ESTIMATING THE MEAN AND VARIANCE OF CENSORED PHOSPHORUS CONCENTRATIONS IN FLORIDA RAINFALL

HOSUNG AHN

WRE-352

Made in United States of America
Reprinted from JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION
Vol. 34, No. 3, June 1998

JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION
VOL. 34, NO. 3 AMERICAN WATER RESOURCES ASSOCIATION JUNE 1998

ESTIMATING THE MEAN AND VARIANCE OF CENSORED PHOSPHORUS CONCENTRATIONS IN FLORIDA RAINFALL[1]

Hosung Ahn[2]

ABSTRACT: The total phosphorous (TP) concentrations in the South Florida rainfall have been recorded in weekly intervals with a detection limit (D_L) of 3.5 µg/L. As a large amount of the data is reported as below the D_L, appropriate statistical methods are needed for data analysis. Thus, an attempt was made to identify an appropriate method to estimate the mean and variance of the data. In particular, a method to separate the statistics for the below D_L portion from the estimated population statistics is proposed. The estimated statistics of the censored data are compared with the statistics of the uncensored data available from the recent years' laboratory records. It was found that the one-step restricted maximum likelihood method is the most accurate for the wet TP data, and that the proposed method to combine the estimated statistics for TP < D_L portion and the sample statistics for TP ≥ D_L portion improves estimates compared to the conventional maximum likelihood estimates.
(KEY TERMS: water quality; censored data; below detection limit; total phosphorus; wet atmospheric deposition; maximum likelihood estimator.)

INTRODUCTION

It has been recognized for the last several decades that the phosphorus level of aquatic systems is directly related not only to their eutrophication but also to the structure of the aquatic vegetation community. Therefore, the management of phosphorus input to the South Florida ecosystem has become an increasing concern resulting in the need for accurate monitoring and analyses of the phosphorus distribution in the region. In South Florida where most water bodies are large and shallow, atmospheric deposition can be one of the significant sources of phosphorus to the system (Chen and Fontaine, 1997). The importance is enhanced by the frequency and magnitude of rainfall in its subtropical climate.

In general, two forms of atmospheric deposition are commonly measured: wet deposition is that in the form of rain, while dry deposition occurs as dustfall under no-rain conditions. The atmospheric deposition data in the region have been collected by the South Florida Water Management District (District) since the early 1980s. The monitoring program was significantly improved in 1992 by deploying wet/dry collectors (Aerochem Metrics Model 301 automatic wet/dry samplers) and adopting a standard operating procedure for data collection and processing. From 19 monitoring sites, atmospheric deposition samples are taken in weekly intervals (every Tuesday) from both wet and dry collectors separately and analyzed at the District's laboratory to determine nutrients and major ions. The data are stored permanently in the District's database with the method detection limit (D_L) of 3.5 µg/L and the reporting accuracy of 1 µg/L.

As population means of the TP concentrations in rainfall approach the D_L and many TP samples have values below the D_L, it is necessary to use appropriate statistical methods to determine the statistics of the data. Thus, the main objective here is to identify an adequate method to estimate the mean and variance of the censored TP concentrations in South Florida's rainfall.

For censored data, earlier studies used maximum likelihood (ML) methods (Aitchison and Brown, 1969; Cohen, 1959; Newman *et al.*, 1989; and Schneider, 1986) which estimate the parameters for the entire distribution (both for the below detection limit (BDL) and above detection limit (ADL) portions simultaneously). Based on the above ML estimates, this paper proposes a method to compute the BDL statistics which are very useful for computing TP loads in

[1]Paper No. 97068 of the *Journal of the American Water Resources Association*. **Discussions are open until February 1, 1999.**
[2]Lead Hydrologist, Resource Assessment Division, WRE, South Florida Water Management District, 3301 Gun Club Road, West Palm Beach, Florida 33406 (E-Mail: hosung.ahn@sfwmd.gov).

which the censored TP concentration values are replaced by the estimated BDL mean. Also, the BDL statistics can be utilized to compute the combined statistics of both BDL and ADL portions as proposed in the following section. An application of the proposed method to compute the statistics for the wet TP deposition data measured in South Florida rainfall is presented in "Results and Discussion."

METHODS FOR SINGLY LEFT-CENSORED LOG-NORMAL DATA

The TP concentrations in atmospheric deposition sources are frequently less than the D_L, below which the data are not sufficiently reliable to report as numerical values. In general, the reason to use the less then D_L indicator in data is that sometimes the signal produced by a sample is too small to discriminate it from the background noise of the instrument. This condition is referred to as "below the detection limit ($< D_L$)", and the data containing less-than (or sometimes greater-than) indicators rather than their exact values are called "censored data."

The main idea of estimating the mean and variance of censored data is that, given a set of observed ADL data and the number of BDL samples, the mean and variance of the data (BDL+ADL) are estimated under the assumption that the probability distribution of the ADL portion is continuous and extendible to that of the BDL portion. Since the TP concentrations in rainfall are mostly positively skewed, lognormal distribution or logarithm (natural) transformation is desirable for data analyses. Common methods of estimating summary statistics for censored lognormal data can be divided into three classes: simple substitution, distributional method, and regression method (Schneider, 1986; Helsel and Hirsch, 1992). Simple substitution methods, which substitute each less than D_L observation with a single value such as $D_L/2$, have less theoretical basis and are apt to have bias in estimated statistics. Thus, some distributional and regression methods are discussed here in detail.

Distributional Methods

Distributional methods use the characteristics of an assumed distribution to estimate summary statistics such as mean and variance under the assumption that a set of data (both BDL and ADL portions) follows a given distribution and that summary statistics are computed by fitting a probability distribution function to the observed ADL data. The most popular maximum likelihood (ML) methods are described below.

ML Estimator of Aitchison and Brown (1969). Let us consider a random variable $X = \{x_i, i = 1,...,n\}$ with n observations. The variable X is assumed to be described adequately by a lognormal distribution $LN(\mu,\sigma^2)$ or equivalently by a normal distribution $N(\mu_y,\sigma_y^2)$ for a transformed variable $Y = \{y_i = \ln(x_i),\ i = 1,...,n\}$. It is further assumed that both X and Y are sorted in order of increasing magnitudes for mathematical convenience. With a single D_L value, X can be divided into two parts as in Figure 1: (m) censored observations $X_B = \{x_i, i = 1,...,m, \text{ for all } x_i < D_L\}$ whose exact values are unknown, and (n-m) uncensored observations $X_A = \{x_i, i = m + 1,...,n, \text{ for all } x_i \geq D_L\}$. If the probability density function (pdf) of Y is given by normal distribution such that

$$f(y) = \frac{1}{\sqrt{2\pi}\,\sigma_y}\exp\left[-\frac{\left(y_i-\mu_y\right)^2}{2\sigma_y^2}\right] \tag{1}$$

Then, the likelihood function of the pdf for the data censored at a log-transformed detection limit ($d = \ln(D_L)$) is given by (Aitchison and Brown, 1969; Cohen, 1959)

$$L = \binom{n}{m}F(\xi)^m \prod_{i=m+1}^{n}\frac{1}{\sqrt{2\pi}\,\sigma_y}\exp\left[-\frac{\left(y_i-\mu_y\right)^2}{2\sigma_y^2}\right] \tag{2}$$

with $p = F(\xi) = \int_{-\infty}^{\xi} f(t)dt$ and $\xi = (d - \mu_y)/\sigma_y$. Taking partial derivatives of the log-likelihood function with respect to μ_y and σ_y^2, respectively, and solving the two simultaneous equations, the estimates (^) of parameters $\hat{\mu}_y$ and $\hat{\sigma}_y^2$ in Equation (1) are obtained by (Aitchison and Brown, 1969)

$$\hat{\mu}_y = d - z\hat{\sigma}_y \tag{3}$$

$$\hat{\sigma}_y = \frac{g(h,z)}{n-m}\sum_{i=m+1}^{n}\left(y_i - d\right) \tag{4}$$

with h = m/n. The functions z and g(h,z) can be obtained from the tables in Hald (1949) and Pearson (1955). After computing $\hat{\mu}_y$ and $\hat{\sigma}_y^2$, the back-transformed mean and variance of original data are obtained by (Aitchison and Brown, 1969)

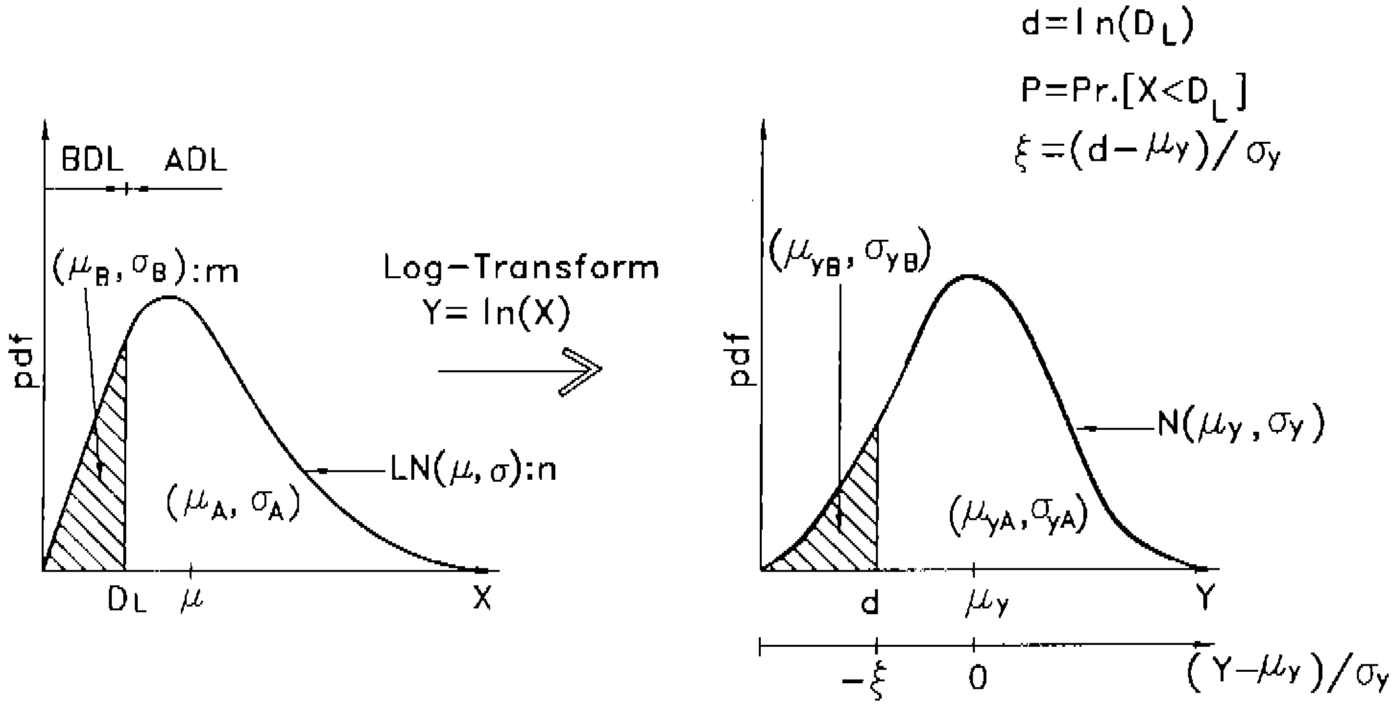


Figure 1. Schematic of a Censored Lognormal Distribution.

$$\hat{\mu} = e^{\hat{\mu}_y + \hat{\sigma}_y^2/2} \tag{5}$$

$$\hat{\sigma}^2 = \hat{\mu}^2 (e^{\hat{\sigma}_y^2} - 1) \tag{6}$$

ML Estimator by Cohen (1959). Defining $r = \hat{\sigma}_y^2/(\hat{\mu}_{yA} - d)$ where $\hat{\mu}_{yA}$ and $\hat{\sigma}_{yA}^2$ are the mean and variance of the log-transformed ADL samples (y_i, $i = m+1,...,n$), Cohen (1959,1961) simplified ML estimators using only one auxiliary function as

$$\hat{\mu}_y = \hat{\mu}_{yA} - \lambda(h,r)(\hat{\mu}_{yA} - d) \tag{7}$$

$$\hat{\sigma}_y^2 = \hat{\sigma}_{yA}^2 + \lambda(h,r)(\hat{\mu}_{yA} - d)^2 \tag{8}$$

The value of $\lambda(h,r)$ as a function of h and r is obtained from the tables in Cohen (1959, 1961), and Equations (5) and (6) are used for back-transformation.

Bias Corrected ML Estimator. For a small sample size ($n < 20$), Shaw (1961) and Schneider and Weissfeld (1986) provided a bias correction method for the previous ML estimates as

$$\hat{\mu}_{yc} = \hat{\mu}_y - \frac{\hat{\sigma}_y}{n+1} \exp\left[2.692 - 5.439\left(\frac{n-m}{n-2m+1}\right)\right] \tag{9}$$

$$\hat{\sigma}_{yc} = \hat{\sigma}_y - \frac{\hat{\sigma}_y}{n+1}\left[0.312 + 0.859\left(\frac{n-m}{n+1}\right)\right] \tag{10}$$

One-Step Restricted ML. The one-step restricted ML method (Persson and Rootzen, 1977) provides the following explicit solutions for the mean and variance by imposing an assumption that BDL samples follow a discrete binomial distribution:

$$\hat{\mu}_y = \hat{\mu}_{yA} - a\sigma^* \tag{11}$$

$$\hat{\sigma}_y^2 = \hat{\sigma}_{yA}^2 - (a\varepsilon - a^2)(\sigma^*)^2 \tag{12}$$

where $a = nF(\varepsilon)/m$, $F(.)$ is the standardized normal pdf, $\sigma^* = 1/2[C^2 + 4\,\hat{\sigma}_{yA}^2 + 4\,(\hat{\mu}_{yA} - d)^2]^{1/2}$, $C = \varepsilon(\hat{\mu}_{yA} - d)$, and $\varepsilon = (d - \hat{\mu}_{yA})/\hat{\sigma}_{yA}^2$.

BDL Statistics

The mean and variance computed in the previous subsection are those for overall data, that is, for both the BDL and ADL portions. However, one may be interested in estimating the statistics of only the BDL portion. For instance, TP loads can be computed accurately from the censored TP concentrations by replacing every censored value with the estimated BDL mean. Also, the mean and variance of a censored data set can be improved by combining the sample statistics for ADL portion and the estimated BDL statistics as will be shown latter.

Once a pdf f(x) of censored data is defined by estimating the parameters μ_x and σ_x^2 using one of the ML methods, the BDL probability p at D_L can be defined and estimated by

$$p = \Pr.\left[X < D_L|\hat{\mu}_x, \hat{\sigma}_x^2\right] = \int_0^{D_L} f(x)dx = \Delta x \sum_{i=1}^{n_1} f(x_i) \quad (13)$$

where Δx is the discrete interval that should be small enough to ensure the accuracy of the numerical summation, and n_1 is the number of effective discretizations of the BDL portion. Using the definition of probability weighted moments (Mood *et al.*, 1974), the BDL mean and standard deviation are computed, respectively, by

$$\hat{\mu}_B = \frac{1}{p}\int_0^{D_L} f(x)xdx = \frac{\Delta x}{p}\sum_{i=1}^{n_1} f(x_i)x_i \quad (14)$$

$$\hat{\sigma}_B^2 = \frac{1}{p}\int_0^{D_L} f(x)(x-\hat{\mu}_B)^2 dx = \frac{\Delta x}{p}\sum_{i=1}^{n_1} f(x_i - \hat{\mu}_B)^2 \quad (15)$$

The analytical solutions of the above three definite integral equations are not known to the author; however, they can be solved efficiently and accurately with the discrete summation schemes shown as the right-hand-sides of Equations (13) through (15). Then, the combined (compound) mean and variance are computed by (Kite, 1988)

$$\hat{\mu} = h\hat{\mu}_B + (1-h)\hat{\mu}_A \quad (16)$$

$$\hat{\sigma}^2 = h\hat{\sigma}_B^2 + (1-h)\hat{\sigma}_A^2 + h(1-h)(\hat{\mu}_B - \hat{\mu}_A)^2 \quad (17)$$

where $\hat{\mu}_y$ and $\hat{\sigma}_y^2$ are one of ML estimates, and the ADL mean and variance are computed by $\hat{\mu}_A = \Sigma_{i=m+1}^{n} x_i/(n-m)$ and $\hat{\sigma}_A^2 = \Sigma_{i=m+1}^{n} [x_i - \hat{\mu}_A]^2/(n-m-1)$, respectively.

Linear Regression Method

If singly censored data after log-transformation follow a normal distribution, it is possible to estimate x_is for $x_i < D_L$ based on the linear relationship of log-transformed ADL values versus the normal scores designated by the plotting positions of the ordered data (Gilliom and Helsel, 1986; Helsel and Cohn, 1988; Helsel and Gilliom, 1986; Helsel, 1990; Helsel and Hirsch, 1992). The normal score (z-score) z_i is an inverse of the standardized cumulative distribution function (cdf) of a plotting position p_i as

$$z_i = \Phi^{-1}(p_i) \quad (18)$$

With the data sorted in increasing order of magnitude, the general expression for the plotting position of the i-th rank (i = m+1,...,n) can be expressed by (Cunnane, 1978; Hirsch and Stedinger, 1987)

$$p_i = \frac{i-\omega}{n+1-\omega} \quad (19)$$

where ϖ is used to correct bias in the extreme (largest and smallest) observations. Depending on ϖ, Equation (19) is called Weibull (ϖ = 0), Blom (ϖ = 3/8), or Hazen (ϖ = 0.5) plotting position. The Blom estimate is selected here since Newman *et al.* (1989, 1995) found that it is the best choice for censored data.

Assuming a simple linear equation adequately represents the relationship between the log-transformed data and the corresponding z-score as

$$y_i = \ln(x_i) = \alpha + \beta z_i, \qquad \text{with } i = 1,...,n \quad (20)$$

the parameters α and β are estimated by regression analysis with observed ADL data. Then, the BDL portion values $\{\hat{x}_i, i = 1,...,m\}$ are estimated by Equation (20) with the z-scores assigned by an appropriate plotting position method. The combined mean and variance of the BDL and ADL portions in this method are obtained, respectively, by

$$\hat{\mu} = \frac{1}{n}\left(\sum_{i=1}^{m}\hat{x}_i + \sum_{i=m+1}^{n} x_i\right) \quad (21)$$

$$\hat{\sigma}^2 = \frac{1}{n-1}\left[\sum_{i=1}^{m}\left(\hat{x}_i - \hat{\mu}\right)^2 + \sum_{i=m+1}^{n}\left(x_i - \hat{\mu}\right)^2\right] \quad (22)$$

It is also possible to estimate the parameters α and β by either Kendall's robust line, tobit regression, or logistic regression (refer to Helsel and Hirsch, 1992), but this study adopted a simple regression method which gives enough accuracy as shown later.

Since the z-score in Equation (20) is the standardized normal variate of log-transformed value, that is, $z_i = [\ln(x_i) - \mu_y]/\sigma_y$, the mean and variance of the censored data may be obtained directly from the relationship of $\ln(x_i) = \mu_y + \sigma_y z_i$ (Gleit, 1985). However, this method was not applied here since it gives more biased results than the regression method by Equations (18) through (22).

RESULTS AND DISCUSSION

Preliminary Statistics of Uncensored TP Data

Among 19 atmospheric deposition monitoring sites, only two sites, namely BG1WET and BG2WET, were selected in this study since the data from the other sites have severe contamination problems which will be addressed separately. Both sites are located in the southeastern marsh of Lake Okeechobee (Latitude 26°43′, Longitude 80°43′), approximately one kilometer east of the agricultural fields. In particular, the wet TP deposition data from 1993 to 1996 were taken for the censored data analysis here because, during the period of record, the uncensored TP data including values for TP < 3.5 μg/L are available from laboratory records. Also available are the uncensored quality control data of atmospheric deposition including equipment blanks, split samples, and replicate samples. Equipment blanks are collected, before the collection of routine samples, at a rate of one in 20 samples (five percent), by passing one liter of deionized water through each piece of sampling equipment. Replicate samples are those collected at the same time and place as the routine samples, while split samples are aliquots (parts) of the same samples used to measure the variability of sampling and the laboratory analyses.

Table 1 summarizes the sample statistics of overall and BDL data, where "overall" means both ADL and BDL portions. As shown in this table and Figure 2, the TP concentrations from each site are highly skewed and thus transformation of the data is needed before the censored data analysis. When the data are log-transformed, the skewness and kurtosis in each site are improved significantly as shown in Table 1. The result of a normality test proposed by Snedecor and Cochran (1980) showed that the kurtosis at each site satisfies the normality assumption, but the skewness at each site is slightly higher than the tabulated value (0.567 for n = 100 and 1 percent significant level). However, the skewness effect to the censored data analysis is diminished when the proposed combined statistics by Equations (16) and (17) are used and thus the log-transformation of the original data is still valid as shown later.

Table 2 presents the sample statistics of the BDL data taken from several different sources collected during 1996 by the District, and Figure 3 displays the frequency distributions of the same BDL data where the negative TP concentrations are the result of instrument noise. These presentations demonstrate that the summary statistics of the BDL wet TP data from different sources are quite consistent. The distribution in the BDL portion has monotonically decreasing TP values so it is deemed to be a left tail of a lognormal distribution. For further comparison, the statistics of 1996-wet-TP-samples (the second column in Table 2) were selected as the typical BDL statistics due to its large sample size.

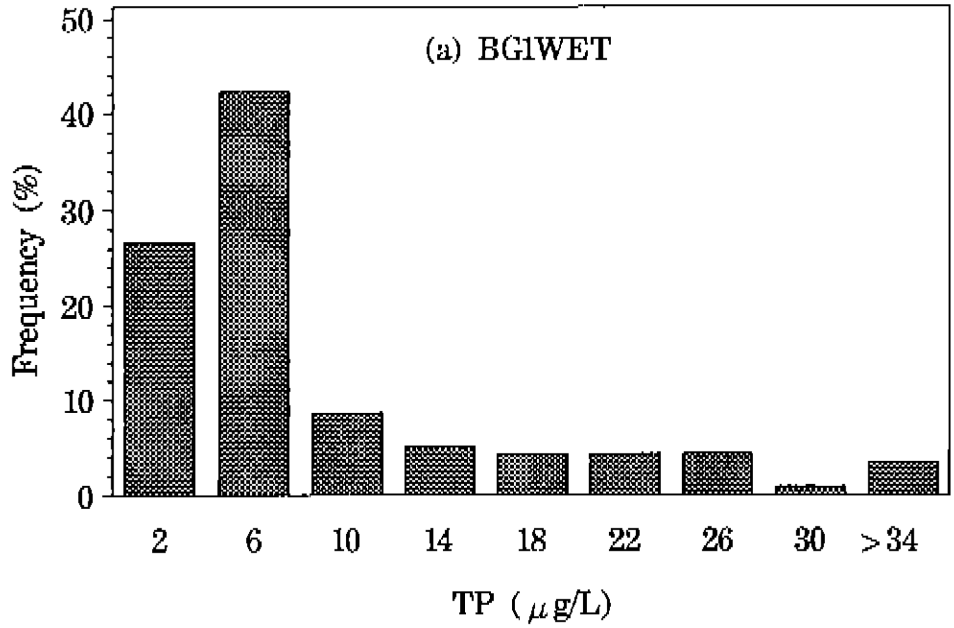
Comparison of Estimators

Table 3 summarizes the means and standard deviations estimated by four ML methods and one regression based method performed in three different ways. The "overall estimated" in Case A represents the statistics of both BDL and ADL portions as outlined by the conventional ML methods. In particular, the

TABLE 1. Sample Statistics of the Uncensored Wet TP Concentrations in South Florida's Rainfall.

Statistics	Overall (ADL+BDL) Data		Data for TP<D_L (3.5 μg/L)	
	BG1WET	BG2WET	BG1WET	BG2WET
Number of Samples	116	109	31 (27%)	25 (23%)
Mean (μg/L)	9.732	13.664	2.074	1.957
S.D. (μg/L)	12.034	20.247	0.917	1.022
Skewness	3.179	3.791	-0.478	-0.467
Skewness for Y = lnX	0.803	0.709	-0.900	-0.895
Kurtosis	12.178	17.22	-0.917	-0.988
Kurtosis for Y = lnX	0.577	0.482	0.343	0.039

*Percent censoring, h = 100 m/n.



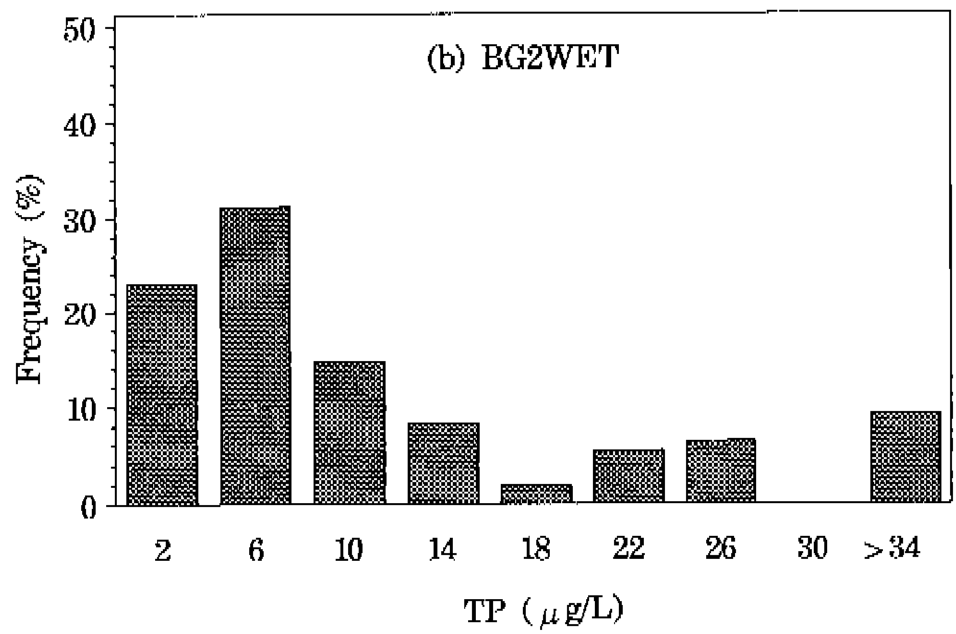


Figure 2. Frequency Distributions of Wet TP Concentration Data at (a) BG1WET (n = 116) and (b) BG2WET (n = 109) Sites.

regression method in Case A computes the sample statistics (moment estimates) of the estimated TP values (both BDL and ADL portions) using the fitted regression equation $x_i = \exp(\alpha + \beta z_i)$. It was observed that a simple linear regression model fits well for ln(TP) versus z-score: R^2s for BG1WET and BG2WET sites are 0.962 and 0.979, respectively. The ML estimates of BDL statistics (for Case B) were obtained by Equations (13), (14), and (15) with the estimated lognormal pdfs listed in Case A, while the regression method in this case computes the sample statistics (moment estimates) of the estimated (filled-in) BDL values $\{x_1,...,x_m\}$. For Case C, Equations (16) and (17) are used to get the combined statistics.

The result in Table 3 indicates that, for the wet TP concentration data, the ML methods outperform the regression method. This differ from the simulation result by Gilliom and Helsel (1986) where they concluded that the regression method is the most robust. To draw a general conclusion of which method

TABLE 2. Statistics of the Uncensored Wet TP Samples for TP < 3.5 μg/L (TP unit in μg/L).

Statistics	1996 Wet TP Samples	Equipment Blank	Replicate Sample	Split Sample
Number of Samples	117	22	13	17
Mean	2.081	2.000	1.833	2.125
Standard Deviation	1.054	1.000	1.030	1.147
Median	2.0	2.0	2.0	2.0
Geometric Mean	2.232	2.075	1.802	2.568
Skewness	-0.828	-0.656	-0.211	-1.182
Pr.[TP < 0 μg/L]* (%)	4.3	2.8	7.7	5.9

*The probability that the selected random BDL samples are less than 0 μg/L due to instrument error (ratio of less than 0 samples to total BDL samples).

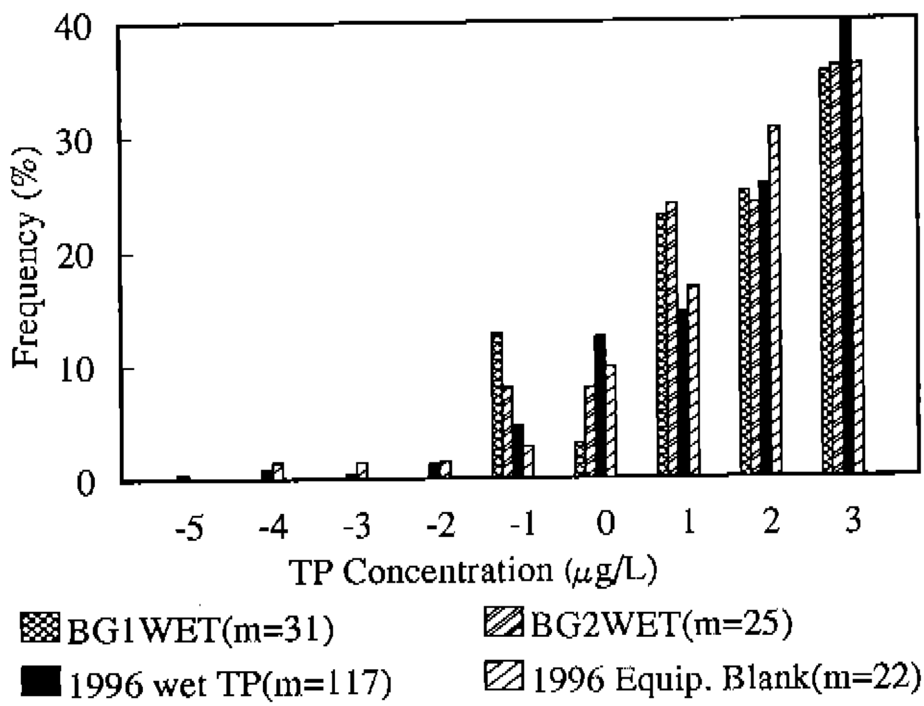


Figure 3. Frequency Distributions of BDL Wet TP Concentration Data From Different Sources, Where m is the BDL Sample Size.

is good for the proposed BDL estimate, comprehensive simulation experiments should be done which is out of the scope here. When the estimated BDL statistics and the ADL sample statistics are combined to obtain the overall statistics, the estimations are improved for four ML methods as well as the regression method. There is no significant difference between the tested ML methods, but the one-step restricted method slightly outperforms the other ML estimates.

In addition, Tables 3 and 4 include the combined statistics estimated by Equations (16) and (17) with replacing the BDL statistics with two simple pdfs: the typical pdf obtained from the 1996-wet-TP-samples (from Table 2) and a uniform distribution. The combined statistics for the typical distribution case are quite comparable to the ML estimates as shown in both Case C in Table 3 and Table 4. The combined standard deviations with a uniform distribution assumption (the last raw in Table 4) seem to be the most accurate, but this result is not defensible since the estimate by Equation (17) was influenced by the underestimated BDL mean.

As shown in Case B in Table 3 and Figure 4, all methods underestimate mean and standard deviation

TABLE 3. Estimated Mean and Standard Deviation of Censored Wet TP Concentration (μg/L).

Estimation Method	BG1WET		BG2WET	
	Mean	S.D.	Mean	S.D.
Case A: Overall Estimated				
Observed	9.732	12.034	13.664	20.247
ML by Aitchison and Brown	9.204	11.215	13.061	19.592
ML by Cohen, Equations (7) and (8)	9.236	11.360	13.115	19.845
Bias Corrected ML, Equations (9) and (10)	9.236	11.470	13.116	20.038
One-Step Restricted ML, Equations (11) and (12)	9.339	11.856	13.262	20.553
Regression Method	8.963	9.751	12.650	15.583
Case B: Estimated for $x_i < D_L$				
Observed	2.074	0.917	1.957	1.022
ML by Aitchison and Brown	2.118	0.868	2.038	0.895
ML by Cohen	2.110	0.874	2.030	0.900
Bias Corrected ML	2.101	0.882	2.021	0.909
One-Step Restricted ML	2.084	0.892	2.010	0.914
Regression Method	1.673	0.702	1.644	0.729
1996 Wet TP Samples*	2.001	1.058	2.001	1.058
Uniform Distribution	1.750	1.021	1.750	1.021
Case C: Statistics for Combined Population of the BDL (Case B) and ADL Portions**				
Observed	9.732	12.034	13.664	20.247
ML by Aitchison and Brown	9.480	11.986	13.467	20.148
ML by Cohen	9.478	11.988	13.466	20.150
Bias Corrected ML	9.475	11.991	13.464	20.152
One-Step Restricted ML	9.471	11.997	13.461	20.154
Regression Method	9.361	11.971	13.377	20.160
1996 Wet TP Samples*	9.470	11.919	13.477	20.122
Uniform Distribution	9.382	11.973	13.344	20.164

*The ADL statistics (mean and SD) for BG1WET and BG2WET are (12.165, 12.895) and (16.869, 21.792) in μg/L, respectively.
**The typical BDL statistics based on the 1996 wet TP samples (from the first column in Table 2).

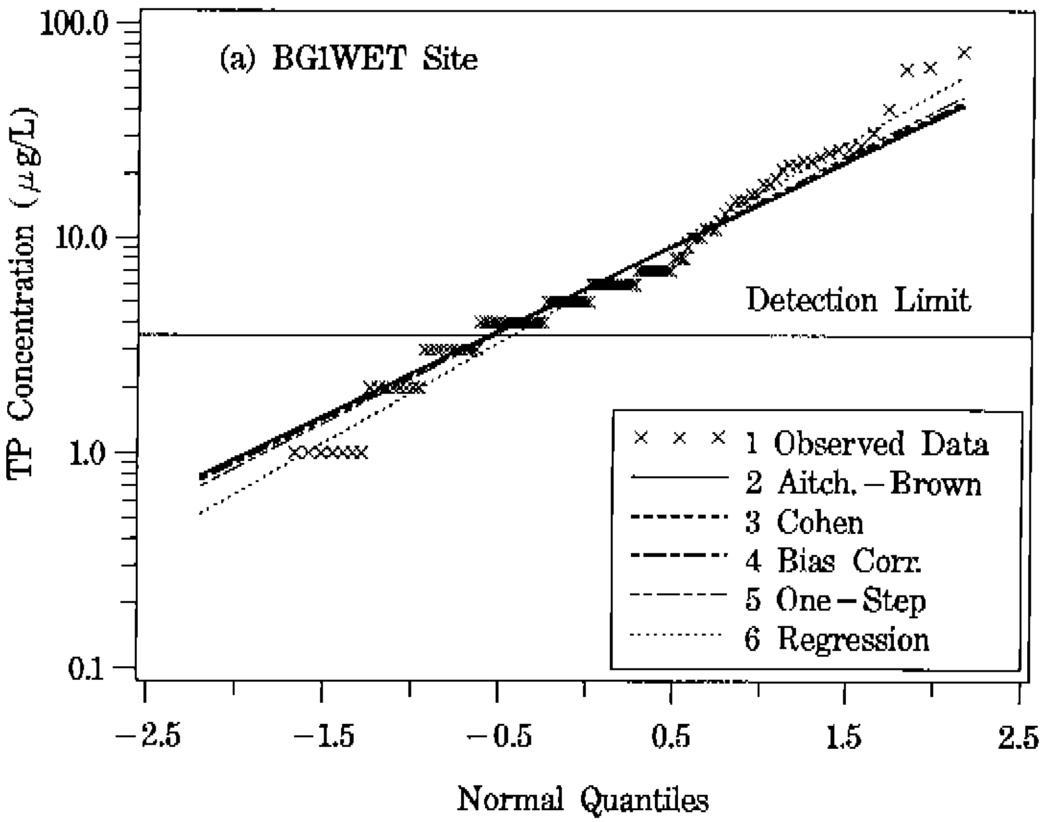
TABLE 4. Comparison of Estimation Errors (=100 | Obs.-Est. | /Obs.) for the BDL Estimates.

Estimation Method	Error (percent)	
	Mean	S.D.
ML (One-Step Restricted)	1.60	6.65
Regression Method	17.66	26.07
1996 Wet TP Samples*	2.89	9.45
Uniform Distributions*	13.10	5.72

*It is assumed that the distribution of the BDL portion follows either the 1996 wet TP sample distribution or a uniform distribution.

compared to the observed statistics of the uncensored data except the standard deviation at BG2WET estimated by the one-step restricted method. The bias in the BDL estimates can also be checked by computing the BDL probability p by Equation (13) and comparing it with h. When the one-step restricted method is used, the estimated p values at BG1WET and BG2WET sites are 0.305 and 0.261, respectively, while the corresponding h values are 0.267 and 0.229, respectively. This difference in each site is the result of overestimation of the left-tail of the fitted pdf compared to the sample distribution.

Comparing the Cases A and C in Table 3, it can be concluded that the estimates can be improved by



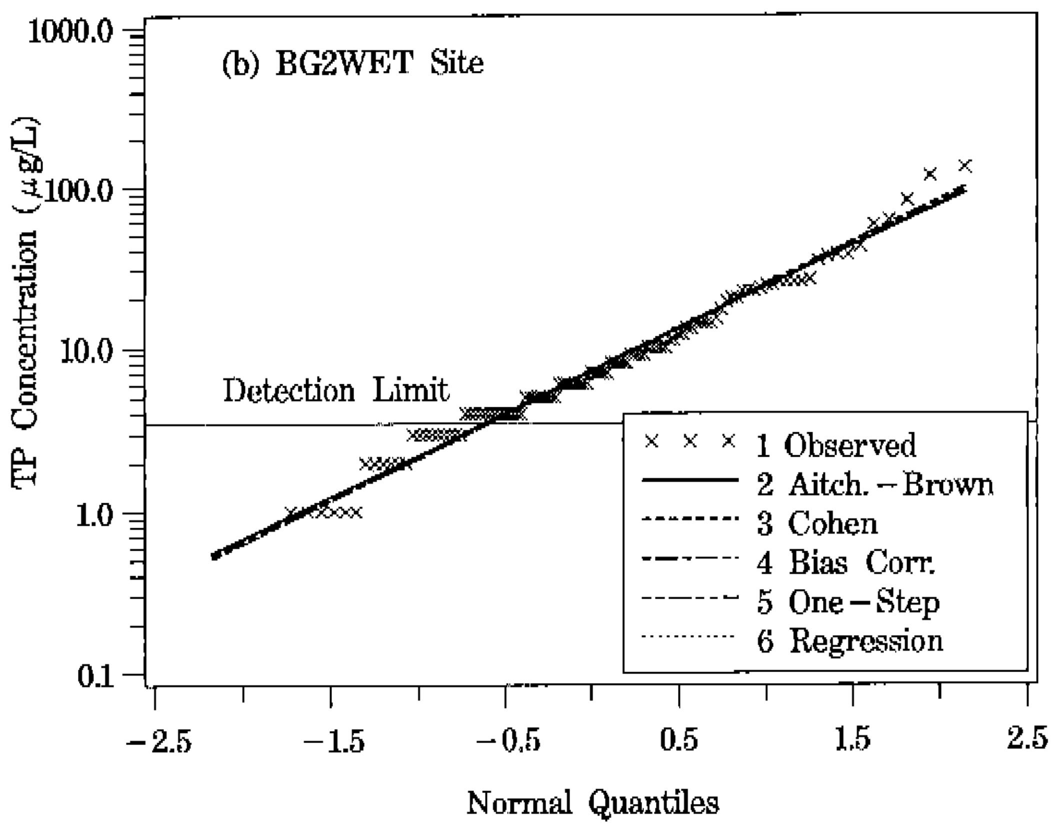


Figure 4. Observed As Well As Estimated CDF Curves of Wet TP Concentrations Versus Standardized Normal Variates (z-scores).

adopting the proposed combined statistics. With the proposed approach, the estimation errors in mean and standard deviation are 2 percent and 7 percent respectively. Moreover, with the typical BDL statistics of a mean of 2.081 μg/L and a standard deviation of 1.054 μg/L, the estimation errors are slightly increased (1 percent and 3 percent, respectively) but are still comparable to the proposed approach.

CONCLUSIONS

The objective of this study was to find an adequate method to compute the mean and variance of the censored TP concentrations in South Florida's rainfall. In particular, this paper proposes an approach to estimate the mean and variance of the BDL data for lognormally distributed singly left-censored data. The proposed approach computes the BDL statistics based on the probability weight moments with the pdf estimated by the conventional maximum likelihood methods. The estimated BDL statistics are useful for calculating TP loads with highly censored TP concentration data as well as for computing the combined statistics by Equations (16) and (17).

The proposed approach was applied to estimate the statistics of the wet TP atmospheric deposition data collected from two monitoring sites in South Florida, where censoring levels of the selected data from both sites are moderate (26.7 percent and 22.9 percent). The combined statistics in each site were computed and compared with those of uncensored data including "recorded" BDL values which are available from laboratory records. The results of analyses herein confirm that the proposed approach to compute combined statistics with the estimated BDL statistics improves estimates compared to the conventional approach to compute the overall statistics.

The one-step restricted method was the most accurate for the wet TP data. The one-step restricted method is relatively simple to use because this method does not rely on iterative solution or tabulated input values. Moreover, this method is based on a ML approach which is widely used for model parameter estimations due to its ability to produce minimum-variance unbiased estimates (Aitchison and Brown, 1969). Therefore, it is highly recommended to use the proposed BDL estimation method in conjunction of the one-step restricted ML method for the censored wet TP concentrations in South Florida rain. With this proposed approach, the estimation errors in mean and standard deviation are about 2 percent and 7 percent, respectively.

Based on the 1996 wet TP samples, the typical mean and standard deviation of the BDL data are about 2.1 μg/L and 1.1 μg/L, respectively. These typical BDL statistics can be used in practice for computing the statistics of the censored wet TP concentrations data in South Florida rain. It was found that, with these typical BDL statistics, the estimation errors in the estimated mean and standard deviation are increased moderately (about 1 percent and 3 percent, respectively) but are still comparable to the proposed approach.

ACKNOWLEDGMENTS

The author expresses sincere appreciation to Maria Loucraft-Manzano, Garth Redfield, Linda Lindstrom, and Tom James, of South Florida Water Management District, for their valuable comments and suggestions on the draft manuscript. I would also like to thank the two anonymous reviewers for their helpful corrections and comments which resulted in an improved presentation of this paper.

LITERATURE CITED

Aitchison, J. and J. A. C. Brown, 1969. The Lognormal Distribution. Cambridge University Press, Cambridge, Massachusetts.

Chen, Z. and T. D. Fontaine, 1997. Simulating the Impact of Rainfall Phosphorus Inputs on the Phosphorus Concentrations in the Everglades Protection Area. *In:* Proceeding of the Conference on Atmospheric Deposition into South Florida: Measuring Net Atmospheric Inputs of Nutrients, October 20-22, 1997. South Florida Water Management District, West Palm Beach, Florida, pp. 19-20.

Cohen, A. C., Jr., 1959. Simplified Estimates for the Normal Distribution When Samples Are Singly Censored or Truncated. Technometrics 1(3):217-237.

Cohen, A. C., Jr., 1961. Tables for Maximum Likelihood Estimates: Singly Truncated and Singly Censored Samples. Technometrics 3(4):535-541.

Cunnane, C., 1978. Unbiased Plotting Positions – A Review. J. Hydrol. 37:205-222.

Gilliom, R. J. and D. R. Helsel, 1986. Estimation of Distributional Parameters for Censored Trace Level Water Quality Data. 1. Estimation Techniques. Water Resources Research 22(2):135-146.

Gleit, A., 1985. Estimation for Small Normal Data Sets With Detection Limits. Environ. Sci. Technol. 19(12):1201-1206.

Hald, A., 1949. Maximum Likelihood Estimation of the Parameters of a Normal Distribution Which is Truncated at a Known Point. Skandinavisk Aktuarietidskrift 32:119-134.

Helsel, D. R., 1990. Statistical Treatment of Data Below Detection Limit, Environ. Sci. Technol 24(12):1767-1774.

Helsel, D. R. and T. A. Cohn, 1988. Estimation of Descriptive Statistics for Multiply Censored Water Quality Data, Water Resources Research 24(2):1997-2004.

Helsel, D. R. and R. J. Gilliom, 1986. Estimation of Distributional Parameters for Censored Trace Level Water Quality Data. 2. Verification and Application. Water Resources Research 22(2):147-155.

Helsel, D. R., and R. M. Hirsch, 1992. Statistical Methods in Water Resources. Elsevier, New York.

Hirsch, R. M., and J. R. Stedinger, 1987. Plotting Positions for Historical Floods and Their Precision. Water Resources Research 23(4):715-727.

Kite, G. W., 1988. Frequency and Risk Analyses in Hydrology, Water Resources Publications. Fort Collins, Colorado.

Mood, A. M., F. A. Graybill, and D. C. Boes, 1974. Introduction to the Theory of Statistics. McGraw Hill Book Co., New York, New York.

Newman, M. C., P. M. Dixon, B. B. Looney, and J. E. Pinder, III, 1989. Estimating Mean and Variance for Environmental Samples With Below Detection Limit Observations. Water Resources Bulletin 25(4):905-916.

Newman, M. C., K. D. Greene, and P. M. Dixon, 1995. UNCENSOR Version 4.0. Savannah River Ecology Laboratory, Aiken, South Carolina.

Pearson, E. S., 1955. The Normal Probability Function: Tables of Certain Area-Ordinate Ratios and of Their Reciprocals. Biometrika 42:217-222.

Persson T. and H Rootzen, 1977. Simple and Highly Efficient Estimators for a Type I Censored Normal Sample. Biometrika 64(1):123-128.

Shaw, J. G., 1961. The Bias of the Maximum Likelihood Estimates of the Location and Scale Parameters Given a Type II Censored Normal Sample. Biometrika 48:448-451.

Schneider, H., 1986. Truncated and Censored Samples from Normal Distributions. Marcel Dekker, Inc., New York, New York.

Schneider, H. and L. Weissfeld, 1986. Inference Based Type II Censored Samples. Biometrics 42:531-536.

Snedecor, W. G and W. G. Cochran, 1980. Statistical Methods, The Iowa State University Press, Ames, Iowa.