UF UNIVERSITY of FLORIDA

Center for Environment & Human Toxicology

PO Box 110885
Gainesville, FL 32611-0885
352-392-2243 Tel
352-392-4707 Fax

January 10, 2023

Leah J. Smith
District and Business Support Program
Division of Waste Management
Florida Department of Environmental Protection
2600 Blair Stone Road
Tallahassee, FL 32399-2400

Re: Review of ProUCL Version 5.2

Dear Ms. Smith:

At your request, I have reviewed the updates to ProUCL included in Version 5.2. I contacted Dr. Phil Goodrum, a statistician and adjunct faculty member at the University of Florida for assistance with the evaluation of Version 5.2. This review includes the ProUCL Version 5.2 software, the *ProUCL Version 5.2.0 User Guide,* and the *ProUCL Version 5.2.0 Technical Guide*. The software was modified by Neptune and Company, Inc. ("Neptune") and was released June 14, 2022. Both supporting documents were prepared by Neptune and are undated but were also released June 14, 2022. The simulations presented in the document *Analysis of UCL Simulations at the Lognormal Distribution Performance of the Chebyshev UCL Estimators and Improved Recommendation Rules* were also reviewed. This document was prepared by Neptune and is dated January 6, 2022. The most significant changes to the software include changes made to the decision logic used to choose the recommended one-sided 95% upper confidence limit on the mean (95 UCL).

Background on UCL Metrics

As background, several methods exist to calculate a one-sided 95 UCL for a data set of interest and their performance in providing an accurate 95 UCL varies depending upon the population sampled (e.g., distribution shape and skewness), properties of the dataset (e.g., number of observations in the sample, frequency of detection), and the metrics used to evaluate 95 UCL method performance (e.g., likelihood that 95 UCL underestimates the true population mean; magnitude of under- or over-estimation). The basis for selecting one method over others to estimate the 95 UCL for a given data set is informed by simulation studies in which the performance of each method is examined when applied to data sets with similar properties. Simulation studies usually attempt to reproduce a range of "real world" sampling conditions by generating thousands of datasets of differing sample sizes and analytical detection limits from populations with different distribution shapes. This is done with numerical simulations that can repeatedly draw a set of observations (at random) from populations with a known distribution

shape, defined by population parameters (e.g., mean, variance). Various 95 UCL estimators are applied to the datasets and the resulting set of 95 UCLs are compared to the known arithmetic mean. These studies identify conditions under which individual methods may over- or underestimate the actual 95 UCL with observed likelihoods and magnitudes. In previous versions of ProUCL, the objective of these tests was to identify methods that, given the skewness and size of the data set of interest, produced an upper confidence limit equal to or exceeding the mean with 95% frequency. This is equivalent to stating that the probability that the UCL will underestimate the population mean (called a "Type I error") is less than or equal to 5%. The observed frequency of the UCL being greater than or equal to the mean is sometimes referred to as the "coverage" metric. Among the methods providing sufficient coverage, the method producing a result closest to the true mean may be recommended so that the resulting value is not any greater than necessary to achieve coverage. The difference between the UCL and the mean can be thought of as the "magnitude of UCL/mean ratio" metric[1]. Both metrics (coverage and magnitude) of the ratio, have been part of the decision logic for ProUCL for many years. When selecting between: a) a method that underestimates coverage but has a low magnitude difference between the UCL and mean or b) a method that almost always achieves the target coverage but can also have a high magnitude difference, professional judgment was used to achieve a balance. ProUCL 5.2 now gives greater weight to UCL methods that minimize the "magnitude" difference at the expense of the coverage (Option A above).

Recommendation to Continue Using ProUCL 5.1

As an example, the 95 UCL estimates for samples sizes less than 28 or with a standard deviation after log transformation (hereafter "SD_{log}") greater than 1.5 default to t-UCLs (i.e., Student's-t and Kaplan Meier (KM)-t tests) in ProUCL 5.2. Previously, the software recommended the Chebyshev method for lognormal (or highly skewed) datasets. The switch from Chebyshev UCLs to t-UCLs results in a lower magnitude UCL/mean ratio, as well as a loss of coverage. Based on simulations performed by Neptune, the 95 UCL estimate suggested by ProUCL Version 5.2 for a dataset with a SD_{log} in the range of 0.5 to 1.5 and a sample size of 10 (the minimum required sample size for the calculation of a 95 UCL in Chapter 62-780, F.A.C.) provides an **80% coverage rate** (e.g., Figures B5 to B7 of the Technical Guide). In other words, the 95 UCL is really an 80 UCL. The user is not warned of this, and in fact would only discover the implications of the new decision tree if they either carefully reviewed the appendix material showing simulations performed by Neptune, or did their own simulation study. In addition to the obvious potential for misunderstanding of what the ProUCL 5.2 recommended values represent, this may also create potential problems for compliance with some Florida rules — specifically those that specify that decisions should be made using the 95 UCL of the data, which presumably means an actual 95% confidence limit and not something lower labeled as a 95 UCL. Due to these concerns and others discussed below, I recommend against using the updated software. Until these concerns are addressed, ProUCL Version 5.1 should continue to be utilized for the calculation of a 95 UCL.

[1] The authors of the ProUCL 5.2 Technical Guide refer to the term "accuracy" when referring to the magnitude of the difference between the 95 UCL and the population mean; however, this terminology sets up a false goal that the upper confidence limit for the sample mean should be equivalent to the population mean. This concept is not accurate because the goal, in fact, is to yield an estimate that is at least equal to the mean, and that doesn't excessively overestimate the mean. Therefore, the ratio of the UCL/mean is more applicable in this context.

Specific Observations Regarding ProUCL 5.2

The following observations regarding the functionality of ProUCL 5.2 may inform how regulatory agencies and the risk assessment community consider options for selecting UCL methods at this time.

1. **The Chebyshev UCL is no longer recommended in ProUCL Version 5.2 for any dataset.** In earlier versions of ProUCL, certain Chebyshev UCL methods (e.g., 95%, 97.5%, or 99% Chebyshev UCL) were recommended for specific conditions, such as lognormally distributed datasets with $SD_{log} \geq 1$[2]. In ProUCL 5.2, all Chebyshev UCLs have been replaced by the Student's-t test for normal distributions and the KM t-test for non-normal distributions, even though prior simulations summarized in earlier ProUCL Technical Manuals and peer-reviewed literature clearly demonstrate the undercoverage of t-UCL methods in many real-world conditions of moderately skewed data. Therefore, skewness is essentially ignored as a distinguishing property of datasets in ProUCL 5.2. Using ProUCL on a dataset with $SD_{log} > 1.5$ will result in a decrease in coverage of the 95 UCL estimate. As stated above, simulations by Neptune (Appendix D of Technical Guide) demonstrate the coverage decreases to 80% for skewed datasets with a low sample size.

2. **The ProUCL 5.2 decision logic (summarized in Appendix C) no longer considers options for changing the confidence coefficient to improve the coverage.** In prior versions of ProUCL, the decision logic was optimized to account for undercoverage by increasing the confidence coefficient from 95% to 97.5% or 99% or decreasing to 90%. This was even applied to Chebyshev under conditions when the Chebyshev 95 UCL failed to achieve 95% coverage. In ProUCL 5.2, only 95% coefficients are included in the decision logic, presumably in an effort to minimize the magnitude of the mean to 95 UCL ratio. Furthermore, the simulations documented in Appendix D, which appear to be the primary basis for recommending the t-UCL methods, are limited to conditions when a distribution is uncensored and lognormal (excluding normal or gamma distributions). Prior releases demonstrate that the degree of censoring (i.e., frequency of detection) contributes to uncertainty in goodness-of-fit evaluations and parameter estimation (both mean and variance) and is an important factor in evaluating overall UCL method performance. Therefore, the simulations appear to be limited to idealized conditions and are not likely to reflect most environmental scenarios.

3. **The ProUCL Version 5.2 output is confusing because it presents more UCL methods than it actually applies to the decision logic.** The full simulation output appears similar to that of ProUCL Version 5.1, giving the user the impression that each UCL is considered in the decision logic for recommending a 95 UCL. Although they are present on the output sheet, Chebyshev UCLs (including 90%, 95%, 97.5%, and 99%) are not included in the decision logic and will no longer be recommended as a 95 UCL estimate. Similarly, none of the bootstrap methods appear to be recommended. Based on the ProUCL Version 5.2 guidance, the few UCLs that are recommended are limited to t-UCLs, H-UCL, and a few of the gamma UCLs. The output sheet should clarify which 95 UCLs are potential candidates in the decision logic, and which are provided simply for informational purposes.

[2] In ProUCL 5.0 and 5.1, USEPA discontinued recommending the use of Chebyshev (MVUE) UCL for lognormally distributed datasets with mild skew (e.g., $0.5 < SD_{log} < 1.5$) (ProUCL 5.1 Technical Guide, page 93).

4. **The revised decision rules are likely to be adopted without careful scrutiny.** Although the updated documentation (e.g., Chapter 9 of the User's Guide) continues to apply a caveat stating that the recommendations are not intended to reflect policy decision, in practice, most users adopt the recommendations, trusting that decision rules have been thoughtfully considered. Important choices regarding statistical methods are included throughout the software functionality - the guidance specifies background tests to use for hypothesis testing, the methodology for identifying hotspots, the methodology for identification of contaminants of potential concern and non-compliance in monitoring wells, and recommends UCLs where 95% coverage of the mean is not the primary factor in choosing a 95 UCL. The result is that ProUCL Version 5.2 appears to promote specific policy decisions through its available options and decision logic rather than the accurate calculation of statistics based on the characteristics of the environmental dataset.

5. **The User's Guide states that these recommendations are based on the professional experience of the developers of the software.** These recommendations are not accompanied by models or simulations that support the changes. Until such simulations are conducted, peer-reviewed, and appropriately presented for users, I do not recommend adopting the proposed changes including the decision logic for 95 UCL selection.

As requested, I have reviewed the reference citations, tables, figures, and Table of Contents, for accuracy. All of these elements were correctly represented in the document. Typographical, formatting, or other editorial errors (if any) were noted above. Conclusions and recommendations were inherent in the document and are addressed in the above comments. Please let me know if you have any questions regarding this review.

Sincerely,

Leah D. Stuchal, Ph.D. .