Workshop 1.3, Basic Frequency Analysis

Streamflow records for the Columbia River at the Dalles (USGS Station number 14105700) have been kept since 1858. Since much of the Columbia River Basin is regulated, reservoir operations must be taken into account when analyzing observed flows so that we can work with a "homogeneous" set of unregulated or natural flows. Table 1 and file **Columbia at the Dalles.xls** list unregulated (natural) annual peak flows for 135 years of record.

In this workshop, we'll work with the first 20 years of data, to make manual computations easier, and then later repeat the analysis with the full data set.

Tasks

(1) Select the first 20 values (1858 to 1877) of the Dalles data set, from Table 1 or Columbia at the Dalles.xls.

(2) **Graphical Distribution.** On Figure 1, graph the histograms of frequency and relative frequency of your sample. Try using an interval width of 100,000 cfs. Graph the cumulative probability distribution on Figure 2.

(3) **Statistics.** Calculate the mean and standard deviation of your sample, using the Table 2 worksheet as a model for the computation. **(Hint: work in 1000 cfs if hand computation.)** You can reproduce the worksheet in Excel for a more convenient computation.

(4) **Normal Distribution.** Assume annual peak discharge follows a Normal distribution. Using your sample statistics from step (3), compute a cumulative Normal distribution using Table 3 to structure your computations, and plot on Figure 2. A tabulation of the Standard Normal distribution is attached. *Consider working in Excel.*

(5) **LogNormal Distribution.** For the 20-member sample, the mean and standard deviation of the base 10 logarithms of the natural flows are:

mean = 2.827; standard deviation = 0.111 (note, these are for 1000 cfs).

Use those statistics and repeat the analysis of step (4), now assuming a logNormal distribution. Use Table 3 for tabulations and Figure 3 for plotting the logNormal cumulative distribution.

(6) Having fitted three distributions (graphical, Normal, LogNormal) to your sample data, determine the following from each curve:

a. The probability of observing a peak flow (Q) between 700,000 and 800,000 cfs, i.e., $P(700{,}000 \leq Q \leq 800{,}000)$.

b. $P(Q \leq 450{,}000)$

c. $P(Q \geq 10^6)$

d. The peak discharge which has a 1% chance of being exceeded in any given year; a 50% chance.

Consider plotting on figures 4 and 5 for the extreme value estimates.

Repeat computations for the entire sample. Use Excel, even if done by hand, above.

Table 1

Columbia River @ the Dalles

Water Year	Natural Peak	Water Year	Natural Peak	Water Year	Natural Peak	Water Year	Natural Peak
		1890	638,000	1930	358,000	1970	671,000
		1891	453,000	1931	338,000	1971	740,000
		1892	612,000	1932	630,000	1972	1,053,000
		1893	685,000	1933	759,000	1973	454,000
		1894	1,246,000	1934	474,000	1974	1,074,000
		1895	482,000	1935	521,000	1975	732,000
		1896	792,000	1936	576,000	1976	685,000
		1897	788,000	1937	414,000	1977	342,000
1858	563,000	1898	658,000	1938	652,000	1978	633,000
1859	847,000	1899	797,000	1939	423,000	1979	542,000
1860	668,000	1900	557,000	1940	411,000	1980	603,000
1861	618,000	1901	673,000	1941	353,000	1981	626,000
1862	948,000	1902	656,000	1942	477,000	1982	793,000
1863	777,000	1903	800,000	1943	586,000	1983	764,000
1864	654,000	1904	643,000	1944	365,000	1984	682,000
1865	714,000	1905	425,000	1945	542,000	1985	614,000
1866	839,000	1906	390,000	1946	622,000	1986	784,000
1867	671,000	1907	602,000	1947	580,000	1987	501,000
1868	483,000	1908	669,000	1948	1,078,000	1988	462,000
1869	328,000	1909	694,000	1949	700,000	1989	572,000
1870	777,000	1910	586,000	1950	864,000	1990	571,000
1871	856,000	1911	598,000	1951	694,000	1991	626,000
1872	737,000	1912	596,000	1952	640,000	1992	393,000
1873	638,000	1913	790,000	1953	720,000		
1874	582,000	1914	524,000	1954	639,000		
1875	684,000	1915	358,000	1955	660,000		
1876	958,000	1916	761,000	1956	972,000		
1877	486,000	1917	764,000	1957	868,000		
1878	486,000	1918	608,000	1958	786,000		
1879	644,000	1919	589,000	1959	691,000		
1880	915,000	1920	446,000	1960	548,000		
1881	600,000	1921	811,000	1961	846,000		
1882	885,000	1922	714,000	1962	566,000		
1883	575,000	1923	617,000	1963	540,000		
1884	700,000	1924	470,000	1964	824,000		
1885	485,000	1925	679,000	1965	731,000		
1886	676,000	1926	303,000	1966	518,000		
1887	899,000	1927	738,000	1967	845,000		
1888	567,000	1928	805,000	1968	594,000		
1889	306,000	1929	501,000	1969	682,000		

Table 2. Worksheet for computing Mean and Standard Deviation

	Flow Q	$(Q-\overline{X}_Q)$	$(Q-\overline{X}_Q)^2$
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20 = N			
$\Sigma Q_i =$		$\Sigma(Q_i-\overline{X}_Q)^2 =$	
$\overline{X}_Q = \Sigma Q_i /N =$		$S^2 = \Sigma(Q_i-\overline{X}_Q)^2 /(n-1) =$	
		$S = \sqrt{S^2} =$	

Table 3

Calculation of Normal and Lognormal Distribution

Sample Statistics:

$\overline{X}_Q$ = ____________ $\overline{X}_{\log Q}$ = ____________.

S_Q = ____________ $S_{\log Q}$ = ____________.

Standard Normal Distribution:

$$P(Q \le q) = P\left(\frac{Q-\mu}{\sigma} \le \frac{q-\mu}{\sigma}\right) = P\left(Z \le \frac{q-\mu}{\sigma}\right)$$

where Q = flow random variable
q = flow value
μ = mean
σ = standard deviation
Z = standard normal deviate

q (1000cfs)	$\frac{q-\overline{X}_Q}{S_Q}$	$P\left(Z \le \frac{q-\overline{X}_Q}{S_Q}\right)$	log q (1000cfs)	$\frac{\log q-\overline{X}_{\log Q}}{S_{\log Q}}$	$P\left(Z \le \frac{\log q-\overline{X}_{\log Q}}{S_{\log Q}}\right)$
100			2.000		
200			2.301		
300			2.477		
400			2.602		
500			2.699		
600			2.778		
700			2.845		
800			2.903		
900			2.954		
1000			3.000		
1100			3.041		
1200			3.079		
1300			3.114		
1400			3.146		

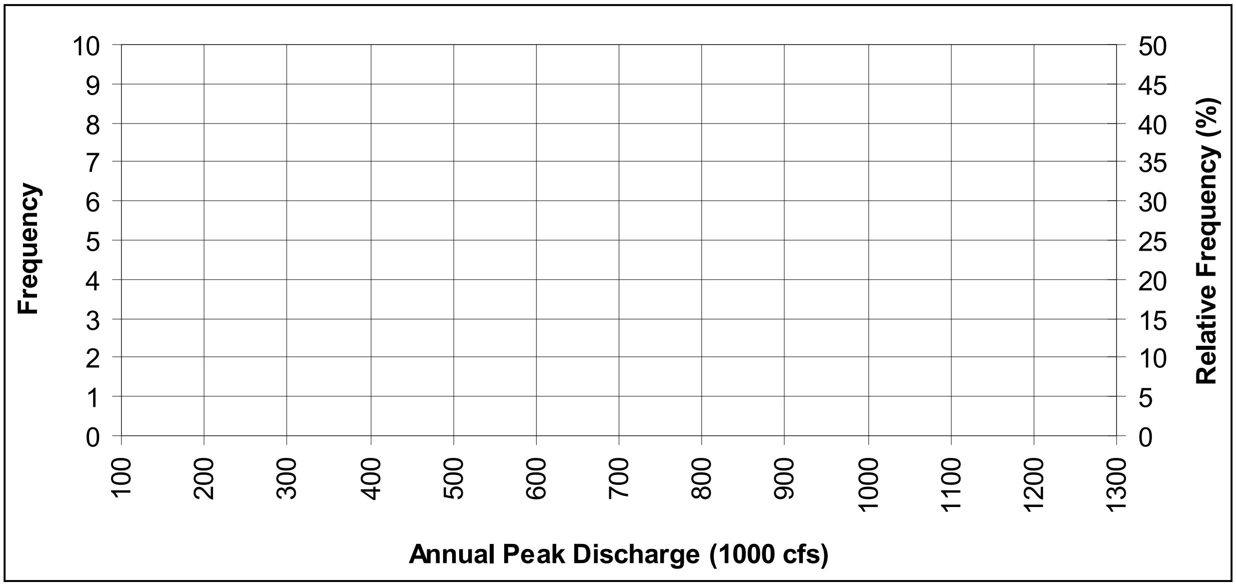

Figure 1. Frequency and Relative Frequency Histograms

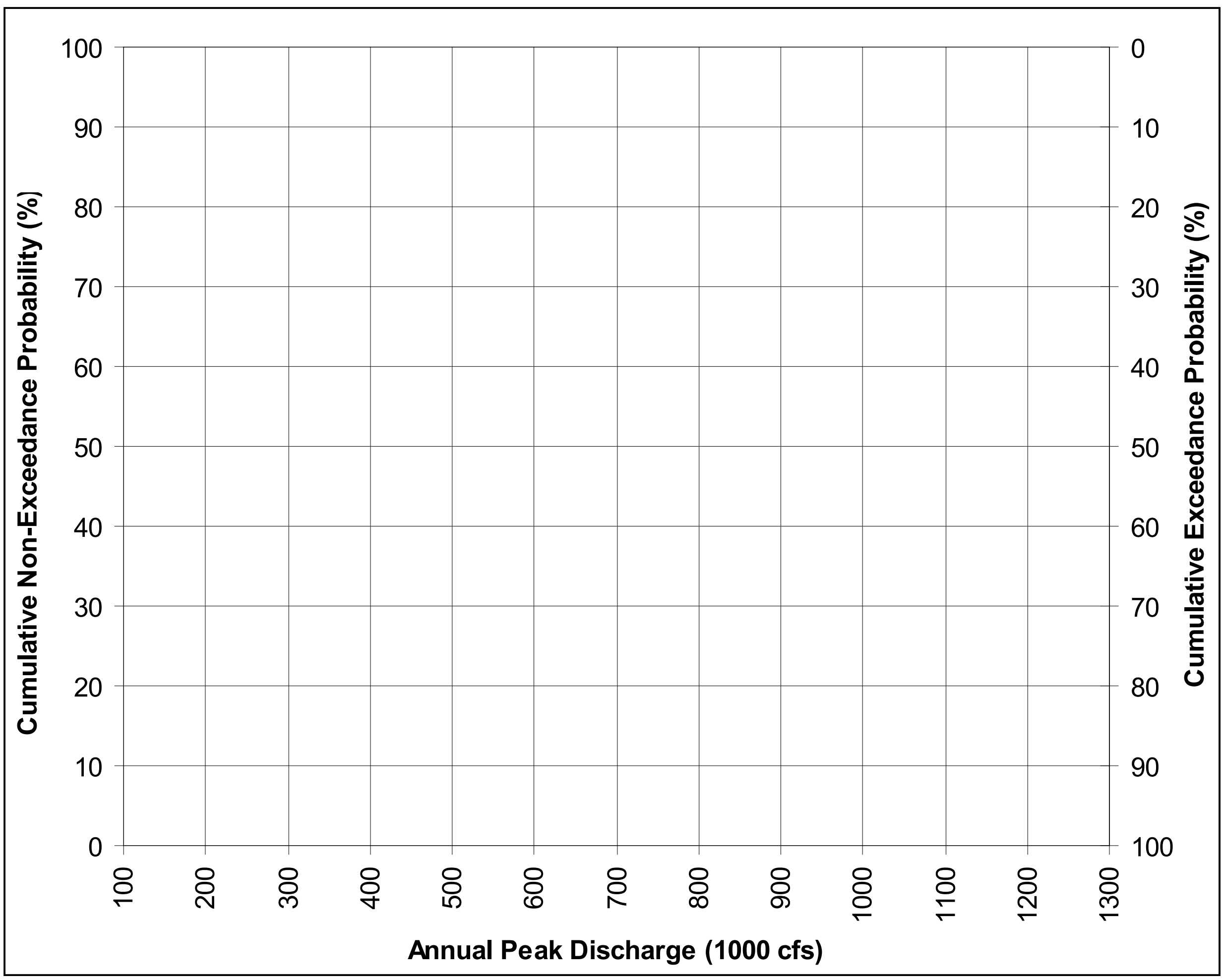

Figure 2. Cumulative Probability, Graphical and Normal Distribution

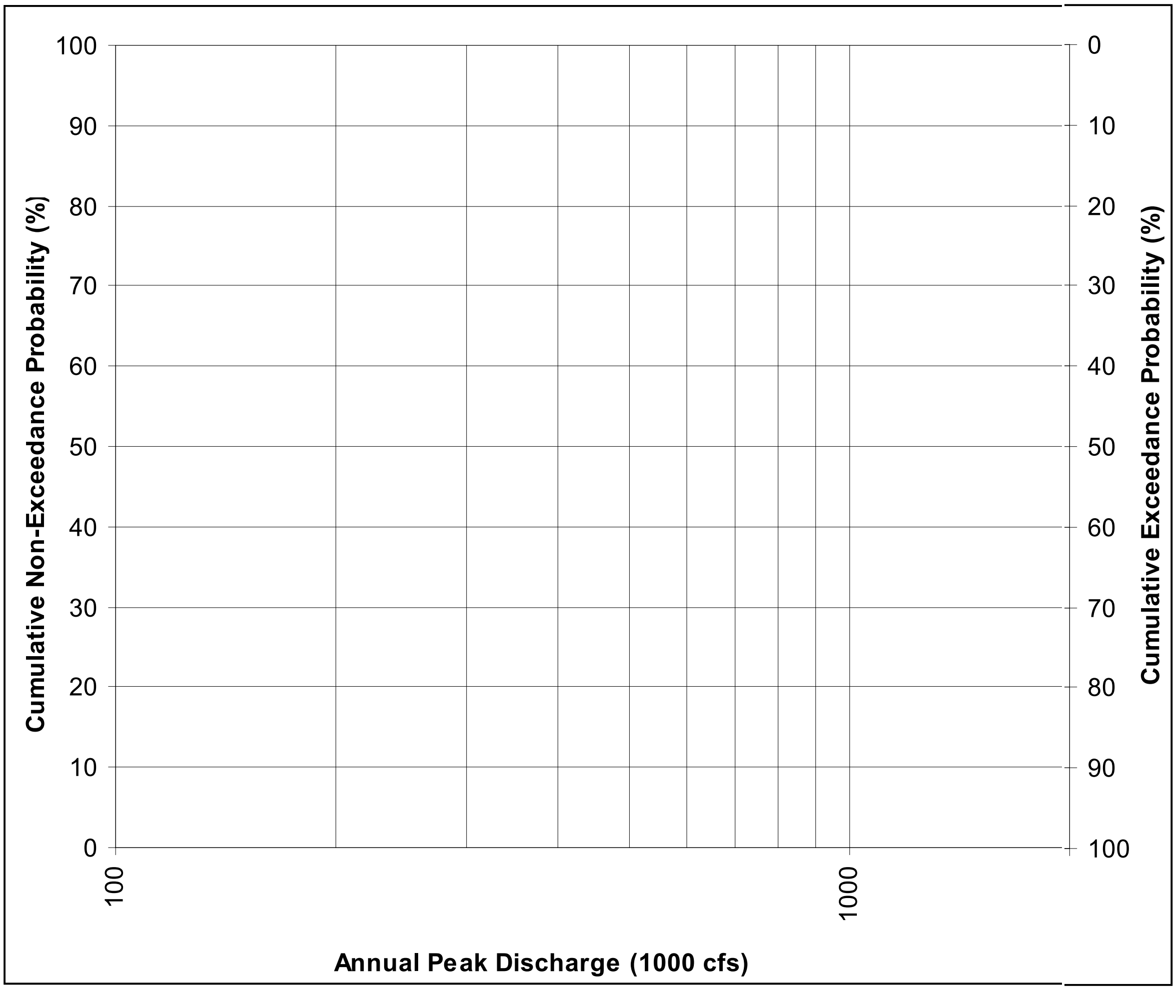

Figure 3. Cumulative Probability, LogNormal Distribution

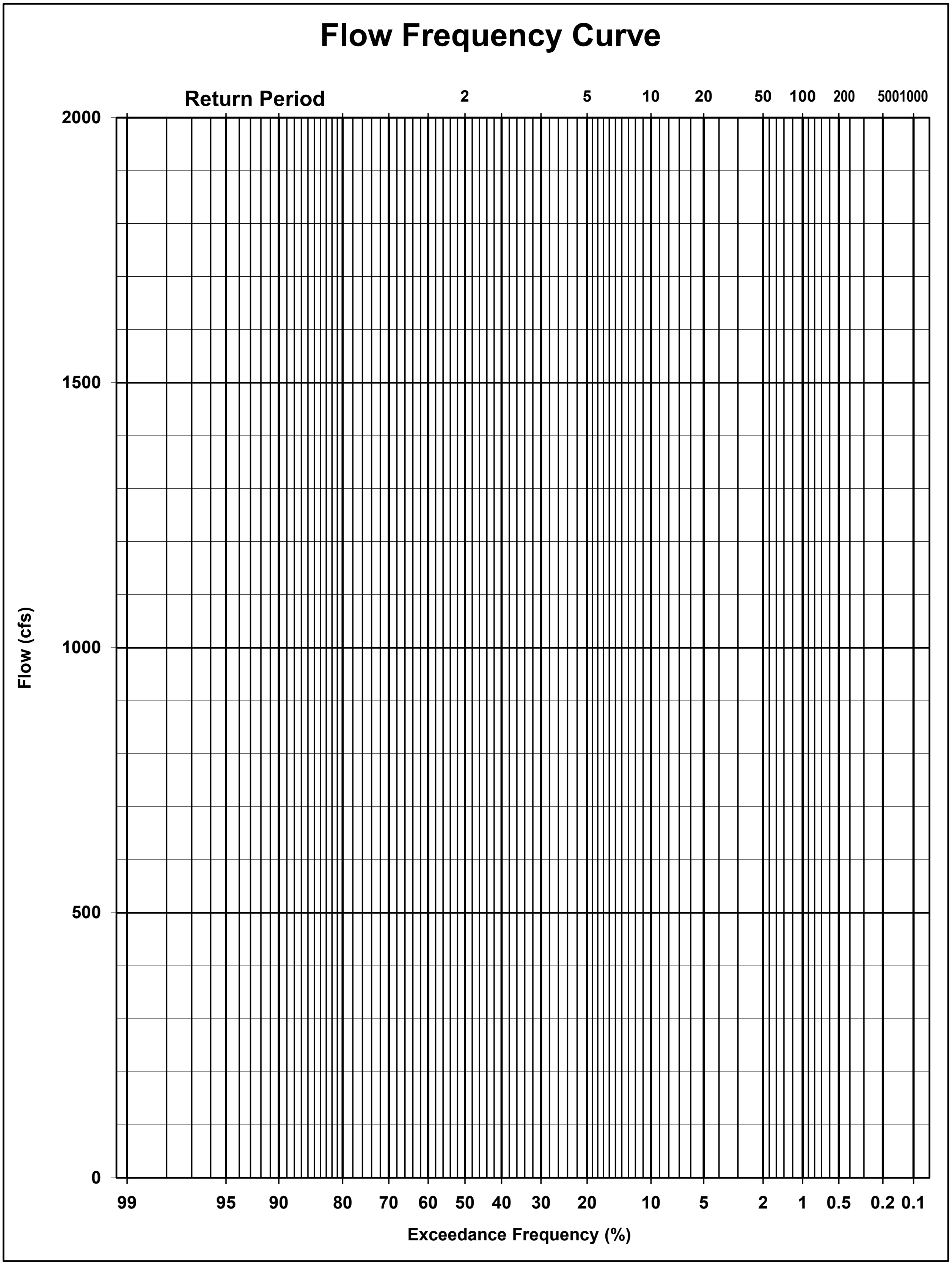

Figure 4. Linear flow versus Normal probability plot, for Graphical and Normal distributions

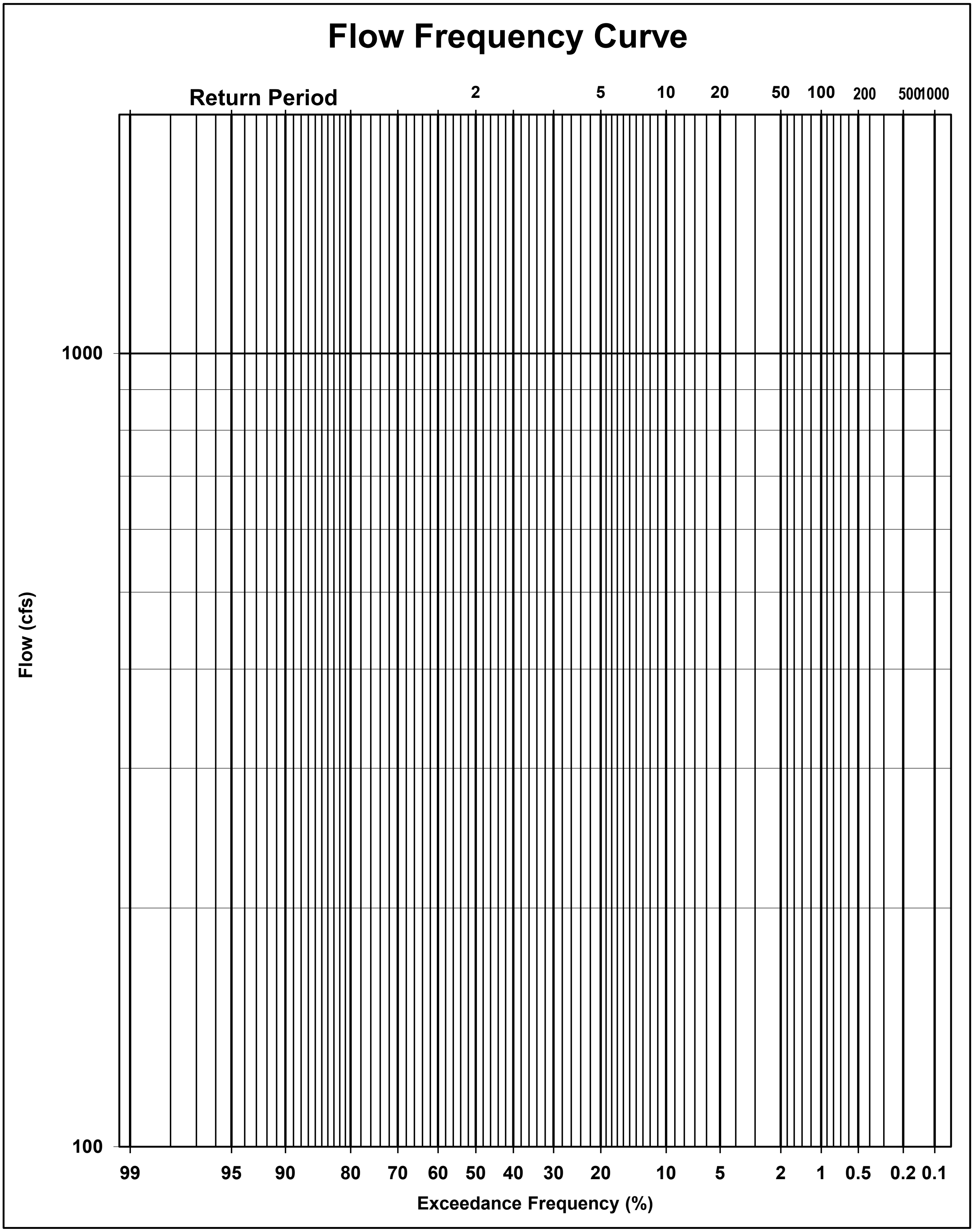

Figure 5. Log flow versus Normal probability plot, for LogNormal distribution