THE CITY OF CALGARY

amec

Frequency Analysis Procedure for Stormwater Design


Submitted to:
The City of Calgary

Submitted by:
AMEC Environment & Infrastructure

April 2014
CW2138

amec.com

FREQUENCY ANALYSIS PROCEDURES
FOR
STORMWATER DESIGN
MANUAL

Submitted to:
The City of Calgary
Calgary, Alberta

Submitted by:
AMEC Environment & Infrastructure
Calgary, Alberta

April 2014

CW2138

Acknowledgements

The authors would like to thank Mr. Bert van Duin, project manager with the City of Calgary, for his invaluable guidance and review comments. The authors would also like to thank Dr. Ed Watt for his insightful review comments on the draft version of this manual.

TABLE OF CONTENTS

PAGE

TABLE OF CONTENTS (*cont'd*)

PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

1.0 INTRODUCTION

1.1 Background

At present, a wide variety of methods are used by the design community in conducting frequency analyses for sizing stormwater infrastructure within The City of Calgary (The City). This manual outlines procedures to be followed when conducting frequency analyses. These procedures address the review and analysis of datasets and how the most suitable probability distribution should be selected. This manual also documents best professional practice and provides worked examples of data analysis procedures.

1.2 Objectives

The objective of frequency analysis is to use an available data series to predict probabilities of occurrence of hydrologic phenomena. The most common hydrologic variables that are subjected to frequency analysis are discharge, precipitation, and volume.

The objective of this manual is to present the theory, assumptions, procedures and guidelines in frequency analysis estimation for use in the design of drainage infrastructure. The general approaches to be used in the design of these facilities are presented in The City's *Stormwater Management and Design Manual*[1].

The audience for this manual is the stormwater professional and/or design engineer with a basic knowledge of statistics. Sources of academic discussion of the topics discussed can be found in the reference section. Due to the nature of the subject, engineering judgement should always be used when following any written guidelines. Preferably, the results of a frequency analysis should always be reviewed by another hydrologist.

1.3 Limitations and Scope

This manual is intended as a practical guide for the practising engineer. It is not intended to be a dissertation on all that there is to know about frequency analysis.

Related topics which are not within the scope of this document are:

- Data in-filling and temporal resolution of data series;
- Flood index methods and regional analyses;
- Historical floods of record in populations and the methods to extend data series;
- Methods to handle significant data gaps; and
- Data quality assurance.

[1] The City of Calgary *Stormwater Management and Design Manual* can be accessed at: http://www.calgary.ca/PDA/DBA/Documents/urban_development/bulletins/2011-stormwater-management-and-Design.pdf

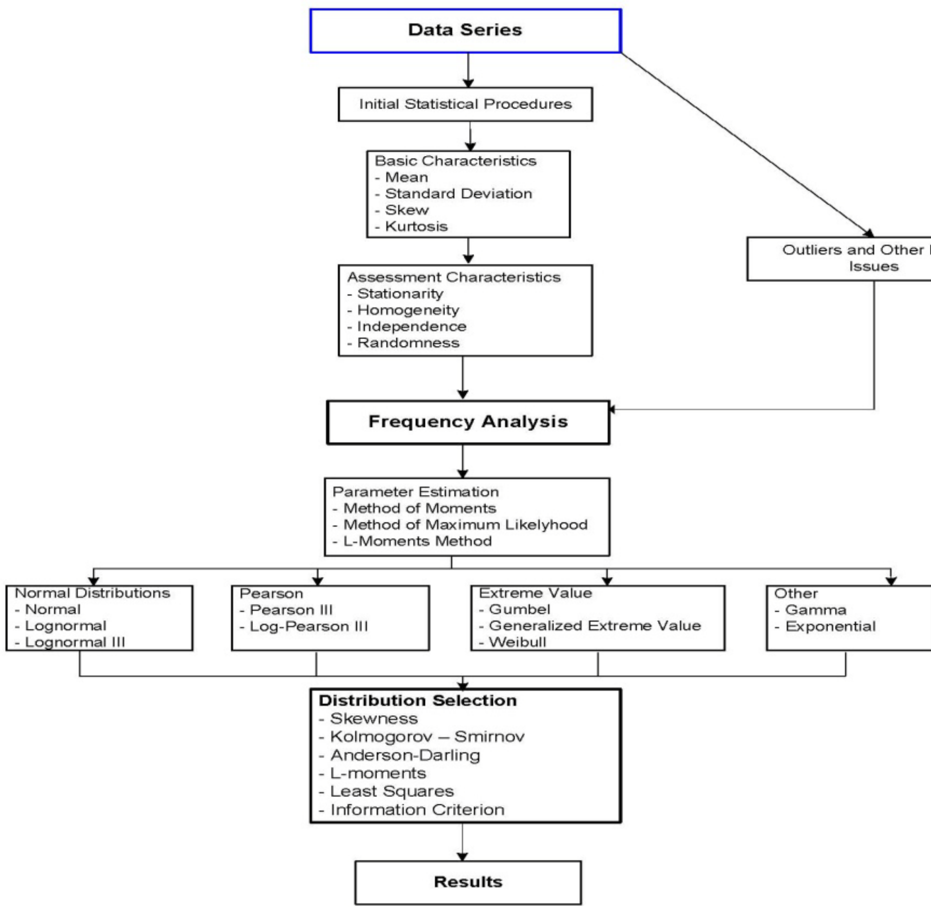
1.4 Report Outline and Approach

This manual has the following three main chapters:

1. Data Series;
2. Frequency Analysis; and
3. Distribution Selection

The following flow chart illustrates the topics addressed within each chapter. It also provides the reader with a reference point as to where the various topics are step-wise within the overall procedure. Each Chapter is preceded by a flow chart depicting where the topics to be discussed fit within the overall sequence, as illustrated on **Figure 1.1**.

Figure 1.1 Overall Flood Frequency Analysis Procedures

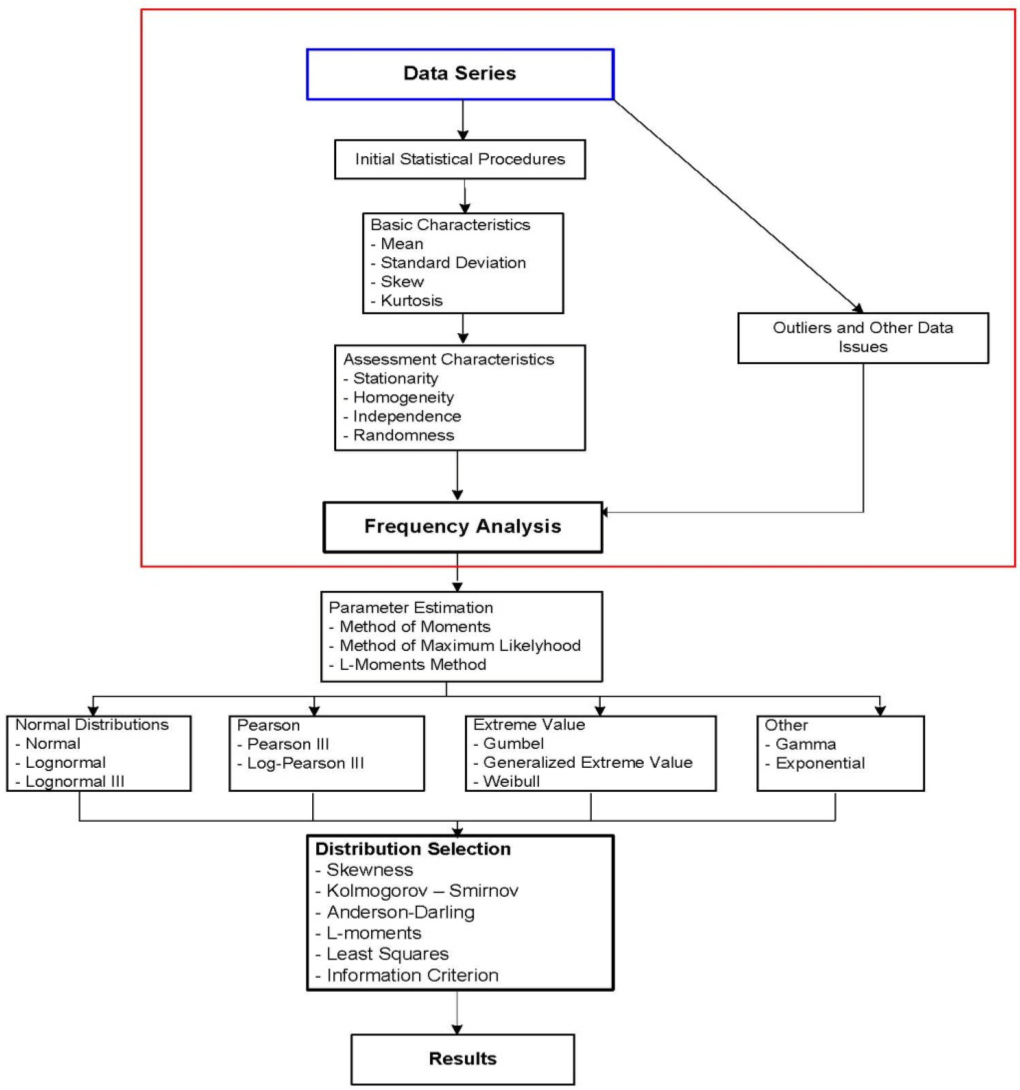


2.0 DATA SERIES

2.1 Background

This section presents a discussion of procedures that should be followed to understand the characteristics of the data series being analyzed. These procedures allow the practitioner to understand the statistical characteristics of the data series and allow for the determination of the suitability of the data series for frequency analysis. **Figure 2.1** illustrates the initial statistical procedures that are discussed in this chapter.

Figure 2.1 Analysis of Data Series



Understanding the characteristics of a data series is an important part of understanding the results of the resultant frequency analysis. A data series can take several different forms depending on the reported parameters. For example, a data series of pond volumes might represent maximum or minimum annual, monthly, seasonal or daily, active or total pond volume. Precipitation data may document rainfall, snowfall or total precipitation with maximum, minimum or mean values for hourly, daily, monthly, seasonal or annual duration.

Hydrologic frequency analysis involves the use of a specific time period or interval within which a hydrologic event can occur. Since hydrologic events vary from season to season, but commonly have a repeating pattern over the hydrologic year, the most common interval is one year. Other time intervals can be selected, such as one season, one month, one day, etc. The probability of occurrence of an event (for example a flood) can then be expressed as a "once in x intervals" event. Thus, if the interval is 1 year, and an event magnitude has an estimated probability of occurrence of say 0.01, then the event probability can be expressed as a "1 in 100-year" event. Note that for this example, the data series underlying the probability estimate must be a series of annual events (one event per annual interval).

Similarly, if the interval is one month, and an event magnitude has an estimated probability of occurrence of say 0.01, then the event probability can be expressed as a "1 in 100-month" event. For this example, the data series underlying the probability estimate must be a series of monthly events (one event per month interval. Only certain months of the year may be valid for inclusion in such an analysis, based on the hydrologic parameter being considered. For example, rainfall would only occur in non-winter months.

A thorough understanding of the input data series and the hydrologic processes governing the data series in question is important for appropriately interpreting the results of the frequency analysis.

2.2 Characteristics of the Data Series

Data series characteristics provide useful information and can be used for determining the appropriate successive steps in the overall frequency analysis procedure.

2.2.1 Basic Characteristics

- **Exceedance probability**, P, is the probability that an event of a given magnitude will be equalled or exceeded in a given period or interval of time. The interval is typically (but not necessarily) one year.
- **Return period**, T, of an event is the long-term average recurrence interval of an event of a given magnitude. For example an event with a return period of 100 years will occur, on average, once in 100 years. The return period equals the reciprocal of the **exceedance probability:**

$$T=1/P$$

Thus an event having an exceedance probability of 0.01 has a return period T of 100 years. Alternatively this can be expressed as a 1:100-year event.

- **Quantile** (or T-year event) is the magnitude corresponding to a particular return period. This may be estimated from a single station analysis, a regional analysis or a combination of both.
- The sample estimate of the **mean** is the average of the sample data:

$$Mean = \bar{x} = \frac{1}{n}\sum_{i=1}^{i=n} x_i$$

- The **standard deviation**, S, is a measure of variability having the same dimensions as the data. Standard deviation is the average of the absolute differences from the mean:

$$Standard\ Deviation = S = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

- **Skewness**, G, is a measure of the symmetry of a distribution about the mean. A negative skew indicates that the tail on the left side of the probability density function is longer or fatter than the tail on the right side. A positive skew indicates that the right tail is longer or fatter than the left tail. The equation to compute skew is:

$$Skew = G = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{S}\right)^3$$

- **Kurtosis**, K, refers to the extent of peakedness or flatness of a probability distribution in comparison with the normal probability distribution (Kurtosis for a normal distribution is 3):

$$Kurtosis = K = \frac{1}{nS^4}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^4$$

A higher kurtosis indicates a peaked distribution, while a distribution with a lower kurtosis will be flatter.

An example for calculating the basic characteristics described above can be found using the sample data in **Appendix A**.

2.2.2 Necessary Characteristics for Frequency Analysis

2.2.2.1 Stationarity

Valid frequency analysis relies on the data series to be stationary. A stationary data series is one that, excluding random fluctuations, is invariant with respect to time. The most recognized types of non-stationarity include jumps, trends and cycles.

Trends in runoff discharges are most often due to gradual changes in land use. Trends in precipitation may result from the effects of climate change. The Spearman rank order correlation coefficient can be used to detect trends in the data set. If the test for trend shows a significant likelihood of a trend, efforts should be made to establish the cause of the trend. Kaliq et al. (2006) review several approaches for the frequency analysis of non-stationary observations. These methods include extremal, r-largest, peaks-over-threshold, time-varying moments, pooled frequency analysis, local likelihood, and quantile regression.

In performing a frequency analysis of a data series determined to be non-stationary, it is important to consider the temporal scale of the processes that create the variations in the data values (e.g., climate change, Pacific Decadal Oscillation, etc.) versus the temporal scale of the quantile estimates (e.g., design return period). If the temporal scale of the changes is much larger than the temporal scale of quantile estimates (e.g., climate change versus annual rainfall), it would be preferable to incorporate the non-stationary climate effects into future estimations of annual rainfall. On the other hand, if one were determining extreme flood frequencies, it would not be necessary to include the El Niño - Southern Oscillation in the analysis because it is already part of the natural variability.

Jumps in stream discharge data are normally due to an abrupt change in a basin or river system such as the construction of a dam. Clarke et al. (2011) states that jumps in precipitation data may be due to change in instrument location, instrument type, and measuring protocol. Jumps can be usually observed from the graphical representation of the data, the Mann-Whitney test for jump or the Wald-Wolfowitz test for jump. If the test shows that the jump is statistically significant (see **Section 2.2.2.6**), and the jump is well understood, it is possible to adjust the data to remove the jump (e.g., subtract the magnitude of the jump from the data values for all data points following the occurrence of the jump) and then undertake a frequency analysis of the new dataset (containing the original data up to the time of the jump, and the adjusted data for the period including and after the jump). The results of the frequency analysis should be studied and understood to ensure that it is representative of the design conditions.

A summary for the various tests for stationarity can be found in **Appendix A**. The full calculations for these tests can be found in the frequency analysis spreadsheet that accompanies this document.

2.2.2.2 Homogeneity

Valid frequency analysis requires the data series to be homogeneous. A homogenous data series originates from a single population. An example of a potentially non-homogeneous data series is peak stream discharges. Spring discharges resulting from snowmelt might not belong to the same population (statistically) as those peak discharges that occur in the summer or fall months, which are due to precipitation storm events.

It is possible that even rainfall data can be non-homogeneous. Rainfall events can be caused by different drivers. Cyclonic storms are caused by warm, moisture-laden air, generally cover more than 2,000 km^2, and last more than 24 hours. Convective storms involve thunderstorm cells, generally cover less than 2,000 km^2, and last less than 24 hours. In rainfall datasets, especially for gauges that report only daily values, it is not always evident which type of storm has occurred. It is important to establish homogeneity in a dataset in order to ensure that the data being used is all of the same population and appropriate for frequency analysis.

To assess the homogeneity of the data series a histogram of the data series should be plotted. It may be useful to determine whether the data series can be used to develop multiple populations for use in frequency analysis (discussion of stream discharge peaks due to snowmelt and peaks due to rainfall; Alberta Transportation 2004) and design consideration given to each population. The most common numerical test for homogeneity is the Mann-Whitney nonparametric test, which checks whether the means of the subsamples differ significantly for chosen levels of significance. Also note that a data series comprising of mixed populations might have a large skew coefficient.

A summary for the test for homogeneity can be found in **Appendix B**. The full calculations for these tests can be found in the frequency analysis spreadsheet.

<u>Dealing with a non-homogeneous data series</u>

For non-homogeneous data series, a common approach is to divide the sample into a series for each of the causes. The frequency curves are then obtained for each separate sub-set. Then the two separate frequency curves can be joined using the following equation for probability (U.S. Army 1982):

$$P_x = P_a + P_b - (P_a \times P_b)$$

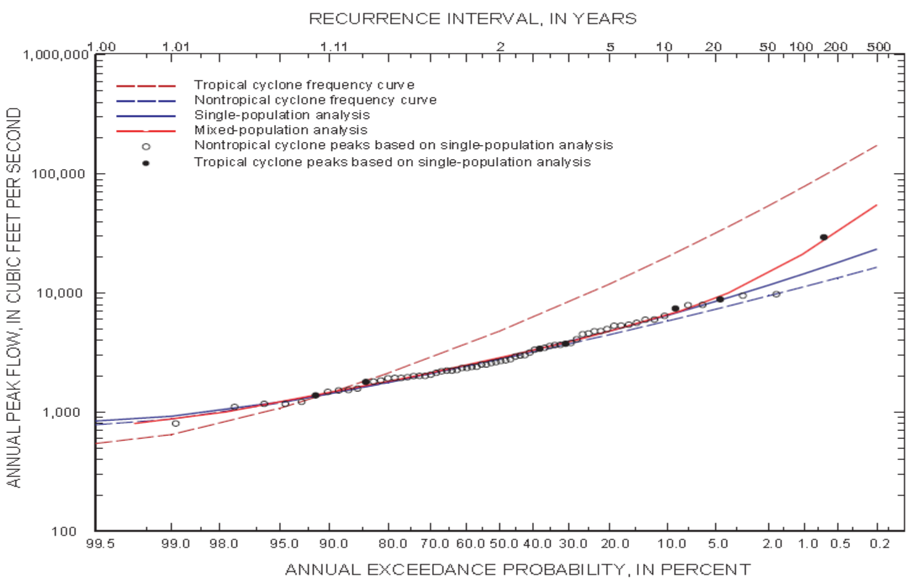
where:
P_x is the probability of an event of magnitude *x*;
P_a is the probability of an event of magnitude *x* occurring due to cause *a*;
P_b is the probability of an event of magnitude *x* occurring due to cause *b*.

The above equation can also be expressed in terms of return periods, T (as shown below), using the same nomenclature for event magnitude, *x*, and causes, *a* and *b*, as discussed in Watt *et al.* (1989).

$$T_x = \frac{T_a \times T_b}{(T_a + T_b - 1)}$$

Figure 2.2 is an example of frequency curve derivation from a mixed population containing data from events caused by tropical cyclones and non-tropical cyclones. Note that the solid blue line depicting the result of a single population analysis does not correctly capture the two distinct causes (dashed red line for tropical cyclones and the dashed blue line for non-tropical cyclones). Extraction of probabilities from each of the dashed lines for a given annual maximum discharge allows the above formulas to produce the solid red line for the mixed population.

Figure 2.2 Frequency Plot for a Mixed Population



From USGS 2003

2.2.2.3 Independence

An independent data series is one where each data point is unaffected by the one before it; even if events are random, they may not be independent. Large natural storages may cause high flows to follow high flows and low flows to follow low flows. Snowmelt may create antecedent moisture conditions that increase runoff from spring and summer rainfall even after the snow cover is gone. Early season peak flows may occur when rainfall occurs on snowpack. Dependence can vary with the interval between successive elements of the series. Dependence among successive daily values tends to be strong, while dependence between annual maximum values is typically weaker. In some cases; however, there may be significant dependence even between annual maximum values (e.g., in the case of very large storages such as evaporation ponds). Two events can be considered independent only if the possibility of occurrence of either is unaffected by the occurrence of the other. Data series that are not independent require special consideration when doing a frequency analysis. Hydrologic variables that may commonly exhibit dependence are discharge and pond volumes.

The extent and scale of the dependence should be determined if dependence is caused by regulation, and this regulation is thoroughly understood, it might be possible to de-regulate the data series prior to frequency analysis. Frequency analysis can then be undertaken on the de-regulated set, and the regulation procedures re-applied to the predicted design events to allow for accurate event prediction.

In a time series independence can be measured by the significance of the correlation coefficient. The correlation coefficient can be calculated both between a data point and the point following (first order correlation) or between a data point and the second point after it (second order correlation). In the case where the first order coefficient is found to be significant, but the second order coefficient is not, the data series can be transformed to an independent data series for analysis as discussed below. In the case where both the first and second order coefficients are found to be significant then another method such as Monte Carlo simulation is necessary.

A summary for the tests for independence for both example data series can be found in **Appendices A** and **B**. The full calculations for these tests can be found in the frequency analysis spreadsheet.

First Order Correlation

Kaliq *et al.* (2006) discuss a method that may be applied to a data series to remove the effect of first order correlation and obtain a data series containing independent observations.
The suggested de-correlation technique is based on the notion that if there are sufficient physical reasons to assume that the current year's observation is dependent on last year's observations then the observations can be de-correlated by eliminating the correlation.
The following formula can be used to totally or partly eliminate correlation between observations by transforming the original data set.

$$y_t' = y_t - \hat{r}_1 y_{t-1}$$

where:
y_t' is the decorrelated value of the observation at time *t*;
$\hat{r}_1$ is the estimated autocorrelation coefficient of order 1 from the sample y_1, y_2, …y_n; and
y_t and y_{t-1} are observations at time points *t* and *t-1*, respectively.

The set of transformed values can be used to fit a frequency distribution. Once that is done, the quantiles can be re-transformed using the following suggested procedure:

$$y_{T} = y'_{T} + (\hat{r}_1 * y_{50})$$

where:
y'_T is the quantile value of the transformed data series for recurrence interval, *T*;
y_T is the inverted (re-transformed) quantile value for recurrence interval, *T*; and
y_{50} is the median value of the original series of observations.

The results of the analysis for the data series in **Appendix B** are presented on **Figure 2.3** and **Figure 2.4**. **Figure 2.3** illustrates the Pearson III distribution fitted to the dependent data series, without consideration for whether or not the series is independent. This approach is incorrect.

For the example plotted in **Figure 2.3**, transformation of the data series to remove serial correlation allows a frequency analysis to be performed on the transformed data series. A check of the transformed data series should indicate that the serial correlation had been successfully removed. The result of the distribution fitting to the transformed data series listed in **Appendix B** is illustrated on **Figure 2.4**.

Figure 2.3 First Order Correlated Data and Best Fit Curve

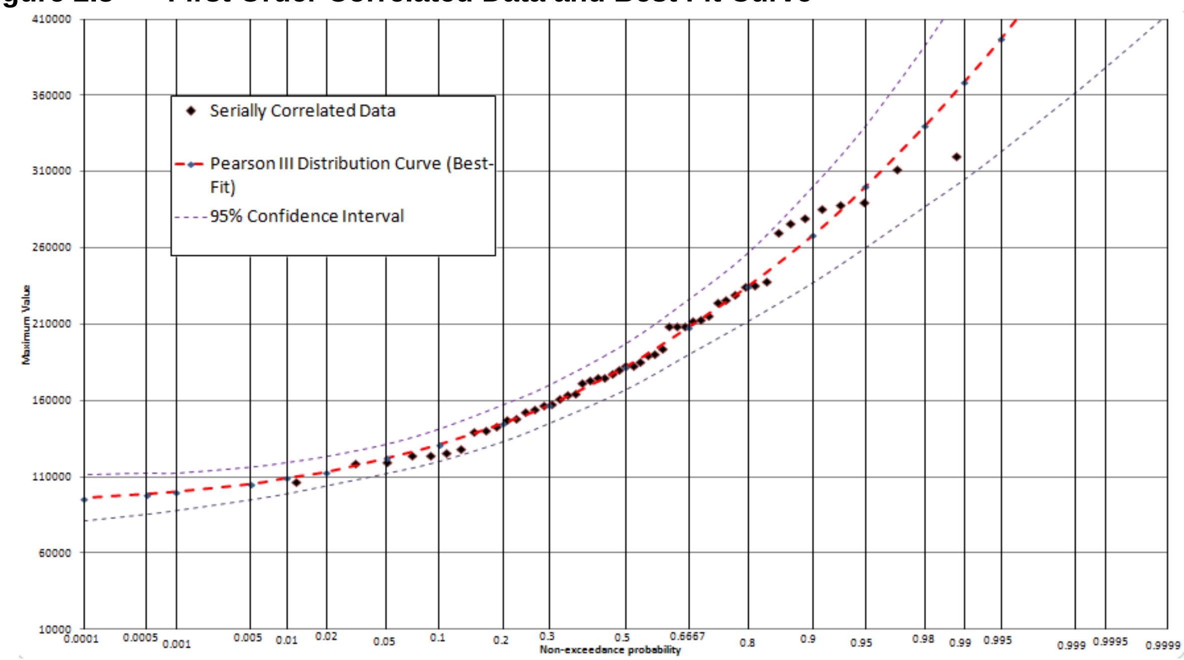
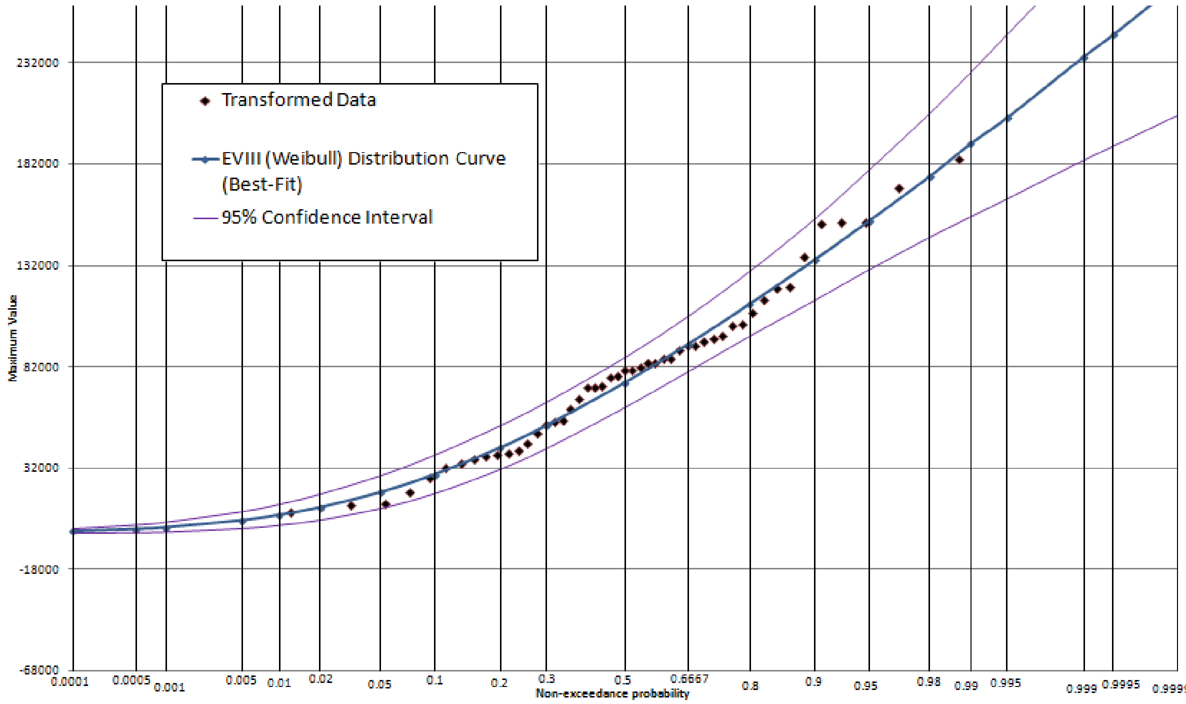


Figure 2.4 Transformed Data with Best Fit Distribution Curve



Note: Using the transformed data series results in a different probability distribution being chosen as best fit. Using a different probability distribution for the transformed data series results in quantile estimates that are markedly different from those determined for the original dependent data series.

Monte Carlo Simulation

The characteristics of a dependent data series can be modelled to produce a large sample of synthetic data, from which frequency estimates can be determined. The following formulas can be used to develop a synthetic data series. For the first value:

$$x_1 = \bar{x} + (s \times \xi)$$

where:
x_1 is the first value of the synthetic data series;
$\bar{x}$ is the mean of the original data series;
s is the standard deviation of the original data series; and
ξ is a random number with mean of zero and standard deviation of unity.

The following formula is used to compute all remaining values (i = 2, 3, 4, …n):

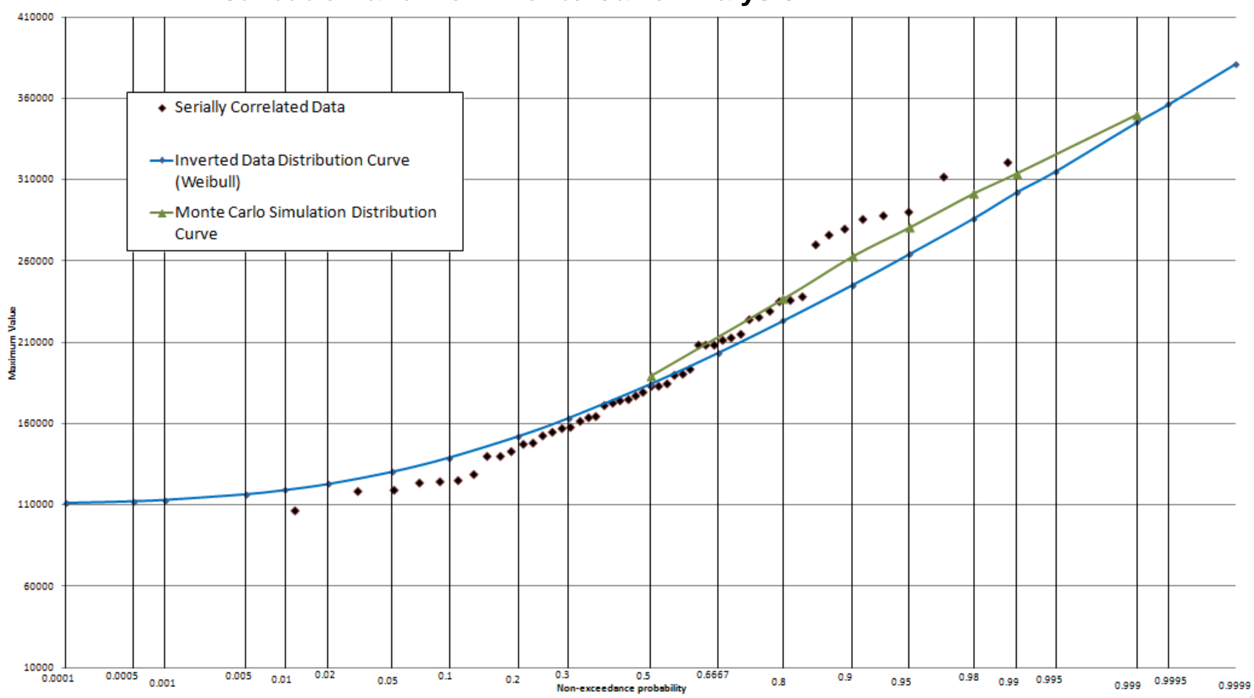
$$x_i = \bar{x} + r_1 \times \left[(x_{i-1} - \bar{x}) + \sqrt{1 - {r_1}^2} \times s \times \xi\right]$$

where:
r_1 is order 1 serial correlation coefficient.

A large sample (greater than 100 values) of synthetic data can be readily generated using this procedure. Once that is done, values can be extracted from the synthetic data series to estimate frequency values of interest. For instance, if a series of 1,000 synthetic data are generated, the data can be ranked from highest to lowest and the tenth highest value selected. This value has a probability of being equalled or exceeded of 1% (10/1,000 = 0.01), and hence provides an estimate of the 1:100-year return period event. Similarly the fiftieth highest value would represent the 1:20-year return period event. An example of the results of a Monte Carlo simulation for a dependent data series is provided in **Appendix B**.

Figure 2.5 shows frequency estimates derived from two methods: 1) inverted quantiles from the frequency distribution presented on **Figure 2.4**; and 2) Monte Carlo simulation. Comparison of the two lines on **Figure 2.5** indicates that although there is some variation in the frequency estimates, the two lines are parallel. Further, the estimate of the 1:100-year event (in the range of 30,000 to 31,000 are consistently less than the (incorrect) estimate that would have been derived from the dependent series illustrated on **Figure 2.3**.

Figure 2.5 Serially Correlated Data Showing Frequency Estimates from Inverted Distribution and from Monte Carlo Analysis



2.2.2.4 Randomness

Valid frequency analysis relies on the data series to be random. In a hydrologic context, randomness means that the fluctuations of the variable arise from natural causes, and that the data series is not the result of human intervention. Precipitation data can usually be assumed to be random. Ponds that operate without active human control can typically be considered to be random, but ponds where water is removed or added by control structures when it reaches a certain threshold do not satisfy the randomness criteria.

No suitable test for randomness in hydrologic set is available, so care must be taken in assessing the provenance of a data set.

2.2.2.5 Tests

The following tests are typically used to determine the characteristics of a data set prior to conducting a frequency analysis.

Stationarity

1. Spearman rank order correlation coefficient test for trend (NERC 1975);
2. Mann-Whitney test for jump (Mann and Whitney 1947); and
3. Wald-Wolfowitz test for jump (Siegel 1956).

Homogeneity

1. Mann-Whitney Test (Mann and Whitney 1947); and
2. Terry Test (Terry 1952, Kite 1977).

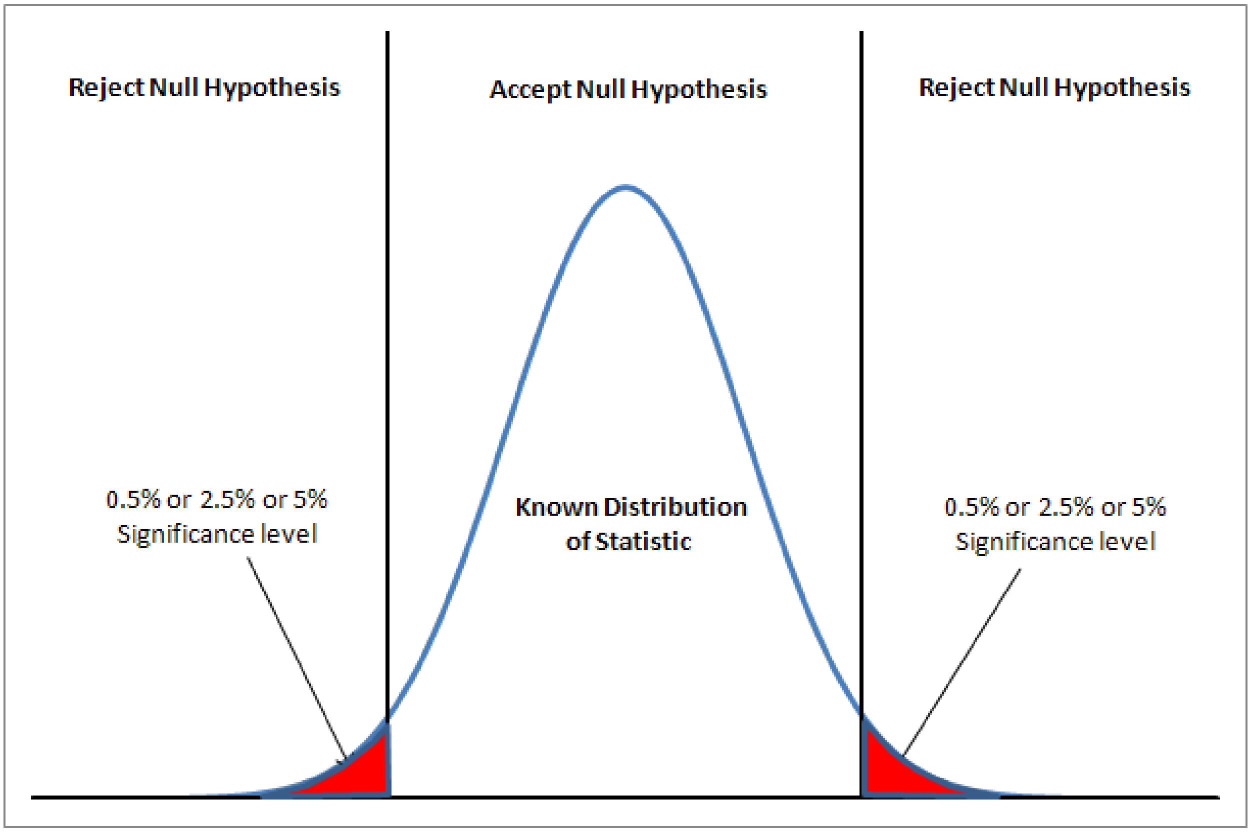
Independence

1. Spearman rank order correlation coefficient test for trend (NERC 1975); and
2. Wald and Wolfowitz (1943) test as described and applied in Bobee and Robitaille (1977).

2.2.2.6 Test Significance Levels

Statistical tests performed on the data series, as well as numerical goodness of fit tests performed on the results of the frequency analyses are subject to significance levels. Significance levels indicate the assessment severity of any particular test. **Figure 2.6** shows a distribution plot of a statistic calculated using a two-tailed hypothesis test. The area under the graph, which is shaded in white, marks the range of statistical probabilities under which the null hypothesis is accepted. The areas outside of a critical value are shaded in red and signify the probabilities for which the hypothesis is rejected. The significance level (alpha) is defined as the percent area where the hypothesis is rejected. Typical values for significance levels are 1%, 5%, and 10%. A 10% significance level is more stringent as it allows the value of a statistic to deviate from the mean within a range of 90% of the probabilities. A 1% significance level, on the other hand, allows the value of a statistic to deviate from the mean within a range of 99% of the probabilities and is therefore less stringent.

Figure 2.6 Probability Distribution Graph of a Two-tailed Hypothesis Test

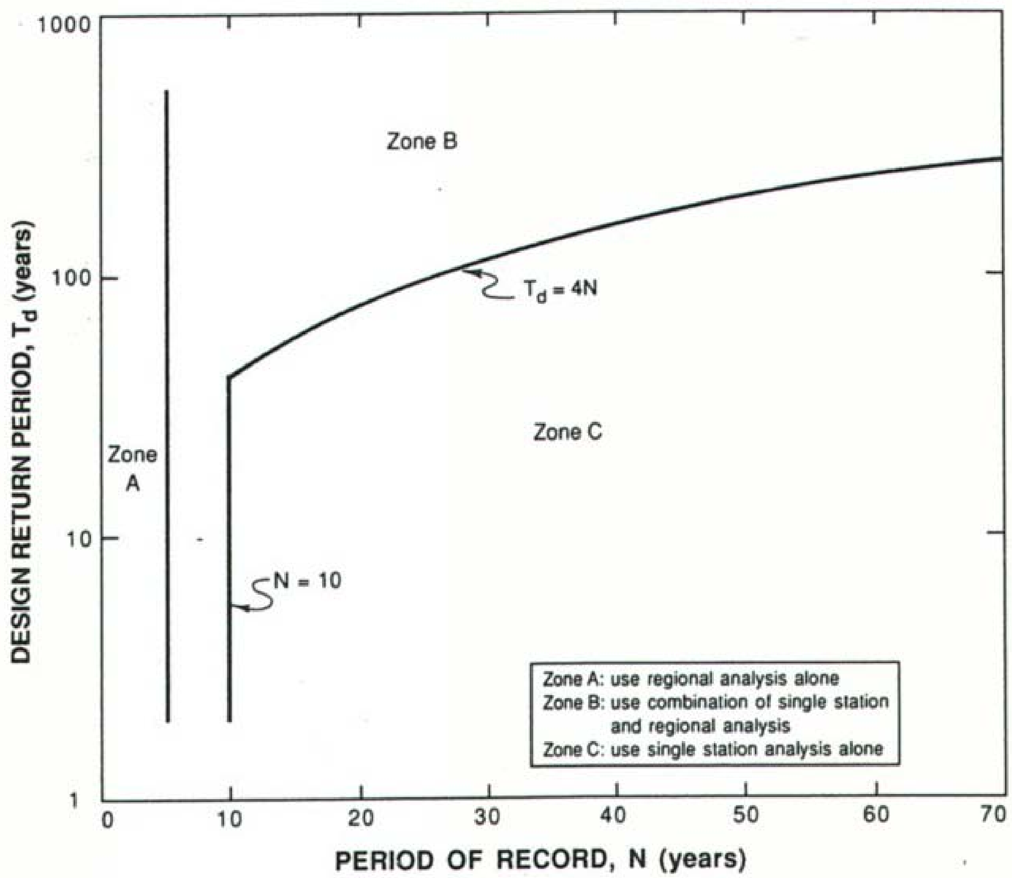


2.3 Record Length

A major encumbrance in hydrology is the size of the data series available from which to compute the magnitude of extreme events. Data series often contain far fewer than 100 values, and often just a few dozen values. Watt *et al.* (1989) provides a practical guideline that can be used to determine whether sufficient data exist from which to compute an extreme event having a desired return period. As illustrated on **Figure 2.7**, Watt *et al*. (1989) suggest using four times the record length as the maximum recurrence interval.

If the data series does not meet a minimum length requirement, gathering of additional regional data, the use of partial duration series or generation of additional data from runoff models could be undertaken to alleviate this shortcoming.

Figure 2.7 Guidance on Period of Record for Estimation of Design Return Period



From Watt *et al*. 1989

2.4 Partial Duration Series

Two basic types of series are used in the frequency analysis of a dataset from a single station:

- The extreme value series consists of the largest event in each equal time interval (i.e., **annual maximum series**); and
- The peaks over threshold series consists of all events above a specified magnitude (i.e., **partial duration series** [PDS]).

2.4.1 Definition and Application

A PDS is typically used for frequency analyses of hydrologic maxima when the available data series is short and/or includes a number of zero or low annual maximum values. In addition, a PDS would be used to better characterize the lower end of the probability-frequency curve (i.e., the recurrence interval of the more common extreme events with return periods in the order of 5 years or less). A PDS of maxima consists of the set of peak data values which exceed a specified threshold value (Stedinger *et al.* 1992). A PDS is sometimes described as a "peaks over threshold" series. Such a data series can include more than one data point for any one year; as well as no data points for years in which the maximum value falls below the threshold.

The decision on whether to use a PDS approach is somewhat subjective. If the available data series is short, a PDS approach may help in adding additional data for analysis. It is advisable to use a PDS if there are more than a couple of zeros in the data series. The need for an analysis of a PDS then becomes more important as the number of zeros increases (> 10% of the data set).

A PDS is not applicable to all hydrologic maxima. For example, spring snowmelt runoff occurs only once per year, thus there is only one such runoff peak discharge event per year. Rainfall runoff events; however, can – and typically do – occur several times per year. Each such runoff event peak discharge is then a potential candidate for inclusion in a partial duration data set, provided that each such event is independent of the preceding event.

2.4.2 Types of Partial Duration Series

There are three general approaches to developing a partial duration data set, based on selection of the threshold value, as follows:

1. Select the magnitude of the threshold value such that the total number of data points equals the total number of years in the period of record. The resulting PDS is termed an "annual exceedance" series.
2. Select the threshold value such that the total number of data points exceeds the number of years in the period of record, but satisfies certain statistical criteria.
3. Select a threshold value of zero, thus including all independent events in the period of record, as well as all annual maxima values equal to zero.

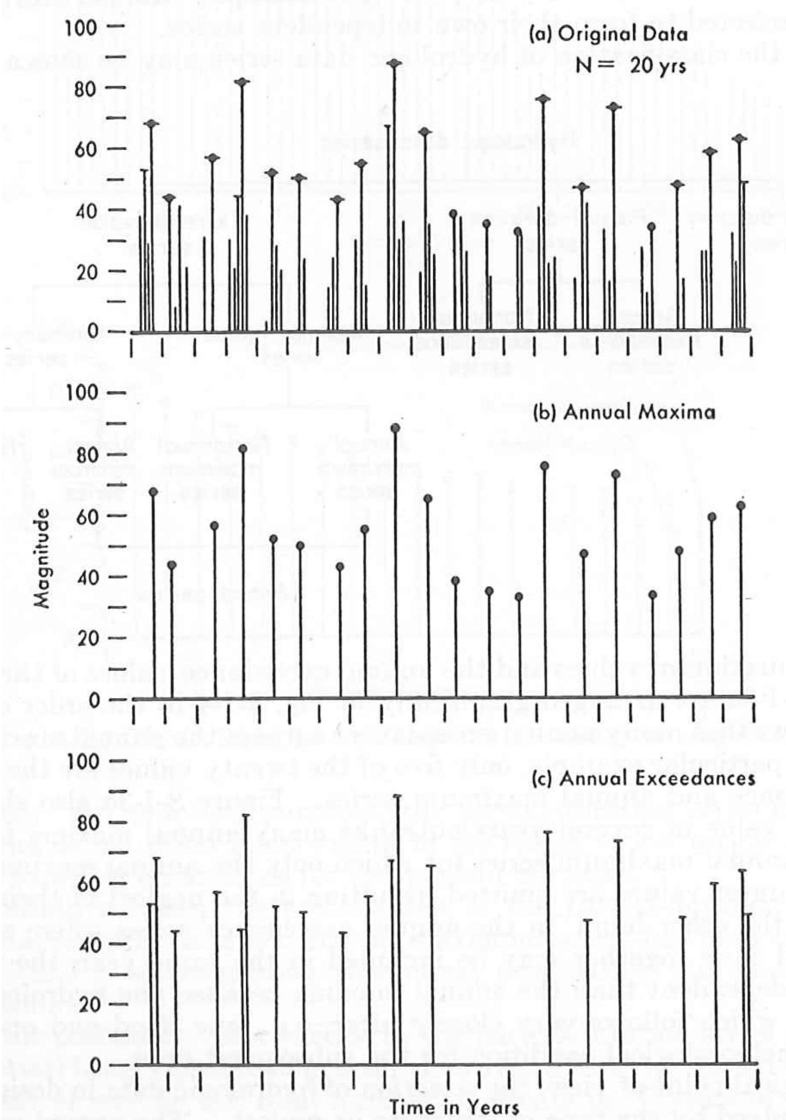
Figure 2.8, adapted from Chow (1964), illustrates the PDS concept. Part (a) of **Figure 2.8** shows the complete data series of peaks over a 20-year period of record, with the annual peaks identified by a dot and the 20 highest peaks identified by a line, at the apex of each. Part (b) shows the series of annual peaks and part (c) shows the PDS of peaks established by using a threshold value of approximately 40. The PDS in this case corresponds to the annual exceedance series and is; therefore, an instance of the first approach listed above.

The second approach is described in Watt *et al.* (1989) and involves the iterative selection of trial threshold values and the application of various statistical tests to evaluate each trial value in order to arrive at the most suitable threshold value.

The third approach has been developed by the Prairie Farm Rehabilitation Administration (1987) for application in the special areas of southeastern Alberta, where it is common to record zero runoff for a significant number of years in the period of record.

Only the first of the three approaches is considered applicable to the design of storm water management facilities within The City. It is also the easiest to use.

Figure 2.8 Relations Between Data Series



From Chow 1964

2.4.3 Exceedance Frequencies

Exceedance frequencies and the associated return periods derived from an annual exceedance PDS can be transposed to the corresponding annual maxima values using the following equation:

$$T_E = 1/\{ \ln T_M - \ln (T_M - 1) \}$$

where:
T_E and T_M are the return periods of the PDS annual exceedance and the annual maxima series, respectively.
This equation is awkward to apply when computing values of T_M to values of T_E, thus a table of values is provided below.

Table 2.1
Comparison of Exceedance Frequencies and
Return Periods for Annual Maxima and Annual Exceedance Series

Partial Duration Annual Exceedance Series		Annual Maxima Series	
Exceedance Frequency	Return Period (T_E)	Exceedance Frequency	Return Period (T_M)
0.01000	100.0	0.00995	100.5
0.01005	99.5	0.0100	100.0
0.0200	50.0	0.0198	50.5
0.0202	49.5	0.0200	50.0
0.0500	20.0	0.0488	20.5
0.0510	19.6	0.0500	20.0
0.100	10.0	0.0952	10.5
0.105	9.52	0.100	10.0
0.200	5.00	0.181	5.52
0.223	4.48	0.200	5.00
0.356	2.81	0.300	3.33
0.510	1.96	0.400	2.50
0.500	2.00	0.394	2.54
0.693	1.44	0.500	2.00
0.917	1.09	0.600	1.67
1.00	1.00	0.630	1.59
1.20	0.833	0.700	1.43
1.61	0.621	0.800	1.25
2.00	0.500	0.862	1.16
2.30	0.435	0.900	1.11
3.00	0.333	0.950	1.05
4.00	0.250	0.981	1.02

The tabulated values indicate that the difference in the return period values is most pronounced at the lower values, and that the value of T_M approaches the value of T_E as the return period increases. Thus, at the 1:5-year frequency, the T_M value is about 10% higher than the T_E value (as indicated by the green highlight), while at the 1:10-year frequency, the T_M value is about 5% higher than the T_E value (as indicated by the turquoise highlight), and at the 1:50-year frequency, the T_M value is only about 1% higher than the T_E value (as indicated by the pink highlight). Thus, it is evident that a PDS frequency analysis would be useful for design of stormwater management facilities for which the 1:5-year event or more frequent events are significant.

If an event can occur in any month of the year, the 1:6-month return period event would have a return period (T_E) of 0.5 years, a 1:3-month return period event would have a return period (T_E) of 0.25 years, etc. However, when considering rainfall, an adjustment would be necessary as the duration of the rainfall season is often less than 1 year. If it is assumed that the normal rainfall runoff season is 9 months long and the PDS analysis gives a return period of 0.333 seasons (not years), the return period would be 0.33 × 9 months = 3 months.

2.5 Outliers

Outliers are data points that depart significantly from the range of the remaining data (i.e., low or high outliers). The magnitude of statistical parameters computed from the data can be significantly affected by the inclusion or deletion of these outliers. Where possible, additional information, (e.g., cause of the event, status of the recording instrument at the time of the event, availability of other data at nearby locations for the same event, etc.) should be used to assess the reliability of outliers.

To determine the effect of outliers on predictions, an analysis both with and without outliers should be conducted and the sensitivity of the results should be evaluated.

Outliers can be identified using the Grubbs and Beck (1972) method or by plotting the data series as a histogram.

A summary for the test for outliers can be found in **Appendix A**. The full calculations for these tests can be found in the frequency analysis spreadsheet.

2.5.1 High Outliers

Values considered to be high outliers should be compared with regional information at nearby sites. If historic regional information is not available to compare to the suspected high outliers, the outliers should be retained in the data set.

2.5.2 Low Outliers

When the sample size of a data series is small, low outliers can affect the skewness (even becoming negative) which creates problems in producing satisfactory results from the frequency analysis. Very low values that depart significantly from the general trend of the data should be removed when determining the sample skew.

Zero values within a data series can also cause issues when using some logarithmic types of distributions as the logarithm of zero is minus infinity.

US Water Resources Council (1982) presents a method for conducting a conditional probability adjustment where the number of zero values is less than 25% of the total values in the data series.

A generally consistent approach has been presented by Wang and Singh (1995), Zhang and Singh (2005), Zhang (2005), and Woo and Wu (2013). This approach is suggested to deal with samples having a varying proportion of zero-value data points. The suggested procedure is presented using the following example of computed water depths in a dry pond, where in many years the pond might remain 'dry' (i.e., have a 'zero' water depth). A worked example using the values discussed below is included in **Appendix C**.

The data set should be divided into the zero-value subset and the non-zero subset.
The probability of an exceedance occurring (i.e., non-zero event) is the ratio of the number of non-zero events divided by the total sample size (e.g., the probability for an exceedance occurring in a sample of 50 with 23 zero values and 27 non-zero values is 27/50 = 0.54).

$$P(X > 0) = \frac{k}{N}$$

Where:
k= number of non-zero values
N = Total number of values
Conversely, the probability of a zero depth occurring is

$$P(X = 0) = \frac{N - k}{N}$$

All non-zero values should be tabulated and ranked from highest to lowest. The probability of a water depth occurring above a depth of interest is computed as the rank of that depth divided by the total number of non-zero values (e.g., where the pond depth of 2.4 m is the fifth highest of the 27 non-zero values, the probability of a water depth being at or above 2.4 m is 5/27 = 18.5%).

$$P_i' = \frac{p_i}{k}$$

Where:
P_i' is the exceedence probability of a non-zero value of rank,i
p_i is the rank of the non-zero value of interest

Then, the likelihood of that depth being equalled or exceeded is the product of the probability of that depth being exceeded times the probability of a non-zero depth occurring (e.g., the probability of the water depth being at or above 2.4 m is 0.185 × 0.54 = 0.10).

$$P_i \ = \ P_i' \times P(X > 0)$$

The method described above is recommended to calculate manually the values so that the plotted zero-value points and the non-zero values are well understood. Thereafter, computational software can be employed to fit a distribution of the non-zero values.
The procedure for this is as described below:

1. Sort the zero and non-zero value data points as described above.
2. Perform a frequency analysis of the non-zero values employing the gamma distribution. The gamma distribution has a lower limit (well-suited to this case where the lower limit is zero).
3. To integrate the probability results from the analysis of non-zero values with the sub-sample of zero-value points, the probability values should be multiplied by the probability of a non-zero level occurring. In effect, this computation shifts the frequency curve to the right (i.e., to a larger non-exceedance probability, or lower exceedance probability). What might be not directly plotted, unless filled in manually, is that the left part of the probability plot has a number of zero values (i.e., a horizontal line extending to the left from the probability value defining the probability of zero-values occurring).

Alternatively, the following five methods have been presented in Alberta Transportation (2004) for dealing with zero values in the data set:

1. Add 1.0 to all the data (when applying log distributions, this data manipulation affects the mean);
2. Add small positive values to all data (e.g., 0.1 or 0.001; affects the mean);
3. Substitute 1.0 in place of all zero readings (when applying log distributions, this affects both the mean and the standard deviation);
4. Substitute small positive values in place of all zero readings (affects the mean and the standard deviation); and
5. Ignore all zero observations.

Each of these methods is a less desirable to the preferred method described at the beginning of this section; method 5 is not recommended. However, each of the first four methods provides a work-around for data sets with zero values. These methods should be tested for the data series under consideration and the resultant frequency curve should be critically appraised to determine if the results are reasonable.

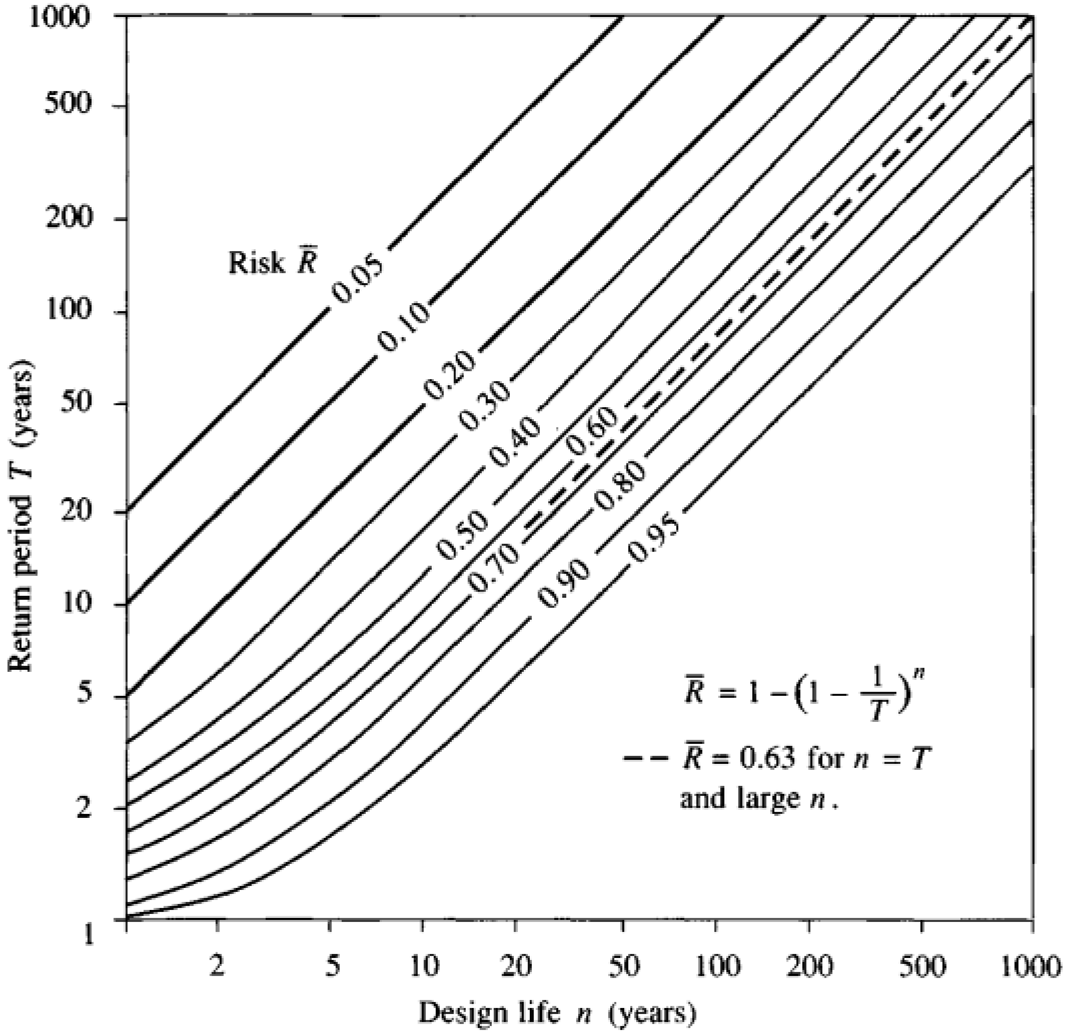
For additional information on the treatment of outliers, one should consult Hawkins (1980), U.S. Water Resources Council (1982) and others.

2.6 Risk

Risk is defined in terms of probability and consequence. The probability, which is the focus of the discussion in this manual, deals with the likelihood of an event of a given magnitude occurring. The consequence of the event occurring, such as a pond overtopping or erosion of a channel bank, is linked to that event and indicates whether or not established performance criteria are met. Watt *et al.* (1989) provide an informative discussion of risk and the design event concept. The definition of risk is the subject of papers by Kaplan and Garrick (1981), Kelman (2003), Baroang *et al.* (2009) and others.

In hydrology we are interested in knowing the probability of one or more events of return period (T) occurring in a duration of periods or years (Y). This aspect is independent of the frequency distribution that might be selected. Chow *et al.* (1988) show that recognition of the probability of a design event occurring within the expected life of a project (n) is an important aspect of the design of stormwater management facilities. This risk of failure (R) can conveniently be illustrated by a family of curves that indicate the risk of at least one event equal to the design event occurring during the life of the structure, as presented in **Figure 2.9**.

Figure 2.9 Risk of Occurrence of Event Occurring During Design Life



From: Chow *et al.* 1988

For example, a structure with a design life of 10 years has a 9.6% chance of at least one 100-year return period event occurring during its 10 year design life.

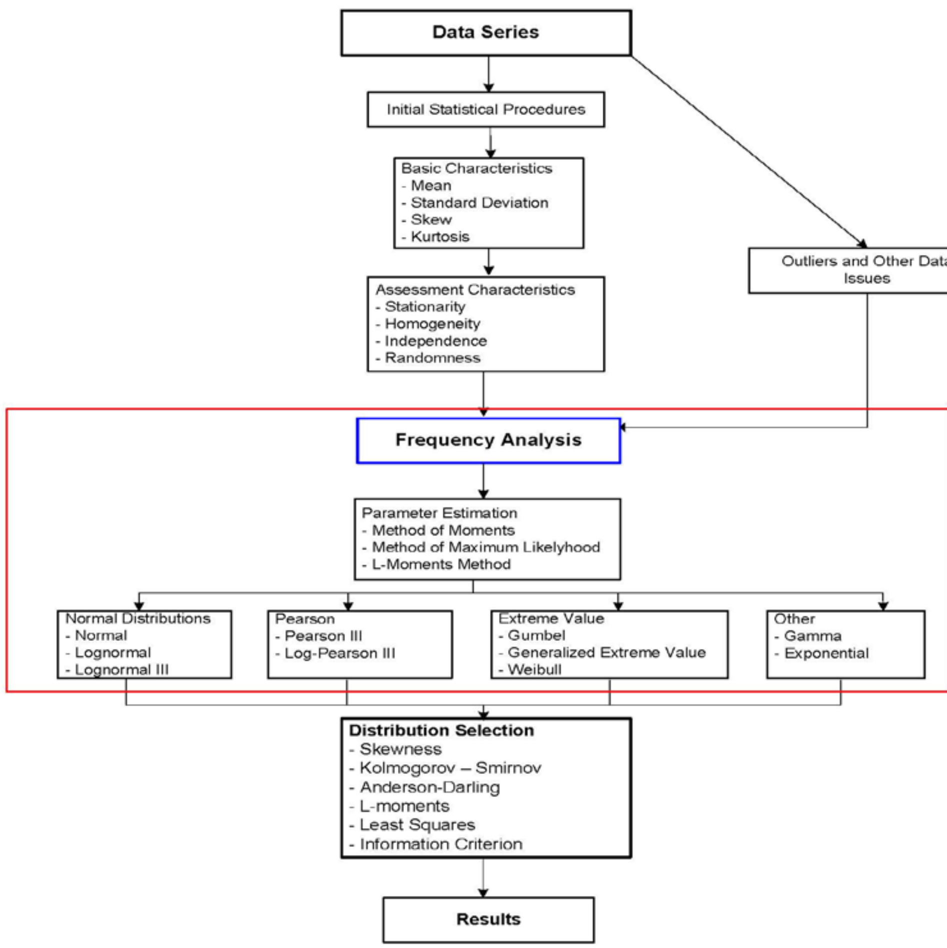
$$R = 1 - (1 - \frac{1}{100})^{10} = 0.096$$

3.0 FREQUENCY ANALYSIS

3.1 Introduction

Frequency analysis is used to interpret a past record of hydrologic events as a basis for predicting probabilities of future occurrences. The procedure consists of selecting a sample in the form of an available data set, fitting a theoretical probability distribution to the sample and "extracting" the prediction rule from the fitted distribution. **Figure 3.1** illustrates the procedures discussed in this chapter.

Figure 3.1 Frequency Analysis



3.2 Plotting Positions

Plots shall be prepared showing the data series and fitted probability distribution. An empirical exceedance probability (p_m) must be calculated for each event in the set using a plotting position formula. The standard formulas used are the Weibull and Cunnane formulas:

Weibull	$p_m = m/(N+1)$;	$T_m = (N+1)/m$
Cunnane	$p_m = (m-0.4)/(N+0.2)$;	$T_m = (N+0.2)/(m-0.4)$

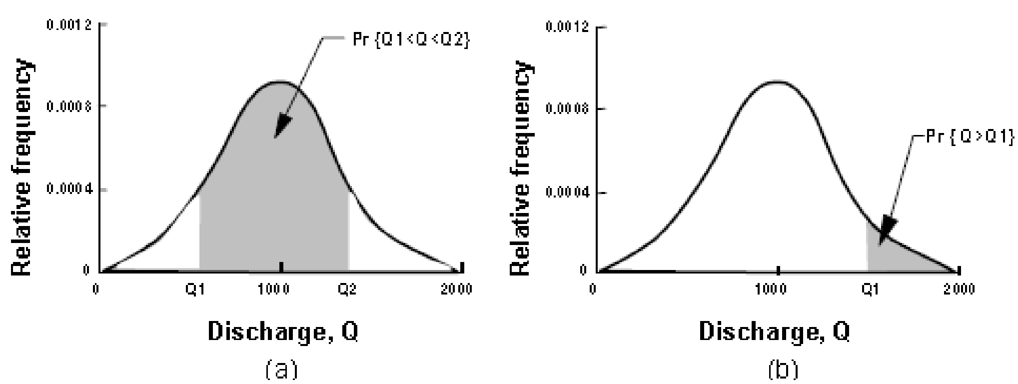
where:
N is the number of points in the data set;
m is the rank of the event in question (i.e., m=1 for the event with the largest magnitude);
pm is the empirical exceedance probability; and
T is the corresponding return period, (T = 1/Pm).

Makkonen (2006) suggests that the only correct formula for plotting position is that of Weibull. Additional information regarding plotting positions can be found in Chow (1964) and Watt *et al*. (1989). The Cunnane formula is typically used by Environment Canada and Watt et al. (1989) note that "it increases significantly the plotted return period of the highest point." **The Cunnane formula is recommended as the default option**.

3.3 Probability Distributions

Probability distributions are mathematical models used for estimating the likelihood and magnitude of events in a frequency analysis. If all the values in a large population data series were plotted in a histogram, smooth curves like those in **Figure 3.2** would result. The shaded area for plot (a) indicates the probability of the value being between Q1 and Q2. In plot (b), the area indicates the probability of the value being greater than Q1. Note that in these plots the area under both curves is unity. These curves are called probability density functions (PDF).

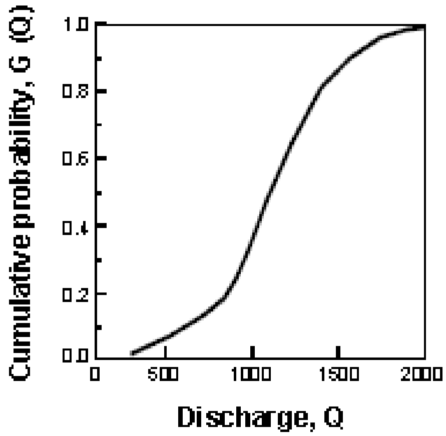
Figure 3.2 Hydrologic Probabilities from Density Functions



From: Bengston 2012

Integrating the area under the curve results (**Figure 3.3)** illustrates the cumulative distribution function (CDF). For this example, it indicates that the probability of the discharge being greater than 1,500 is approximately 15%.

Figure 3.3 Cumulative Distribution Function



From Bengston 2012

Different probability functions generate different shapes when plotted as PDFs and CDFs. Each probability function will predict a different probability for a given event, based on that distribution's probability density function. The distributions used in hydrologic frequency analysis are usually defined by two (mean and standard deviation) or three (mean, standard deviation, and skewness) parameters. Probability distributions are fit to the data series of interest by estimating the distribution parameters from the data series to be analysed.

Log-normal, gamma and extreme value I are two-parameter distributions with a positive skew. The most common three-parameter distributions are: log-Pearson III, three parameter log-normal, and the generalized extreme value (GEV) distributions. The main advantages of two-parameter distributions are their simplicity and that they may be insensitive to sampling error. Some authors argue that two-parameter distributions "impose" less bias (e.g., an inaccuracy in the resultant fitted distribution derived from the fitting technique) than the three-parameter distributions. The three-parameter distributions are very flexible in shape, which is one of their advantages. However, when dealing with small samples (say less than 25 data points) three-parameter distributions should be avoided since they are very sensitive to the coefficient of skewness, which may not be properly estimated because of the small sample size.

3.3.1 Parameter Estimation Methods

Parameters to describe the various distributions can be calculated directly in the case of some of the basic distributions and need to be estimated in the case of other, more complex distributions. The two most common methods for estimation of distribution parameters from the sample data are the method of moments (MOM) and the method of maximum likelihood (MLE). Often, MLE estimators cannot be described by simple formulas, so numerical methods have to be used. Details of parameter estimation by the MOM or MLE can be found in Kite (1977).

A newer method of parameter estimation is the L-moments method introduced by Hosking (1990). Sample estimators of L-moments are linear combinations of the ranked observations so they do not involve “powering” (squaring, cubing, etc.) of observations, like the MOM method. The computed standard deviation and skewness are almost unbiased. L-moments can be written as functions of probability-weighted moments (PWM). The PWM method was developed first and used to define effective statistics for fitting distributions. Later, the PWM method was expressed as L-moments which are more easily interpreted.

Due to the complexity of parameter estimation methods, use of statistical software packages (as discussed in **Section 4.0**) is recommended. The best parameter estimation method to use if informed by the type of distribution and size of the data set. Some guidelines based on distribution type are discussed in **Section 3.3.2**, but in general the use of PWM or L-moments is preferred, with the MLE not recommended for small sample sizes (< 25) (Hansen 2011).

3.3.2 Common Hydrologic Probability Distributions

Table 3.1 summarizes probability distributions typically used in hydrologic frequency analysis

Table 3.1
Common Hydrologic Probability Distributions

Family	Distribution	Probability Density Function	Parameters	Notes
Normal	Normal	$p(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-u}{\sigma}\right)^2}$	μ is population mean σ is standard deviation	The normal distribution is a two parameter density function that is defined by the mean and the standard deviation of the data set. The probability distribution function is symmetric around the mean. It has the characteristic of having zero skew.
	Log-Normal	$p(x)=\frac{1}{x\sigma_y\sqrt{2\pi}}e^{-\frac{(\ln x-\mu_y)^2}{2\sigma_y^2}}$	μ_y is mean of the natural logarithms of x σ_y is the mean and standard deviation of the natural logarithms of x	The log-normal distribution implies a normal distribution of the log-transformed data. Log-normal distributions have received wide usage in hydrology, since most of the hydrologic variables are bounded by zero on the left and are positively skewed. Stedinger (1980) made the following conclusions regarding the fitting of log-normal distributions: • The maximum likelihood method is generally best for fitting two-parameter log-normal distribution for samples of 25 or more. • For a smaller samples (less than 25), the choice of method (MOM or MLE) is not important.
	Log-Normal III	$p(x)=\frac{1}{(x-a)\sigma_y\sqrt{2\pi}}e^{-\frac{(\ln(x-a)-\mu_y)^2}{2\sigma_y^2}}$	a is the lower boundary and μ_y and σ_y are the mean and standard deviation of the natural logarithms of x, respectively	The three parameter log-normal distribution is a normal distribution of the log-transformed data, with a lower boundary. It is difficult to obtain good estimations for parameter estimates, but the MOM performs best for log-normal distributions with low skewness coefficient (say less than 0.4).

Family	Distribution	Probability Density Function	Parameters	Notes
Pearson	Pearson Type III	$p(x)=\frac{a^{\lambda}}{\Gamma(\lambda)}(x-m)^{\lambda-1}e^{-a(x-m)}$	Γ(λ) is the gamma function and α, λ and m are the scale, shape and location parameters calculated by one of the parameter estimation methods	This distribution has been used extensively in the U.S. and a publication from the *U.S. Water Resources Council* (1982) provides all the details. Bobee (1975) showed that this distribution can have many different shapes, depending on the combination of distribution parameters, although some of them are very rare in hydrology. The Pearson Type III distribution can have a positive lower boundary and be unbounded above, or it can have a zero lower boundary with a positive upper bound (this latter characteristic could be reasonable from a hydrological perspective if a physical upper limit exists for the parameter described by the data set, such as the pond water level upstream of a dam).
	Log-Pearson Type III	$p(x)=\frac{a^{\lambda}}{x\Gamma(\lambda)}(\ln x-m)^{\lambda-1}e^{-a(\ln x-m)}$	Γ(λ) is the gamma function and α, λ and m are the scale, shape and location parameters calculated by one of the parameter estimation methods	As with the Pearson Type III distribution the Log-Pearson III distribution can have many different shapes. Bobee (1975) shows that some of these shapes are not applicable to a hydrologic frequency analysis. The log-Pearson III applies to hydrologic frequency analysis only when λ > 1 and when 1/ α > 0 (Kite 1977).

Family	Distribution	Probability Density Function	Parameters	Notes
Extreme Value	Generalized Extreme Value	$p(x)=\frac{1}{a}\left(1-\frac{k}{a}(x-u)\right)^{\frac{1}{k}-1}e^{-\left(1-\frac{k}{a}(x-u)\right)^{\frac{1}{k}}}$	u is the location parameter, α is the scale parameter, and k is the shape parameter	The GEV distribution encompasses all types of extreme value distributions, including the Gumbel and Weibull distributions. Hosking *et al.* (1985) recommends parameter estimation by L-moments method, while Martins and Stedinger (2000) recommend the MLE for hydrologic application.
	Gumbel (EV II)	$p(x)=\frac{1}{a}e^{\left(-\frac{x-u}{a}-e^{-\frac{x-u}{a}}\right)}$	*u* is the location parameter and α is the scale parameter.	The Gumbel distribution is a special case (EVI) of the EV family of distributions with a constant skewness of 1.1396, which is its major setback. The Gumbell distribution has a lower bound. The most common applications of Gumbel distribution are for rainfall intensity-duration-frequency analysis by Environment Canada and the distribution of annual maximum daily discharges Ritzema (1994). Hosking (1990) shows that the L-moments method provides accurate quantile estimates for hydrologic data sets.
	Weibull (EV III)	$p(x)=\frac{c}{a}\left(\frac{x}{a}\right)^{c-1}e^{-\left(\frac{x}{a}\right)^{c}}$	*c* is the shape parameter and *α* is the scale parameter.	The Weibull distribution has an upper bound and should not be used for extreme value estimation unless a physical upper limit exists for the parameter described by the data set, such as the pond water level upstream of a dam. The Weibull distribution has a skewness coefficient less than 1.1396. Both L-moments method and MLE should be used for parameter estimation.

Family	Distribution	Probability Density Function	Parameters	Notes
Gamma		$p(x)=\frac{a^{\lambda}}{\Gamma(\lambda)}x^{\lambda-1}e^{-ax}$	where Γ(λ) is the gamma function, α is the scale parameter and λ is the shape parameter as calculated by one of the parameter estimation methods	The application of the two-parameter Gamma distribution is common in extreme value frequency analysis. Markovic (1965) shows that three-parameter gamma distributions have no significant fitting advantages over two-parameter distributions.
Exponential		$p(x)=\frac{a^{\lambda}}{\Gamma(\lambda)}x^{\lambda-1}e^{-ax}$	*α* is the scale parameter and *m* is the location parameter.	In hydrology, the inter-event time (the time between separate hydrological events) is described by the exponential distribution. It may also be used to model the amount of rain falling into a rain gauge over a short interval, such as an hour (Pegram 2010). The exponential probability distribution function is asymptotic at both axes, and as such is un-fit for use in low flow frequency analyses.

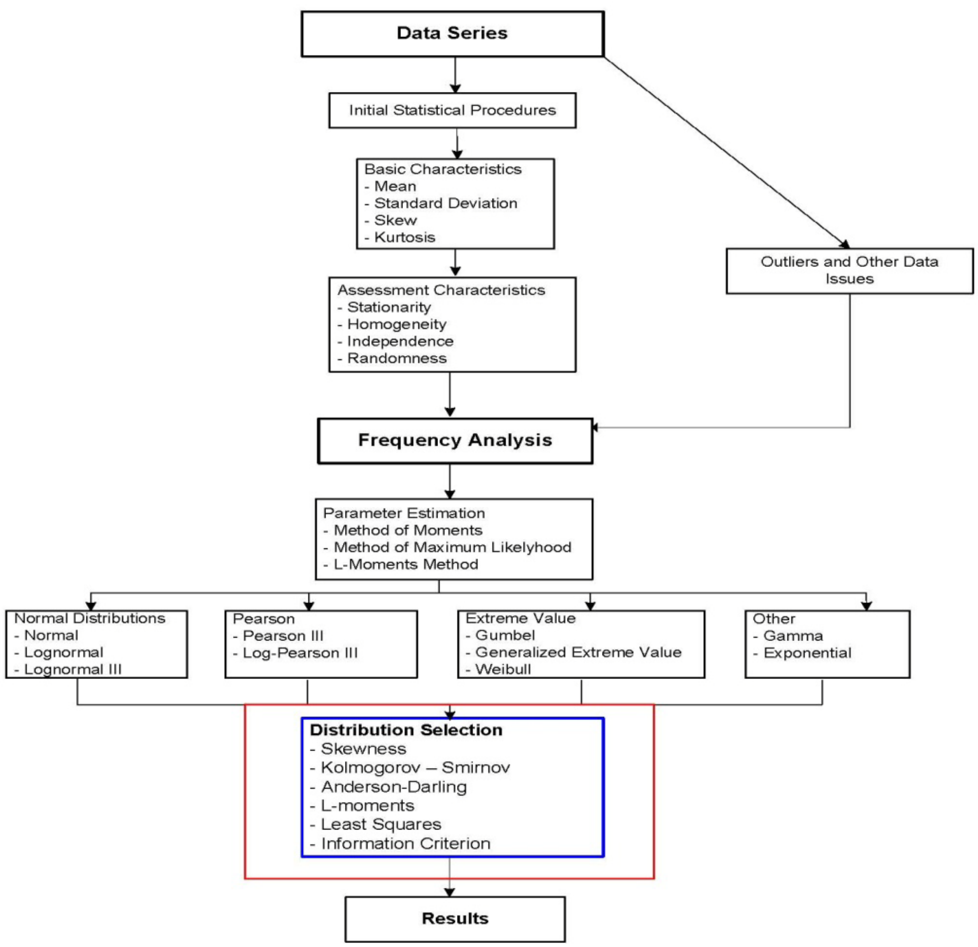
Note: If the user chooses to examine a different distribution from those presented in **Table 3.1**, the use of that distribution should be justified.

4.0 DISTRIBUTION SELECTION

4.1 Introduction

The selection of a distribution is most commonly based on an assessment of fit (i.e., goodness-of-fit) of a theoretical distribution with respect to a sample using graphical assessments, numerical tests, and information criteria. The purpose of using these methods is to derive a defensible probability estimate. Some engineers average the quantiles derived from fitting of various distributions to the same population. This manual avoids this conceptually problematic practice by providing the methods to achieve a defensible estimate. **Figure 4.1** illustrates the procedures discussed in this Chapter.

Figure 4.1 Selection of a Distribution

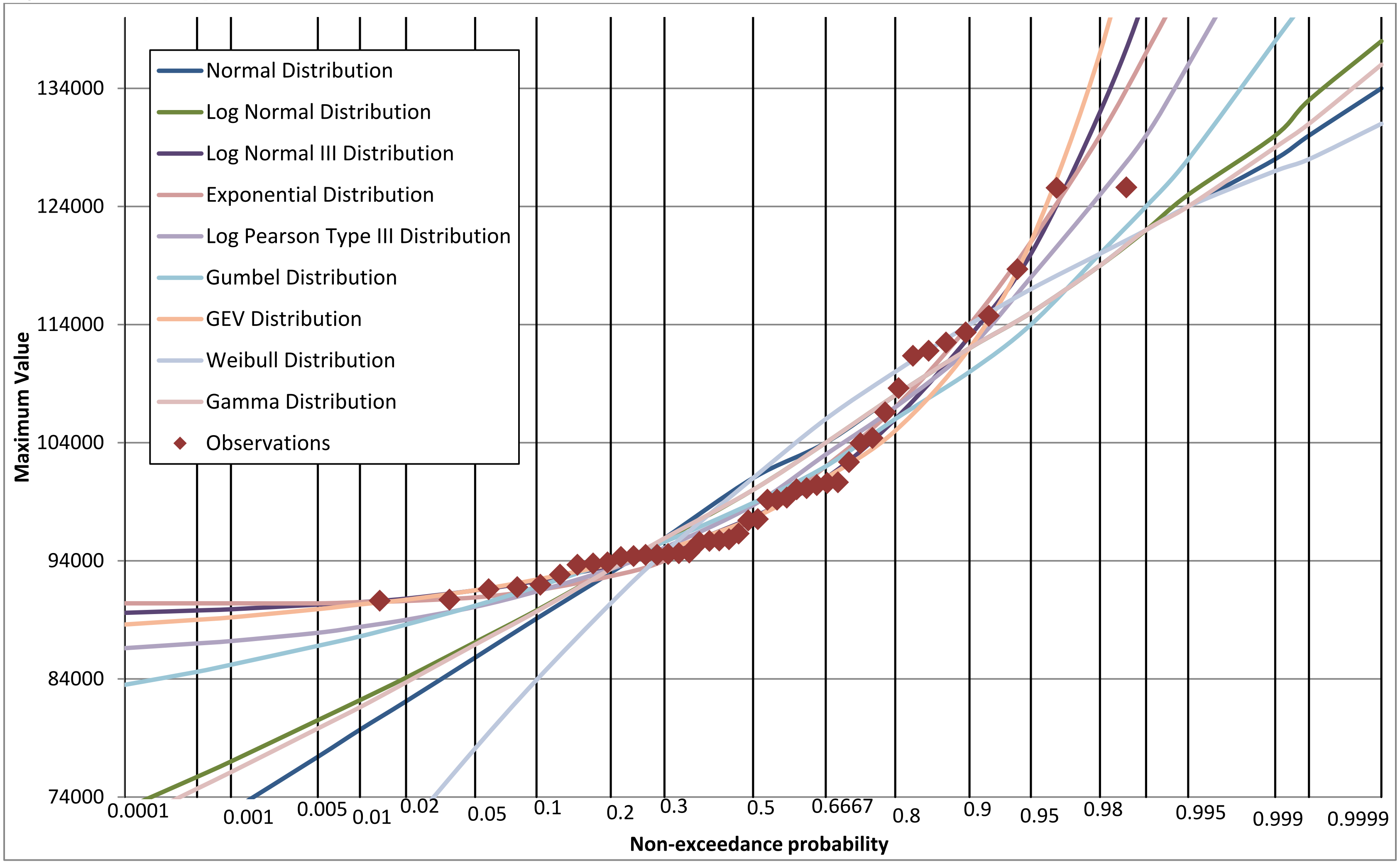


4.2 Graphical Assessment

It is very helpful to observe the graphical representation of the theoretical distribution versus the plotting positions (sample points). For example, in normal and EVI distributions, the linear trend of sample points should closely match these distributions. In a three-parameter distribution, asymptotic distribution behaviour (e.g., the curve asymptotically approaching an upper bound) can be identified. This aspect should be considered if there are physical limitations of the data series under consideration (e.g., pond spill-over based on the local topography for pond volume).

If the sample set has a few very low values, the skewness coefficient of the transformed set may be reduced. This can produce situations where the upper bound of the fitted distribution is even less than the maximum observed event, common in situations with negative skew. Care should be taken to observe the presence of an upper bound from the graphical results of the frequency analysis or from the warnings and results provided by numerical analysis programs and reach a conclusion as to whether or not a particular distribution is reasonable for the given data set. An example of graphical assessment, using all the previously discussed distributions, for the data in **Appendix A** is shown in **Figure 4.2**. Using graphical assessment it becomes immediately obvious that several of these distributions are not suitable for use in fitting this data series (for example the Weibull, Normal, and Gamma distributions in **Figure 4.2**).

Figure 4.2 Graphical Assessment of Fit



4.3 Goodness-of-Fit Tests

4.3.1 Distribution Values of Skewness Coefficient

For the samples of moderate size (e.g., 25 to 50 points), a rough indication of the two-parameter distribution adequacy is the comparison between the theoretical and computed sample value of the coefficient of skewness (i.e., normal and log-normal skewness = 0, EV I skewness = 1.14). Checking the sample skewness should not be done without checking the visual fit of the distribution or without doing other numerical tests. Also, since the sampling error of the coefficient of skewness can be very large, this method should only be used as a general indication of the appropriate distribution. Hence, it may be informative for large populations (e.g., for 50 or more sample points). Zero skew only indicates symmetry about the mean.

4.3.2 Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test is a numerical goodness-of-fit test usable for any of the discussed statistical distributions. To apply this test, the empirical probability of the data series normalized by sample size and the cumulative probability, based on the distribution, are calculated.
The discrepancies between the empirical probability and the probability distribution are then calculated for each of the observed values. The maximum discrepancy is called the D-statistic (D_n), which is then compared to the critical D-statistic (Yevjevich 1972).

$$F_n(x) = \frac{1}{n}\,[Number\ of\ observations\ \leq x\,]$$
$$D_n = maximum\ |F_n(x) - F(x)|$$

The critical D-statistic depends only on the sample size (n). Gray (1973) provides a detailed table of critical D-Statistic values for a range of sample sizes. The values may be approximated by the formulas presented in **Table 4.1** below.

Table 4.1
Acceptance Limits for the Kolomogorov-Smirnov Test of Goodness-of-Fit

Significance Level[1]	Critical D_n
10%	$\frac{1.22}{\sqrt{n}}$
5%	$\frac{1.36}{\sqrt{n}}$
1%	$\frac{1.63}{\sqrt{n}}$

Note: [1] The significance level indicates whether observations fit a pattern or are due to chance. A significance level of 10% indicates that there is a 10% chance that the value is due to chance. Hence, the lower the significance level chosen, the stronger the criterion (i.e., the lower the likelihood of the value being due to chance). While there are no rules for establishing significance level, a level of 5% is often used.

If the calculated D-statistic is greater than the critical statistic, the frequency distribution does not match the sample set.

4.3.3 Anderson-Darling Test

The Anderson-Darling test is a numerical goodness-of-fit test usable for the normal, exponential, Weibull, Gumbell and lognormal distributions. As with the Kolmogorov-Smirnov test, a statistic A is compared to a table of values based on sample size and significance level to determine if the data series fits with the compared probability distribution.
The Anderson-Darling test statistic is calculated using the following:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln\big(F(X_i)\big) + \ln\left(1 - F(X_{n-i+1})\right)\right]$$

where:
n is sample size and *i* is an index of the data set.

Critical values of the Anderson-Darling test statistic depend on the specific distribution being tested, but as tables of critical values for many of the distributions discussed do not exist, the same critical values for all distributions are sometimes used (D'Agostino and Stevens 1986). This results in a test less likely to reject a good fit.

4.3.4 L-Moments Diagrams

Since their introduction by Hosking (1990), the L-moments diagrams have been used often to assess the goodness-of-fit of various probability distributions, mostly for regional samples of streamflow and precipitation. An L-moment diagram compares the theoretical with sample estimates of the L-moment ratios (L-variation, L-skew, L-kurtosis) for a range of assumed distributions. Two main advantages of this method are that one can compare the fit of several distributions using a single graph, and that the L-moment ratios are approximately unbiased for all probability distributions. Further information can be found in Hosking (1990). No commercial software using this method is available at this time, which limits its practical use.

Hosking also indicates that this method is most often used for regional analysis. The application for stormwater design in Calgary is very limited and the inclusion here is only for completeness.

4.3.5 Least Squares Method

The Least Squares Method compares the fit of multiple distributions to a single data sample. The method involves calculating of the sum of squares of the differences between calculated and observed discharges. The disadvantage of this method is that it is dependent on the plotting position equation used to compute the return periods of the observed data. Fortunately, this dependence only governs the absolute value of the sum of squares and it does not affect the relative ranking of the distributions. A ranking of the distributions by order of least standard error based on the sum of squares reflects the ranking of the goodness-of-fit of each distribution.

The standard error for the j^{th} distribution is calculated based on the following formula:

$$SE_j = \sqrt{\frac{1}{n - m_j} \sum_{i=1}^{n} (x_i - y_i)^2}$$

where:
x_i are the data set, y_i are the event magnitudes computed from the j^{th} probability distribution, n the number of events in the data set, and m_j the number of parameters estimated for the j^{th} distribution (Kite, 1977).

4.3.6 Information Criterion

Information criterion methods are a measure of relative goodness-of-fit for statistical models and provide a means for selection of a preferred distribution. They do not test how well the model fits, but are able to rank models based on their goodness-of-fit. Burnham and Anderson (2004) show two most commonly used information criterion methods for frequency analysis are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Both the AIC and BIC are based partially on the likelihood function, and penalize prospective distributions as they increase the number of estimated parameters.

$$AIC = 2k - 2\ln(L)$$
$$BIC = -2\ln(L) + k\ln(L)$$

where:
k is the number of parameters in the statistical model;
L is the maximized value of the likelihood function for the estimated model;
the preferred model is the model with the minimum AIC/BIC value.

In hydrological practice AIC and BIC are applied to a single return period, and the probability functions ranked according to the results for that single return period Chow and Watt (1992) used the AIC to assess probability distributions at 42 long-term Canadian hydrometric stations, and provide guidance on its use.

4.4 Uncertainty

Estimates derived from frequency analysis are subject to some uncertainty, and this uncertainty is greater for longer return periods (i.e., further away from mean or median values). Understanding the estimated uncertainty is important for application of estimates for project design. Uncertainty in frequency analysis arises from sampling uncertainty, distribution uncertainty, and measurement or modelling uncertainty. The following from Alberta Transportation (2004) formula can be used to calculate the combined sampling and distribution uncertainty:

$$SE_t = \sqrt{SE_s^{\ 2} + SE_d^{\ 2}}$$

where:
SE_t is the total uncertainty, SE_s is the sampling uncertainty, and SE_d is the distribution uncertainty.

Uncertainty should be considered by the engineer as an indicator of the range of values within which the "true" value of the parameter of interest might exist. With reference to **Figure 4.3**, while the best fit curve indicates a value of 120,000 for the 1:20-year (0.05 exceedence probability) high event, values in the range of 107,000 to 132,000 may occur 95% of the time. Taking narrower confidence limits, values between 112,000 and 129,000 would occur 85% of the time. In the case of the latter, there is a 7.5% probability (0.5 × [1 - 85 percent]) that the 1:20-year magnitude event would have a value greater than 129,000. Thus, the uncertainty in estimates should be evaluated in making decisions about design values when the implications of a greater than design value represent an appreciable risk.

4.4.1 Sampling Uncertainty

Sampling uncertainty describes how well the available data series represents the actual population of data. This uncertainty decreases as the number of sample values increases (data series size), but increases with the length of the return period estimated (i.e., extreme probabilities). Sampling error is generally synonymous with confidence limits.

The standard error due to sampling uncertainty is a function of the parameter estimation method, as well as the probability distribution. Kite (1977) discusses the calculation of standard error for the normal, lognormal, 3-parameter lognormal, Gumbel, Pearson-III and log Pearson-III distributions.

4.4.2 Distribution Uncertainty

Distribution uncertainty describes the uncertainty in the selection of the most appropriate distribution for the data and can be estimated by comparing the estimates provided by various distributions for a given return period. Distribution uncertainty for a given return period is often calculated as 25% of the range of estimates for that return period given by acceptable fitting distributions (Alberta Transportation 2004).

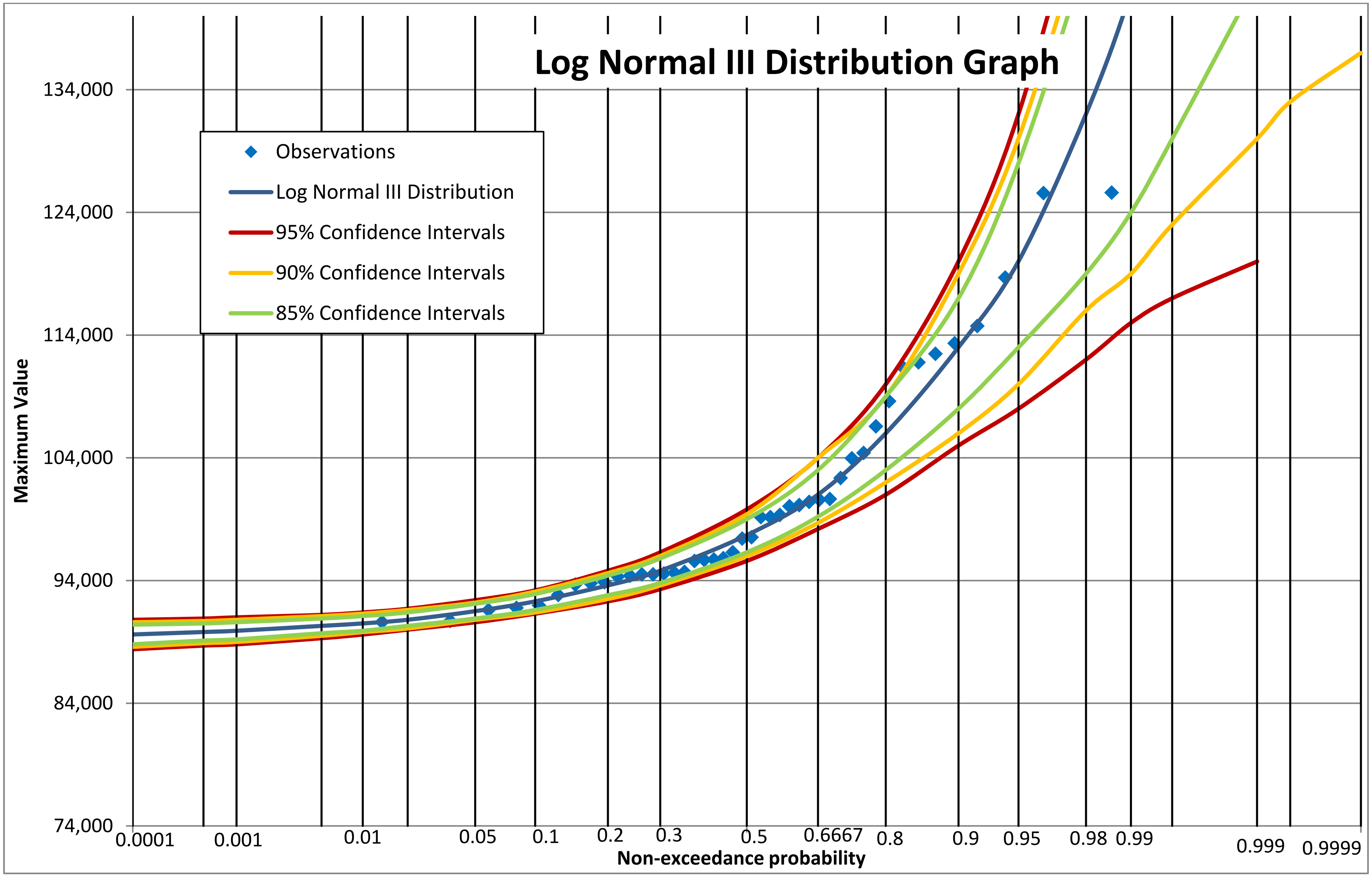
4.4.3 Measurement or Modelling Uncertainty

Measurement or modelling uncertainty is uncertainty in the accuracy of the measured or modelled data used in the data series. This uncertainty is directly related to the source of the data series being analyzed. Measurement uncertainty can be determined from technical specifications for monitoring instruments and measurement procedures. Modelling uncertainty would be a function of the confidence in the modelled data series, and dependent on the model used and the parameters estimated for that model.

4.4.4 Confidence Intervals

Confidence intervals present a range of statistical estimates where the true values are reasonably expected to lie, and illustrate the reliability of estimates of a fitted frequency distribution, as shown on **Figure 4.3**. The upper and lower boundary values of the confidence interval are called confidence limits. The size of confidence interval depends on the confidence level typically 99%, 95%, or 90%. Typically, a 95% confidence interval is used. Confidence intervals are a function of uncertainty, typically only sampling uncertainty, but can be calculated to include other sources of uncertainty if their standard errors are known.

Figure 4.3 Log Normal III Distribution Graph



5.0 NUMERICAL TOOLS

Due to the data intensive nature of frequency analysis, the use of software packages can significantly ease the frequency analysis process. It is important; however, to make sure that the data series being entered into numerical tools meets the assumptions used in frequency analysis (stationarity, homogeneity, randomness, independence). Commonly used analytical software packages are discussed in the following sections.

5.1 Consolidated Frequency Analysis

This package was developed by Environment Canada and is a hydrologic software package for the analysis and graphing of extreme data. Consolidated Frequency Analysis (CFA) is an MS-DOS program that performs both parametric and non-parametric analysis of extreme daily and instantaneous data from the Water Survey of Canada Hydrometric Database (HYDAT). CFA might not run successfully on some current operating systems. It can be obtained at http://www.trentu.ca/iws/software.php.

5.2 Hydrologic Frequency Analysis Plus

Hydrologic Frequency Analysis Plus (HYFRAN+) is a software package used to fit statistical distributions. It includes a number of mathematical tools that can be used for the statistical analysis of extreme events. It provides the verification of statistical hypothesis for independence and homogeneity, supports a large number of statistical probability distributions (including all those discussed in this manual), and several fitting methods. HYFRAN+ also includes support for goodness-of-fit tests and information criterion tests for distribution ranking. It can be obtained at http://www.wrpllc.com/index.html.

5.3 Low-flow Frequency Analysis

The Low-flow Frequency Analysis (LFA) package was developed by Environment Canada and is a hydrologic software package for the analysis and graphing of the frequency of occurrence of selected low flows, based on streamflow records at one or more gauging sites. The LFA uses exported streamflow data from the Environment Canada's HYDAT to evaluate low flows for a variety of surface water studies using a Gumbel Type III distribution. It can be obtained at http://www.trentu.ca/iws/software.php.

5.4 Hydrologic Engineering Center Flood Frequency Analysis

The Hydrologic Engineering Center Flood Frequency Analysis (HEC-FFA) package was developed by the U.S. Army Corps of Engineers to reflect the techniques described in the *Guidelines for Determining Flood Flow Frequency* (U.S. Water Resources Council 1982). Data can be plotted using the Weibull, median or Hazen formulae and it uses the log-Pearson Type III distribution for computation of the frequency curve. The HEC-FFA package is no longer supplied by the U.S. Army Corps of Engineers, but at the time of writing this manual can be found at http://www.hec.usace.army.mil/publications/ComputerProgramDocumentation/CPD-13.pdf.

5.5 Hydrologic Engineering Center Statistical Software Package

The Hydrologic Engineering Center Statistical Software Package (HEC-SSP) performs statistical analyses of hydrologic data. This package can perform flood flow frequency analysis based on Bulletin 17B (U.S. Water Resources Council 1982), similar to HEC-FFA and PeakFQ. It also has the capability to do a generalized frequency analysis on not only flow data but other hydrologic data as well, a volume frequency analysis on high and low flows, a duration analysis, a coincident frequency analysis and a curve combination analysis. It is available free from http://www.hec.usace.army.mil/software/hec-ssp/index.html.

5.6 United States Geologic Survey PeakFQ

The U.S. Geologic Survey (USGS-PeakFQ) package was developed by the U.S.Geologic Survey to reflect the techniques described in the *Guidelines for Determining Flood Flow Frequency* (U.S. Water Resources Council 1982). It provides estimates of instantaneous annual peak flows for a range of recurrence intervals, including 1.5, 2, 2.33, 5, 10, 25, 50, 100, 200, and 500 years. The Pearson Type III frequency distribution is fit to the logarithms of the instantaneous annual peak values. The parameters of the Pearson Type III frequency curve are estimated by the logarithmic sample moments (mean, standard deviation and coefficient of skewness) with adjustments for low outliers, high outliers, historic peaks, and generalized skew. It can be obtained at http://water.usgs.gov/software/PeakFQ/.

5.7 EasyFit

EasyFit is a program by MathWave Technologies used to fit statistical distributions. It includes a number of mathematical tools that can be used for the statistical analysis of extreme events. A large number of statistical probability distributions, including all those discussed in this manual, are supported. Kolmogorov-Smirnov, Anderson-Darling and Chi-Squared goodness-of-fit tests are included. EasyFit does not provide the option of choosing the parameter estimation method, nor does it test for independence, homogeneity, stationarity or randomness. It can be obtained at www.mathwave.com.

5.8 TREND - eWater Toolkit

TREND is statistical software designed to test for trend, change and randomness. It uses 12 statistical tests based on the WMO/UNESCO Expert Workshop on Trend/Change detection. Tests include Mann-Kendall, Spearman's Rho, Linear Regression, Distribution-Free CUSUM, Cumulative Deviation, Worsley Likelyhood Ratio, Rank-Sum, Student's t, Median Crossing, Turning Points, Rank Difference and Autocorrelation. It can be obtained at www.toolkit.net.au/trend.

5.9 Recommended Tool

As the numerical tool with the most complete toolset with respect to hydrologic frequency analysis, HYFRAN+ is recommended. Depending on the complexity of the data set, it is possible to complete an entire frequency analysis using only the HYFRAN+ tool. Use of the HYFRAN+ tool does require judgment in creating the data series and interpretation of the output, but it allows the user to quickly compare multiple frequency distributions, parameter estimation methods and provides some goodness-of-fit of fit and data series characteristic tests to aid in this judgement.

6.0 SPREADSHEET

A spreadsheet was prepared as a part of this manual to provide users with a consistent format for analysing data. The spreadsheet provides tests for stationarity, homogeneity, independence and outliers.

Using the output data from HYFRAN+, the spreadsheet assesses the goodness-of-fit using the Anderson-Darling Test, the Kolmogorov-Smirnov Test, expected probability analysis and least squares ranking. It also provides a sheet to visually compare goodness-of-fit.

This tool does not replace the need for sound judgment when completing a frequency analysis, but supplements HYFRAN+ to provide additional information to allow for informed decisions to be made.

The spreadsheet is available at The City of Calgary Water Resources - Development Approvals webpage, see http://www.calgary.ca/UEP/Water/Pages/Specifications/Submission-for-approval-/Development-Approvals-Submissions.aspx.

7.0 CLOSURE

This report has been prepared for the exclusive use of **The City of Calgary**. This report is based on, and limited by, the interpretation of data, circumstances and conditions available at the time of completion of the work as referenced throughout the report. It has been prepared in accordance with generally accepted engineering practices. No other warranty, express or implied, is made.

Yours truly,
AMEC Environment & Infrastructure

Phil McMechan, P.Eng.
Water Resources Engineer

Neil van der Gugten, M.A.Sc., P.Eng.
Senior Hydrotechnical Engineer

Charles Wojcik, B.A.Sc., M.B.A, P.Eng.
Water Resources Engineer

Gary R.E. Beckstead, M.Sc., P.Eng.
Principal Water Resources Engineer

Reviewed by:

Monica Wagner
Unit Manager – Water Resources

PM/NvdG/GB/dh/tb/jp/elf

Permit to Practice No. P-4546

8.0 LITERATURE CITED

Alberta Transportation. 2004. *Guidelines on Extreme Flood Analysis*. This document is available at http://www.transportation.alberta.ca/Content/doctype125/Production/gdlnextrmfld.pdf.

Baroang, KM., M. Hellmuth and P. Block. 2009. *Identifying Uncertainty and Defining Risk in the Context of the 4th World Water Development Report*. Discussion Paper, United Nations World Water Assessment Programme, Perugia, Italy. p. 1-33.

Bengston, H. H., 2012. *Hydrology (Part 2) – Frequency Analysis of Flood Data*. Course No: C05-013. Continuing Education and Development Inc., Stony Point, New York, United States of America. This document is available at http://www.cedengineering.com/upload/Hydrology%202%20-%20Flood%20Data.pdf.

Bobee, B. 1975. The log-Pearson type 3 distribution and its application in hydrology. Water Resources Res., 11(5), pp. 681-689.

Bobee, B. and R. Robitaille. 1977. *The Use of the Pearson Type 3 and Log Pearson Type 3 Distributions Revisited.* Water Resources Res., 13(2), pp. 427-443.

Burnham, K.P. and D.R. Anderson. 2004. *Multimodel Inference: Understanding AIC and BIC in Model Selection*, Sociological Methods and Research, 33: 261-304.

Clarke, C., M. Hulley, J. Marsalek, and E. Watt. 2011. Stationarity of AMAX Series of Short-Duration Rainfall for Long-Term Canadian Stations: Detection of Jumps and Trends. Canadian Journal of Civil Engineering. 38: 1175-1184.

Chow, K. C. A, and W. E. Watt. 1992. *Use of Aikake Information Criterion for Selection of Flood Frequency Distribution*. Canadian Journal of Civil Engineering, 19 616-626.

Chow, V. T. 1964. *Handbook of Applied Hydrology*. McGraw-Hill, New York, New York, pp. 8-18 to 8-23.

Chow, V.T., D.R. Maidment, L.W. Mays. 1988. *Applied Hydrology*. McGraw-Hill.

D'Agostino, R.B. and M.A. Stephens. 1986. *Goodness-of-Fit Techniques*. New York, United States of America: Marcel Dekker.

Gray, D.M. (Ed.). 1973. *Handbook on the Principles of Hydrology*. Water Information Center, Port Washington, New York.

Grubbs E.E. and G. Beck. 1972. Extension of sample sizes and percentage points for significance tests of outlying observations. Technometrics, 14(4): 847-854

Hansen, C. 2011. *Frequency Analysis of Annual Maximum Rainfall for Selected Sites in Alberta, Using At-Stie and Regional Approaches*. Edmonton, Alberta.

Hawkins, D.M. 1980. *Identification of Outliers*. Monographs on Applied Probability and Statistics, Chapman and Hall, London, England.

Hosking, J.R.M., J.R. Wallis, and E.F. Wood. 1985. Estimation of the Generalized Extreme - Value Distribution by the Method of Probability - Weighted Moments, Technometrics, Vol. 27, No. 3, pp: 251- 260.

Hosking, J.R.M. 1990. *L-moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics*. Journal of the Royal Statistical Society, Series B52: 105-124.

Kaliq, M.N., T.B.M.J. Ouarda, J.C. Gachon and B. Bobee. 2006. *Frequency Analysis of a Sequence of Dependent and/or Non-Stationary Hydro-meteorological Observations: a Review*. Journal of Hydrology, Vol. 339, Issue 3-4, P. 534 to 552.

Kaplan, S. and B.J. Garrick. 1981. *On the Quantitative Definition of Risk*. Risk Analysis, Volume 1, No. 1. Society for Risk Analysis, pp. 11 to 27.

Kelman, I. 2003. *Defining Risk*. FloodRiskNet Newsletter, Issue 2, pp. 6 to 8.

Kite, G.W., 1977 *Frequency and Risk Analyses in Hydrology,* Water Resources Publications, Colorado, United States of America.

Makkonen, L. *Plotting Positions in Extreme Value Analysis*. Journal of Applied Meteorology and Climatology, Volume 45, pp. 334 to 340.

Mann, H.B. and Whitney, D.R. 1947. On the test of whether one of two random variables is stochastically larger than the other. An. Math. Stat., 18, pp. 50-60.

Markovic, R.D. August 1965. *Probability Functions of Best Fit to Distributions of Annual Precipitation and Runoff*. Colorado State University Hydrology Paper No. 8. Fort Collins, Colorado.

Martins, E.S and Stedinger, J.R. March 2000. Generalized maximum-likelihood generalized extreme- value quintile estimators for hydrologic data. Water Resources Research 36(3): 737-744.

National Environment Research Council. 1975. *Flood Studies Report: Volume 1, Hydrological Studies.* Natural Environment Research Council, London, England.

Pegram, G.G.S. Probabilistic Methods and Stochastic Hydrology. 2010. Volume I in Hydraulic Structures, Equipment and Water Data Acquisition Systems. Encyclopaedia™ of Life Support Systems. Accessed 30 October 2012 from: www.eolss.net/Sample-Chapters/C07/E2-15-02-05.pdf.

Prairie Farm Rehabilitation Administration, Unpublished Course Notes. 1987. *Current Practice in Hydrologic Evaluation*. Regina, Saskatchewan.

Ritzema H.P (Ed.). 1994. *Frequency and Regression Analysis*. Chapter 6 in: D*rainage Principles and Applications*, Publication 16, International Institute for Land Reclamation and Improvement, Wageningen, the Netherlands.

Siegel, S. 1956. *Nonparametric statistics for the behavioural sciences.* McGraw-Hill, New York, United States of America.

Stedinger, J.R. 1980. *Fitting Log Normal-Distributions to Hydrologic Data*. Water Resources Research 16 (3): 481-490.

Stedinger J.R., Vogel, R.M. and Foufoula-Georgiou, E. 1992. *Frequency Analysis of Hydrologic Data*. In Handbook of Hydrology, D.R. Maidment (ed.), McGraw-Hill, Inc. New York, United States of America.

Terry, M.E. 1952. Some Rank Order Tests which are Most Powerful Against Specific Parametric Alternatives. Annals of Mathematical Statistics 23, 346-366.

United States Army (U.S. Army). 1982. *Mixed Population Frequency Analysis*. Institute for Water Resources, Hydrologic Engineering Center, Training Document 17, Davis California.

United States Geological Survey (USGS). 2003. *Peak-Flow Frequency Estimates for United States. Geological Survey Streamflow-Gaging Stations in Connecticut*. United States Geological Survey Water-Resources Investigations Report 03-4196. East Hartford, Connecticut. This document is available at:
http://pubs.usgs.gov/wri/wri034196/wrir03-4196.pdf

United States Water Resources Council (U.S. Water Research Council). 1982. *Guidelines for Determining Flood Flow Frequency*. Bulletin 17B. Interagency Advisory Committee on Water Data.

Wald, A. and Wolfowitz, J. 1943. An exact test for randomness in the non-parametric case based on serial correlation. An. Math. Stat., 14, pp. 378-388.

Wang, S. X. and V. P. Singh, 1995. Frequency Estimation for Hydrological Samples with Zero Values. *Journal of Water Resources Planning and Management*, Vol. 121, No. 1, January/February 1995.

Watt, W.E., *et al.* 1989. *Hydrology of Floods in Canada: A Guide to Planning and Design*. National Research Council of Canada, Associate Committee on Hydrology. Ottawa, Ontario.

Woo, M-K, and K Wu, 1989. Fitting Annual floods with Zero Flows. *Canadian Water Resources Journal*, 14:2, 10-16. Published online 23 January 32013 at http://dx.doi.org/10.4296/cwrj1402010

Yevjevich, V. 1972. *Probability and Statistics in Hydrology*. Water Resources Publications, LLC. Highlands Ranch, Colorado.

Zhang, L. and Singh, V.P., 2005. Frequency Analysis of Flood Damage. *Journal of Hydrologic Engineering*, Vol. 10, No. 2, March 1, 2005

Zhang, Lan, 2005. Multivariate Hydrological Frequency Analysis and Risk Mapping**.** A Dissertation Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree Doctor of Philosophy in the Department of Civil and Environmental Engineering.

Appendix A

Stormwater Pond

APPENDIX A

Stormwater Pond

This data set represents the annual maximum series for a pond with a maximum discharge rate of 2.5 L/s/ha within the City of Calgary.

Year	Maximum Annual Volume (m^3)	Year (continued)	Maximum Annual Volume (m^3)
1964	103965	1986	104405
1965	108623	1987	96312
1966	111361	1988	111787
1967	91776	1989	91977
1968	100415	1990	95597
1969	100653	1991	97534
1970	90714	1992	118695
1971	100058	1993	97426
1972	106573	1994	94507
1973	94512	1995	94577
1974	95828	1996	94707
1975	90625	1997	114748
1976	102362	1998	112484
1977	93868	1999	95720
1978	113341	2000	95687
1979	94328	2001	100594
1980	99352	2002	92831
1981	94405	2003	93670
1982	93780	2004	99179
1983	94649	2005	125582
1984	99176	2006	100155
1985	125620	2007	91601

Basic Characteristics

The **mean** of the sample is:

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$
$$= \frac{103965 + 108623 + \ldots + 91601}{44}$$
$$= \frac{4425759}{44}$$
$$= 100585$$

The **standard deviation** of the sample is:

$$S = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}$$
$$= \sqrt{\frac{1}{(44-1)} \left(\left(103965 - 100585\right)^2 + \left(108623 - 100585\right)^2 + \ldots + \left(91601 - 100585\right)^2\right.}$$
$$= 8996$$

The **skew** of the sample is:

$$G = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{S}\right)^3$$
$$= \frac{44}{(44-1)(44-2)} \left(\left(\frac{103965 - 100585}{8996}\right)^3 + \left(\frac{108623 - 100585}{8996}\right)^3 + \ldots + \left(\frac{91601 - 100585}{8996}\right)^3\right)$$
$$= 1.33$$

The **kurtosis** of the sample is:

$$K = \frac{1}{nS^4} \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^4$$
$$= \frac{1}{44 \times 8996^4} \left(\left(103965 - 100585\right)^4 + \left(108623 - 100585\right)^4 + \ldots + \left(91601 - 100585\right)^4\right)$$
$$= 3.72$$

DFASCC

Data and Frequency Analysis Spreadsheet for the City of Calgary
Version 1.2

PROJECT INFORMATION SHEET

Project Name:	South Pond - Shepard Landfill
Project Description:	Stormwater Management Facility for Shepard Landfill (Maximum Annual Pond Volumes) - Independent Dataset Example
Location:	Calgary
Date:	12/12/2012
Designed by:	Charles Wojcik
Company Name:	AMEC
Reviewed by:	-

Clear Project Information Sheet

Hydrologic Data Series Input

LEGEND

User Input	Negative Result
Calculated Cells	Positive Result

NOTES

- This Spreadsheet is designed for a maximum of 10,000 entries (if more are required then formulas need to be adjusted)
-Input dataset must be based on uniform time distribution (i.e.: daily, weekly, monthly, yearly) and must not include multiple values for any of the time steps (duplicates highlighted in red)
-Input dataset must not have any missing cells of data (cells with "0" will be treated as having a value of 0)

Index	Date	Value	Empirical Probability of Non-Exceedance
1	1964	103965	0.738
2	1965	108623	0.805
3	1966	111361	0.828
4	1967	91776.3	0.081
5	1968	100415	0.647
6	1969	100653	0.692
7	1970	90714.3	0.036
8	1971	100058	0.602
9	1972	106573	0.783
10	1973	94512.4	0.285
11	1974	95828.3	0.443
12	1975	90624.9	0.014
13	1976	102362	0.715
14	1977	93867.6	0.195
15	1978	113341	0.896
16	1979	94328.1	0.217
17	1980	99351.6	0.579
18	1981	94404.6	0.240
19	1982	93780	0.172
20	1983	94648.7	0.330
21	1984	99175.6	0.534
22	1985	125620	0.986
23	1986	104405	0.760
24	1987	96311.5	0.466
25	1988	111787	0.851
26	1989	91977.1	0.104
27	1990	95597.3	0.376
28	1991	97534.3	0.511
29	1992	118695	0.941
30	1993	97426.2	0.489
31	1994	94506.9	0.262
32	1995	94576.9	0.308
33	1996	94707	0.353
34	1997	114748	0.919
35	1998	112484	0.873
36	1999	95720	0.421
37	2000	95686.8	0.398
38	2001	100594	0.670
39	2002	92830.6	0.127
40	2003	93670	0.149
41	2004	99178.8	0.557
42	2005	125582	0.964
43	2006	100155	0.624
44	2007	91601.4	0.059

Clear All Input Data

Basic Characteristics	
Number of Data Entries	44
Maximum Value	126000
Minimum Value	90600
Average (Mean) Value	101000
Median Value	97500
Standard Deviation	9000
Variance	80900000
Variation coefficient (Cv)	0.0894
Skewness coefficient (Cs)	1.33
Kurtosis	3.72

*Values assumed to be sample not full population

Empirical Probability of Non-Exceedance (Plotting Position) based on:
$F(x(k)) = (k-a)/(n-2a+1)$, $0 <= a <= 0.5$

a =	0.4	Cunnane (1978)
k=	rank of the even in question (in ascending order)	
n=	44	

Stationarity

Test for Trend:

Choose Significance Level (alpha):	5%

1) Spearman Rank Order Correlation Coefficient

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}.$$

H_0= Data has no trend

Spearman Correlation Coefficient:	-0.044

based on z	**No Significant Trend at 0.05 Significance Level**
based on t	**No Significant Trend at 0.05 Significance Level**

When there are no ties in rankings:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

Spearman Correlation Coefficient:	-0.044
t-distribution value	-0.287
Degrees of freedom	42

T (Adjustment for ties) =	0
Standard Normal (z)=	-0.284

Tests for Jump:

2) Mann-Whitney Test for jump (a.k.a. Mann-Whitney U test)

Index number of subsample divide	22

H_0= Independent samples drawn from the same population (No Jump)

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}$$

Number of values in sample 1 n_1=	22
Number of values in sample 2 n_2=	22
Total of Ranking in sample 1 R_1=	487
Total of Ranking in sample 2 R_2=	503
U_1=	234

No Jump at 0.05 Significance Level

$$U_1 + U_2 = n_1 n_2.$$

U_2=	250
U (Minimum of U_1 and U_2)=	234
Standard Normal (z)=	-0.188

3) Wald-Wolfowitz Test (The runs test)

$$\mu - \frac{2\,N_+\,N_-}{N} + 1\,, \qquad \sigma^2 = \frac{2\,N_+\,N_-\,(2\,N_+\,N_- - N)}{N^2\,(N-1)} = \frac{(\mu - 1)(\mu - 2)}{N - 1}\,.$$

Number of data greater than median N_+ =	22
Number of data less than median N_- =	22
Total number of runs =	24

H_0= Data represent sample of single independently distributed random variable (No Jump)

Mean =	23.0
Variance =	10.7
Standard Normal (z)=	0.2

No Jump at 0.05 Significance Level

Homogeneity

Choose Significance Level (alpha):	5%

Mann-Whitney Test for homogeneity (a.k.a. Mann-Whitney U test)

Index number of subsample divide	22

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$

H_0= There is homogeneity between samples with respect to probability of random drawing of a larger observation

Sample is Homogeneous at 0.05 Significance Level

Number of values in sample 1 n_1=	22
Number of values in sample 2 n_2=	22
Total of Ranking in sample 1 R_1=	487
Total of Ranking in sample 1 R_2=	503
U_1=	234

$$U_1 + U_2 = n_1 n_2.$$

U_2=	250
U (Minimum of U_1 and U_2)=	234
Standard Normal (z)=	-0.188

Terry Test

Index number of subsample divide	22

H_0= There is homogeneity between samples with respect to probability of random drawing of a larger observation

Total sample size	44
Subsample 1 (m)	22
Subsample 2 (n)	22

Sample is Homogeneous at 0.05 Significance Level

Standard Deviation =	3.260
Sum of ranks in first subsample c =	1.163
z =	0.357

Independence

Choose Significance Level (alpha): 5%

1) Spearman Rank Order Correlation Coefficient

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}.$$

H_0= Data is independent

Spearman Correlation Coefficient:	-0.06

Data is independent at 0.05 Significance Level

When there are no ties in rankings:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

Spearman Correlation Coefficient:	-0.06
t-distribution value	-0.40
Degrees of freedom	42

2) Wald-Wolfowitz Test

$$R = \sum_{i=1}^{N-1} x_i x_{i+1} + x_1 x_N$$

Statistic R	4.45E+11
Mean	4.45E+11
Variance	2.56E+17

H_0= Data is independent

Data is independent at 0.05 Significance Level

Standard Normal (z)=	0.1

2) Anderson Test

$$r_1 = \left[\sum_{i=1}^{N-1} x_i x_{i+1} + x_1 x_N - \left(\sum_{i=1}^{N} x_i \right)^2 \Big/ N \right] \Big/ \left[\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2 \Big/ N \right]$$

Statistic r	-0.014
Mean	-0.023
Variance	0.023

H_0= Data is independent

Data is independent at 0.05 Significance Level

Standard Normal (z)=	0.1

Outliers

Significance Level (alpha): 10%

Grubbs and Beck test for Outliers

1) High Outliers

Assumption: logarithms of sample are normally distributed

X_h = exp (xmean+K_nS)
K(n) =-3.62201+6.2844N^1/4-2.49835N^1/2+0.491436N^3/4-0.037911N
K(n) = -0.9043+3.345*SQRT(log(n))-0.4046log(n) for 5<n<150

Sample Size (n) =	44	
K(n) =	2.72	
K(n) for 5<n<150 =	2.72	
X_h=	126000	< Any value higher than X_h is considered a high outlier
Maximum Value	126000	
High Outliers	No High Outliers Present	

2) Low Outliers

X_h = exp (xmean-K_nS)
K(n) =-3.62201+6.2844N^1/4-2.49835N^1/2+0.491436N^3/4-0.037911N
K(n) = -0.9043+3.345*SQRT(log(n))-0.4046log(n) for 5<n<150

Sample Size (n) =	44	
K(n) =	2.72	
K(n) for 5<n<150 =	2.72	
X_h=	79500	< Any value lower than X_h is considered a low outlier
Minimum Value	90600	
Low Outliers	No Low Outliers Present	

Dependent Dataset

Choose Significance Level (alpha):	5%

Autocorrelation coefficient

$$R_c(\tau)=\frac{\sum_{i=1}^{N-|\tau|}X_iY_{i+\tau}-\frac{1}{N-|\tau|}\left(\sum_{i=1}^{N-|\tau|}X_i\right)\left(\sum_{i=\tau+1}^{N}Y_i\right)}{\left[\sum_{i=1}^{N-|\tau|}X_i^2-\frac{1}{N-|\tau|}\left(\sum_{i=1}^{N-|\tau|}X_i\right)^2\right]^{0.5}\left[\sum_{i=1+|\tau|}^{N}Y_i^2-\frac{1}{N-|\tau|}\left(\sum_{i-1+|\tau|}^{N}Y_i\right)^2\right]^{0.5}}$$

H_0 - The data is not serially correlated

One Time Period Offset		
Autocorrelation coefficient offset by one time period	r(1) =	-0.005
t-distribution values for one time period offset	t =	-0.034

No Serial Correlation at 0.05 Significance Level

Two Time Periods Offset		
Autocorrelation coefficient offset by two time periods	r(2) =	-0.245
t-distribution values for two time periods offset	t =	-1.639

No Serial Correlation at 0.05 Significance Level

Instructions:
Compare the results of the autocorrelation tests for one time period offset and for the two time period offset. One of the following 2 scenarios will result:
1. The finding for the one period time step is serially correlated, and the finding for the two time step is also serially correlated. In this case, transposing the data series is unlikely to produce an independent data set suitable for frequency analysis. In this case, other methods, such as the Monte Carlo simulation are necessary.
2. The finding for the one period time step is serially correlated, and the finding for the two time step is NOT serially correlated. In this case, the data series should be transposed to produce an independent data set suitable for frequency analysis.

Frequency Analysis Results Input

LEGEND

User Input	Negative Result
Calculated Cells	Positive Result

Clear All Input Data

NOTES

- This spreadsheet designed to accept the results of 10 specific Frequency Analysis outputs
- The input data must be in the same format as the output table from Hyfran (either copied and pasted special as text in the top left cell of each yellow input box, or manually input as distribution results and hyfran calculated paremeters in specified areas.
- Input dataset must be complete (only one method of estimation per distribution type, refer to Section 3.3.1 and 3.3.2 of the Frequency Analysis Procedures for Stormwater Design Manual when choosing methods of estimation)
- An additional 11th Frequency Analysis ouput can be copied into the last input box. This output will be displayed in the visual goodness of fit tab, however no numerical goodness of fit tests will be performed on it.

Normal (Gaussian) type of distributions:

Normal Distribution:

Paste Normal Distribution Hyfran Output in Cell Below (A15)

Rivière Harricana à Amos

Results of the fitting

Normal (Maximum Likelihood)

Number of observations 44

Parameters

mu	100585.414
sigma	8995.9516

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)	
10000	0.9999	1.34E+05	3.85E+03	1.26E+05	1.42E+05
2000	0.9995	1.30E+05	3.47E+03	1.23E+05	1.37E+05
1000	0.999	1.28E+05	3.29E+03	1.22E+05	1.35E+05
200	0.995	1.24E+05	2.84E+03	1.18E+05	1.29E+05
100	0.99	1.22E+05	2.63E+03	1.16E+05	1.27E+05
50	0.98	1.19E+05	2.41E+03	1.14E+05	1.24E+05
20	0.95	1.15E+05	2090	1.11E+05	1.19E+05
10	0.9	1.12E+05	1840	1.09E+05	1.16E+05
5	0.8	1.08E+05	1580	1.05E+05	1.11E+05
3	0.6667	1.04E+05	1420	1.02E+05	1.07E+05
2	0.5	1.01E+05	1360	9.79E+04	1.03E+05
1.4286	0.3	9.59E+04	1450	9.30E+04	9.87E+04
1.25	0.2	9.30E+04	1580	8.99E+04	9.61E+04
1.1111	0.1	8.91E+04	1840	8.54E+04	9.27E+04
1.0526	0.05	8.58E+04	2090	8.17E+04	8.99E+04
1.0204	0.02	8.21E+04	2.41E+03	7.74E+04	8.68E+04
1.0101	0.01	7.97E+04	2.63E+03	7.45E+04	8.48E+04
1.005	0.005	7.74E+04	2.84E+03	7.18E+04	8.30E+04
1.001	0.001	7.28E+04	3.29E+03	6.63E+04	7.92E+04
1.0005	0.0005	7.10E+04	3.47E+03	6.42E+04	7.78E+04
1.0001	0.0001	6.71E+04	3.85E+03	5.96E+04	7.47E+04

Lognormal Distribution:

Paste Lognormal Distribution Output from Hyfran in Cell Below (A57)

Rivière Harricana à Amos

Results of the fitting

Lognormal (Maximum Likelihood)

Number of observations 44

Parameters

mu	11.515099
sigma	0.085235

Quantiles
q = F(X) : non-exceedance probability

T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)	
10000	0.9999	1.38E+05	5.02E+03	1.28E+05	1.47E+05
2000	0.9995	1.33E+05	4.36E+03	1.24E+05	1.41E+05
1000	0.999	1.30E+05	4.07E+03	1.22E+05	1.38E+05
200	0.995	1.25E+05	3.36E+03	1.18E+05	1.31E+05
100	0.99	1.22E+05	3.05E+03	1.16E+05	1.28E+05
50	0.98	1.19E+05	2.73E+03	1.14E+05	1.25E+05
20	0.95	1.15E+05	2.29E+03	1.11E+05	1.20E+05
10	0.9	1.12E+05	1.95E+03	1.08E+05	1.16E+05
5	0.8	1.08E+05	1.61E+03	1.05E+05	1.11E+05
3	0.6667	1.04E+05	1400	1.01E+05	1.07E+05
2	0.5	1.00E+05	1290	9.77E+04	1.03E+05
1.4286	0.3	9.58E+04	1320	9.33E+04	9.84E+04
1.25	0.2	9.33E+04	1400	9.05E+04	9.60E+04
1.1111	0.1	8.98E+04	1570	8.68E+04	9.29E+04
1.0526	0.05	8.71E+04	1730	8.37E+04	9.05E+04
1.0204	0.02	8.41E+04	1920	8.04E+04	8.79E+04
1.0101	0.01	8.22E+04	2050	7.82E+04	8.62E+04
1.005	0.005	8.05E+04	2170	7.62E+04	8.47E+04
1.001	0.001	7.70E+04	2400	7.23E+04	8.17E+04
1.0005	0.0005	7.57E+04	2490	7.08E+04	8.06E+04
1.0001	0.0001	7.30E+04	2670	6.78E+04	7.82E+04

Lognormal III Distribution

Paste Lognormal III Distribution Output from Hyfran in Cell Below (A99)

Rivière Harricana à Amos

Results of the fitting

3-parameter lognormal (Maximum Likelihood)

Number of observations 44

Parameters

m	89146.9291
mu	9.05288
sigma	0.785637

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)		
10000	0.9999	2.48E+05	6.77E+04	N/D	N/D	
2000	0.9995	2.02E+05	4.23E+04	N/D	N/D	
1000	0.999	1.86E+05	3.38E+04	1.20E+05	2.52E+05	
200	0.995	1.54E+05	1.87E+04	1.17E+05	1.90E+05	
100	0.99	1.42E+05	1.39E+04	1.15E+05	1.70E+05	
50	0.98	1.32E+05	9.99E+03	1.12E+05	1.52E+05	
20	0.95	1.20E+05	6.03E+03	1.08E+05	1.32E+05	
10	0.9	1.13E+05	3.84E+03	1.05E+05	1.20E+05	
5	0.8	1.06E+05	2280	1.01E+05	1.10E+05	
3	0.6667	1.01E+05	1500	9.82E+04	1.04E+05	
2	0.5	9.77E+04	1060	9.56E+04	9.98E+04	
1.4286	0.3	9.48E+04	752	9.33E+04	9.63E+04	
1.25	0.2	9.36E+04	624	9.23E+04	9.48E+04	
1.1111	0.1	9.23E+04	499	9.13E+04	9.32E+04	
1.0526	0.05	9.15E+04	443	9.06E+04	9.24E+04	
1.0204	0.02	9.08E+04	433	9.00E+04	9.17E+04	
1.0101	0.01	9.05E+04	451	8.96E+04	9.14E+04	
1.005	0.005	9.03E+04	477	8.93E+04	9.12E+04	
1.001	0.001	8.99E+04	5.44E+02	8.88E+04	9.10E+04	
1.0005	0.0005	8.98E+04	5.71E+02	8.87E+04	9.09E+04	
1.0001	0.0001	8.96E+04	6.25E+02	8.84E+04	9.08E+04	

Exponential and Pearson type of distributions:

Exponential Distribution

Paste Exponential Distribution Output from Hyfran in Cell Below (A142)

Rivière Harricana à Amos

Results of the fitting

Exponential (Maximum Likelihood)

Number of observations 44

Parameters

alpha	10192.1535
m	90393.2601

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)		
10000	0.9999	1.84E+05	1.43E+04	1.56E+05	2.12E+05	
2000	0.9995	1.68E+05	1.18E+04	1.45E+05	1.91E+05	
1000	0.999	1.61E+05	1.07E+04	1.40E+05	1.82E+05	
200	0.995	1.44E+05	8.20E+03	1.28E+05	1.60E+05	
100	0.99	1.37E+05	7.13E+03	1.23E+05	1.51E+05	
50	0.98	1.30E+05	6.05E+03	1.18E+05	1.42E+05	
20	0.95	1.21E+05	4.63E+03	1.12E+05	1.30E+05	
10	0.9	1.14E+05	3.55E+03	1.07E+05	1.21E+05	
5	0.8	1.07E+05	2.48E+03	1.02E+05	1.12E+05	
3	0.6667	1.02E+05	1.69E+03	9.83E+04	1.05E+05	
2	0.5	9.75E+04	1070	9.54E+04	9.96E+04	
1.4286	0.3	9.40E+04	568	9.29E+04	9.51E+04	
1.25	0.2	9.27E+04	388	9.19E+04	9.34E+04	
1.1111	0.1	9.15E+04	265	9.09E+04	9.20E+04	
1.0526	0.05	9.09E+04	236	9.05E+04	9.14E+04	
1.0204	0.02	9.06E+04	232	9.01E+04	9.11E+04	
1.0101	0.01	9.05E+04	232	9.00E+04	9.10E+04	
1.005	0.005	9.04E+04	233	9.00E+04	9.09E+04	
1.001	0.001	9.04E+04	234	8.99E+04	9.09E+04	
1.0005	0.0005	9.04E+04	234	8.99E+04	9.09E+04	
1.0001	0.0001	9.04E+04	234	8.99E+04	9.09E+04	

Pearson Type III Distribution

Paste Pearson III Distribution Output from Hyfran in Cell Below (A184)

Rivière Harricana à Amos

Results of the fitting

Pearson type III (Method of moments)

Number of observations 44

Parameters

alpha	0.000167
lambda	2.259593
m	87062.7494

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)		
10000	0.9999	1.61E+05	2.32E+04	1.15E+05	2.06E+05	
2000	0.9995	1.50E+05	1.78E+04	1.15E+05	1.85E+05	
1000	0.999	1.46E+05	1.55E+04	1.15E+05	1.76E+05	
200	0.995	1.35E+05	1.05E+04	1.14E+05	1.55E+05	
100	0.99	1.30E+05	8.46E+03	1.13E+05	1.46E+05	
50	0.98	1.25E+05	6.54E+03	1.12E+05	1.37E+05	
20	0.95	1.18E+05	4.28E+03	1.10E+05	1.26E+05	
10	0.9	1.13E+05	2.94E+03	1.07E+05	1.18E+05	
5	0.8	1.07E+05	2150	1.03E+05	1.11E+05	
3	0.6667	1.03E+05	1900	9.90E+04	1.06E+05	
2	0.5	9.86E+04	1710	9.53E+04	1.02E+05	
1.4286	0.3	9.49E+04	1220	9.25E+04	9.73E+04	
1.25	0.2	9.31E+04	969	9.12E+04	9.50E+04	
1.1111	0.1	9.11E+04	1320	8.85E+04	9.37E+04	
1.0526	0.05	8.99E+04	2110	N/D	N/D	
1.0204	0.02	8.88E+04	3.11E+03	N/D	N/D	
1.0101	0.01	8.83E+04	3.77E+03	N/D	N/D	
1.005	0.005	8.79E+04	4.33E+03	N/D	N/D	
1.001	0.001	8.74E+04	5.36E+03	N/D	N/D	
1.0005	0.0005	8.72E+04	5.71E+03	N/D	N/D	
1.0001	0.0001	8.70E+04	6.35E+03	N/D	N/D	

Log-Pearson Type III Distribution

Paste Log Pearson III Distribution Output from Hyfran in Cell Below (A226)

Rivière Harricana à Amos

Results of the fitting

Log-Pearson type III (Méthode SAM)

Number of observations 44

Parameters

alpha	48.316743
lambda	3.123498
m	4.936298

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)		
10000	0.9999	1.70E+05	3.19E+04	N/D	N/D	
2000	0.9995	1.55E+05	2.25E+04	N/D	N/D	
1000	0.999	1.49E+05	1.89E+04	N/D	N/D	
200	0.995	1.36E+05	1.18E+04	1.13E+05	1.59E+05	
100	0.99	1.30E+05	9.22E+03	1.12E+05	1.48E+05	
50	0.98	1.25E+05	6.92E+03	1.11E+05	1.38E+05	
20	0.95	1.18E+05	4.38E+03	1.09E+05	1.26E+05	
10	0.9	1.12E+05	2.94E+03	1.06E+05	1.18E+05	
5	0.8	1.07E+05	2.05E+03	1.03E+05	1.11E+05	
3	0.6667	1.03E+05	1720	9.91E+04	1.06E+05	
2	0.5	9.87E+04	1520	9.57E+04	1.02E+05	
1.4286	0.3	9.51E+04	1140	9.28E+04	9.73E+04	
1.25	0.2	9.33E+04	931	9.15E+04	9.51E+04	
1.1111	0.1	9.14E+04	1060	8.93E+04	9.34E+04	
1.0526	0.05	9.01E+04	1590	8.70E+04	9.32E+04	
1.0204	0.02	8.90E+04	2350	N/D	N/D	
1.0101	0.01	8.84E+04	2880	N/D	N/D	
1.005	0.005	8.79E+04	3350	N/D	N/D	
1.001	0.001	8.72E+04	4260	N/D	N/D	
1.0005	0.0005	8.70E+04	4580	N/D	N/D	
1.0001	0.0001	8.66E+04	5220	N/D	N/D	

Extreme Value type of distributions:

EVI (Gumbel) Distribution

Paste EV Distribution Output from Hyfran in Cell Below (A269)

Rivière Harricana à Amos

Results of the fitting

Gumbel (Maximum Likelihood)

Number of observations 44

Parameters

u	96724.892
alpha	5959.72509

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)	
10000	0.9999	1.52E+05	6.92E+03	1.38E+05	1.65E+05
2000	0.9995	1.42E+05	5.78E+03	1.31E+05	1.53E+05
1000	0.999	1.38E+05	5.29E+03	1.28E+05	1.48E+05
200	0.995	1.28E+05	4.16E+03	1.20E+05	1.36E+05
100	0.99	1.24E+05	3.68E+03	1.17E+05	1.31E+05
50	0.98	1.20E+05	3.20E+03	1.14E+05	1.26E+05
20	0.95	1.14E+05	2.57E+03	1.09E+05	1.19E+05
10	0.9	1.10E+05	2.09E+03	1.06E+05	1.14E+05
5	0.8	1.06E+05	1630	1.02E+05	1.09E+05
3	0.6667	1.02E+05	1290	9.96E+04	1.05E+05
2	0.5	9.89E+04	1050	9.68E+04	1.01E+05
1.4286	0.3	9.56E+04	911	9.38E+04	9.74E+04
1.25	0.2	9.39E+04	900	9.21E+04	9.57E+04
1.1111	0.1	9.18E+04	950	8.99E+04	9.36E+04
1.0526	0.05	9.02E+04	1030	8.82E+04	9.22E+04
1.0204	0.02	8.86E+04	1130	8.64E+04	9.08E+04
1.0101	0.01	8.76E+04	1210	8.53E+04	9.00E+04
1.005	0.005	8.68E+04	1270	8.43E+04	8.93E+04
1.001	0.001	8.52E+04	1420	8.24E+04	8.80E+04
1.0005	0.0005	8.46E+04	1470	8.18E+04	8.75E+04
1.0001	0.0001	8.35E+04	1580	8.04E+04	8.66E+04

GEV (General Extreme Value) Distribution

Paste GEV Distribution Output from Hyfran in Cell Below (A311)

Rivière Harricana à Amos

Results of the fitting

GEV (Maximum Likelihood)

Number of observations 44

Parameters

alpha	4713.24077
k	-0.370941
u	95749.0007

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)		
10000	0.9999	4.70E+05	3.54E+05	N/D	N/D	
2000	0.9995	2.96E+05	1.51E+05	N/D	N/D	
1000	0.999	2.48E+05	1.03E+05	N/D	N/D	
200	0.995	1.74E+05	3.92E+04	N/D	N/D	
100	0.99	1.53E+05	2.49E+04	N/D	N/D	
50	0.98	1.37E+05	1.53E+04	1.07E+05	1.67E+05	
20	0.95	1.21E+05	7.57E+03	1.06E+05	1.36E+05	
10	0.9	1.12E+05	4210	1.04E+05	1.21E+05	
5	0.8	1.05E+05	2260	1.01E+05	1.10E+05	
3	0.6667	1.01E+05	1440	9.80E+04	1.04E+05	
2	0.5	9.76E+04	1010	9.56E+04	9.96E+04	
1.4286	0.3	9.49E+04	724	9.35E+04	9.63E+04	
1.25	0.2	9.37E+04	620	9.25E+04	9.49E+04	
1.1111	0.1	9.24E+04	548	9.13E+04	9.34E+04	
1.0526	0.05	9.15E+04	548	9.04E+04	9.26E+04	
1.0204	0.02	9.07E+04	598	8.95E+04	9.19E+04	
1.0101	0.01	9.03E+04	650	8.90E+04	9.15E+04	
1.005	0.005	8.99E+04	7.05E+02	8.85E+04	9.13E+04	
1.001	0.001	8.92E+04	8.27E+02	8.76E+04	9.09E+04	
1.0005	0.0005	8.90E+04	8.76E+02	8.73E+04	9.07E+04	
1.0001	0.0001	8.86E+04	9.79E+02	8.67E+04	9.05E+04	

EVIII (Weibull) Distribution

Paste Weibull Distribution Output from Hyfran in Cell Below (A353)

Rivière Harricana à Amos

Results of the fitting

Weibull (Maximum Likelihood)

Number of observations 44

Parameters

alpha	104949.171
c	10.066386

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)		
10000	0.9999	1.31E+05	3.37E+03	1.24E+05	1.37E+05	
2000	0.9995	1.28E+05	3.08E+03	1.22E+05	1.34E+05	
1000	0.999	1.27E+05	2.94E+03	1.21E+05	1.33E+05	
200	0.995	1.24E+05	2.58E+03	1.19E+05	1.29E+05	
100	0.99	1.22E+05	2.41E+03	1.17E+05	1.27E+05	
50	0.98	1.20E+05	2.23E+03	1.16E+05	1.25E+05	
20	0.95	1.17E+05	1.98E+03	1.13E+05	1.21E+05	
10	0.9	1.14E+05	1.79E+03	1.10E+05	1.18E+05	
5	0.8	1.10E+05	1650	1.07E+05	1.13E+05	
3	0.6667	1.06E+05	1640	1.03E+05	1.09E+05	
2	0.5	1.01E+05	1780	9.77E+04	1.05E+05	
1.4286	0.3	9.47E+04	2140	9.05E+04	9.89E+04	
1.25	0.2	9.04E+04	2440	8.56E+04	9.52E+04	
1.1111	0.1	8.39E+04	2910	7.82E+04	8.96E+04	
1.0526	0.05	7.81E+04	3310	7.16E+04	8.46E+04	
1.0204	0.02	7.12E+04	3750	6.39E+04	7.86E+04	
1.0101	0.01	6.65E+04	4020	5.86E+04	7.43E+04	
1.005	0.005	6.20E+04	4240	5.37E+04	7.03E+04	
1.001	0.001	5.28E+04	4590	4.38E+04	6.18E+04	
1.0005	0.0005	4.93E+04	4680	4.02E+04	5.85E+04	
1.0001	0.0001	4.20E+04	4770	3.27E+04	5.14E+04	

Gamma type of distributions:

Gamma Distribution

Paste Gamma Distribution Output from Hyfran in Cell Below (A396)

Rivière Harricana à Amos

Results of the fitting

Gamma (Maximum Likelihood)

Number of observations 44

Parameters

alpha	0.001358
lambda	136.63822

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

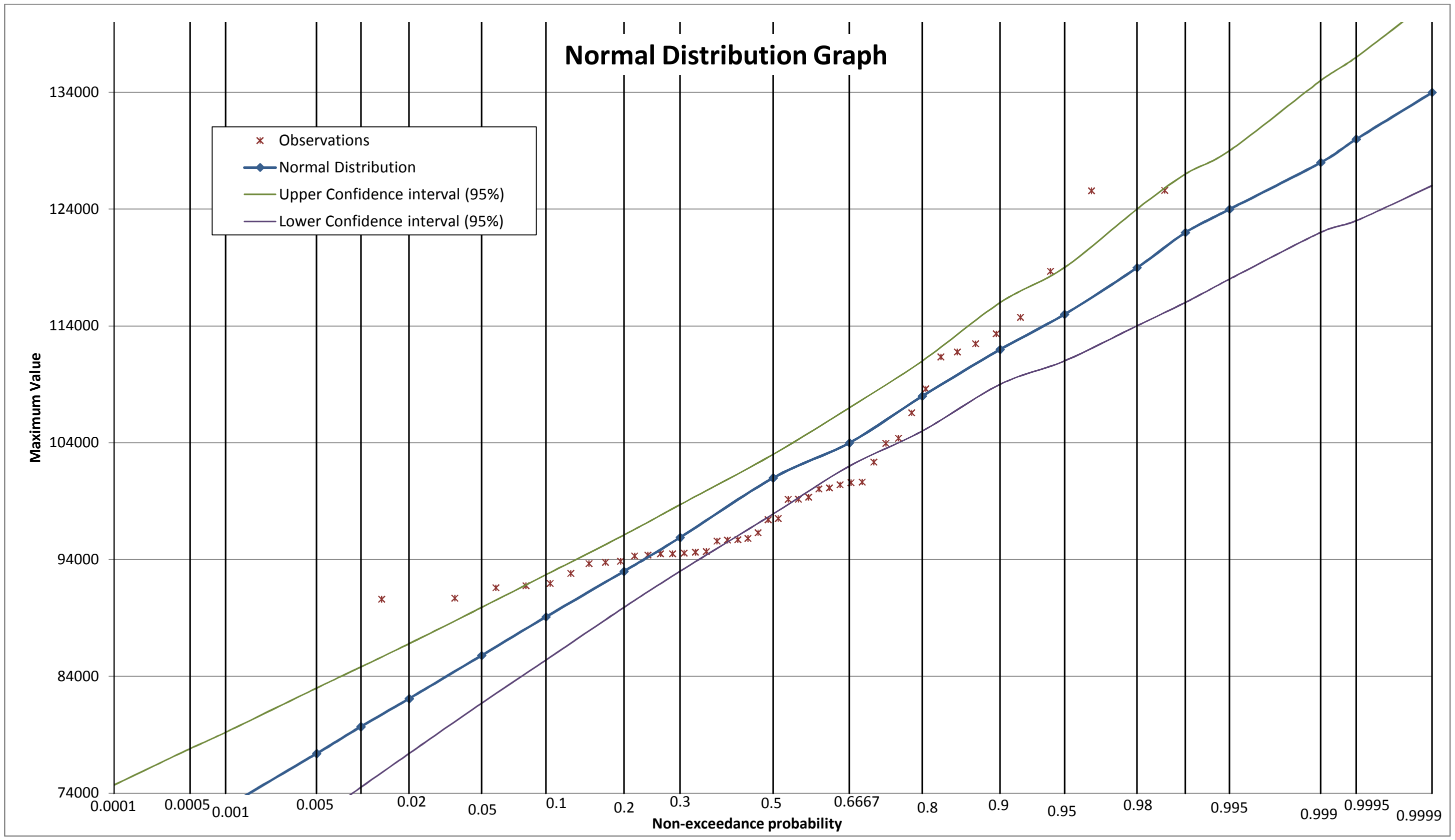
T	q	XT	Standard deviation	Confidence interval (95%)	
10000	0.9999	1.36E+05	4.45E+03	1.27E+05	1.45E+05
2000	0.9995	1.31E+05	3.92E+03	1.24E+05	1.39E+05
1000	0.999	1.29E+05	3.68E+03	1.22E+05	1.37E+05
200	0.995	1.24E+05	3.10E+03	1.18E+05	1.30E+05
100	0.99	1.22E+05	2.83E+03	1.16E+05	1.27E+05
50	0.98	1.19E+05	2.56E+03	1.14E+05	1.24E+05
20	0.95	1.15E+05	2.18E+03	1.11E+05	1.19E+05
10	0.9	1.12E+05	1.88E+03	1.08E+05	1.15E+05
5	0.8	1.08E+05	1580	1.05E+05	1.11E+05
3	0.6667	1.04E+05	1390	1.01E+05	1.07E+05
2	0.5	1.00E+05	1300	9.78E+04	1.03E+05
1.4286	0.3	9.59E+04	1340	9.33E+04	9.85E+04
1.25	0.2	9.33E+04	1440	9.05E+04	9.61E+04
1.1111	0.1	8.97E+04	1620	8.65E+04	9.29E+04
1.0526	0.05	8.69E+04	1810	8.33E+04	9.04E+04
1.0204	0.02	8.37E+04	2030	7.97E+04	8.77E+04
1.0101	0.01	8.16E+04	2170	7.74E+04	8.59E+04
1.005	0.005	7.98E+04	2310	7.53E+04	8.43E+04
1.001	0.001	7.61E+04	2570	7.10E+04	8.11E+04
1.0005	0.0005	7.47E+04	2670	6.94E+04	7.99E+04
1.0001	0.0001	7.17E+04	2880	6.61E+04	7.74E+04

User Defined Distribution

Paste User Defined Distribution Output from Hyfran in Cell Below (A439) or Input Calculated Values in Designated Cells

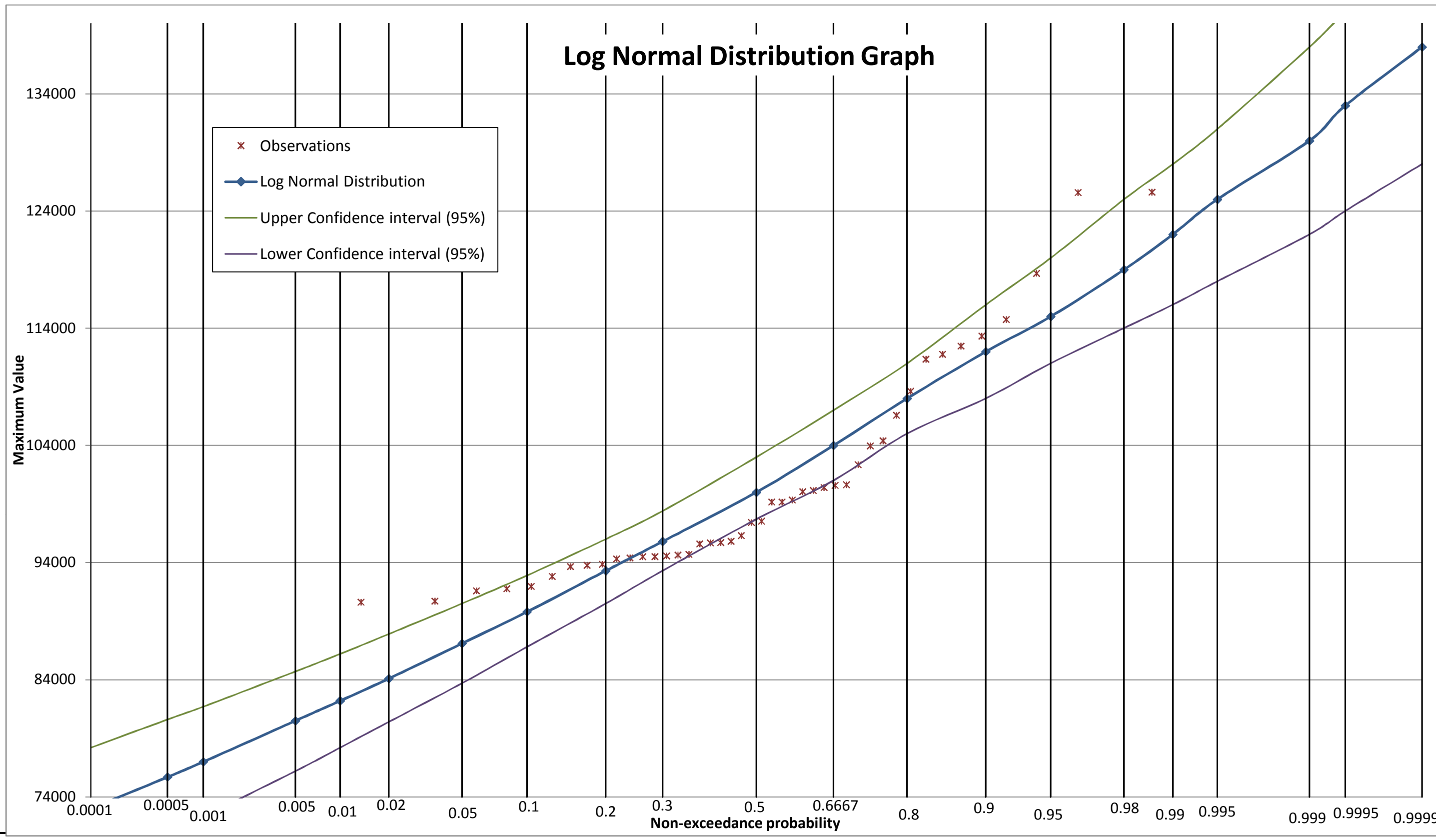
				Lower	**Upper**	
T	q	XT	Standard deviation	Confidence interval (95%)		

Visual Goodness of Fit



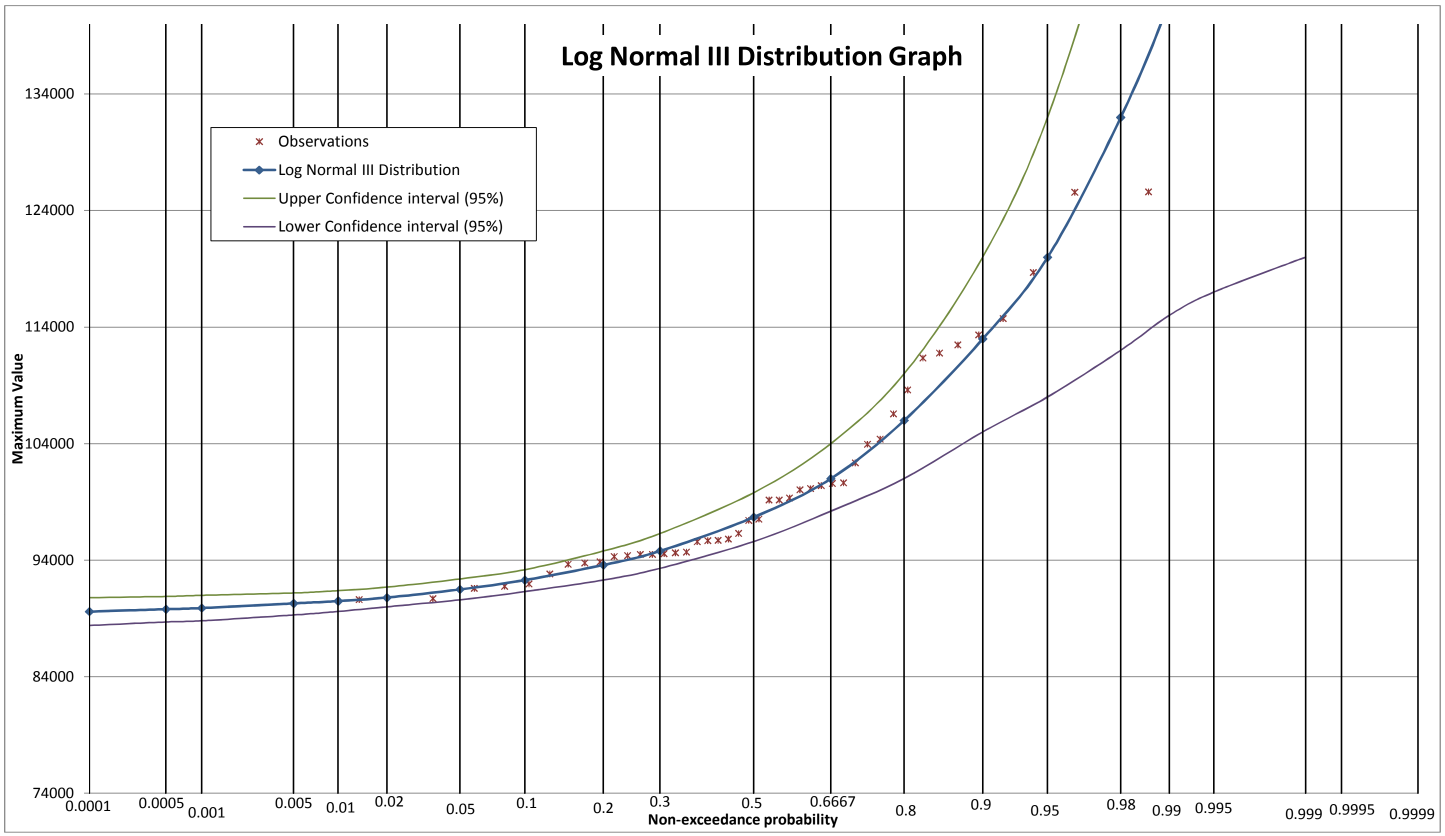
Normal Distribution
Visual Goodness-of-Fit Test Notes:

Several points above the upper confidence interval at both low and high end of non-exceedance probability and below the lower confidence interal near 0.5 non-exceedance probability



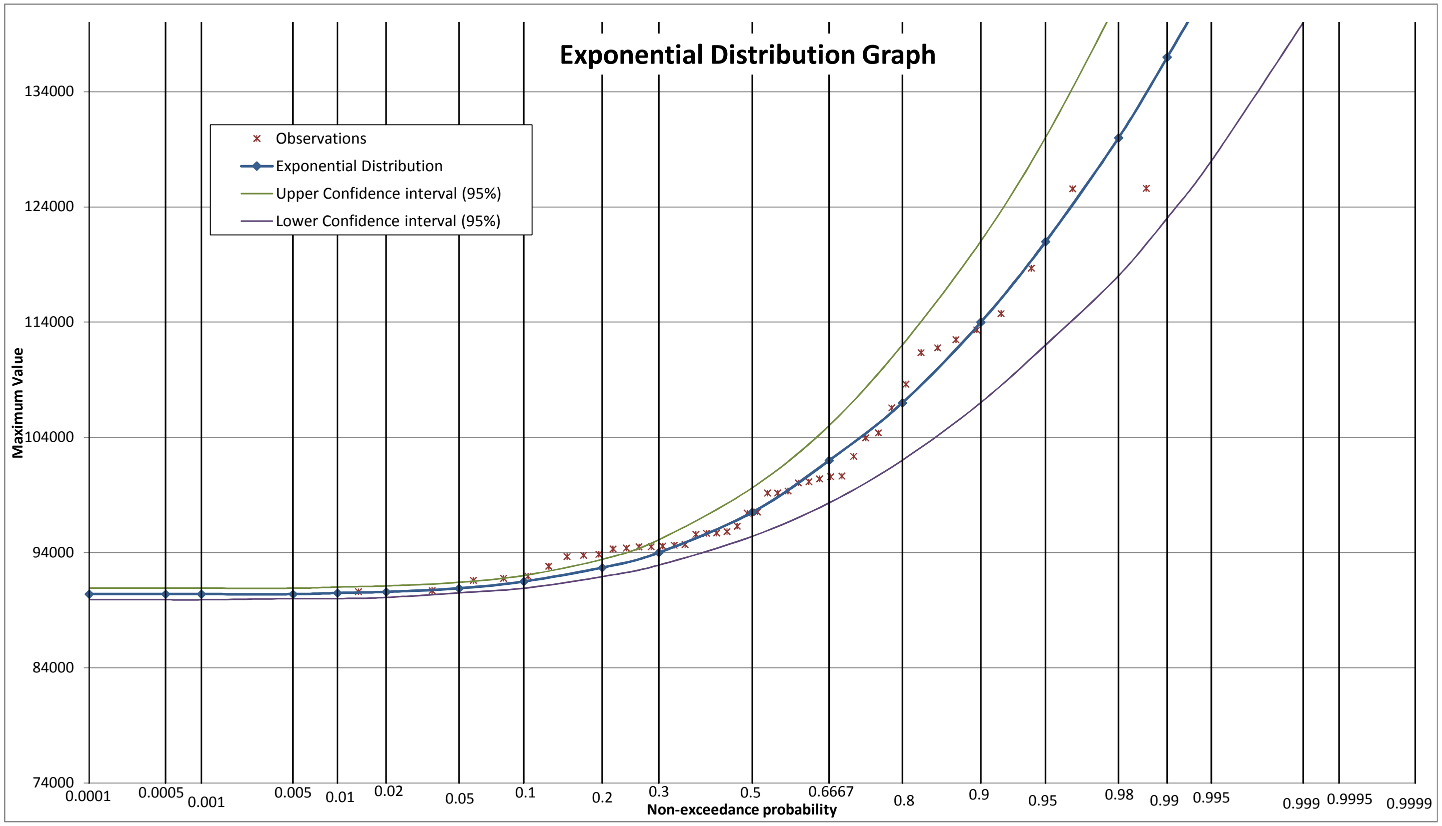
Log Normal Distribution
Visual Goodness-of-Fit Test Notes:

A few points above the upper confidence interval at both low and high end of non-exceedance probability and some below the lower confidence interal near 0.5 non-exceedance probability



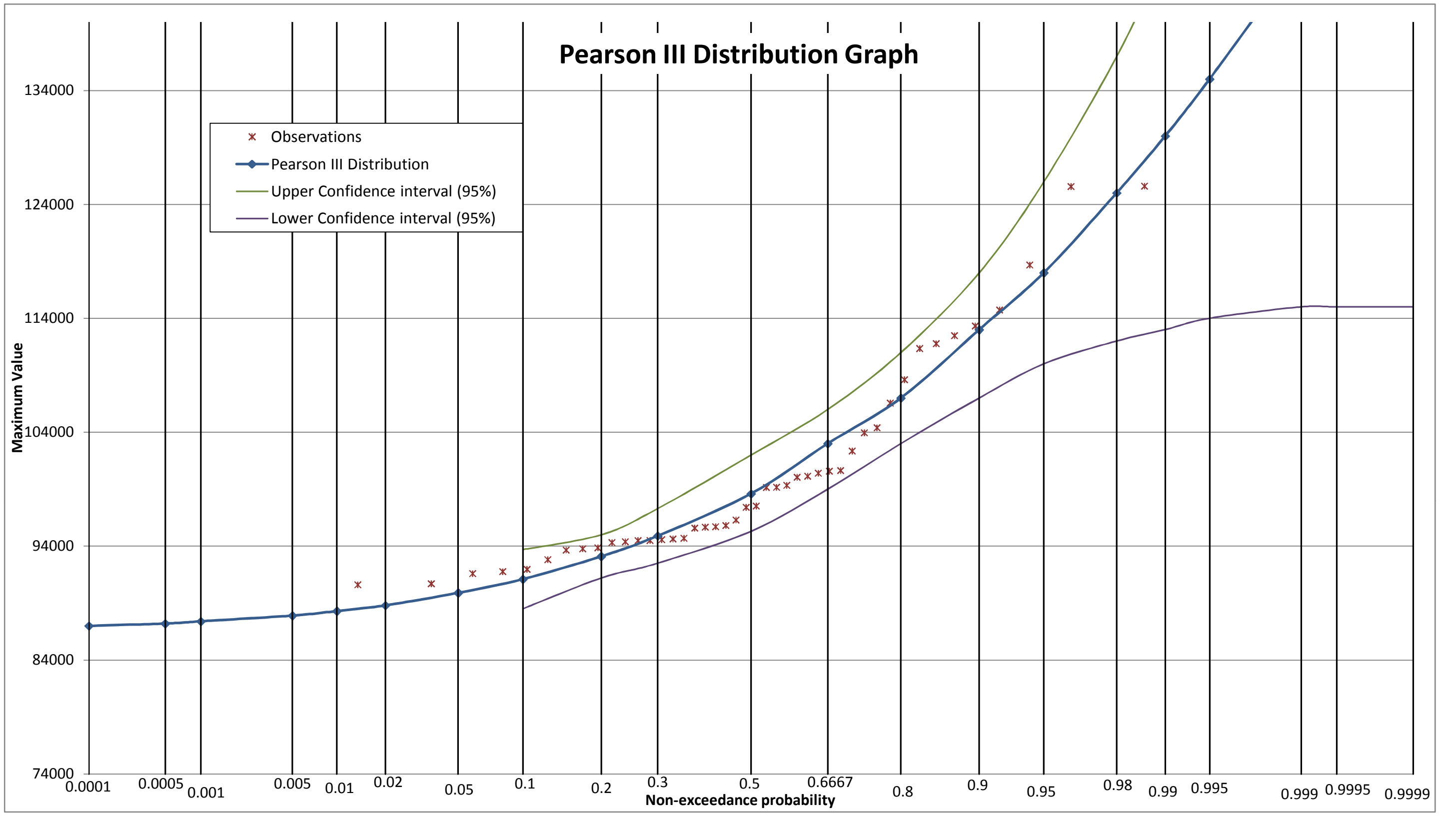
Log Normal III Distribution
Visual Goodness-of-Fit Test Notes:

All points within confidence interval, very good fit for majority of points with the exception of the largest event.
Good fit, may be over estimating values at high non-exceedance probabilities.
As the values are for pond volumes, it might be possible that the pond overflowed. The data were checked and the pond design information indicates that these levels are below the spillway crest.

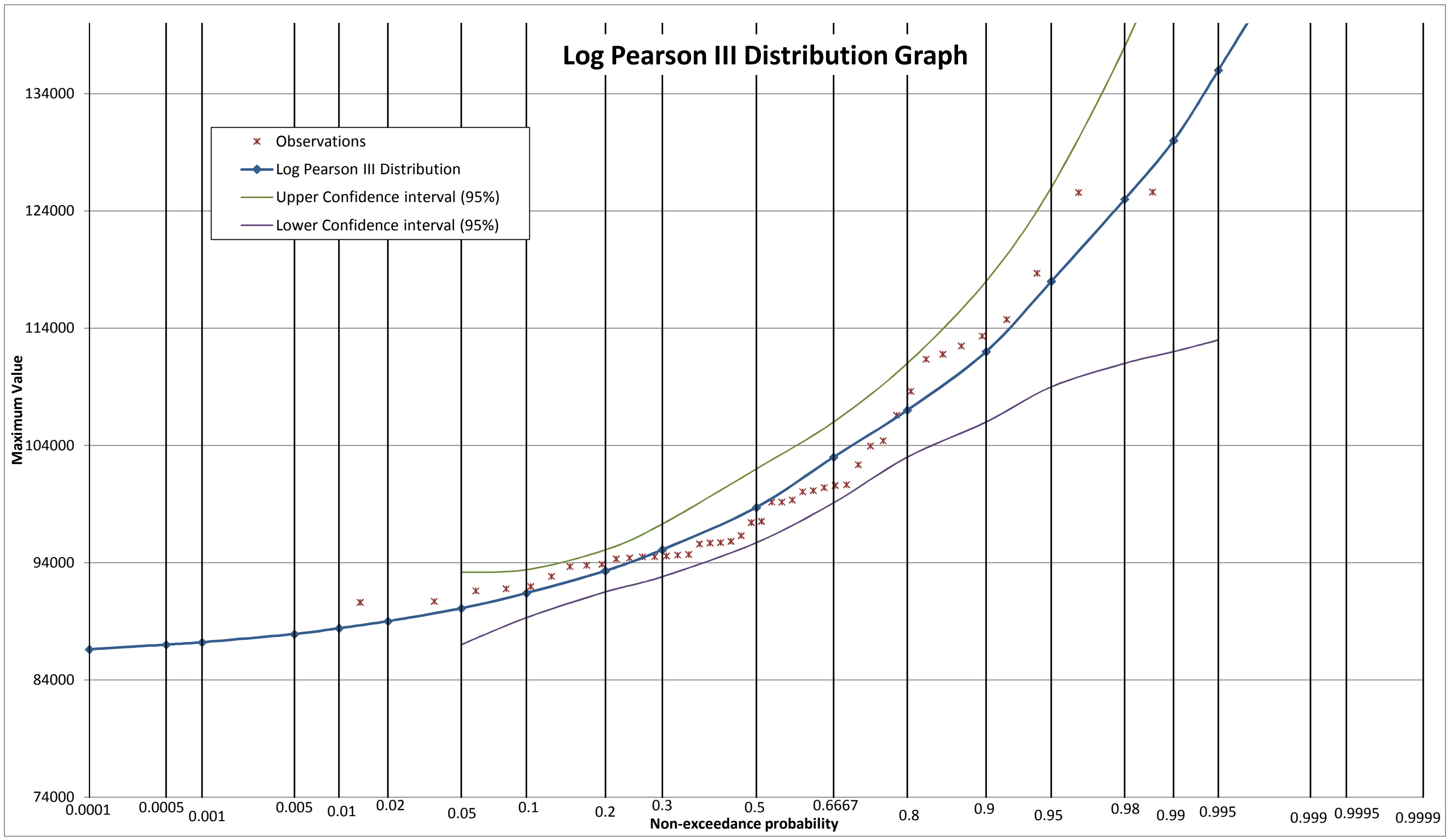


Exponential Distribution
Visual Goodness-of-Fit Test Notes:

Some points outside of confience interval for lower non-exceedance probabilities. Good fit at higher non-exceedance probabilities.



Pearson III Distribution Visual Goodness-of-Fit Test Notes:
Good visual fit throughout distribution

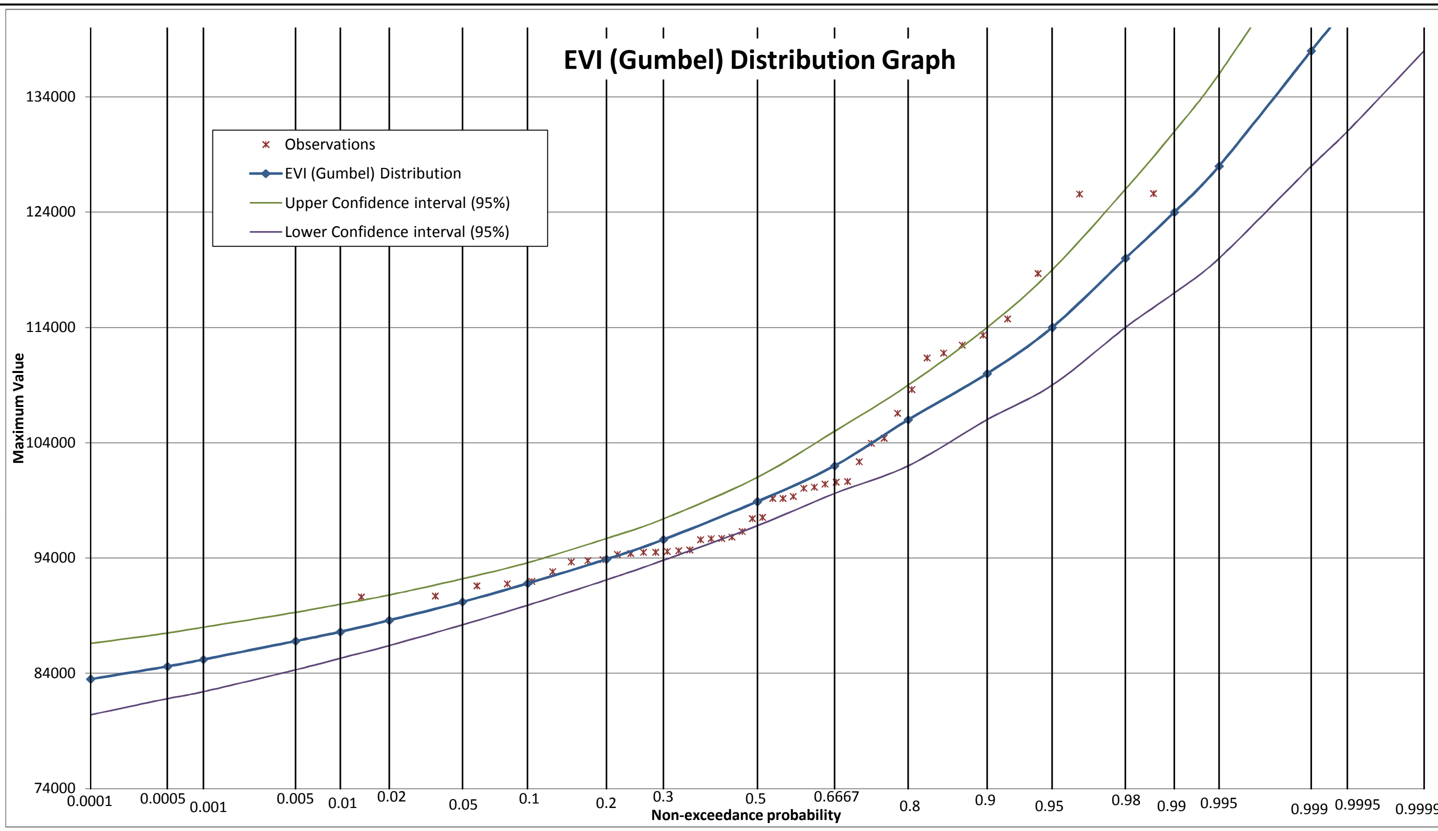


Log Pearson III Distribution
Visual Goodness-of-Fit Test Notes:

Good visual fit throughout distribution

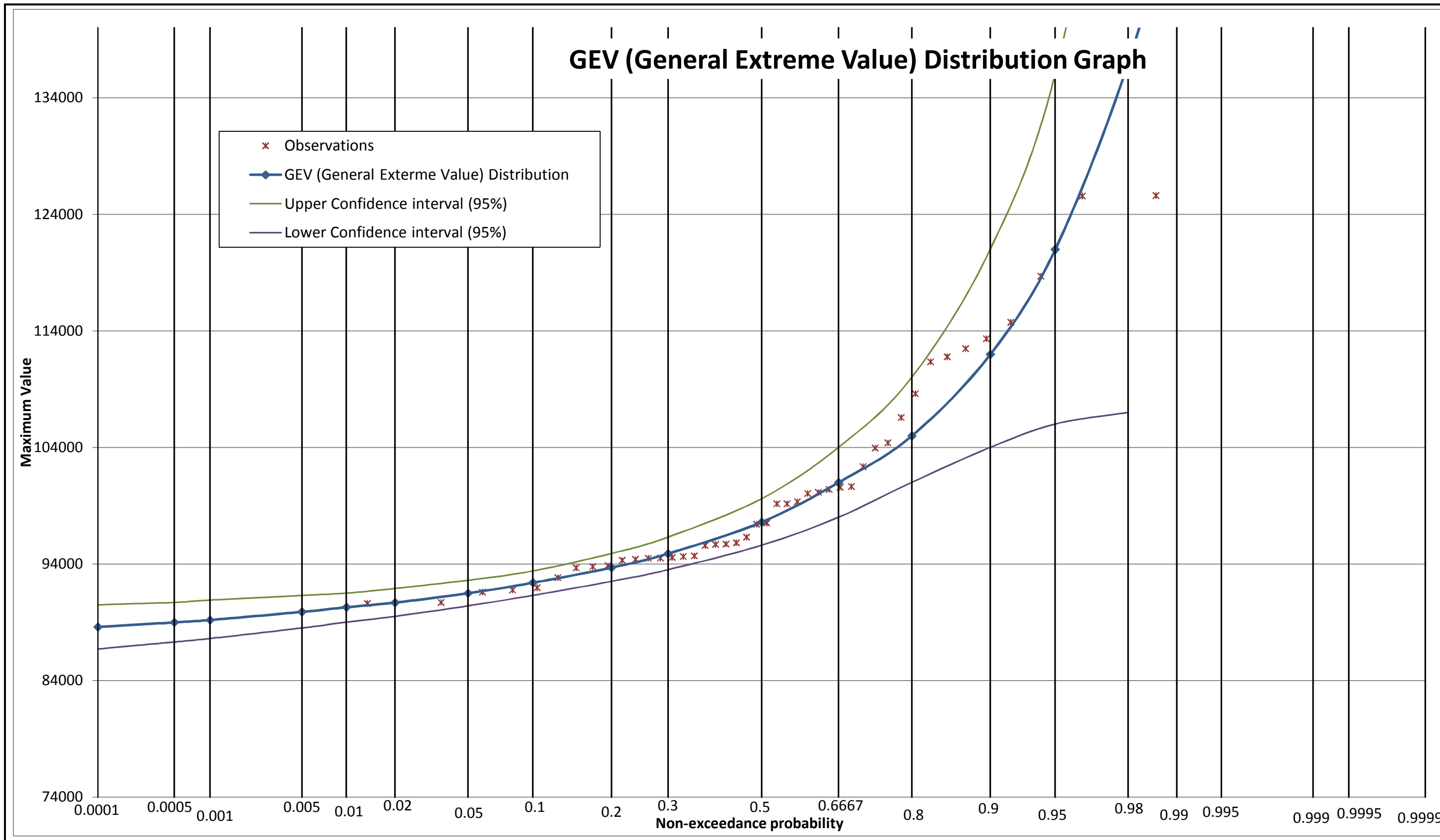
The log-Pearson III applies to hydrologic frequency analsisys only when the following shape tests pass

Lambda shape test	Pass
Alpha shape test	Pass



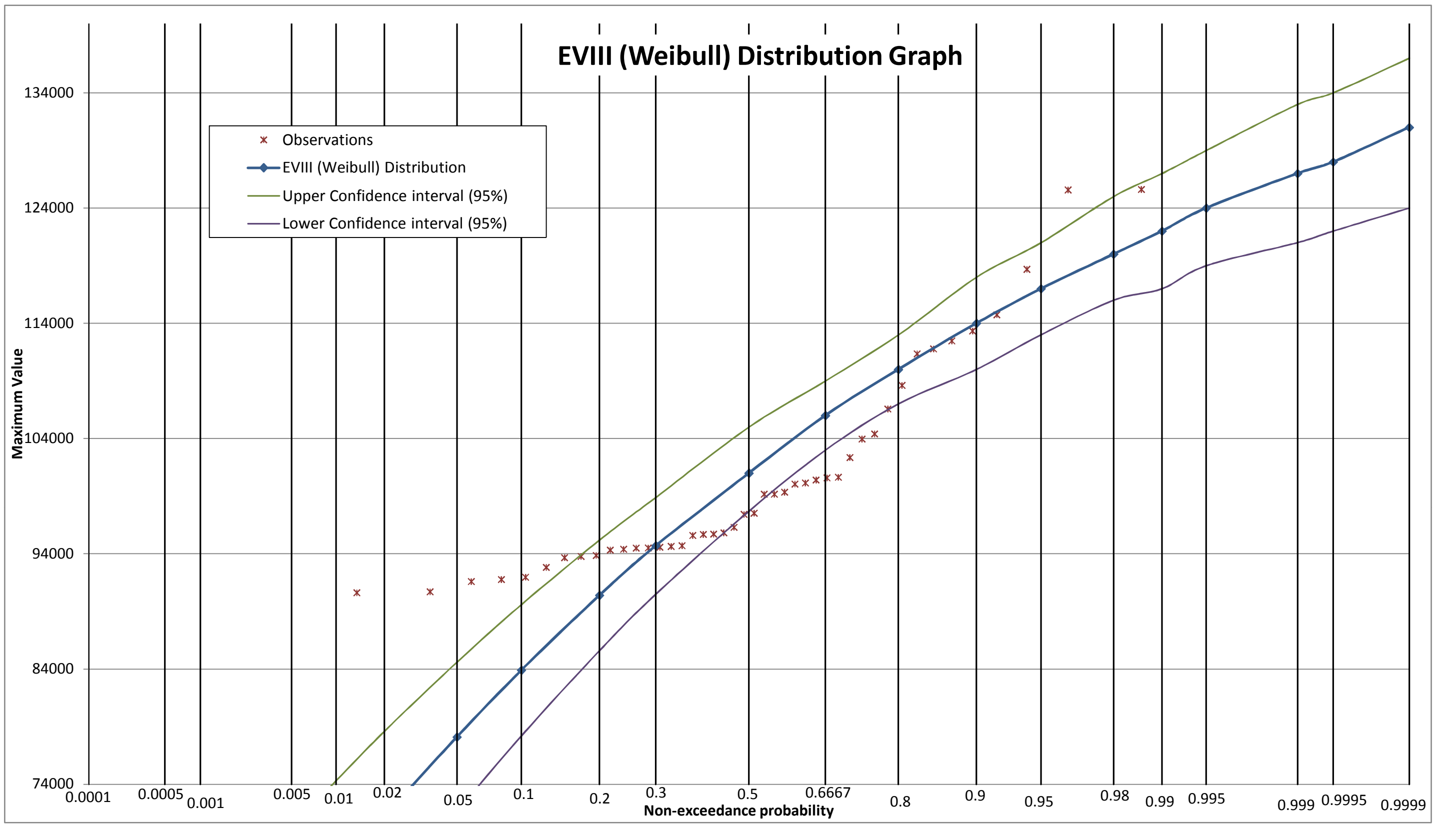
EVI (Gumbel) Distribution
Visual Goodness-of-Fit Test Notes:

Distribution appears to be underestimating values at high exceedance probabilities



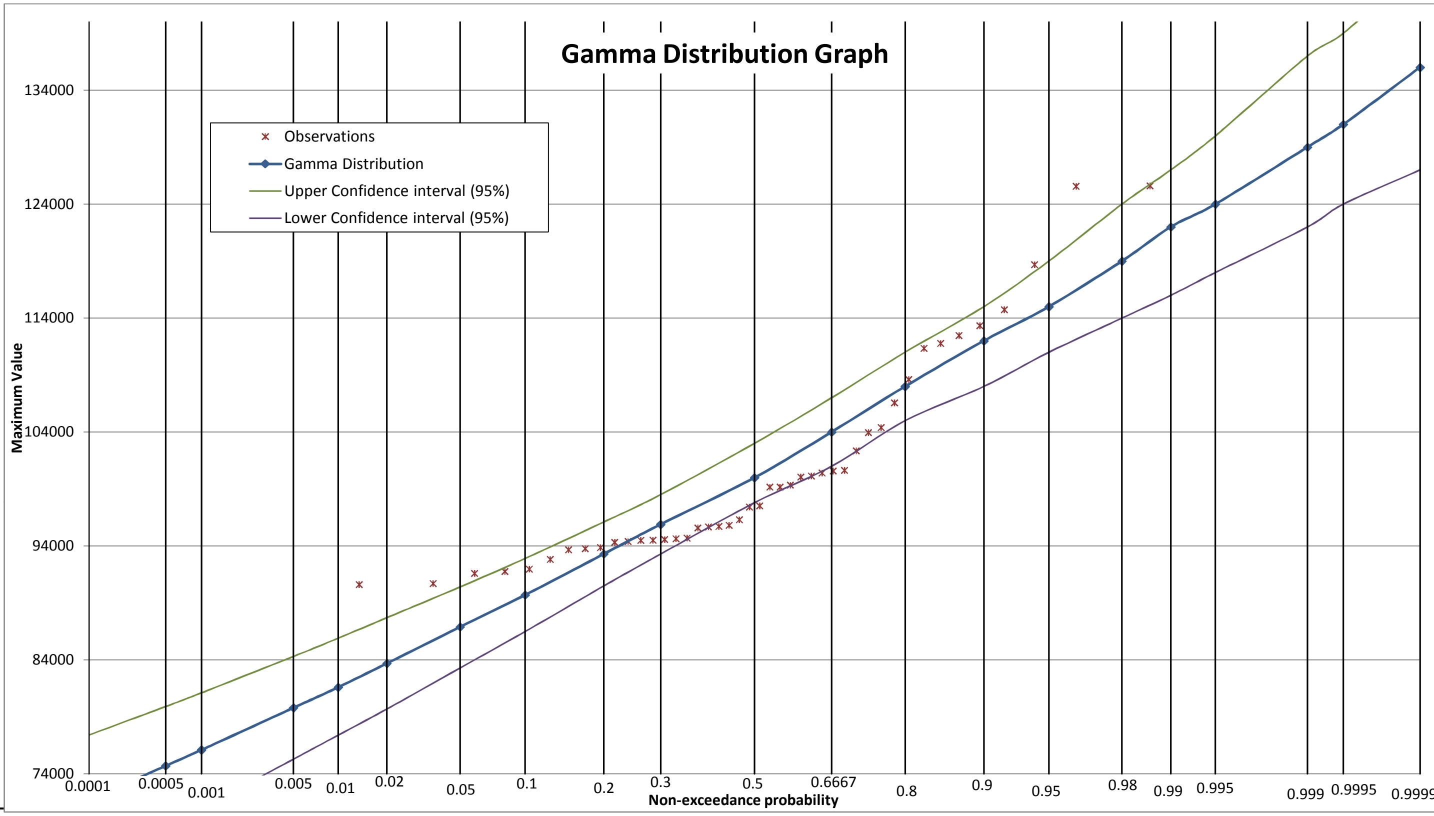
GEV (General Extreme Value) Distribution Visual Goodness-of-Fit Test Notes:

Good fit, may be over estimating values at high non-exceedance probabilities



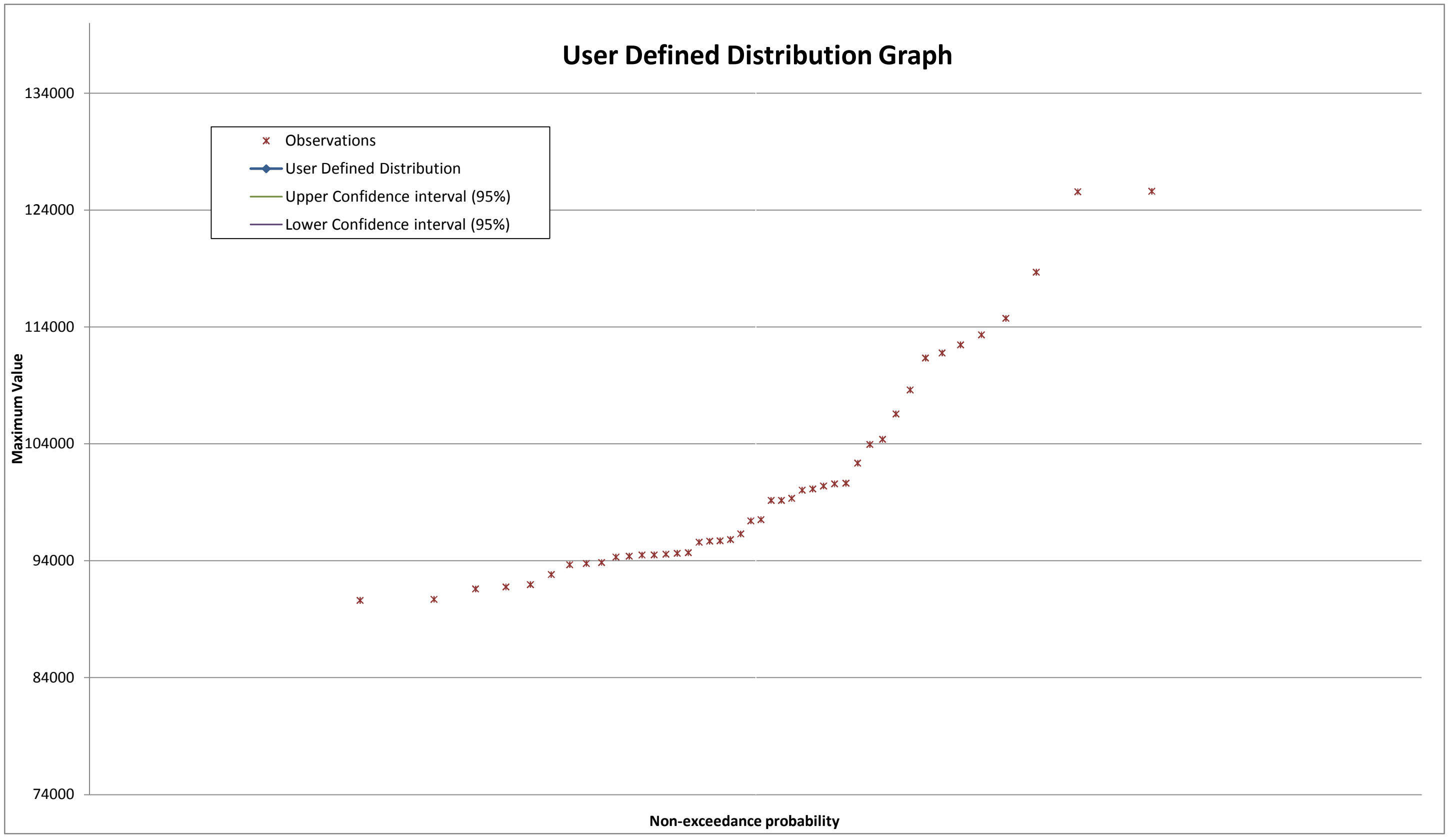
EVIII (Weibull) Distribution
Visual Goodness-of-Fit Test Notes:

Most observations fall outside of confidence intervals with likely under-estimating values at high non-exceedance probabilities



Gamma Distribution
Visual Goodness-of-Fit Test Notes:

Most observations fall outside of confidence intervals with likely under-estimating values at high non-exceedance probabilities



User Defined Distribution
Visual Goodness-of-Fit Test Notes:

Input comments on Visual Goodness-of-Fit here

Numerical Tests

Choose Significance Level (alpha) : **5%**

1) Anderson-Darling Test (1952)

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\cdot[\ln F(X_i) + \ln(1-F(X_{n-i+1}))]$$

H0= Data follows specified distribution
HA= Data does not follow the specified distribution

Distribution Type:	Critical Value at 10%	Critical Value at 5%	Critical Value at 1%	A2	Hypothesis	Rank (1 = best fit)
Normal	1.929	2.502	3.907	2.321	Accept H0	9
Lognormal	1.929	2.502	3.907	1.981	Accept H0	7
Lognormal III	1.929	2.502	3.907	0.314	Accept H0	1
Exponential	1.929	2.502	3.907	0.591	Accept H0	3
Pearson III	1.929	2.502	3.907	0.775	Accept H0	5
Log Pearson III	1.929	2.502	3.907	0.701	Accept H0	4
Gumbel	1.929	2.502	3.907	1.091	Accept H0	6
GEV	1.929	2.502	3.907	0.339	Accept H0	2
Weibull	1.929	2.502	3.907	3.154	Reject H0	10
Gamma	1.929	2.502	3.907	2.126	Accept H0	8

*Critical values based on values calculated by EasyFit Software

2) Kolmogorov-Smirnov Test (1933)

$$F_n(x) = \frac{1}{n}\cdot\left[\text{Number of observations} \le x\right] \qquad D_n = \sup_x |F_n(x) - F(x)|$$

H0= Data follows specified distribution
HA= Data does not follow the specified distribution

Distribution Type:	Critical Value at 10%	Critical Value at 5%	Critical Value at 1%	Dn	Hypothesis	Rank (1 = best fit)
Normal	0.184	0.205	0.246	0.179	Accept H0	9
Lognormal	0.184	0.205	0.246	0.162	Accept H0	7
Lognormal III	0.184	0.205	0.246	0.073	Accept H0	1
Exponential	0.184	0.205	0.246	0.139	Accept H0	6
Pearson III	0.184	0.205	0.246	0.097	Accept H0	4
Log Pearson III	0.184	0.205	0.246	0.092	Accept H0	3
Gumbel	0.184	0.205	0.246	0.119	Accept H0	5
GEV	0.184	0.205	0.246	0.073	Accept H0	2
Weibull	0.184	0.205	0.246	0.204	Accept H0	10
Gamma	0.184	0.205	0.246	0.169	Accept H0	8

Expected Probability Analysis (Alberta Environment 1981)

Chance of Occurrence of Expected Flood Events: 83%

$$P_{nr} = p_n q_r - \frac{n\,(r!)}{(r-n)!\,n!}$$

Return Period (Years)	Range of Expected Number of Flood Events Greater than or Equal to the Return Period		Number of Flood Events Greater than or Equal to the Return Period based on Distribution									
	Low	High	Normal	Lognormal	Lognormal III	Exponential	Pearson III	Log Pearson III	Gumbel	GEV	Weibull	Gamma
10000	0	0	0	0	0	0	0	0	0	0	0	0
2000	0	0	0	0	0	0	0	0	0	0	0	0
1000	0	0	0	0	0	0	0	0	0	0	0	0
200	0	1	2	2	0	0	0	0	0	0	2	2
100	0	1	2	2	0	0	0	0	2	0	2	2
50	0	2	2	2	0	0	2	2	2	0	2	2
20	0	5	3	3	2	2	3	3	4	2	3	3
10	1	7	6	6	5	4	5	6	8	6	4	6
5	5	13	9	9	10	9	9	9	10	10	8	9
3	10	19	11	11	13	13	12	12	13	13	10	11
2	17	27	13	18	21	22	21	21	21	21	13	18
1.4286	27	36	24	25	28	35	28	28	27	28	29	24
1.25	31	39	38	38	38	39	38	38	35	37	44	38
1.1111	37	43	44	44	39	42	42	42	40	39	44	44
1.0526	39	44	44	44	42	42	44	44	44	42	44	44
1.0204	42	44	44	44	42	44	44	44	44	43	44	44
1.0101	43	44	44	44	44	44	44	44	44	44	44	44
1.005	43	44	44	44	44	44	44	44	44	44	44	44
1.001	44	44	44	44	44	44	44	44	44	44	44	44
1.0005	44	44	44	44	44	44	44	44	44	44	44	44
1.0001	44	44	44	44	44	44	44	44	44	44	44	44
		# our of range	5	4	0	0	0	0	2	0	5	4
		Rank	9	7	1	1	1	1	6	1	9	7

Least Squares Ranking

Distribution Type:	Standard Error	Rank
Normal	3549	9
Lognormal	3098	6
Lognormal III	2215	4
Exponential	1662	1
Pearson III	1667	3
Log Pearson III	1666	2
Gumbel	2493	5
GEV	3374	8
Weibull	6074	10
Gamma	3233	7

$$SE_j = \sqrt{\frac{1}{n-m_j}\sum_{i=1}^{n}(x_i - y_i)^2}$$

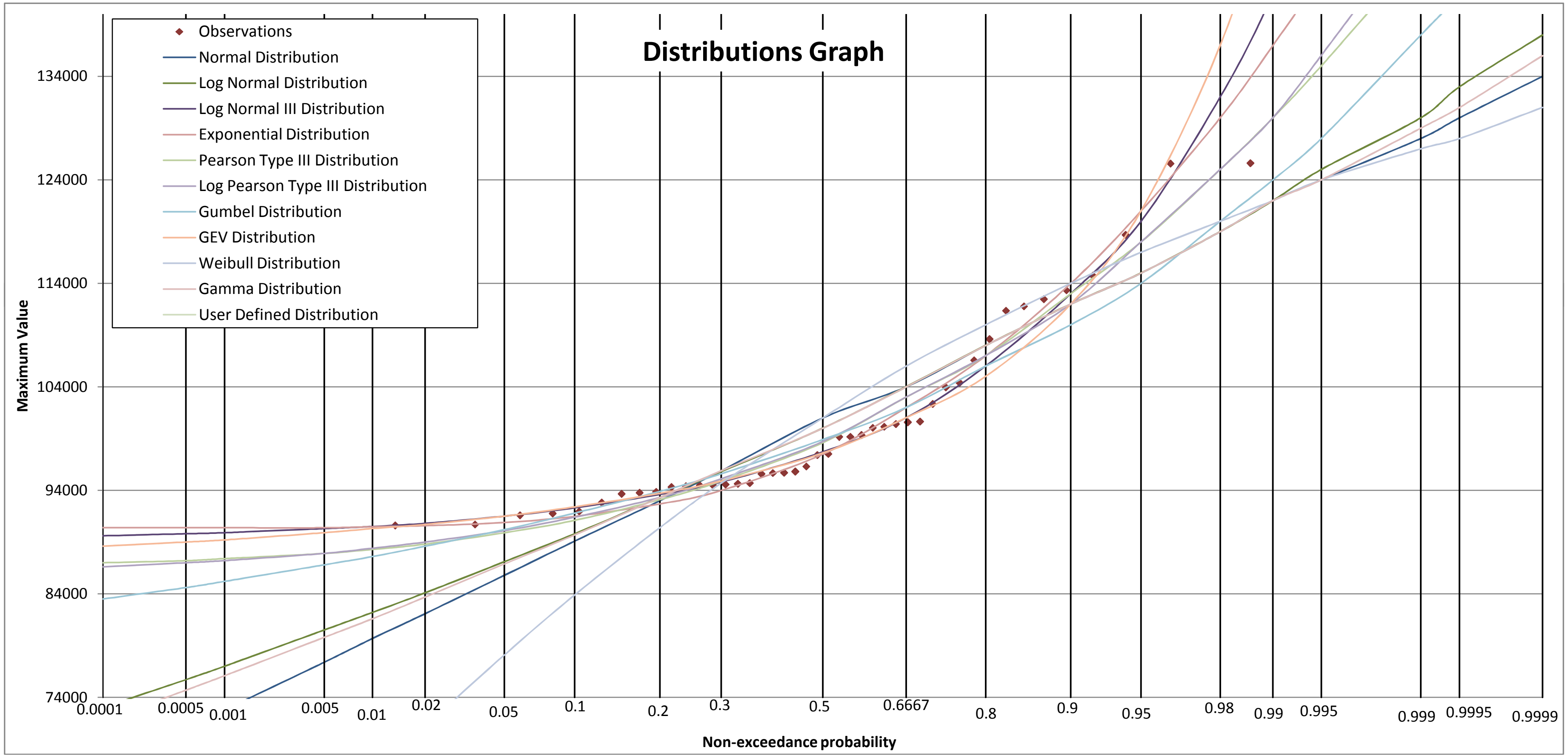
Sampling and Distribution Uncertainty

NOTES

- Select a distribution type and a return period of interest
- The Sample Uncertainty, Distribution Uncertainty and Total Uncertainty will be displayed on the right
- The plot below displays all the distributions input in the Frequency Analysis Input Tab

Return Period of Interest (Years)	50
Distribution Type	Lognormal III
Corresponding Value	132000

Sampling Uncertainty at (95%) Confidence Interval ±	20000
Distribution Uncertainty ±	2250
Total Uncertainty ±	22300



Summary Sheet

Initial Statistical Tests:

Tests for Stationarity

Test	Result
Spearman Rank Order Correlation Coefficient	No Significant Trend at 0.05 Significance Level
Mann-Whitney Test for jump (a.k.a. Mann-Whitney U test)	No Jump at 0.05 Significance Level
Wald-Wolfowitz Test (The runs test)	

Tests for Homogeneity

Test	Result
Mann-Whitney Test for jump (a.k.a. Mann-Whitney U test)	Sample is Homogeneous at 0.05 Significance Level
Terry Test	Sample is Homogeneous at 0.05 Significance Level

Tests for Independence

Test	Result
Spearman Rank Order Correlation Coefficient	Data is independent at 0.05 Significance Level
Wald-Wolfowitz Test for Independence	Data is independent at 0.05 Significance Level
Anderson Test	Data is independent at 0.05 Significance Level

Test for Outliers

Test	Result
Grubbs and Beck Test for Outliers	
Are any high outliers present?	No High Outliers Present
Are and low outliers present?	No Low Outliers Present

Project Information

Project Name:	South Pond - Shepard Landfill
Project Description:	Stormwater Management Facility for Shepard Landfill (Maximum Annual Pond Volumes) - Independent Dataset Example
Location:	Calgary
Date:	12/12/2012
Designed by:	Charles Wojcik
Company Name:	AMEC
Reviewed by:	-

Numerical Goodness-of-fit Tests Results

Distribution Type	Numerical Goodness-of-fit Tests from Spreadsheet				Average of Ranks	Ranking from Numerical Tests	Numerical Goodness-of-fit Tests from Hyfran (Input by user)		Notes from Visual Goodness-of-fit Test
	A-D Test	K-S Test	Expected Probability	Least Squares Ranking			BIC	AIC	
Normal	9	9	9	9	9.00	9	9	9	Several points above the upper confidence interval at both low and high end of non-exceedance probability and below the lower confidence interal near 0.5 non-exceedance probability
Lognormal	7	7	7	6	6.75	7	7	7	A few points above the upper confidence interval at both low and high end of non-exceedance probability and some below the lower confidence interal near 0.5 non-exceedance probability
Lognormal III	1	1	1	4	1.75	1	2	2	All points within confidence interval, very good fit for majority of points with the exception of the largest event. Good fit, may be over estimating values at high non-exceedance probabilities. As the values are for pond volumes, it might be possible that the pond overflowed. The data were checked and the pond design information indicates that these levels are below the spillway crest.
Exponential	3	6	1	1	2.75	3	1	1	Some points outside of confience interval for lower non-exceedance probabilities. Good fit at higher non-exceedance probabilities.
Pearson III	5	4	1	3	3.25	4	6	5	Good visual fit throughout distribution
Log Pearson III	4	3	1	2	2.50	2	5	4	Good visual fit throughout distribution
Gumbel	6	5	6	5	5.50	6	4	6	Distribution appears to be underestimating values at high exceedance probabilities
GEV	2	2	1	8	3.25	4	3	3	Good fit, may be over estimating values at high non-exceedance probabilities
Weibull	10	10	9	10	9.75	10	10	10	Most observations fall outside of confidence intervals with likely under-estimating values at high non-exceedance probabilities
Gamma	8	8	7	7	7.50	8	8	8	Most observations fall outside of confidence intervals with likely under-estimating values at high non-exceedance probabilities

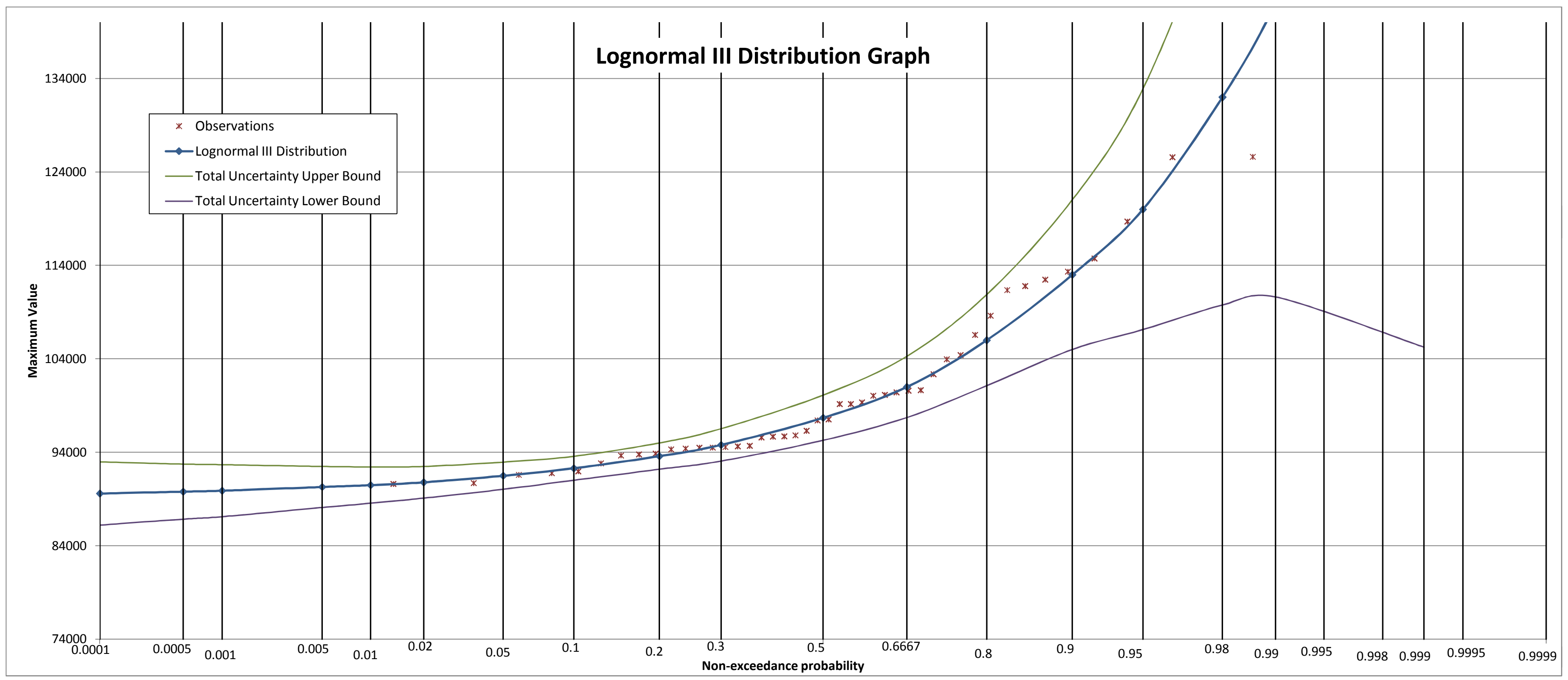
Selected Distribution and Results

Distribution type chosen based on visual and numerical goodness-of-fit tests: **Lognormal III**

Instructions:
- Based on the results of the numerical and visual goodness-of-fit tests presented above, choose the preferred distribution in the cell on the left

Return Period	Probability	Magnitude	Total Uncertainty (Upper Bound)	Total Uncertainty (Lower Bound)
10000	0.9999	248000	#N/A	#N/A
2000	0.9995	202000	#N/A	#N/A
1000	0.9990	186000	267000	105000
500	0.9980	171000	238000	107000
200	0.9950	154000	197000	111000
100	0.9900	142000	173000	111000
50	0.9800	132000	154000	110000
20	0.9500	120000	133000	107000
10	0.9000	113000	121000	105000
5	0.8000	106000	111000	101000
3	0.6667	101000	104000	97700
2	0.5000	97700	100000	95300
1.4286	0.3000	94800	96500	93100
1.25	0.2000	93600	95000	92200
1.1111	0.1000	92300	93600	91000
1.0526	0.0500	91500	93000	90100
1.0204	0.0200	90800	92500	89100
1.0101	0.0100	90500	92400	88600
1.005	0.0050	90300	92500	88100
1.001	0.0010	89900	92700	87100
1.0005	0.0005	89800	92700	86900
1.0001	0.0001	89600	93000	86200

*Total uncertainty is based on sampling uncertainty at ((95%) Confidence Interval) plus distribution uncertainty of Top 4 distributions (based on numerical goodness of fit tests)



Errors and Warnings

Cumulative distribution function warning
No warning
No warning
No warning
No warning
No warning
No warning
No warning
No warning
No warning
No warning

If a warning is present, please check if hyfran output results were pasted correctly. If hyfran results were pasted correctly the warning signifies that the Continious Distribution Function (CDF) used in this workbook does not produce same output values as the input frequency analysis results, which in turn indicates that the numerical goodness-of-fit tests calculated by this spreadsheet for this distribution may be based on inaccurate numbers. Another possible solution would be to use a different method of estimating the CDF parameters for example: method of weighted moments.

Appendix B

Evaporation Pond

APPENDIX B

Evaporation Pond

This data set represents the annual maximum series for an evaporation pond within the City of Calgary.

Year	Maximum Annual Volume (m^3)	Year (continued)	Maximum Annual Volume (m^3)
1960	175429.2	1986	212838.7
1961	140225.6	1987	208496.3
1962	139969.8	1988	163642.9
1963	128655.6	1989	164685.2
1964	158122.4	1990	148228.9
1965	279614.0	1991	161572.3
1966	320651.9	1992	189779.9
1967	285739.6	1993	235806.2
1968	147232.6	1994	225649.3
1969	182995.3	1995	174598.4
1970	193868.8	1996	191010.8
1971	157185.5	1997	223957.9
1972	171580.3	1998	288364.5
1973	185033.1	1999	276010.4
1974	173016.2	2000	237963.6
1975	124281.4	2001	179862.1
1976	107020.9	2002	118502.1
1977	119620.4	2003	142927.9
1978	208577.6	2004	125533.2
1979	211793.0	2005	229135.5
1980	154689.1	2006	235101.8
1981	208363.4	2007	311667.8
1982	215528.3	2008	290380.9
1983	182988.1	2009	270393.0
1984	124045.3	2010	177500.1
1985	152469.4		

Evaporation Pond

The below data set is the result of the maximum annual evaporation pond volumes being de-correlated, a transformation through which the data set has become independent. The results of the frequency analysis on the independent data set will have to be inverted to give the distributed maximum annual evaporation volumes.

Year	De-correlated Values	Year (continued)	De-correlated Values
1960		1986	121075
1961	34643	1987	80399
1962	55575	1988	38159
1963	44415	1989	66197
1964	80691	1990	49113
1965	184448	1991	72361
1966	152366	1992	92537
1967	92755	1993	121587
1968	-24740	1994	83729
1969	94383	1995	38791
1970	83733	1996	85929
1971	40505	1997	108998
1972	76978	1998	153575
1973	81767	1999	102458
1974	61654	2000	71847
1975	20152	2001	36644
1976	32222	2002	10252
1977	55210	2003	71607
1978	136584	2004	39512
1979	86261	2005	153583
1980	27221	2006	97196
1981	115264	2007	170172
1982	90125	2008	102803
1983	53272	2009	95627
1984	13914	2010	14764
1985	77813		

*Shaded value has been removed because it failed the Low Outlier Test

Results of De-Correlated Value Distribution (Weibull)

T	q	Volume	Inverted Volumes
10000	0.9999	271000	381132
2000	0.9995	246000	356132
1000	0.9990	235000	345132
200	0.9950	205000	315132
100	0.9900	192000	302132
50	0.9800	176000	286132
20	0.9500	154000	264132
10	0.9000	135000	245132
5	0.8000	113000	223132
3	0.6667	93400	203532
2	0.5000	74200	184332
1.4286	0.3000	53200	163332
1.25	0.2000	42000	152132
1.1111	0.1000	28900	139032
1.0526	0.0500	20100	130232
1.0204	0.0200	12600	122732
1.0101	0.0100	8900	119032
1.005	0.0050	6280	116412
1.001	0.0010	2800	112932
1.0005	0.0005	1980	112112
1.0001	0.0001	884	111016

DFASCC

Data and Frequency Analysis Spreadsheet for the City of Calgary
Version 1.2

PROJECT INFORMATION SHEET

Project Name:	Evaporation Pond - Example #1 (Dependent data set)
Project Description:	A typical evaporation pond somewhere in Calgary. The area has a size of 100 hectares of which 50% is developed area, with the remainder undeveloped. All runoff is to be fully contained. In the first example, we have only two catchment areas, one for the developed areas (say 50 ha with 60% hard) and landscaped and one for the area around the pond (also 50ha). The pond is assumed at a zero m2 footprint at the bottom, with 0.57% sideslopes. No irrigation of the landscaped areas.
Location:	Say somewhere in northeast or southeast Calgary
Date:	09/11/2012
Designed by:	Bert van Duin
Company Name:	City of Calgary Water Resources - Infrastructure Planning
Reviewed by:	Who volunteers?

Clear Project Information Sheet

Hydrologic Data Series Input

LEGEND

User Input	Negative Result
Calculated Cells	Positive Result

NOTES

- This Spreadsheet is designed for a maximum of 10,000 entries (if more are required then formulas need to be adjusted)
-Input dataset must be based on uniform time distribution (i.e.: daily, weekly, monthly, yearly) and must not include multiple values for any of the time steps (duplicates highlighted in red)
-Input dataset must not have any missing cells of data (cells with "0" will be treated as having a value of 0)

Clear All Input Data

Index	Date	Value	Empirical Probability of Non-Exceedance
1	1960	175429.2	0.441
2	1961	140225.6	0.168
3	1962	139969.8	0.148
4	1963	128655.6	0.129
5	1964	158122.4	0.305
6	1965	279614	0.891
7	1966	320651.9	0.988
8	1967	285739.6	0.910
9	1968	147232.6	0.207
10	1969	182995.3	0.520
11	1970	193868.8	0.598
12	1971	157185.5	0.285
13	1972	171580.3	0.383
14	1973	185033.1	0.539
15	1974	173016.2	0.402
16	1975	124281.4	0.090
17	1976	107020.9	0.012
18	1977	119620.4	0.051
19	1978	208577.6	0.656
20	1979	211793	0.676
21	1980	154689.1	0.266
22	1981	208363.4	0.617
23	1982	215528.3	0.715
24	1983	182988.1	0.500
25	1984	124045.3	0.070
26	1985	152469.4	0.246
27	1986	212838.7	0.695
28	1987	208496.3	0.637
29	1988	163642.9	0.344
30	1989	164685.2	0.363
31	1990	148228.9	0.227
32	1991	161572.3	0.324
33	1992	189779.9	0.559
34	1993	235806.2	0.813
35	1994	225649.3	0.754
36	1995	174598.4	0.422
37	1996	191010.8	0.578
38	1997	223957.9	0.734
39	1998	288364.5	0.930
40	1999	276010.4	0.871
41	2000	237963.6	0.832
42	2001	179862.1	0.480
43	2002	118502.1	0.031
44	2003	142927.9	0.188
45	2004	125533.2	0.109
46	2005	229135.5	0.773
47	2006	235101.8	0.793
48	2007	311667.8	0.969
49	2008	290380.9	0.949
50	2009	270393	0.852
51	2010	177500.1	0.461

Basic Characteristics	
Number of Data Entries	51
Maximum Value	321000
Minimum Value	107000
Average (Mean) Value	193000
Median Value	183000
Standard Deviation	54400
Variance	2960000000
Variation coefficient (Cv)	0.282
Skewness coefficient (Cs)	0.634
Kurtosis	2.48

*Values assumed to be sample not full population

Empirical Probability of Non-Exceedance (Plotting Position) based on:
F(x(k)) = (k-a)/ (n-2a+1), 0 <=a<=0.5

a =	0.4	Cunnane (1978)
k=	rank of the even in question (in ascending order)	
n=	51	

Stationarity

Test for Trend:

Choose Significance Level (alpha):	5%

1) Spearman Rank Order Correlation Coefficient

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}.$$

H_0= Data has no trend

Spearman Correlation Coefficient:	0.294

based on z	Trend Detected at 0.05 Significance Level
based on t	Trend Detected at 0.05 Significance Level

When there are no ties in rankings:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

Spearman Correlation Coefficient:	0.294
t-distribution value	2.150
Degrees of freedom	49

T (Adjustment for ties) =	0
Standard Normal (z)=	2.099

Tests for Jump:

2) Mann-Whitney Test for jump (a.k.a. Mann-Whitney U test)

Index number of subsample divide	26

H_0= Independent samples drawn from the same population (No Jump)

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}$$

Presence of Jump Possible at 0.05 Significance Level

Number of values in sample 1 n_1=	26
Number of values in sample 2 n_2=	25
Total of Ranking in sample 1 R_1=	564
Total of Ranking in sample 2 R_2=	762
U_1=	213

$$U_1 + U_2 = n_1 n_2.$$

U_2=	437
U (Minimum of U_1 and U_2)=	213
Standard Normal (z)=	-2.110

3) Wald-Wolfowitz Test (The runs test)

$$\mu = \frac{2\, N_+\, N_-}{N} + 1, \qquad \sigma^2 = \frac{2\, N_+\, N_-\, (2\, N_+\, N_- - N)}{N^2\, (N-1)} = \frac{(\mu - 1)(\mu - 2)}{N - 1}.$$

Number of data greater than median N_+ =	25
Number of data less than median N_- =	25
Total number of runs =	19

H_0= Data represent sample of single independently distributed random variable (No Jump)

Mean =	26.0
Variance =	12.2
Standard Normal (z)=	-2.1

Presence of Jump Possible at 0.05 Significance Level

Homogeneity

Choose Significance Level (alpha): 5%

Mann-Whitney Test for homogeneity (a.k.a. Mann-Whitney U test)

Index number of subsample divide	26

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$

H_0= There is homogeneity between samples with respect to probability of random drawing of a larger observation

Sample is not Homogeneous at 0.05 Significance Level

Number of values in sample 1 n_1=	26
Number of values in sample 2 n_2=	25
Total of Ranking in sample 1 R_1=	564
Total of Ranking in sample 1 R_2=	762
U_1=	213

$$U_1 + U_2 = n_1 n_2.$$

U_2=	437
U (Minimum of U_1 and U_2)=	213
Standard Normal (z)=	-2.110

Terry Test

Index number of subsample divide	26

H_0= There is homogeneity between samples with respect to probability of random drawing of a larger observation

Total sample size	51
Subsample 1 (m)	26
Subsample 2 (n)	25

Sample is Homogeneous at 0.05 Significance Level

Standard Deviation =	3.516
Sum of ranks in first subsample c =	6.801
z =	1.934

Independence

Choose Significance Level (alpha):	5%

1) Spearman Rank Order Correlation Coefficient

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}.$$

H_0= Data is independent

Spearman Correlation Coefficient:	0.54

Non-independence Detected at 0.05 Significance Level

When there are no ties in rankings:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

Spearman Correlation Coefficient:	0.54
t-distribution value	5.33
Degrees of freedom	49

2) Wald-Wolfowitz Test

$$R = \sum_{i=1}^{N-1} x_i x_{i+1} + x_1 x_N$$

Statistic R	1.98E+12
Mean	1.89E+12
Variance	4.15E+20

H_0= Data is independent

Non-independence Detected at 0.05 Significance Level

Standard Normal (z)=	4.5

2) Anderson Test

$$r_1 = \left[\sum_{i=1}^{N-1} x_i x_{i+1} + x_1 x_N - \left(\sum_{i=1}^{N} x_i \right)^2 \Big/ N \right] \Big/ \left[\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2 \Big/ N \right]$$

Statistic r	0.603
Mean	-0.020
Variance	0.020

H_0= Data is independent

Non-independence Detected at 0.05 Significance Level

Standard Normal (z)=	4.4

Outliers

Significance Level (alpha):	10%

Grubbs and Beck test for Outliers

1) High Outliers

Assumption: logarithms of sample are normally distributed

X_h = exp (xmean+K_nS)
K(n) =-3.62201+6.2844N^1/4-2.49835N^1/2+0.491436N^3/4-0.037911N
K(n) = -0.9043+3.345*SQRT(log(n))-0.4046log(n)   for 5<n<150

Sample Size (n) =	51	
K(n) =	2.78	
K(n) for 5<n<150 =	2.78	
X_h=	401000	< Any value higher than X_h is considered a high outlier
Maximum Value	321000	
High Outliers	No High Outliers Present	

2) Low Outliers

X_h = exp (xmean-K_nS)
K(n) =-3.62201+6.2844N^1/4-2.49835N^1/2+0.491436N^3/4-0.037911N
K(n) = -0.9043+3.345*SQRT(log(n))-0.4046log(n)   for 5<n<150

Sample Size (n) =	51	
K(n) =	2.78	
K(n) for 5<n<150 =	2.78	
X_h=	86000	< Any value lower than X_h is considered a low outlier
Minimum Value	107000	
Low Outliers	No Low Outliers Present	

Dependent Dataset

Choose Significance Level (alpha):	5%

Autocorrelation coefficient

$$R_c(\tau)=\frac{\sum_{i=1}^{N-|\tau|}X_iY_{i+\tau}-\frac{1}{N-|\tau|}\left(\sum_{i=1}^{N-|\tau|}X_i\right)\left(\sum_{i=\tau+1}^{N}Y_i\right)}{\left[\sum_{i=1}^{N-|\tau|}X_i^2-\frac{1}{N-|\tau|}\left(\sum_{i=1}^{N-|\tau|}X_i\right)^2\right]^{0.5}\left[\sum_{i=1+|\tau|}^{N}Y_i^2-\frac{1}{N-|\tau|}\left(\sum_{i-1+|\tau|}^{N}Y_i\right)^2\right]^{0.5}}$$

H_0 - The data is not serially correlated

One Time Period Offset

Autocorrelation coefficient offset by one time period	r(1) =	0.602
t-distribution values for one time period offset	t =	5.275

Serial Correlation Detected at 0.05 Significance Level

Two Time Periods Offset

Autocorrelation coefficient offset by two time periods	r(2) =	0.115
t-distribution values for two time periods offset	t =	0.807

No Serial Correlation at 0.05 Significance Level

Instructions:
Compare the results of the autocorrelation tests for one time period offset and for the two time period offset. One of the following 2 scenarios will result:
1. The finding for the one period time step is serially correlated, and the finding for the two time step is also serially correlated. In this case, transposing the data series is unlikely to produce an independent data set suitable for frequency analysis. In this case, other methods, such as the Monte Carlo simulation are necessary.
2. The finding for the one period time step is serially correlated, and the finding for the two time step is NOT serially correlated. In this case, the data series should be transposed to produce an independent data set suitable for frequency analysis.

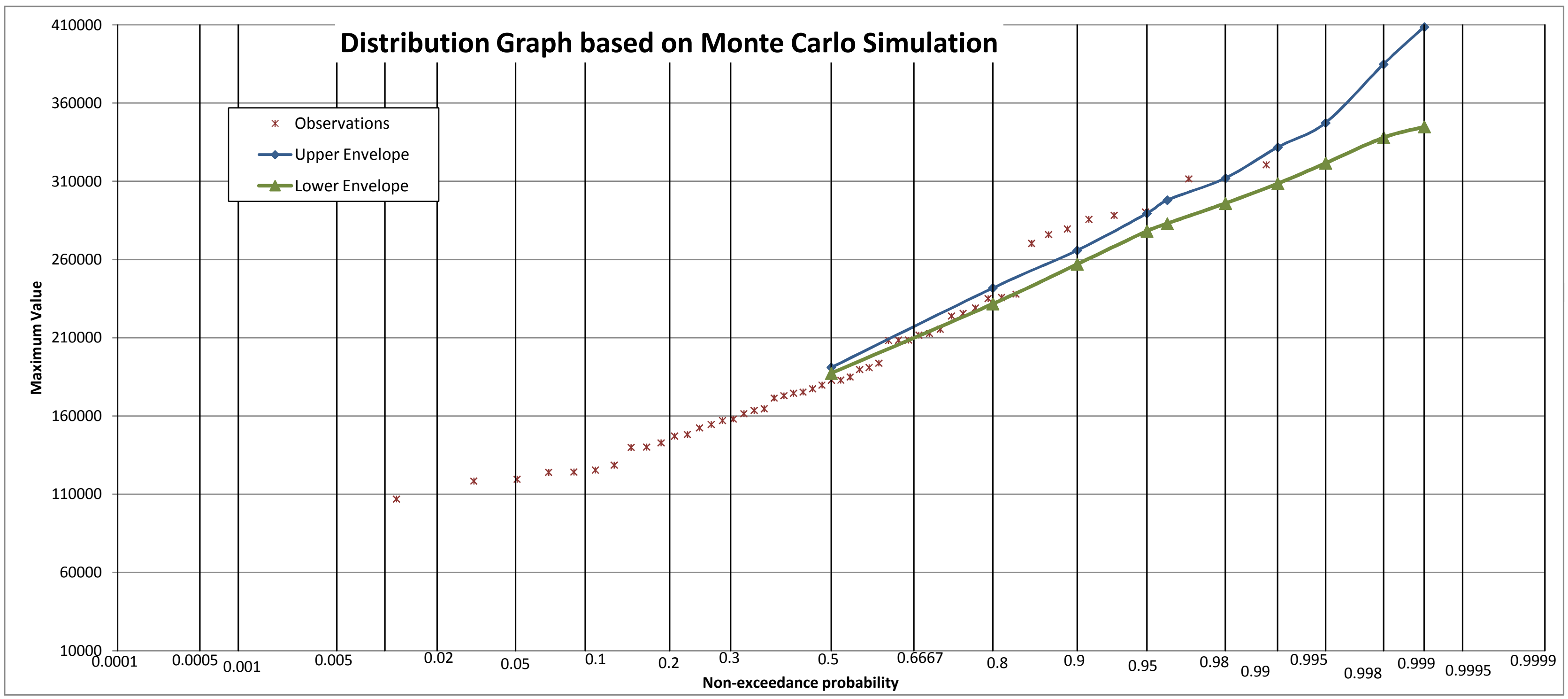
Monte Carlo Simulation

Monte Carlo Simulation Results			
344654.8	Largest	1/1000	0.001
343985.0	2nd Largest	1/500	0.002
321487.5	5th	1/200	0.005
310456.4	10th	1/100	0.01
299945.5	20th	1/50	0.02
281805.5	40th	1/25	0.04
278461.8	50th	1/20	0.05
261968.7	100th	1/10	0.1
237041.6	200th	1/5	0.2
191509.2	500th	1/2	0.5

Results Sample 1	Results Sample 2	Results Sample 3	Results Sample 4	Results Sample 5	Maximum	Minimum	Average
408711.6	373584.5	344796.1	368690.5	375848.4	408711.6	344796.1	374326.2
384846.1	363744.0	337923.4	350775.9	365784.6	384846.1	337923.4	360614.8
345502.7	337844.1	321564.7	347352.7	343011.2	347352.7	321564.7	339055.1
324406.7	315085.9	308495.8	331790.6	328867.6	331790.6	308495.8	321729.3
308390.0	301040.4	295938.6	312050.8	302375.7	312050.8	295938.6	303959.1
289202.0	283001.0	284147.2	297966.0	284128.6	297966.0	283001.0	287689.0
281639.2	278674.1	280645.8	289542.9	278154.4	289542.9	278154.4	281731.3
264546.6	261825.3	256999.8	265877.1	258527.3	265877.1	256999.8	261555.2
237730.5	239759.9	231855.8	241829.3	231627.3	241829.3	231627.3	236560.6
188190.8	190505.6	189764.5	191132.6	187310.6	191132.6	187310.6	189380.8

Take 5 New Samples

1 in 100 year



DFASCC

Data and Frequency Analysis Spreadsheet for the City of Calgary
Version 1.2

PROJECT INFORMATION SHEET

Project Name:	Evaporation Pond - Example #1 (De-correlated data set)
Project Description:	A typical evaporation pond somewhere in Calgary. The area has a size of 100 hectares of which 50% is developed area, with the remainder undeveloped. All runoff is to be fully contained. In the first example, we have only two catchment areas, one for the developed areas (say 50 ha with 60% hard) and landscaped and one for the area around the pond (also 50ha). The pond is assumed at a zero m2 footprint at the bottom, with 0.57% sideslopes. No irrigation of the landscaped areas.
Location:	Say somewhere in northeast or southeast Calgary
Date:	09/11/2012
Designed by:	Bert van Duin
Company Name:	City of Calgary Water Resources - Infrastructure Planning
Reviewed by:	Who volunteers?

Clear Project Information Sheet

Hydrologic Data Series Input

LEGEND

User Input	Negative Result
Calculated Cells	Positive Result

NOTES

- This Spreadsheet is designed for a maximum of 10,000 entries (if more are required then formulas need to be adjusted)
-Input dataset must be based on uniform time distribution (i.e.: daily, weekly, monthly, yearly) and must not include multiple values for any of the time steps (duplicates highlighted in red)
-Input dataset must not have any missing cells of data (cells with "0" will be treated as having a value of 0)

Index	Date	Value	Empirical Probability of Non-Exceedance
1	1961	34643	0.134
2	1962	55575	0.337
3	1963	44415	0.256
4	1964	80691	0.520
5	1965	184448	0.988
6	1966	152366	0.907
7	1967	92755	0.683
8	1969	94383	0.703
9	1970	83733	0.581
10	1971	40505	0.236
11	1972	76978	0.459
12	1973	81767	0.541
13	1974	61654	0.358
14	1975	20152	0.073
15	1976	32222	0.114
16	1977	55210	0.317
17	1978	136584	0.886
18	1979	86261	0.622
19	1980	27221	0.093
20	1981	115264	0.825
21	1982	90125	0.642
22	1983	53272	0.297
23	1984	13914	0.033
24	1985	77813	0.480
25	1986	121075	0.846
26	1987	80399	0.500
27	1988	38159	0.175
28	1989	66197	0.378
29	1990	49113	0.276
30	1991	72361	0.439
31	1992	92537	0.663
32	1993	121587	0.866
33	1994	83729	0.561
34	1995	38791	0.195
35	1996	85929	0.602
36	1997	108998	0.805
37	1998	153575	0.927
38	1999	102458	0.764
39	2000	71847	0.419
40	2001	36644	0.154
41	2002	10252	0.012
42	2003	71607	0.398
43	2004	39512	0.215
44	2005	153583	0.947
45	2006	97196	0.744
46	2007	170172	0.967
47	2008	102803	0.785
48	2009	95627	0.724
49	2010	14764	0.053

Clear All Input Data

Basic Characteristics	
Number of Data Entries	49
Maximum Value	184000
Minimum Value	10300
Average (Mean) Value	79000
Median Value	80400
Standard Deviation	41700
Variance	1740000000
Variation coefficient (Cv)	0.528
Skewness coefficient (Cs)	0.53
Kurtosis	2.73

*Values assumed to be sample not full population

Empirical Probability of Non-Exceedance (Plotting Position) based on:
$F(x(k)) = (k-a)/ (n-2a+1)$, $0 <= a <= 0.5$

a = 0.4 Cunnane (1978)
k= rank of the even in question (in ascending order)
n= 49

Stationarity

Test for Trend:

Choose Significance Level (alpha):	1%

1) Spearman Rank Order Correlation Coefficient

H_0= Data has no trend

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}.$$

Spearman Correlation Coefficient:	0.103

based on z	No Significant Trend at 0.01 Significance Level
based on t	No Significant Trend at 0.01 Significance Level

When there are no ties in rankings:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

Spearman Correlation Coefficient:	0.103
t-distribution value	0.712
Degrees of freedom	47

T (Adjustment for ties) =	0
Standard Normal (z)=	0.705

Tests for Jump:

2) Mann-Whitney Test for jump (a.k.a. Mann-Whitney U test)

Index number of subsample divide	25

H_0= Independent samples drawn from the same population (No Jump)

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}$$

Number of values in sample 1 n_1=	25
Number of values in sample 2 n_2=	24
Total of Ranking in sample 1 R_1=	597
Total of Ranking in sample 2 R_2=	628
U_1=	272

No Jump at 0.01 Significance Level

$$U_1 + U_2 = n_1 n_2.$$

U_2=	328
U (Minimum of U_1 and U_2)=	272
Standard Normal (z)=	-0.560

3) Wald-Wolfowitz Test (The runs test)

$$\mu - \frac{2\,N_+\,N_-}{N} + 1\,, \qquad \sigma^2 = \frac{2\,N_+\,N_-\,(2\,N_+\,N_- - N)}{N^2\,(N-1)} = \frac{(\mu-1)(\mu-2)}{N-1}\,.$$

Number of data greater than median N_+ =	24
Number of data less than median N_- =	24
Total number of runs =	17

H_0= Data represent sample of single independently distributed random variable (No Jump)

Mean =	25.0
Variance =	11.7
Standard Normal (z)=	-2.5

No Jump at 0.01 Significance Level

Homogeneity

Choose Significance Level (alpha):	5%

Mann-Whitney Test for homogeneity (a.k.a. Mann-Whitney U test)

Index number of subsample divide	25

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}$$

H_0= There is homogeneity between samples with respect to probability of random drawing of a larger observation

Sample is Homogeneous at 0.05 Significance Level

Number of values in sample 1 n_1=	25
Number of values in sample 2 n_2=	24
Total of Ranking in sample 1 R_1=	597
Total of Ranking in sample 1 R_2=	628
U_1=	272

$$U_1 + U_2 = n_1 n_2.$$

U_2=	328
U (Minimum of U_1 and U_2)=	272
Standard Normal (z)=	-0.560

Terry Test

Index number of subsample divide	25

H_0= There is homogeneity between samples with respect to probability of random drawing of a larger observation

Total sample size	49
Subsample 1 (m)	25
Subsample 2 (n)	24

Sample is Homogeneous at 0.05 Significance Level

Standard Deviation =	3.444
Sum of ranks in first subsample c =	1.480
z =	0.430

Independence

Choose Significance Level (alpha): 1%

1) Spearman Rank Order Correlation Coefficient

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}.$$

H_0= Data is independent

Spearman Correlation Coefficient:	0.37

Non-independence Detected at 0.01 Significance Level

When there are no ties in rankings:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}.$$

Spearman Correlation Coefficient:	0.37
t-distribution value	2.93
Degrees of freedom	47

2) Wald-Wolfowitz Test

$$R = \sum_{i=1}^{N-1} x_i x_{i+1} + x_1 x_N$$

Statistic R	3.34E+11
Mean	3.04E+11
Variance	1.37E+20

H_0= Data is independent

Data is independent at 0.01 Significance Level

Standard Normal (z)=	2.6

2) Anderson Test

$$r_1 = \left[\sum_{i=1}^{N-1} x_i x_{i+1} + x_1 x_N - \left(\sum_{i=1}^{N} x_i \right)^2 \Big/ N \right] \Big/ \left[\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2 \Big/ N \right]$$

Statistic r	0.336
Mean	-0.021
Variance	0.020

H_0= Data is independent

Data is independent at 0.01 Significance Level

Standard Normal (z)=	2.5

Outliers

Significance Level (alpha): 10%

Grubbs and Beck test for Outliers

1) High Outliers

Assumption: logarithms of sample are normally distributed

X_h = exp (xmean+K_nS)
K(n) =-3.62201+6.2844N^1/4-2.49835N^1/2+0.491436N^3/4-0.037911N
K(n) = -0.9043+3.345*SQRT(log(n))-0.4046log(n) for 5<n<150

Sample Size (n) =	49
K(n) =	2.76
K(n) for 5<n<150 =	2.76
X_h=	404000
Maximum Value	184000
High Outliers	No High Outliers Present

< Any value higher than X_h is considered a high outlier

2) Low Outliers

X_h = exp (xmean-K_nS)
K(n) =-3.62201+6.2844N^1/4-2.49835N^1/2+0.491436N^3/4-0.037911N
K(n) = -0.9043+3.345*SQRT(log(n))-0.4046log(n) for 5<n<150

Sample Size (n) =	49
K(n) =	2.76
K(n) for 5<n<150 =	2.76
X_h=	10900
Minimum Value	10300
Low Outliers	Low Outlier May Be Present

< Any value lower than X_h is considered a low outlier

Dependent Dataset

Choose Significance Level (alpha): 1%

Autocorrelation coefficient

$$R_c(\tau) = \frac{\sum_{i=1}^{N-|\tau|} X_i Y_{i+\tau} - \frac{1}{N-|\tau|}\left(\sum_{i=1}^{N-|\tau|} X_i\right)\left(\sum_{i=\tau+1}^{N} Y_i\right)}{\left[\sum_{i=1}^{N-|\tau|} X_i^2 - \frac{1}{N-|\tau|}\left(\sum_{i=1}^{N-|\tau|} X_i\right)^2\right]^{0.5}\left[\sum_{i=1+|\tau|}^{N} Y_i^2 - \frac{1}{N-|\tau|}\left(\sum_{i-1+|\tau|}^{N} Y_i\right)^2\right]^{0.5}}$$

H_0 - The data is not serially correlated

One Time Period Offset		
Autocorrelation coefficient offset by one time period	r(1) =	0.313
t-distribution values for one time period offset	t =	2.260

No Serial Correlation at 0.01 Significance Level

Two Time Periods Offset		
Autocorrelation coefficient offset by two time periods	r(2) =	-0.042
t-distribution values for two time periods offset	t =	-0.290

No Serial Correlation at 0.01 Significance Level

Instructions:
Compare the results of the autocorrelation tests for one time period offset and for the two time period offset. One of the following 2 scenarios will result:
1. The finding for the one period time step is serially correlated, and the finding for the two time step is also serially correlated. In this case, transposing the data series is unlikely to produce an independent data set suitable for frequency analysis. In this case, other methods, such as the Monte Carlo simulation are necessary.
2. The finding for the one period time step is serially correlated, and the finding for the two time step is NOT serially correlated. In this case, the data series should be transposed to produce an independent data set suitable for frequency analysis.

Frequency Analysis Results Input

LEGEND

User Input	Negative Result
Calculated Cells	Positive Result

Clear All Input Data

NOTES

- This spreadsheet designed to accept the results of 10 specific Frequency Analysis outputs
- The input data must be in the same format as the output table from Hyfran (either copied and pasted special as text in the top left cell of each yellow input box, or manually input as distribution results and hyfran calculated paremeters in specified areas.
- Input dataset must be complete (only one method of estimation per distribution type, refer to Section 3.3.1 and 3.3.2 of the Frequency Analysis Procedures for Stormwater Design Manual when choosing methods of estimation)
- An additional 11th Frequency Analysis ouput can be copied into the last input box. This output will be displayed in the visual goodness of fit tab, however no numerical goodness of fit tests will be performed on it.

Normal (Gaussian) type of distributions:

Normal Distribution:

Paste Normal Distribution Hyfran Output in Cell Below (A15)

Results of the fitting

Normal (Maximum Likelihood)

Number of observations 49

Parameters

mu	78997.2653
sigma	41749.0137

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)	
10000	0.9999	2.34E+05	1.69E+04	2.01E+05	2.67E+05
2000	0.9995	2.16E+05	1.52E+04	1.87E+05	2.46E+05
1000	0.999	2.08E+05	1.45E+04	1.80E+05	2.36E+05
200	0.995	1.87E+05	1.25E+04	1.62E+05	2.11E+05
100	0.99	1.76E+05	1.16E+04	1.53E+05	1.99E+05
50	0.98	1.65E+05	1.06E+04	1.44E+05	1.86E+05
20	0.95	1.48E+05	9200	1.30E+05	1.66E+05
10	0.9	1.33E+05	8090	1.17E+05	1.48E+05
5	0.8	1.14E+05	6960	1.00E+05	1.28E+05
3	0.6667	9.70E+04	6240	8.47E+04	1.09E+05
2	0.5	7.90E+04	5960	6.73E+04	9.07E+04
1.4286	0.3	5.71E+04	6370	4.46E+04	6.96E+04
1.25	0.2	4.39E+04	6960	3.02E+04	5.75E+04
1.1111	0.1	2.55E+04	8090	9.63E+03	4.13E+04
1.0526	0.05	1.03E+04	9200	-7.73E+03	2.84E+04
1.0204	0.02	-6.76E+03	1.06E+04	-2.75E+04	1.40E+04
1.0101	0.01	-1.81E+04	1.16E+04	-4.08E+04	4.54E+03
1.005	0.005	-2.86E+04	1.25E+04	-5.30E+04	-4.07E+03
1.001	0.001	-5.00E+04	1.45E+04	-7.84E+04	-2.17E+04
1.0005	0.0005	-5.84E+04	1.52E+04	-8.83E+04	-2.85E+04
1.0001	0.0001	-7.63E+04	1.69E+04	-1.09E+05	-4.31E+04

Lognormal Distribution:

Paste Lognormal Distribution Output from Hyfran in Cell Below (A57)

Results of the fitting

Lognormal (Maximum Likelihood)

Number of observations 49

Parameters

mu	11.104548
sigma	0.654105

Quantiles
q = F(X) : non-exceedance probability

T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)	
10000	0.9999	7.57E+05	2.01E+05	3.63E+05	1.15E+06
2000	0.9995	5.72E+05	1.37E+05	3.04E+05	8.40E+05
1000	0.999	5.02E+05	1.14E+05	2.79E+05	7.25E+05
200	0.995	3.58E+05	7.02E+04	2.21E+05	4.96E+05
100	0.99	3.05E+05	5.52E+04	1.96E+05	4.13E+05
50	0.98	2.55E+05	4.23E+04	1.72E+05	3.38E+05
20	0.95	1.95E+05	2.81E+04	1.40E+05	2.50E+05
10	0.9	1.54E+05	1.95E+04	1.16E+05	1.92E+05
5	0.8	1.15E+05	1.26E+04	9.06E+04	1.40E+05
3	0.6667	8.81E+04	8610	7.12E+04	1.05E+05
2	0.5	6.65E+04	6210	5.43E+04	7.86E+04
1.4286	0.3	4.72E+04	4710	3.80E+04	5.64E+04
1.25	0.2	3.83E+04	4180	3.01E+04	4.65E+04
1.1111	0.1	2.87E+04	3640	2.16E+04	3.59E+04
1.0526	0.05	2.27E+04	3270	1.63E+04	2.91E+04
1.0204	0.02	1.73E+04	2880	1.17E+04	2.30E+04
1.0101	0.01	1.45E+04	2630	9.35E+03	1.97E+04
1.005	0.005	1.23E+04	2410	7.60E+03	1.71E+04
1.001	0.001	8.80E+03	1990	4.90E+03	1.27E+04
1.0005	0.0005	7.72E+03	1840	4.11E+03	1.13E+04
1.0001	0.0001	5.84E+03	1550	2.80E+03	8.87E+03

Lognormal III Distribution

Paste Lognormal III Distribution Output from Hyfran in Cell Below (A99)

Results of the fitting

3-parameter lognormal (Maximum Likelihood)

Number of observations 49

Parameters

m	-85591.2535
mu	11.979911
sigma	0.251011

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)	
10000	0.9999	3.20E+05	7.14E+04	1.80E+05	4.60E+05
2000	0.9995	2.79E+05	5.33E+04	1.74E+05	3.83E+05
1000	0.999	2.61E+05	4.62E+04	1.70E+05	3.51E+05
200	0.995	2.19E+05	3.13E+04	1.58E+05	2.80E+05
100	0.99	2.00E+05	2.56E+04	1.50E+05	2.51E+05
50	0.98	1.82E+05	2.05E+04	1.41E+05	2.22E+05
20	0.95	1.55E+05	1.45E+04	1.27E+05	1.84E+05
10	0.9	1.34E+05	1.09E+04	1.13E+05	1.56E+05
5	0.8	1.11E+05	8230	9.53E+04	1.28E+05
3	0.6667	9.21E+04	6900	7.86E+04	1.06E+05
2	0.5	7.39E+04	6150	6.19E+04	8.60E+04
1.4286	0.3	5.43E+04	5610	4.33E+04	6.53E+04
1.25	0.2	4.36E+04	5460	3.29E+04	5.42E+04
1.1111	0.1	3.00E+04	5650	1.90E+04	4.11E+04
1.0526	0.05	2.00E+04	6320	7.58E+03	3.23E+04
1.0204	0.02	9.66E+03	7620	-5.29E+03	2.46E+04
1.0101	0.01	3.36E+03	8770	-1.38E+04	2.06E+04
1.005	0.005	-2.04E+03	9970	-2.16E+04	1.75E+04
1.001	0.001	-1.22E+04	1.28E+04	-3.72E+04	1.28E+04
1.0005	0.0005	-1.58E+04	1.39E+04	-4.31E+04	1.15E+04
1.0001	0.0001	-2.29E+04	1.65E+04	-5.53E+04	9.51E+03

Exponential and Pearson type of distributions:

Exponential Distribution

Paste Exponential Distribution Output from Hyfran in Cell Below (A142)

Results of the fitting

Exponential (Maximum Likelihood)

Number of observations 49

Parameters

alpha	70177.4583
m	8819.80697

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)	
10000	0.9999	6.55E+05	9.31E+04	4.73E+05	8.38E+05
2000	0.9995	5.42E+05	7.68E+04	3.92E+05	6.93E+05
1000	0.999	4.94E+05	6.98E+04	3.57E+05	6.30E+05
200	0.995	3.81E+05	5.35E+04	2.76E+05	4.85E+05
100	0.99	3.32E+05	4.65E+04	2.41E+05	4.23E+05
50	0.98	2.83E+05	3.94E+04	2.06E+05	3.61E+05
20	0.95	2.19E+05	3.02E+04	1.60E+05	2.78E+05
10	0.9	1.70E+05	2.32E+04	1.25E+05	2.16E+05
5	0.8	1.22E+05	1.62E+04	9.01E+04	1.53E+05
3	0.6667	8.59E+04	1.10E+04	6.43E+04	1.08E+05
2	0.5	5.75E+04	6960	4.38E+04	7.11E+04
1.4286	0.3	3.39E+04	3690	2.66E+04	4.11E+04
1.25	0.2	2.45E+04	2500	1.96E+04	2.94E+04
1.1111	0.1	1.62E+04	1670	1.29E+04	1.95E+04
1.0526	0.05	1.24E+04	1470	9.55E+03	1.53E+04
1.0204	0.02	1.02E+04	1430	7.43E+03	1.30E+04
1.0101	0.01	9.53E+03	1440	6.71E+03	1.23E+04
1.005	0.005	9.17E+03	1440	6.35E+03	1.20E+04
1.001	0.001	8.89E+03	1450	6.06E+03	1.17E+04
1.0005	0.0005	8.85E+03	1450	6.02E+03	1.17E+04
1.0001	0.0001	8.83E+03	1450	5.99E+03	1.17E+04

Pearson Type III Distribution

Paste Pearson III Distribution Output from Hyfran in Cell Below (A184)

Results of the fitting

Pearson type III (Method of moments)

Number of observations 49

Parameters

alpha	0.00009
lambda	14.242331
m	-78559.3918

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)		
10000	0.9999	2.83E+05	5.36E+04	1.78E+05	3.88E+05	
2000	0.9995	2.53E+05	4.23E+04	1.70E+05	3.36E+05	
1000	0.999	2.40E+05	3.76E+04	1.66E+05	3.14E+05	
200	0.995	2.07E+05	2.69E+04	1.54E+05	2.60E+05	
100	0.99	1.92E+05	2.25E+04	1.48E+05	2.36E+05	
50	0.98	1.76E+05	1.84E+04	1.40E+05	2.12E+05	
20	0.95	1.53E+05	1.34E+04	1.27E+05	1.80E+05	
10	0.9	1.34E+05	1.02E+04	1.14E+05	1.54E+05	
5	0.8	1.13E+05	7930	9.71E+04	1.28E+05	
3	0.6667	9.42E+04	6990	8.05E+04	1.08E+05	
2	0.5	7.54E+04	6590	6.24E+04	8.83E+04	
1.4286	0.3	5.48E+04	6160	4.27E+04	6.68E+04	
1.25	0.2	4.32E+04	5980	3.15E+04	5.49E+04	
1.1111	0.1	2.83E+04	6360	1.59E+04	4.08E+04	
1.0526	0.05	1.70E+04	7730	1890	3.22E+04	
1.0204	0.02	5.44E+03	1.05E+04	-1.51E+04	2.60E+04	
1.0101	0.01	-1.65E+03	1.29E+04	-2.70E+04	2.36E+04	
1.005	0.005	-7.71E+03	1.54E+04	-3.80E+04	2.25E+04	
1.001	0.001	-1.90E+04	2.13E+04	N/D	N/D	
1.0005	0.0005	-2.29E+04	2.38E+04	N/D	N/D	
1.0001	0.0001	-3.08E+04	2.94E+04	N/D	N/D	

Log-Pearson Type III Distribution

Paste Log Pearson III Distribution Output from Hyfran in Cell Below (A226)

Results of the fitting

Log-Pearson type III (Méthode SAM)

Number of observations 49

Parameters

alpha	-8.180007
lambda	5.158023
m	5.453209

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)		
10000	0.9999	2.48E+05	1.34E+05	N/D	N/D	
2000	0.9995	2.34E+05	1.06E+05	N/D	N/D	
1000	0.999	2.27E+05	9.30E+04	N/D	N/D	
200	0.995	2.05E+05	6.21E+04	N/D	N/D	
100	0.99	1.94E+05	4.86E+04	N/D	N/D	
50	0.98	1.80E+05	3.54E+04	N/D	N/D	
20	0.95	1.59E+05	1.98E+04	1.20E+05	1.97E+05	
10	0.9	1.39E+05	1.17E+04	1.16E+05	1.62E+05	
5	0.8	1.15E+05	9.53E+03	9.61E+04	1.34E+05	
3	0.6667	9.34E+04	9450	7.49E+04	1.12E+05	
2	0.5	7.29E+04	8500	5.63E+04	8.96E+04	
1.4286	0.3	5.15E+04	6200	3.93E+04	6.36E+04	
1.25	0.2	4.06E+04	5290	3.02E+04	5.10E+04	
1.1111	0.1	2.82E+04	5150	1.82E+04	3.83E+04	
1.0526	0.05	2.03E+04	5420	9.68E+03	3.09E+04	
1.0204	0.02	1.35E+04	5470	2810	2.43E+04	
1.0101	0.01	1.01E+04	5240	-161	2.04E+04	
1.005	0.005	7.64E+03	4870	-1920	1.72E+04	
1.001	0.001	4.09E+03	3810	N/D	N/D	
1.0005	0.0005	3.16E+03	3350	N/D	N/D	
1.0001	0.0001	1.76E+03	2410	-2970	6480	

Extreme Value type of distributions:

EVI (Gumbel) Distribution

Paste EV Distribution Output from Hyfran in Cell Below (A269)

Results of the fitting

Gumbel (Maximum Likelihood)

Number of observations 49

Parameters

u	59102.8738
alpha	35577.0447

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)	
10000	0.9999	3.87E+05	3.91E+04	3.10E+05	4.63E+05
2000	0.9995	3.30E+05	3.26E+04	2.66E+05	3.93E+05
1000	0.999	3.05E+05	2.99E+04	2.46E+05	3.63E+05
200	0.995	2.48E+05	2.35E+04	2.01E+05	2.94E+05
100	0.99	2.23E+05	2.08E+04	1.82E+05	2.64E+05
50	0.98	1.98E+05	1.81E+04	1.62E+05	2.33E+05
20	0.95	1.65E+05	1.45E+04	1.36E+05	1.93E+05
10	0.9	1.39E+05	1.18E+04	1.16E+05	1.62E+05
5	0.8	1.12E+05	9190	9.44E+04	1.30E+05
3	0.6667	9.12E+04	7310	7.69E+04	1.06E+05
2	0.5	7.21E+04	5950	6.05E+04	8.38E+04
1.4286	0.3	5.25E+04	5160	4.24E+04	6.26E+04
1.25	0.2	4.22E+04	5090	3.22E+04	5.22E+04
1.1111	0.1	2.94E+04	5370	1.89E+04	4.00E+04
1.0526	0.05	2.01E+04	5800	8.69E+03	3.14E+04
1.0204	0.02	1.06E+04	6390	-1.96E+03	2.31E+04
1.0101	0.01	4.77E+03	6810	-8.58E+03	1.81E+04
1.005	0.005	-2.18E+02	7200	-1.43E+04	1.39E+04
1.001	0.001	-9.65E+03	7990	-2.53E+04	6.01E+03
1.0005	0.0005	-1.31E+04	8290	-2.93E+04	3.20E+03
1.0001	0.0001	-1.99E+04	8920	-3.74E+04	-2.41E+03

GEV (General Extreme Value) Distribution

Paste GEV Distribution Output from Hyfran in Cell Below (A311)

Results of the fitting

GEV (Maximum Likelihood)

Number of observations 49

Parameters

alpha	36186.9044
k	0.109691
u	61461.3155

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)		
10000	0.9999	2.71E+05	6.54E+04	N/D	N/D	
2000	0.9995	2.48E+05	4.93E+04	N/D	N/D	
1000	0.999	2.37E+05	4.26E+04	1.53E+05	3.20E+05	
200	0.995	2.07E+05	2.82E+04	1.51E+05	2.62E+05	
100	0.99	1.92E+05	2.28E+04	1.48E+05	2.37E+05	
50	0.98	1.76E+05	1.79E+04	1.41E+05	2.11E+05	
20	0.95	1.53E+05	1.27E+04	1.28E+05	1.78E+05	
10	0.9	1.34E+05	9830	1.14E+05	1.53E+05	
5	0.8	1.12E+05	7900	9.60E+04	1.27E+05	
3	0.6667	9.26E+04	6880	7.91E+04	1.06E+05	
2	0.5	7.45E+04	6140	6.24E+04	8.65E+04	
1.4286	0.3	5.47E+04	5570	4.37E+04	6.56E+04	
1.25	0.2	4.38E+04	5520	3.30E+04	5.46E+04	
1.1111	0.1	2.99E+04	5960	1.82E+04	4.15E+04	
1.0526	0.05	1.93E+04	6770	6.01E+03	3.25E+04	
1.0204	0.02	8.22E+03	8050	-7.57E+03	2.40E+04	
1.0101	0.01	1.30E+03	9070	-1.65E+04	1.91E+04	
1.005	0.005	-4.75E+03	1.01E+04	-2.45E+04	1.50E+04	
1.001	0.001	-1.64E+04	1.23E+04	-4.06E+04	7.73E+03	
1.0005	0.0005	-2.07E+04	1.32E+04	-4.67E+04	5.22E+03	
1.0001	0.0001	-2.95E+04	1.52E+04	-5.94E+04	3.75E+02	

EVIII (Weibull) Distribution

Paste Weibull Distribution Output from Hyfran in Cell Below (A353)

Results of the fitting

Weibull (Maximum Likelihood)

Number of observations 49

Parameters

alpha	89119.1126
c	1.996513

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

T	q	XT	Standard deviation	Confidence interval (95%)		
10000	0.9999	2.71E+05	3.34E+04	2.06E+05	3.36E+05	
2000	0.9995	2.46E+05	2.82E+04	1.91E+05	3.01E+05	
1000	0.999	2.35E+05	2.59E+04	1.84E+05	2.85E+05	
200	0.995	2.05E+05	2.05E+04	1.65E+05	2.46E+05	
100	0.99	1.92E+05	1.81E+04	1.56E+05	2.27E+05	
50	0.98	1.76E+05	1.57E+04	1.46E+05	2.07E+05	
20	0.95	1.54E+05	1.25E+04	1.30E+05	1.79E+05	
10	0.9	1.35E+05	1.02E+04	1.15E+05	1.55E+05	
5	0.8	1.13E+05	8100	9.72E+04	1.29E+05	
3	0.6667	9.34E+04	6900	7.99E+04	1.07E+05	
2	0.5	7.42E+04	6230	6.20E+04	8.64E+04	
1.4286	0.3	5.32E+04	5750	4.19E+04	6.45E+04	
1.25	0.2	4.20E+04	5420	3.14E+04	5.27E+04	
1.1111	0.1	2.89E+04	4780	1.95E+04	3.82E+04	
1.0526	0.05	2.01E+04	4070	1.21E+04	2.81E+04	
1.0204	0.02	1.26E+04	3180	6.40E+03	1.89E+04	
1.0101	0.01	8.90E+03	2570	3.85E+03	1.39E+04	
1.005	0.005	6.28E+03	2050	2.25E+03	1.03E+04	
1.001	0.001	2.80E+03	1160	5.22E+02	5.08E+03	
1.0005	0.0005	1.98E+03	897	2.20E+02	3.74E+03	
1.0001	0.0001	8.84E+02	479	-5.56E+01	1.82E+03	

Gamma type of distributions:

Gamma Distribution

Paste Gamma Distribution Output from Hyfran in Cell Below (A396)

Results of the fitting

Gamma (Maximum Likelihood)

Number of observations 49

Parameters

alpha	0.000039
lambda	3.053841

Quantiles
q = F(X) : non-exceedance probability
T = 1/(1-q)

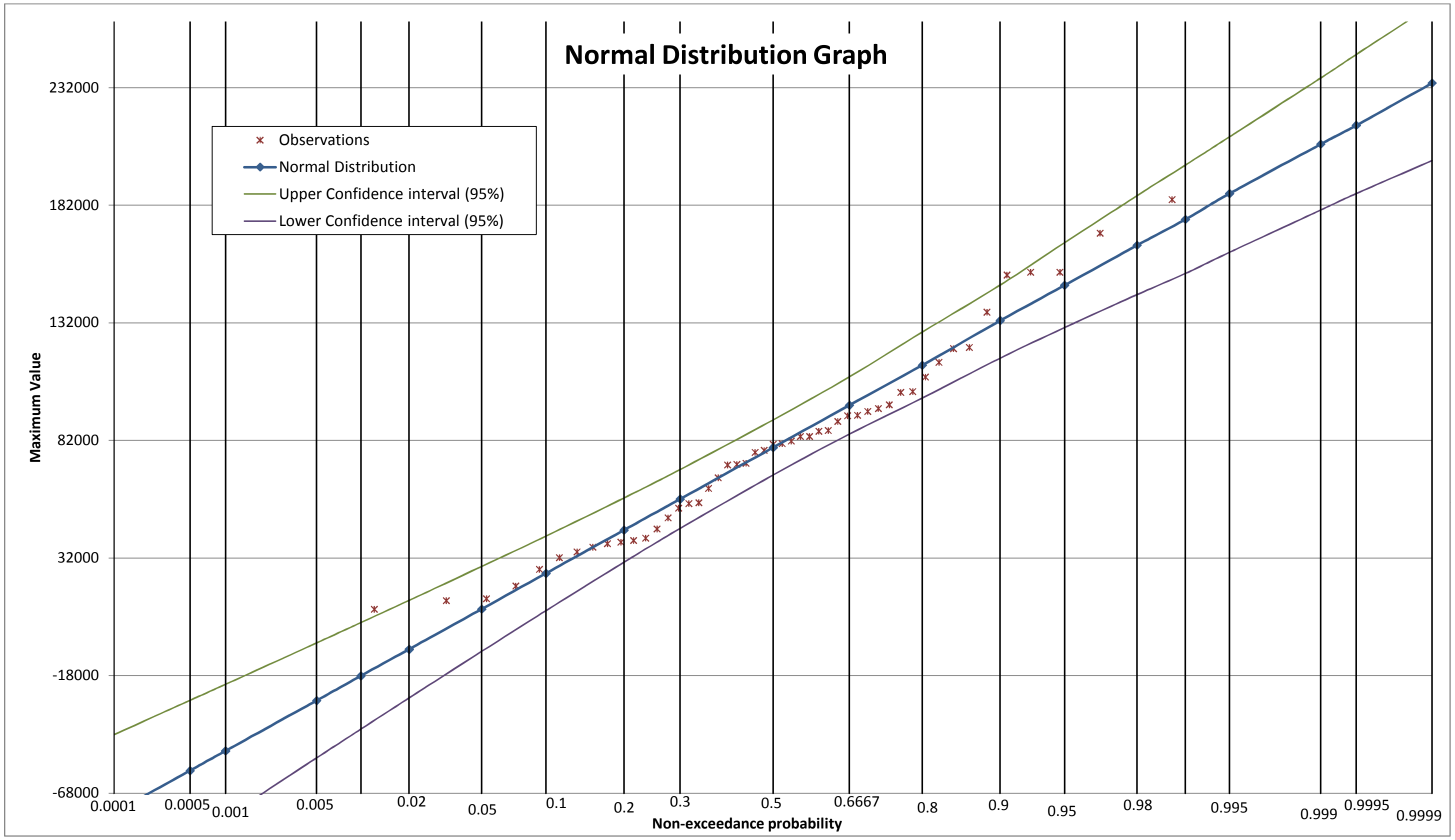
T	q	XT	Standard deviation	Confidence interval (95%)		
10000	0.9999	3.63E+05	4.87E+04	2.68E+05	4.59E+05	
2000	0.9995	3.14E+05	4.02E+04	2.36E+05	3.93E+05	
1000	0.999	2.93E+05	3.65E+04	2.21E+05	3.65E+05	
200	0.995	2.42E+05	2.81E+04	1.87E+05	2.97E+05	
100	0.99	2.20E+05	2.45E+04	1.72E+05	2.68E+05	
50	0.98	1.97E+05	2.09E+04	1.56E+05	2.38E+05	
20	0.95	1.65E+05	1.62E+04	1.33E+05	1.97E+05	
10	0.9	1.40E+05	1.28E+04	1.15E+05	1.65E+05	
5	0.8	1.12E+05	9540	9.37E+04	1.31E+05	
3	0.6667	9.10E+04	7450	7.64E+04	1.06E+05	
2	0.5	7.05E+04	5980	5.88E+04	8.23E+04	
1.4286	0.3	5.07E+04	5110	4.06E+04	6.07E+04	
1.25	0.2	4.07E+04	4790	3.13E+04	5.01E+04	
1.1111	0.1	2.94E+04	4400	2.08E+04	3.80E+04	
1.0526	0.05	2.20E+04	4030	1.41E+04	2.99E+04	
1.0204	0.02	1.53E+04	3560	8.37E+03	2.23E+04	
1.0101	0.01	1.18E+04	3220	5.50E+03	1.81E+04	
1.005	0.005	9.11E+03	2910	3.41E+03	1.48E+04	
1.001	0.001	4.86E+03	2270	4.09E+02	9.32E+03	
1.0005	0.0005	3.61E+03	2030	-3.68E+02	7.60E+03	
1.0001	0.0001	1.56E+03	1540	-1.46E+03	4.58E+03	

User Defined Distribution

Paste User Defined Distribution Output from Hyfran in Cell Below (A439) or Input Calculated Values in Designated Cells

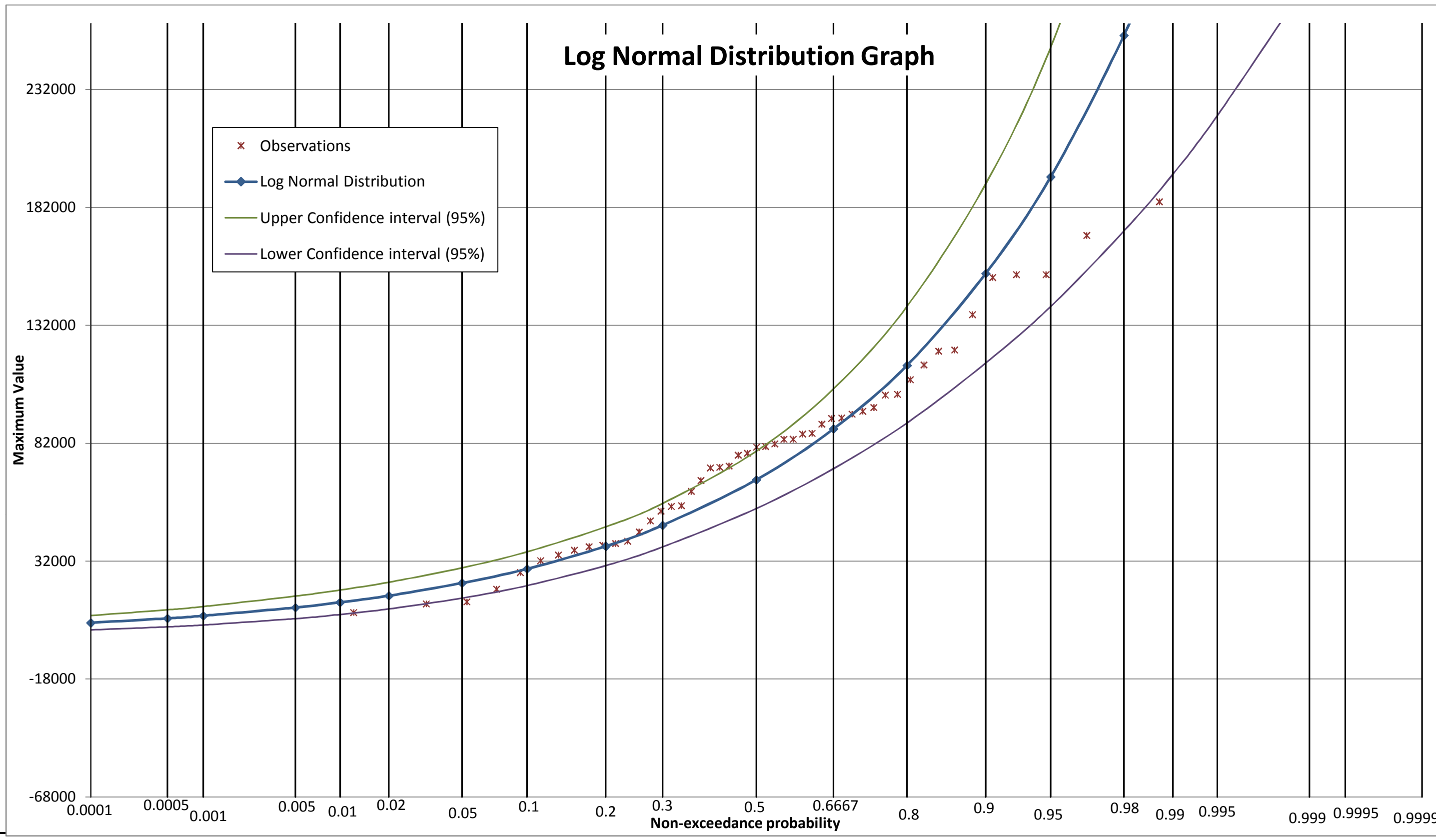
				Lower	**Upper**	
T	q	XT	Standard deviation	Confidence interval (95%)		

Visual Goodness of Fit



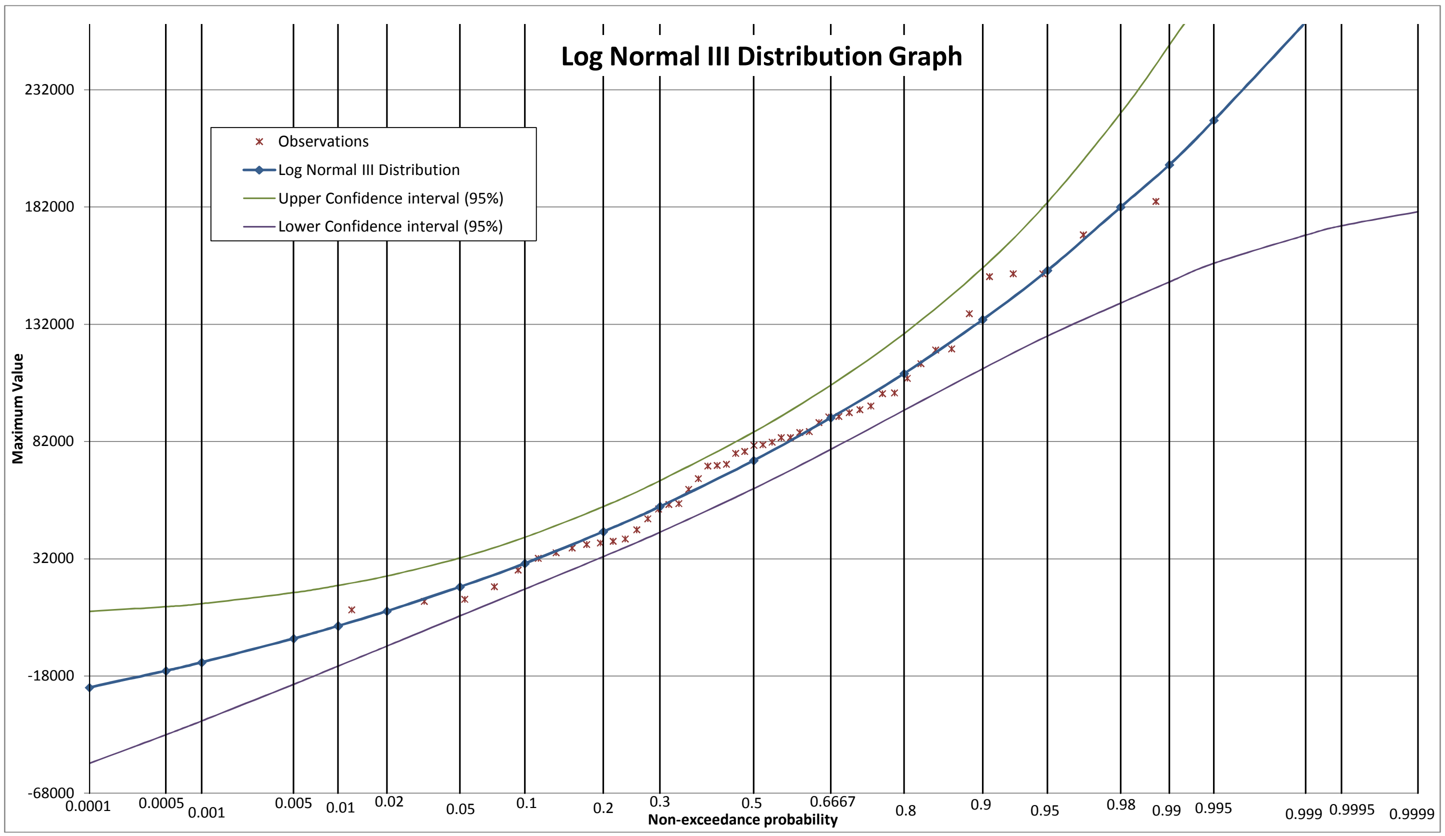
Normal Distribution
Visual Goodness-of-Fit Test Notes:

Several points above the upper confidence interval at both low and high end of non-exceedance probability and below the lower confidence interal near 0.5 non-exceedance probability



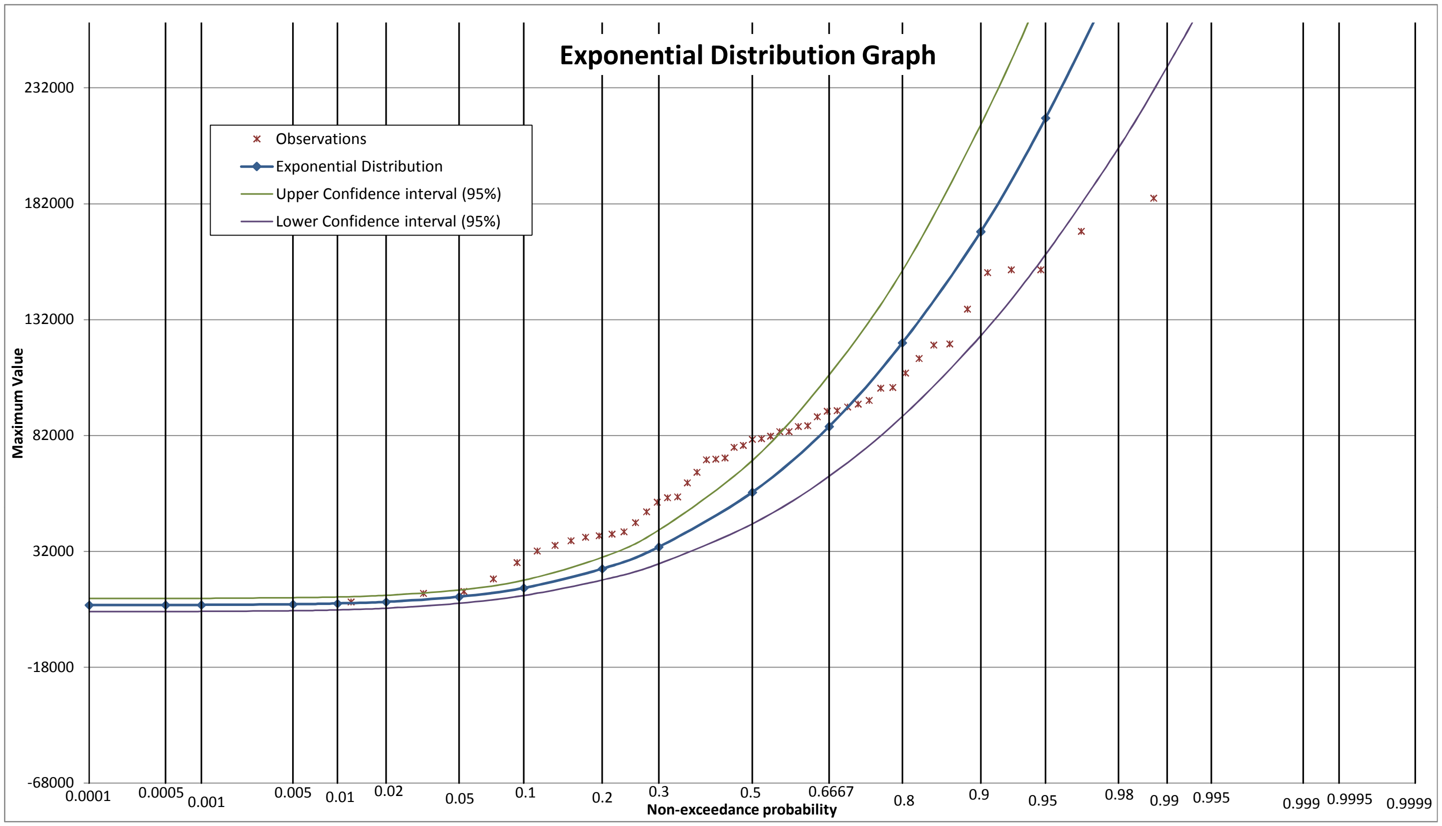
Log Normal Distribution
Visual Goodness-of-Fit Test Notes:

A few points above the upper confidence interval at both low and high end of non-exceedance probability and some below the lower confidence interal near 0.5 non-exceedance probability



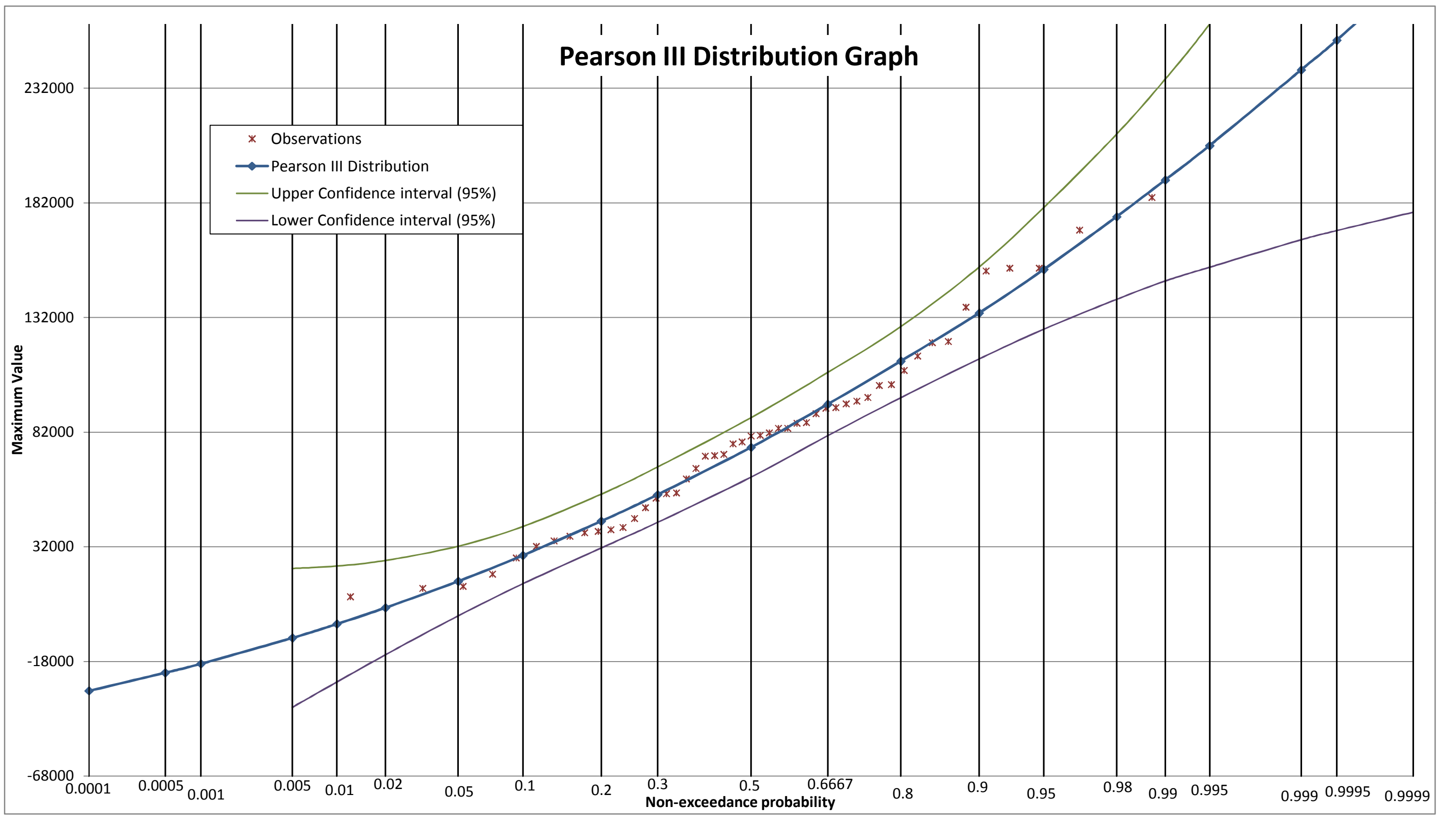
Log Normal III Distribution
Visual Goodness-of-Fit Test Notes:

All points within confidence interval, very good fit for majority of points with the exception of the largest event.
Good fit, may be over estimating values at high non-exceedance probabilities.
As the values are for pond volumes, it might be possible that the pond overflowed. The data were checked and the pond design information indicates that these levels are below the spillway crest.



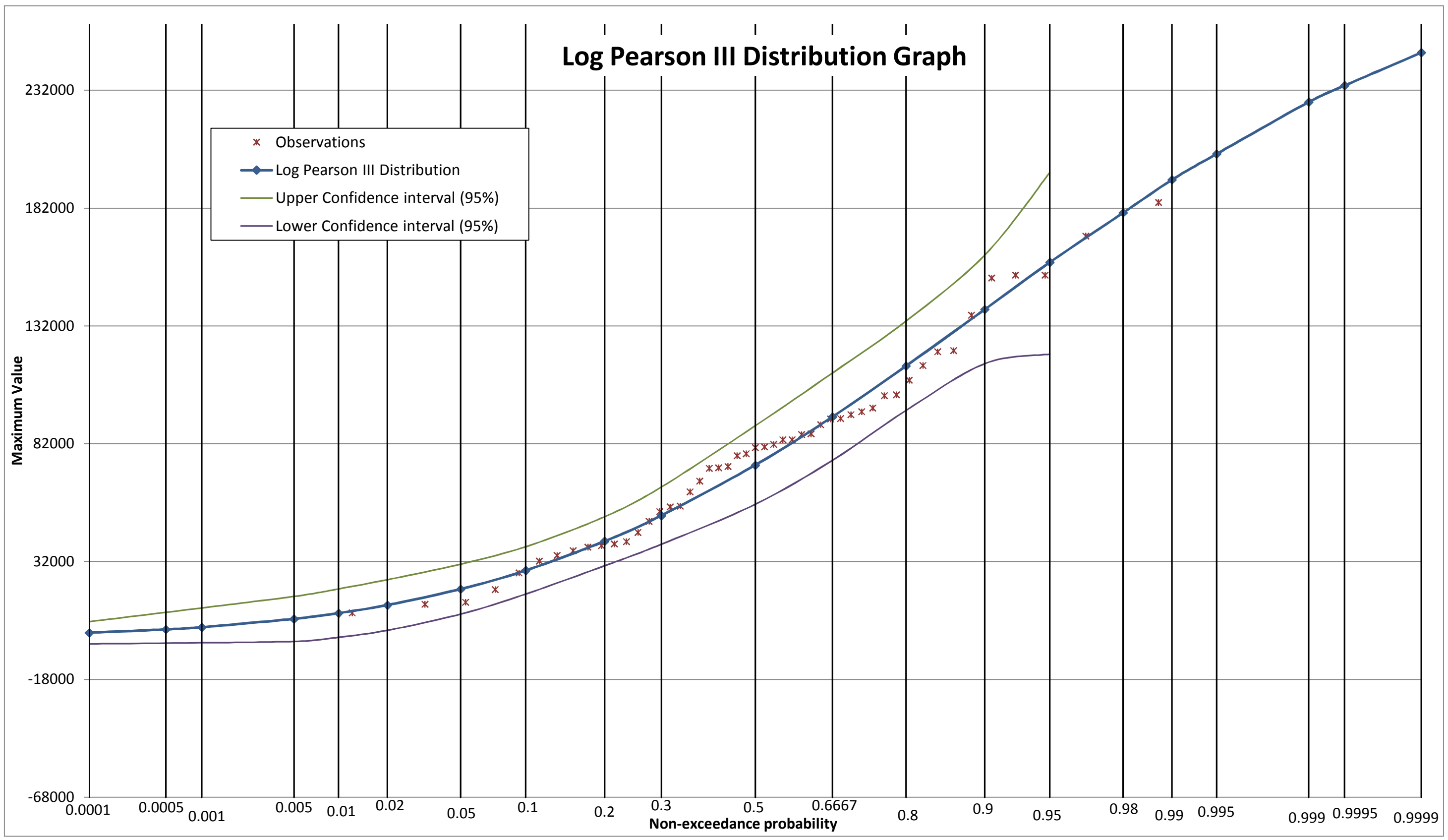
Exponential Distribution
Visual Goodness-of-Fit Test Notes:

Some points outside of confience interval for lower non-exceedance probabilities. Good fit at higher non-exceedance probabilities.

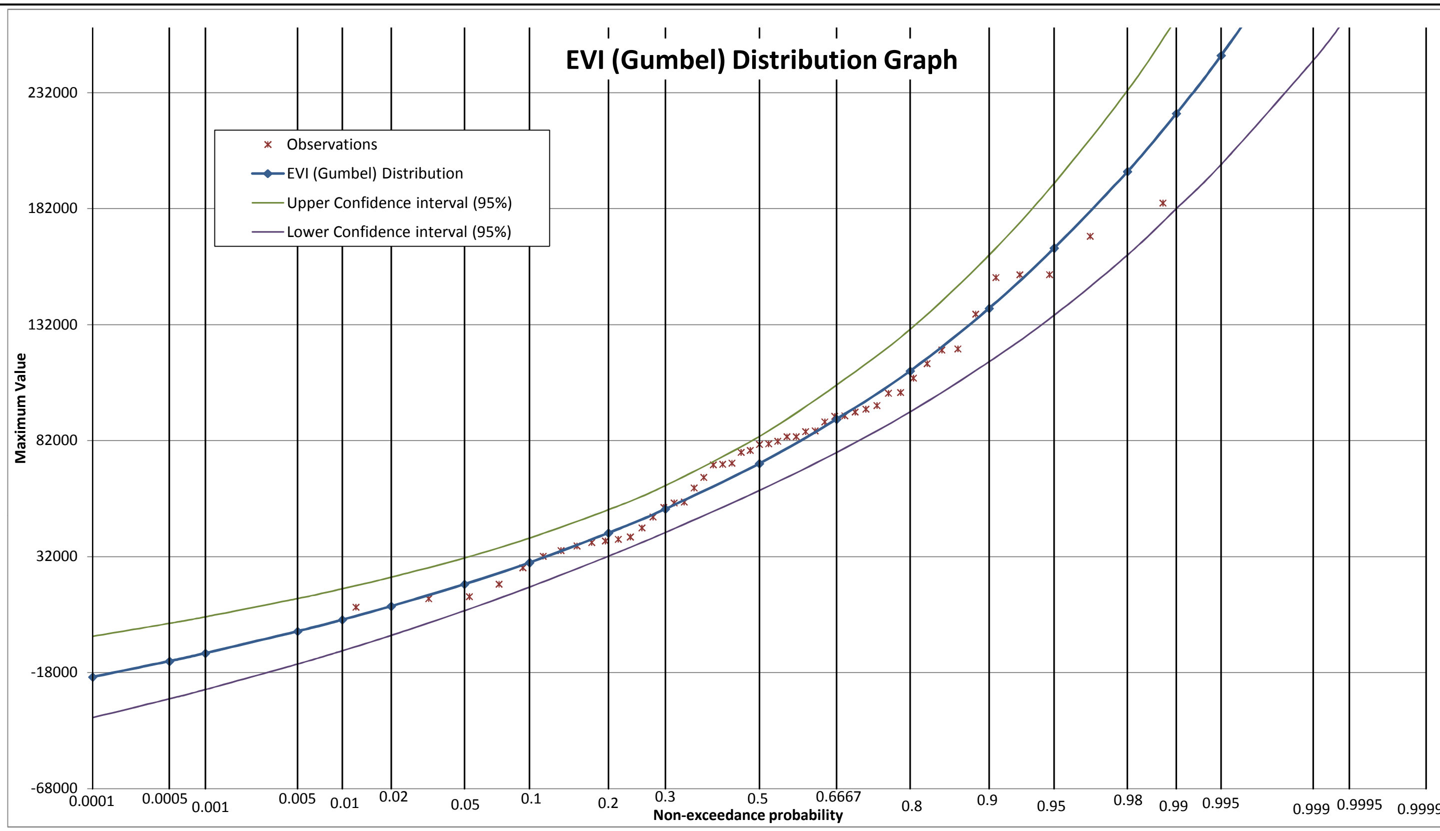


Pearson III Distribution
Visual Goodness-of-Fit Test Notes:

Good visual fit throughout distribution

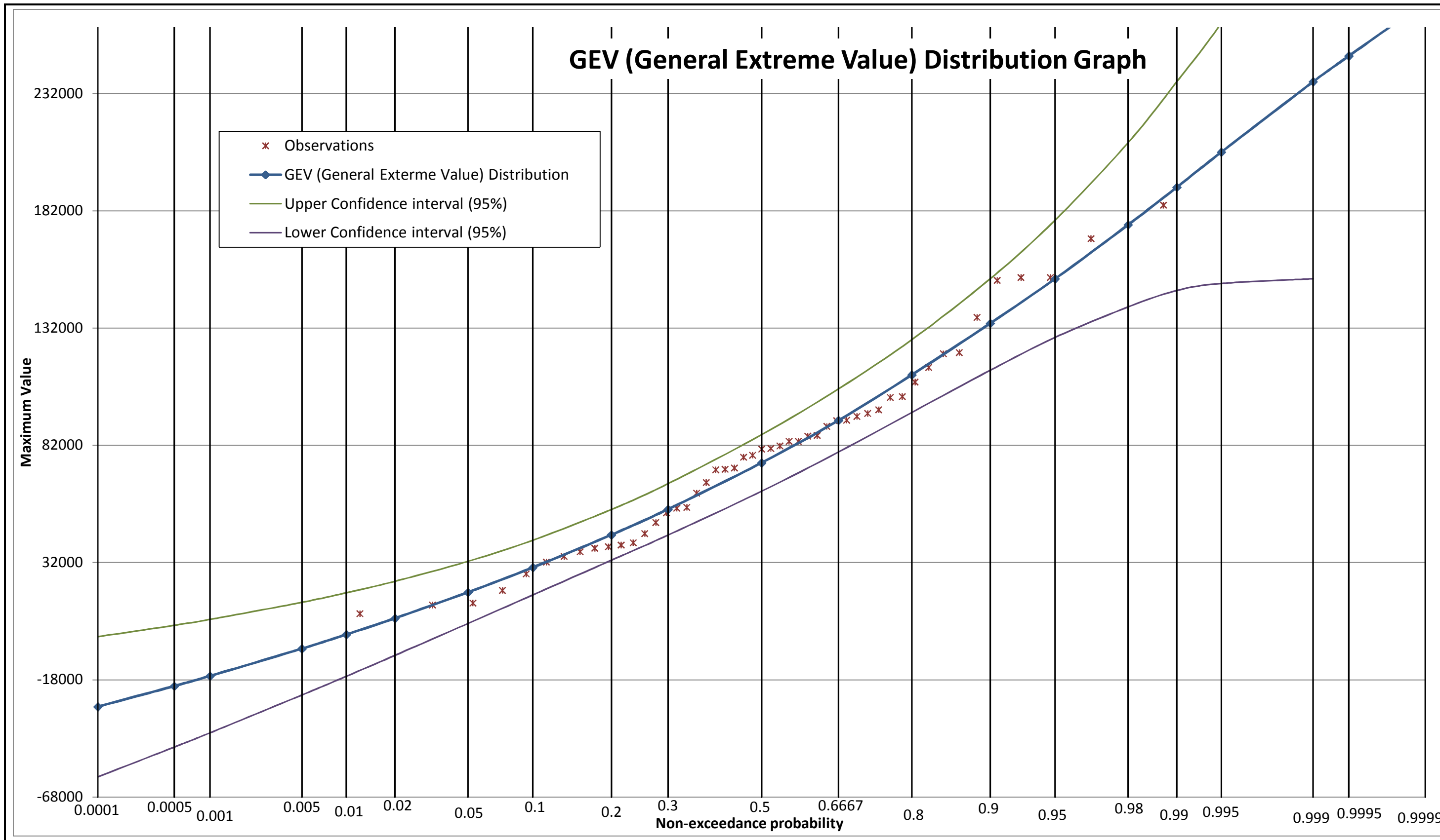


Log Pearson III Distribution Visual Goodness-of-Fit Test Notes:
Good visual fit throughout distribution
The log-Pearson III applies to hydrologic frequency analsisys only when the following shape tests pass
Lambda shape test
Pass
Alpha shape test
Fail

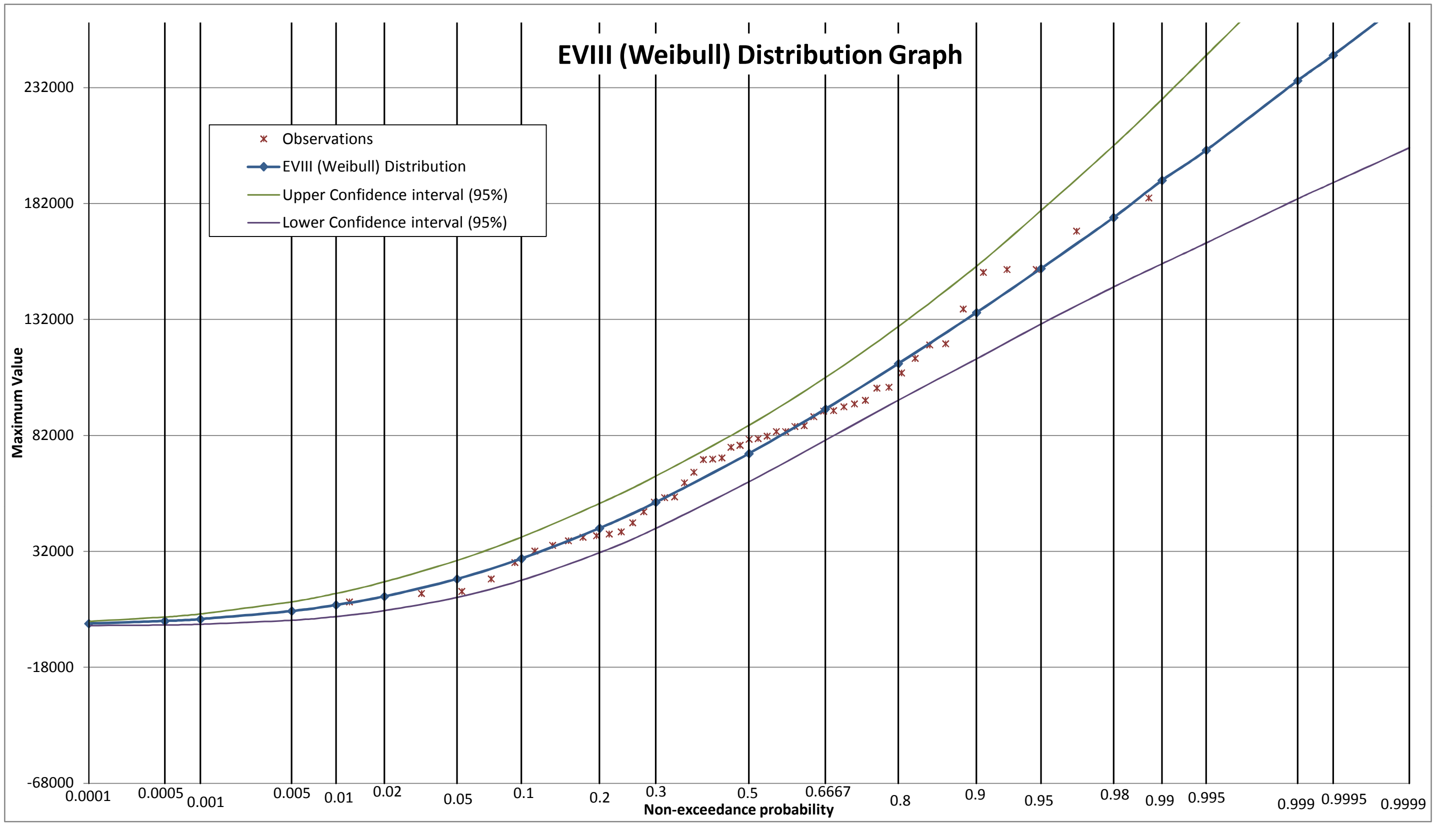


EVI (Gumbel) Distribution Visual Goodness-of-Fit Test Notes:

Distribution appears to be underestimating values at high exceedance probabilities

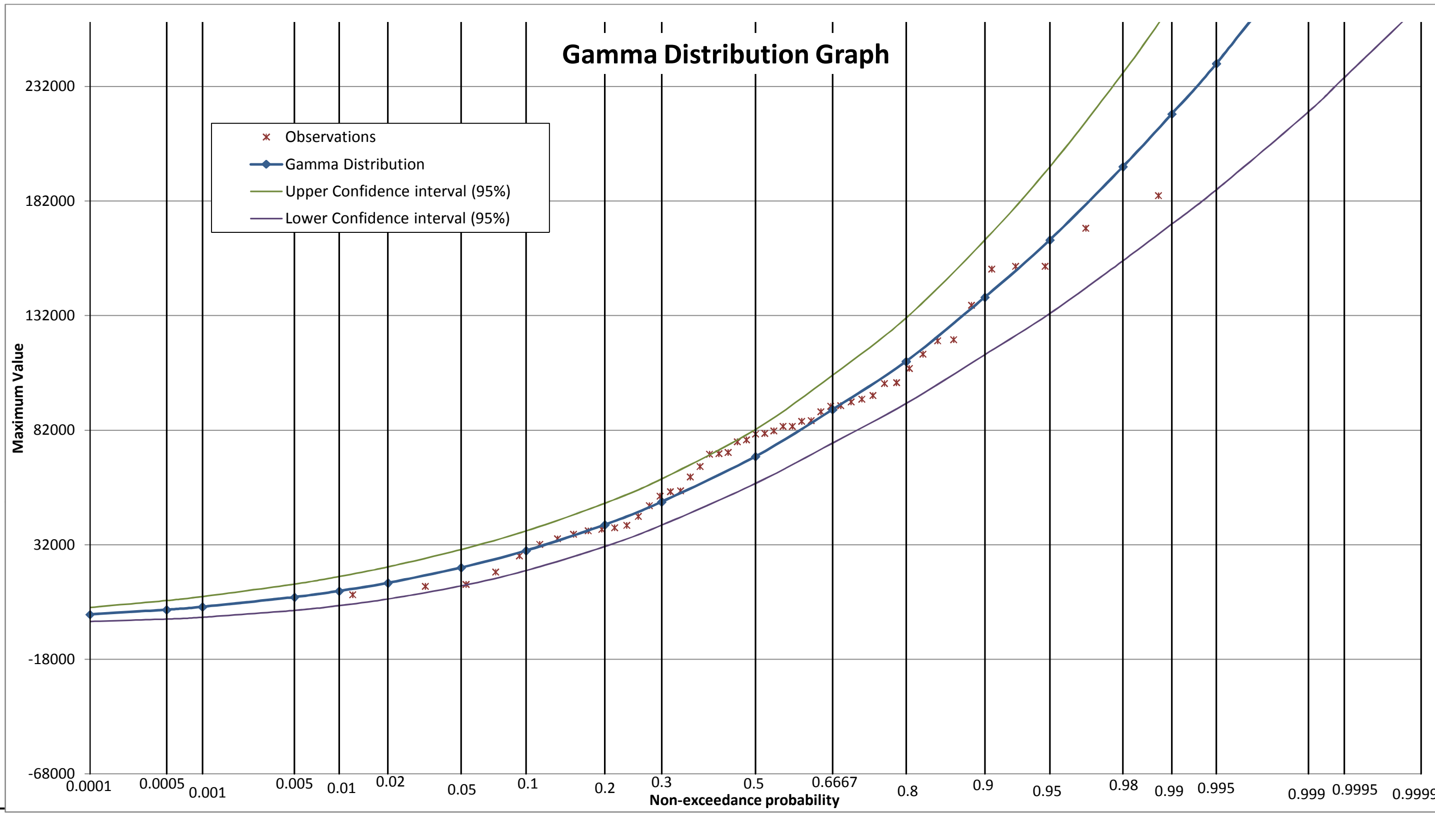


GEV (General Extreme Value) Distribution Visual Goodness-of-Fit Test Notes:
Good fit, may be over estimating values at high non-exceedance probabilities



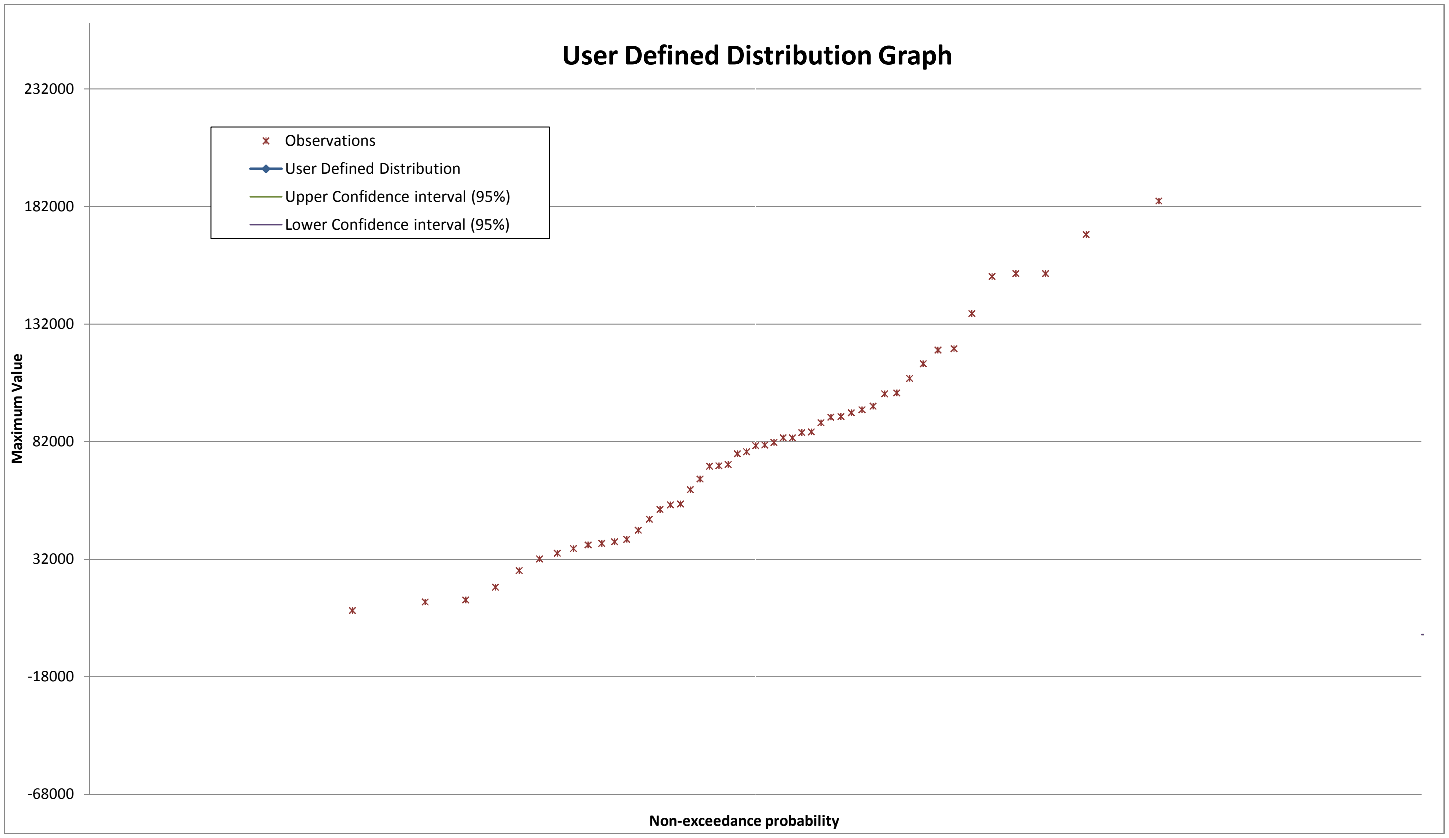
EVIII (Weibull) Distribution
Visual Goodness-of-Fit Test Notes:

Most observations fall outside of confidence intervals with likely under-estimating values at high non-exceedance probabilities



Gamma Distribution
Visual Goodness-of-Fit Test Notes:

Most observations fall outside of confidence intervals with likely under-estimating values at high non-exceedance probabilities



User Defined Distribution
Visual Goodness-of-Fit Test Notes:

Input comments on Visual Goodness-of-Fit here

Numerical Tests

Choose Significance Level (alpha) : **5%**

1) Anderson-Darling Test (1952)

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\cdot[\ln F(X_i) + \ln(1 - F(X_{n-i+1}))]$$

H0= Data follows specified distribution
HA= Data does not follow the specified distribution

Distribution Type:	Critical Value at 10%	Critical Value at 5%	Critical Value at 1%	A2	Hypothesis	Rank (1 = best fit)
Normal	1.929	2.502	3.907	0.470	Accept H0	7
Lognormal	1.929	2.502	3.907	1.080	Accept H0	9
Lognormal III	1.929	2.502	3.907	0.307	Accept H0	4
Exponential	1.929	2.502	3.907	3.235	Reject H0	10
Pearson III	1.929	2.502	3.907	0.288	Accept H0	1
Log Pearson III	1.929	2.502	3.907	0.338	Accept H0	5
Gumbel	1.929	2.502	3.907	0.366	Accept H0	6
GEV	1.929	2.502	3.907	0.302	Accept H0	3
Weibull	1.929	2.502	3.907	0.291	Accept H0	2
Gamma	1.929	2.502	3.907	0.530	Accept H0	8

*Critical values based on values calculated by EasyFit Software

2) Kolmogorov-Smirnov Test (1933)

$$F_n(x) = \frac{1}{n}\cdot\left[\text{Number of observations} \le x\right] \qquad D_n = \sup_x |F_n(x) - F(x)|$$

H0= Data follows specified distribution
HA= Data does not follow the specified distribution

Distribution Type:	Critical Value at 10%	Critical Value at 5%	Critical Value at 1%	Dn	Hypothesis	Rank (1 = best fit)
Normal	0.174	0.194	0.233	0.066	Accept H0	1
Lognormal	0.174	0.194	0.233	0.158	Accept H0	9
Lognormal III	0.174	0.194	0.233	0.089	Accept H0	5
Exponential	0.174	0.194	0.233	0.204	Reject H0	10
Pearson III	0.174	0.194	0.233	0.070	Accept H0	2
Log Pearson III	0.174	0.194	0.233	0.100	Accept H0	6
Gumbel	0.174	0.194	0.233	0.107	Accept H0	7
GEV	0.174	0.194	0.233	0.084	Accept H0	3
Weibull	0.174	0.194	0.233	0.088	Accept H0	4
Gamma	0.174	0.194	0.233	0.128	Accept H0	8

Expected Probability Analysis (Alberta Environment 1981)

Chance of Occurrence of Expected Flood Events	83%

$$P_{nr} = p_n q_r - \frac{n\,(r!)}{(r-n)!\,n!}$$

Return Period (Years)	Range of Expected Number of Flood Events Greater than or Equal to the Return Period		Number of Flood Events Greater than or Equal to the Return Period based on Distribution									
	Low	High	Normal	Lognormal	Lognormal III	Exponential	Pearson III	Log Pearson III	Gumbel	GEV	Weibull	Gamma
10000	0	0	0	0	0	0	0	0	0	0	0	0
2000	0	0	0	0	0	0	0	0	0	0	0	0
1000	0	0	0	0	0	0	0	0	0	0	0	0
200	0	1	0	0	0	0	0	0	0	0	0	0
100	0	1	1	0	0	0	0	0	0	0	0	0
50	0	3	2	0	1	0	1	1	0	1	1	0
20	0	5	5	0	2	0	4	2	2	4	2	2
10	1	7	6	2	6	2	6	5	5	6	6	5
5	5	13	9	9	9	6	9	9	9	9	9	9
3	11	21	13	18	17	20	15	15	17	16	15	17
2	19	29	25	30	27	32	27	27	28	27	27	30
1.4286	30	39	32	36	34	43	34	35	35	34	35	35
1.25	36	45	37	40	37	45	37	37	37	37	37	37
1.1111	42	49	45	44	44	46	44	44	44	44	44	44
1.0526	44	49	48	45	46	48	46	45	46	46	46	45
1.0204	47	49	49	46	49	49	49	48	48	49	48	46
1.0101	48	49	49	47	49	49	49	49	49	49	49	48
1.005	48	49	49	48	49	49	49	49	49	49	49	49
1.001	49	49	49	49	49	49	49	49	49	49	49	49
1.0005	49	49	49	49	49	49	49	49	49	49	49	49
1.0001	49	49	49	49	49	49	49	49	49	49	49	49
		# our of range	0	3	0	2	0	0	0	0	0	2
		Rank	1	10	1	8	1	1	1	1	1	8

Least Squares Ranking

Distribution Type:	Standard Error	Rank
Normal	7440	8
Lognormal	19845	9
Lognormal III	5054	4
Exponential	29916	10
Pearson III	4985	2
Log Pearson III	5104	5
Gumbel	6818	6
GEV	5012	3
Weibull	4772	1
Gamma	6991	7

$$SE_j = \sqrt{\frac{1}{n-m_j}\sum_{i=1}^{n}(x_i - y_i)^2}$$

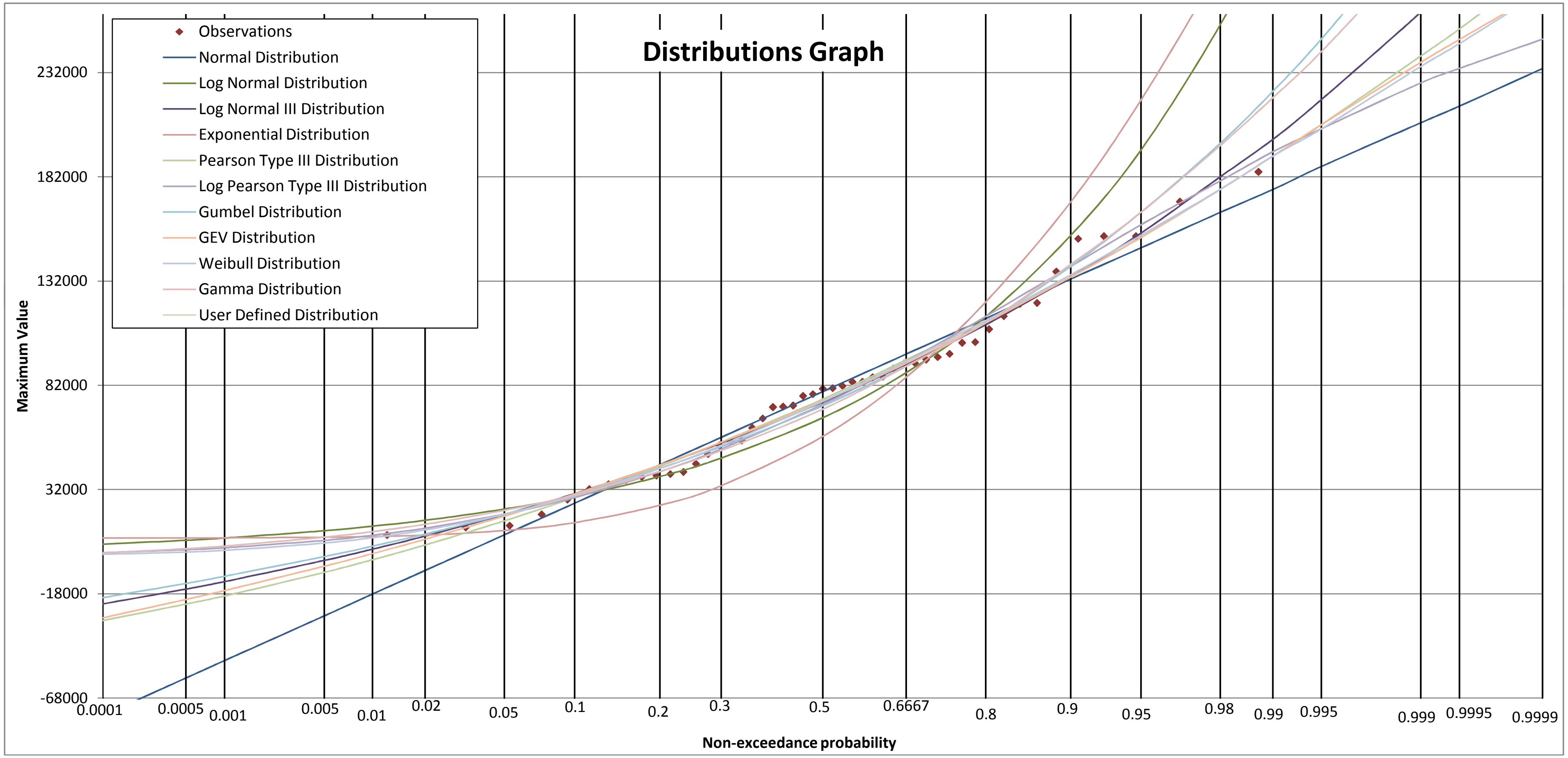
Sampling and Distribution Uncertainty

NOTES

- Select a distribution type and a return period of interest
- The Sample Uncertainty, Distribution Uncertainty and Total Uncertainty will be displayed on the right
- The plot below displays all the distributions input in the Frequency Analysis Input Tab

Return Period of Interest (Years)	50
Distribution Type	Lognormal III
Corresponding Value	182000

Sampling Uncertainty at (95%) Confidence Interval ±	40500
Distribution Uncertainty ±	4130
Total Uncertainty ±	44600



Summary Sheet

Initial Statistical Tests:

Tests for Stationarity

Test	Result
Spearman Rank Order Correlation Coefficient	No Significant Trend at 0.05 Significance Level
Mann-Whitney Test for jump (a.k.a. Mann-Whitney U test)	No Jump at 0.05 Significance Level
Wald-Wolfowitz Test (The runs test)	

Tests for Homogeneity

Test	Result
Mann-Whitney Test for jump (a.k.a. Mann-Whitney U test)	Sample is Homogeneous at 0.05 Significance Level
Terry Test	Sample is Homogeneous at 0.05 Significance Level

Tests for Independence

Test	Result
Spearman Rank Order Correlation Coefficient	Non-independence Detected at 0.01 Significance Level
Wald-Wolfowitz Test for Independence	Data is independent at 0.01 Significance Level
Anderson Test	Data is independent at 0.01 Significance Level

Test for Outliers

Test	Result
Grubbs and Beck Test for Outliers	
Are any high outliers present?	No High Outliers Present
Are and low outliers present?	Low Outlier May Be Present

Project Information

Project Name:	Evaporation Pond - Example #1 (De-correlated data set)
Project Description:	A typical evaporation pond somewhere in Calgary. The area has a size of 100 hectares of which 50% is developed area, with the remainder undeveloped. All runoff is to be fully contained. In the first example, we have only two catchment areas, one for the developed areas (say 50 ha with 60% hard) and landscaped and one for the area around the pond (also 50ha). The pond is assumed at a zero m2 footprint at the bottom, with 0.57% sideslopes. No irrigation of the landscaped areas.
Location:	Say somewhere in northeast or southeast Calgary
Date:	09/11/2012
Designed by:	Bert van Duin
Company Name:	City of Calgary Water Resources - Infrastructure Planning
Reviewed by:	Who volunteers?

Numerical Goodness-of-fit Tests Results

Distribution Type	Numerical Goodness-of-fit Tests from Spreadsheet				Average of Ranks	Ranking from Numerical Tests	Numerical Goodness-of-fit Tests from Hyfran (Input by user)		Notes from Visual Goodness-of-fit Test
	A-D Test	K-S Test	Expected Probability	Least Squares Ranking			BIC	AIC	
Normal	7	1	1	8	4.25	5	5	8	Several points above the upper confidence interval at both low and high end of non-exceedance probability and below the lower confidence interal near 0.5 non-exceedance probability
Lognormal	9	9	10	9	9.25	9	9	9	A few points above the upper confidence interval at both low and high end of non-exceedance probability and some below the lower confidence interal near 0.5 non-exceedance probability
Lognormal III	4	5	1	4	3.50	4	7	6	All points within confidence interval, very good fit for majority of points with the exception of the largest event. Good fit, may be over estimating values at high non-exceedance probabilities. As the values are for pond volumes, it might be possible that the pond overflowed. The data were checked and the pond design information indicates that these levels are below the spillway crest.
Exponential	10	10	8	10	9.50	10	10	10	Some points outside of confience interval for lower non-exceedance probabilities. Good fit at higher non-exceedance probabilities.
Pearson III	1	2	1	2	1.50	1	8	7	Good visual fit throughout distribution
Log Pearson III	5	6	1	5	4.25	5	4	3	Good visual fit throughout distribution
Gumbel	6	7	1	6	5.00	7	3	4	Distribution appears to be underestimating values at high exceedance probabilities
GEV	3	3	1	3	2.50	3	6	5	Good fit, may be over estimating values at high non-exceedance probabilities
Weibull	2	4	1	1	2.00	2	1	1	Most observations fall outside of confidence intervals with likely under-estimating values at high non-exceedance probabilities
Gamma	8	8	8	7	7.75	8	2	2	Most observations fall outside of confidence intervals with likely under-estimating values at high non-exceedance probabilities

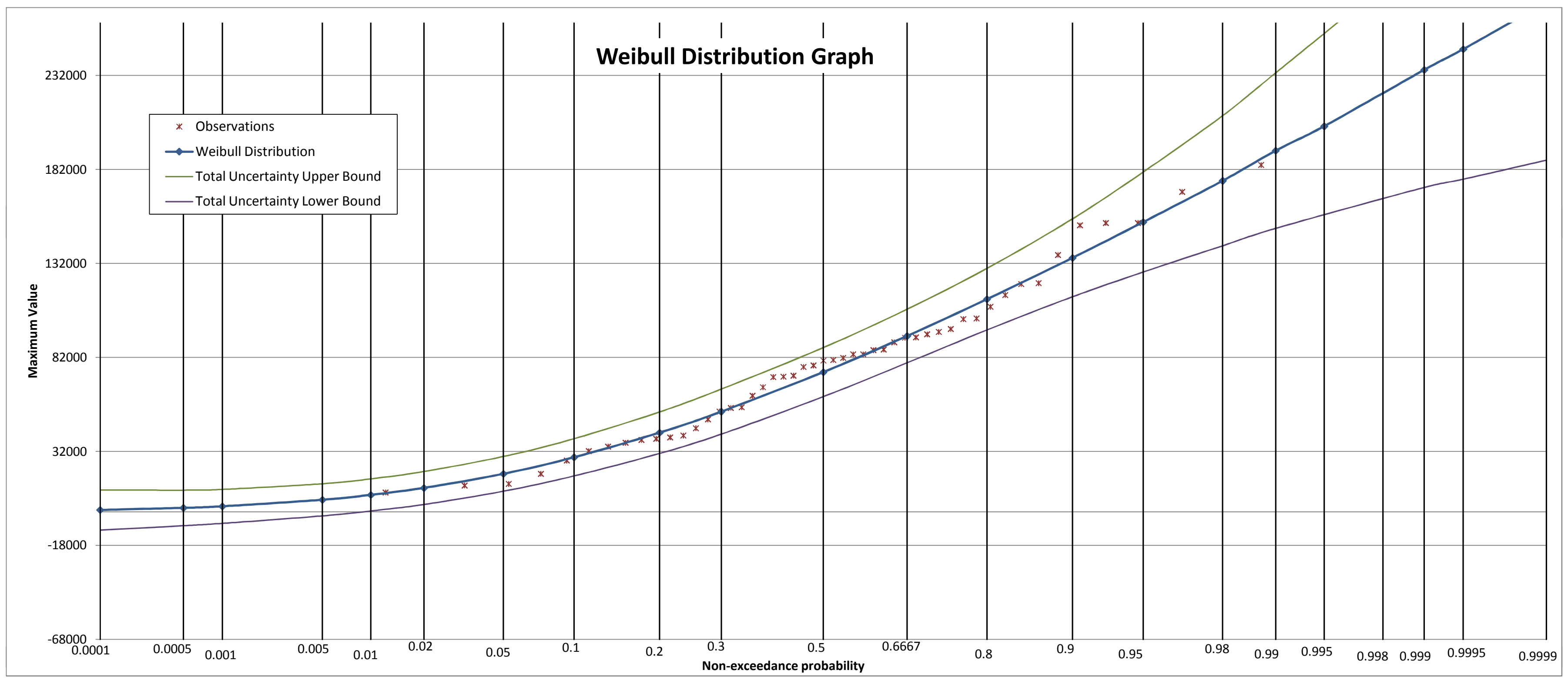
Selected Distribution and Results

Distribution type chosen based on visual and numerical goodness-of-fit tests:	Weibull

Instructions:
- Based on the results of the numerical and visual goodness-of-fit tests presented above, choose the preferred distribution in the cell on the left

Return Period	Probability	Magnitude	Total Uncertainty (Upper Bound)	Total Uncertainty (Lower Bound)
10000	0.9999	271000	355000	187000
2000	0.9995	246000	315000	177000
1000	0.9990	235000	298000	172000
500	0.9980	223000	279000	166000
200	0.9950	205000	253000	157000
100	0.9900	192000	233000	151000
50	0.9800	176000	211000	141000
20	0.9500	154000	181000	127000
10	0.9000	135000	156000	114000
5	0.8000	113000	129000	96600
3	0.6667	93400	108000	79100
2	0.5000	74200	87300	61100
1.4286	0.3000	53200	65200	41200
1.25	0.2000	42000	53100	30900
1.1111	0.1000	28900	38800	19000
1.0526	0.0500	20100	29400	10900
1.0204	0.0200	12600	21400	3820
1.0101	0.0100	8900	17500	350
1.005	0.0050	6280	14800	-2280
1.001	0.0010	2800	11800	-6240
1.0005	0.0005	1980	11400	-7480
1.0001	0.0001	884	11600	-9810

*Total uncertainty is based on sampling uncertainty at ((95%) Confidence Interval) plus distribution uncertainty of Top 4 distributions (based on numerical goodness of fit tests)



Errors and Warnings

Cumulative distribution function warning
No warning
No warning
No warning
No warning
No warning
No warning
No warning
No warning
No warning
No warning

If a warning is present, please check if hyfran output results were pasted correctly. If hyfran results were pasted correctly the warning signifies that the Continious Distribution Function (CDF) used in this workbook does not produce same output values as the input frequency analysis results, which in turn indicates that the numerical goodness-of-fit tests calculated by this spreadsheet for this distribution may be based on inaccurate numbers. Another possible solution would be to use a different method of estimating the CDF parameters for example: method of weighted moments.

Appendix C

Zero Values

Raw Data		
Year	Depth (m)	Elevation (m)
1964	0.0	647.6
1965	0.0	647.6
1966	2.6	650.2
1967	0.7	648.3
1968	2.9	650.5
1969	0.0	647.6
1970	0.0	647.6
1971	1.3	648.9
1972	0.4	648.0
1973	0.0	647.6
1974	1.3	648.9
1975	0.0	647.6
1976	0.0	647.6
1977	0.0	647.6
1978	0.0	647.6
1979	0.0	647.6
1980	1.5	649.1
1981	0.4	648.0
1982	0.0	647.6
1983	2.2	649.8
1984	0.0	647.6
1985	0.0	647.6
1986	0.8	648.4
1987	0.0	647.6
1988	2.7	650.3
1989	2.3	649.9
1990	0.3	647.9
1991	1.4	649.0
1992	0.0	647.6
1993	0.0	647.6
1994	0.9	648.5
1995	0.4	648.0
1996	2.1	649.7
1997	2.0	649.6
1998	0.0	647.6
1999	0.9	648.5
2000	0.0	647.6
2001	0.2	647.8
2002	0.0	647.6
2003	2.4	650.0
2004	0.2	647.8
2005	0.0	647.6
2006	0.4	648.0
2007	0.0	647.6
2008	0.0	647.6
2009	0.0	647.6
2010	2.0	649.6
2011	1.4	649.0
2012	0.7	648.3
2013	2.8	650.4

Preliminary Analysis			
Year	Depth (m)	Elevation (m)	Rank
1964	0.0	647.6	50
1965	0.0	647.6	49
1969	0.0	647.6	48
1970	0.0	647.6	47
1973	0.0	647.6	46
1975	0.0	647.6	45
1976	0.0	647.6	44
1977	0.0	647.6	43
1978	0.0	647.6	42
1979	0.0	647.6	41
1982	0.0	647.6	40
1984	0.0	647.6	39
1985	0.0	647.6	38
1987	0.0	647.6	37
1992	0.0	647.6	36
1993	0.0	647.6	35
1998	0.0	647.6	34
2000	0.0	647.6	33
2002	0.0	647.6	32
2005	0.0	647.6	31
2007	0.0	647.6	30
2008	0.0	647.6	29
2009	0.0	647.6	28
2004	0.2	647.8	27
2001	0.2	647.8	26
1990	0.3	647.9	25
2006	0.4	648.0	24
1995	0.4	648.0	23
1981	0.4	648.0	22
1972	0.4	648.0	21
1967	0.7	648.3	20
2012	0.7	648.3	19
1986	0.8	648.4	18
1999	0.9	648.5	17
1994	0.9	648.5	16
1971	1.3	648.9	15
1974	1.3	648.9	14
1991	1.4	649.0	13
2011	1.4	649.0	12
1980	1.5	649.1	11
2010	2.0	649.6	10
1997	2.0	649.6	9
1996	2.1	649.7	8
1983	2.2	649.8	7
1989	2.3	649.9	6
2003	2.4	650.0	5
1966	2.6	650.2	4
1988	2.7	650.3	3
2013	2.8	650.4	2
1968	2.9	650.5	1

Cunnane Plotting Position (p_m)	
for non-zero values	for full data set
	0.988
	0.968
	0.948
	0.928
	0.908
	0.888
	0.869
	0.849
	0.829
	0.809
	0.789
	0.769
	0.749
	0.729
	0.709
	0.689
	0.669
	0.649
	0.629
	0.610
	0.590
	0.570
	0.550
0.978	0.530
0.941	0.510
0.904	0.490
0.868	0.470
0.831	0.450
0.794	0.430
0.757	0.410
0.721	0.390
0.684	0.371
0.647	0.351
0.610	0.331
0.574	0.311
0.537	0.291
0.500	0.271
0.463	0.251
0.426	0.231
0.390	0.211
0.353	0.191
0.316	0.171
0.279	0.151
0.243	0.131
0.206	0.112
0.169	0.092
0.132	0.072
0.096	0.052
0.059	0.032
0.022	0.012

Non-zero	27
Total	50
Probability of the depth exceeding zero	0.54
Above 2.4 (650 m) or above	5
Probability of an exceedance of 2.4	18.5%

zero values
non zero values

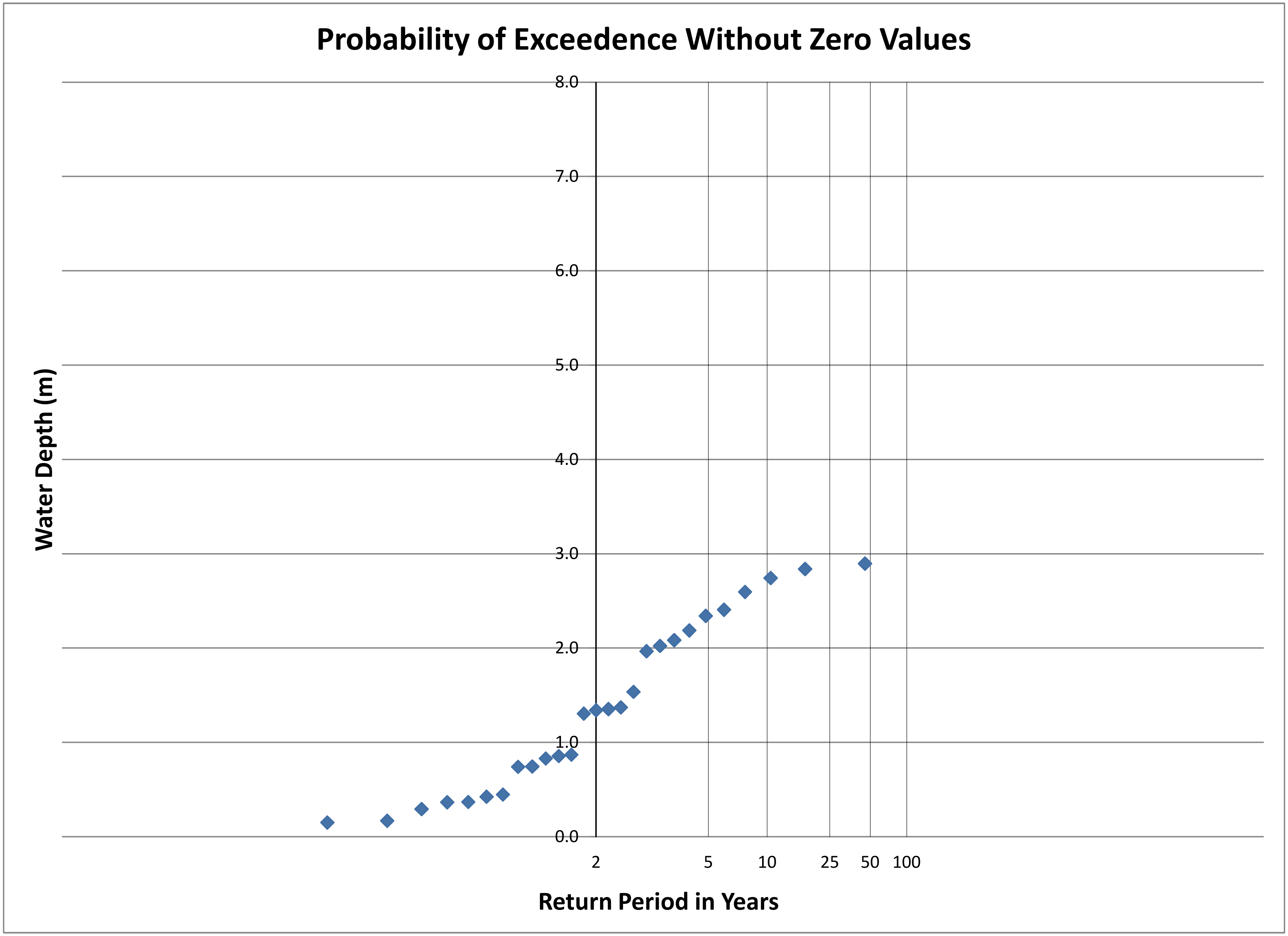
$p_m = (m-0.4)/(N+0.2)$
N = total number of non-zero values

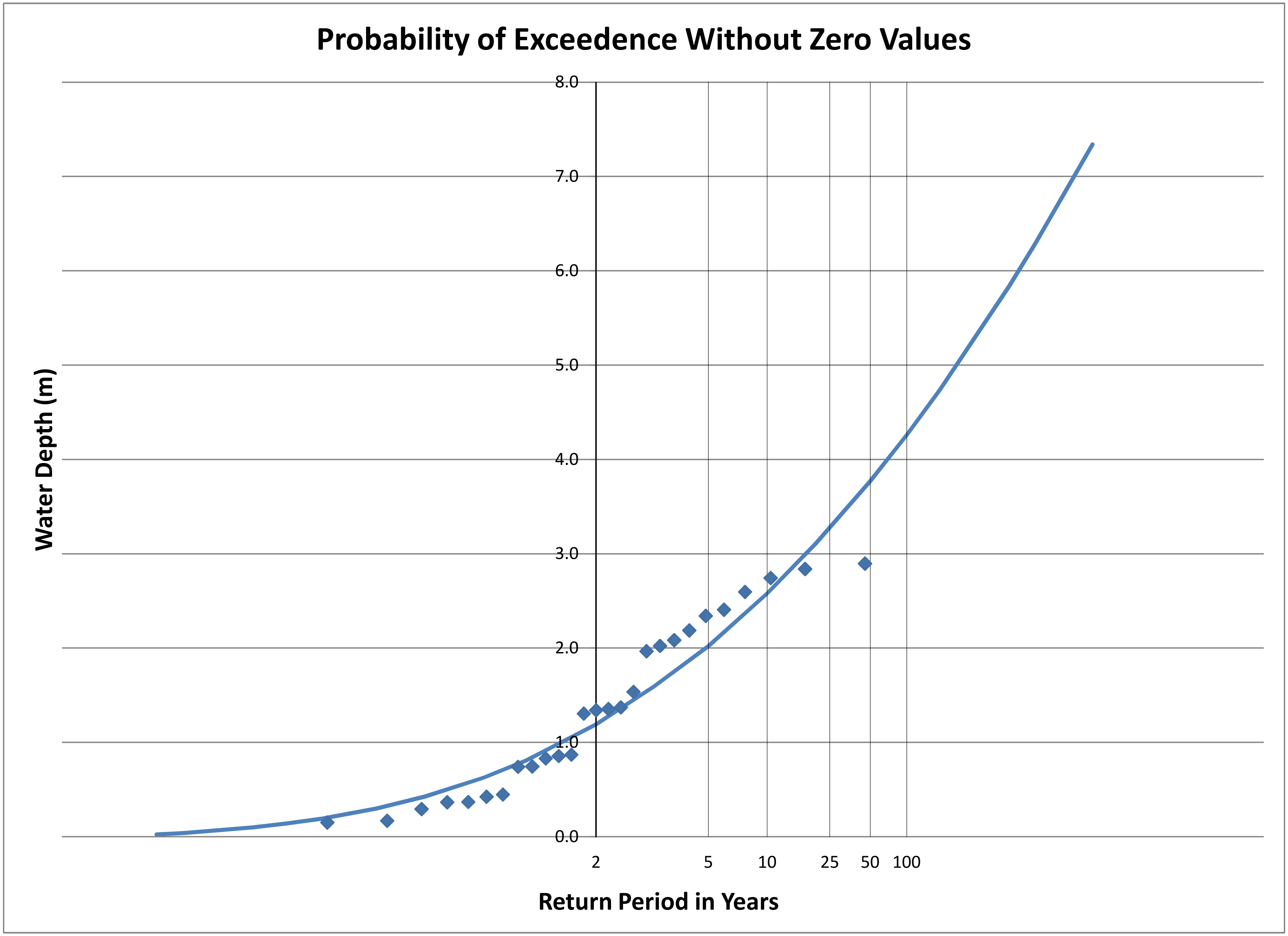
m = rank of event in question

Fitted Gamma Distribution		Probability by Fitting a Gamma Distribution			Normal Probability Plotting Position			
XT	**q**	**for non-zero values**	**for full data set**		**from gamma distribution**		**from Cunnane**	
7.34	0.9999	1E-04	5.4E-05		3.72	3.87		-2.26
6.29	0.9995	0.0005	0.00027		3.29	3.46		-1.85
5.83	0.999	0.001	0.00054		3.09	3.27		-1.63
4.74	0.995	0.005	0.0027		2.58	2.78		-1.46
4.26	0.99	0.01	0.0054		2.33	2.55		-1.33
3.77	0.98	0.02	0.0108		2.05	2.30		-1.22
3.11	0.95	0.05	0.027		1.64	1.93		-1.12
2.58	0.9	0.1	0.054		1.28	1.61		-1.03
2.02	0.8	0.2	0.108		0.84	1.24		-0.95
1.59	0.6667	0.3333	0.179982		0.43	0.92		-0.87
1.19	0.5	0.5	0.27		0.00	0.61		-0.80
0.808	0.3	0.7	0.378		-0.52	0.31		-0.74
0.626	0.2	0.8	0.432		-0.84	0.17		-0.67
0.425	0.1	0.9	0.486		-1.28	0.04		-0.61
0.299	0.05	0.95	0.513		-1.64	-0.03		-0.55
0.192	0.02	0.98	0.5292		-2.05	-0.07		-0.49
0.138	0.01	0.99	0.5346		-2.33	-0.09		-0.44
0.0982	0.005	0.995	0.5373		-2.58	-0.09		-0.38
0.0393	0.001	0.999	0.53946		-3.09	-0.10		-0.33
0.0233	0.0005	0.9995	0.53973		-3.29	-0.10		-0.28
0.0001	0.0001	0.9999	0.539946		-3.72	-0.10		-0.23
Note: approximate values for this lower bound shown in the line above								-0.18
		1-q	(1-q)*0.54					-0.13
							-2.01	-0.07
							-1.56	-0.02
							-1.31	0.02
							-1.12	0.07
							-0.96	0.13
							-0.82	0.18
							-0.70	0.23
							-0.58	0.28
							-0.48	0.33
							-0.38	0.38
							-0.28	0.44
							-0.19	0.49
							-0.09	0.55
							0.00	0.61
							0.09	0.67
							0.19	0.74
							0.28	0.80
							0.38	0.87
							0.48	0.95
							0.58	1.03
							0.70	1.12
							0.82	1.22
							0.96	1.33
							1.12	1.46
							1.31	1.63
							1.56	1.85
							2.01	2.26

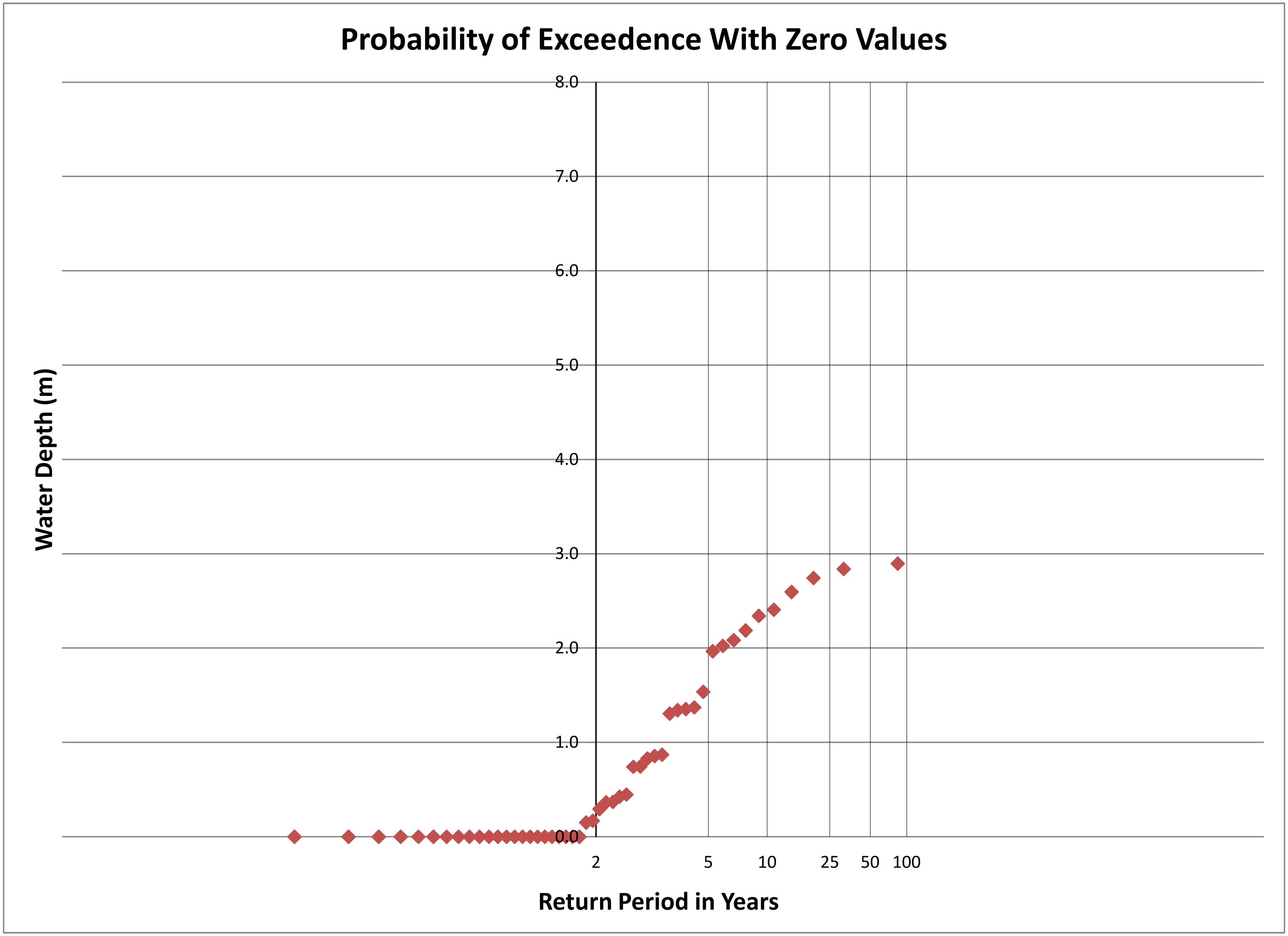
The exceedence probability for the the full data set is equal to the exceedence probability value in column N times the probability of a non-zero depth occuring (0.54).

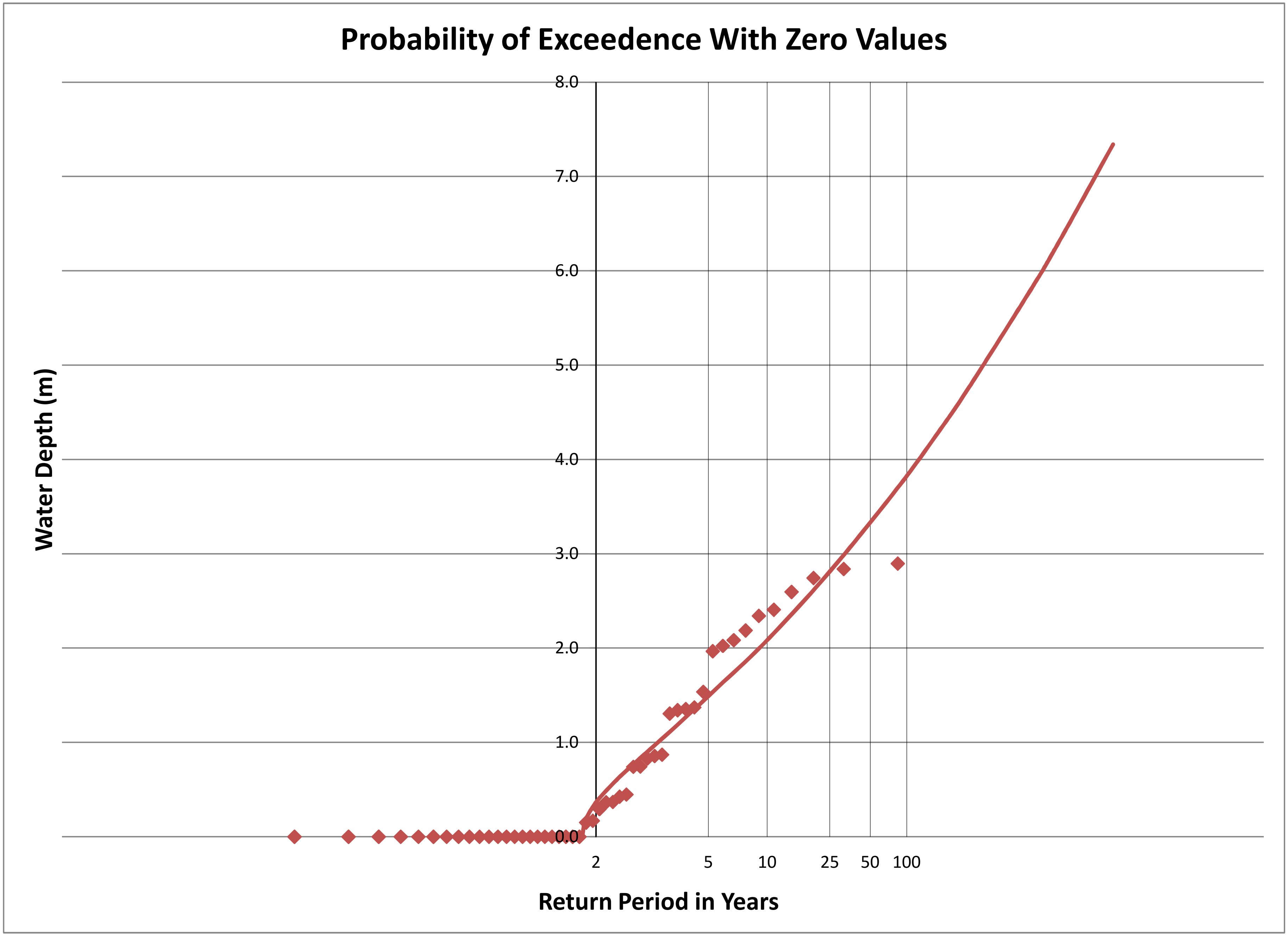
These are for the Return Period Lines on the plots		
	1000	0
2	-1.39E-16	-1.39E-16
5	8.42E-01	8.42E-01
10	1.28E+00	1.28E+00
25	1.75E+00	1.75E+00
50	2.05E+00	2.05E+00
100	2.33E+00	2.33E+00


Probability of Exceedence Without Zero Values
Water Depth (m)
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
2
5
10
25
50
100
Return Period in Years



Probability of Exceedence Without Zero Values
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
Water Depth (m)
2
5
10
25
50
100
Return Period in Years


Probability of Exceedence With Zero Values
Water Depth (m)
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
2
5
10
25
50
100
Return Period in Years



Probability of Exceedence With Zero Values
Water Depth (m)
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
2
5
10
25
50
100
Return Period in Years