Effective Modeling of Urban Water Systems Conference Proceedings, Monograph 13. (edited by W. James, K.N. Irvine, E.A. McBean, and R.E. Pitt). CHI. Guelph, Ontario, pp. 247 – 274. 2005.

STORMWATER QUALITY DESCRIPTIONS USING THE THREE PARAMETER LOGNORMAL DISTRIBUTION.

Alexander Maestre, Robert Pitt, S. Rocky Durrans, Subhabrata Chakraborti
University of Alabama. Tuscaloosa

Abstract

The cumulative probability distribution used to describe the variability of stormwater pollutant concentrations has been a matter of interest in recent years. Many predictive models attempt to estimate appropriate stormwater constituent concentrations based on land use and the amount of impervious area. The most important study that characterized stormwater was the Nationwide Urban Runoff Program (NURP) (EPA 1983). NURP was conducted throughout the U.S. and included about 2300 events from 1978 thru 1982. One of the conclusions of the final NURP report was that the event mean concentrations (EMCs) of stormwater constituents were described by lognormal distributions. This finding has been re-evaluated recently, with the conclusion that not all stormwater constituents were adequately described by lognormal distributions (Van Buren, 1997; Beherra, 2000).

Stormwater managers have generally accepted the assumption of lognormality of stormwater constituent concentrations between the 5th and 95th percentiles. Based on this assumption, it is common to use the log-transformed EMC values to evaluate differences between landuse categories and other characteristics. Statistical inference methods, like estimation and test of hypothesis, and analysis of variance (ANOVA) require statistical information about the distribution of the EMC to evaluate these differences. The use of the log-transformed data usually includes the location and scale parameter, but a lower bound parameter is usually neglected. In this paper, a large database, the National Stormwater Quality Database v.1.1 (NSQD) (Pitt, et al. 2003), will be used to evaluate a three-parameter lognormal distribution for stormwater constituent concentrations for different landuses. The NSQD is a compilation of the phase 1 data from the stormwater permit program. This paper will also evaluate this national data for the presence of unusual elevated values and their effect on the goodness of fit for the three parameter lognormal distribution.

1.0 Introduction

The National Stormwater Quality Database v. 1.1 (NSQD) contains water quality characteristics from the monitoring required by the NPDES Phase 1 stormwater permit applications and subsequent permits, during the period of 1992 to 2002. This database contains about 3770 events from 256 sites in 66 communities from throughout the U.S. For each site, much additional data, including the percentage of each land use in the catchment, the total area, the percentage of imperviousness, the geographical location and the season, has been included in the database. Information about the characteristics of each event is also included. Total precipitation, precipitation intensity, total runoff and antecedent dry period are also included, if collected. The database only contains data collected at the outfall, in-stream samples were not included in the database. Water quality characteristics where divided in four main groups: Common parameters, nutrients, metals and others (ie. pesticides and organic compounds). Much time and effort was spent in reviewing this data for QA/QC problems and correcting the information.

The Nationwide Urban Runoff Program (NURP) evaluated the characteristics of stormwater discharges at 81 outfalls in 28 communities throughout the U.S. (EPA, 1983). One of the conclusions of NURP was that stormwater constituent concentrations could be described using a lognormal distribution. Recently, Van Buren (1997) found that stormwater concentrations were described using a lognormal distribution for almost all constituents, with the exception of dissolved constituents that were better described with a normal distribution. Beherra (2000) also found that some stormwater constituent concentrations were better described using a lognormal distribution, while others were better described with gamma or exponential distributions. The constituents that were best described with a gamma distribution were: total solids, total Kjeldahl nitrogen (TKN), total phosphorous, chemical oxygen demand (COD), barium and copper. The constituents that were best described with an exponential distribution were: suspended solids, nitrates and aluminum. In both of these studies, fewer than 50 samples (collected at the same site) were available for evaluation.

During the research reported in this paper, statistical test were used to evaluate the log-normality of a selection of the constituents in the NSQD database. Statistical descriptions were obtained of each set of data including box and probability plots for each land use category and for the pooled dataset. It was found in almost all cases that the log-transformed data followed a straight line between the 5^{th} and 95^{th} percentile, as illustrated in Figure 1 for total dissolved solids (TDS) in residential areas.

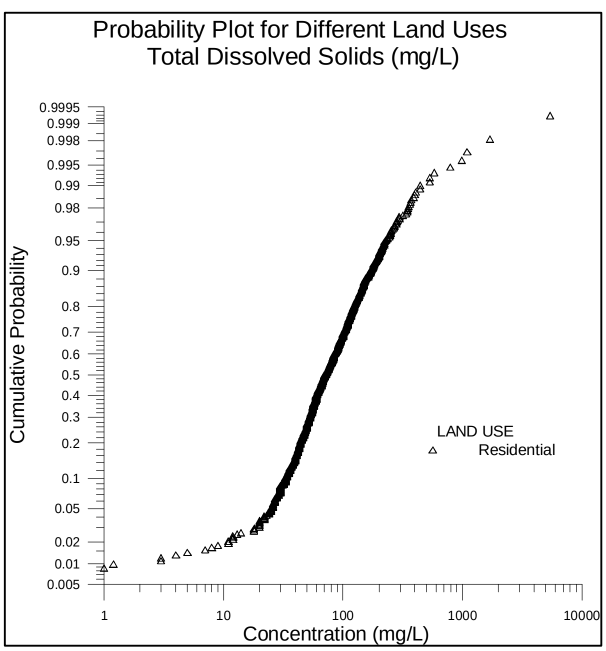


Figure 1. Probability plot of total dissolved solids in residential land uses.

For many statistical tests focusing on the central tendency (such as for determining the average concentration that is used for mass balance calculations), this may be a suitable fit. As an example, WinSLAMM, the Source Loading and Management Model (Pitt 1986; Pitt and Voorhees 1995), uses a Monte Carlo component to describe the likely variability of stormwater source flow pollutant concentrations using either lognormal or normal probability distributions for each constituent. However, if the extreme values are of importance (such as when dealing with the influence of many non-detectable values on the predicted concentrations, or determining the frequency of observations exceeding a numerical standard), a better description of the extreme values may be important.

The NSQD underwent an extensive data evaluation process, including multiple comparisons of the all data values in the database to original documents. In some cases, data was available from the local agency in electronic form. These spreadsheets were reformatted to be consistent to the NSQD format. However, it was found that all of the submitted electronic data needed to be verified against original data sheets and reports. When reviewing the NSQD, it was assumed that some of the events in the upper and lower tails of the distributions were caused by errors, most likely due to faulty transcription of the data (such as mislabeling the units for heavy metals or nutrients as mg/L instead of μg/L, for example). Unusual values were verified with the original reports and datasets. While some values (less than 5% of the complete dataset) were found to be in error and were corrected, most of the suspected values were found to be stormwater observations. Besides the targeted extreme values, all reported values were also examined in relationship to other related constituents (COD vs. BOD; total metal concentrations vs. dissolved metal concentrations; TKN vs. NH_3; TDS vs. specific conductivity; SS vs. turbidity; etc) and unusual behavior was further checked and corrected, as necessary. In some cases, unusual values could not be verified and were therefore eliminated from the dataset, although this was very unusual.

After the extensive QA/QC activities and corrections were made to the NSQD, the next step was to conduct a sensitivity analysis to determine the effects of the remaining unusual high and low values on the probability distribution parameters.

2.0 The Effects of Unusual High and Low Values on Probability Distribution Parameters

For this evaluation, 10,000 sets of 200 samples each were randomly generated following a lognormal distribution (1, 1), but having differing amounts of extreme values in each data set. For each set, the mean, variance and coefficient of variation were calculated. Two main factors were analyzed using this data: the extreme value factor and percentage of extreme values in each sample. The following percentages of extreme values were selected for evaluation: 0.5, 1, 5, 10, 25 and 50%. For each percentage of extreme values, the following factors were analyzed: 0.001, 0.01, 0.1, 10, 100, 1.000, 10,000, 100,000 and 1,000,000. For example (5%, 100) indicates that in each set, five percent of the data were increased by a factor of 100. The coefficient of variation was then calculated for each set of data. The medians of the coefficients of variation for the 10,000 runs are shown in Figure 2 for each level of extreme values.

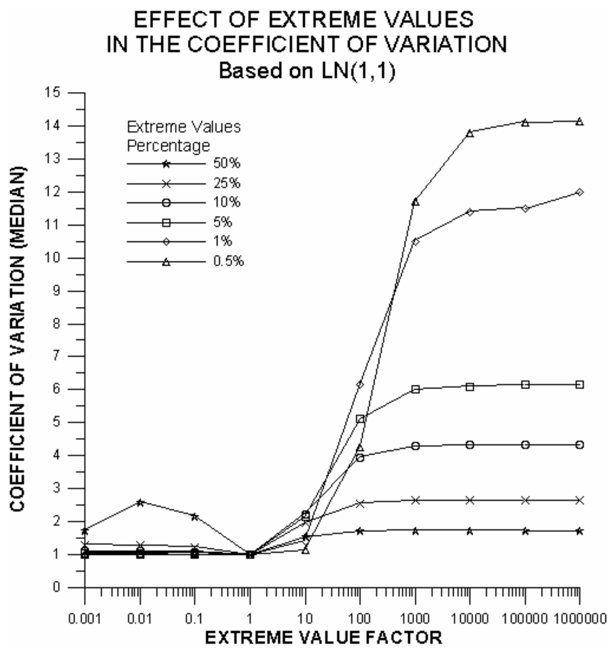


Figure 2. Effect of unusual values on the coefficient of variation

For a lognormal distribution (1,1) the coefficient of variation is one. Figure 2 shows how this original value is changed for different amounts of extreme values in the data sets, and for different factors in these extreme values. The horizontal axis represents the factor used in the extreme values. As an example, many of the incorrect extreme values observed in the NSQD for heavy metals were because the units were originally incorrectly reported as mg/L in the submitted information, while the correct units were actually µg/L. This would be an extreme value factor of 1,000. Extreme value factors of 10 were also fairly common and were associated with simple misplacements of decimal points in the data.

This figure shows that for small error factors (0.1, 0.01 and 0.001) there is not a large effect in the coefficient of variation for percentages smaller than 10%. For larger percentages the effect in the coefficient of variation is important. When 50% of the data is affected by an error factor of 0.01, the coefficient of variation was increased by almost three times.

High extreme value factors can have an important effect on the coefficient of variation. When 10 percent of the data was increased by a factor of 10, the coefficient of variation was increased almost three times. Notice that affecting 10 percent of the data by a factor of 10 has almost the same effect as affecting 50% of the data by a factor of a hundredth. This effect is reduced when the percentage of elevated values in the dataset is smaller than 10%.

For factors larger than a hundred, the effect on the coefficient of variation is much greater. Very low percentages of elevated values can increase the coefficient of variation by up to 15 times. For example, when only 0.5% of the sample is affected by a factor of a thousand, the coefficient of variation increases almost 12 times more than the correct value. As noted earlier this is important because it is not unusual to find reported values affected by a factor larger than a hundred (See Figure 1). Some of these values can be due to incorrect reporting units, but in many cases they were considered as valid observations because they were supported by similarly high values of other closely related constituents. For factors greater than 10^4 the multiplying value of the coefficient of variation remains stable at the maximum value obtained.

The above analyses indicate that in lognormal distributions, the presence of just a few unusual elevated values is important and can dramatically affect the reported coefficient of variation for the distribution of concentration. This observation is critical in the relatively common case were one or a very few observations are affected by a factor larger than a hundred. In the other extreme, factors smaller than one do not have a large impact on the reported coefficient of variation, except when the percentage of extreme values is greater than 50% (obviously, there are many other problems with that data set too).

The effect of extreme values on the mean and standard deviation was also analyzed. Figure 3 shows the effect of the extreme values on calculated standard deviation.

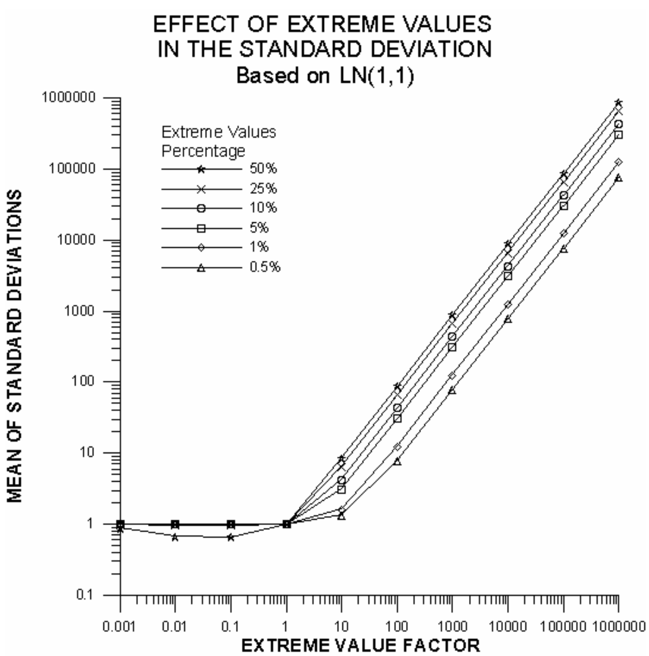


Figure 3. Effect of unusual values on the standard deviation

For large extreme value factors (larger than one) the standard deviation increases as the percentage of extreme values increases. For small extreme value factors, percentages smaller than 25% do not have an important effect on the standard deviation. For a specific extreme value factor, changing the extreme value percentages from 0.5% to 50% increases the standard deviation close to 10 times.

The effect of the presence of extreme value on the distribution mean is shown in Figure 4. For small extreme value factors, the mean is reduced almost 80 percent when the extreme value percentage is close to 50%. This is expected because in a lognormal (1,1) most of the values are located in the lower tail of the distribution. For extreme value occurrences less than 25%, the mean value is reduced by less than 20%.

Large extreme values factor have much larger effects on the distribution means. As the extreme value percentage increases, the calculated means also increase. If 0.5% of the values are affected by a factor of a hundred, the mean value is doubled. If 50% of the values are affected by the same factor, the mean values are increased by almost 50 times. For factors larger than a thousand, increasing the percentages of extreme values from 0.5% to 50% increases the mean values by up to two orders of magnitude.

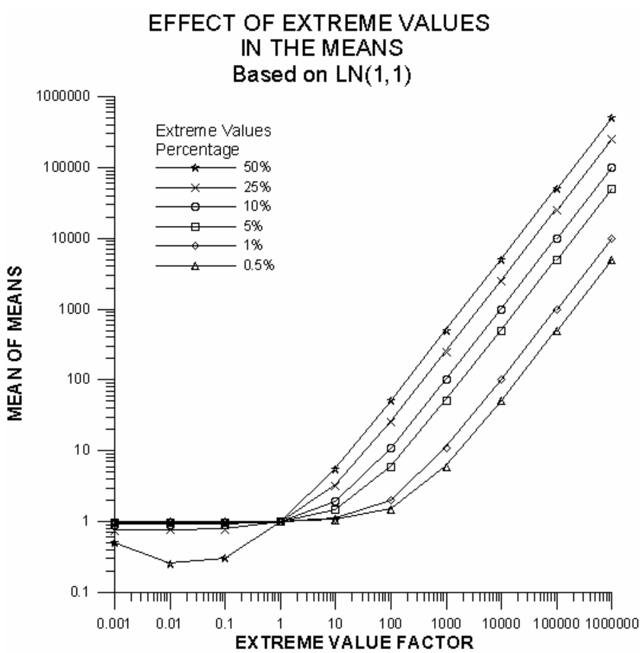


Figure 4. Effect of unusual values on the mean

These evaluations are important because it points out that for a lognormal distribution, the effects of few elevated values in the upper tail have much greater effects on common statistics than unusual values in the lower tail. Many stormwater researchers have focused on the lower tail, especially when determining how to handle the detection limits and unreported data. Stormwater constituents usually have unusual values in both tails of the probability distribution. It is a common to delete elevated values from the observations assuming they are expendable “outliers”. This practice is not recommended unless is there sufficient evidence that the observed values are a mistake. Actual elevated values can have a large effect on the calculated distribution parameters. If these are arbitrary removed, the data analyses will likely be flawed.

Effective Modeling of Urban Water Systems Conference Proceedings, Monograph 13.
(edited by W. James, K.N. Irvine, E.A. McBean, and R.E. Pitt). CHI. Guelph, Ontario,
pp. 247 – 274. 2005.

3.0 Analysis of Lognormality of Stormwater Constituents

The goodness of fitness of twenty nine stormwater constituent probability distributions was evaluated using the Kolmogorov-Smirnov test. Figure 5 shows how the test accepts or rejects the null hypothesis that the empirical and the estimated distributions are the same. If the null hypothesis is valid, then the constituent can be adequately represented by the lognormal distribution. The observations are sorted and a probability is assigned by its rank. The distribution generated by this ranking is known as the empirical distribution. The estimated distribution function is also compared on the same plot. The estimated distribution function is calculated with the mean and standard deviation of the original data. If the distance between the empirical and the estimated distributions is higher than a critical value d_α or D_{max}, the hypothesis of lognormality is rejected. Notice in Figure 5 that the horizontal axis has a logarithmic scale.

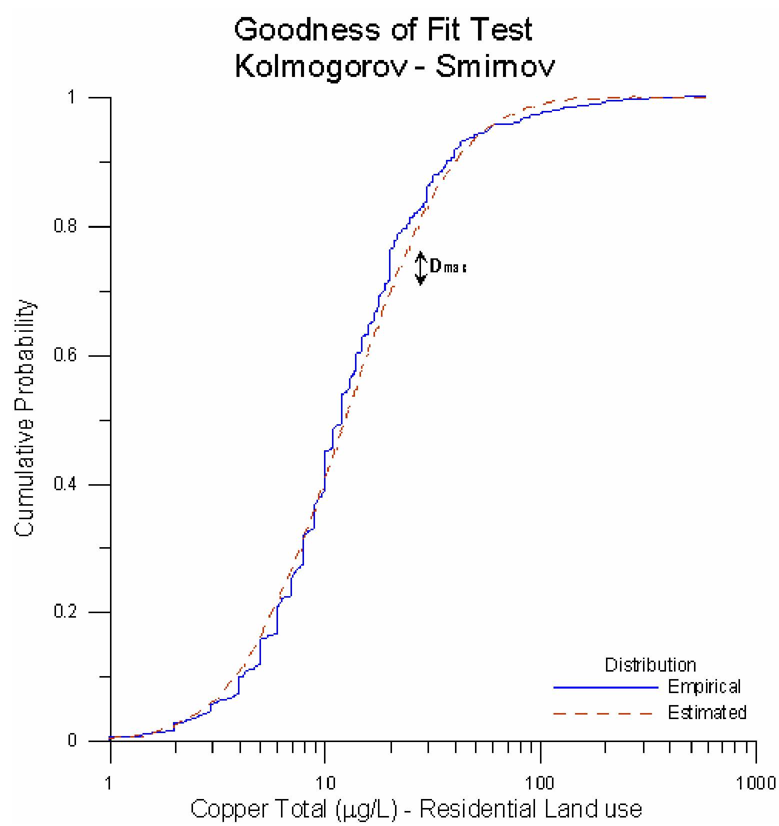


Figure 5. Cumulative and empirical probability distributions of total copper for residential land use data.

There are many options to assign the probability based on the ranks. Most methods assign the probability as a percentage of the total range. The probability of the observation is calculated as its rank divided by the number of observations. Kottegoda (1998) suggested that for extreme events analysis the plotting position can be calculated as:

$$p = \frac{i - 0.5}{n}$$

Where p is the cumulative probability of the observation, *i* is the rank of the observation and n is the total number of observations. We used this plotting position for these analyses because it does not restrict the probability of the largest observation to be one.

In the Kolmogorov-Smirnov test, the null hypothesis is that the observed data follow a lognormal distribution. If the sample size is small, and the distance between the empirical and the observed distributions is smaller than the critical value D_{max}, the test is interpreted as "there is not enough evidence to reject the hypothesis that the distribution is lognormal." In most cases, the NSQD contains enough samples to be able to accept or reject the null hypothesis with acceptable levels of confidence and power.

The NSQD contains many factors for each sampled event that likely affect the observed concentrations. These include such factors as seasons, geographical zones, rain intensities, etc. These factors may affect the shape of the probability distribution. As more data become available, the critical value D_{max} is reduced in the test. There will always be a specific number of samples that will lead to rejection of the null hypothesis because the maximum distance between the empirical and estimated probability distributions became larger than the critical value D_{max}. The only way to evaluate the required number of samples in each category is using the power of the test. Power is the probability that the test statistic will lead to a rejection of the null hypothesis (Gibbons and Chakraborti, 2003).

Masey (1950) states that the power of the Kolmogorov-Smirnov Test can be written as:

$$power = 1 - \Pr\left(\frac{-d_\alpha \pm \Delta\sqrt{n}}{\sqrt{F_1(x_0)(1 - F_1(x_0))}} < \frac{\{S_n(x_0) - F_1(x_0)\}\sqrt{n}}{\sqrt{F_1(x_0)(1 - F_1(x_0))}} < \frac{d_\alpha \pm \Delta\sqrt{n}}{\sqrt{F_1(x_0)(1 - F_1(x_0))}} \right)$$

where:

$d_\alpha = D_{max}$: critical distance at the level of significance α (confidence of the test)
S_n: Cumulative empirical probability distribution
F_1: Cumulative alternative probability distribution
Δ: maximum absolute difference between the cumulative estimated probability distribution and the alternative cumulative probability distribution

Massey also found that for large sample sizes, the power can be never be smaller than

$$power > 1 - \int_{2(-d_\alpha \pm \Delta\sqrt{n})}^{2(d_\alpha \pm \Delta\sqrt{n})} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

This reduced expression can be used to calculate the number of samples required to reject the null hypothesis with a desired power. Figure 6 shows the power of the d test for 1% and 5 % levels of confidence of the test (Massey, 1951). For example, assume that the maximum distance between the alternative cumulative and the estimated cumulative probability distributions is 0.2, and we want an 80% power against the alternative at a 5 percent level of confidence. To calculate the number of required samples, we read that $\Delta N^{0.5}$ is 1.8 for a power of 0.8 and 5% level of confidence. Solving for $N = (1.8/0.2)^2 =$ 81 samples. If we want to calculate the number of samples when the difference between the alternative cumulative and the estimated cumulative probability function is 0.05, with the same power and level of confidence, then 1,296 samples would be required. When the lines are very close together, it is obviously very difficult to statistically show that they are different.

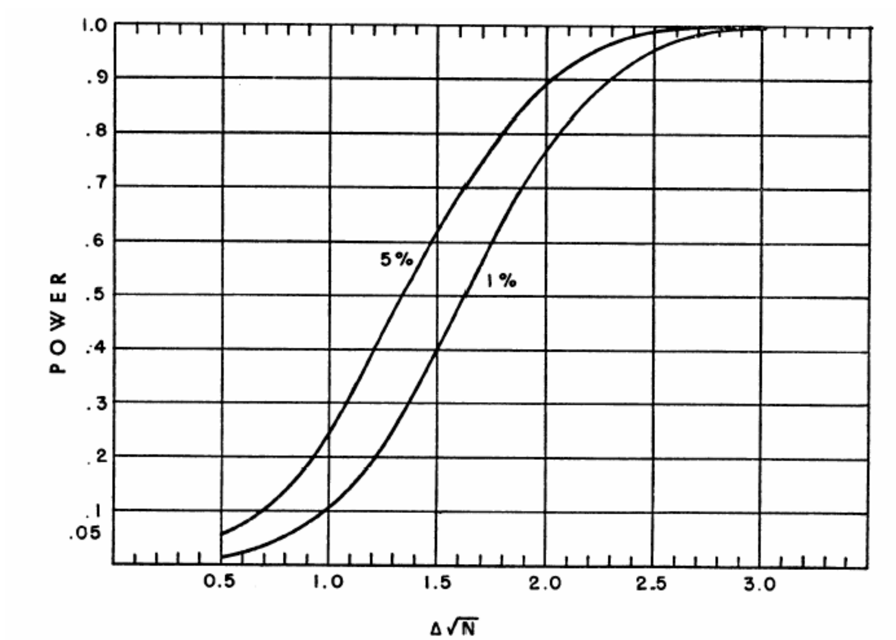


Figure 6. Lower bounds for the power of the d test for α=0.01 and α=0.05. Massey (1951).

In the NSQD, most of the data were from residential land uses. The Kolmogorov-Smirnov test was used to indicate if the cumulative empirical probability distribution of the residential stormwater constituents can be adequately represented with a lognormal distribution. Table 1 shows the resulting power of the test for Δ=0.05 and Δ=0.1, when applied to selected constituents that had very high levels of detection in residential land uses.

Table 1. Power of the test when applied to selected constituents in residential land uses

CONSTITUENT	**N**	**Percentage Detected**	**$\Delta N^{0.5}$ (Δ=0.05)**	**Power (Δ=0.05,α = 5%)**	**$\Delta N^{0.5}$ (Δ=0.1)**	**Power (Δ=0.1,α = 5%)**
TDS (mg/L)	861	99.2	1.46	0.60	2.92	1
TSS (mg/L)	991	98.6	1.56	0.65	3.12	1
BOD (mg/L)	941	97.6	1.52	0.65	3.04	1
COD (mg/L)	796	98.9	1.40	0.55	2.80	1
NO2+NO3 (mg/L)	927	97.4	1.50	0.60	3.00	1
TKN (mg/L)	957	96.8	1.52	0.65	3.04	1

Effective Modeling of Urban Water Systems Conference Proceedings, Monograph 13.
(edited by W. James, K.N. Irvine, E.A. McBean, and R.E. Pitt). CHI. Guelph, Ontario,
pp. 247 – 274. 2005.

TP (mg/L)	963	96.9	1.53	0.65	3.06	1
Total Copper (μg/L)	799	83.6	1.29	0.50	2.58	1
Total Lead (μg/L)	788	71.3	1.19	0.40	2.38	1
Total Zinc (μg/L)	810	96.4	1.40	0.55	2.80	1

Table 1 shows that the number of collected samples is sufficient to detect if the empirical distribution is located inside an interval of width 0.1 above and below the estimated cumulative probability distribution. If the interval is reduced to 0.05, the power varies between 40 and 65%. To estimate the interval width, 10 cumulative distributions of 1,000 random data points, having a lognormal (1, 1) distribution, were compared with the estimated cumulative distribution for normal, gamma and exponential distributions. The maximum distance between the cumulative lognormal and the cumulative normal distributions was 0.25. The maximum distance with cumulative gamma (the same for exponential in this case) was 0.28. An interval width of 0.1 was considered appropriate for the analysis.

Another factor that must be considered is the importance of relatively small errors in the selected distribution and the problems of a false negative determination. It may not be practical to collect as many data observations as needed when the distributions are close (such as when the width interval is 0.05). Therefore, it is important to understand what types of further statistical and analysis problems may be caused by having fewer samples than optimal. For example, Figure 7 (Total phosphorus in residential area) shows that most of the data fall along the straight line (indicating a lognormal fit), with fewer than 10 observations (out of 933) in the tails being outside of the obvious path of the line.

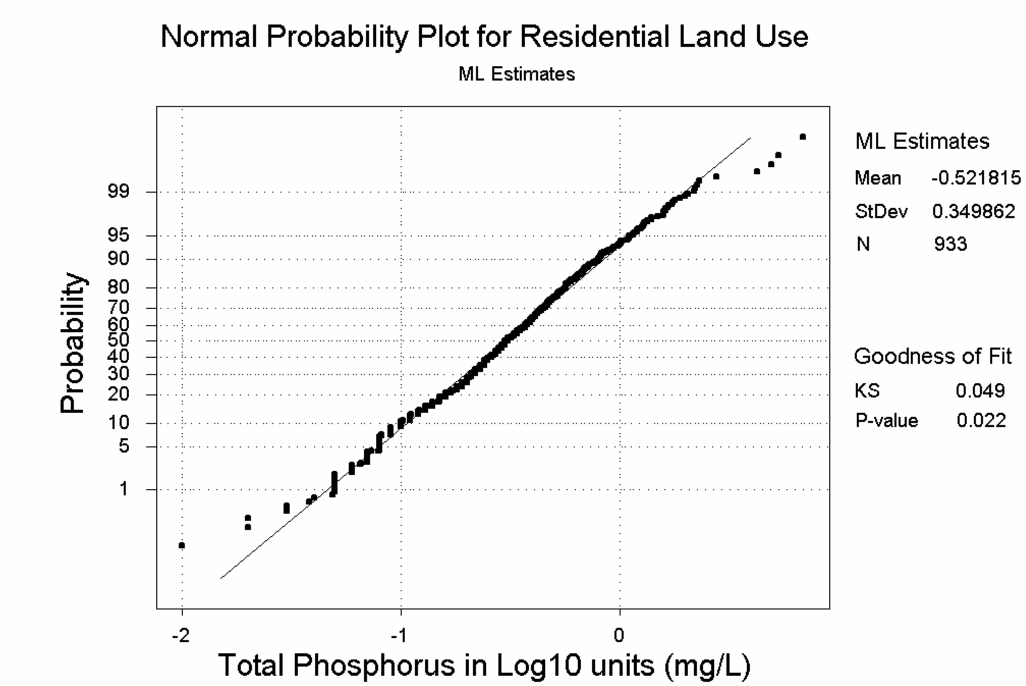


Figure 7. Normality test for Total Phosphorus in residential land uses using the NSQD

The calculated p-value for the Kolmogorov-Smirnov test is 0.022, indicating that the null hypothesis could be rejected and that there is not enough evidence that the empirical distribution is adequately represented by a lognormal distribution. Notice that the departures of any observations on the tails are smaller than 0.049. However, the tails are not responsible for the rejection of the null hypothesis (See Figure 8).

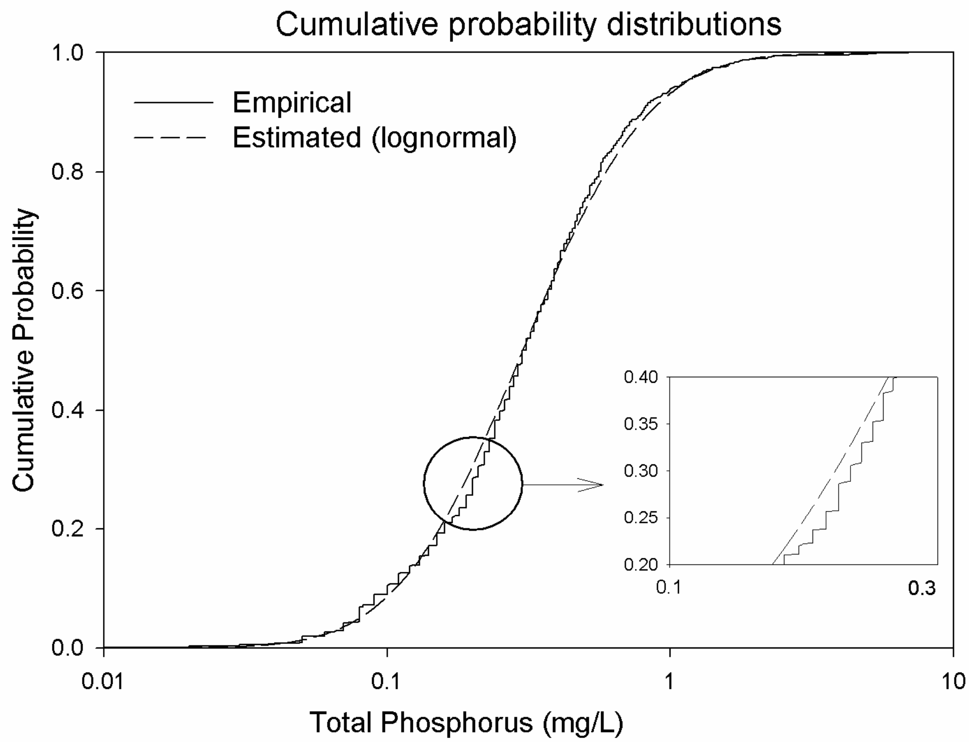


Figure 8. D_{max} was located in the middle of the distribution

In this case, D_{max} is located close to a total phosphorus concentration of 0.2 mg/L (-0.7 in log scale). As in this case, the hypothesized distributions are usually rejected because of the departures in the middle of the distribution, not in the tails. However, as previously pointed out, a small number of observations in the upper tail can change the shape of the estimated cumulative probability distribution by affecting the mean and standard deviation of the data.

The methods used previously by Van Buren and Beherra evaluated the probability distributions only using two parameters, the median and the standard deviation. They suggested the gamma and exponential distribution as alternatives to the lognormal for some stormwater constituents. Table 2 shows the comparison for the goodness of fit using the 2-parameter gamma, exponential and lognormal distributions using the method of moments.

Table 2. Comparison of goodness of fit for gamma, exponential and lognormal distributions using the NSQD v.1.1

CONSTITUENT	PDF	RESIDENTIAL			COMMERCIAL			INDUSTRIAL			OPEN SPACE			FREEWAYS		
		N_{Det}	Dmax	P-value	N_{Det}	Dmax	P-value	N_{Det}	Dmax	P-value	N_{Det}	Dmax	P-value	N_{Det}	Dmax	P-value
Conductivity (mS/cm)	Gamma	106	0.381	0	66	0.230	0.002	108	0.348	0	2	-	-	86	0.238	0
	Exponential	100%	0.195	0.001	100	0.237	0.001	100	0.228	0	100	-	-	100	0.232	0
	Lognormal		0.081	0.493		0.100	0.530		0.074	0.619		-	-		0.129	0.113
Hardness (mg/L)	Gamma	250	0.217	0	139	0.141	0.008	138	0.323	0	8	0.304	0.458	127	0.451	0
	Exponential	100%	0.203	0	100	0.115	0.067	96.4	0.133	0.018	100	0.369	0.228	100	0.161	0.003
	Lognormal		0.071	0.166		0.090	0.206		0.080	0.369		0.354	0.268		0.077	0.447
Oil and Grease (mg/L)	Gamma	533	0.876	0	308	0.629	0	327	0.939	0	19	0.210	1.080	60	0.103	0.810
	Exponential	57.8%	0.514	0	70.8	0.304	0	65.1	0.697	0	36.84	0.265	0.750	71.7	0.286	0.002
	Lognormal		0.112	0.001		0.103	0.019		0.098	0.032		0.202	1.127		0.101	0.827
Total Dissolved Solids (mg/L)	Gamma	861	0.234	0	399	0.457	0	413	0.645	0	45	0.109	0.698	97	0.082	0.553
	Exponential	99.3%	0.207	0	99.5	0.150	0	99.5	0.172	0	97.8	0.195	0.070	99	0.171	0.007
	Lognormal		0.050	0.029		0.049	0.303		0.066	0.053		0.120	0.561		0.054	1.136
Total Suspended Solids (mg/L)	Gamma	991	0.288	0	458	0.363	0	428	0.206	0	44	0.132	0.464	134	0.534	0
	Exponential	98.6%	0.141	0	98.3	0.214	0	99.1	0.108	0	95.5	0.289	0.002	99.3	0.168	0.011
	Lognormal		0.032	0.280		0.064	0.053		0.029	0.995		0.113	0.683		0.066	0.627
BOD5 (mg/L)	Gamma	941	0.321	0	432	0.191	0	406	0.921	0	44	0.112	0.770	26	0.272	0.076
	Exponential	97.6%	0.140	0	97.5	0.142	0	95.3	0.355	0	86.4	0.261	0.011	84.6	0.168	0.580
	Lognormal		0.058	0.004		0.054	0.166		0.105	0		0.114	0.746		0.103	1.252
COD (mg/L)	Gamma	796	0.129	0	373	0.137	0	362	0.216	0	43	0.373	0	67	0.163	0.061
	Exponential	98.9%	0.161	0	98.4	0.136	0	98.9	0.119	0	76.7	0.168	0.312	98.5	0.139	0.157
	Lognormal		0.036	0.250		0.038	0.695		0.074	0.040		0.128	0.684		0.107	0.445
Fecal Coliform (Colonies/100 mL)	Gamma	446	0.655	0	233	0.333	0	297	-	-	23	0.179	0.520	49	0.239	0.007
	Exponential	88.3%	0.374	0	88	0.396	0	87.9	0.504	0	91.3	0.208	0.324	100	0.355	0
	Lognormal		0.080	0.013		0.076	0.192		0.051	0.510		0.181	0.503		0.105	0.677
Fecal Streptococcus (Colonies/100 mL)	Gamma	305	0.158	0	181	0.354	0	195	-	-	22	0.144	0.869	25	0.096	1.262
	Exponential	89.5%	0.202	0	91.7	0.278	0	93.8	0.399	0	90.9	0.142	0.892	100	0.164	0.518
	Lognormal		0.077	0.081		0.097	0.091		0.083	0.161		0.181	0.538		0.119	0.990
Ammonia (mg/L)	Gamma	595	0.132	0	299	0.131	0	254	0.154	0	32	-	-	79	0.216	0.003
	Exponential	81.5%	0.101	0	83.3	0.066	0.228	85.8	0.071	0.221	18.7	-	-	87.3	0.105	0.440
	Lognormal		0.044	0.305		0.050	0.589		0.047	0.758		-	-		0.133	0.173
NO2+NO3 (mg/L)	Gamma	927	0.197	0	425	0.147	0	418	0.080	0.011	44	0.123	0.654	25	0.274	0.055
	Exponential	97.4%	0.141	0	98.1	0.120	0	96.2	0.132	0	84.1	0.120	0.686	96	0.177	0.443
	Lognormal		0.070	0		0.040	0.531		0.080	0.011		0.141	0.463		0.139	0.789
TKN (mg/L)	Gamma	957	0.203	0	449	0.127	0	440	0.195	0	45	0.169	0.323	125	0.280	0
	Exponential	96.8%	0.182	0	97.3	0.156	0	95.9	0.134	0	71.1	0.141	0.556	96.8	0.138	0.020
	Lognormal		0.035	0.218		0.042	0.423		0.048	0.292		0.147	0.500		0.074	0.539

* P-values greater than one are used only for comparison. NDet: Number of collected samples and percentage detected

Table 2. Comparison of goodness of fit for gamma, exponential and lognormal distributions using the NSQD v.1.1 (Continuation)

CONSTITUENT	PDF	RESIDENTIAL			COMMERCIAL			INDUSTRIAL			OPEN SPACE			FREEWAYS		
		N_{Det}	Dmax	P-value	N_{Det}	Dmax	P-value	N_{Det}	Dmax	P-value	N_{Det}	Dmax	P-value	N_{Det}	Dmax	P-value
Dissolved Phosphorus (mg/L)	Gamma	738	0.117	0	323	0.177	0	325	0.200	0	44	0.154	0.127	22	0.449	0
	Exponential	84.1	0.144	0	81.1	0.129	0	87.1	0.135	0	79.5	0.384	0.657	95.5	0.350	0.012
	Lognormal		0.043	0.199		0.075	0.104		0.124	0.682		0.124	0.682		0.170	0.593
Total Phosphorus (mg/L)	Gamma	963	0.184	0	446	0.179	0	434	0.227	0	46	0.666	0	128	0.456	0
	Exponential	96.9	0.129	0	95.7	0.114	0	96.3	0.107	0	84.8	0.320	0.001	99.2	0.187	0
	Lognormal		0.049	0.022		0.038	0.582		0.049	0.273		0.116	0.696		0.085	0.325
Total Antimony (µg/L)	Gamma	288	0.268	0.636	142	-	-	164	0.282	0.045	17	-	-	14	0.423	0.164
	Exponential	2.8	0.417	0.213	2.1	-	-	14.6	0.173	0.473	0	-	-	50	0.465	0.096
	Lognormal		0.233	0.841		-	-		0.096	1.279		-	-		0.419	0.171
Total Arsenic (µg/L)	Gamma	426	0.531	0	213	0.643	0	267	0.291	0	19	0.271	0.828	61	0.125	0.694
	Exponential	42	0.224	0	32.9	0.249	0	54.3	0.141	0.006	31.6	0.462	0.154	55.7	0.266	0.016
	Lognormal		0.154	0		0.164	0.046		0.129	0.016		0.273	0.819		0.149	0.441
Total Beryllium (µg/L)	Gamma	301	0.464	0	163	0.305	0.542	209	0.390	0.002	19	-	-	12	-	-
	Exponential	7.3	0.471	0	4.29	0.530	0.039	10.5	0.539	0	0	-	-	16.7	-	-
	Lognormal		0.200	0.342		0.205	1.108		0.163	0.620		-	-		-	-
Total Cadmium (µg/L)	Gamma	723	0.643	0	358	0.511	0	395	0.445	0	38	0.295	0.051	95	0.110	0.388
	Exponential	30.3	0.358	0	43	0.311	0	49.4	0.237	0	55.3	0.560	0	71.6	0.153	0.083
	Lognormal		0.120	0.004		0.113	0.039		0.083	0.136		0.206	0.338		0.052	1.380
Total Chromium (µg/L)	Gamma	435	0.292	0	235	0.151	0.004	256	0.122	0.008	36	0.252	0.386	76	0.058	1.208
	Exponential	55.4	0.132	0	58.7	0.201	0	72.7	0.067	0.381	36.1	0.272	0.290	98.7	0.176	0.019
	Lognormal		0.069	0.206		0.086	0.262		0.062	0.480		0.180	0.861		0.084	0.685
Total Copper (µg/L)	Gamma	799	0.394	0	387	0.296	0	416	0.408	0	39	0.107	0.226	97	0.451	0
	Exponential	83.6	0.149	0	92.8	0.137	0	89.9	0.177	0	74.4	0.127	0.092	99	0.231	0.090
	Lognormal		0.067	0.005		0.070	0.060		0.080	0.017		0.131	0.742		0.038	1.507
Total Lead (µg/L)	Gamma	788	0.300	0	377	0.297	0	412	0.276	0	45	0.177	0.608	107	0.203	0
	Exponential	71.3	0.173	0	85.4	0.136	0	76.5	0.225	0	42.2	0.389	0.006	100	0.125	0.072
	Lognormal		0.044	0.218		0.057	0.250		0.059	0.223		0.132	1.034		0.039	1.451
Total Nickel (µg/L)	Gamma	419	0.292	0	232	0.260	0	250	0.090	0.159	38	0.164	1.373	99	0.188	0.004
	Exponential	45.3	0.203	0	59.5	0.176	0	62.4	0.111	0.044	18.4	0.261	0.772	89.9	0.227	0
	Lognormal		0.081	0.160		0.056	0.831		0.065	0.525		0.166	1.360		0.091	0.460
Total Selenium (µg/L)	Gamma	318	0.263	0.095	169	0.169	0.952	203	0.434	0.022	19	-	-	16	-	-
	Exponential	6.9	0.254	0.117	7.7	0.174	0.907	5.9	0.256	0.416	21.1	-	-	6.3	-	-
	Lognormal		0.253	0.119		0.196	0.735		0.190	0.841		-	-		-	-
Total Silver (µg/L)	Gamma	406	0.421	0	222	0.143	0.718	287	0.263	0.002	19	-	-	21	-	-
	Exponential	12.6	0.333	0	11.3	0.159	0.563	17.4	0.340	0	5.3	-	-	19	-	-
	Lognormal		0.271	0.001		0.184	0.370		0.146	0.236		-	-		-	-
Total Zinc (µg/L)	Gamma	810	0.244	0	392	0.234	0	433	0.273	0	45	0.180	0.253	93	0.158	0.023
	Exponential	96.4	0.122	0	99	0.141	0	98.6	0.083	0.005	71.1	0.167	0.336	96.8	0.155	0.027
	Lognormal		0.054	0.020		0.040	0.585		0.044	0.389		0.105	0.981		0.063	0.985

* P-values greater than one are used only for comparison. NDet: Number of collected samples and percentage detected

Table 2 shows that for residential, commercial and industrial land uses, the lognormal distribution better fits the empirical data, except for selenium and silver in commercial land uses. In open space land uses, about 50% of the constituents were adequately fitted by the lognormal distribution, 30% by the gamma distribution and the remaining by the exponential distribution. In freeway areas, lognormal distributions better fit most of the constituents, except that fecal streptococcus, total arsenic and total chromium were better fitted by the gamma distribution and ammonia was better fitted by the exponential distribution. Also note in Table 2 that residential, commercial and industrial land uses had larger sample sizes than the other two land uses. It seems that for small sample sizes, gamma and exponential distributions better represent actual stormwater constituent distributions, but once the number of samples increases, the lognormal distribution is best. The few cases were the gamma distribution was a better fit was for NO_2+NO_3 in industrial land uses, and chromium in freeway areas. The exponential distribution better represents total ammonia in freeway areas (with around 70 detected samples) than the other two distribution types.

Other transformations were also tested, such as the square root, and other power functions, but the results were not improved. It was therefore decided to investigate if a three-parameter lognormal distribution function can be used to improve the overall goodness of fit for stormwater constituent probability distributions. As shown in the following section, this third parameter, in some cases, allows a much better fit of the cumulative empirical and estimated probability distributions.

4.0 Three Parameter Lognormal Calculations.

Goodness of fit was evaluated using three-parameter lognormal probability distribution. The probability distributions were created for residential, commercial, industrial, open space, and freeways land uses. The distribution parameters were calculated using the maximum likelihood and the L-moments methods. The maximum likelihood method requires that it be solved iteratively using three equations. The initial parameters were obtained using the method of moments. The results were compared with the two parameter standard model and the actual data. The model with the smaller maximum distance between the empirical and the estimated function was selected as the best model. All the calculations were made using only the detected values. The percentage of non-detected values was also calculated for each dataset.

In general, the L-moments method provided a better fit for the upper tail of the distribution whereas the maximum likelihood method provided a better fit for the lower tail. Figure 9 shows the three estimated models for TSS in industrial land uses.

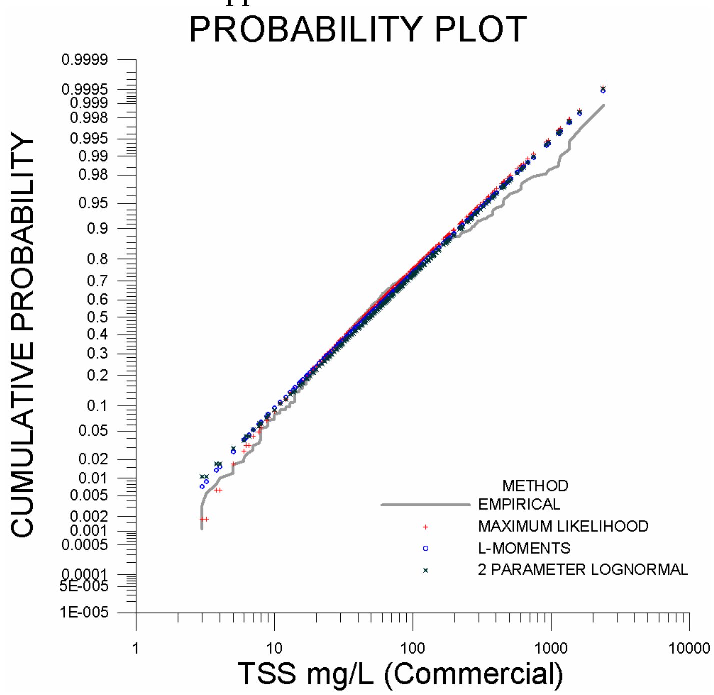


Figure 9. Estimated models for TSS in commercial land uses.

In this graph it is observed that the empirical distribution has higher values in the upper tail compared with any of the three models. In the lower tail, the maximum likelihood method using the three parameters better fit the observed values best. In this case the maximum likelihood method was better than the other two models, although none of the methods adequately represented the extreme high values.

The L-moments method generally betters fits the upper tail distribution, but typically trims or overestimate the lower tail. Figure 10 shows the results for TDS in industrial land uses. The L-moments better fits the empirical distribution in the upper tail, but it trims any observation smaller than 35 mg/L (almost 20 percent of the total dataset) in the lower tail. The 2 parameter lognormal and the maximum likelihood method provide better results although both were worse in the more critical upper tail region.

Effective Modeling of Urban Water Systems Conference Proceedings, Monograph 13. (edited by W. James, K.N. Irvine, E.A. McBean, and R.E. Pitt). CHI. Guelph, Ontario, pp. 247 – 274. 2005.

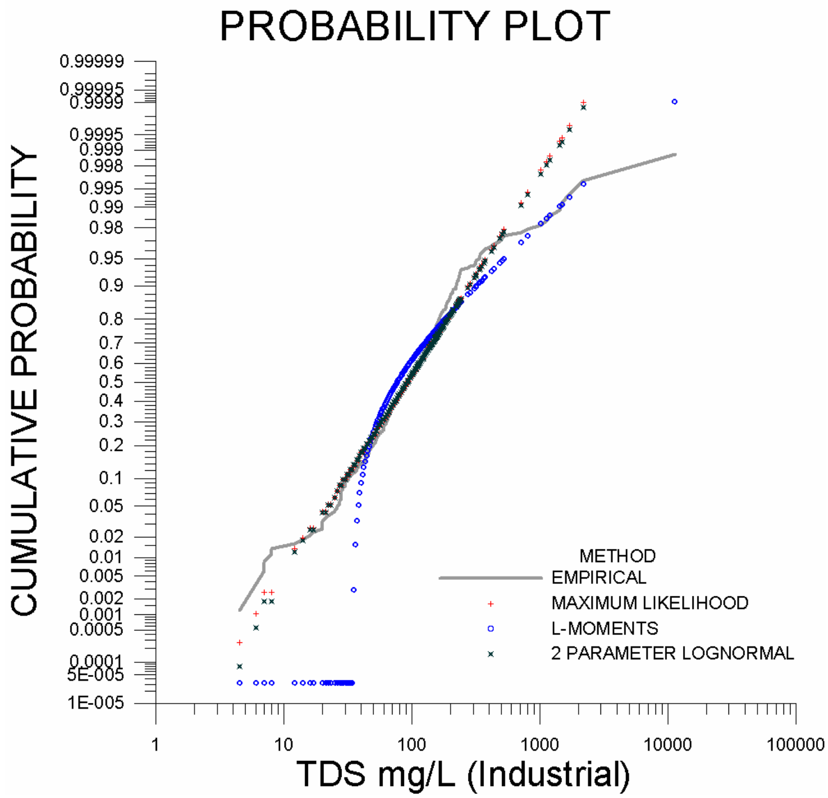


Figure 10. Estimated models for TDS in industrial land uses.

Table 3 presents the results for 15 constituents in five landuses. For each of the three methods, the p-value was calculated. The higher the p-value, the better is the fit between the empirical and the estimated function. Some of the p-values in the table are larger than one. When the number of samples is large, the p-value is calculated as a chi square distribution with 2 degrees of freedom. This probability is calculated only with one tail of the chi square distribution. The p-value is two times this probability. The maximum p-value is one, but for effects of comparison this presents two times the probability calculated from a one tail chi square distribution.

The maximum likelihood method with 3 parameters, and the lognormal 2-parameter distribution, produced the best descriptions for most of the constituents. For almost all constituents the function estimated by the L-moments method failed the lognormal assumption. Low p-values were obtained because the function was truncated and does not estimate the lower tail of the distribution.

It seems that when the numbers of samples increase, the L-moments method tends to truncate the function. The maximum likelihood method seems to improve the fit of the distribution, but when the number of samples is large, the cumulative estimated probability distribution is far from the cumulative empirical probability distribution, or no convergence is possible during the iteration process.

Table 3. Goodness of fit for different land uses. (p-value larger than one are used only for comparison)

RESIDENTIAL			Critical	Observed		2 Parameter Log		3 parameter maximum likelihood					3 parameter L-moments				
CONSTITUENT	**N**	**% Detected**	**$D_{0.05}$**	**μ**	**σ**	**D2**	**p-value**	**μ**	**σ**	**c**	**D3m**	**p-value**	**κ**	**α**	**ξ**	**D3l**	**p-value**
Conductivity (μS/cm)	106	100	0.132	4.638	0.710	0.081	0.493	4.327	0.919	20.767	0.052	1.133	-1.240	61.461	88.366	0.087	0.398
Hardness (mg/L)	250	100	0.086	3.497	0.706	0.071	0.166	3.539	0.675	-1.114	0.066	0.231	-	-	-	0.635	0
Oil & Grease (mg/L)	533	57.8	0.077	1.428	1.204	0.112	0.001	1.356	1.267	0.164	0.102	0.003	-2.559	2.065	2.482	0.208	0
Fecal Coliform (Col/100 ml)	446	88.3	0.069	8.7205	2.40448	0.080	0.013	8.734	2.367	-2.992	0.078	0.017	-1.929	17283.541	8236.423	0.096	0.001
Fecal Streptococcus (C/100 ml)	305	89.5	0.082	9.8344	1.88029	0.077	0.081	9.907	1.725	-210.924	0.066	0.190	-1.309	38433.290	24834.760	0.077	0.078
Ammonia (mg/L)	595	81.5	0.062	-1.1672	0.9166	0.044	0.305	0.015	0.220	-0.684	0.139	0.000	-	-	-	-	-
DP (mg/L)	738	84.1	0.055	-1.8303	0.85689	0.043	0.199	0.003	0.056	-0.877	0.327	0.000	-	-	-	-	-
Antimony (μg/L)	288	2.8	0.454	1.3554	1.71904	0.233	0.841	6.569	0.023	-700.000	0.368	0.228	-1.117	9.631	5.281	0.270	0.626
Arsenic (μg/L)	426	42	0.102	1.2098	0.85043	0.154	0.000	1.047	0.971	0.356	0.166	0.000	-1.588	1.981	2.632	0.208	0
Beryllium (μg/L)	301	7.3	0.281	0.0283	1.5958	0.200	0.342	-0.423	2.040	0.136	0.186	0.436	-2.063	1.154	0.675	0.227	0.207
Total Cadmium (μg/L)	723	30.3	0.092	-0.3532	1.21891	0.120	0.004	-0.444	1.301	0.033	0.110	0.010	-1.880	0.626	0.511	0.093	0.045
Total Chromium (μg/L)	435	55.4	0.088	1.5794	0.89199	0.069	0.206	1.473	0.980	0.328	0.067	0.236	-1.157	4.003	4.382	0.066	0.242
Total Nickel (μg/L)	419	45.3	0.099	1.7909	0.75669	0.081	0.160	1.601	0.890	0.768	0.083	0.147	-1.078	4.163	5.384	0.084	0.137
Total Selenium (μg/L)	318	6.9	0.281	1.0969	0.83323	0.253	0.119	0.479	1.348	0.876	0.259	0.061	-1.178	2.103	2.577	0.257	0.11
Total Silver (μg/L)	406	12.6	0.19	1.0686	1.3707	0.271	0.001	0.984	1.469	0.089	0.278	0.001	-1.522	3.390	2.767	0.294	0

COMMERCIAL			Observed			2 Parameter Log		3 parameter maximum likelihood					3 parameter L-moments				
CONSTITUENT	**N**	**% Detected**	**μ**	**σ**	**$D_{0.05}$**	**D2**	**p-value**	**μ**	**σ**	**c**	**D3m**	**p-value**	**κ**	**α**	**ξ**	**D3l**	**p-value**
Conductivity (μS/cm)	66	100	4.779	0.721	0.167	0.100	0.530	4.736	0.746	3.865	0.097	0.581	-1.011	76.386	108.055	0.093	0.633
Hardness (mg/L)	139	100	3.689	0.988	0.115	0.090	0.206	3.828	0.844	-3.808	0.063	0.653	-0.935	36.911	40.394	0.072	0.474
Oil & Grease (mg/L)	308	70.8	1.609	1.070	0.092	0.103	0.019	1.300	1.358	0.737	0.092	0.062	-1.853	3.648	3.638	0.126	0.002
TDS (mg/L)	399	99.5	4.332	0.791	0.068	0.049	0.303	4.393	0.741	-3.495	0.049	0.289	-1.066	55.106	70.423	0.071	0.035
TSS (mg/L)	458	98.3	3.883	1.180	0.064	0.064	0.053	3.735	1.218	1.988	0.042	0.416	-1.207	55.082	46.245	0.048	0.250
BOD (mg/L)	432	97.5	2.493	0.868	0.066	0.054	0.166	2.302	1.026	1.396	0.040	0.527	-1.002	10.321	11.447	0.044	0.380
COD (mg/L)	373	98.4	4.167	0.865	0.071	0.038	0.695	4.163	0.868	0.194	0.037	0.719	-0.911	54.903	63.127	0.034	0.860
Fecal Coliform (Col/100 ml)	233	88	8.202	2.380	0.095	0.076	0.192	8.191	2.398	1.870	0.077	0.175	-1.768	11370.330	5408.770	0.150	0.000
Fecal Streptococcus (C/100 ml)	181	91.7	8.940	2.061	0.106	0.097	0.091	8.936	2.061	2.494	0.096	0.093	-1.640	16795.600	9532.540	0.056	0.702
Ammonia (mg/L)	299	83.3	-0.706	1.083	0.086	0.050	0.589	-0.697	1.072	-0.002	0.048	0.632	-0.947	0.549	0.522	0.040	0.888
NO2+NO3 (mg/L)	425	98.1	-0.523	0.882	0.067	0.040	0.531	-0.432	0.800	-0.039	0.034	0.837	-0.849	0.510	0.600	0.030	0.954
TKN (mg/L)	449	97.3	0.471	0.828	0.065	0.042	0.423	0.575	0.734	-0.126	0.050	0.228	-0.866	1.246	1.571	0.032	0.816
DP (mg/L)	323	81.1	-2.077	1.016	0.084	0.075	0.104	-2.157	1.092	0.006	0.062	0.273	-1.077	0.127	0.121	0.059	0.315
TP (mg/L)	446	95.7	-1.473	0.881	0.066	0.038	0.582	-1.537	0.935	0.010	0.041	0.466	-0.991	0.196	0.220	0.049	0.264
Arsenic (μg/L)	213	32.9	0.9336	0.92361	0.163	0.164	0.046	0.729	1.098	0.301	0.195	0.010	-1.736	1.565	1.935	0.280	0
Total Cadmium (μg/L)	358	43	0.047	1.309	0.11	0.113	0.039	0.023	1.335	0.011	0.109	0.052	-1.640	1.180	0.879	0.063	0.591
Total Chromium (μg/L)	235	58.7	1.8134	0.71608	0.116	0.086	0.262	1.711	0.787	0.453	0.089	0.225	-0.734	4.462	6.100	0.083	0.295
Total Copper (μg/L)	387	92.8	2.8829	0.90117	0.072	0.070	0.060	2.807	0.965	0.874	0.063	0.117	-1.251	14.103	15.641	0.069	0.067
Total Lead (μg/L)	377	85.4	3.0328	1.03226	0.076	0.057	0.250	3.015	1.049	0.220	0.058	0.225	-1.251	19.633	18.805	0.053	0.329
Total Nickel (μg/L)	232	59.5	1.9782	0.8075	0.116	0.056	0.831	1.668	1.070	1.282	0.089	0.220	-0.969	5.691	6.842	0.076	0.406
Total Selenium (μg/L)	169	7.7	1.366	0.953	0.361	0.196	0.735	0.829	1.492	0.863	0.167	0.670	-0.803	3.940	3.621	0.210	0.638
Total Silver (μg/L)	222	11.3	0.9637	1.35108	0.272	0.184	0.370	1.080	1.174	-0.141	0.182	0.379	-0.911	3.587	3.133	0.165	0.513
Total Zinc (μg/L)	392	99	5.0388	0.84183	0.069	0.040	0.585	5.082	0.803	-4.834	0.039	0.619	-1.021	120.091	144.868	0.052	0.243

INDUSTRIAL			Observed			2 Parameter Log		3 parameter maximum likelihood					3 parameter L-moments				
CONSTITUENT	**N**	**% Detected**	**μ**	**σ**	$D_{0.05}$	**D2**	**p-value**	**μ**	**σ**	**c**	**D3m**	**p-value**	**κ**	**α**	**ξ**	**D3l**	**p-value**
Conductivity (μS/cm)	108	100	5.011	0.673	0.131	0.074	0.619	4.743	0.848	27.365	0.067	0.767	-1.197	83.673	129.715	0.106	0.174
Hardness (mg/L)	138	96.4	3.794	0.842	0.118	0.080	0.369	3.701	0.914	2.758	0.081	0.346	-1.272	31.221	38.285	0.119	0.047
Oil & Grease (mg/L)	327	65.1	1.623	1.153	0.093	0.098	0.032	1.456	1.298	0.447	0.093	0.048	-3.227	1.068	2.750	0.313	0.000
TDS (mg/L)	413	99.5	4.516	0.870	0.067	0.066	0.053	4.539	0.849	-1.484	0.065	0.064	-1.496	62.528	76.123	0.111	0.000
TSS (mg/L)	428	99.1	4.287	1.200	0.066	0.029	0.995	4.292	1.193	-0.169	0.028	1.023	-1.133	88.174	74.697	0.026	1.119
BOD (mg/L)	406	95.3	2.4121	0.992	0.069	0.105	0.000	2.303	1.085	0.729	0.095	0.002	-2.246	5.718	7.369	0.184	0.000
COD (mg/L)	362	98.9	4.2217	0.91441	0.072	0.074	0.040	4.237	0.899	-0.714	0.076	0.032	-1.096	57.774	63.159	0.046	0.437
Fecal Coliform (Col/100 ml)	297	87.9	7.6064	2.65965	0.084	0.051	0.510	7.574	2.732	1.560	0.055	0.417	-2.356	5638.993	2369.488	0.045	0.688
Fecal Streptococcus (C/100 ml)	195	93.8	9.1491	1.82906	0.101	0.083	0.161	9.190	1.741	-64.211	0.073	0.280	-2.253	11045.202	7378.455	0.199	0.000
Ammonia (mg/L)	254	85.8	-0.7071	1.00903	0.092	0.047	0.758	-0.685	0.985	-0.007	0.049	0.715	-0.864	0.518	0.524	0.046	0.789
NO2+NO3 (mg/L)	418	96.2	-0.3857	0.93116	0.068	0.080	0.011	-0.142	0.703	-0.132	0.043	0.454	-0.689	0.608	0.739	0.045	0.406
TKN (mg/L)	440	95.9	0.4238	0.88424	0.066	0.048	0.292	0.471	0.840	-0.050	0.050	0.239	-1.023	1.256	1.444	0.040	0.502
DP (mg/L)	325	87.1	-2.1766	0.87102	0.081	0.124	0.682	-2.141	0.837	-0.003	0.051	0.450	-1.002	0.093	0.108	0.063	0.211
TP (mg/L)	434	96.3	-1.2683	0.98202	0.067	0.049	0.273	-1.299	1.010	0.005	0.044	0.387	-1.068	0.271	0.271	0.035	0.724
Antimony (μg/L)	164	14.6	1.4793	1.01264	0.269	0.096	1.279	1.275	1.183	0.479	0.113	1.088	-1.334	3.747	3.661	0.150	0.684
Arsenic (μg/L)	267	54.3	1.5218	0.95205	0.113	0.129	0.016	1.19121.2628		0.752	0.128	0.018	-1.069	4.359	4.359	0.116	0.039
Beryllium (μg/L)	209	10.5	-0.3588	1.94765	0.281	0.163	0.620	-0.892	2.658	0.060	0.197	0.362	-2.074	1.346	0.568	0.231	0.191
Total Cadmium (μg/L)	395	49.4	0.7417	1.12552	0.097	0.083	0.136	0.588	1.276	0.161	0.095	0.060	-1.611	1.898	1.686	0.115	0.012
Total Chromium (μg/L)	256	72.7	2.5512	1.06906	0.1	0.062	0.480	2.599	1.015	-0.359	0.059	0.543	-0.911	13.859	13.657	0.050	0.803
Total Copper (μg/L)	416	89.9	3.2275	1.02866	0.07	0.080	0.017	3.179	1.076	0.716	0.073	0.030	-1.343	22.969	22.081	0.057	0.172
Total Lead (μg/L)	412	76.5	3.3651	1.32824	0.077	0.059	0.223	3.333	1.367	0.379	0.057	0.263	-1.374	38.025	27.991	0.054	0.316
Total Nickel (μg/L)	250	62.4	2.8058	0.97034	0.109	0.065	0.525	2.802	0.971	0.042	0.066	0.512	-0.772	17.094	17.834	0.088	0.182
Total Selenium (μg/L)	203	5.9	1.1472	1.27671	0.375	0.190	0.841	6.916	0.014	-1000.000	0.364	0.083	-1.851	2.381	2.025	0.202	0.753
Total Silver (μg/L)	287	17.4	0.113	1.80819	0.192	0.146	0.236	0.158	1.708	-0.010	0.153	0.194	-1.358	2.181	1.394	0.157	0.170
Total Zinc (μg/L)	433	98.6	5.305	0.970	0.066	0.044	0.389	5.359	0.915	-7.026	0.034	0.743	-0.899	198.198	208.408	0.030	0.951

OPEN SPACE			Observed			2 Parameter Log		3 parameter maximum likelihood					3 parameter L-moments				
CONSTITUENT	**N**	**% Detected**	**μ**	**σ**	$D_{0.05}$	**D2**	**p-value**	**μ**	**σ**	**c**	**D3m**	**p-value**	**κ**	**α**	**ξ**	**D3l**	**p-value**
TDS (mg/L)	45	97.8	4.762	0.744	0.205	0.120	0.561	4.480	0.962	20.400	0.115	0.621	-	-	-	0.759	0
TSS (mg/L)	44	95.5	3.945	1.717	0.21	0.113	0.683	3.672	2.096	2.777	0.095	0.942	-1.184	121.932	72.238	0.173	0.162
BOD (mg/L)	44	86.4	1.6211	0.66954	0.215	0.114	0.746	1.624	0.659	-0.013	0.115	0.733	-0.614	3.421	5.098	0.110	0.801
COD (mg/L)	43	76.7	3.548	0.785	0.231	0.128	0.684	3.323	0.946	5.000	0.151	0.441	-1.221	23.447	-16.159	0.184	0.215
Fecal Coliform (Col/100 ml)	23	91.3	8.9527	1.43209	0.287	0.181	0.503	8.725	1.692	534.506	0.217	0.278	-0.791	13914.660	10684.240	0.187	0.458
Fecal Streptococcus (C/100 ml)	22	90.9	9.6472	1.62819	0.294	0.181	0.538	9.847	1.248	-1070.380	0.165	0.676	-0.599	27514.897	24175.705	0.139	0.921
NO2+NO3 (mg/L)	44	84.1	-0.478	1.07398	0.234	0.141	0.463	-0.378	1.173	0.135	0.141	0.464	-0.579	0.746	0.759	0.122	0.664
TKN (mg/L)	45	71.1	-0.097	0.92495	0.234	0.147	0.500	-0.378	1.173	0.135	0.140	0.569	-0.716	0.918	0.976	0.160	0.389
DP (mg/L)	44	79.5	-2.1844	1.08407	0.224	0.124	0.682	-2.137	1.017	-0.003	0.131	0.606	-0.595	0.135	0.137	0.142	0.484
TP (mg/L)	46	84.8	-1.4264	1.12592	0.215	0.116	0.696	-1.499	1.188	0.009	0.126	0.584	-1.999	0.161	0.171	0.211	0.062
Total Cadmium (μg/L)	38	55.3	0.0291	2.62317	0.287	0.206	0.338	6.921	0.244	-1000.000	0.327	0.022	-1.561	7.207	2.856	0.319	0.028
Total Chromium (μg/L)	36	36.1	-1.79	1.66154	0.325	0.799	0.000	6.926	0.030	-1000.000	0.275	0.279	-1.446	10.225	5.944	0.193	0.756
Total Copper (μg/L)	39	74.4	2.1184	1.20044	0.246	0.131	0.742	6.926	0.035	-1000.000	0.334	0.003	-1.539	8.380	6.792	0.187	0.262
Total Lead (μg/L)	45	42.2	1.8882	1.95431	0.301	0.132	1.034	1.634	2.355	0.188	0.203	0.417	-1.406	16.381	8.735	0.174	0.632
Total Zinc (μg/L)	45	71.1	3.6048	1.19833	0.234	0.105	0.981	3.315	1.527	3.980	0.132	0.658	-1.142	44.515	36.591	0.113	0.885

FREEWAYS			Observed			2 Parameter Log		3 parameter maximum likelihood					3 parameter L-moments				
CONSTITUENT	**N**	**% Detected**	**μ**	**σ**	$D_{0.05}$	**D2**	**p-value**	**μ**	**σ**	**c**	**D3m**	**p-value**	**κ**	**α**	**ξ**	**D3l**	**p-value**
Conductivity (μS/cm)	86	100	4.586	0.681	0.147	0.129	0.113	4.404	0.795	12.765	0.113	0.226	-1.060	58.592	87.231	0.096	0.411
Hardness (mg/L)	127	100	3.604	0.791	0.121	0.077	0.447	3.485	0.875	3.045	0.077	0.440	-1.395	22.761	30.345	0.114	0.074
Oil & Grease (mg/L)	60	71.7	1.974	0.575	0.207	0.101	0.827	1.395	0.976	2.433	0.161	0.214	-0.543	4.299	7.228	0.098	0.881
TDS (mg/L)	97	99	4.279	0.771	0.139	0.054	1.136	4.342	0.720	-3.500	0.051	1.210	-0.660	56.217	74.574	0.044	1.368
TSS (mg/L)	134	99.3	4.464	1.068	0.118	0.066	0.627	4.513	1.010	-2.484	0.059	0.797	-1.447	76.753	75.337	0.120	0.043
BOD (mg/L)	26	84.6	2.262	0.894	0.281	0.103	1.252	2.013	1.084	1.385	0.099	1.304	-1.212	7.478	8.170	0.117	1.097
COD (mg/L)	67	98.5	4.552	0.980	0.167	0.107	0.445	4.798	0.715	-17.925	0.073	0.999	-0.881	81.577	97.057	0.082	0.817
Fecal Coliform (Col/100 ml)	49	100	7.585	1.716	0.194	0.105	0.677	7.514	1.811	30.900	0.128	0.398	-1.569	3543.87	2053.58	0.108	0.635
Fecal Streptococcus (C/100 ml)	25	100	9.263	1.662	0.264	0.119	0.990	8.854	2.363	540.500	0.211	0.215	-0.896	21116.59	15451.91	0.132	0.833
Ammonia (mg/L)	79	87.3	0.009	1.094	0.164	0.133	0.173	0.141	0.942	-0.083	0.106	0.423	-1.065	1.00	1.01	0.123	0.252
NO2+NO3 (mg/L)	25	96	-1.097	0.868	0.269	0.139	0.789	-1.628	1.301	0.087	0.114	1.073	-1.275	0.241	0.275	0.117	1.038
TKN (mg/L)	125	96.8	0.750	0.887	0.124	0.074	0.539	0.723	0.907	0.038	0.075	0.512	-1.159	1.669	1.907	0.071	0.598
DP (mg/L)	22	95.5	-1.226	1.188	0.287	0.170	0.593	-1.670	1.595	0.053	0.207	0.333	-1.992	0.181	0.190	0.255	0.131
TP (mg/L)	128	99.2	-1.266	0.772	0.121	0.085	0.325	-1.545	0.961	0.051	0.065	0.686	-1.394	0.171	0.231	0.089	0.27
Antimony (μg/L)	14	50	0.967	0.284	0.483	0.419	0.171	4.415	0.008	-80.000	0.397	0.220	0.660	0.563	2.922	0.270	0.719
Arsenic (μg/L)	61	55.7	0.952	0.666	0.227	0.149	0.441	0.290	1.180	0.886	0.106	0.937	-0.679	1.787	2.547	0.096	1.07
Total Cadmium (μg/L)	95	71.6	0.016	0.838	0.165	0.052	1.380	-0.108	0.937	0.081	0.067	1.087	-0.781	0.028	1.574	0.051	1.409
Total Chromium (μg/L)	76	98.7	2.096	0.734	0.157	0.084	0.685	2.165	0.680	-0.450	0.075	0.854	-0.555	6.127	8.575	0.055	1.272
Total Copper (μg/L)	97	99.0	3.525	0.842	0.139	0.038	1.507	3.433	0.915	2.047	0.048	1.295	-0.857	28.576	33.493	0.041	1.443
Total Lead (μg/L)	107	100	3.226	1.164	0.131	0.039	1.451	3.178	1.212	0.587	0.045	1.285	-1.155	28.993	24.965	0.040	1.424
Total Nickel (μg/L)	99	89.9	2.325	0.673	0.144	0.091	0.460	1.989	0.896	2.228	0.062	1.013	-0.960	6.390	9.308	0.073	0.769
Total Zinc (μg/L)	93	96.8	5.273	0.877	0.143	0.063	0.985	5.392	0.757	-17.226	0.054	1.000	-0.920	156.885	189.601	0.064	0.959

In commercial, industrial and freeways land uses, the numbers of samples available were between 100 and 500 samples. According to the prior discussion, this number of samples will result in an analysis having a power close or above 0.5. In these cases, most of the better fits were obtained using the L-moments method. In commercial and industrial land uses, more than half of the constituents also had the highest p-values when the L-moments method was used.

In open space areas, there were not many samples available. The small number of samples results in a low power. In this case, the higher p-values results were observed when the two parameter lognormal distribution was used. The use of the third parameter in constituents having small numbers of sample observations did not improve the fit of the estimated cumulative probability distribution. In the attached Appendix, the p-values for each land use and constituent are shown for the three methods.

5.0 CONCLUSIONS

Most of the stormwater constituents can be assumed to follow a lognormal distribution with little error. The use of the third parameter in the estimated lognormal distribution is recommended, depending on the number of samples. When the number of samples is large per category (approximately more than 400 samples) the maximum likelihood and the 2-parameter lognormal distribution better fit the empirical distribution. For large sample sizes, the L-moments method usually unacceptably truncates the distribution in the lower tail. However, when the sample size is more moderate per category (approximately between 100 and 400 samples), the 3-parameter lognormal method, estimated by L-moments, better fits the empirical distribution. When the sample size is small (<100 samples), the use of the third parameter does not improve the fit with the empirical distribution and the 2 parameter lognormal distribution produces a better fit than the other two methods.

Some constituents (such as TKN, TP, COD and Cu) show an increase in the p-value when the number of samples is acceptable and the 3-parameter lognormal probability distribution is used. The use of the lognormal distribution also has an advantage over the other distributions because it can be easily transformed to a normal distribution.

The few cases were gamma distribution seems to be a better model was for NO_2+NO_3 in industrial land uses, and chromium in freeway areas. The exponential distribution better fit total ammonia in freeway areas. The remaining constituents were well represented by the lognormal distribution.

Future studies will involve the development of multivariate and general linear models. Some of the requirements of these models are that the residuals have the same variance and that they are normally distributed. The use of lognormal transformations will facilitate the development of the predictive models.

Effective Modeling of Urban Water Systems Conference Proceedings, Monograph 13. (edited by W. James, K.N. Irvine, E.A. McBean, and R.E. Pitt). CHI. Guelph, Ontario, pp. 247 – 274. 2005.

6.0 REFERENCES

Behera, P. K. Li, J. L. Adams B. "Characterization of Urban Runoff Quality: A Toronto Case Study". In: *Applied Modeling of Urban Water Systems*, Vol. 8 (Edited by William James) CHI. Guelph, Ontario, 2000.

Gibbons, J. Chakraborti, S. *Nonparametric Statistical Inference.* Marcel Dekker Inc. New York, NY. Fourth Edition (2003) 645 pages.

Kottegoda N.T. and Rosso R. *Statistics, Probability and Reliability for Civil and Environmental Engineers.* WCB McGraw-Hill. New York, NY. 1998

Massey , F. J. Jr. "A Note on the Power of a Non-Parametric Test". *Annals of Mathematical Statistics*, 21 (1950) 440-443

Massey , F. J. Jr. "The Kolmogorov-Smirnov Test for Goodness of Fit." *Journal of the American Statistical Association*, 46, No 253 (1951) 67-78

Pitt, R. "The Incorporation of Urban Runoff Controls in the Wisconsin Priority Watershed Program." In: *Advanced Topics in Urban Runoff Research*, (Edited by B. Urbonas and L.A. Roesner). Engineering Foundation and ASCE, New York. pp. 290-313. 1986.

Pitt, R. and J. Voorhees. "Source loading and management model (SLAMM)." *Seminar Publication: National Conference on Urban Runoff Management: Enhancing Urban Watershed Management at the Local, County, and State Levels*. March 30 – April 2, 1993. Center for Environmental Research Information, U.S. Environmental Protection Agency. EPA/625/R-95/003. Cincinnati. Ohio. pp. 225-243. April 1995.

Pitt, R. Maestre A. Morquecho R. "Evaluation of NPDES Phase I Municipal Stormwater Monitoring Data" in: *National Conference on Urban Stormwater: Enhancing the Programs at the Local Level*. EPA/625/R-03/003. February 2003.

U.S. Environmental Protection Agency. *Results of the Nationwide Urban Runoff Program.* Water Planning Division, PB 84-185552, Washington, D.C. 1983

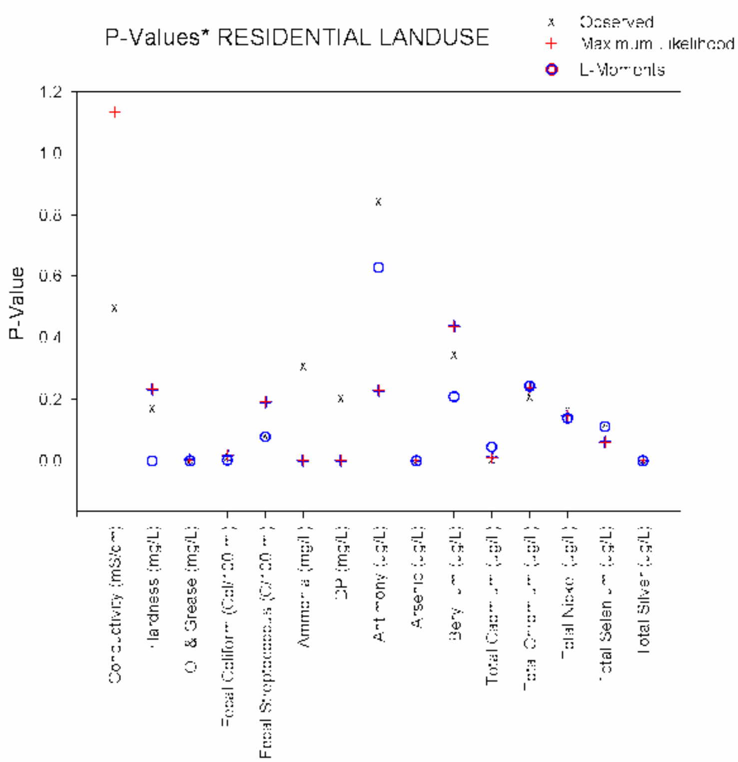
Van Buren M. A. Watt, W. E. Marsalek J. "Application of the Log-normal and Normal Distributions to Stormwater Quality Parameters". *Water Research* Vol. 31 No. 1 pp. 95-104, 1997.

Textbook: Statistics, Probability and Reliability for Civil and Environmental Engineers, by N.T. Kottegoda and R. Rosso (WCB McGraw-Hill 1997)
References: Probability, Statistics & Decision for Civil Engineers, Benjamin & Cornell, McGraw-Hill 1970
Probability and Statistics for Engineers & Scientists, Walpole, Myers and Myers, Prentice Hall, 6th Ed. 1998

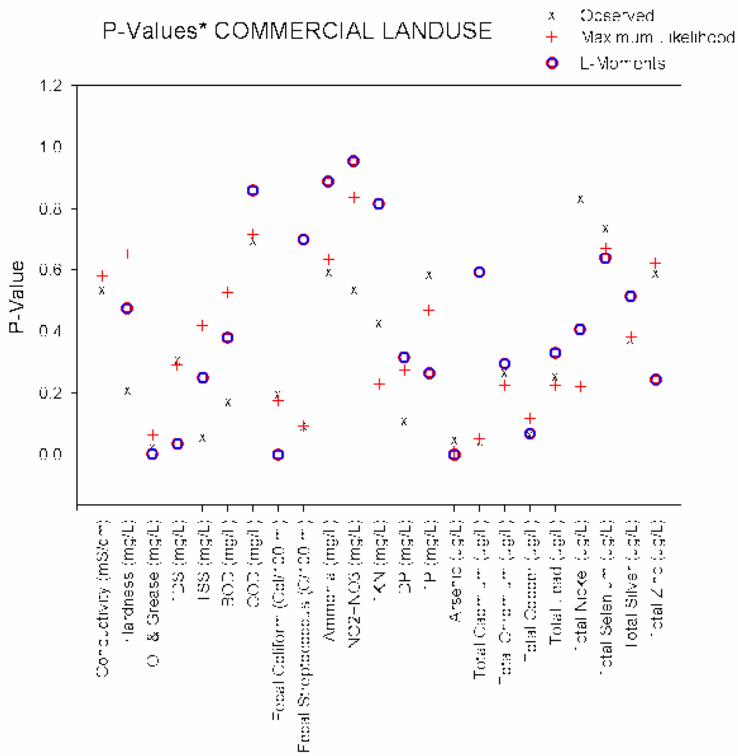
Effective Modeling of Urban Water Systems Conference Proceedings, Monograph 13.
(edited by W. James, K.N. Irvine, E.A. McBean, and R.E. Pitt). CHI. Guelph, Ontario,
pp. 247 – 274. 2005.

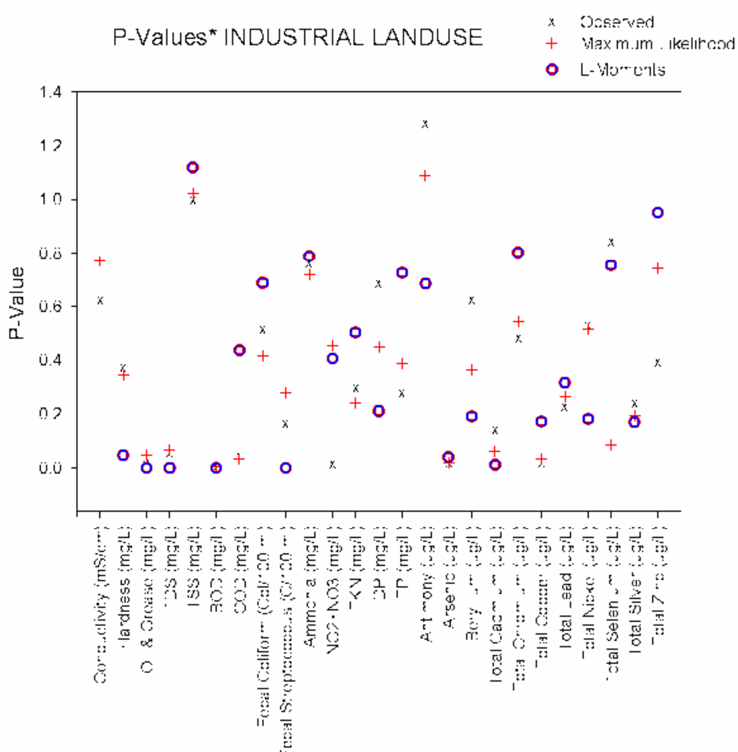
APPENDIX

Figure 11 shows the p-values obtained for each constituent and for each method. The labels are organized by the following groups: common constituents, nutrients and metals.



Effective Modeling of Urban Water Systems Conference Proceedings, Monograph 13.
(edited by W. James, K.N. Irvine, E.A. McBean, and R.E. Pitt). CHI. Guelph, Ontario,
pp. 247 – 274. 2005.





Effective Modeling of Urban Water Systems Conference Proceedings, Monograph 13.
(edited by W. James, K.N. Irvine, E.A. McBean, and R.E. Pitt). CHI. Guelph, Ontario,
pp. 247 – 274. 2005.