Module 4:
Probability and Extreme Floods

CWR 3540: Water Resources Engineering
FIU Department of Civil & Environmental Engineering
Professor Fuentes

Background

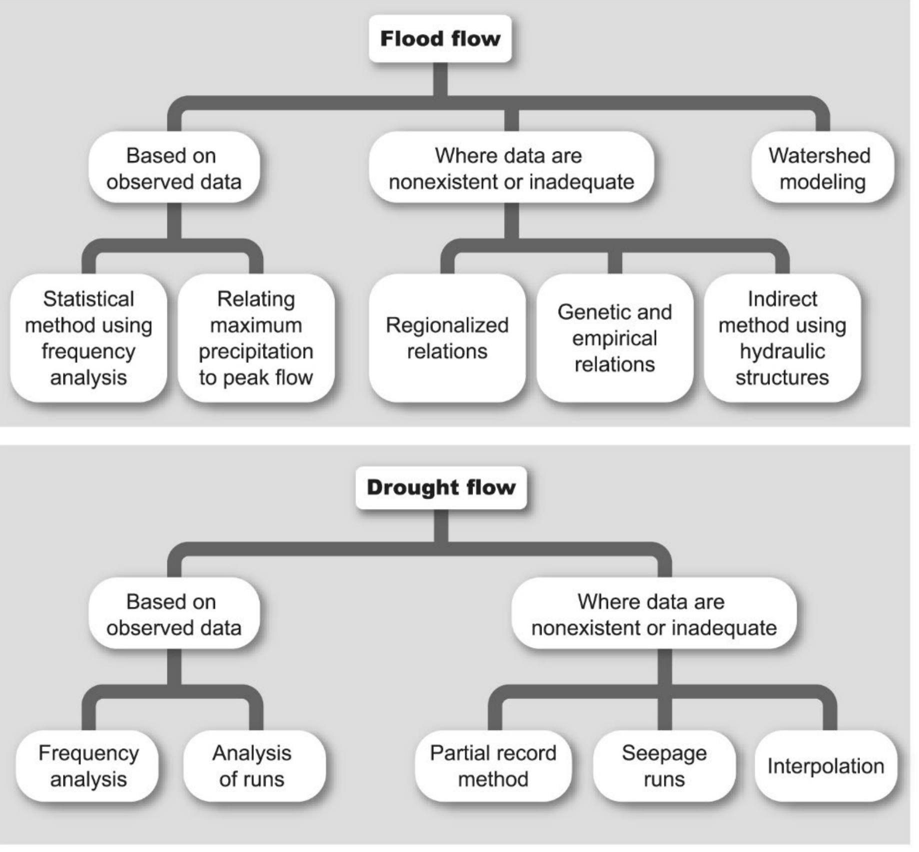
- Floods and droughts are extreme hydrological events.
  - *If available*, streamflow and/or precipitation records are the basis to estimate floods and droughts:
    - When records are not long enough, extrapolation and statistical techniques are used to “estimate” extreme values
  - If not available, empirical and other methods are also used to estimate extreme values

Recurrence Interval or Return Period, T, and Probability of Occurrence

- Floods (in high-flow analysis) will be equal or exceeded in a period of time (i.e., T)
  - $T = 1/P_{\geq}$, where P = probability of occurrence
- Droughts (in low-flow analysis) will be equaled or less:
  - $T = 1/P_{\leq}$, where P = probability of occurrence

Methods for Extreme (Flood) Flow & Drought Computation

Figure 11.1 Methods for flood flow computation and procedures used in drought analysis.



Basic Definitions

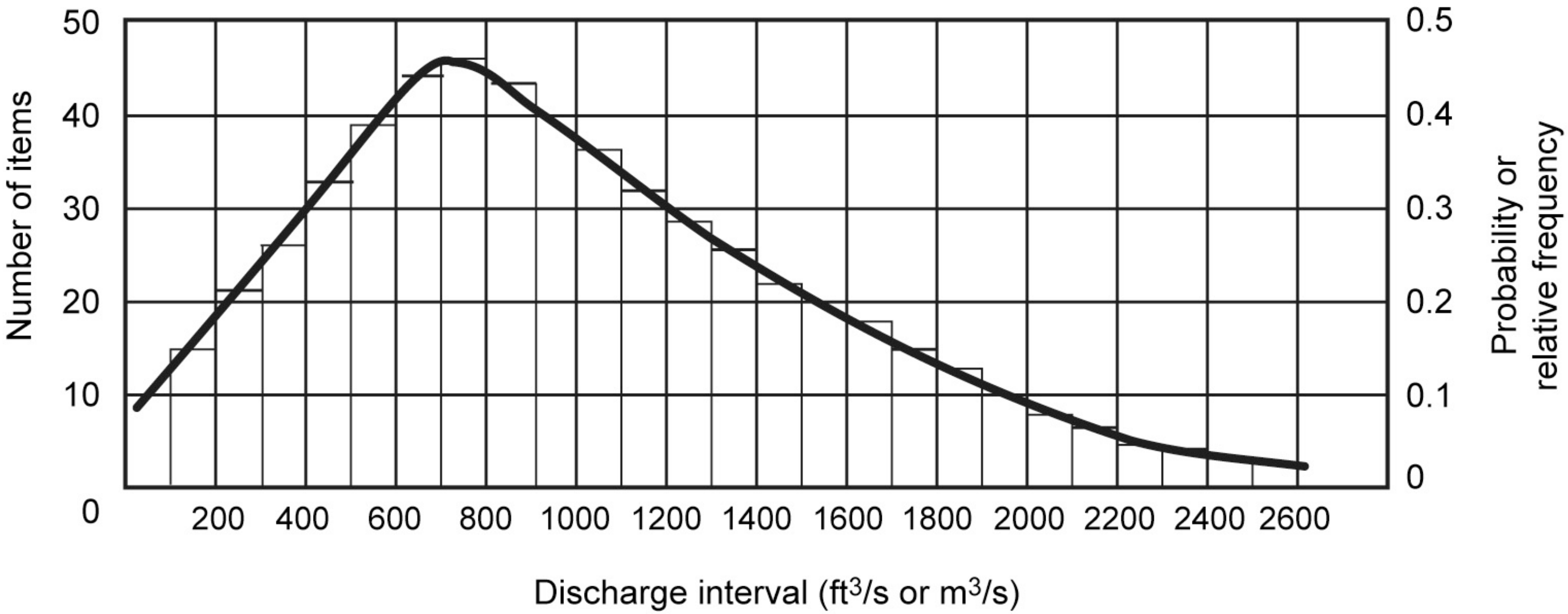
- Discrete versus continuous random variables (e.g., floods, droughts, precipitation measurements, infiltration rates, etc.)
- *Variate* = individual observation or value (e.g., a discharge of 1 ft^3/s measured at a particular location at a specific time)
- Time series = an array of *variates*
- Classes = equal intervals of groups of *variates* (e.g., 0-10, 10-20, 20-30, 30-40, …, ft^3/s)
- Frequency = number of items (or *variates*) in a class

Frequency Distribution Curve:

n_i = number of items in ith-class; N = total number of items in a series

$p = n_i/N$ [Eq. 11.1]

Figure 11.2 Frequency distribution curve.



Probability Distributions

- For continuous random variable, x:
  - PDF or Probability Density Function
  - CPD or Cumulative Distribution Function

- Common PDFs (See 11.5 for detailed list with properties):
  - Normal,
  - Lognormal
  - Extreme value
  - Log-Pearson Type III (Gamma Type)

Common PDFs

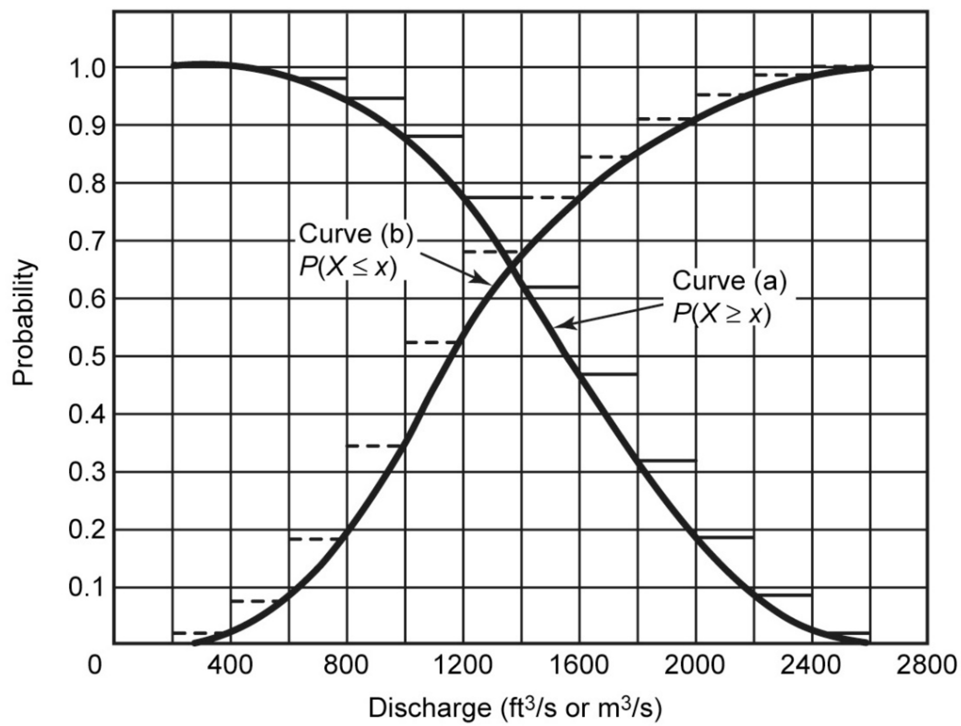
Table 11.5 Properties of Common Distributions

Distribution	Probability Density Function (PDF), $p(X)$	Cumulative Density Function (CDF), $P(X \le x)$	Range	Mean μ or $\bar{X}$	Standard Deviation σ or S
1. Normal	$\frac{1}{\sqrt{2\pi}\sigma} e^{-(X-\mu)^2/2\sigma^2}$ or $\frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ where $z = \frac{X-\mu}{\sigma}$	$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx$	$-\infty \le x \le \infty$	μ	σ
2. Lognormal	$\frac{1}{\sqrt{2\pi}\sigma_y} e^{-(y-\mu_y)^2/2\sigma_y^2}$	$\int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-(y-\mu_y)^2/2\sigma_y^2} dy$	$-\infty \le y \le \infty$	μ_y	σ_y
$y = \ln x$			$0 \le x \le \infty$		
3. Extreme value Type I, $y = (x-\beta)/\alpha$	$\frac{1}{\alpha} e^{-y-e^{-y}}$	e^{-e-y}	$-\infty \le x \le \infty$	$\beta + 0.577\alpha$	1.283α
Type III	$\alpha x^{\alpha-1} \beta^{-\alpha} e^{-(x/\beta)\alpha}$	$1 - e^{-(x/\beta)\alpha}$	$x \ge 0$	$\beta\Gamma(1 + 1/\alpha)$	$\beta[\Gamma(1 + 2/\alpha) - \Gamma^2(1 + 1/\alpha)]^{1/2}$
4. Log-Pearson	$p_0(1 + y/\alpha)^c e^{-cy/\alpha}$	$\int_{-\infty}^{y} p_0(1 + y/\alpha)^c e^{-cy/\alpha} dy$	$-\infty \le y \le \infty$	$(c+1)\frac{\alpha}{c}$	$\sqrt{c+1}\frac{\alpha}{c}$
Type III	where	(known as incomplete gamma function)	$0 \le x \le \infty$		
$y = \ln x$	p_0 = prob. at the mode $= \frac{N}{\alpha} \frac{c^{c+1}}{e^c \Gamma(c+1)}$				

Γ is the gamma function; $\Gamma(n) = (n-1)!$.
α and β are evaluated from relations shown under the columns of mean and standard deviation.
c and α are evaluated from relations shown under the columns of mean and standard deviation.

Cumulative Probability Curve

Figure 11.3 Cumulative probability curve.



Design Flood for Structures

- Acceptable Level of Risk:
  - Probable Maximum Flood
  - Optimum Design Flood for a Return Period, T
- Economic Factors:
  - For instance, peak flow rate at a Return Period, T, that minimizes the average annual cost (construction cost, O&M, damage cost)
- Standard Practice:
  - Based on
    - type of structure,
    - importance of the structure, and
    - development of the area subject to flooding

Probability of at Least One Flood in n-years

- f_x (exactly k events in n years) =

  $C^n_k P^k (1 - P)^{n-k}$

  where,

  $C^n_k = n! \div k! \, (n = k)!$ (see Eq. 11.3 for definitions)

  See Example 11.1

- f_x (at least one flood in n years) =

  $= 1 - (1 - P)^n$, where $P_{\geq} = 1/T$

  and Risk = R = f_x x 100 *(see Example 11.2 and 11.3)*

Risk level = f(Return Period)

Table 11.1 Return Period, 1/*P*, For Various Risk Levels [eq. (11.4)]

Acceptable Level of Risk, *R* (%)	Project Life, *n* (years) 25	50	100
	Return Period		
1	2440	5260	9950
25	87	175	345
50	37	72	145
75	18	37	72
99	6	11	27

Probability Paper

- General Purpose:
  - To “linearize” the plot (i.e., Y = MX + N) for different CDFs
- Probability paper for
  - Normal
  - Lognormal
  - Type I extreme value or Gumbel
  - Type III extreme value or Weibull

Methods of Flood Frequency Analysis

- Graphical Method
  - Usable for long records (not that common)
  - Based on plotting of a Frequency Distribution Curve (see Figure 11.2) that is data organized by classes and their frequencies
  - Common paper: lognormal probability
- Empirical Method
  - Also graphical
  - Based on a ranking of variates from largest to smallest to estimate P or T for a plotting position (see Example 11.5)
- Analytical Method
  - Based on linearized functions, such as Equation 11.10, and the K-T relationships for each PDF of choice (see tables 11.6, 11.7 and 11.7)

Graphical Method

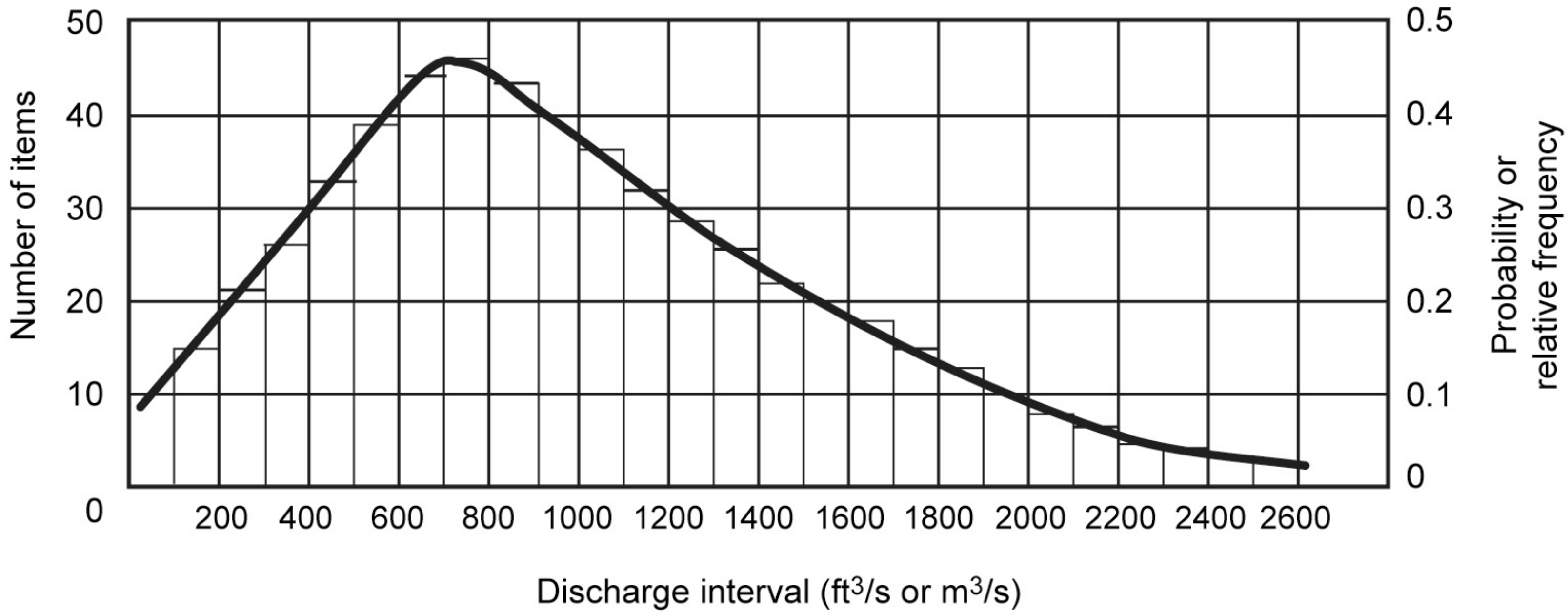
- Just like the development of a ”frequency distribution” curve/plot as illustrated in Figure 11.2 and examples.
- Procedure:
  - Flood flows are arranged into several class intervals of equal range in discharge;
  - Cumulate the number of occurrences in each class, starting with the highest value;
  - Determine the % of the class occurrences per total occurrences;
  - Plot the % versus the lower discharge limit of each class on probability paper (i.e., commonly lognormal probability paper)

Example: Empirical Method:

n_i = number of items in ith-class; N = total number of items in a series

$p = n_i/N$ [Eq. 11.1]

Figure 11.2 Frequency distribution curve.



Empirical Method

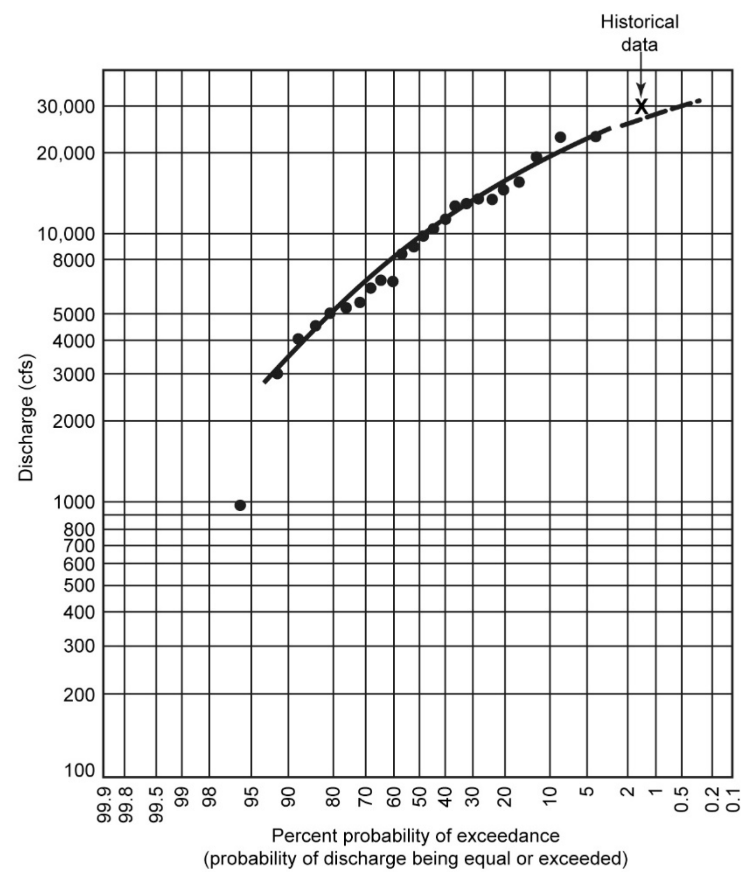
- Also, a graphical method, as follows:
  - Organize an array of n flood flow values in descending order of magnitude starting with the highest value;
  - Assign a rank m to each value, 1 to n, form the highest to the smallest one;
  - Calculate the *"plotting position"* using Weibull's formula (1939):
    - $P_m = m/(n + 1)$ or Equation 11.9, p. 436
  - Plot flow (i.e., discharge) versus the plotting position (or *percent probability of exceedance)on* lognormal probability paper (see example of Figure 11.6)
  - See application of Example 11.5.

Example 11.5: Empirical Method

Table 11.4 Annual Peak Flows of the River in Example 11.5

(1) Year	(2) Flow (cfs)	(3) Rank	(4) Plotting Position (%)	(1) Year	(2) Flow (cfs)	(3) Rank	(4) Plotting Position (%)
1991	14,400	5	20	2003	6,240	17	68
1992	6,720	16	64	2004	22,700	1	4
1993	13,390	7	28	2005	11,140	10	40
1994	15,360	4	16	2006	4,560	21	84
1995	8,856	13	52	2007	5,376	19	76
1996	5,136	20	80	2008	12,480	9	36
1997	6,770	15	60	2009	19,200	3	12
1998	9,600	12	48	2010	12,984	8	32
1999	980	24	96	2011	5,450	18	72
2000	4,030	22	88	2012	13,440	6	24
2001	10,440	11	44	2013	22,680	2	8
2002	3,100	23	92	2014	8,400	14	56

Figure 11.6 Flood-frequency curve for Example 11.5.



Analytical Method Distributions

(i.e., commonly used in hydrologic engineering)

- Normal (Gaussian) Distribution
- Lognormal Distribution
- Extreme Value Distribution (or Gumbel)
- Log-Pearson Type III (Gamma—Type) Distribution – *adopted as standard by U.S. federal agencies for flood analysis*

Analytical Method

- In classical analysis, the CDF is solved in tables that provide the cumulative (exceedance) probability for a desired flow value **or**
- In hydrological engineering, we used Chow simplified approach (1951), with *Y = log X*
  - $X = \ddot{X} + KS$ (L^3T^{-1}) for the Normal (Gaussian) distribution; *K = frequency factor of used PDF*
  - $Y = \bar{Y} + KS$ (L^3T^{-1}) for the Lognormal, Log-Pearson Type III (Gamma-type) and Type I extreme value (Gumbel) distributions; *K = frequency factor of used PDF*
  - *See steps in Section 11.10.1 – Use of frequency factors*

K - Normal Probability Distribution

Table 11.6 Frequency Factor for Normal Distribution

Exceedance Probability	Return Period	*K*	Exceedance Probability	Return Period	*K*
0.0001	10,000	3.719	0.450	2.22	0.126
0.0005	2,000	3.291	0.500	2.00	0.000
0.001	1,000	3.090	0.550	1.82	–0.126
0.002	500	2.88	0.600	1.67	–0.253
0.003	333	2.76	0.650	1.54	–0.385
0.004	250	2.65	0.700	1.43	–0.524
0.005	200	2.576	0.750	1.33	–0.674
0.010	100	2.326	0.800	1.25	–0.842
0.025	40	1.960	0.850	1.18	–1.036
0.050	20	1.645	0.900	1.11	–1.282
0.100	10	1.282	0.950	1.053	–1.645
0.150	6.67	1.036	0.975	1.026	–1.960
0.200	5.00	0.842	0.990	1.010	–2.326
0.250	4.00	0.674	0.995	1.005	–2.576
0.300	3.33	0.524	0.999	1.001	–3.090
0.350	2.86	0.385	0.9995	1.0005	–3.291
0.400	2.50	0.253	0.9999	1.0001	–3.719

K - Log-Pearson Type III Distribution

Table 11.7 Frequency Factors for Log-Pearson Type III Distribution

Skew Coefficient, *g*	Probability							
	0.99	0.80	0.50	0.20	0.10	0.04	0.02	0.01
	Return Period							
	1.0101	1.2500	2	5	10	25	50	100
3.0	−0.667	−0.636	−0.396	0.420	1.180	2.278	3.152	4.051
2.8	−0.714	−0.666	−0.384	0.460	1.210	2.275	3.114	3.973
2.6	−0.769	−0.696	−0.368	0.499	1.238	2.267	3.071	3.889
2.4	−0.832	−0.725	−0.351	0.537	1.262	2.256	3.023	3.800
2.2	−0.905	−0.752	−0.330	0.574	1.284	2.240	2.970	3.705
2.0	−0.990	−0.777	−0.307	0.609	1.302	2.219	2.912	3.605
1.8	−1.087	−0.799	−0.282	0.643	1.318	2.193	2.848	3.499
1.6	−1.197	−0.817	−0.254	0.675	1.329	2.163	2.780	3.388
1.4	−1.318	−0.832	−0.225	0.705	1.337	2.128	2.706	3.271
1.2	−1.449	−0.844	−0.195	0.732	1.340	2.087	2.626	3.149
1.0	−1.588	−0.852	−0.164	0.758	1.340	2.043	2.542	3.022
0.8	−1.733	−0.856	−0.132	0.780	1.336	1.993	2.453	2.891
0.6	−1.880	−0.857	−0.099	0.800	1.328	1.939	2.359	2.755
0.4	−2.029	−0.855	−0.066	0.816	1.317	1.880	2.261	2.615
0.2	−2.178	−0.850	−0.033	0.830	1.301	1.818	2.159	2.472
0	−2.326	−0.842	0	0.842	1.282	1.751	2.054	2.326
−0.2	−2.472	−0.830	0.033	0.850	1.258	1.680	1.945	2.178
−0.4	−2.615	−0.816	0.066	0.855	1.231	1.606	1.834	2.029
−0.6	−2.755	−0.800	0.099	0.857	1.200	1.528	1.720	1.880
−0.8	−2.891	−0.780	0.132	0.856	1.166	1.448	1.606	1.733
−1.0	−3.022	−0.758	0.164	0.852	1.128	1.366	1.492	1.588
−1.2	−3.149	−0.732	0.195	0.844	1.086	1.282	1.379	1.449
−1.4	−3.271	−0.705	0.225	0.832	1.041	1.198	1.270	1.318
−1.6	−3.388	−0.675	0.254	0.817	0.994	1.116	1.166	1.197
−1.8	−3.499	−0.643	0.282	0.799	0.945	1.035	1.069	1.087
−2.0	−3.605	−0.609	0.307	0.777	0.895	0.959	0.980	0.990
−2.2	−3.705	−0.574	0.330	0.752	0.844	0.888	0.900	0.905
−2.4	−3.800	−0.537	0.351	0.725	0.795	0.823	0.830	0.832
−2.6	−3.889	−0.499	0.368	0.696	0.747	0.764	0.768	0.769
−2.8	−3.973	−0.460	0.384	0.666	0.702	0.712	0.714	0.714
−3.0	−4.051	−0.420	0.396	0.636	0.660	0.666	0.666	0.667

K - Extreme Value Type I Distribution

Table 11.8 Frequency Factors for Extreme Value Type I Distribution

	Probability								
	0.2	0.1	0.067	0.05	0.04	0.02	0.0133	0.01	0.001
Sample Size, n	Return Period								
	5	10	15	20	25	50	75	100	1000
15	0.967	1.703	2.117	2.410	2.632	3.321	3.721	4.005	6.265
20	0.919	1.625	2.023	2.302	2.517	3.179	3.563	3.836	6.006
25	0.888	1.575	1.963	2.235	2.444	3.088	3.463	3.729	5.842
30	0.866	1.541	1.922	2.188	2.393	3.026	3.393	3.653	5.727
35	0.851	1.516	1.891	2.152	2.354	2.979	3.341	3.598	
40	0.838	1.495	1.866	2.126	2.326	2.943	3.301	3.554	5.576
45	0.829	1.478	1.847	2.104	2.303	2.913	3.268	3.520	
50	0.820	1.466	1.831	2.086	2.283	2.889	3.241	3.491	5.478
55	0.813	1.455	1.818	2.071	2.267	2.869	3.219	3.467	
60	0.807	1.446	1.806	2.059	2.253	2.852	3.200	3.446	
65	0.801	1.437	1.796	2.048	2.241	2.837	3.183	3.429	
70	0.797	1.430	1.788	2.038	2.230	2.824	3.169	3.413	5.359
75	0.792	1.423	1.780	2.029	2.220	2.812	3.155	3.400	
80	0.788	1.417	1.773	2.020	2.212	2.802	3.145	3.387	
85	0.785	1.413	1.767	2.013	2.205	2.793	3.135	3.376	
90	0.782	1.409	1.762	2.007	2.198	2.785	3.125	3.367	
95	0.780	1.405	1.757	2.002	2.193	2.777	3.116	3.357	
100	0.779	1.401	1.752	1.998	2.187	2.770	3.109	3.349	5.261
∞ [a]	0.719	1.305	1.635	1.866	2.044	2.592	2.911	3.137	4.936

[a]Additional data for $n = \infty$:

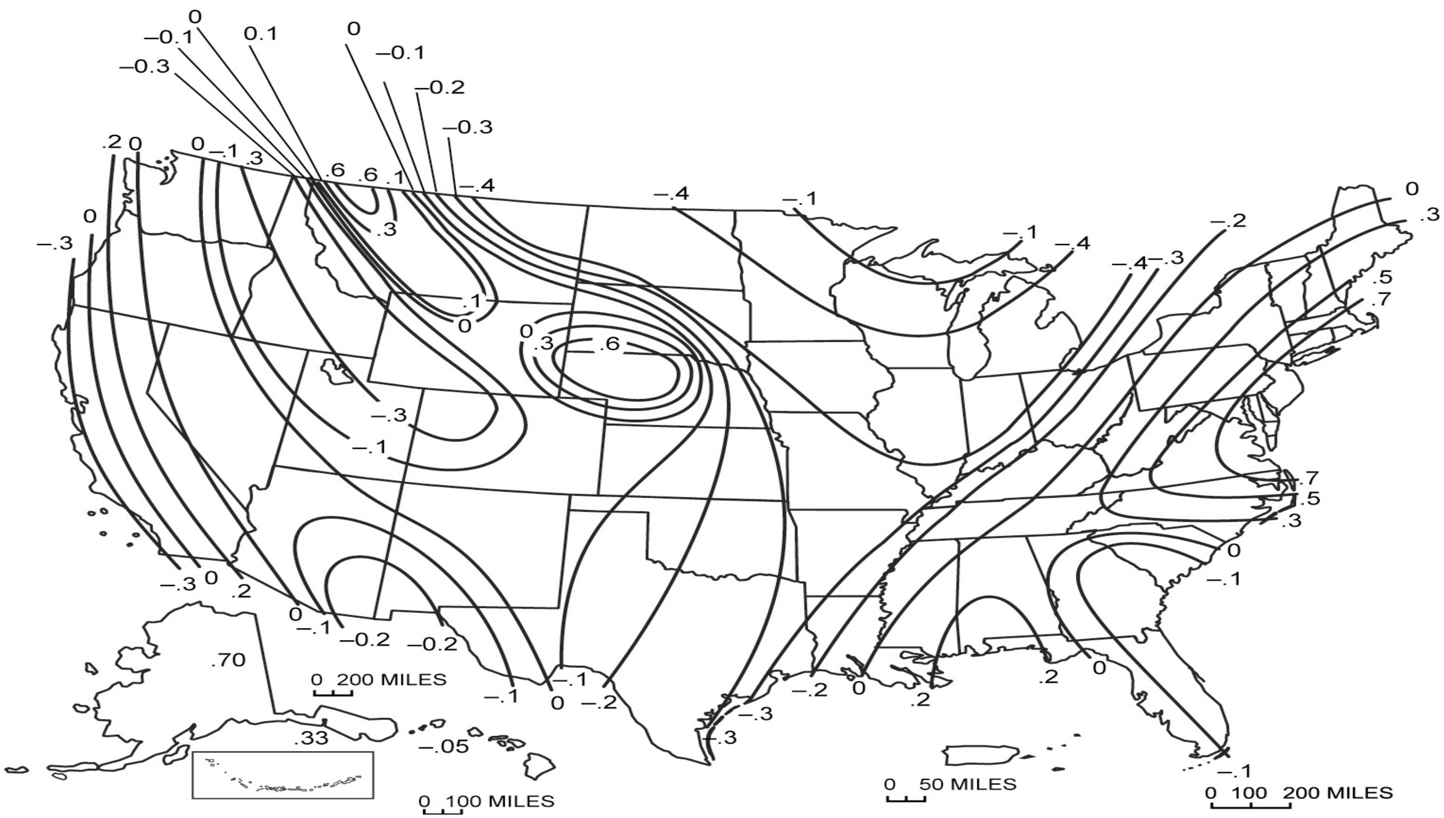
Probability	K
0.3	0.354
0.4	0.0737
0.5	−0.164
0.6	−0.383
0.8	−0.821
0.9	−1.100

Generalized Skew Coefficient

- $G = W g_s + (1 - W)g_m$
- $W = V(g_m)/ [V(g_s) = V (gm)]$
  - Where
    - g = generalized skew coefficient
    - W weighted factor
    - g_s sample skew coefficient
    - g_m map (regional) skew coefficient
    - V(   ) = mean squared error of the variable in parentheses

Map Skew Coefficients: Log Annual Maximum Streamflow

Figure 11.7 Map skew coefficients of logarithmic annual maximum streamflow (Interagency Advisory Committee on Water Data, 1982).



Confidence Limits: Error Limits

Table 11.10 Error Limits for Frequency Curve

Years of Record, *N*	Percent Exceedance Frequency (at 5% Level of Significance)[a]						
	0.1	1	10	50	90	99	99.9
5	4.41	3.41	2.12	0.95	0.76	1.00	1.22
10	2.11	1.65	1.07	0.58	0.57	0.76	0.94
15	1.52	1.19	0.79	0.46	0.48	0.65	0.80
20	1.23	0.97	0.64	0.39	0.42	0.58	0.71
30	0.93	0.74	0.50	0.31	0.35	0.49	0.60
40	0.77	0.61	0.42	0.27	0.31	0.43	0.53
50	0.67	0.54	0.36	0.24	0.28	0.39	0.49
70	0.55	0.44	0.30	0.20	0.24	0.34	0.42
100	0.45	0.36	0.25	0.17	0.21	0.29	0.37
	99.9	99	90	50	10	1	0.1
	Percent Exceedance Frequency (at 95% Level of Significance)[a]						

[a] Chance of true value being greater than the value represented by the error curve.
Source: Beard (1962).

Confidence Limits: Probability Adjustment

Table 11.11 Error Limits and Probability Adjustments

(a) P_α (%) (select)	0.1	1	10	50	90	99	99.9
(b) 5% level (from Table 11.10) ($N = 24$)	1.11	0.88	0.58	0.36	0.39	0.54	0.67
(c) Error limit, [(b) × S]	0.342	0.271	0.179	0.111	0.120	0.166	0.206
(d) Log value [X^a + (c)]		4.702	4.452	4.077	3.627	3.175	
(e) Curve value, [$\log^{-1}$ (d)] (cfs)		50,350	28,310	11,940	4240	1496	
(f) 95% level (from Table 11.10)	0.67	0.54	0.39	0.36	0.58	0.88	1.11
(g) Error limit, [(f) × S]	0.206	0.166	0.120	0.111	0.179	0.271	0.342
(h) Log value, [X^a – (g)]		4.265	4.153	3.855	3.328	2.738	
(i) Curve value, [$\log^{-1}$ (h)] (cfs)		18,410	14,220	7160	2130	550	
(j) P_N (for N – 1 = 23) (from Table 11.12)	0.29	1.58	11.0	50.0	89.0	98.42	99.71

S = Standard deviation of flood sample

[a] = Column 3 of Table 11.9

Confidence Limits: P-adjustment for Small Samples

Table 11.12 P_n Versus $P\infty$ for Normal Distribution (Percent)[a] for Expected Probability Adjustment

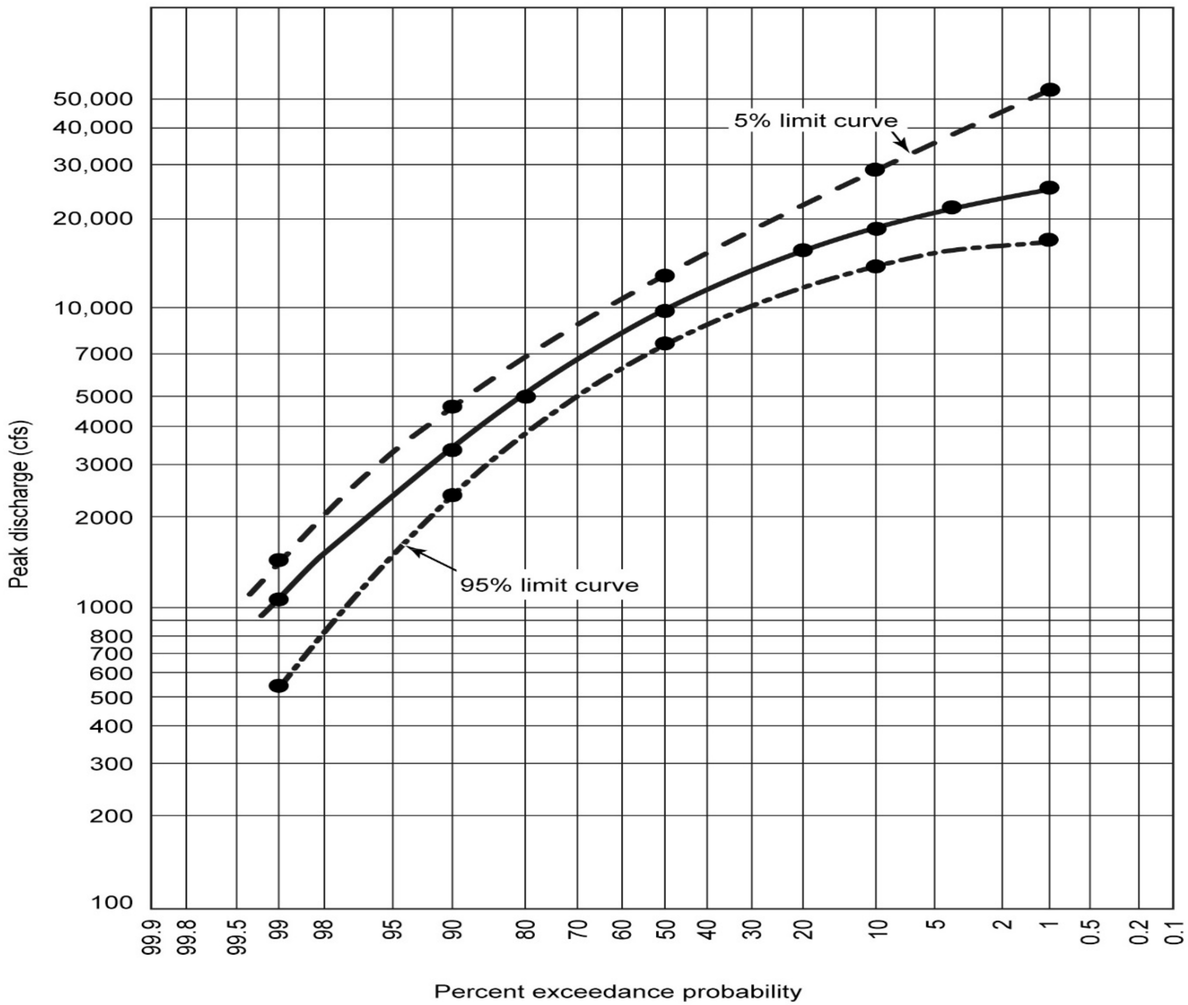
	$P\infty$						
$N-1$	50	30	10	5	1	0.1	0.01
	Adjusted Probability, P_n						
5	50.0	32.5	14.6	9.4	4.2	1.79	0.92
10	50.0	31.5	12.5	7.3	2.5	0.72	0.25
15	50.0	31.1	11.7	6.6	1.96	0.45	0.13
20	50.0	30.8	11.3	6.2	1.7	0.34	0.084
25	50.0	30.7	11.0	5.9	1.55	0.28	0.06
30	50.0	30.6	10.8	5.8	1.45	0.24	0.046
40	50.0	30.4	10.6	5.6	1.33	0.20	0.034
60	50.0	30.3	10.4	5.4	1.22	0.16	0.025
120	50.0	30.2	10.2	5.2	1.11	0.13	0.017
∞	50.0	30.0	10.0	5.0	1.0	0.10	0.01

[a] Values for probability > 50 by subtraction from 100 [i.e., $P_{90} = (100 - P_{10})$].

Source: Beard (1962).

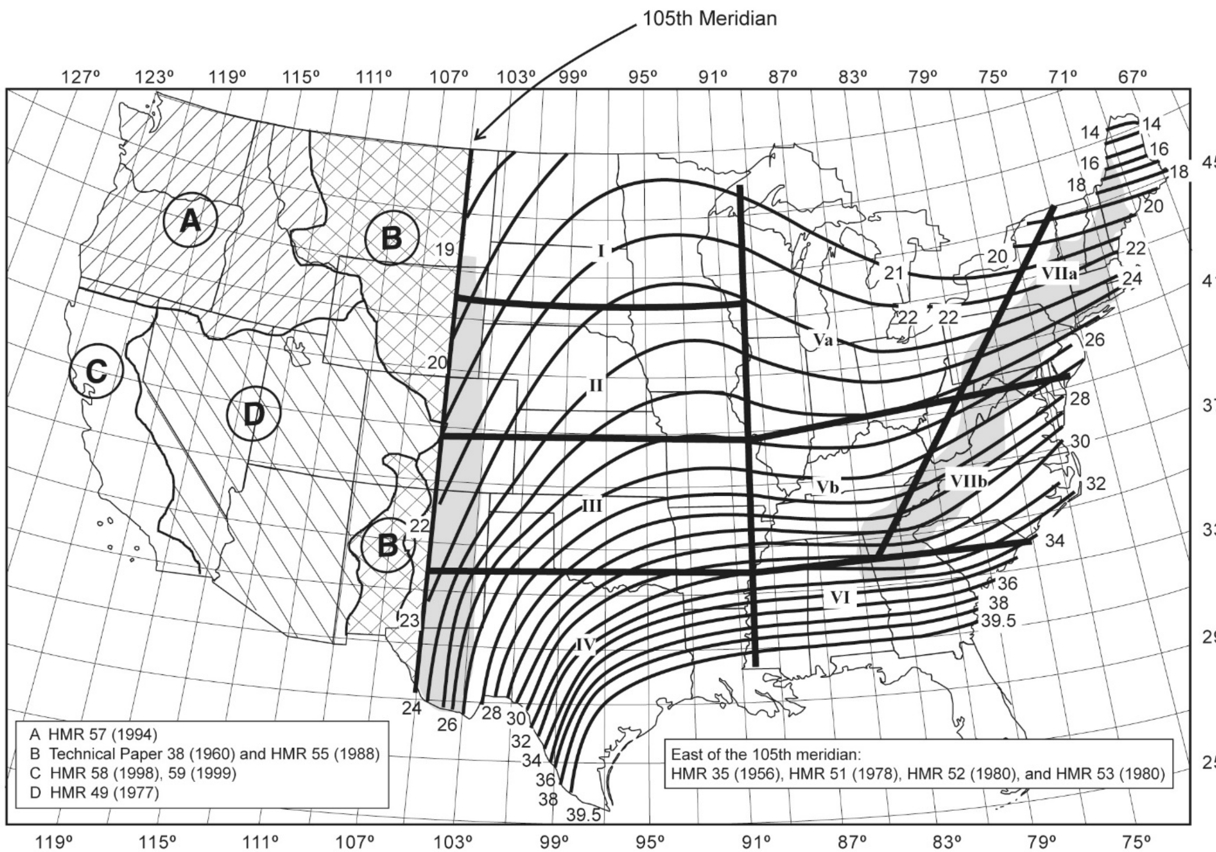
Log-Pearson Type III Curve: Example 11.6 with Confidence limits

Figure 11.8 Log-Pearson type III frequency curve and a reliability band.



Estimation of PMP
US National Weather Service

Figure 11.10 All-season PMP (in.) for 200 mi², 24 hr.



Depth-Area-Duration Relation Probable Maximum Precipitation (PMP)

Table 11.14 Depth-Area-Duration Relation of Maximum Probable Precipitation[a]

Storm Area		Duration	Regions								
mi^2	km^2	(hr)	I	II	III	IV	Va	Vb	VI	VIIa	VIIb
10	26	6	1.00	1.09	1.03	0.93	1.04	1.01	0.90	1.04	1.00
		12	1.20	1.29	1.22	1.10	1.26	1.18	1.07	1.21	1.16
		24	1.28	1.38	1.31	1.25	1.34	1.31	1.25	1.34	1.33
		48	1.38	1.50	1.45	1.40	1.50	1.45	1.40	1.50	1.45
		72	1.47	1.60	1.55	1.50	1.52	1.53	1.50	1.52	1.53
200	518	6	0.75	0.78	0.74	0.66	0.76	0.72	0.67	0.73	0.68
		12	0.90	0.93	0.87	0.82	0.93	0.86	0.81	0.87	0.85
		24	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		48	1.10	1.12	1.14	1.16	1.13	1.14	1.16	1.17	1.15
		72	1.15	1.20	1.20	1.22	1.22	1.23	1.23	1.23	1.25
1,000	2,590	6	0.57	0.56	0.54	0.50	0.56	0.52	0.50	0.52	0.52
		12	0.67	0.71	0.66	0.63	0.70	0.69	0.63	0.63	0.66
		24	0.77	0.80	0.79	0.79	0.80	0.79	0.83	0.80	0.80
		48	0.85	0.90	0.92	0.93	0.90	0.92	0.94	0.93	0.93
		72	0.96	0.97	0.98	1.00	0.97	0.98	1.04	0.98	0.98
5,000	12,950	6	0.36	0.36	0.31	0.28	0.36	0.31	0.28	0.33	0.31
		12	0.45	0.47	0.43	0.39	0.48	0.43	0.40	0.45	0.43
		24	0.52	0.54	0.54	0.55	0.54	0.54	0.55	0.56	0.56
		48	0.63	0.67	0.68	0.65	0.67	0.65	0.68	0.70	0.69
		72	0.70	0.74	0.76	0.76	0.74	0.76	0.78	0.74	0.76
10,000	25,900	6	0.26	0.27	0.23	0.21	0.28	0.23	0.22	0.28	0.23
		12	0.36	0.37	0.33	0.30	0.38	0.35	0.32	0.37	0.35
		24	0.42	0.45	0.43	0.43	0.47	0.44	0.45	0.47	0.45
		48	0.50	0.58	0.54	0.55	0.58	0.57	0.58	0.60	0.60
		72	0.60	0.62	0.64	0.65	0.66	0.64	0.67	0.67	0.65
20,000	51,800	6	0.18	0.20	0.17	0.16	0.20	0.17	0.16	0.20	0.16
		12	0.27	0.28	0.25	0.23	0.30	0.28	0.25	0.33	0.28
		24	0.35	0.36	0.35	0.32	0.38	0.36	0.36	0.40	0.37
		48	0.45	0.47	0.45	0.45	0.48	0.47	0.48	0.50	0.49
		72	0.50	0.55	0.55	0.55	0.56	0.55	0.56	0.57	0.55

[a] Factors derived by the author from the figures in National Weather Service (1978) to be applied to 24-hour values on 200-mi^2 area of Figure 11.10.