Statistical Primer – Analyzing Flow Time Series

Hydrologic Analysis for Ecosystem Restoration

Lecture 1.3

March 2022

1

Summarizing & Evaluating Time-Series

- One task in restoration analysis is looking at a time-series of flow and evaluating how successful it would be in meeting restoration goals
  - "success" is based on relationships developed between flow thresholds, their frequency, and the system ecology
- Need to **compare** different strategies based on statistical summaries of the flow time-series that results from them
- Make probability statements about future flows based on historical record as it was, and modified by strategy

2

2

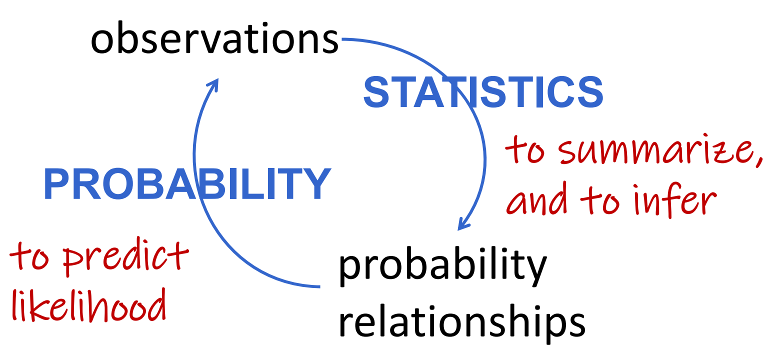
Objective

- Discuss mathematical tools for summarizing data and for inferring probability
- Consider statistical measures of streamflow that are useful for restoration analysis
  - Consider high flow and low flow
  - Consider both independent and persistent data
- Discuss the computation and use of these measures

3

3

A relationship between Probability and Statistics



4

4

Topics

- Probability
- Flow Frequency Curves — annual probability
  - Empirical (Graphical) Distributions
  - Analytical Distributions — less necessary because we don't extrapolate
- Volume-~~Duration~~ Frequency Curves — annual probability
- Flow *Duration* Curves — daily probability

5

5

Probability

Describes the likelihood of occurrence of future observations of a Random Variable

won't happen $0 \leq$ Prob [occurrence] ≤ 1 will happen

$0\% \leq$ Prob [occurrence] $\leq 100\%$

What's a Random Variable...?

... a number you don't know yet,
whose value varies randomly in time

6

6

Flow Probability (or, Frequency)

Typical Question:

What is the probability that streamflow will exceed (**or not exceed**) a particular magnitude during some period of time?

will be less than

(or, how often will...?)

Examples: What is the chance that a daily flow in the fall will be less than 100 cfs? (or, how often will...?)

What is the chance that the annual maximum flow will exceed 1000 cfs? (or, how often does...?)

7

7

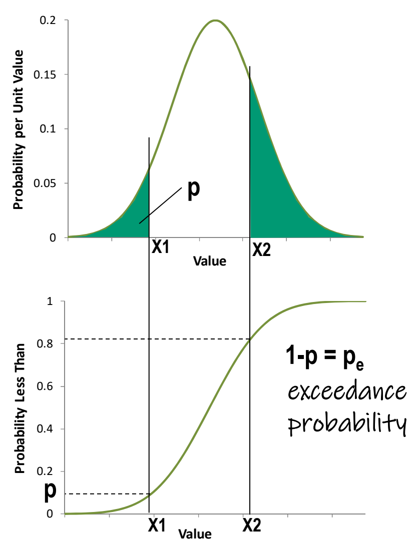
Definition of PDF and CDF

A Probability Density Function (PDF), f(x), defines the probability of occurrence for a continuous random variable.

area under curve = probability

The Cumulative Distribution Function (CDF), F(x) is the probability the random variable is less than some value

curve = probability



8

8

Estimating Probability from Observations

→ this is "Statistics"

Estimate the probability of flow levels from the relative frequency of flows observed *(empirical)*

observed frequency

$$\text{Relative Frequency} = \frac{\#\ \text{successes}}{\#\ \text{trials}}$$

estimates probability when observations are independent

success = value within the interval of interest
# trials = size of sample (independent observations)

$$\text{Relative Frequency} = \frac{\text{count in range}}{\text{total count}}$$

9

9

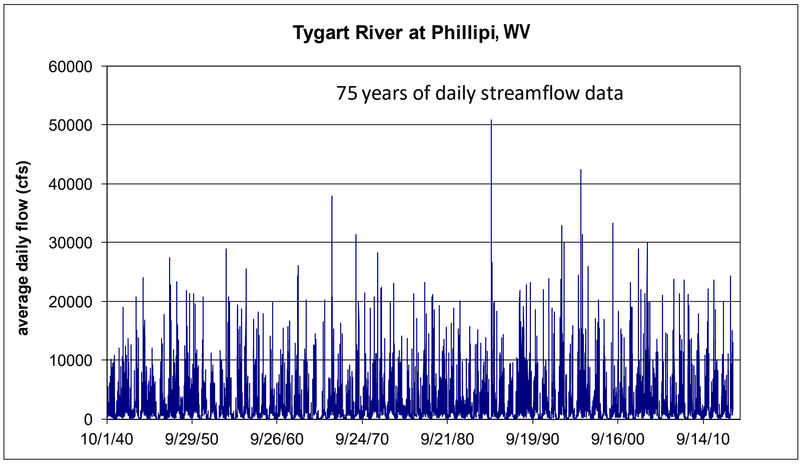
Topics

- Probability
- Flow Frequency Curves annual probability
  - Empirical (Graphical) Distributions
  - Analytical Distributions *less necessary because we don't extrapolate*
- Volume-~~Duration~~ Frequency Curves annual probability
- Flow *Duration* Curves daily probability

10

10

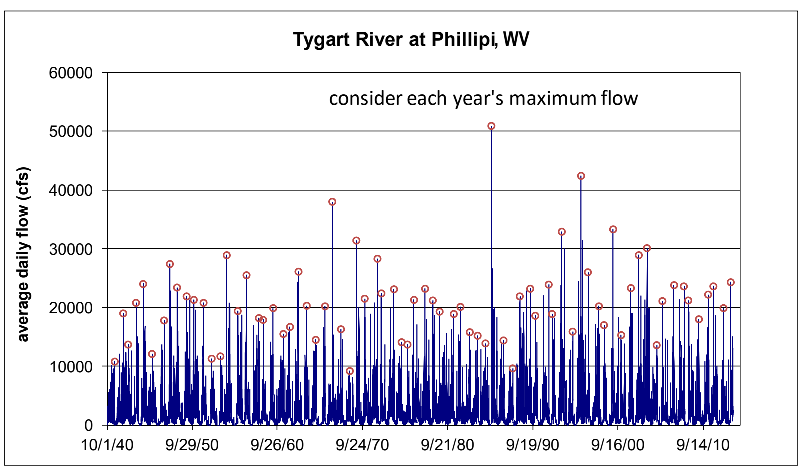
Streamflow gage data, daily flow record



11

11

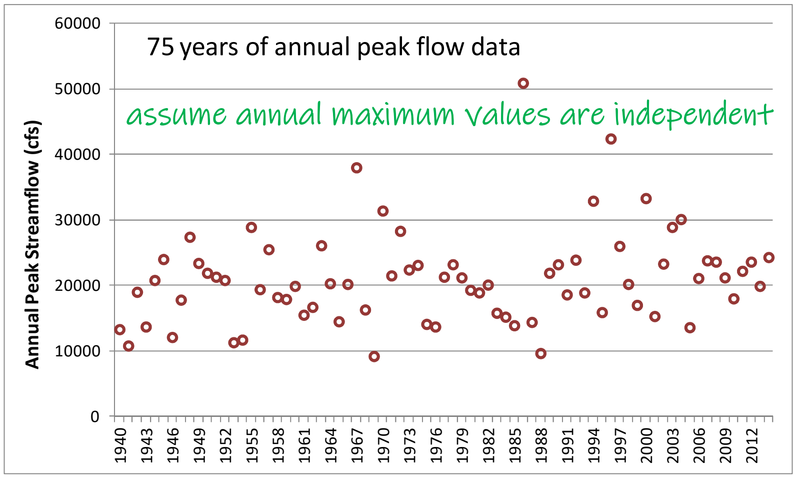
Streamflow gage data, find annual max



12

12

Tygart at Phillipi, ANNUAL MAX

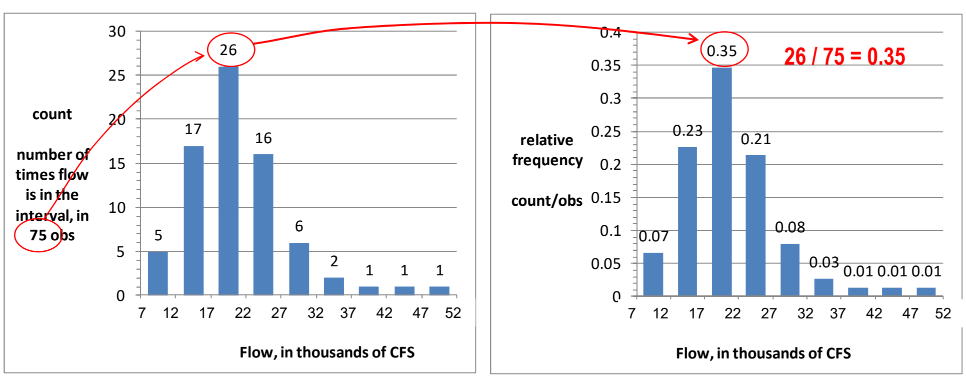


13

13

Histogram

divide data range into bins, count number of observations in each bin (count), divide by total (relative frequency)



this shape is similar to a PDF, probability density function

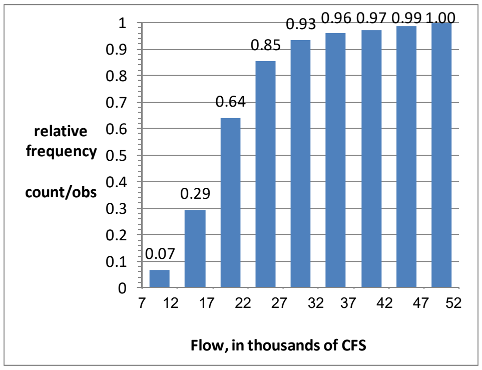
14

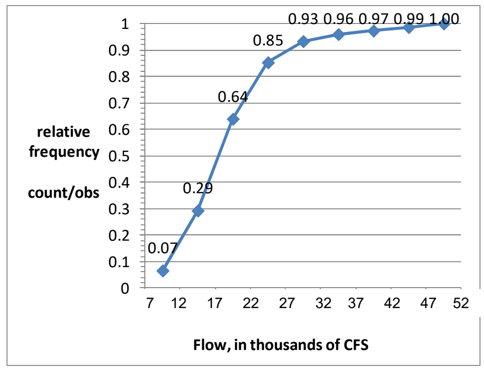
14

Cumulative Histogram

Sum the relative frequencies to get a cumulative frequency

(how many are <u>less than or equal to</u> a given value)





this curve is a CDF, cumulative distribution function

15

15

“relying on observation alone”

Empirical Probability Distributions

- Generalize *histogram* approach to <u>plotting position</u> (or empirical non-exceedance probability)
  - Size the bins to contain <u>one</u> observation each -- not equal
  - Gives estimate of non-exceedance probability of each observation

$$P[X \leq x_i] = \frac{m_i}{N}$$

m_i = rank of ordered observation i
(with m = 1 as smallest value)
N = total # of observations

ie, we estimate probability that will be smaller than observation x_i
by the fraction of the observed sample that <u>is</u> smaller

16

16

Empirical Probability Distributions

plotting position

Plotting Position:

General Formula

$$p = \frac{m + a}{N + b}$$

Weibull: a = 0.0, b = 1.0

Median: a = -0.3, b = 0.4

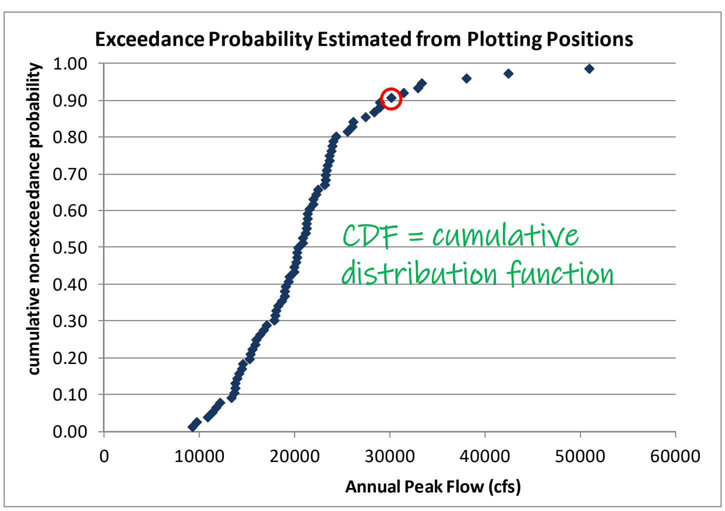
Weibull $P = \frac{m}{N + 1}$

Year	Rank	Flow	Non-Excdnc Probability	Year	Rank	Flow	Non-Excdnc Probability
1986	75	50900	0.99	1965	14	14500	0.18
1996	74	42400	0.97	1987	13	14400	0.17
1967	73	38000	0.96	1975	12	14100	0.16
2000	72	33300	0.95	1985	11	13900	0.14
1994	71	32900	0.93	1943	10	13700	0.13
1970	70	31400	0.92	1943	9	13700	0.12
2004	69	30100	0.91	2005	8	13600	0.11
1955	68	28900	0.89	1940	7	13300	0.09
1955	67	28900	0.88	1946	6	12100	0.08
1972	66	28300	0.87	1954	5	11700	0.07
1948	65	27400	0.86	1953	4	11300	0.05
1963	64	26100	0.84	1941	3	10800	0.04
1997	63	26000	0.83	1988	2	9650	0.03
1957	62	25500	0.82	1969	1	9200	0.01

17

...also called a frequency curve...

Empirical (Graphical) Distribution

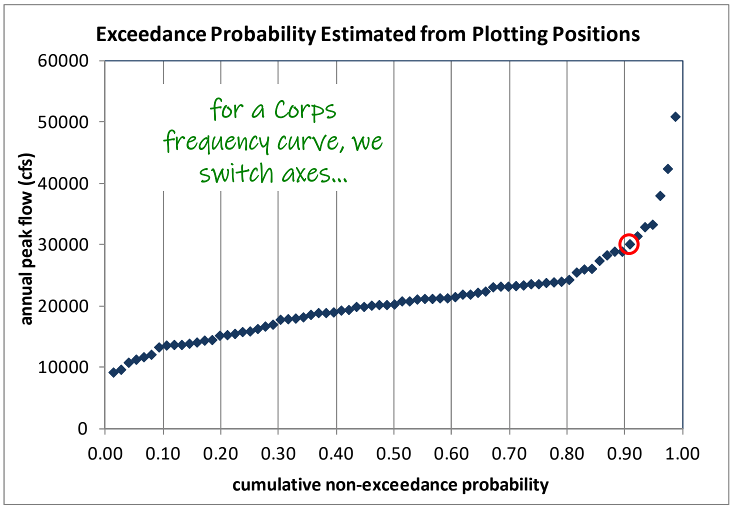


18

18

...also called a frequency curve...

Empirical (Graphical) Distribution

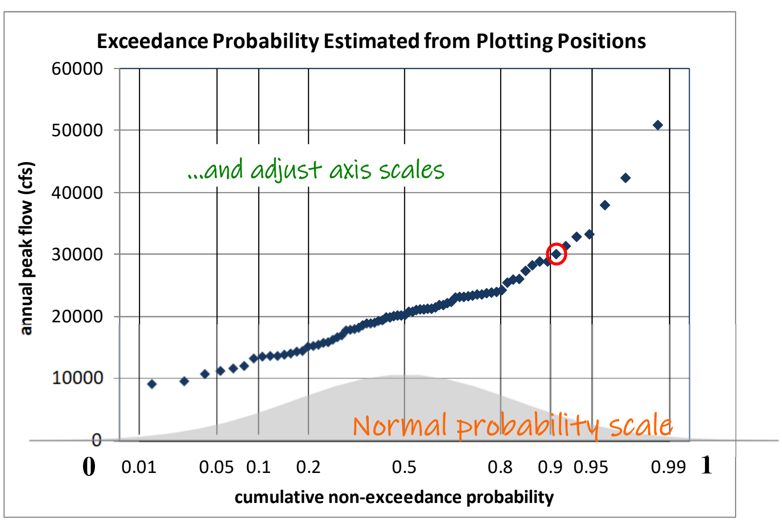


19

19

...also called a frequency curve...

Empirical (Graphical) Distribution

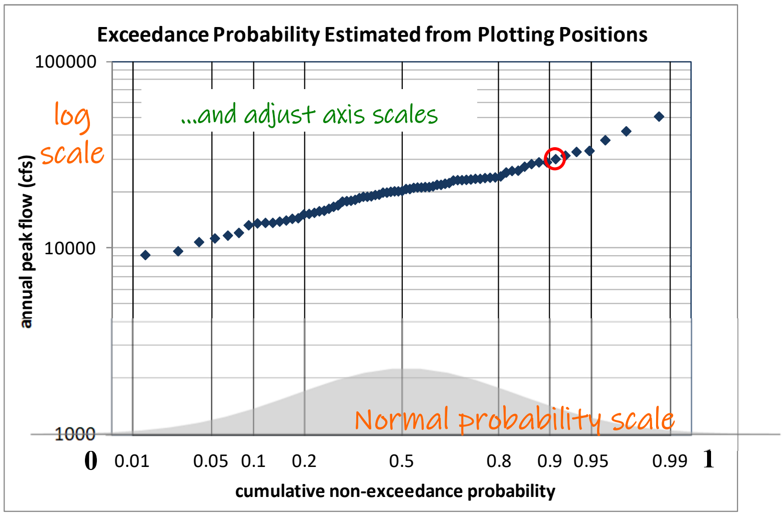


20

20

...also called a frequency curve...

Empirical (Graphical) Distribution

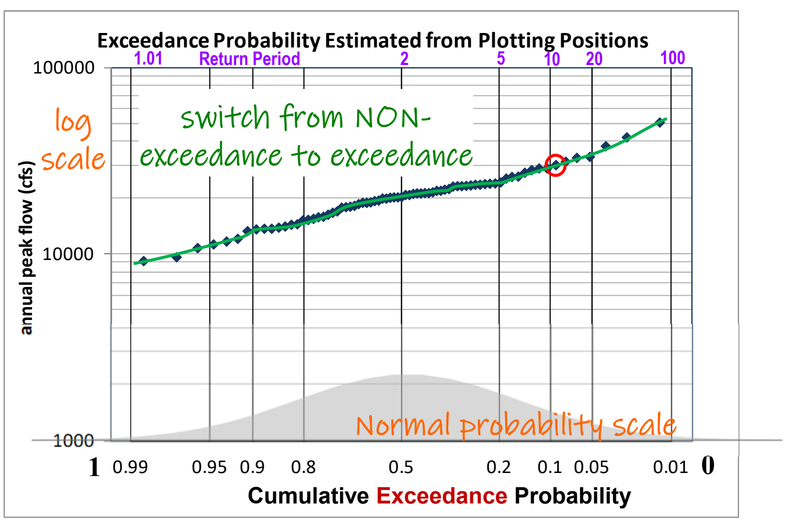


21

21

...also called a frequency curve...

Empirical (Graphical) Distribution



How often might we exceed 40,000 cfs?

22

22

Topics

- Probability
- Flow Frequency Curves annual probability
  - Empirical (Graphical) Distributions
  - Analytical Distributions less necessary because we don't extrapolate
- Volume-~~Duration~~ Frequency Curves annual probability
- Flow *Duration* Curves daily probability

23

23

Analytical Probability Distributions

- Some analytical relationship between probability and flow

$$P = aQ^b$$

a and b are distribution parameters

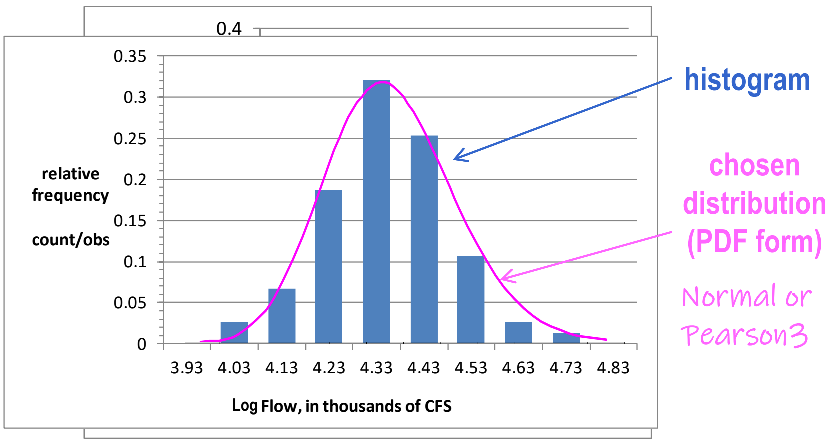
- Common Distributions
  - Normal
  - Log-Normal (ie, the logs of the values fit Normal)
  - Log-Pearson III

24

24

Fitting a Probability Distribution

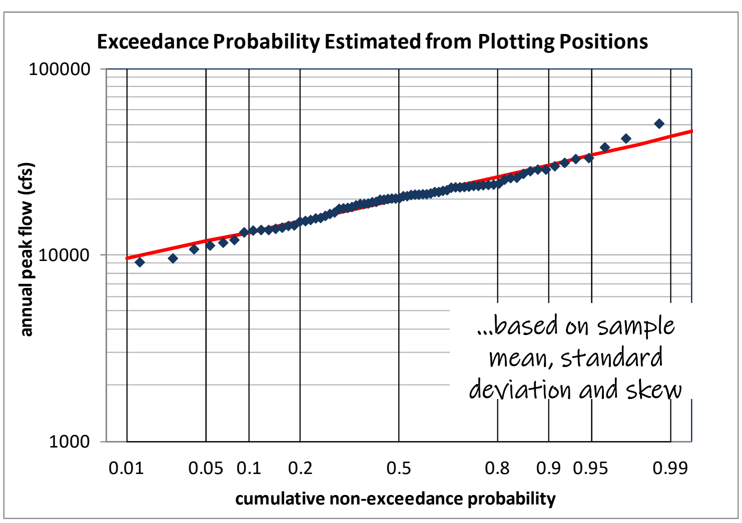
based on log(flow)



PDF = probability density function

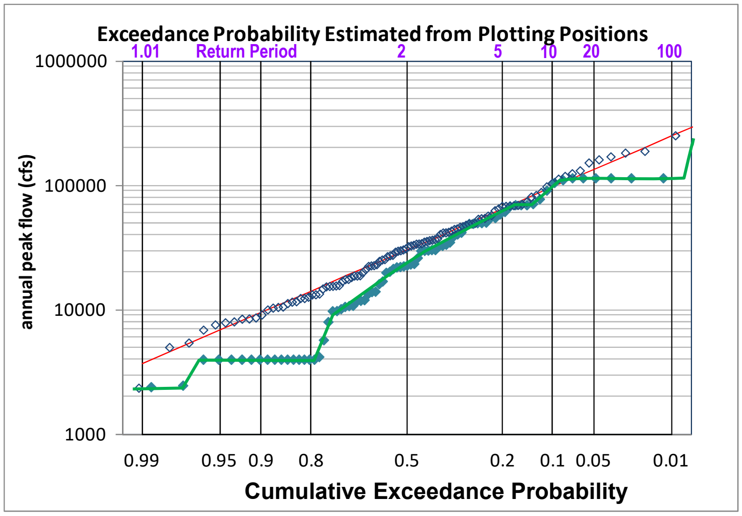
25

Tygart @ Phillipi Log Pearson III Distribution



26

What about regulated flows?



27

27

Alternative Probability Axis Labels

annual

non-exceedance probability = cumulative probability

exceedance probability = 1 – cumulative probability

%-chance exceedance = 100 * exceedance probability

exceedance frequency = 100 * exceedance probability

return period = 1 / exceedance probability

recurrence interval = return period

28

28

Topics

- Probability
- Flow Frequency Curves — annual probability
  - Empirical (Graphical) Distributions
  - Analytical Distributions
- Volume-~~Duration~~ Frequency Curves — annual probability
- Flow *Duration* Curves — daily probability

29

29

Volume-Duration Frequency Curves → VDF

Definition

Probability that the **X-day** volume or average flow will exceed (or not exceed) a particular level

(or, how often will...?)

Example:

Find the 7-day/10-year low-flow

the 7-consecutive-day volume (or average flow) that has 10% annual non-exceedance probability

30

30

Estimating VDF curves

- Obtain the maximum or minimum X-day consecutive flow volumes (or average flow) for each year, extracted from X-day moving volume (or average)

  - these annual values are independent, so can perform a standard frequency analysis

- Compute plotting positions and distribution parameters from period-of-record sample

31

31

Annual Maximum Flow of Various Durations

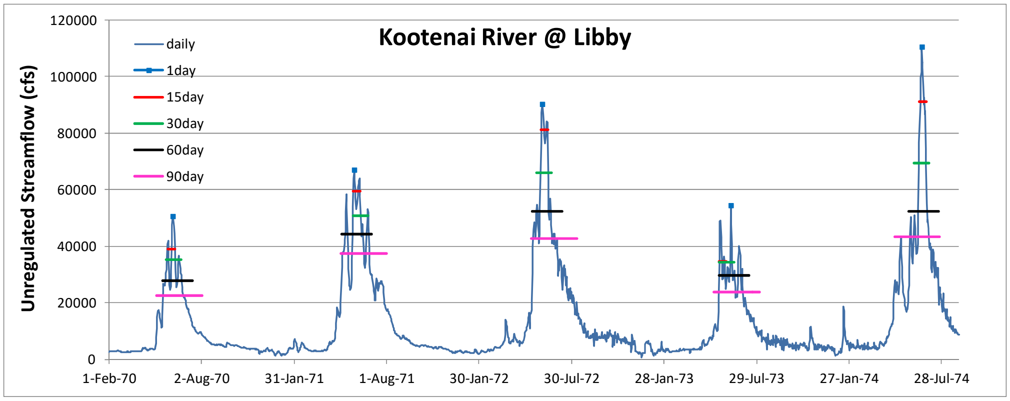
	highest mean value for duration, daily average cfs					
year	1-day		3-day		5-day	
	date	avg flow	date	avg flow	date	avg flow
1941	8/27/1941	10800	4/7/1941	8627	4/10/1941	5444
1942	8/24/1942	19000	8/25/1942	12077	8/28/1942	6967
1943	3/20/1943	13700	12/31/1942	11333	12/30/1942	8897
1944	2/23/1944	20800	2/25/1944	13337	2/25/1944	7901
1945	12/26/1944	24000	12/28/1944	18900	1/1/1945	13570
1946	1/8/1946	12100	1/9/1946	9450	1/11/1946	6200
1947	6/9/1947	17800	6/10/1947	11137	3/31/1947	6179
1948	2/14/1948	27400	2/15/1948	18847	2/19/1948	11229
1949	12/16/1948	23400	12/18/1948	18143	1/29/1949	11316
1950	2/1/1950	21900	2/2/1950	19400	2/5/1950	13043
1951	12/4/1950	21300	12/5/1950	16800	12/9/1950	11414
1952	1/28/1952	20800	1/29/1952	16133	2/1/1952	8869
1953	1/9/1953	11300	1/10/1953	8337	1/14/1953	5147
1954	1/17/1954	11700	1/18/1954	9460	3/17/1954	6137
1955	10/16/1954	28900	10/17/1954	16730	3/7/1955	11816
1956	1/30/1956	19400	8/8/1956	15447	2/5/1956	13387
1957	2/10/1957	25500	2/12/1957	16320	2/4/1957	11766
1958	7/24/1958	18200	7/25/1958	15667	7/29/1958	10013
1959	1/22/1959	17900	1/23/1959	13933	1/23/1959	7940
1960	3/31/1960	19900	4/1/1960	18033	4/4/1960	13083
1961	6/10/1961	15500	6/11/1961	11407	2/20/1961	8373
1962	3/22/1962	16700	3/1/1962	11457	3/3/1962	7710
1963	3/20/1963	26100	3/7/1963	21167	3/8/1963	12963
1964	3/5/1964	20300	3/6/1964	14933	3/9/1964	9911
1965	3/26/1965	14500	3/27/1965	12300	3/30/1965	8123
1966	5/1/1966	20200	5/3/1966	13273	2/16/1966	9860
1967	3/7/1967	38000	3/8/1967	28267	3/9/1967	16029

X-day average flows

32

32

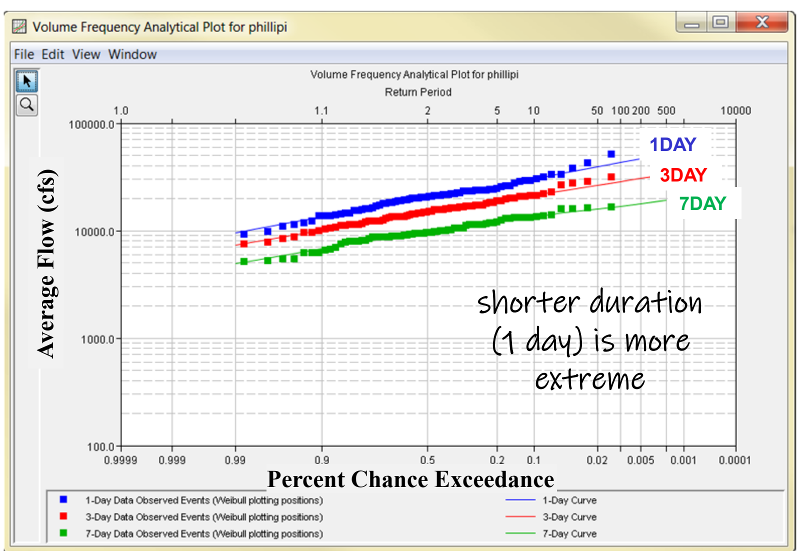
Longer Duration Max Average Flows



33

33

Example High Flow VDF Curves

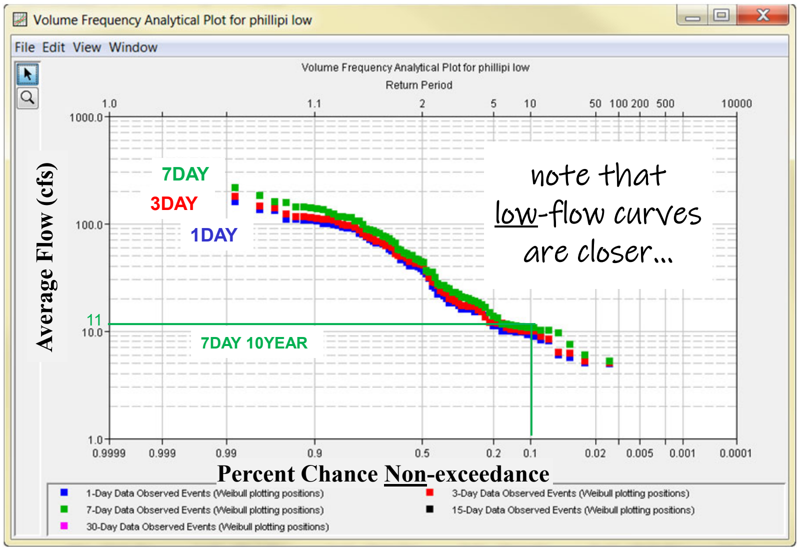


average flows can be plotted closer together than volumes

34

34

Example Low Flow VDF Curves



35

35

High vs Low Flows

- Usually, can fit an analytical distribution to *unregulated* high flows
  - But not to *regulated* high flows
- Low flows can be more non-linear (effects of groundwater, etc), and might only allow graphical curve (empirical distribution)
  - Unreported diversions cause problems in low-flow analysis

36

36

Topics

- Probability
- Flow Frequency Curves   annual probability
  - Empirical (Graphical) Distributions
  - Analytical Distributions
- Volume-~~*Duration*~~ Frequency Curves   annual probability
- Flow *Duration* Curves   daily probability

37

37

Flow/Stage Duration Analysis

Fraction of days exceeding a particular level, referred to as percent of time exceeded

OR

not the annual extreme...

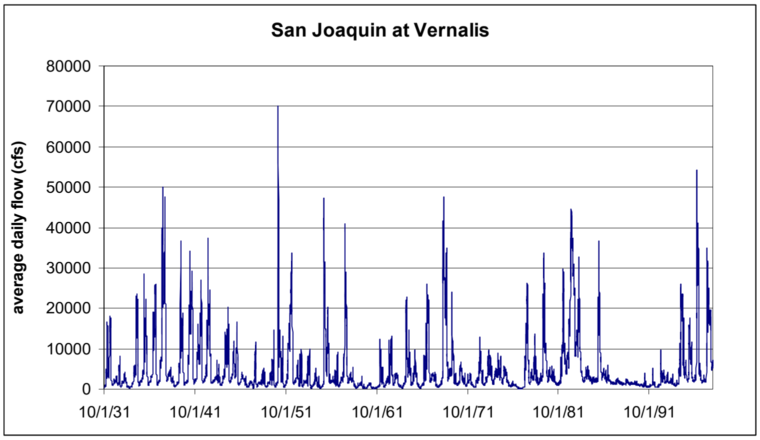
Estimate of the likelihood that a daily flow will exceed a particular level (refers to future, not soon...)

- Graphical display of stream characteristics
- Restoration, run-of-river hydropower, water supply

38

38

Starting with a daily flow record…

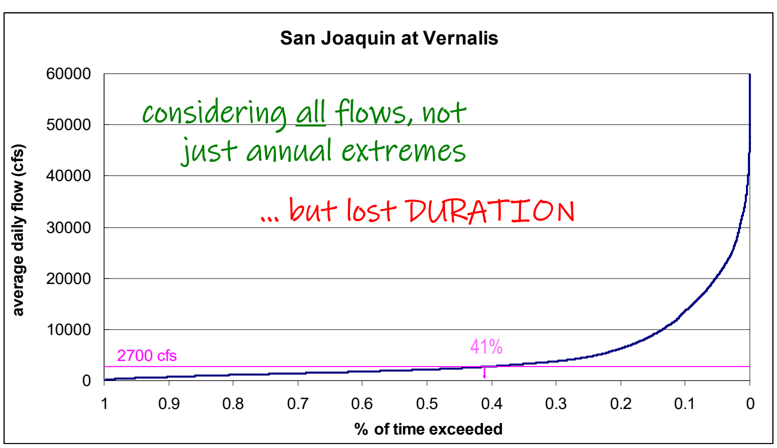


Question: How much of the time can we expect future flows to exceed 2700 cfs?

39

39

Order the data from high to low, and compute relative frequency of exceedance

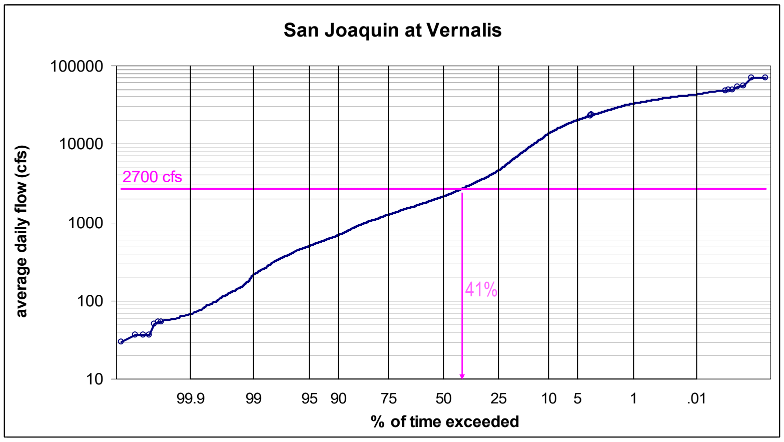


2700 cfs is exceeded 41% of the time

40

40

Same curve, different axis scales

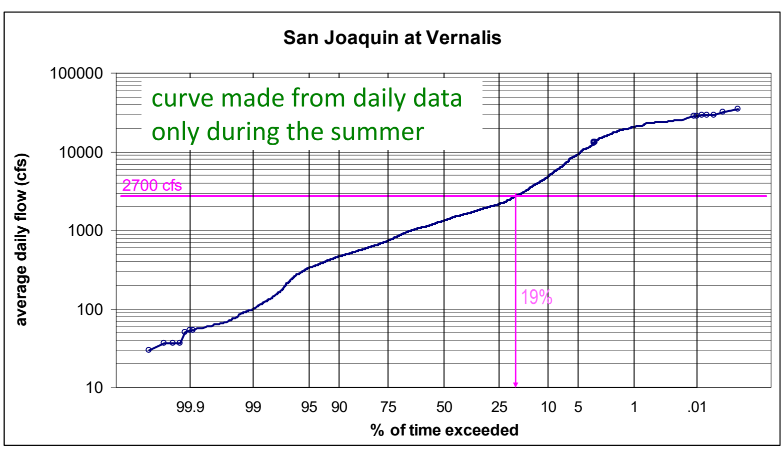


2700 cfs is exceeded 41% of the time

41

41

What if we're interested in the 2700 cfs threshold only during the summer?



2700 cfs is exceeded 19% of the time <u>in the summer</u>

42

42

Flow Duration Analysis Cautions

- Data points are not independent, so frequency does not approximate probability in the near term
  - approximates daily probability (not annual) in future
- Analytical distributions do not fit data
- Need sufficient record length (at least 10 years) to obtain 95% to 5% confidence intervals
- Estimate of distribution tails is not very reliable

43

43

Review

- Probability
- Flow Frequency Curves — annual probability
  - Empirical (Graphical) Distributions
  - Analytical Distributions
- Volume-~~Duration~~ Frequency Curves — annual probability
- Flow *Duration* Curves — daily probability

44

44