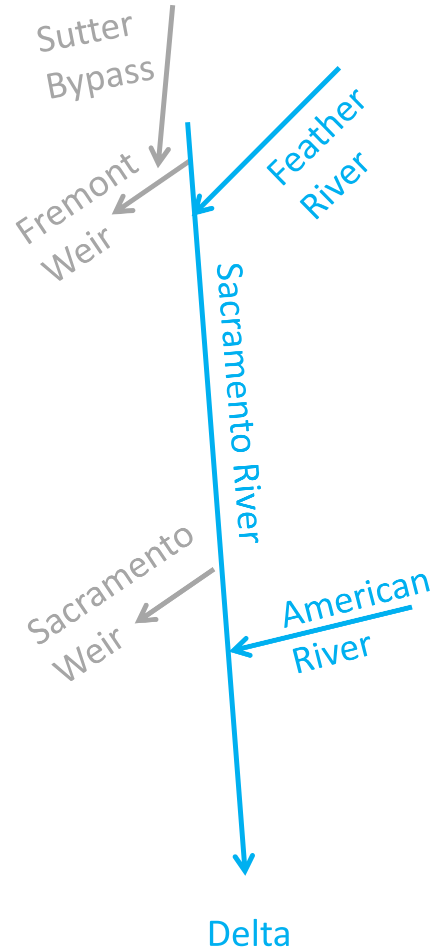


Hydraulic Modeling of the Period of Record



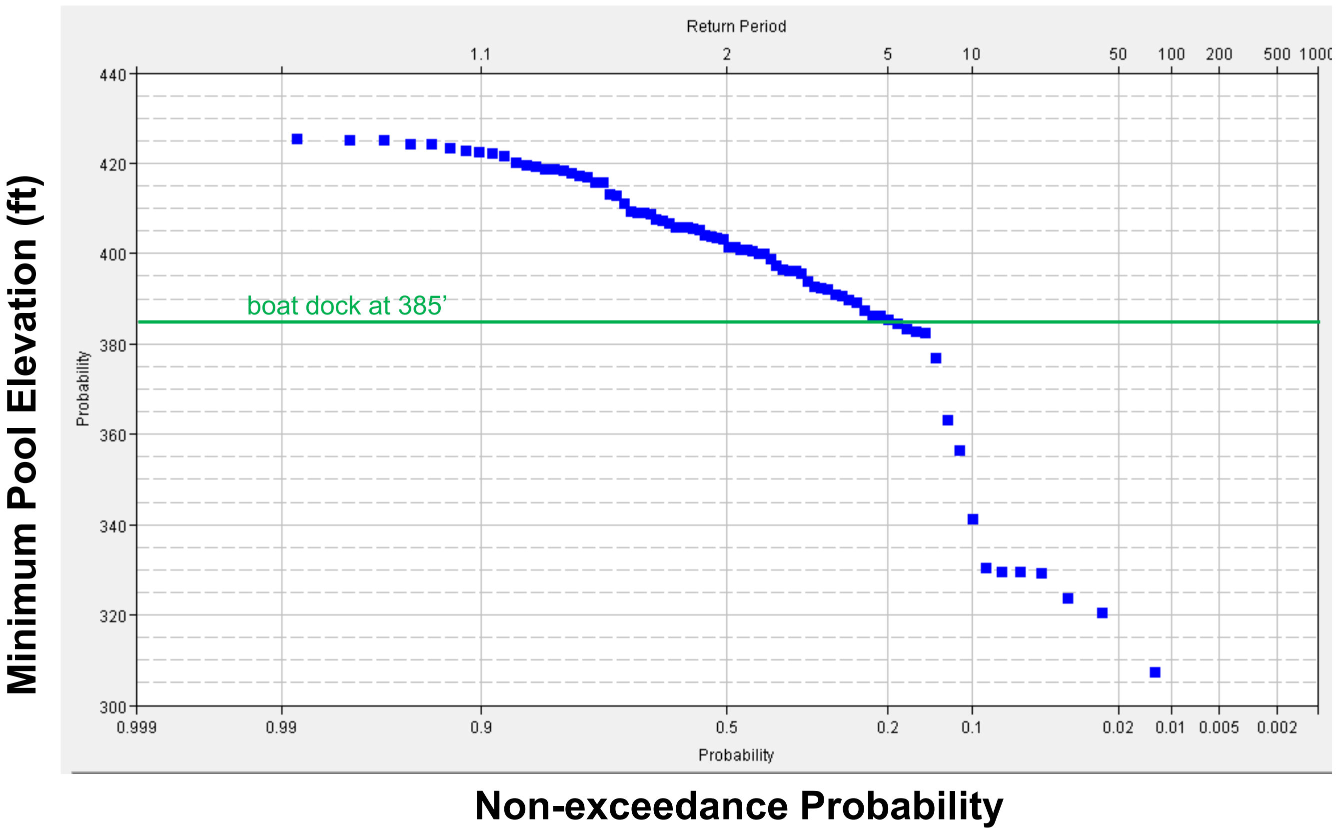




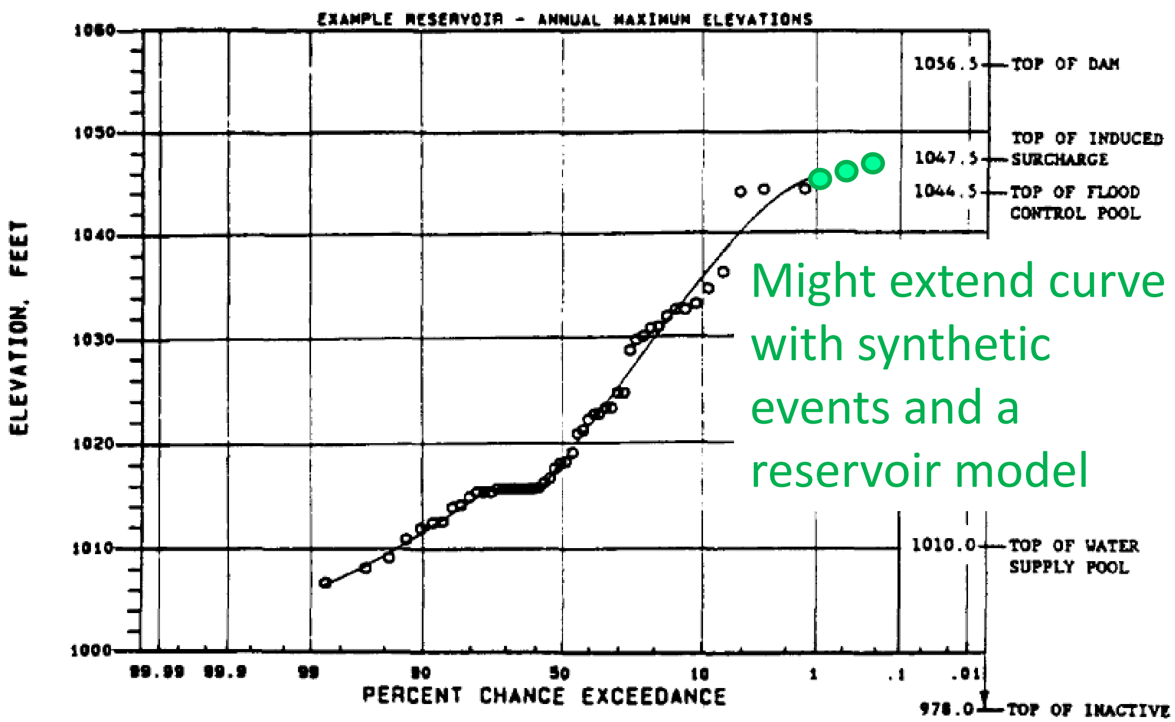


Sacramento-SanJoaquin Delta

Reservoir Min Stage vs frequency



Reservoir Max Stage vs frequency



Might extend curve with synthetic events and a reservoir model

might do Monte Carlo simulation for many synthetic events!

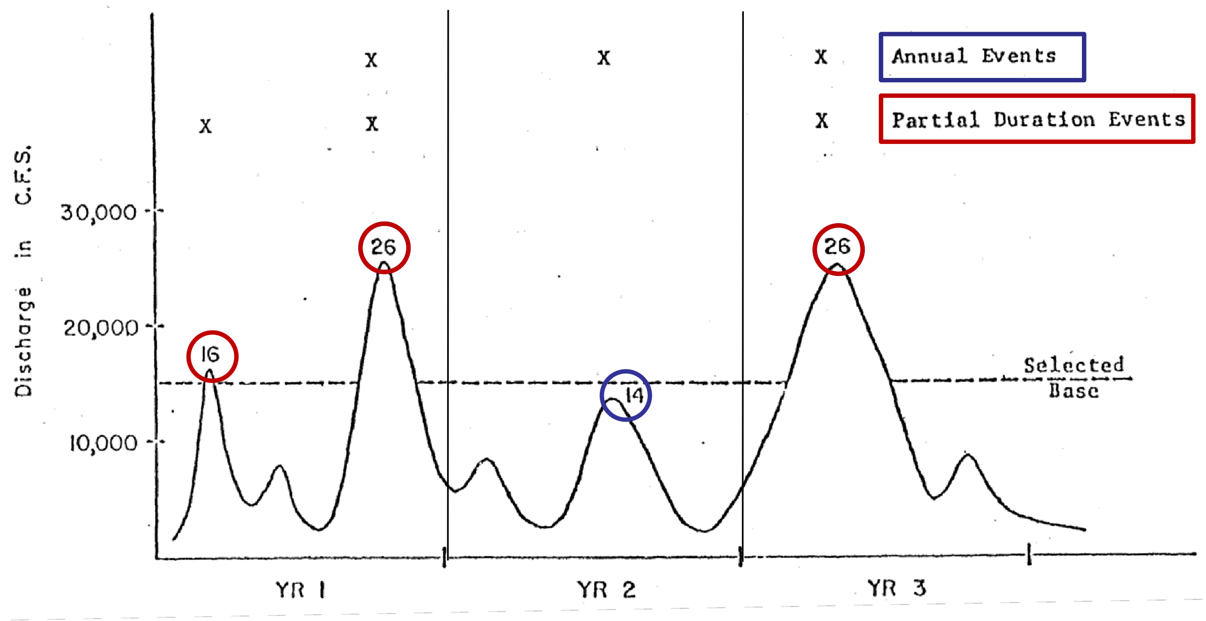
Outline

- Graphical Frequency Analysis
- Plotting Positions
  - Statistical estimation theory
  - Sampling Distributions to define uncertainty
- Non-analytic (Graphical) Frequency Curves
  - Regulated Flow Frequency Curves
  - Stage Frequency Curves
- Partial Duration Series *(peaks over threshold)*
  - as graphical, rather than analytical

Partial Duration Series

All peak events above a particular threshold

Annual Series vs Partial Duration



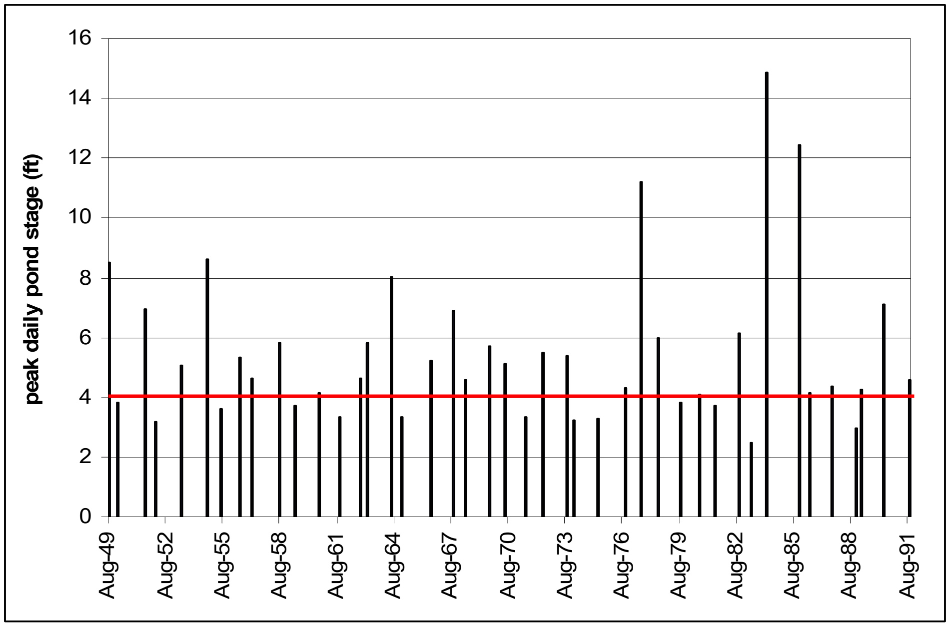
Partial Duration Series

All peak events above a particular threshold

- Select a base level for identifying peaks
  - The base level should be large enough so that each peak selected is an independent event
- Compile the peaks greater than the base level from the historic record
- Rank the events from largest to smallest
  - Note that the number of events might be larger than the number of years of record, **N**

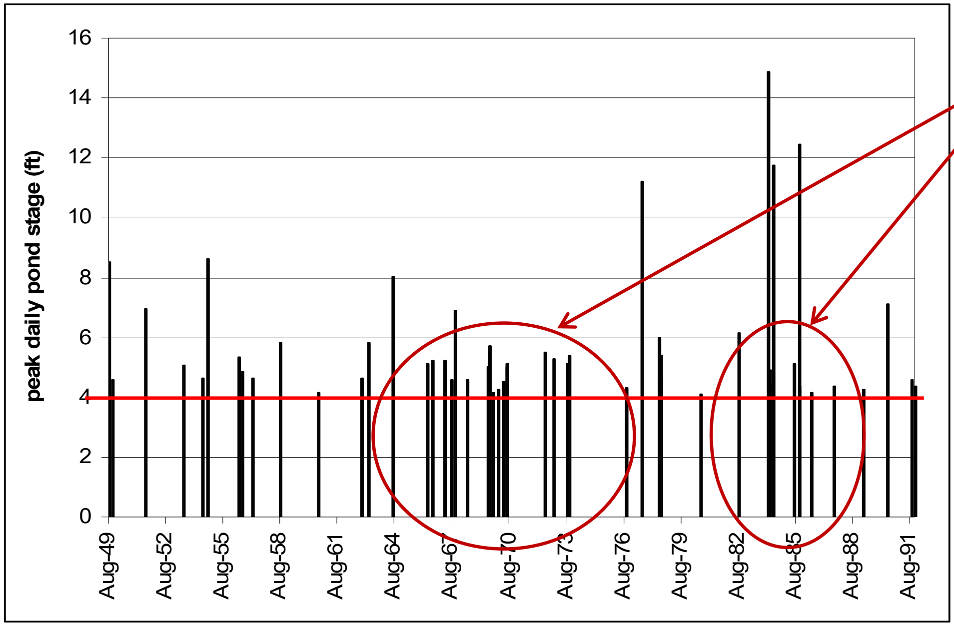
Annual Maximum Series

- If compile the peak event of each year, forming the annual series, there might be several values below interest, and interesting values missed



Partial Duration Series

- If more than one event per year is relevant, and low values are not relevant, use the partial duration series, or all peaks above a threshold



ensure all values are separate, independent events

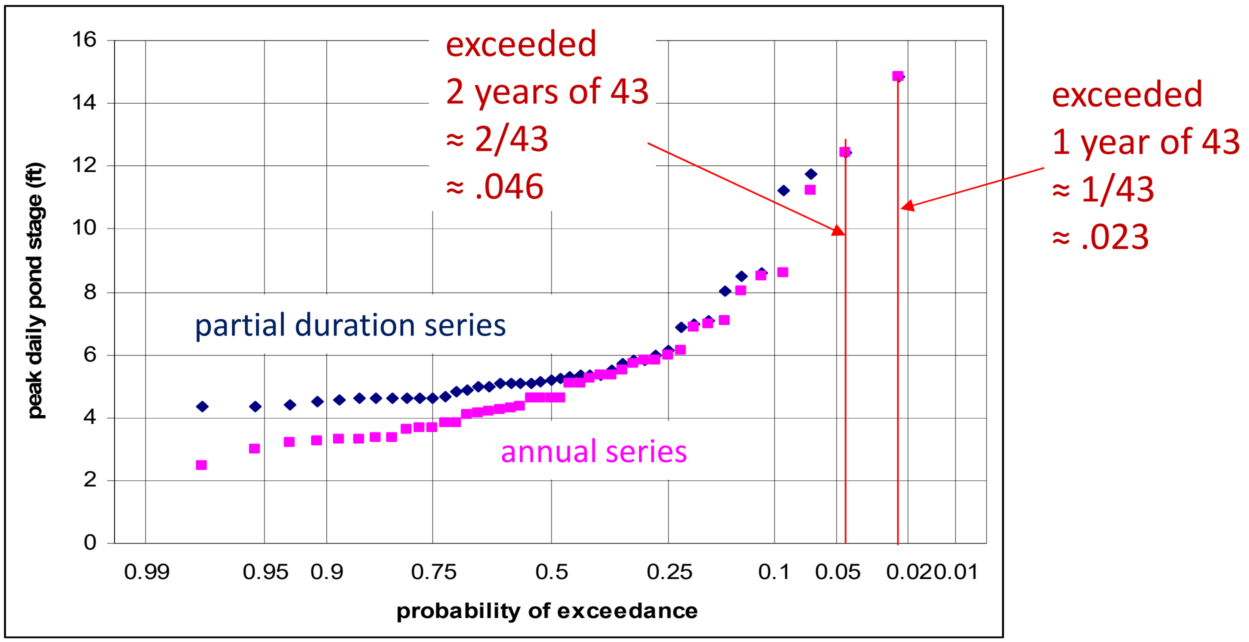
Partial Duration Series

- Calculate plotting positions
  - If there are a total of **k** partial duration series events, then the plotting position for the smallest event is ~ **k/N**
    - Note that if there are more events than years (k>N) this might be a number greater than 1.0
    - The plotting position for this event is interpreted as **k/N** events per year

<u>Example</u> Smallest event = 4 ft k=61, N=42	Then the plotting position estimate states that on the average it is expected that 1.5 events per year will exceed 4 ft

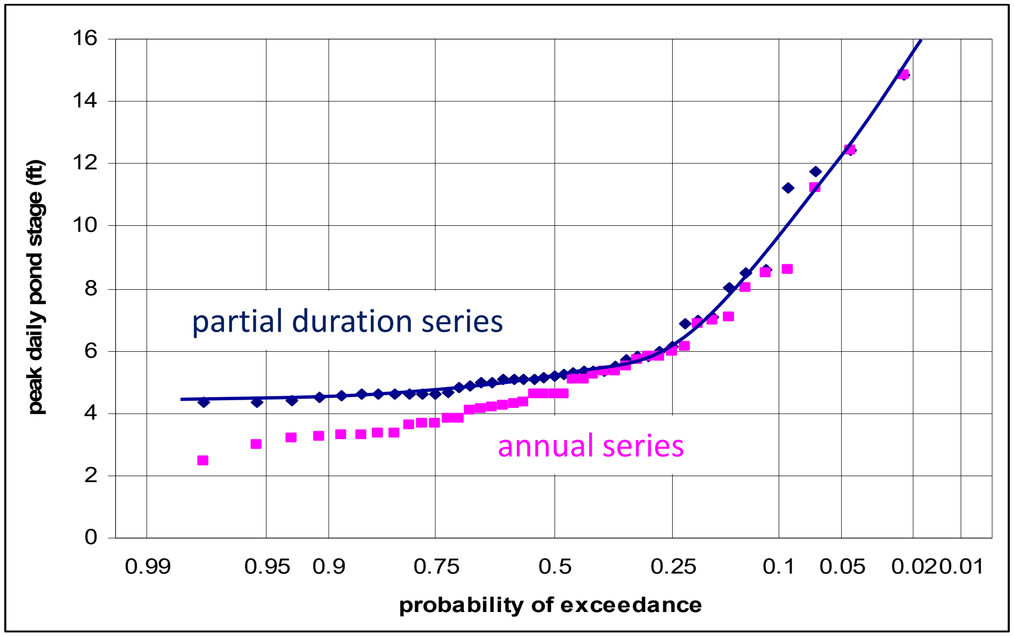
Frequency Analysis

- Compute estimated probability as
  relative frequency = # occurrences greater / # years



Frequency Analysis

- Generally, LP3 distribution does not fit a partial duration series frequency curve. Usually use graphical, or…..



Why use Partial Duration Series?

- Estimating expected damage when more than one damaging event per year
- Concerned about the low end of the curve, for example for riparian ecosystem restoration studies

Outline

- Graphical Frequency Analysis
- Plotting Positions
  - Statistical estimation theory
  - Sampling Distributions to define uncertainty
- Non-analytic (Graphical) Frequency Curves
  - Regulated Flow Frequency Curves
  - Stage Frequency Curves
- Partial Duration Series *(peaks over threshold)*