Σ²Π

S. S. PAPADOPULOS & ASSOCIATES, INC.
Environmental & Water-Resource Consultants

Memorandum

Date: March 19, 2002

From: Karen Lewis, Debbie Hathaway

To: Kevin Flanigan, NMISC
Mike Velasquez, USACE

Subject: Tech memo – comparison of statistical distributions for mainstem and tributary inflow for use in MRG3 Risk modeling

SSPA has reevaluated probability distribution functions (PDFs) for several of the inflow terms used in the @Risk modeling. This exercise was conducted in response to Phase 2 comments, in which some reviewers questioned why selected distributions were used. The reviewers suggested that the use of lognormal distributions be explored as an alternative to the several different types used in the Phase 2 study. Original and alternative PDFs are presented here for comparison.

The original and alternative distributions are:

Flow	Original distribution	Alternative distribution
• Otowi index flow, 1950-1998	Beta	Lognormal
• Otowi index flow, 1950-1998		Beta and Lognormal using 8 classes
• Otowi supplemented index, 1896-1998		Beta and Lognormal
• Otowi supplemented index, 1919-1998		Beta and Lognormal
• Jemez, 1950-1998	Beta	Lognormal
• Galisteo, 1970-1998	Gamma	Lognormal
• Rio Puerco, 1950-1998	PearsonVI	Lognormal
• Rio Salado, 1950-1998[1]	Pearson VI	Lognormal

Probability distributions were evaluated using the software *BestFit, Probability Distribution Fitting for Windows*, June 1997, distributed by Palisade Corporation.
For each of these terms, a pair of probability density function graphs, a table of goodness-of-fit statistics, and a probability plot showing the original data points and both original and alternative PDFs are given. These latter plots are collected at the end of the memo.

The pair of probability density function graphs shows the measured flow data plotted as a histogram, overlaid with a line representing the probability distribution function chosen for the data. The upper plot shows the distribution function used in the previous Middle Rio Grande study; the lower plot shows a potential new distribution function for the data, generally a lognormal. In all cases, with the exception of the Otowi index flow using 8 classes, the data histograms and function fitting use data divided into 10 classes.

Below each pair of graphs is the table indicating the goodness of fit between the specified distribution function and the raw data histogram. Test statistics and critical values at α=0.05 are given for the Chi-Square,

[1] The Rio Salado flow is based on measured flows from 1950-1984, and reconstructed flows from 1985-1998. The reconstructed flows are based on a linear regression between the Rio Salado and Rio Puerco from 1950-1984. The correlation between Rio Salado and Rio Puerco flow for that period is 0.57.

Kolmogorov-Smirnov, and Anderson Darling tests. Any probability distribution fit that has a value of the test statistic above the critical value is generally rejected, while PDF fits with test statistic values below the critical value are accepted. In all cases, the lognormal function does not fit as well as the originally chosen function. For several flows, the lognormal is rejected on the basis of the test statistics. (Note: the Kolmogorov-Smirnov statistics are normalized by the BestFit software used for the statistical analysis. Both test and critical values have been multiplied by $\sqrt{n}$ where n is the number of data points.)

Finally, the probability plots are collected together at the end of the memo. There is one plot for each term, showing both of the fitted PDFs in black and grey lines, and the original measured data as black points. These graphs provide a quick visual comparison between the fit of the PDFs to the data, and the comparison between old and new PDFs.

Particular attention has been paid to the Otowi flow in this analysis. In addition to presenting a lognormal PDF for the original data, the effects of changing the number of classes the data is divided into prior to fitting is given in the Otowi index flow 8-class analysis. The effects of using an extended data set are also shown. Both the full available data set for annual Otowi flow, starting in 1896 and running to 1998, with gaps from 1906-1909 and 1914-1918, and the continuous annual Otowi flow from 1919-1998 were fitted, both with a beta function (the best fitting function as specified by the BestFit software) and a lognormal. These records were compiled by using the raw USGS annual flows measured at Otowi Bridge prior to 1940, and the Otowi index flow from 1940 to 1998. The average flows represented by these time periods are:

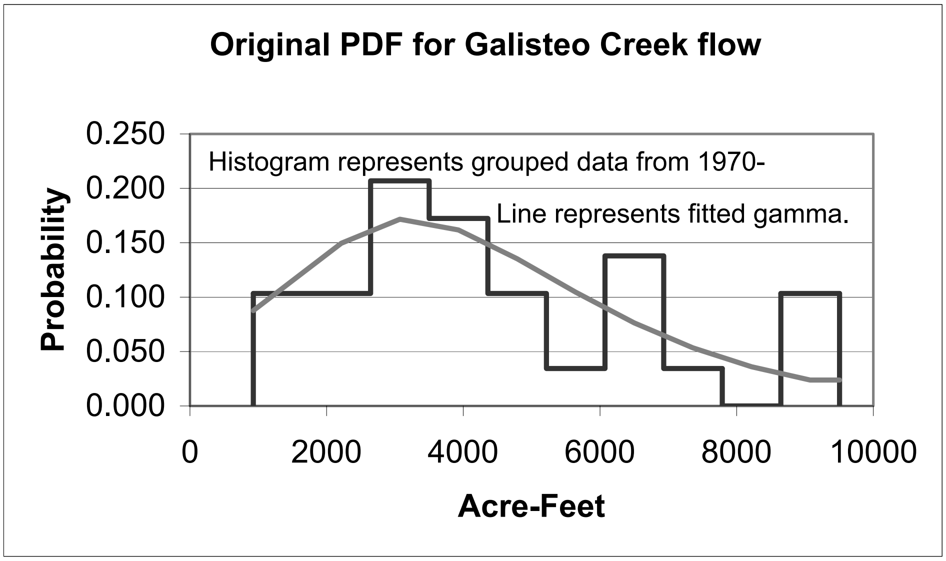
Time span	Flow (acre-feet/year)	Number of years
Average 1919-1998	1,058,102	80
Average 1896-1998	1,074,996	94
Average 1950-1998	963,576	49

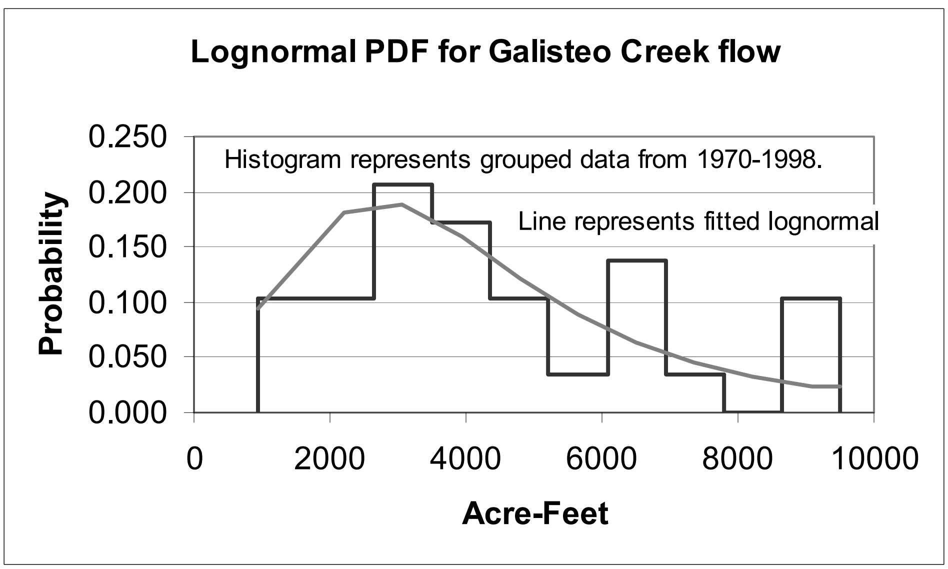
As can be seen in the probability density function graphs and probability plots, the difference between the originally chosen distributions and the lognormal are generally small for the Galisteo, Rio Puerco and Rio Salado data. Though the test statistics indicate that the original distribution generally fits the data slightly better, a lognormal could justifiably be substituted.

For the Jemez River and Otowi index flows, the difference between the original (beta) distribution and the lognormal distribution is larger. In particular, for both flows the lognormal function peaks at a point in the data histogram where the measured data indicates a gap. Consequently, the Chi-Square test statistic is significantly poorer for the lognormal distribution than the beta distribution, though it is still below the critical value. For the Otowi index flow, the chi-square test fails for the lognormal distribution, though it passes for the beta distribution. This is true regardless of the period of record used or the number of classes the data is divided into.

Overall, unless there is a compelling reason to replace the beta distributions used for the Jemez and Otowi index flows, SSPA is inclined to retain the original beta distributions used in the modeling since we believe these distributions best represent the measured data. For the Galisteo, Rio Puerco and Rio Salado flows, we are comfortable using either the originally chosen distributions or lognormal distributions.

Galisteo Creek PDF

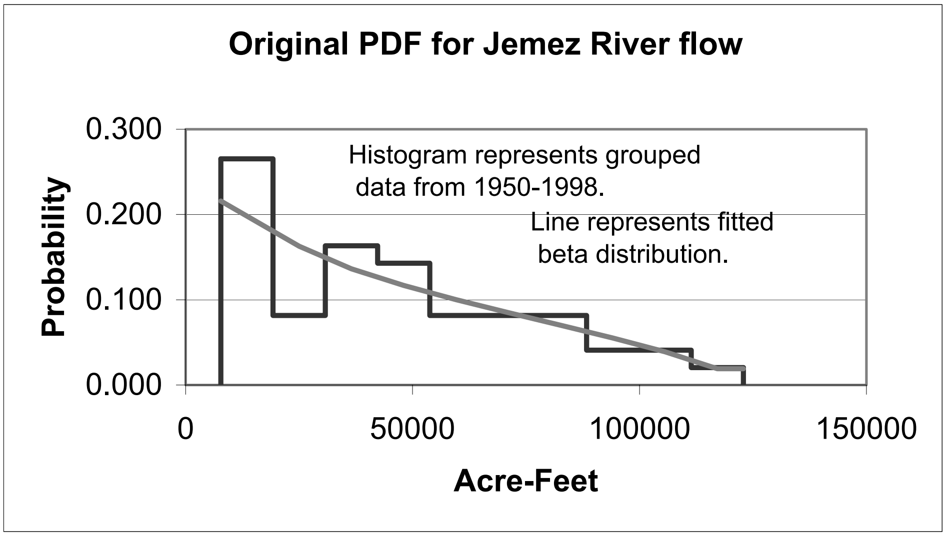


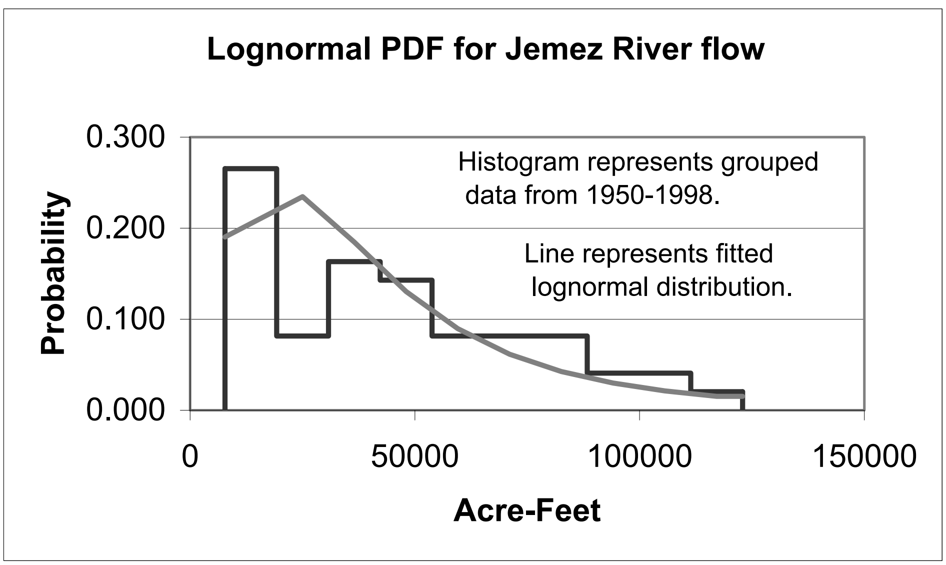


Statistics – critical values specified for α=0.05

	Chi-Square		Kolmogorov-Smirnov		Anderson-Darling	
	statistic	crit.value	statistic	crit.value	statistic	crit.value
Gamma	13.51	14.07	0.349	1.358	0.153	2.492
Lognormal	14.69	14.07	0.511	1.358	0.282	2.492

Jemez River PDF

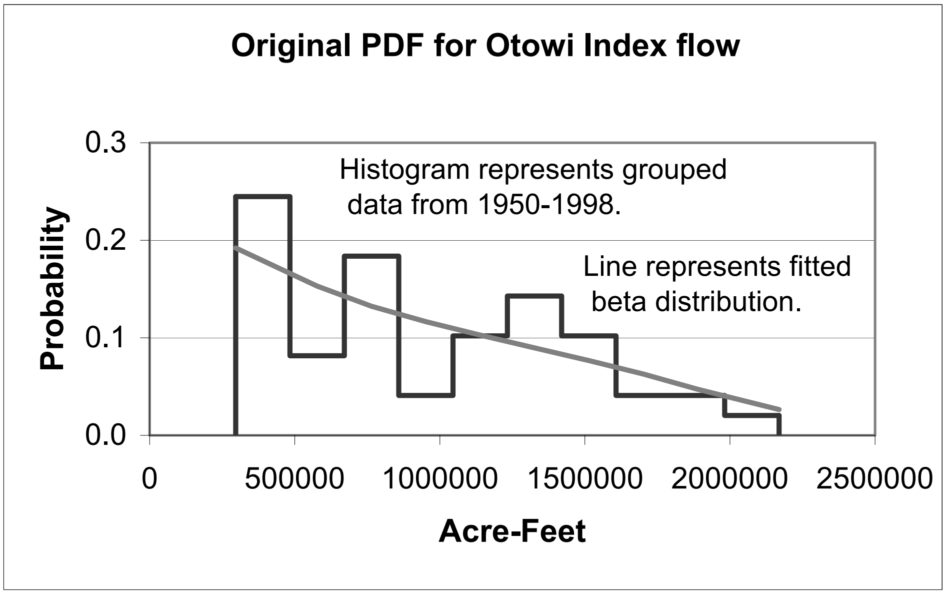


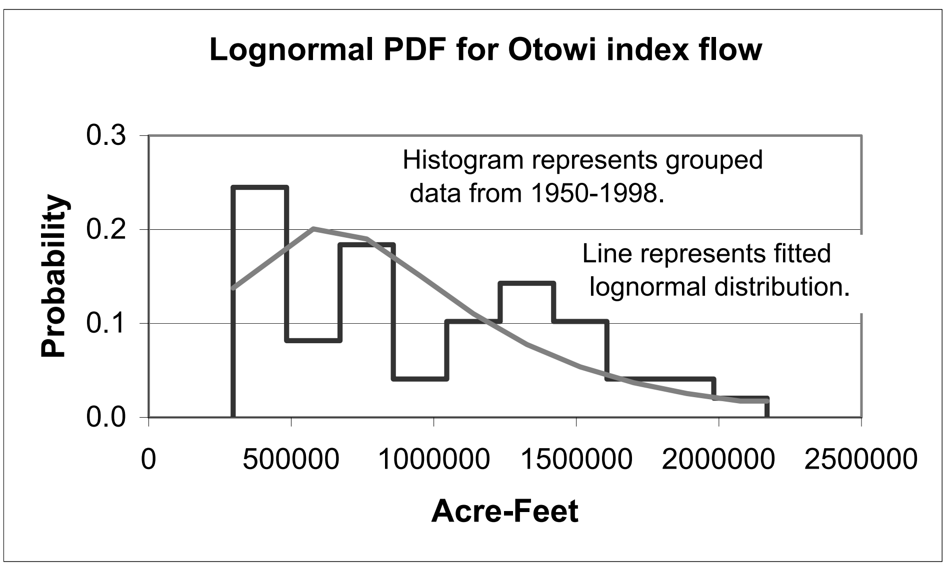


Statistics – critical values specified for α=0.05

	Chi-Square		Kolmogorov-Smirnov		Anderson-Darling	
	statistic	crit.value	statistic	crit.value	statistic	crit.value
Beta	3.63	14.07	0.44	1.358	0.410	2.492
Lognormal	10.41	14.07	0.62	1.358	0.568	2.492

Otowi Index flow PDF

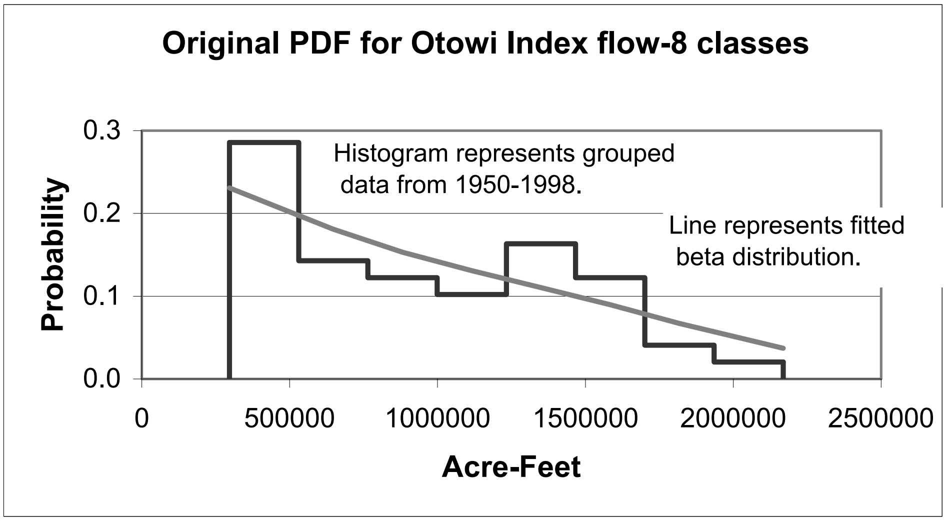


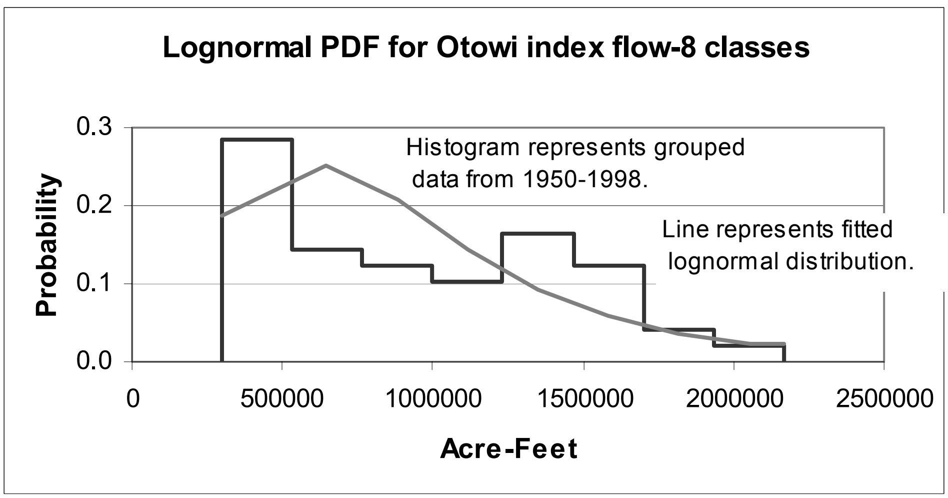


Statistics – critical values specified for α=0.05

	Chi-Square		Kolmogorov-Smirnov		Anderson-Darling	
	statistic	crit.value	statistic	crit.value	statistic	crit.value
Beta	8.31	14.07	0.55	1.358	0.464	2.492
Lognormal	18.15	14.07	0.98	1.358	1.053	2.492

Otowi Index flow PDF – 8 classes

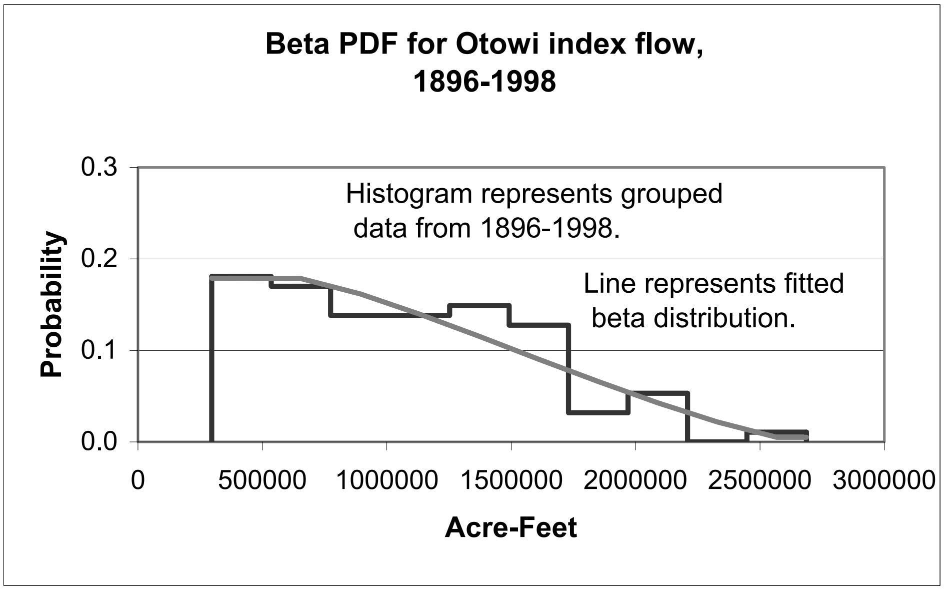


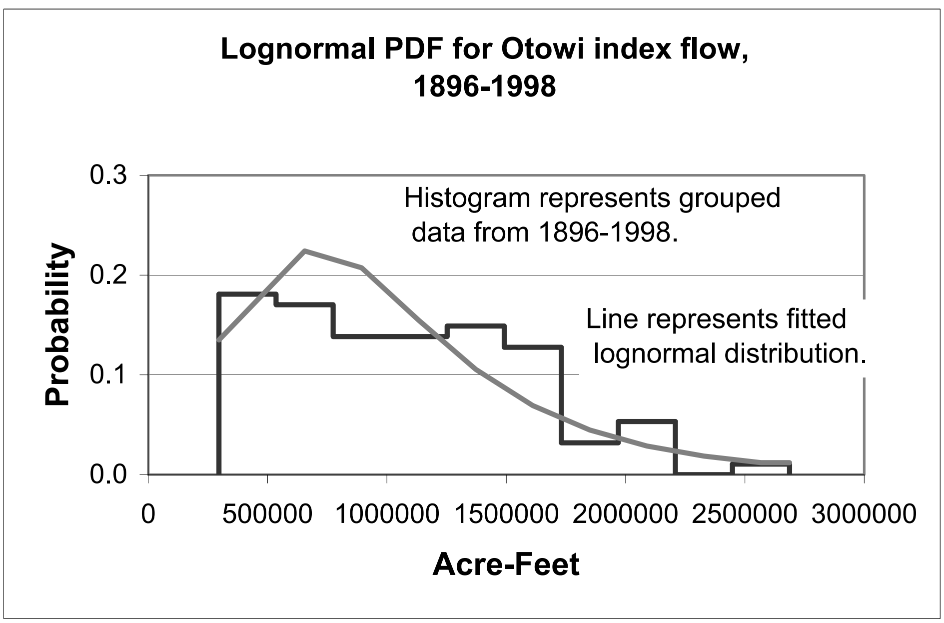


Statistics – critical values specified for $\alpha=0.05$

	Chi-Square		Kolmogorov-Smirnov		Anderson-Darling	
	statistic	crit.value	statistic	crit.value	statistic	crit.value
Beta	4.43	11.1	0.55	1.358	0.464	2.492
Lognormal	14.34	11.1	0.98	1.358	1.053	2.492

Otowi Index flow PDF – all data, 1896-1998

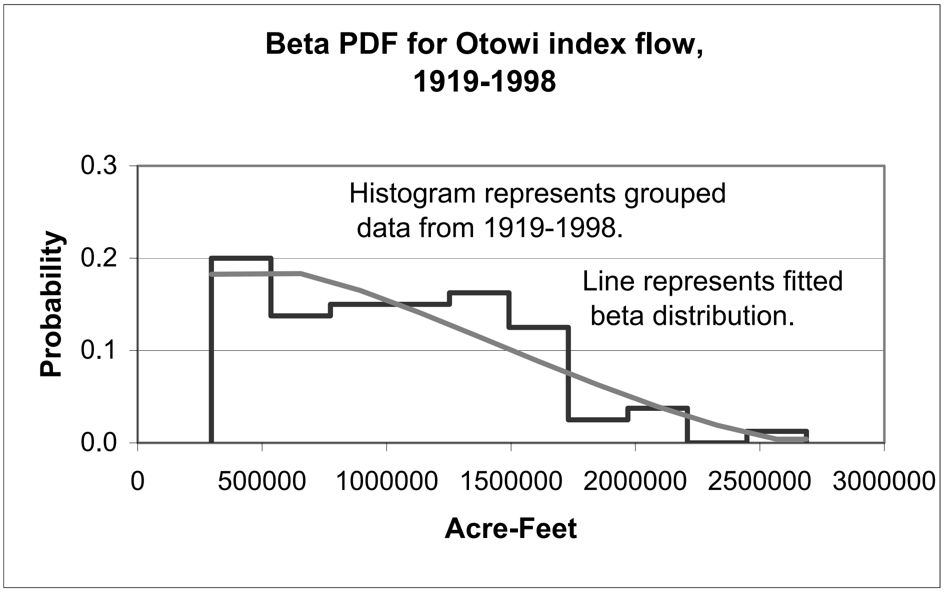


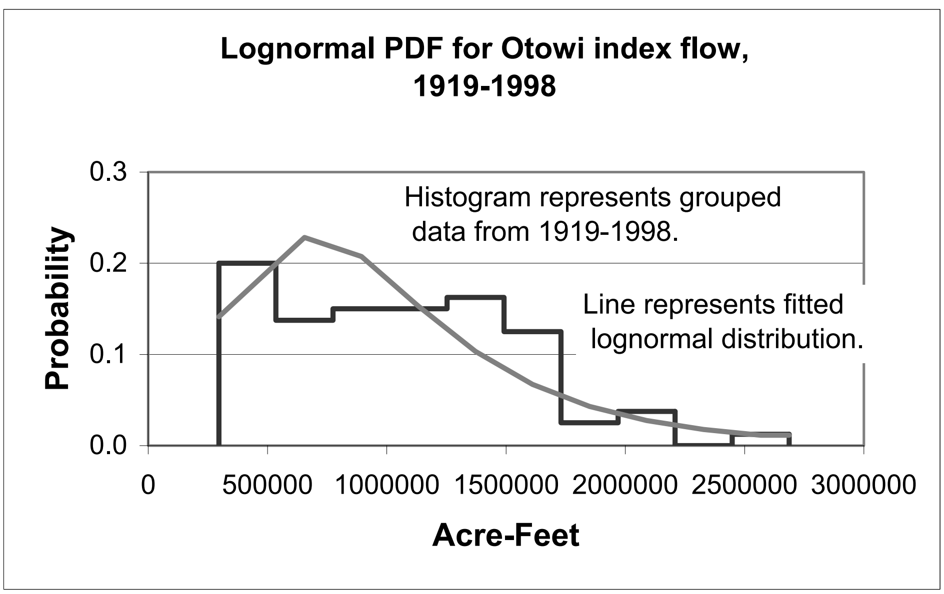


Statistics – critical values specified for α=0.05

	Chi-Square		Kolmogorov-Smirnov		Anderson-Darling	
	statistic	crit.value	statistic	crit.value	statistic	crit.value
Beta	7.20	14.07	0.562	1.358	0.45	2.492
Lognormal	15.97	14.07	1.037	1.358	1.22	2.492

Otowi Index flow PDF – 1919-1998

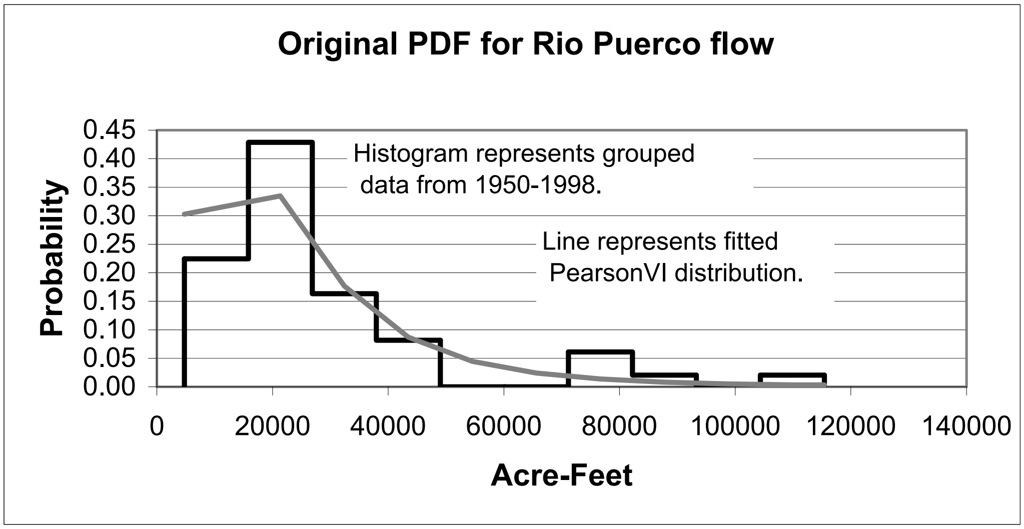


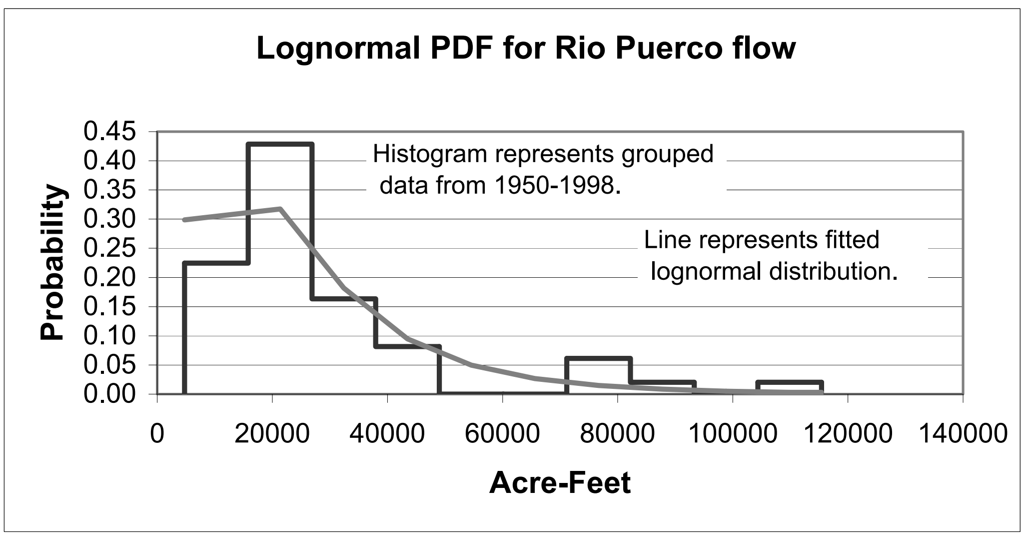


Statistics – critical values specified for α=0.05

	Chi-Square		Kolmogorov-Smirnov		Anderson-Darling	
	statistic	crit.value	statistic	crit.value	statistic	crit.value
Beta	8.80	14.07	0.65	1.358	0.58	2.492
Lognormal	15.70	14.07	1.06	1.358	1.25	2.492

Rio Puerco PDF

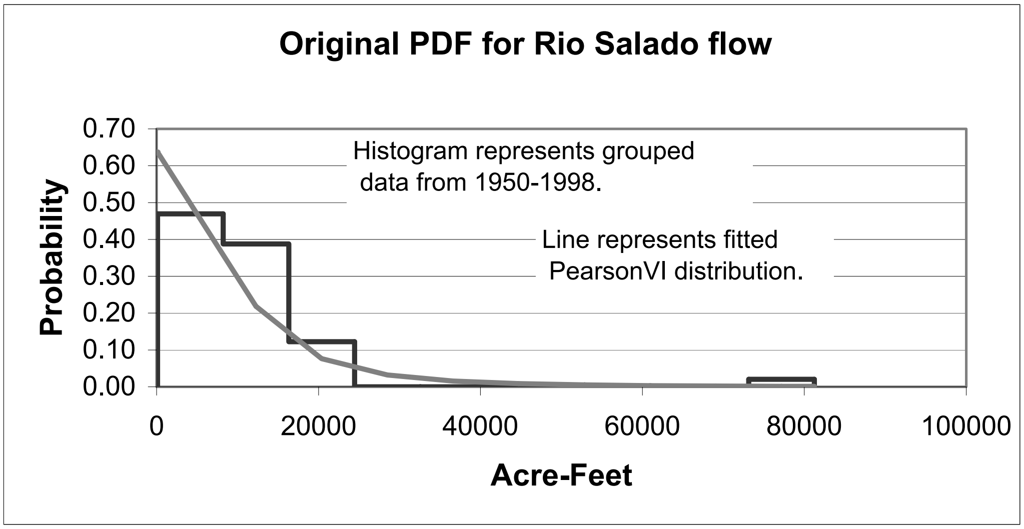


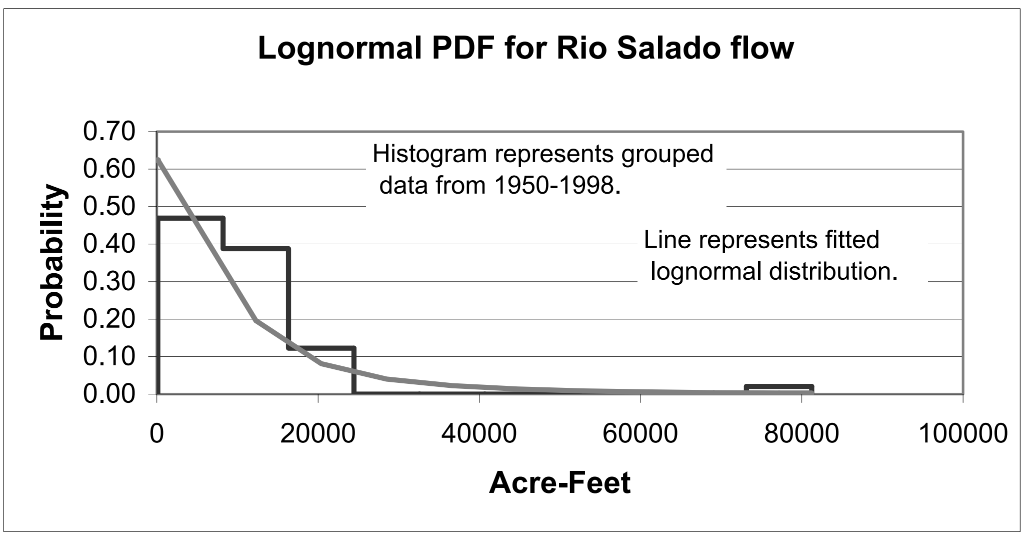


Statistics – critical values specified for α=0.05

	Chi-Square		Kolmogorov-Smirnov		Anderson-Darling	
	statistic	crit.value	statistic	crit.value	statistic	crit.value
PearsonVI	18.95	12.59	0.624	1.358	0.474	2.492
Lognormal	19.55	12.59	0.735	1.358	0.632	2.492

Rio Salado PDF

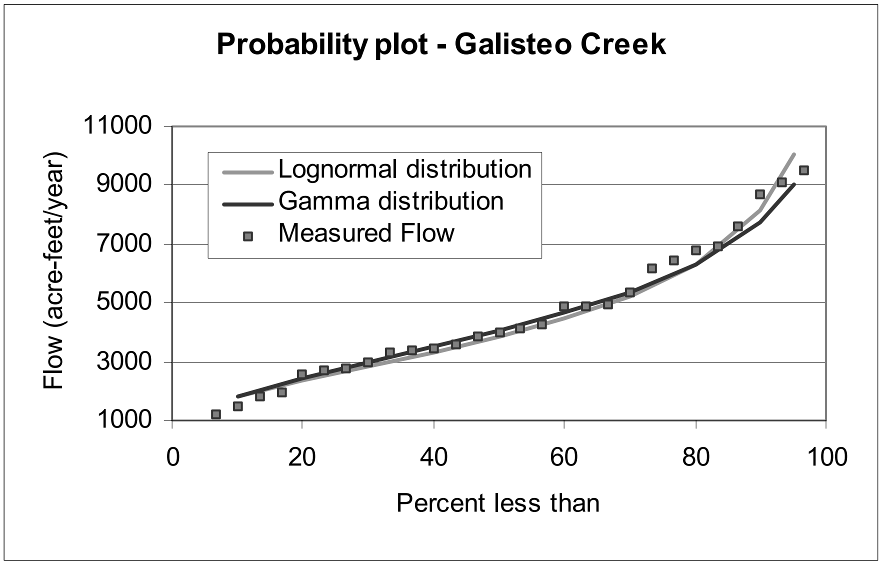


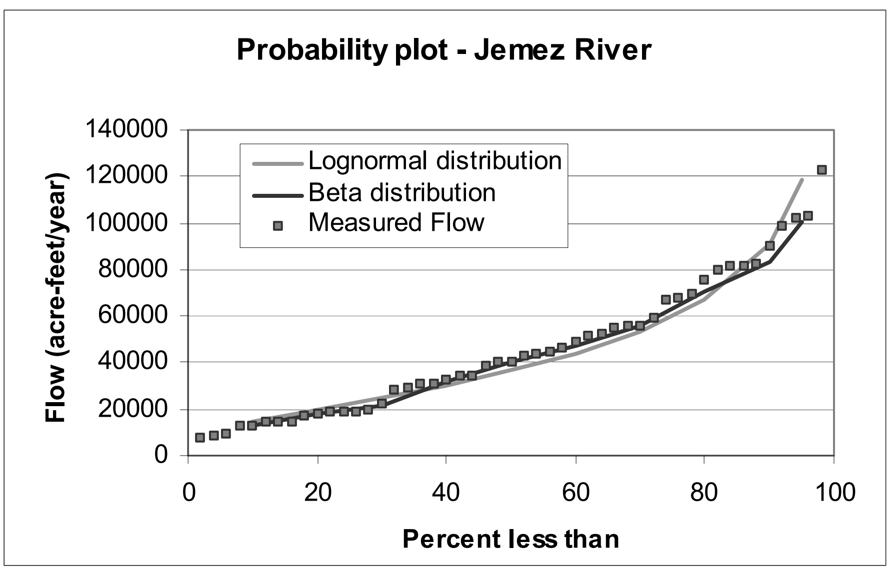


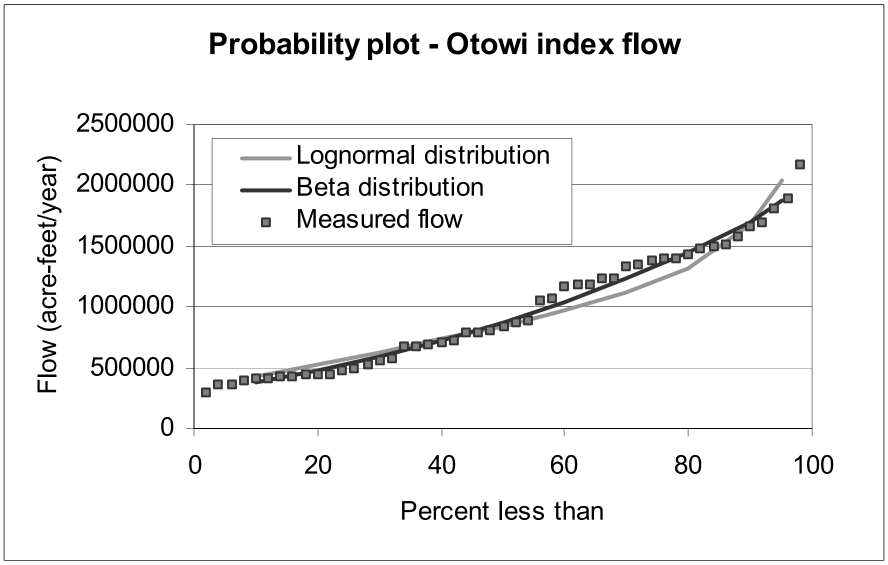
Statistics – critical values specified for α=0.05

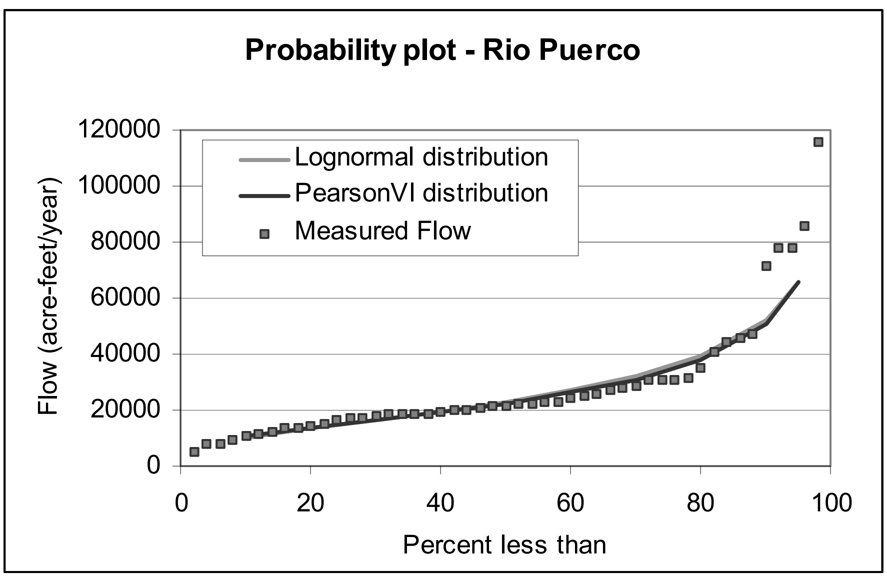
	Chi-Square		Kolmogorov-Smirnov		Anderson-Darling	
	statistic	crit.value	statistic	crit.value	statistic	crit.value
PearsonVI	23.52	12.59	1.063	1.358	1.575	2.492
Lognormal	21.13	12.59	1.398	1.358	2.186	2.492

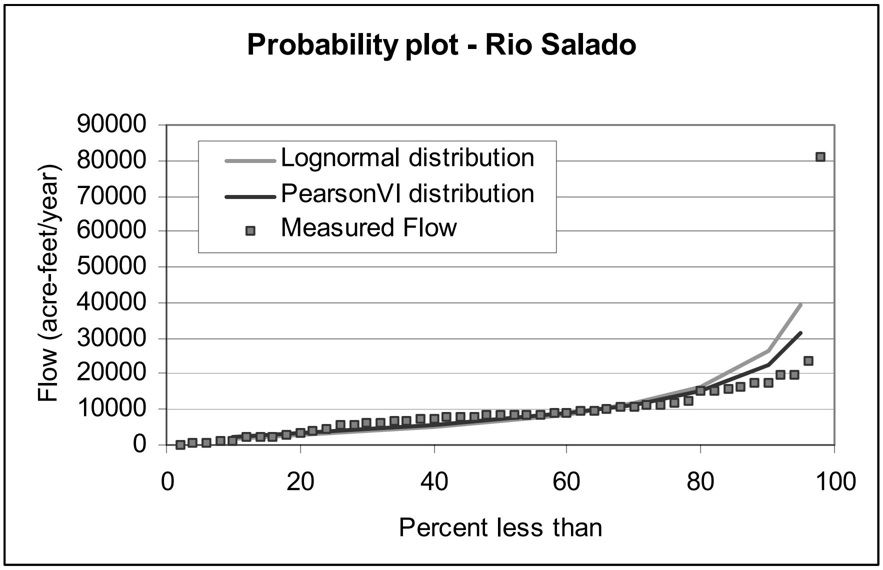
Rio Salado 1950-1984 measured data only:
PearsonVI(1.69,3.83,1.77e^4) CS 26.40 / KS 0.877 / AD 0.79
Lognormal(1.31e^4,2.58e^4) CS 19.08 / KS 1.110 / AD 1.21


Probability plot - Galisteo Creek
Lognormal distribution
Gamma distribution
Measured Flow
Flow (acre-feet/year)
11000
9000
7000
5000
3000
1000
0
20
40
60
80
100
Percent less than


Probability plot - Jemez River
Lognormal distribution
Beta distribution
Measured Flow
Flow (acre-feet/year)
140000
120000
100000
80000
60000
40000
20000
0
0
20
40
60
80
100
Percent less than


Probability plot - Otowi index flow
Lognormal distribution
Beta distribution
Measured flow
Flow (acre-feet/year)
2500000
2000000
1500000
1000000
500000
0
0
20
40
60
80
100
Percent less than


Probability plot - Rio Puerco
Lognormal distribution
PearsonVI distribution
Measured Flow
Flow (acre-feet/year)
120000
100000
80000
60000
40000
20000
0
0
20
40
60
80
100
Percent less than


Probability plot - Rio Salado
Flow (acre-feet/year)
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
Lognormal distribution
PearsonVI distribution
Measured Flow
0
20
40
60
80
100
Percent less than