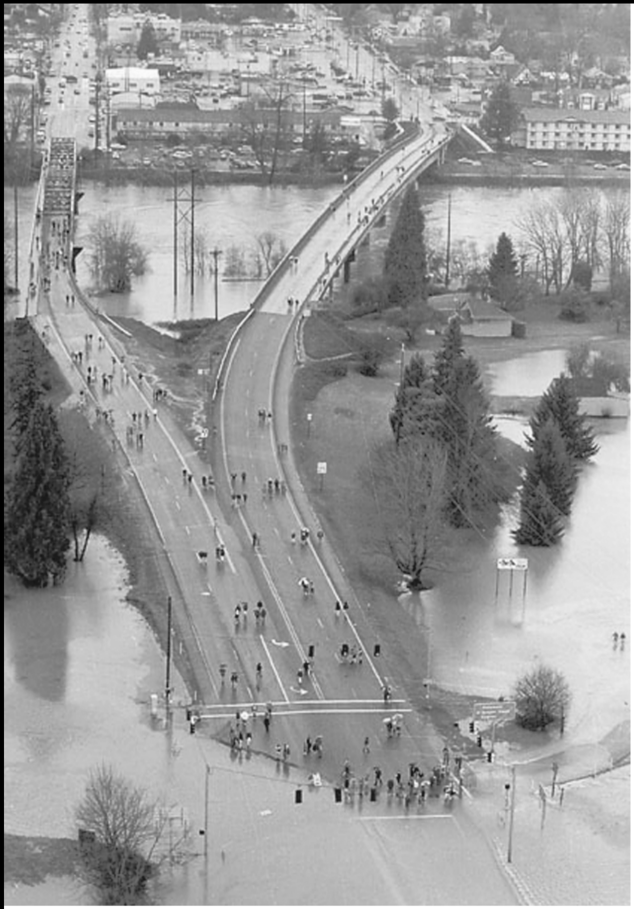
Statistics &

Flood Frequency
Chapter 3

Dr. Philip B. Bedient

Predicting FLOODS



Oregon flood scene

Houston flood scene (photo courtesy of Houston chronicle)

Flood Frequency Analysis

- Statistical Methods to evaluate probability exceeding a particular outcome - P (X >20,000 cfs) = 10%
- Used to determine return periods of rainfall or flows
- Used to determine specific frequency flows for floodplain mapping purposes (10, 25, 50, 100 yr)
- Used for datasets that have no obvious trends
- Used to statistically extend data sets

Random Variables

- Parameter that cannot be predicted with certainty
- Outcome of a random or uncertain process - flipping a coin or picking out a card from deck
- Can be discrete or continuous
- Data are usually discrete or quantized
- Usually easier to apply continuous distribution to discrete data that has been organized into bins

Typical CDF

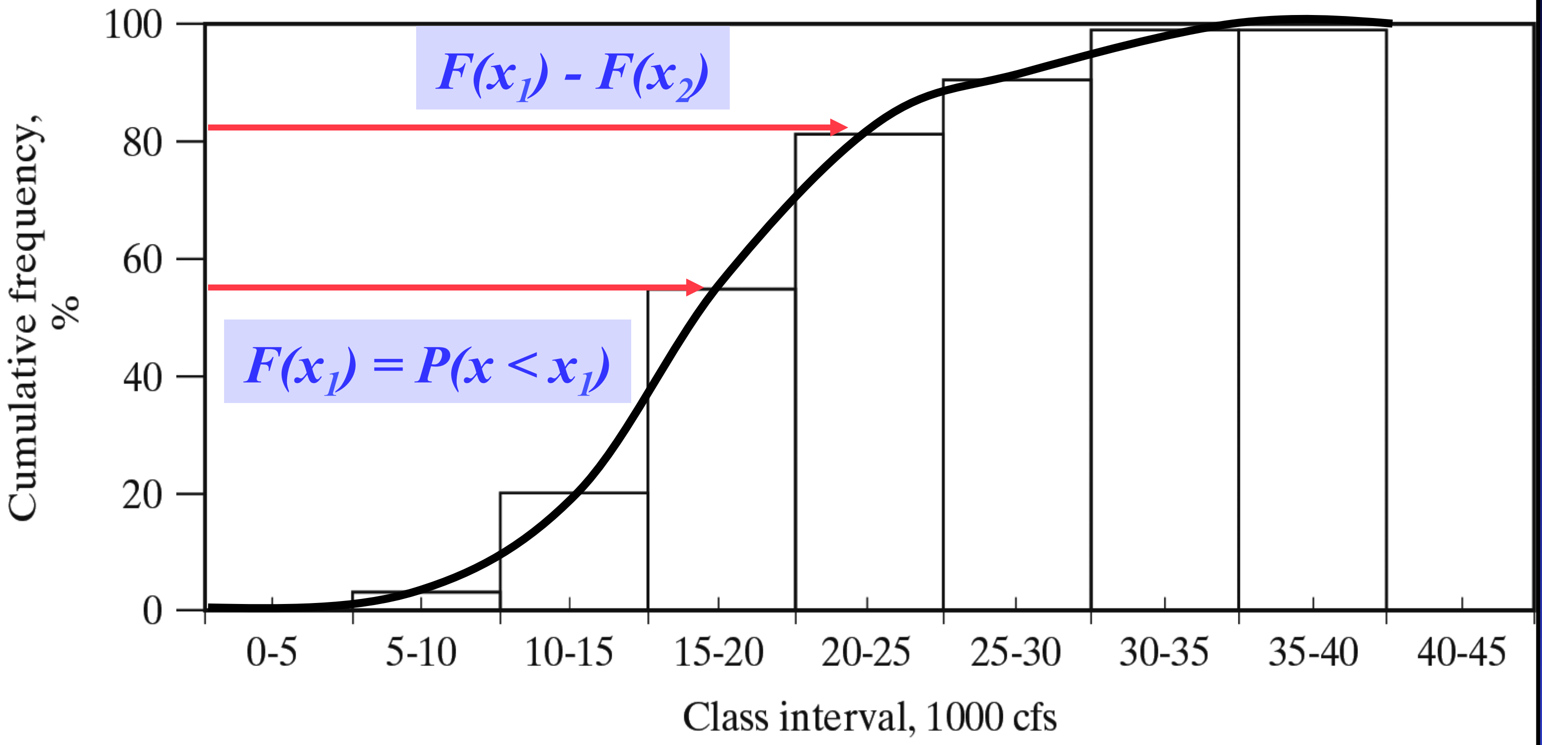


Figure 3.5

Cumulative frequency histogram for the Siletz River, plotted vs. class intervals.

Freq Histogram of Flows

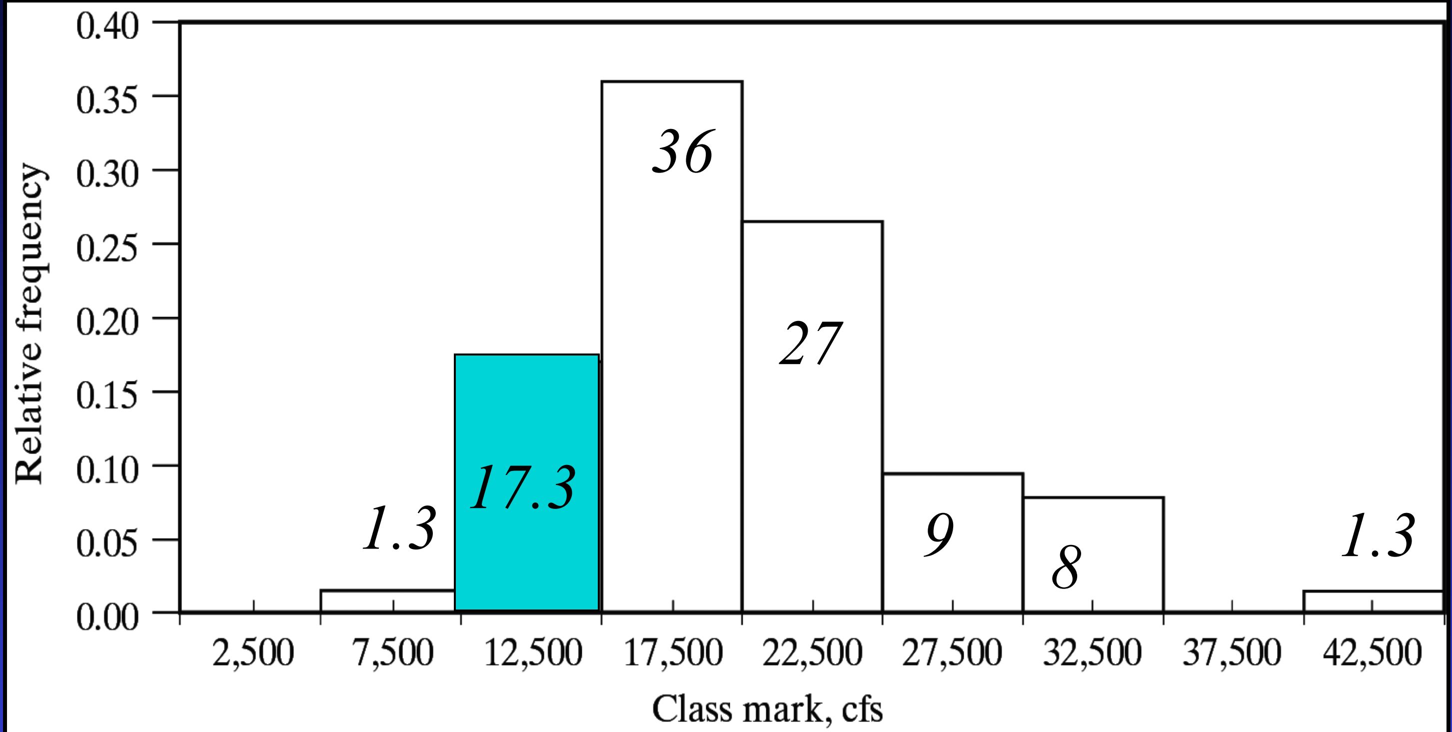


Figure 3.4

Relative frequencies (probabilities) for the Siletz River, plotted vs. their class mark.

Probability that Q is 10,000 to 15, 000 = 17.3%

Prob that Q < 20,000 = 1.3 + 17.3 + 36 = 54.6%

Probability Distributions

CDF is the most useful form for analysis

$$F(x) = P(X \le x) = \sum_i P(x_i)$$

$$F(x_1) = P(-\infty \le x \le x_1) = \int_{-\infty}^{x1} f(x)dx$$

$$P(x_1 \le x \le x_2) = F(x_2) - F(x_1)$$

Moments of a Distribution

Used to characterize a distribution or set of data

Moments taken about the origin (1st) or the mean (2nd, 3rd, etc)

Discrete $P(x_i)$

$$\mu_N^{'} = \sum_{-\infty}^{\infty} x_i^N P(x_i)$$

Continuous $f(x)$

$$\mu_N^{'} = \int_{-\infty}^{\infty} x^N f(x)\,dx$$

Moments of a Distribution

First Moment about the Origin - Mean

$$E(x) = \mu = \sum x_i P(x_i)$$

Discrete

$$E(x) = \mu = \int_{-\infty}^{\infty} x f(x) dx$$

Continuous

Var(x) = Variance
Second moment about mean

$$Var(x) = \sigma^2 = \sum_{-\infty}^{\infty} (x_i - \mu)^2 P(x_i)$$

$$Var(x) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

$$Var(x) = E(x^2) - (E(x))^2$$

$$cv = \frac{\sigma}{\mu} = \text{Coeff. of Variation}$$

Estimates of Moments from a Dataset

$$\bar{x} = \frac{1}{n}\sum_{i}^{n} x_i \Rightarrow \text{Mean of Data}$$

$$s_x^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 \Rightarrow \text{Variance}$$

Std Dev. $S_x = (S_x{}^2)^{1/2}$

Skewness Coefficient

Used to evaluate high or low data points - flood or drought data

$$\text{Skewness} \rightarrow \frac{\mu_3}{\sigma^3} \rightarrow \text{third central moment}$$

$$C_s = \frac{n}{(n-1)(n-2)} \frac{\sum (x_i - \bar{x})^3}{s_x^3} \quad \text{skewness coeff.}$$

$$\text{Coeff of Var} = \frac{\sigma}{\mu}$$

Mean, Median, Mode

- Positive Skew moves mean to right
- Negative Skew moves mean to left
- Normal Dist’n has mean = median = mode
- Median has highest prob. of occurrence

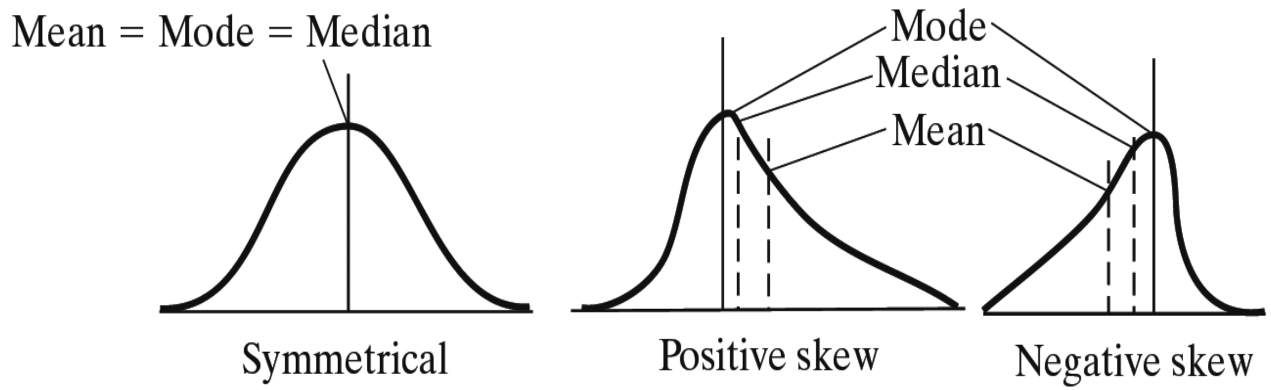


Figure 3.9

Effect of skewness on PDF and relative locations of mean, median, and mode. (From Haan, 1977, Fig. 3.3.).

Skewed PDF - Long Right Tail

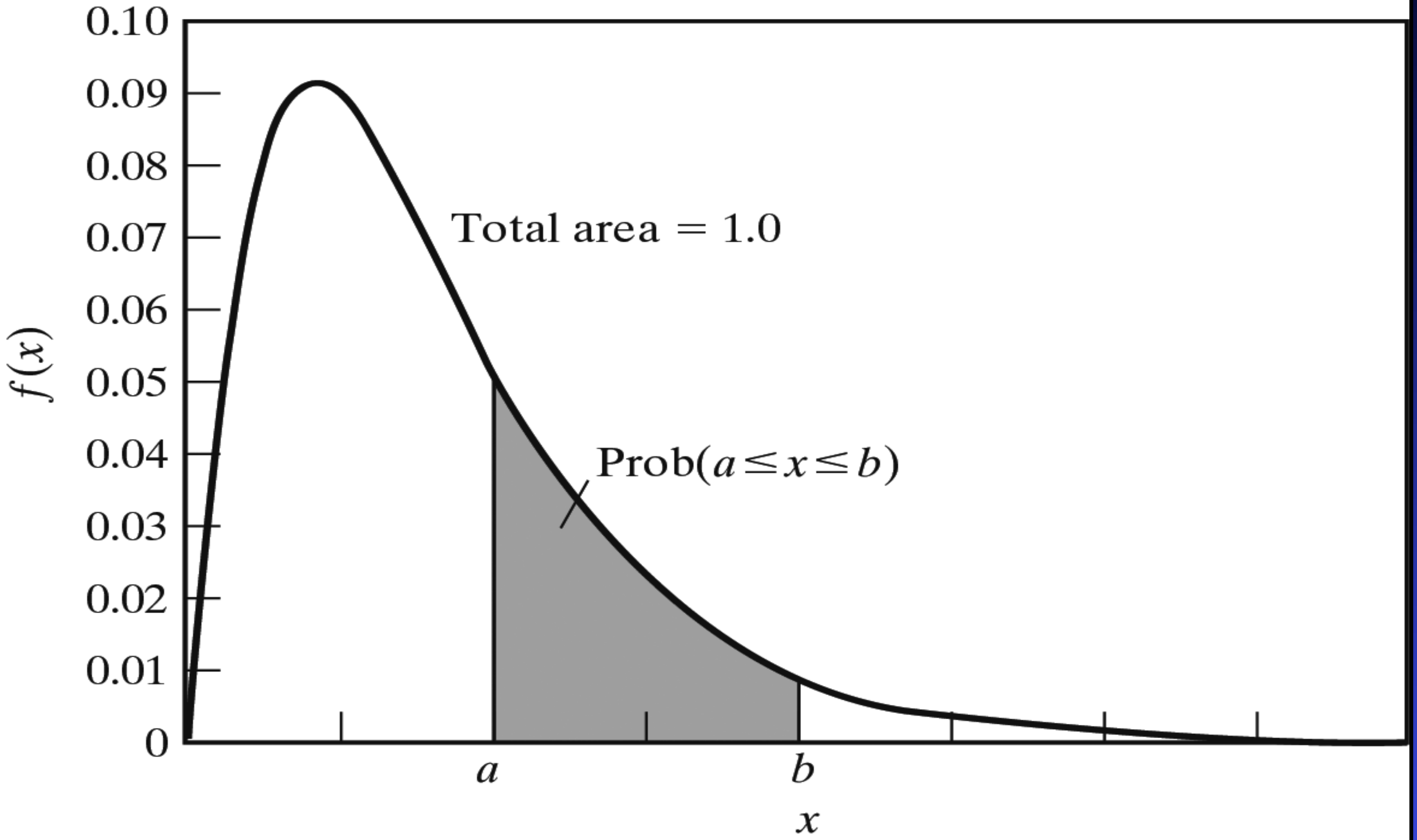
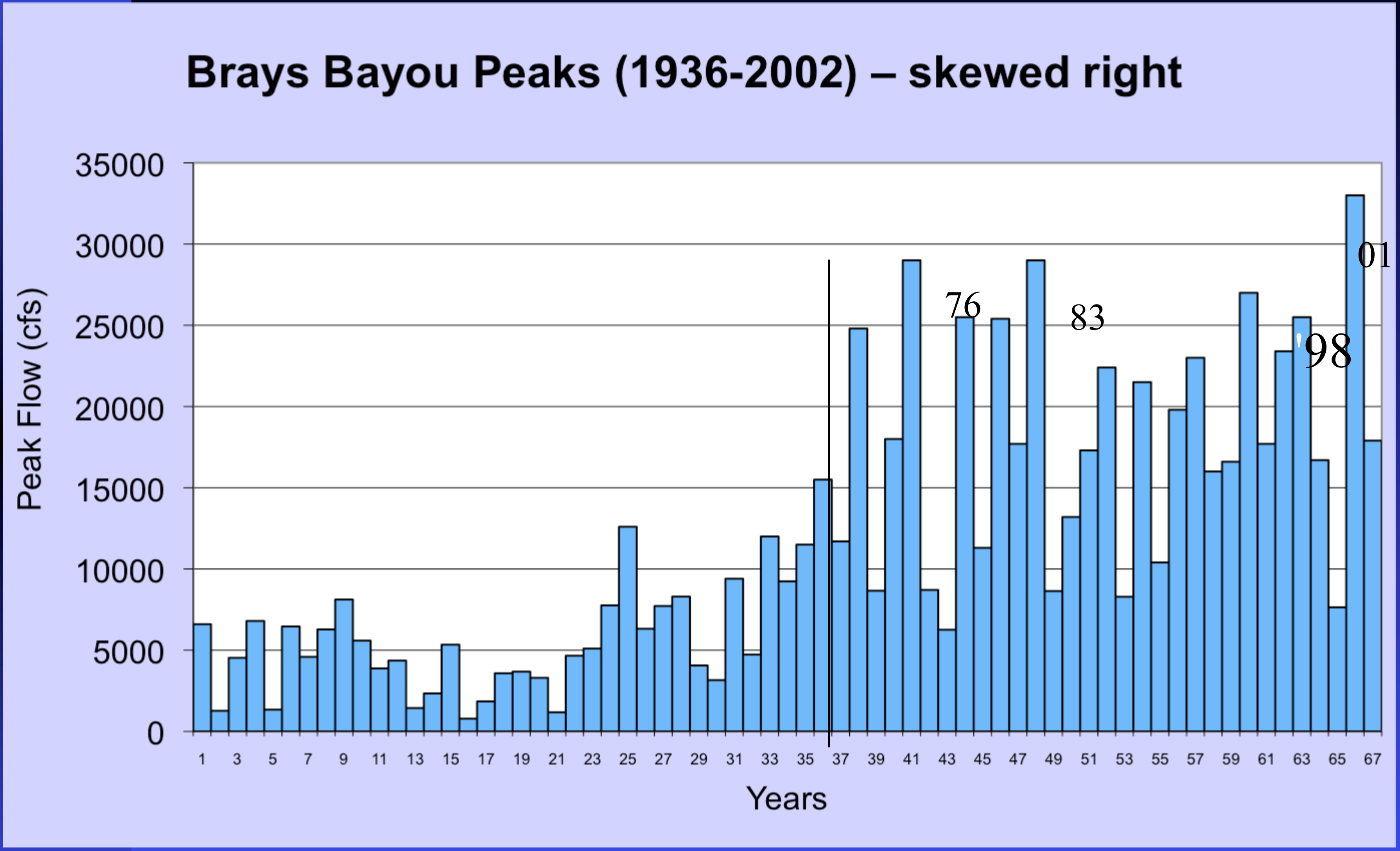
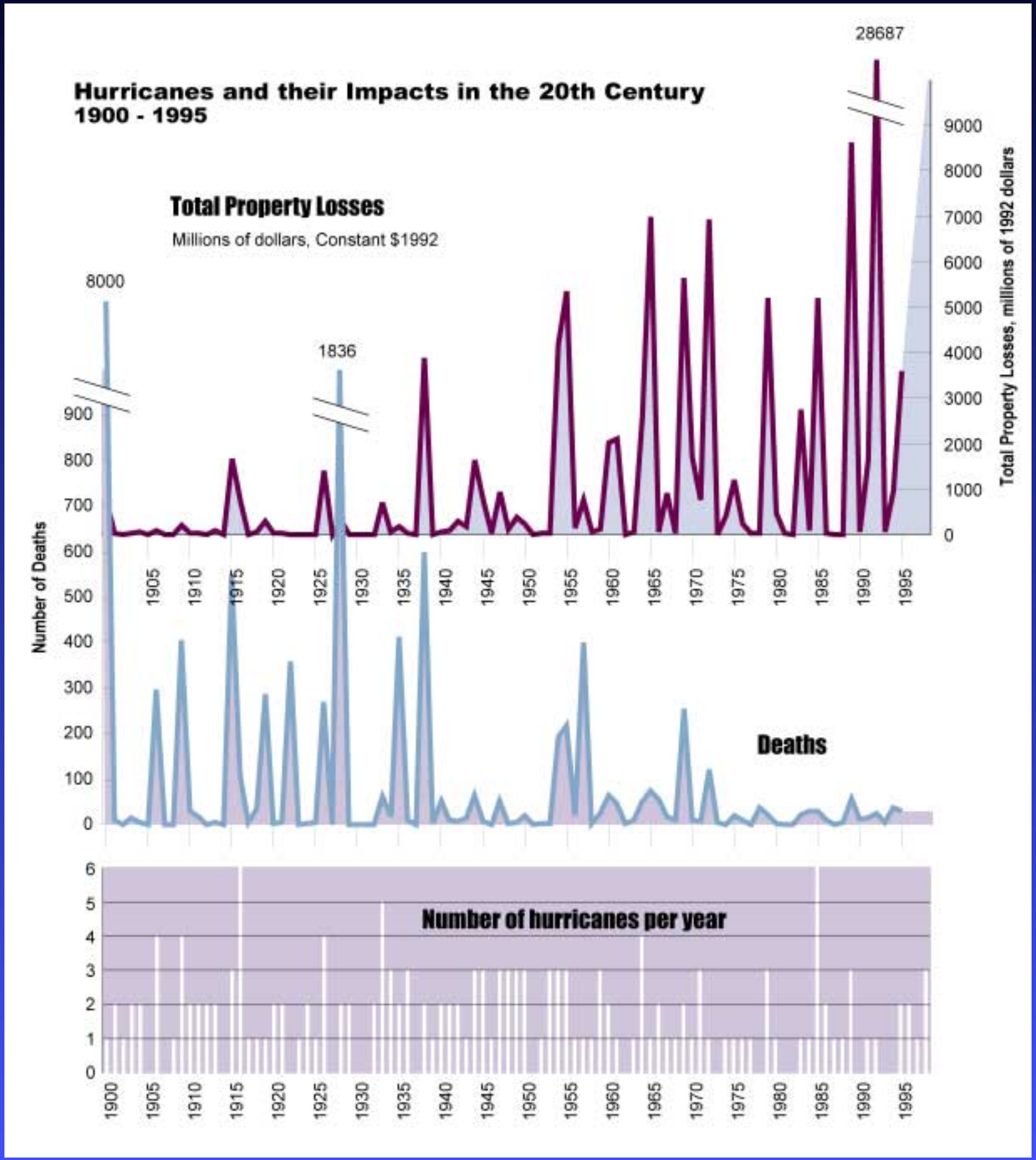


Figure 3.7

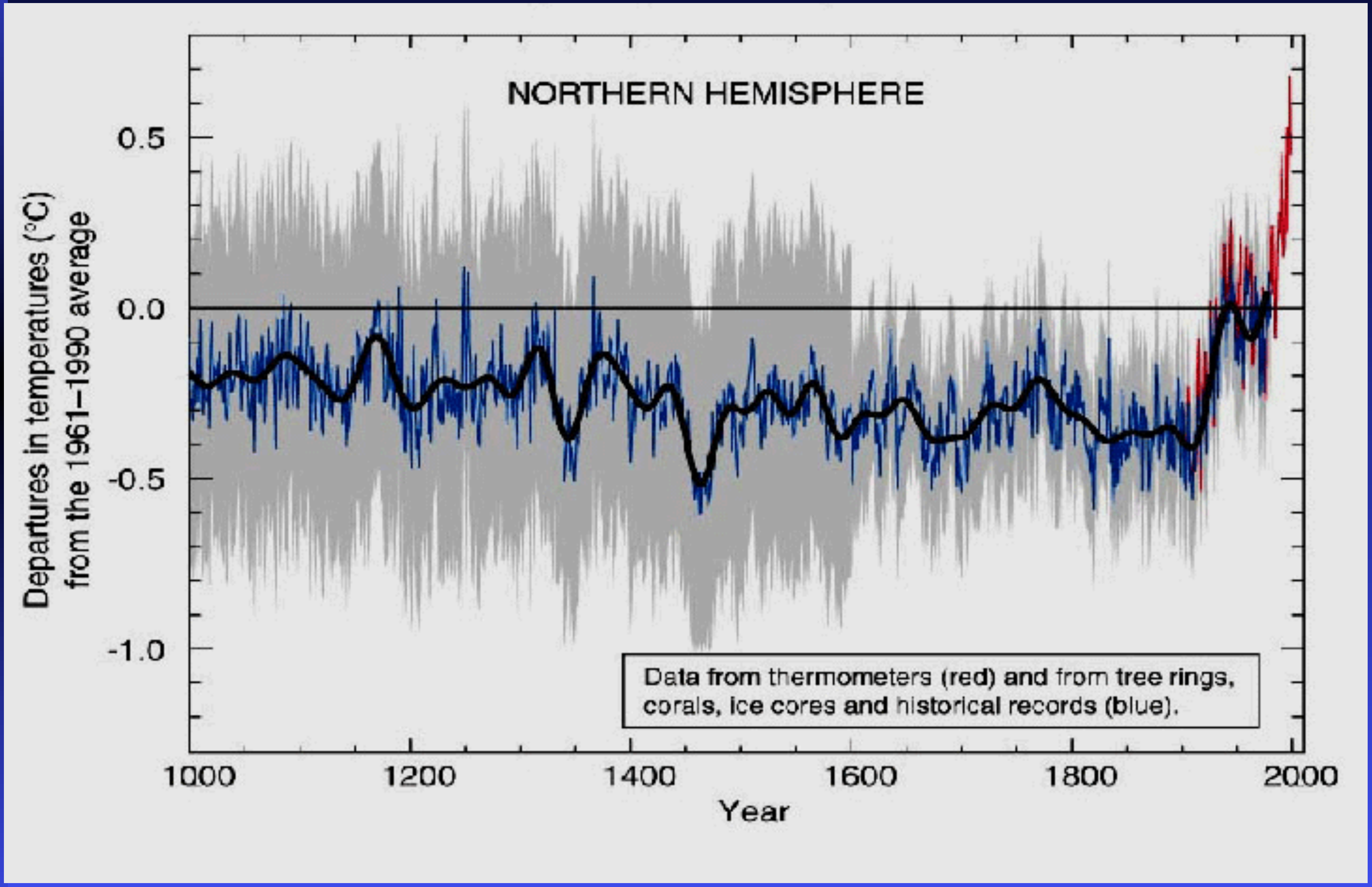
Continuous probability density function.


Brays Bayou Peaks (1936-2002) – skewed right
Peak Flow (cfs)
35000
30000
25000
20000
15000
10000
5000
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67
Years
76
83
'98
01

Skewed Data


Hurricanes and their Impacts in the 20th Century
1900 - 1995
Total Property Losses
Millions of dollars, Constant $1992
8000
1836
28687
Total Property Losses, millions of 1992 dollars
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Number of Deaths
0
100
200
300
400
500
600
700
800
900
Deaths
Number of hurricanes per year
0
1
2
3
4
5
6
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995

Climate Change Data



Siletz River Data

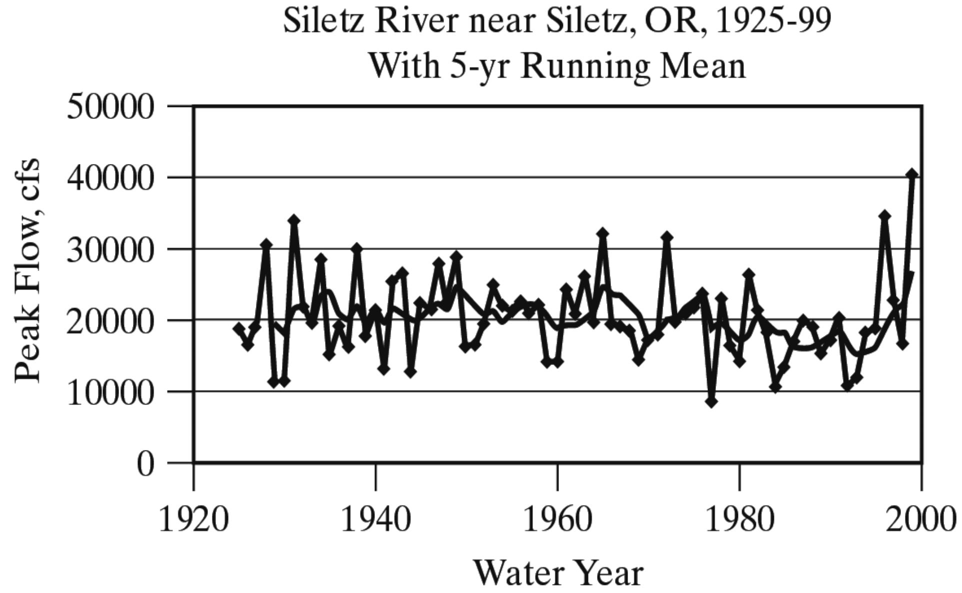


Figure 3.2

Time series of annual maximum peak flows for the Siletz River, near Siletz, Oregon. Also shown is the 5-yr running mean, from which longerterm trends can sometimes be discerned. Only quantitative methods of time series analysis can determine for sure whether or not there are periodicities or nonstationary components in the data, but none are obvious visually.

Stationary Data Showing No Obvious Trends

Data with Trends

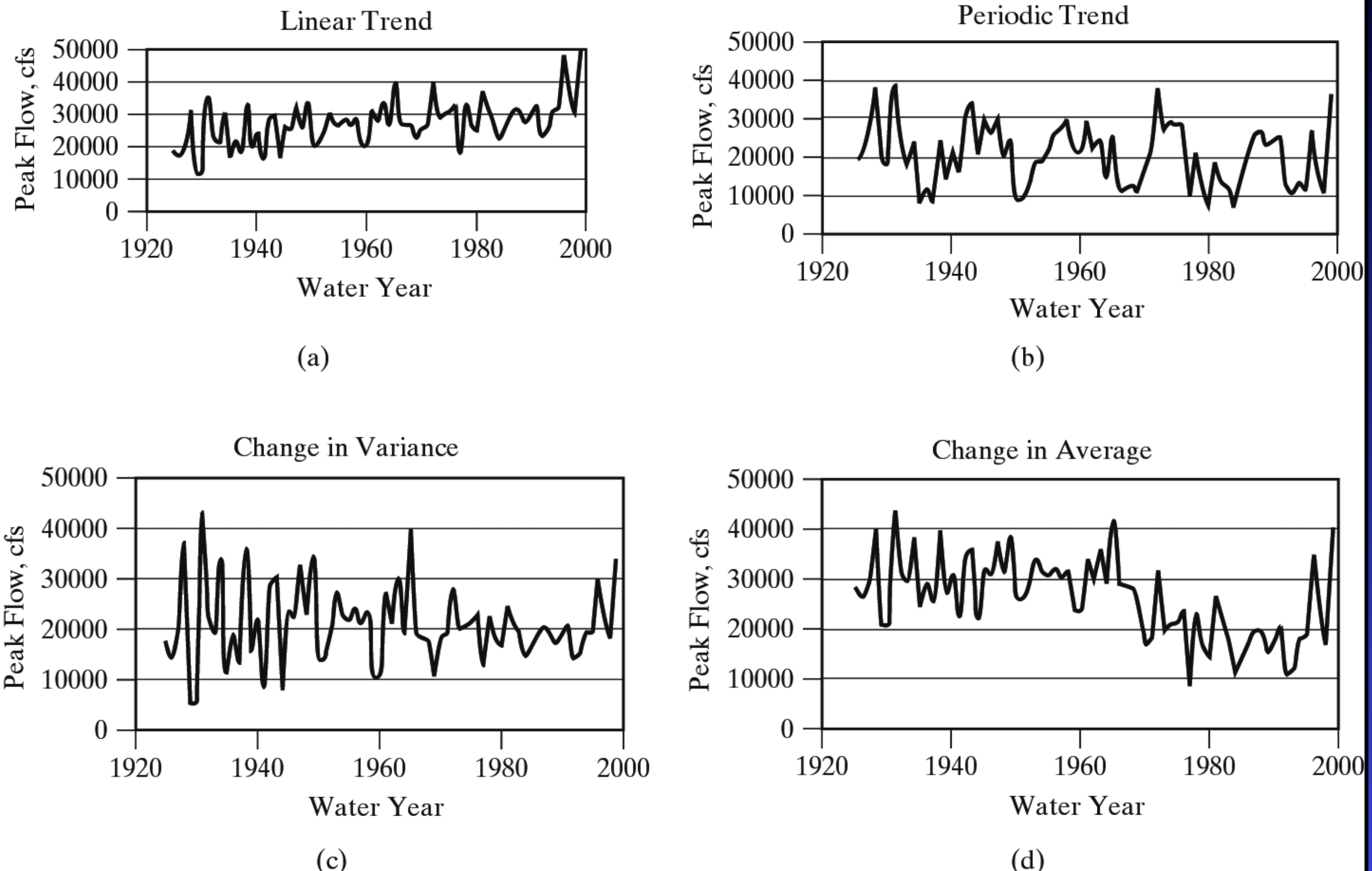


Figure 3.3

Hypothetical examples of nonstationary time series. (a) Linear trend. (b) Periodic trend. (c) Change in variance. (d) Change in average.

Frequency Histogram

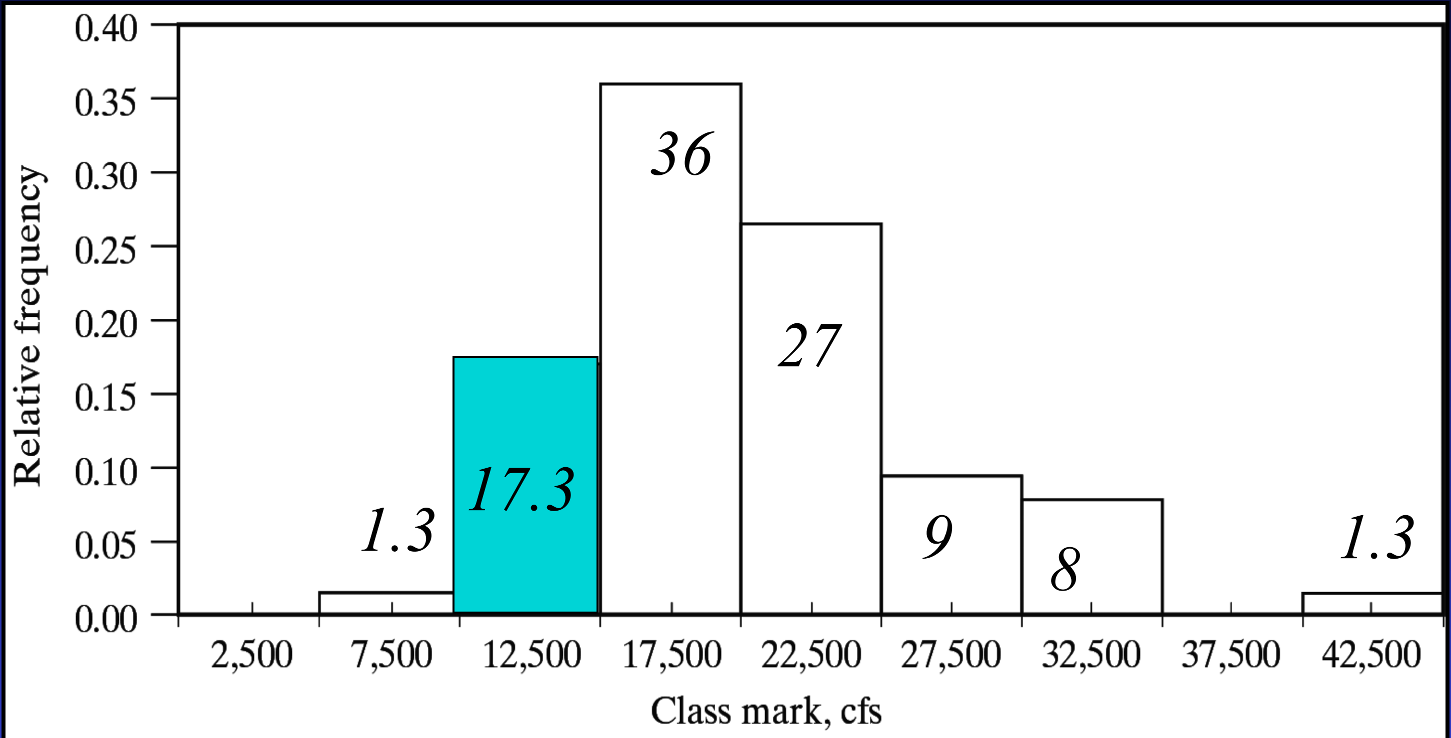


Figure 3.4

Relative frequencies (probabilities) for the Siletz River, plotted vs. their class mark.

Probability that Q is 10,000 to 15, 000 = 17.3%

Prob that Q < 20,000 = 1.3 + 17.3 + 36 = 54.6%

Cumulative Histogram

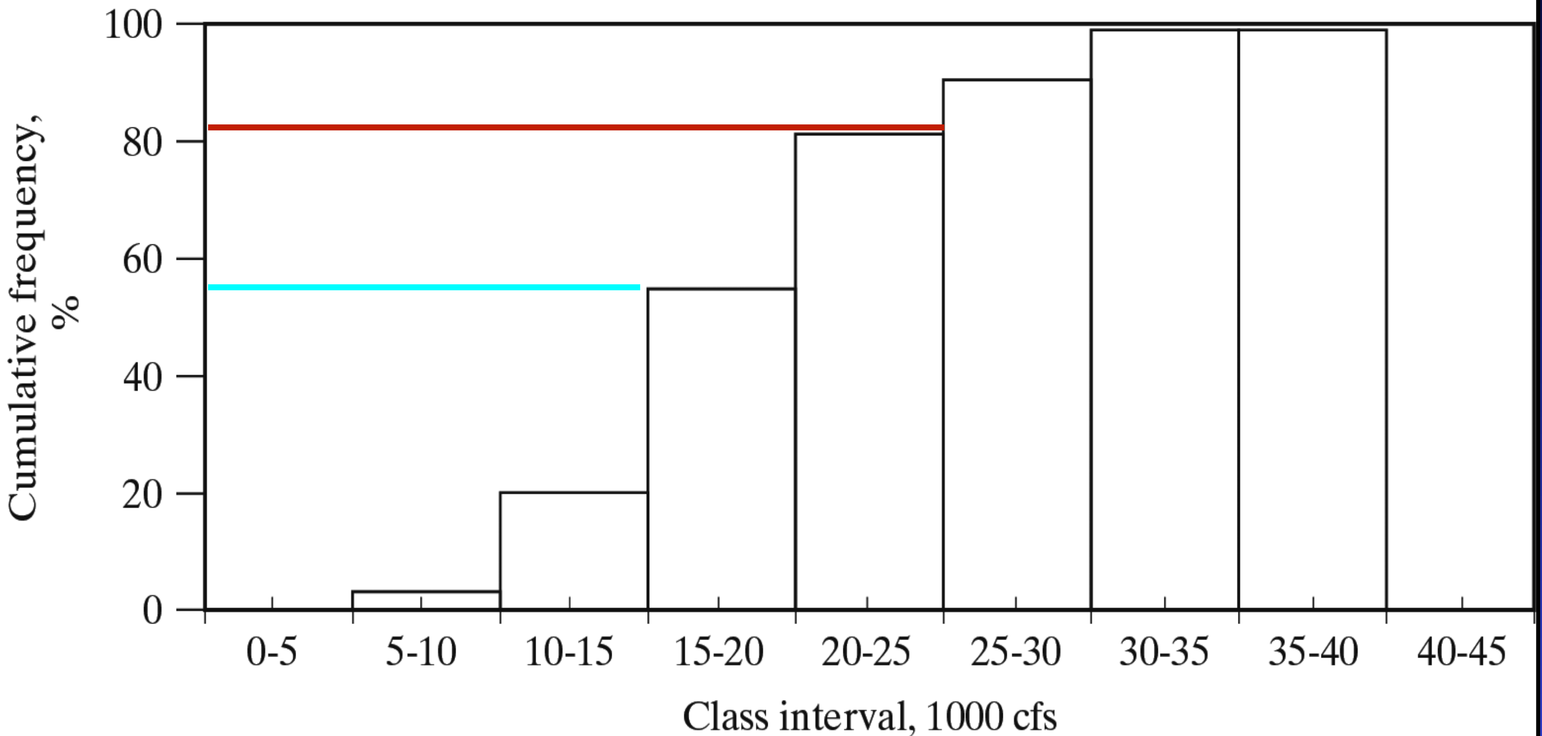


Figure 3.5

Cumulative frequency histogram for the Siletz River, plotted vs. class intervals.

Probability that Q < 20,000 is 54.6 %
Probability that Q > 25,000 is 19 %

PDF - Gamma Dist

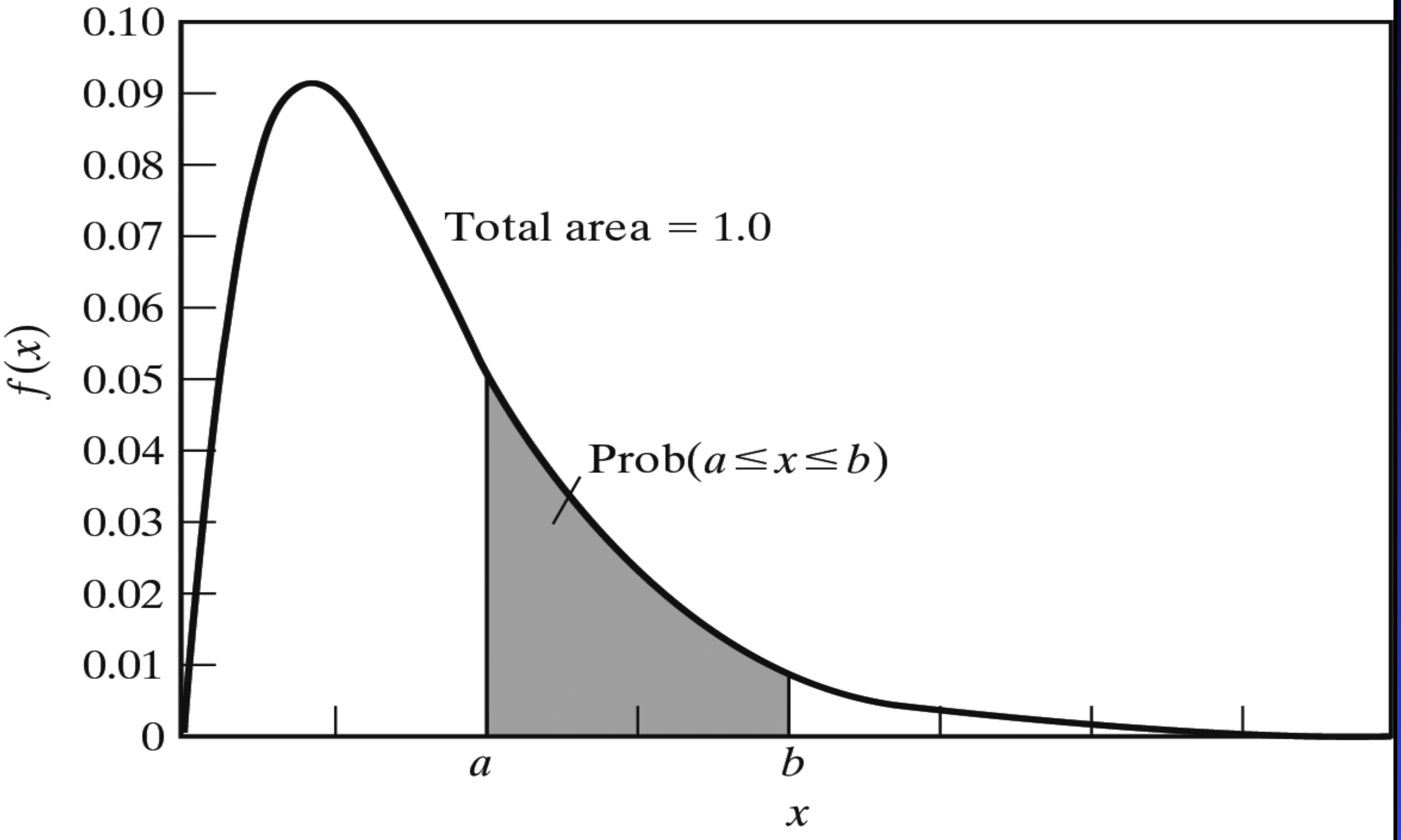


Figure 3.7

Continuous probability density function.

Major Distributions

- Binomial - P (x successes in n trials)
- Exponential - decays rapidly to low probability - event arrival times
- Normal - Symmetric based on μ and σ
- Lognormal - Log data are normally dist' d
- Gamma - skewed distribution - hydro data
- Log Pearson III -skewed logs -recommended by the IAC on water data - most often used

Binomial Distribution

The probability of getting x successes followed by n-x failures is the product of prob of n independent events:

$$p^x (1-p)^{n-x}$$

This could be used to represent the case of flooding – a success is exceeding a certain level while a failure is falling below that level in any given year. Thus, over a 25 year period, one would just add up the number of successes and the number of failures by year.

Binomial Distribution

The probability of getting x successes followed by n-x failures is the product of prob of n independent events: $p^x (1-p)^{n-x}$

This represents only one possible outcome. The number of ways of choosing x successes out of n events is the binomial coeff. The resulting distribution is the Binomial or B(n,p).

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$x = 0, 1, 2, 3, \ldots, n$$

Bin. Coeff for single success in 3 years = 3(2)(1) / 2(1) = 3

For 3 success in 3 years = 6 / (3) (2)(1) = 1

Binomial Dist' n B(n,p)

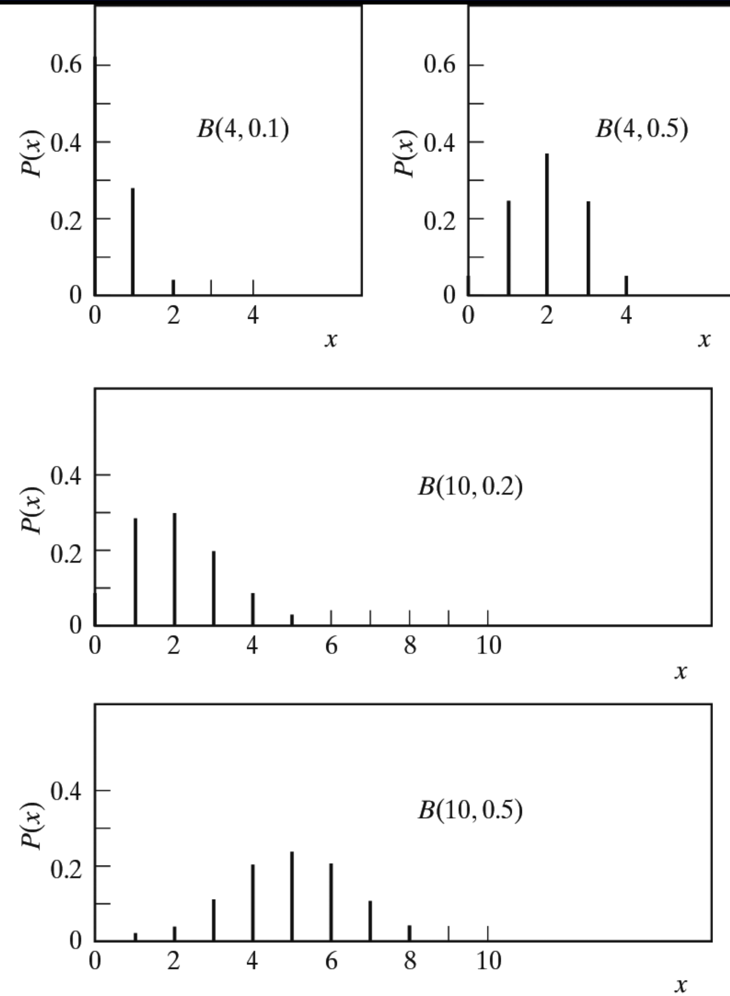


Figure 3.13

Binomial Probability mass function (PMF). (From Benjamin and Cornell, 1970, Fig. 3.3.1.)

Risk and Reliability

The probability of at least one success in n years, where the probability of success in any year is 1/T, is called the RISK.

Prob success = $p = 1/T$ *and* Prob failure = $1-p$

$$
\begin{aligned}
RISK &= 1 - P(0) \\
&= 1 - Prob(\text{no success in } n \text{ years}) \\
&= 1 - (1-p)^n \\
&= 1 - (1 - 1/T)^n
\end{aligned}
$$

$$Reliability = (1 - 1/T)^n$$

Design Periods vs RISK and Design Life

Expected Design Life (Years)

Risk %	5	10	25	50	100
75	4.1	7.7	18.5	36.6	72.6
50	7.7	14.9	36.6	72.6	144.8
20	22.9	45.3	112.5	224.6	448.6
10	48	95.4	237.8	475.1	949.6

x 2

x 3

Risk Example

What is the probability of at least one 50 yr flood in a 30 year mortgage period, where the probability of success in any year is 1/T = 1.50 = 0.02

$$\text{RISK} = 1 - (1 - 1/T)^n = 1 - (1 - 0.02)^{30}$$

$$= 1 - (0.98)^{30} = 0.455 \text{ or } 46\%$$

If this is too large a risk, then increase design level to the 100 year where p = 0.01

$$\text{RISK} = 1 - (0.99)^{30} = 0.26 \text{ or } 26\%$$

Important Probability Distributions

Normal – mean and std dev. – zero skew

Log Normal (Log data) – same as normal

Gamma – skewed data

Exponential- constant skew

Normal, LogN, LPIII

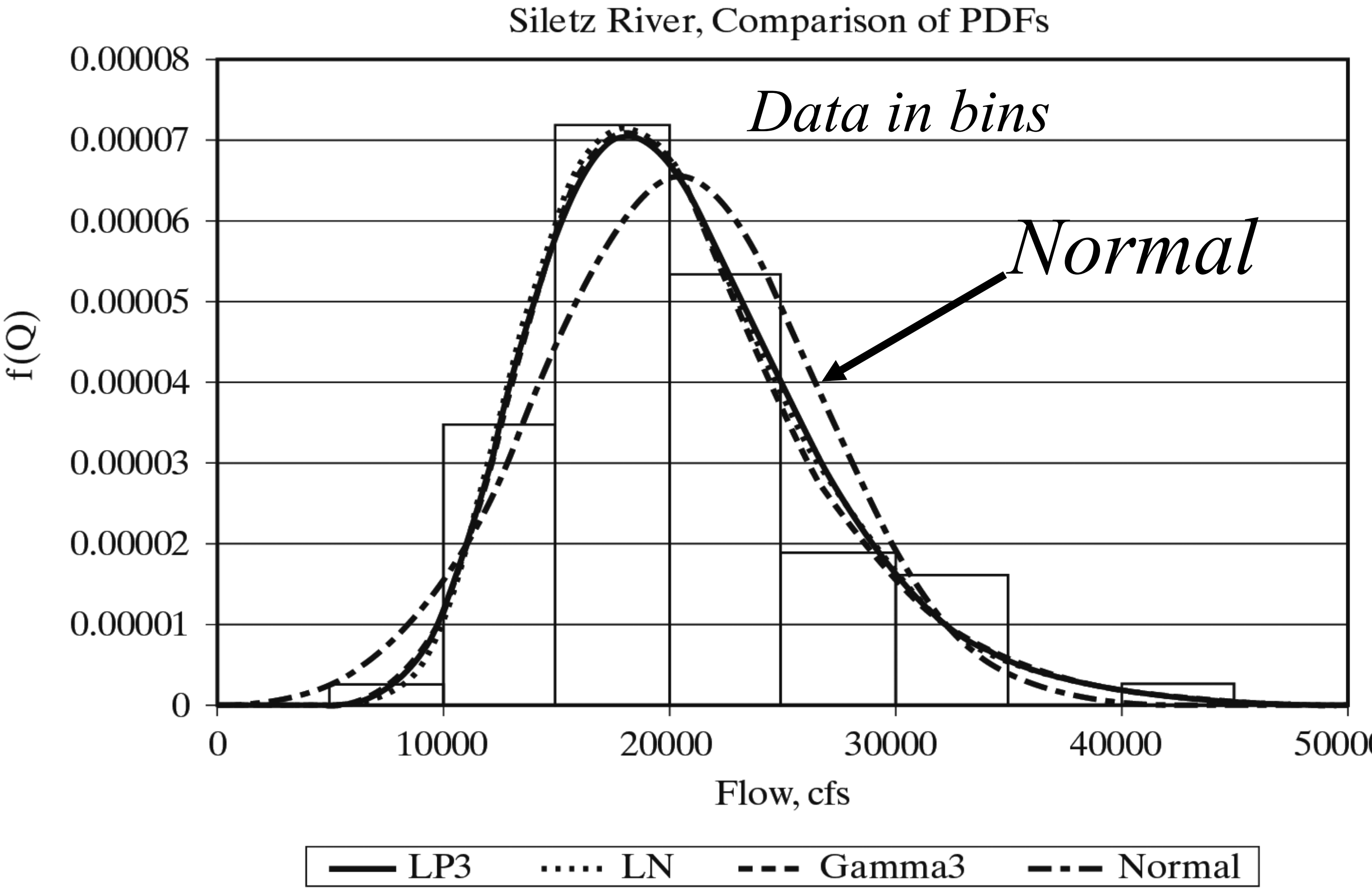


Figure 3.15

Four PDFs fit to data for the Siletz River. Fit is by the method of moments, as shown in the text, with moments given in Example 3.3.

Normal Prob Paper

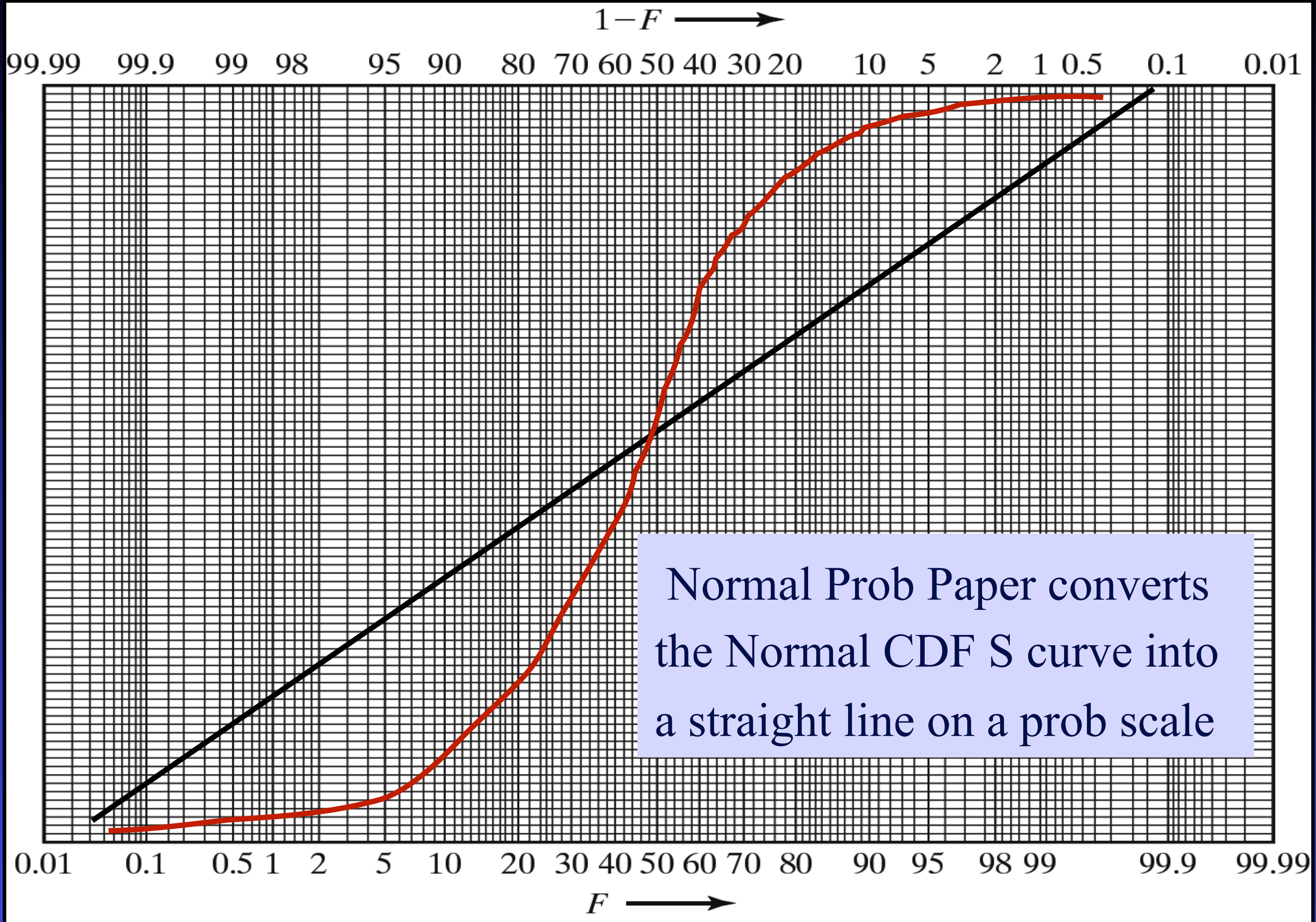


Figure 3.16

Normal probability paper.

Normal Prob Paper

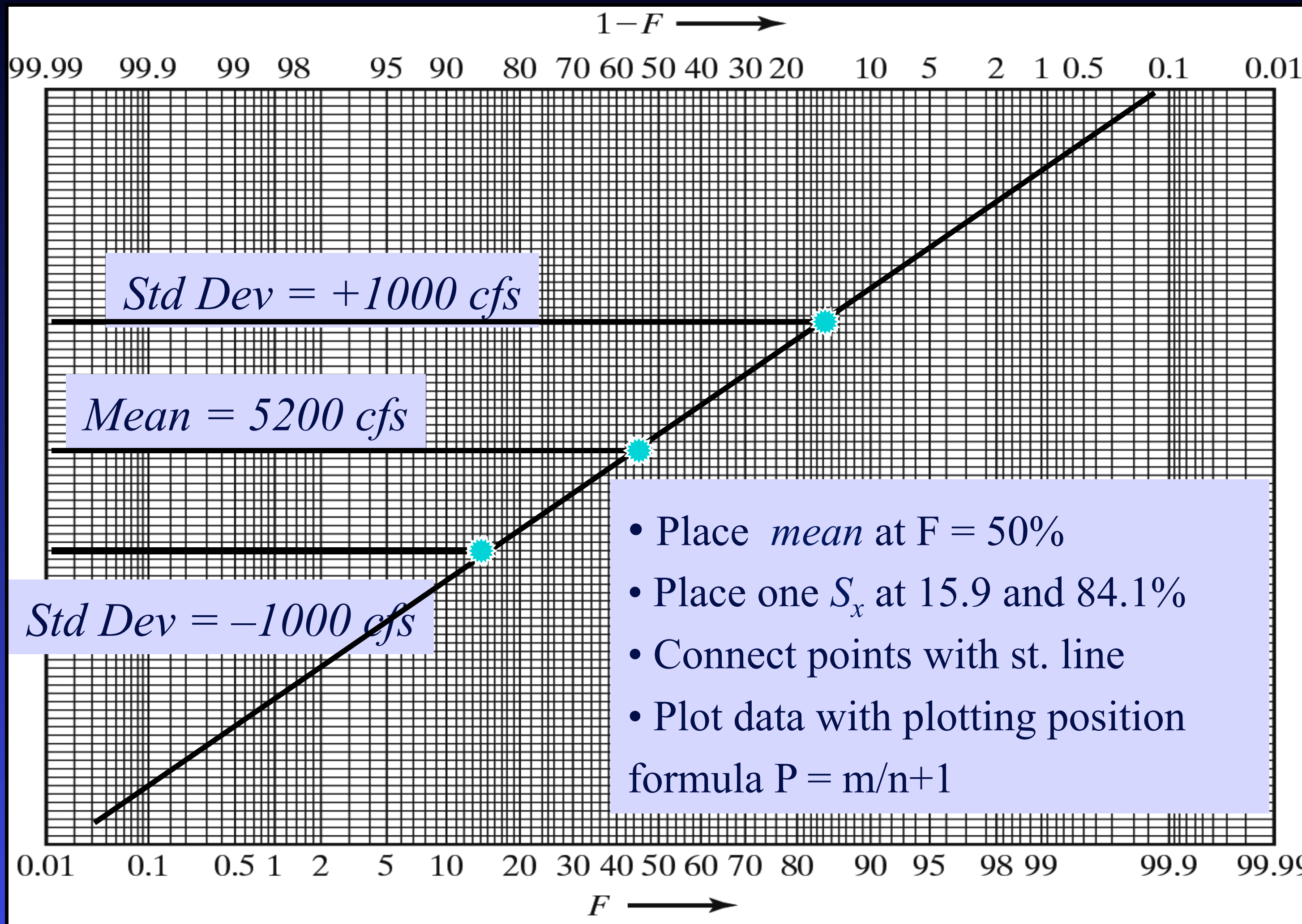


Figure 3.16

Normal probability paper.

Normal Dist'n Fit

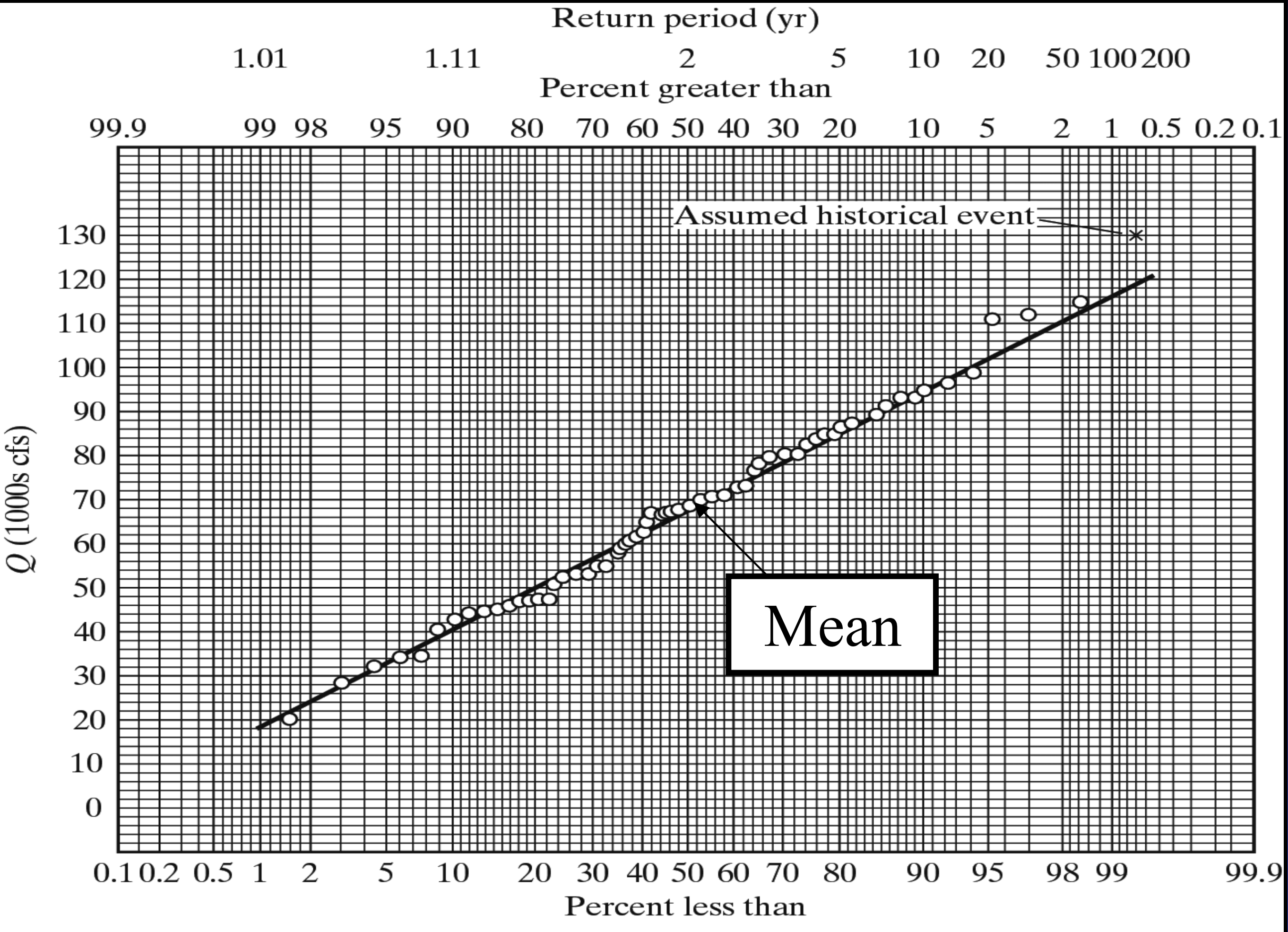


Figure P3.19

Normal probability plot for Kentucky River data. (From Haan, 1977, p. 137.)

Exponential Dist'n

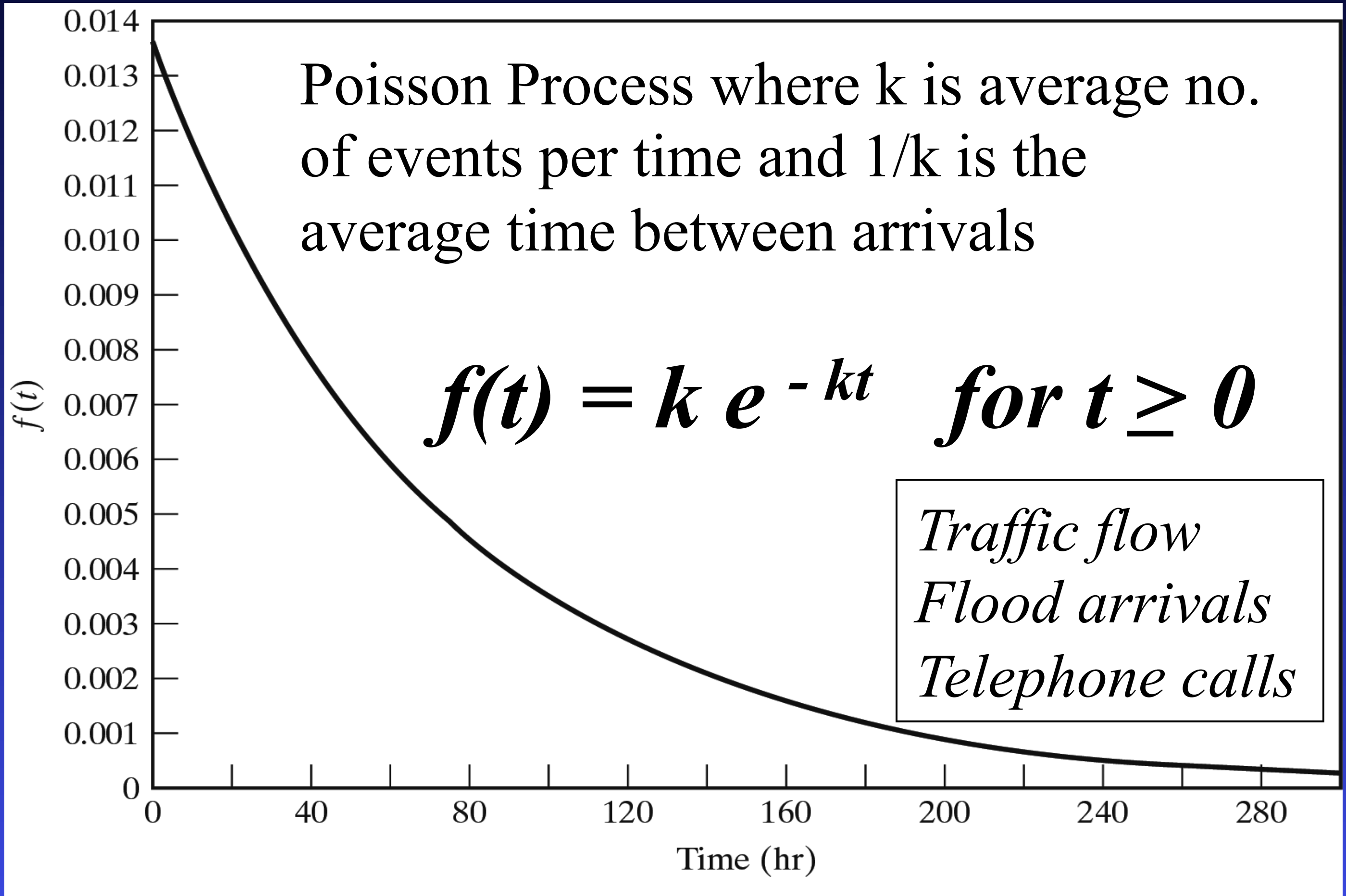


Figure 3.14

Exponential PDF. Parameters for Example 3.9.

Exponential Dist'n

$$f(t) = k\, e^{-kt} \quad \text{for } t \geq 0$$

$$F(t) = 1 - e^{-kt}$$

Avg Time Between Events

$$E(t) = \int_0^\infty (tk) e^{-kt}\, dt$$

$$\text{Letting } u = kt$$

$$\text{Mean or } E(t) = \frac{1}{k} \int_0^\infty u e^{-u}\, du = \frac{1}{k}$$

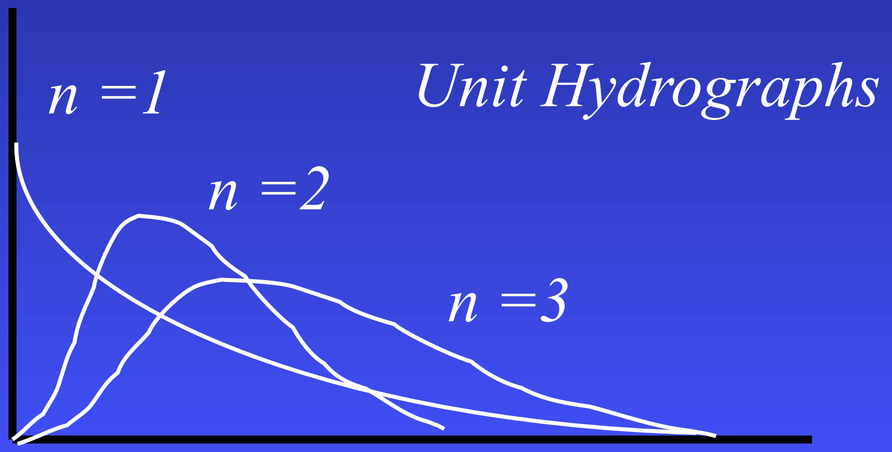
$$Var = \frac{1}{k^2}$$

Gamma Dist’n

$$Q_n = \frac{1}{K\Gamma(n)}\left(\frac{t}{k}\right)^{n-1} e^{-t/K}$$

$$Mean \quad or \quad E(t) = n\mathrm{K}$$

$$Var = nK^2 \qquad where\, \Gamma(n) = (n-1)!$$



Parameters of Dist'n

Distribution	Normal x	LogN $Y = \log x$	Gamma x	Exp t
Mean	μ_x	μ_y	nk	1/k
Variance	σ_x^2	σ_y^2	nk^2	$1/k^2$
Skewness	zero	zero	$2/n^{0.5}$	2

Frequency Analysis of Peak Flow Data

Year	Rank	Ordered cfs
1940	1	42,700
1925	2	31,100
1932	3	20,700
1966	4	19,300
1969	5	14,200
1982	6	14,200
1988	7	12,100
1995	8	10,300
2000	……	…….

Frequency Analysis of Peak Flow Data

- Take Mean and Variance (S.D.) of ranked data
- Take Skewness C_s of data (3rd moment about mean)
- If C_s near zero, assume normal dist' n
- If C_s large, convert Y = Log x - (Mean and Var of Y)
- Take Skewness of Log data - $C_s(Y)$
- If C_s near zero, then fits Lognormal
- If C_s not zero, fit data to Log Pearson III

Siletz River Example
75 data points - Excel Tools

	Original Q		Y = Log Q	
Mean	20,452		4.2921	
Std Dev	6089		0.129	
Skew	0.7889		- 0.1565	
Coef of Variation	0.298		0.03	

Siletz River Example - Fit Normal and LogN

Normal Distribution $Q = Q_m + z\,S_Q$

$$Q_{100} = 20452 + 2.326(6089) = \underline{34{,}620\ cfs}$$

Mean + z (S.D.)

Where z = std normal variate - tables

Log N Distribution $Y = Y_m + k\,S_Y$

$$Y_{100} = 4.29209 + 2.326(0.129) = 4.5923$$

k = freq factor and $Q = 10^Y = \underline{39{,}100\ cfs}$

Log Pearson Type III

Log Pearson Type III $Y = Y_m + k\,S_Y$

K is a function of Cs and Recurrence Interval
Table 3.4 lists values for pos and neg skews

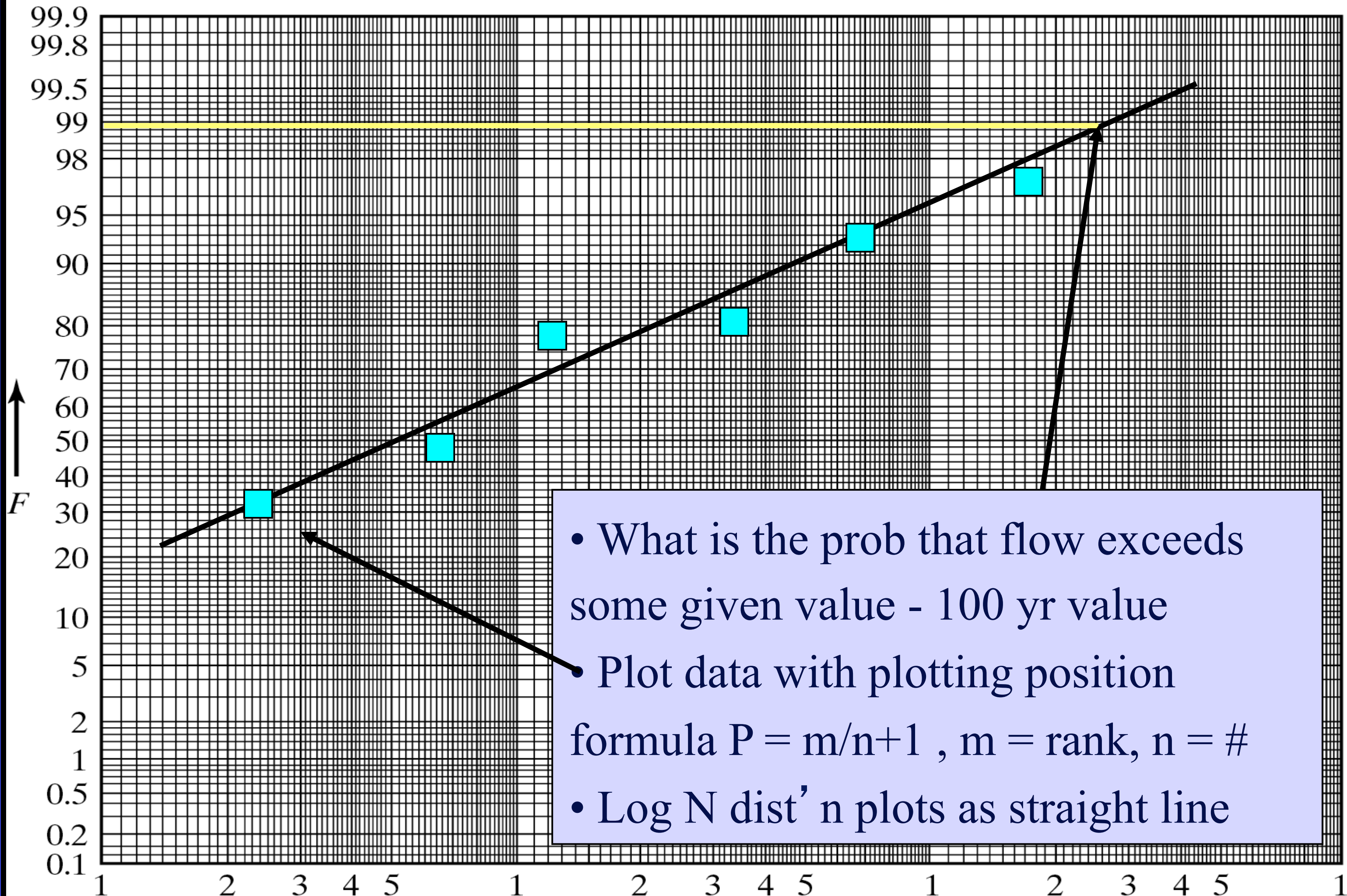
For Cs = -0.15, thus K = 2.15 from Table 3.4

$$Y_{100} = 4.29209 + 2.15(0.129) = 4.567$$

$Q = 10^Y = 36{,}927$ *cfs for LP III*

Plot several points on Log Prob paper

LogN Prob Paper for CDF



- What is the prob that flow exceeds some given value - 100 yr value
- Plot data with plotting position formula P = m/n+1 , m = rank, n = #
- Log N dist'n plots as straight line

Figure 3.17

Lognormal probability paper.

LogN Plot of Siletz R.

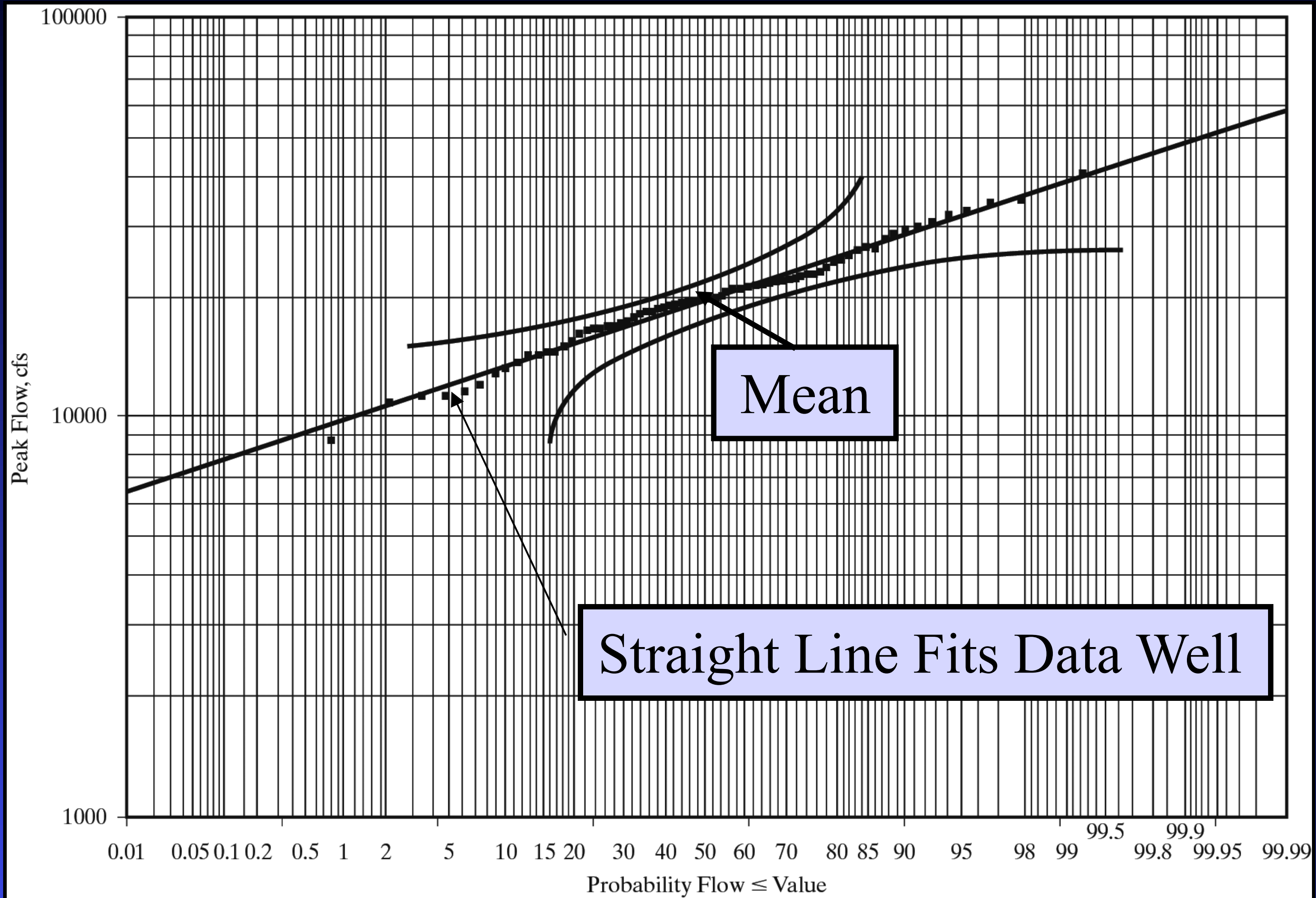


Figure 3.18

Lognormal plot of Siletz River flows, with 90% confidence intervals. Only every fifth value is plotted in the middle of the ranked series, for additional clarity.

Siletz River Flow Data

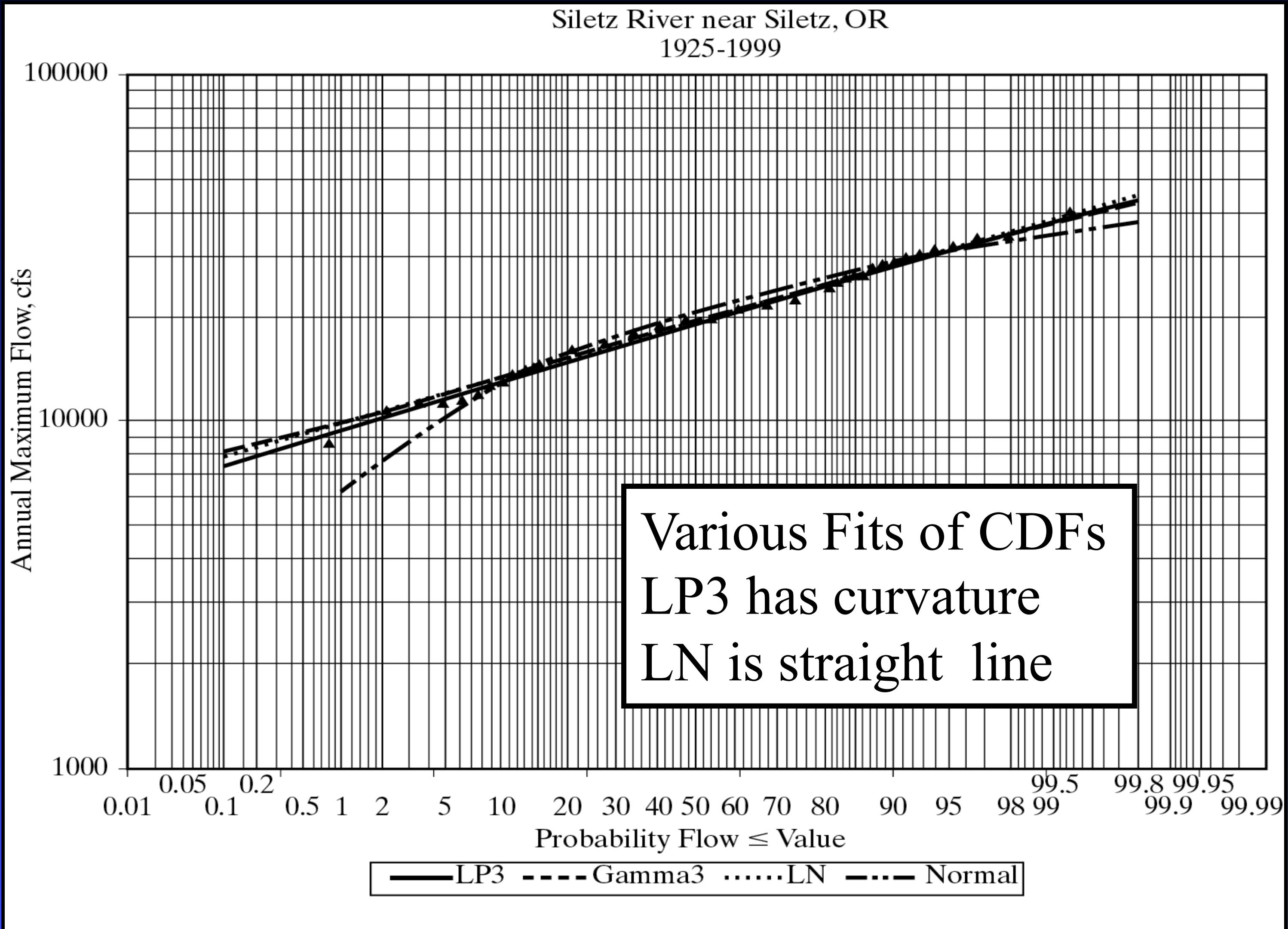


Figure 3.20

Comparison of four fitted CDFs for Siletz River flows 1925–1999.

Flow Duration Curves

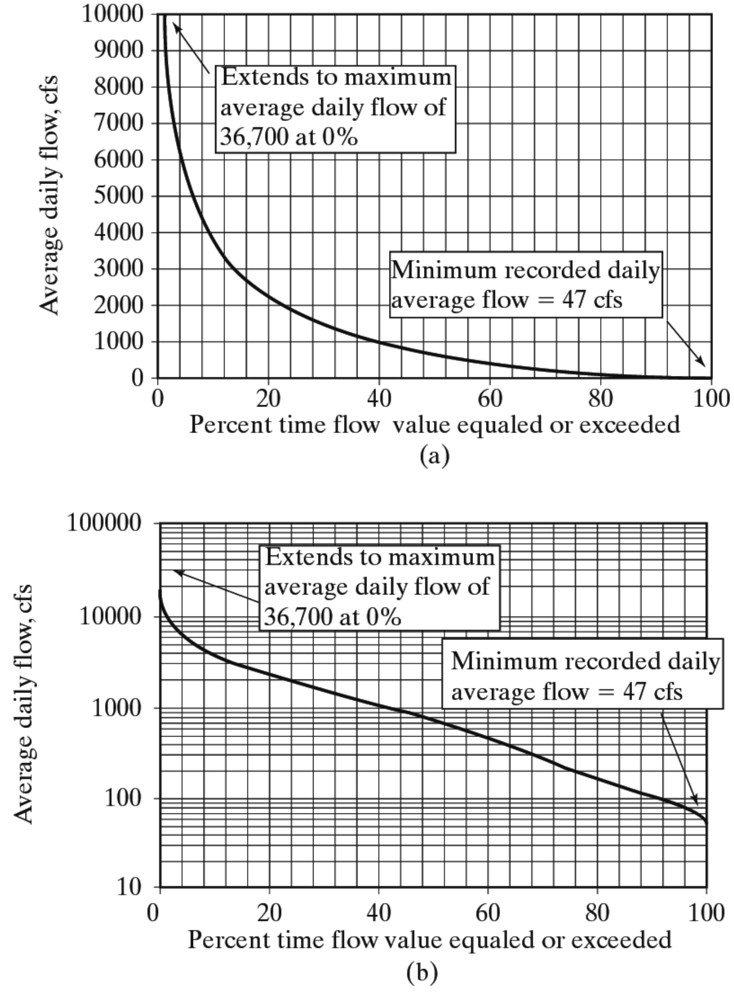


Figure 3.12

Flow-duration curve for the Siletz River. (a) Arithmetic scale, used for analysis of yield for water supply. (b) Logarithmic scale, useful when maximum and minimum flows have large separation.

Trends in data have to be removed before any Frequency Analysis