The Sky's the Limit

May Raad

HDR, 300 Richmond Road, Suite 200, Ottawa, ON K1Z 6X6

CONFERENCE: 2022 World of Coal Ash – (www.worldofcoalash.org)

KEYWORDS: Censored data, background concentration level, left-censored data, limit of detection, method detection limit, multiply censored data, nondetects

ABSTRACT

The Sky's the Limit – A Review of Laboratory Limits of Detection and Implications in the Statistical Analysis of Groundwater Samples

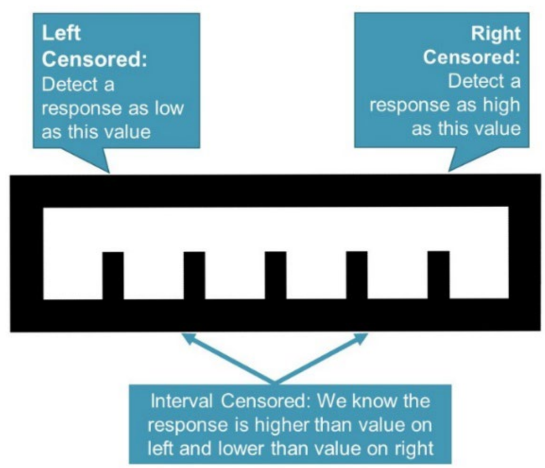
Agencies that have been in step with the Environmental Protection Agency's (EPA) Hazardous and Solid Waste Management System; Disposal of Coal Combustion Residuals from Electric Utilities; Final Rule since 2015 are most likely as of 2022 in assessment monitoring, corrective action or are in the process of conducting closure or retro-fit of CCR Units. In any of these three phases, and including the detection monitoring phase, the Final Rule clearly acknowledges the issues brought about by censored or nondetects constituents when estimating background concentration levels or testing for statistically significant increases (SSIs) or statistically significant levels (SSLs). Part 257.93 (g)(5) specifies that appropriate statistical methods be used when analyzing data below the "limit of detection". What is not clear in the Final Rule is exactly what is a "limit of detection". The method detection limit (MDL) is the most well-known limit, but there are others such as the reporting limit (RL), practical quantitation limit (PQL), limit of detection (LOD), etc. The statistical methods which account for censored or nondetect data, including multiply censored data, are well documented in EPA's Unified Guidance and other notable references; however, the issue of what to do when a laboratory provides multiple types of limits or does not provide the MDL is not satisfactorily addressed. This paper will review the common types of limits issued from laboratories and discuss how varying limit values can impact data quality and credibility of background concentration levels and statistical results when testing for SSIs or SSLs.

INTRODUCTION

Part 257.93 of the Hazardous and Solid Waste Management System; Disposal of Coal Combustion Residuals from Electric Utilities; Final Rule (CCR Rule) describes the sampling and analytical requirements to ensure monitoring results can provide accurate representation of groundwater quality at the background and downgradient wells.[13] The rule provides statistical options to test for differences in constituent concentrations between background or groundwater protection standards (GWPS) and monitoring wells. These include analysis of variance tests, prediction and tolerance limits and control charts with the option to use another other method that meets the performance standards of 257.93(g). For the purpose of this paper, only prediction and tolerance limits will be used as methods to monitor groundwater quality (257.93 (g)(4)) and demonstrate the impact of varying analytical limits.

Assumptions concerning the nature and quality of the groundwater samples are also covered in 257.93(g). The successful application of any statistical test depends on valid assumptions regarding data distribution 257.93 (g)(1) and independence of observations over time and area 257.93 (g)(6). The issue with analyzing data that have been censored is highlighted in 257.93 (g)(5). It specifies that appropriate statistical methods be used when analyzing data below the "limit of detection".

From a statistical perspective, limits of detection represent censoring limits. Groundwater samples whose concentrations are below the limits of detection applicable to the laboratory facility taking the measurements are deemed as nondetects.



In general, censored data represent "partial" measurement values because it is known the values are either below a certain value (i.e., left censored) or above a certain value (i.e., right censored). In some studies, values are known to be above one value but below another value (i.e., interval censored).

Measurement devices in laboratories are tested and calibrated to detect values down to a limit below which the measured values are unclear, though one cannot assess a zero concentration is confirmed. Hence, groundwater samples with nondetects represent left-censored data. While the actual concentration below the censored value is unknown, knowing what the censored value is provides significant insight. To delete such data from a statistical analysis would be a significant loss of information.

The CCR Rule does not indicate what types of statistical methods one should use in the presence of censored data. However, it does briefly describe the quality of the censored value in Part 257.93(g)(5):

“Any practical quantitation limit that is used in the statistical method shall be the lowest concentration level that can be reliably achieved with specified limits of precision and accuracy during routine laboratory operating conditions that are available to the facility.”

The CCR Rule does not address the real challenges laboratories have when calibrating their equipment to measure samples with typically very low concentrations. These challenges result in a myriad of detection limits, and for the same class of limits, differing censored values from the lab even on the same day, resulting in the outcome of multiply censored data.

Are there statistical procedures that can accommodate censored data based on different types of limits or different values even within the same class of limits?

Fortunately, there are some and this paper presents data examples of such methods. However, these methods cannot make-up for the quality issues brought about when laboratories select types of limits that are not the “lowest concentration level that can be reliably achieved”. By working closely with the laboratory and clearly specifying quality expectations in terms of precision and reproducibility, can quality data be provided on a routine basis for the purpose of CCR Rule compliance for parts 257.93, 257.94, 257.95, and 257.98 (c) 1),(2), and 257.102(c).

OVERVIEW OF ANALYTICAL LIMITS

Early researchers from the 1960s such as Bernard Altshuler, Bernard Pasternack and Lloyd Currier established methods to determine statistically based limits of detection for measurement devices and methods. The statistics were necessary to estimate the level of confidence in the numerical value of the limit being calculated.

In particular, Lloyd Currie now retired from the National Institute of Standards and Technology led the development of the 1995 International Union of Pure and Applied Chemistry’s (IUPAC) “Nomenclature in Evaluation of Analytical Methods Including Detection and Quantification Capabilities”.

The nomenclature from the IUPAC was incorporated into the ISO 11843 series of standards which is followed in Europe. However, in North America, a varying range of limits of detection and quantification have arisen to meet different needs. The situation now exists where different laboratories use different types of censoring limits and reporting standards for nondetects. These differing methods, definitions and statistical derivations for determining analytical limits subsequently can lead to different outcomes and interpretations.[2] In summary, there is no standard industry practice in North America for setting censoring limits.[7]

A consistent concept in analytical limits is the method detection limit (MDL) though this nomenclature for the same concept may be termed differently by different laboratories. It is a key analytical limit as it conveys the precision and reproducibility of a laboratory's process or 'method' to measure analytes under routine laboratory operations. The USEPA's revised definition of the MDL is as follows:

"The MDL is the minimum measured concentration of a substance that can be reported with 99% confidence that the measured concentration is distinguishable from method blank results." [14]

A related statistic which uses the similar testing protocols as the MDL and has basically the same definition is the instrument detection limit (IDL). It is a statistical measurement that is used to measure the performance of the instrument. The main difference between the IDL and the MDL is how the testing samples are prepared. For the IDL, the samples are dissolved "neat" in a solvent while for the MDL, the samples are dissolved in a sample matrix.

The IDL is estimated through the method to establish an instrument's ability to characterize the amount of analyte (e.g., sample concentration) required to be 99% confident that a measured signal is real and not baseline noise. The IDL sets an instrument's performance while the MDL is used for routine laboratory testing.[1]

The following is a list of analytical limits while by no means exhaustive demonstrates the range in different limits and the source of possible confusion when conducting statistical analyses of analytes in groundwater samples.

EDL – Estimated Detection Limit

EQL – Estimated Quantitation Limit

IDL - Instrument Detection Limit

LLD – Lower Limit of Detection (comparable to IDL)

LLQ – Lower Limit of Quantitation (comparable to EQL)

LOD – Limit of Detection (comparable to IDL[9,10], or MDL[2])

LOQ – Limit of Quantification (comparable to EQL)

MDL – Method Detection Limit

MQL – Method Quantitation Limit (comparable to EQL)

PQL – Practical Quantification Limit (comparable to EQL)

RL – Reporting Limit, DLR – Detection Limit for Purposes of Reporting (comparable to EQL, PQL)

SDL – The MDL adjusted to reflect sample-specific actions such as but not limited to dilution or smaller aliquot size

SQL – Sample Quantification Limit - the EQL,PQL or RL adjusted to reflect sample specific actions such as but not limited to dilution or smaller aliquot size

There is an important distinction between the use of 'detection' and 'quantitation'. Detection limits indicate minimum concentration values that can be detected with a certain level of statistical confidence as distinguishable from instrument noise or method blanks. Quantitation limits indicate minimum values that a laboratory feels confident enough on which to quantify and report.

While detection limits are based on statistical methods, quantitation limits are often arbitrary, and each laboratory may have different methods to set their quantitation limits. Multiples of the IDL, EDL or MDL are used to set the applicable quantitation limits with an outcome of higher and higher quantitation limits over the numbers of laboratories conducting environmental testing. As there is no industry standard followed in the United States regarding how high these limits can be, are we in an era of the "sky's the limit"?[1]

The laboratory is required, however, to set the quantitation value to be the concentration of the lowest standard analyzed for a sample set based on the low standard used for instrument calibration.[2] In reality, laboratories will take multiples of a detection limit such as the MDL that go beyond the lowest standard to minimize the chance of questionable reported values. Hence, the often arbitrariness of the quantitation limit.

The complement of setting the lowest concentration that can be reliably reported is the highest concentration that can be reliably report. During instrument calibration, the standard samples also have a maximum concentration assumed by the laboratory. If testing samples arrive and have higher concentrations than those set by the sample standards, the laboratory then dilutes the testing samples by factors that enable the instruments to reliably report based on instrument calibration. When reporting such samples, the laboratory may flag the measurements with SDLs and SQLs along with the dilution factor. If they do not use SDLs or SQLs, they may still report MDLs or PQLs but with values that vary as a function of the dilution factor.

[1] The author of this article chose the title independently of published research. However, through research for article, she came across another group that also described the growth in detection and quantitation limits and wondered as well if the "sky's the limit". See the article by Michael Brisson and Derek Popp "Detected or Not" in reference 2 on the last page. This is probably more than coincidental and underlies the frustration of many in the range and definitions of detection and quantitation limits.

It is helpful if the lab provides measurement results prior to and after dilution to better understand the impact of the dilution on the sample. Through a question and answer forum on the USEPA's website, EPA recommended dilution factors of five or less.[16]

Even if laboratories use dilution factors higher than five, it is still preferable to have the laboratory provide results after dilution than for the laboratory to report on a value that falls outside the calibration range.

The USEPA recommends reporting values between the detection limit and the quantitation limit as estimates.[12, 15] Laboratories will qualify such estimates as 'J' qualifiers. While there is some uncertainty in the measurements, there is a quantity that can be measured as opposed to just being detected. Only those analyte samples that cannot be detected below the detection limit such as the MDL are assigned a nondetect status. Laboratories will qualify a nondetect result with a 'U' qualifier.

For the purpose of consistency with the CCR Rule, the PQL will be referenced in this paper in place of the more common reporting limit (RL) or the less common DLR or EQL.

This paper will present some of the more common detection and quantitation limits and compare how key statistical parameters such as upper prediction limits and upper tolerance limits are impacted by the different statistical methods used in the presence of censored or nondetect data and different limits of detection or quantitation such as the MDL and PQL.

STATISTICAL METHODS FOR CENSORED DATA

An entire field of statistics is regulated to the analysis of data with censored or nondetect values. The medical field relies on such methods to model survival times based on the presence of certain diseases. Medical data from clinical or epidemiological studies are examples of right censored data since after the study ends, death is expected to happen but the time of which it will happen is unknown.

The transportation field uses these approaches to develop mathematical functions that relate a travel mode's cost, reliability, convenience, and comfort among other mode attributes to how a commuter or driver values those attributes. Transportation planners use these models to build networks of travel demand for a municipality or region. Data for this use is a mix of left, interval and right censored data.

In the environmental sector, especially in the analysis of constituents in groundwater samples, it is the lowest concentration that can be reliably detected or reported which is the goal. This is important as it minimizes the chance of false negatives. However, due to limitation in instrumentation or process, concentrations below a threshold cannot be detected. This does not rule out that the constituent is still present in the sample. Hence, that threshold value is an example of left-censored data.

As to the statistical methods used to estimate sample parameters such as averages, standard deviations, confidence, prediction, or tolerance limits, significant literature is available to practitioners. The author of this paper recommends research completed by D.R. Helsel for those that wish to have a comprehensive understanding of these methods.[3,4,5,6] For the purpose of this paper the common and accessible Kaplan-Meier (KM), Regression on Order Statistics (also called Robust Regression on Order Statistics) (ROS) and the Maximum Likelihood Method (MLE) are used to demonstrate the differences in estimates for example datasets. USEPA's Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities (Unified Guidance) and ProUCL Technical Guide Version 5.1 have multiple examples that explain and demonstrate the KM and ROS Methods. [11, 12] The MLE method is not demonstrated in these documents, but the algorithm can be accessed using software such as R, Phyton, Matlab or SAS among other statistical applications.

The methods that will be demonstrated in this paper have an important advantage of being able to accommodate data with multiple censored values (also referred to as multiply censored data) within the same data set. They are not appropriate in circumstances where more than 50 percent of the data consists of censored or nondetect data. Nonparametric techniques based on higher order statistics should be used with data sample with such high levels of censored data.[11, 12]

Naïve statistical methods such as simple substitution using the censored value or ½ the censored value while discouraged by researchers, are unfortunately still being used today.[3, 4, 5, 6, 11, 12] The problem with this method is that no information from the detected values is used to represent what the nondetect constituents could be. Because of the arbitrary nature when assuming the concentration of the nondetect is the censored value or a fraction of the censored value, an unknown level of bias enters into the calculations. Key statistics such as the upper prediction or tolerance limits as referenced in the CCR Rule may be higher or lower than what the underlying distribution of constituent concentrations would suggest. By leveraging all the available censored and non-censored sampled data to compute what proportion of observations would fall below differing lower thresholds, resulting sample parameters would better reflect the nature of the underlying distributions.

Basic decision inputs such as sample size and the type of sample distribution help practitioners decide which of the three recommended statistical methods to use.

The KM method is traditionally used to model data with a mix of censored and non-censored values that cannot be explained using parametric distributions. This approach assumes the non-censored or censored data arise from the same underlying distribution, regardless of the nature of that distribution. The KM method is able to estimate the sample mean and standard deviation from the cumulative distribution function that is outputs.

After sorting the censored and non-censored values from smallest value to largest, it calculates the approximate percentage of observations below each distinct value until 100 percent of the data is covered. It does not require the actual concentration for each nondetect observation below its respective censored threshold value. The KM method works well with small datasets less than 50 observations; however, it is not generally recommended for datasets with only one censored value.

The ROS method is a parametric method and uses a fitted model based on the underlying parametric distribution to substitute or ‘impute’ a censored value with a modeled value. Examples of parametric distributions which can be used with the ROS method are normal, lognormal or gamma. The method assumes concentrations are a linear function of quantiles from the data for the appropriate distribution. The imputed values are only used to estimate the sample mean and variance from which other statistics such as prediction or tolerance limits are derived. This method is advantageous for multiply censored data. Unlike the KM method, it can be successfully used when the data set has only one censored value.

The MLE method assumes a parametric distribution. It can accommodate a wide variety of parametric distributions. The lognormal or gamma are common distributions which provide better functions to model data from samples containing groundwater concentrations for various metals. Other distributions which can be tested include normal, log-logistic or Weibull among other distributions from the exponential family. It does require however, a large sample size of approximately 50 or more. It uses the non-censored data, the proportion of censored data and the distribution assumption to solve for the data set’s summary statistics. The MLE method is numerically iterative and solves for the data set’s mean and variance that maximizes the probability of observing such data. From a methodological perspective, it is the best approach; however, significant data are required. The availability of large samples is most likely not to be the case with groundwater samples.

The next section demonstrates the popular imputation methods using fictional data with a mix of limits to estimate summary statistics from which other statistics such as UPLs and UTLs are produced. The impact of deleting nondetects from the analysis is also included in some of the data examples. The software package ProUCL available from USEPA is used to calculate the UPLs and UTLs.[11] In particular the 95 percent UPL (95UPL) for one future observation and the 95 percent UTL with 95 percent coverage (95UPL95) are calculated for illustrative purposes. The first example which used the MLE to estimate UPLs and UTLs was generated using the software package NCSS.[8] Because of the smaller sample sizes used in examples 2 to 5, the MLE method was not applied.

EXAMPLE 1: SAMPLE WITH 20 PERCENT NONDETECTS, ONE DISTINCT MDL VALUE

The data represent metal concentrations from surface water samples taken from a river over a 30 year period under conditions of minimal anthropogenic activity in the basin in which it is situated. There are 19 censored values with a threshold of 0.100 ug/L. A scatter plot of the observations is shown in **Figure *1***. Based on Lilliefors test for normality at the 5 percent significance level, a lognormal distribution fits the data. The shape of the histogram of the data in **Figure 2** supports this GOF test outcome. **Table *1*** contains the summary statistics based on different imputation methods. Deleting all nondetects causes significant shift in the distribution with the highest sample mean and largest values for the 95UPL and 95UTL95. Using these statistics to represent background or a GWPS, respectively, can significantly inflate the underlying natural upper limits from baseline data.

Interestingly, imputing using the censored value, ½ the censored value, KM, ROS or MLE methods, produced comparable 95UPLs and 95UTL95s. Given the relatively large sample size and average level of censorship, all methods provided reasonable estimates. Of note is the MLE method which is considered the most robust and produced the lowest estimates of the methods for the 95UPL and 95UTL95. The skewness of the data set is high since the standard deviation of the detected logged data is greater than 1 or 1.24.[11] A risk with using the lognormal distribution for environmental data is that the modeled tails tend to be long producing higher upper limits than what the underlying distribution's data can justify. The MLE method's ability to appropriately consider the proportion of the data less than censored value of 0.100 was able to control for the tendency of the other methods to output higher upper limits.

Figure 1: Scatter plot of example 1 data (ug/L) vs time with one distinct MDL value

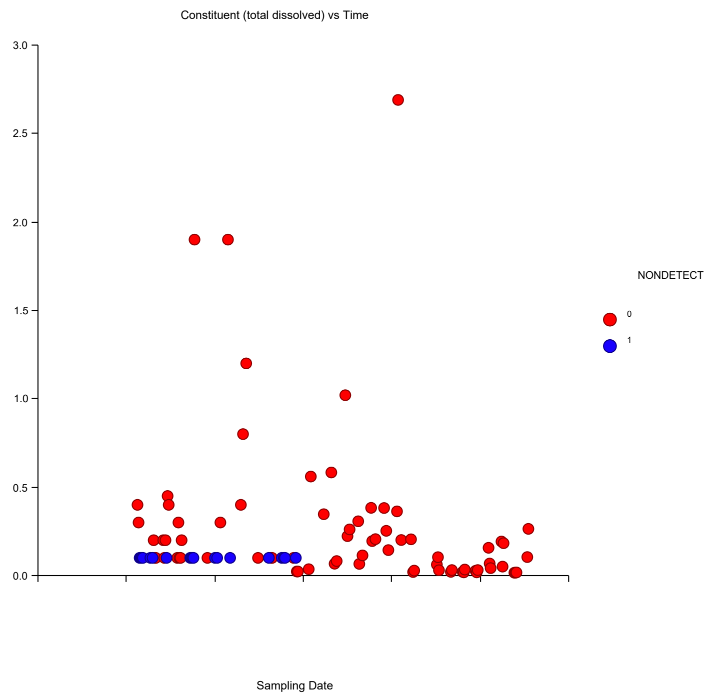


Figure 2: Histogram of example 1 data (detect and nondetect)

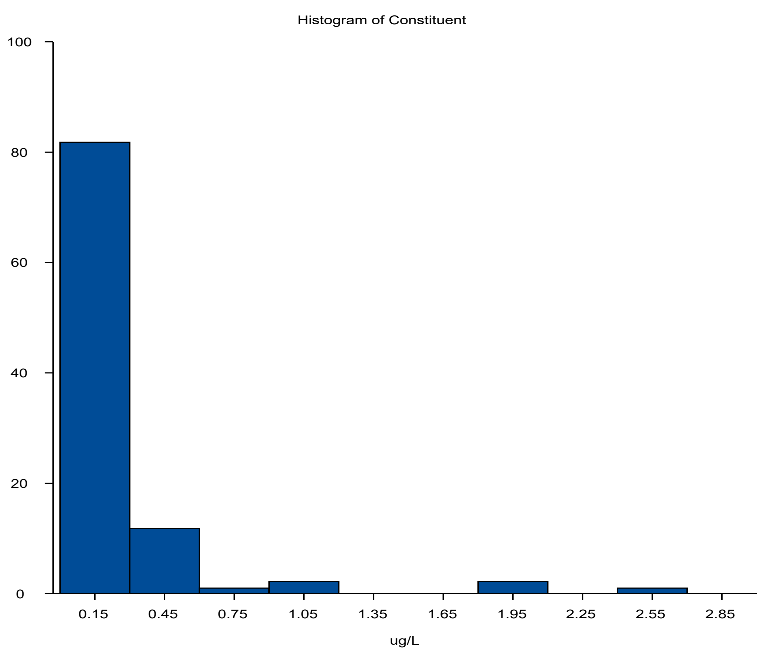


Table 1: Example 1 data (ug/L) summary statistics and 95% UPL and 95% UTL with 95% coverage

Method	Mean	SD	95UPL	95UTL95
Censored Value	0.241	0.410	0.779	1.05
1/2 Censored Value	0.231	0.414	0.747	1.02
Deletion	0.278	0.453	1.03	1.49
KM	0.0958	3.53	0.788	1.10
ROS	0.229	0.415	0.816	1.14
MLE	0.122	2.92	0.688	0.91

EXAMPLE 2: SAMPLE WITH 28 PERCENT NONDETECTS, FIVE DISTINCT MDL VALUES

This constituent had three nondetects and three measurements with 'J' qualifiers. The three nondetects corresponded to MDL values of 2.48, 1.28 and 0.128 mg/L. The MDL values for each sample ranged from as low as 0.128 to as high as 128 mg/L. The laboratory that provided these results diluted most of the samples using factors of 5, 10, 20 and 1000. The measurement which was diluted at a 1000 factor level corresponds to the highest detected 'J' estimate value of 128 mg/L. A scatter plot of the values is shown in **Figure 3** below. The GOF test based on the Shapiro-Wilk test indicated that the normal distribution appropriately fit the detected data at the 5 percent level of significance. A histogram of the detect and nondetect data is in **Figure 4**.

Figure 3: Scatter plot of example 2 data (mg/L) vs time with one distinct MDL value

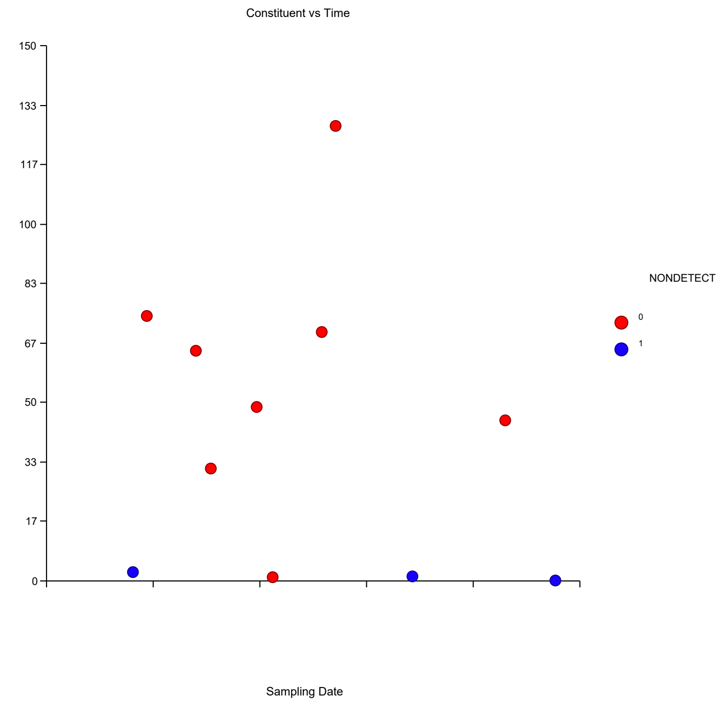
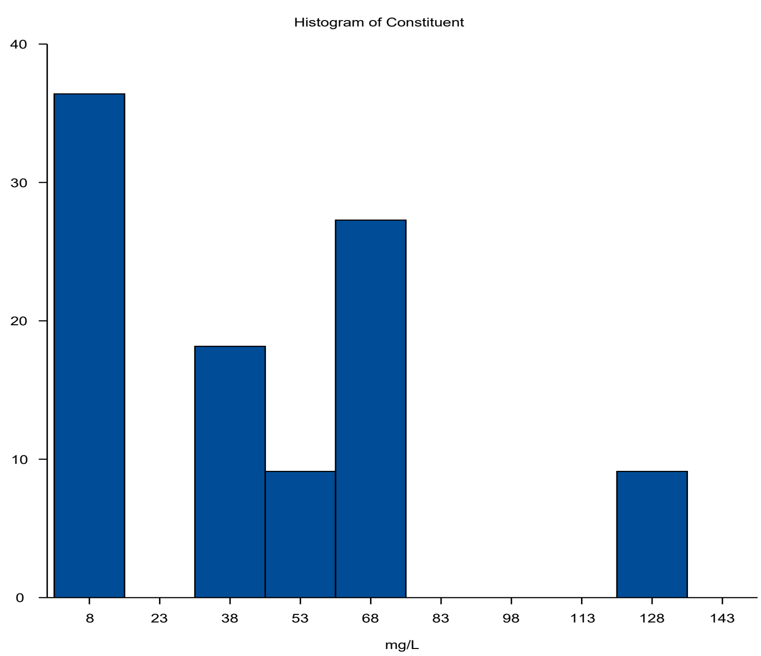


Figure 4: Histogram of example 2 data (detect and nondetect)



Given the data followed a normal distribution, ProUCL only output summary and test statistics using the ½ substitution and KM method. (ProUCL removed the ROS feature for normally distributed data since the ROS method tended to impute lower censored values with negative numbers, potentially biasing sample summary statistics). The results using both methods are comparable as shown in **Table 2**; however, the KM method being able to leverage data patterns in the detected data, provides estimates that are lower than those provided by the simple substitution method. A concern of many researchers in the statistical analysis of groundwater samples is how the simple substitution method can bias results due to the arbitrariness of imputed value. This concern is demonstrated in **Table *2***, albeit the shift in estimates is relatively small. A larger sample with the same or higher level of censorship, could exhibit larger differences in the estimates.

Table 2: Example 2 data (mg/L) summary statistics and 95% UPL and 95% UTL with 95% coverage

Method	Mean	SD	95UPL	95UTL95
1/2 Censored Value	42.2	40.7	119	157
KM	42.2	38.9	116	152

Sometimes, laboratories only provide the PQL and not both the MDL and its corresponding PQL. Without the MDL, the PQL, a significantly higher value, becomes the threshold for censored data.

To demonstrate the impact of only having the PQL and a higher analytical limit, the same data set is used with all values below the PQL treated as nondetects. With this change, now there are 6 nondetects (55 percent nondetects) and the data follow a lognormal distribution based on the Shapiro-Wilk GOF test at the 5 percent significance level. The PQLs have the following values: 0.375, 1.88, 3.75, 7.50 and 375 mg/L. **Table 3** below shows the results of estimating summary statistics for this dataset assuming data follow the lognormal distribution. With this higher level of censorship, notable differences in the UPLs and UTLs appear across all three methods. The simple substitution method shifts the location 95UPL and 95UTL95 to the right relative to the ROS method. Compare a 95UPL of 717 from the simple substitution method to that of the 86.4 from the ROS method. Similarly so for the 95UTL95. The simple substitution method produces a value of 5,650 while the ROS method produces a value of 143. The KM method also exhibits a high level of bias in its summary and test statistics with the lowest sample mean and standard deviation of 4.36 and 11.7 mg/L, respectively. The 95UPL and 95UTL94 this method outputs are also notably higher than the same statistics output using the ROS method by a factor of 5 for the 95UPL and a factor of 31 for the 95UTL05.

Table 3: Example 2 data (mg/L) summary statistics and 95% UPL and 95% UTL with 95% coverage using a higher limit of detection

Method	Mean	SD	95UPL	95UTL95
1/2 Censored Value	41.8	55.7	717	5,650
KM	4.36	11.7	460	4,450
ROS	35.3	20.1	86.4	143

As final note on this data example, the Unified Guidance recommends using nonparametric methods to estimate test statistics such as UPLs and UTLs. The above methods demonstrated in **Table 3** were used solely to show how high values for detection limits can impact the quality of the analysis.

The 95UPL and 95UTL95 using nonparametric methods for this dataset produce an estimate of 375 mg/L.

Regardless of methods used to analyze this dataset with higher limit values, all methods produced 95UPLs and 95UTL95s higher than using the dataset with lower detections limits shown in **Table 2**.

Given the potential bias that can result in the statistical analysis of this dataset when using high censored values, practitioners should insist that detection limits be based on statistical methods which follow recognized standards such as USEPA's 'Definition and Procedure for the Determination of the Method Detection Limit, Revision 2'[14] and that quantitation limits reflect the concentration of the lowest standard analyzed for a sample given the low standard used for instrument calibration.

EXAMPLE 3: SAMPLE WITH 45 PERCENT NONDETECTS, FIFTEEN DISTINCT MDL VALUES

The MDL values for this sample range from a low of 0.0865 to a high of 0.612. There are no observations with 'J' qualifiers. While the multitude of differing censored values may seem problematic, they are not. The MDL values fall within the lower half range of the detected values. See **Figure 5**. A normal distribution appropriately fit the detected data based on the Shapiro-Wilk GOF test at the five percent significance level. A histogram of the detect and nondetect data is in **Figure 6**. ProUCL was used to estimate upper limits and results are in **Table 4**. As the data are assumed to be normally distributed, ProUCL only outputs results for the simple substitution method and KM method.

Interestingly, the highly discouraged simple substitution method produced similar summary statistics and upper limits to that produced using the KM method. This is probably due to the fact that the normal distribution fit the data well providing a symmetry that made up for the arbitrariness of the simple substitution method based on ½ the MDL. Nonetheless, the KM or ROS method where data quality and quantity permit should be used to avoid unintentionally estimating very high and potentially non-representative upper limits of natural variability from background samples. Example 2 provides a good demonstration where simple substitution produced higher than expected values for upper limits.

Figure 5: Scatter plot of example 3 data (mg/L) vs time with fifteen distinct MDL values

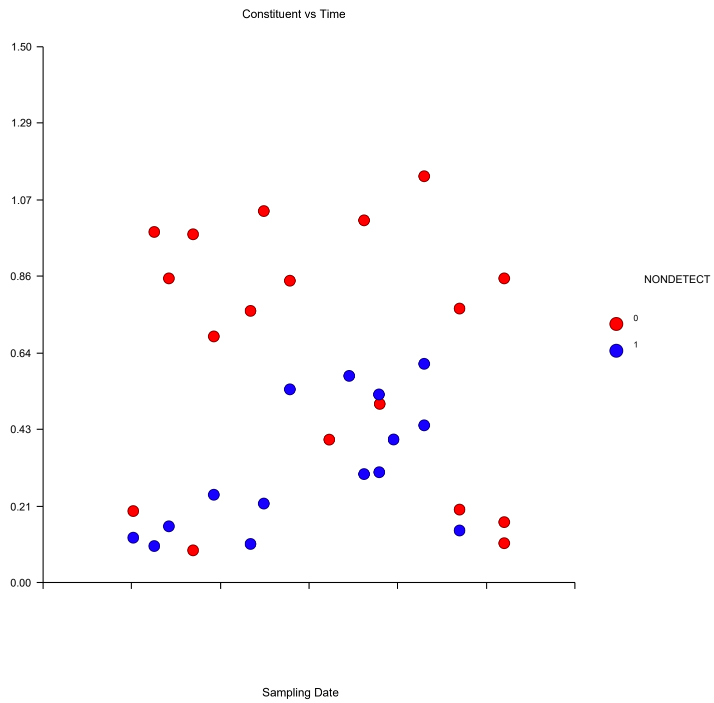


Figure 6: Histogram of example 3 data (detect and nondetect)

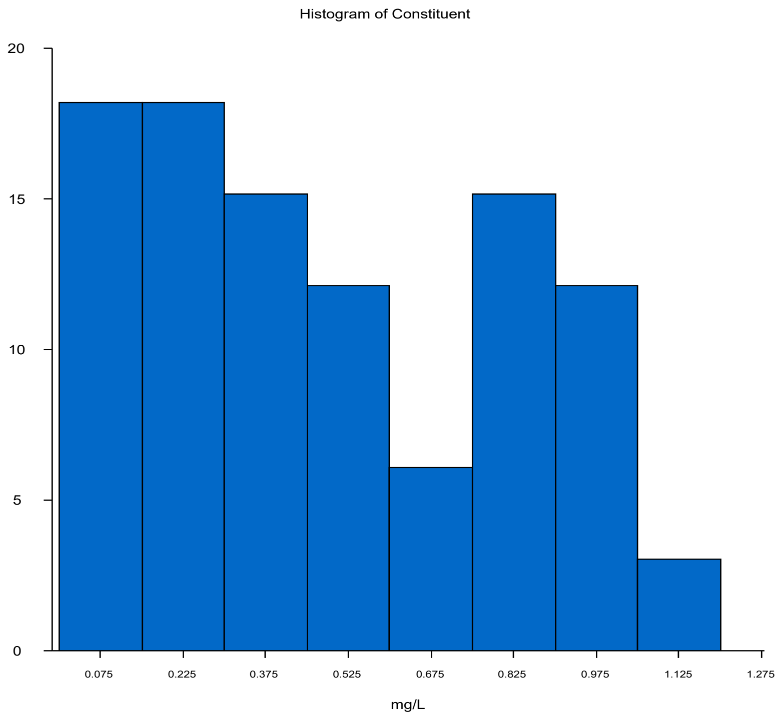


Table 4: Example 3 data (mg/L) summary statistics and 95% UPL and 95% UTL with 95% coverage

Method	Mean	SD	95UPL	95UTL95
1/2 Censored Value	**0.424**	**0.364**	**1.04**	**1.21**
KM	**0.408**	**0.369**	**1.05**	**1.25**

EXAMPLE 4: SAMPLE WITH 4 PERCENT NONDETECTS, ONE DISTINCT PQL VALUE

This sample consisted of 27 samples of a particular congener with only 1 value labeled as nondetect with a “U” qualifier. However, this one nondetect was associated with a PQL that was higher in value than all detects in the sample with a value 190 pg/L, over two times as large as the second highest detected value of 82.3 pg/L (see **Figure 7**). No MDL was provided. The detected data follow a lognormal distribution at the 5 percent significance level based on the Shapiro-Wilk GOF test. A histogram of the data is in **Figure 8**.

Figure 7: Scatter plot of example 4 data (pg/L) vs time with one distinct MDL value

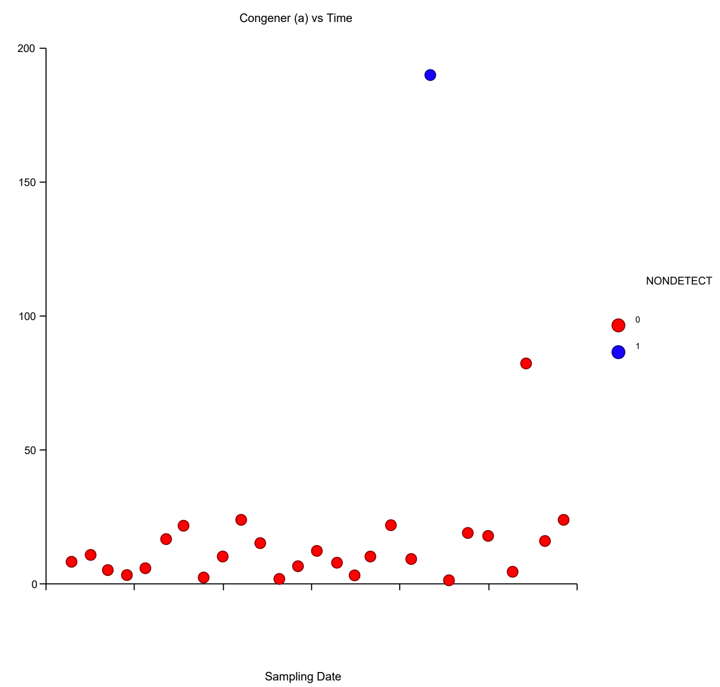
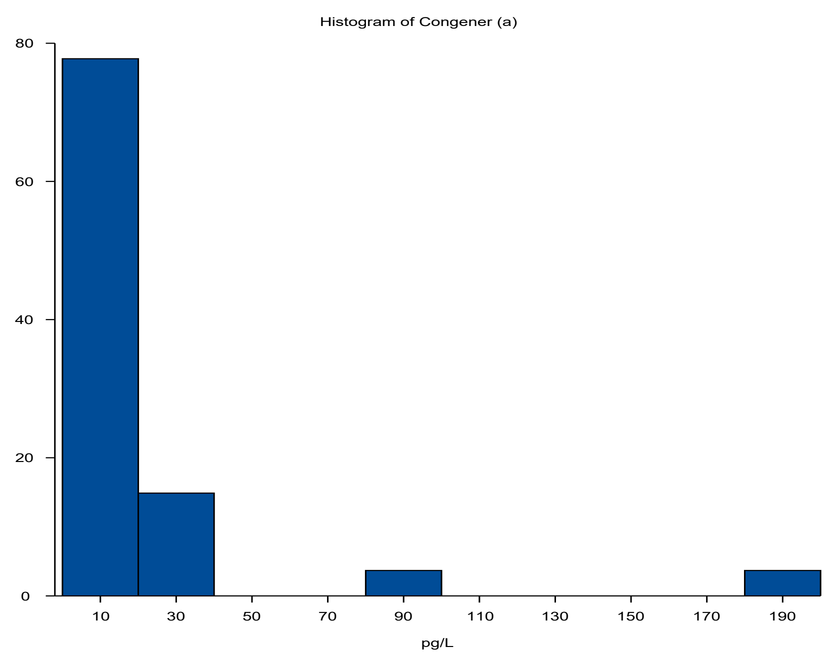


Figure 8: Histogram of example 4 data (detect and nondetect)



The three methods of the simple substitution, KM and ROS used to estimate the 95UPL and 95UTL95 are provided in **Table *5***. The KM and ROS methods provide similar results which are related to the fact there is only one nondetect value. The sample mean and standard deviations are comparable between the simple substitution method and the KM or ROS methods. As mentioned in the previous examples, the use of the simple substitution method should be avoided as in this case it shifts the summary statistics and the 95UPL and 95UTL95 test statistics to the right relative to what the more advantages KM and ROS methods produce. An investigation is recommended as to whether or not the nondetect observation should be removed as an outlier.

Table 5: Example 4 data (pg/L) summary statistics and 95% UPL and 95% UTL with 95% coverage

Method	Mean	SD	95UPL	95UTL95
1/2 Censored Value	16.9	21.9	59.2	101
KM/ROS	13.7	15.4	45.4	73.2

EXAMPLE 5: SAMPLE WITH NO NONDETECTS, FOUR DISTINCE PQL VALUES

This example uses another type of congener that has two detects flagged as ‘J’ estimates. In total, there are four different PQL values of 275, 199, 192 and 109 pg/L. The two concentrations flagged as ‘J’ qualifiers were associated with PQL values of 275 and 192 pg/L. These two ‘J’ estimates were the two largest values (42,600, 17,300 pg/L) in the sample of 26 observations (see **Figure 9**). Laboratory notes for these two observations indicate the concentrations in the samples are outside the instrument’s calibration range. Parametric distributions to this dataset can not be fit, hence, nonparametric order statistics are used to estimate the upper limits. Visually, it is difficult to detect a parametric distribution based on histogram of the data in **Figure 10**.

Using nonparametric order statistics produced reasonable values for the 95UPL and 95UTL95 in **Table 6** relative to the sample mean considering the sample standard deviation. However, what would the results look like if the largest value was removed? The ‘J’ estimate value of 42,300 pg/L which is over double the value of the second largest concentration, also a ‘J’ estimate, is suspect. Dropping the largest ‘J’ estimate and re-testing for GOF identified a normal distribution. A histogram of the revised data is in **Figure 11** below. The summary statistics and upper limits in Table after the largest value has been removed show notable reductions in values for the standard deviation and the 95UPL and 95UTL95. For example, the standard deviation is just over half as large as the sample with the largest value. It is good practice to follow-up with the laboratory to better understand this value and how representative it is of the concentrations at the time of sampling.

Figure 9: Scatter plot of example 5 data (pg/L) vs time

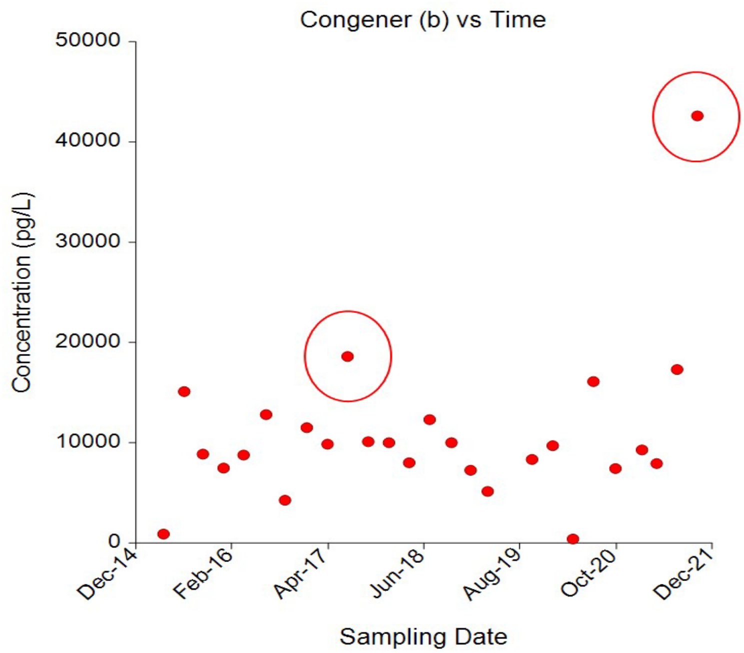


Figure 10: Histogram of example 5 data

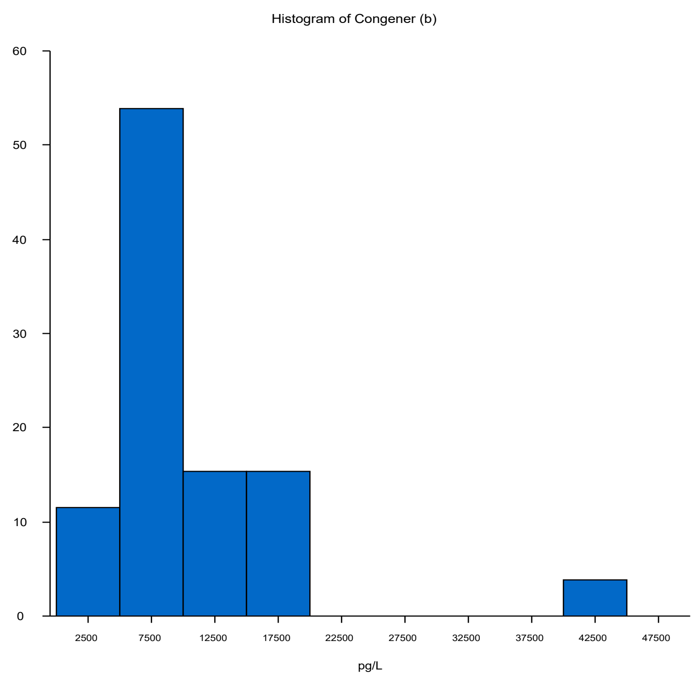


Figure 11: Histogram of example 5 data (pg/L) without the largest observation

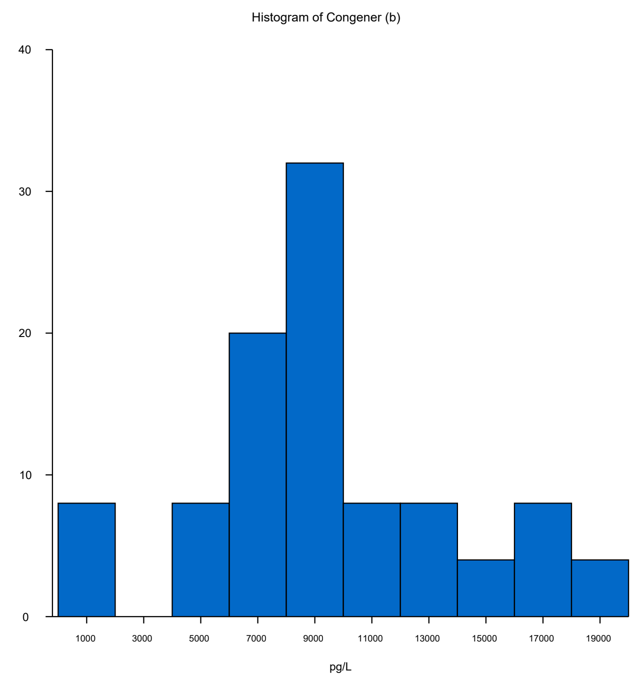


Table 6: Example 5 data (pg/L) summary statistics and 95% UPL and 95% UTL with 95% coverage

Method	Mean	SD	95UPL	95UTL95
Nonparametric	10,767	7,799	34,200	42,600
Normal (with 42,600 removed)	9,493	4,409	17,186	19,598

RECOMMENDATIONS

Based on a review of available literature on the statistical analysis of multiply censored data for groundwater samples and the CCR Rule requirement that quantitation limits used in the statistical method “be the lowest concentration level that can be reliably achieved with specified limits of precision and accuracy during routine laboratory operating conditions that are available to the facility”, statistical methods over naïve or simple substitution methods should be used to estimate key sample summary statistics in the presence of censored data. The use of simple substitution methods or deletion of nondetects shifts the upper tails of the data towards the right producing higher (and potentially biased) UPLs and UTLs than warranted.

With respect to the three popular statistical methods to estimate summary statistics from data sets with censored data, the MLE method is the most preferable given its additional feature of searching for summary statistics such as the mean and variance which are most likely to predict the observed data. However, it does require larger sample sizes to facilitate its iterative, computational algorithms.

The ROS method like the MLE method is parametric and can produce estimates from smaller datasets. It can address both multiply and single censored data.

The KM method is a distribution free method. It is useful when datasets are small and when it is unclear what distribution describes the data set. However, it may not be ideal when there is only one level of censoring and a high percentage of nondetects (e.g., greater than 20%).

All three methods assume that less than 50 percent of the data are censored. Otherwise, nonparametric, higher order statistics methods are used. Nonparametric methods make no distinction between the censored values and noncensored values. Hence, care must be taken to ensure that some of the detected values are larger than all of the nondetect values and “J” estimates.

In conjunction with using good statistics to analyze data sets, practitioners should partner with laboratories to reach a common understanding that the quantitation limits must represent the lowest concentration standards analyzed for a given analyte and instrument. Choosing limits that are arbitrarily higher than the lowest standard impacts data quality.

Practitioners need to understand the detection and quantitation concepts and the definitions of data qualifiers as practiced by the laboratory so that the data are analyzed correctly.

Understanding how and when the laboratory calibrates its instruments and protocols for dilution can reduce confusion when estimated reported values (e.g., ‘J’ estimates) or censored values are significantly higher than the highest confirmed detect values.

The wide range of analytical limits that laboratories are using can leave practitioners with the attitude that the “sky’s the limit”. One may question how much confidence can be placed in laboratory results when limit values can vary within the same sample or when they yield nondetect or ‘J’ estimate values higher than detected values.

Partnering with laboratories and agreeing to the analytical terms and methods for limits prior to the sample analysis will reduce the confusion on what censored values, statistics and data should be used to estimate the necessary statistics to test for SSIs or SSLs.

In addition, data quality will be improved producing results that are defensible and helping to confirm that the statistical analysis method used to evaluate groundwater monitoring data based on the prediction and tolerance limits or any other statistical method is as “at least as effective as any other approach in this section for evaluating groundwater data “.[13]

REFERENCES

[1] Agilent, Instrument Detection Limit (IDL), https://www.agilent.com/en/support/liquid-chromatography-mass-spectrometry-lc-ms/instrument-detection-limit, accessed May 13, 2022

[2] Brisson M, Popp D, "Detected or Not?": Defining Laboratory Analytical Limits, the Synergist, https://synergist.aiha.org/201711-detected-or-not, accessed May 13, 2022

[3] Helsel DR, Cohn TA. (1998) Estimation of descriptive statistics for multiply censored water quality data. Water Resources Res; 24: 1997–2004

[4] Helsel DR, Hirsch RM. (2002) Statistical methods in water resources. Department of the Interior, United States Geological Survey, Reston, Virginia. Available from: http://water.usgs.gov/pubs/twri/twri4a3/

[5] Helsel DR. (2005) Nondetects and data analysis. New York: John Wiley & Sons, Inc

[6] Helsel, D.R., 2010. Much Ado About Next to Nothing: Incorporating Nondetects in Science. Ann Work Exposures and Health 54, 257-262

[7] ITRC (Interstate Technology & Regulatory Council). 2013. Groundwater Statistics and Monitoring Compliance, Statistical Tools for the Project Life Cycle. GSMC-1. Washington, D.C.: Interstate Technology & Regulatory Council, Groundwater Statistics and Monitoring Compliance Team. https://projects.itrcweb.org/gsmc-1/., accessed May 13, 2022

[8] NCSS 12 Statistical Software (2018). NCSS, LLC. Kaysville, Utah, USA, ncss.com/software/ncss.

[9] Qaallc, Understanding Laboratory Reporting Limits, https://qaallc.com/article-2.html, accessed May 13, 2022

[10] Scion Instruments, Instrument Detection Limit (IDL) for the Scion SQ GC-MS, https://scioninstruments.com/instrument-detection-limit-idl-for-the-scion-sq-gc-ms/, accessed May 13, 2022

[11] Singh, A. and Ashok Singh. ProUCL Version 5.1.00 Technical Guide. USEPA, EPA/600/R-07/041, October, 2015

[12] U.S. Environmental Protection Agency (USEPA). Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities: Unified Guidance. Office of Resource Conservation and Recovery, Program Implementation and Information Division, USEPA, EPA 530/R-09-007, 2009

[13] U.S. Environmental Protection Agency (USEPA). Hazardous and Solid Waste Management System; Disposal of Coal Combustion Residuals From Electric Utilities; Final Rule. 40 CFR Parts 257 and 261, April 17, 2015

[14] U.S. Environmental Protection Agency (USEPA). Definition and Procedure for the Determination of the Method Detection Limit, Revision 2. EPA 821-R-16-006, December 2016

[15] U.S. Environmental Protection Agency (USEPA). DL, MDL, PQL, What the "L" is Going On? What Does All This Alphabet Soup Really Mean?, https://19january2017snapshot.epa.gov/quality/idl-mdl-pql-what-l-going-what-does-all-alphabet-soup-really-mean_.html, accessed May 13, 2022

[16] U.S. Environmental Protection Agency (USEPA). Very Helpful Information and Frequently Asked Questions, Region 1, NPDES, Remediation. https://www3.epa.gov/region1/npdes/remediation/RGP-FREQ-021406.pdf, accessed May 13, 2022