PERFORMANCE EVALUATION OF UCL ESTIMATION METHODS:

WHEN DATA ARE NOT NORMAL AND NOT LOGNORMAL

November 10, 2013

Prepared for:
Michigan Department of Environmental Quality
Remediation and Redevelopment Division

Prepared By:
Kern Statistical Services, Inc.
5175 NE River Rd.
Sauk Rapids, MN 56379

Introduction

The Michigan Department of Environmental Quality (MDEQ) provides statistical guidance and analysis tools supporting Part 201 Cleanup Criteria of the Natural Resources and Environmental Protection Act, 1994 PA 451, as amended in the Sampling Strategies and Statistics Training Materials (S3TM; MDEQ, 2002). The S3TM includes recommendations for analysis of data that are normally or lognormally distributed. Unfortunately, analytical data are often not distributed normally or lognormally, and modern statistical methods for this situation are not well understood, particularly for small sample sizes. Methods have been proposed, with promising candidates including relatively new, non-parametric procedures known as bootstrapping (Efron, 1982; Efron and Tibshirani, 1993) as well as parametric methods based on maximum likelihood and generalized linear models (McCullagh and Nelder, 1984). These parametric methods extend beyond the normal and lognormal models to include the gamma distribution, which is a more flexible probability model.

As early as the 1970s, other specialized methods for estimating confidence limits for skewed distributions were also proposed (Grice and Bain, 1980; Parkin et al., 1990; Hall, 1992; Chen, 1995) but little consensus was developed with regard to how best to select amongst them, and their robustness to departures from underlying assumptions is not well known. Starting in the late 1990s, United States Environmental Protection Agency (U.S. EPA) began efforts to understand the pros and cons of the plethora of methods that had been proposed in efforts to develop coherent guidance for estimating confidence limits for right-skewed distributions (U.S. EPA, 2002). As part of these research efforts, U.S. EPA developed software (ProUCL©; U.S. EPA 2004 and 2007) implementing at least 15 methods for estimating upper confidence limits (UCLs) for the mean (U.S. EPA, 2007). Singh et al. (2002) and Singh and Singh (2003) provide recommendations for selection among the 15 methods contingent on the observed characteristics of the sample data. Their recommendations are included in ProUCL© output to help practitioners to select appropriate methods based on underlying statistical distributions. However, in many instances when studies are conducted to support compliance demonstrations under Part 201, sample sizes may be too small to reliably distinguish between lognormal, gamma, and other right-skewed distributions, making it potentially difficult to follow these recommendations.

U.S. EPA recommendations are derived from simulation studies based primarily on lognormal, normal, or gamma distributions (Singh et al., 2002 and 2003). Few studies have considered performance based on field data from actual empirical distributions, so the robustness of common UCL procedures is largely unknown. Because data often do not conform to standard theoretical statistical distributions - even the highly flexible gamma - it is important to understand the robustness of proposed statistical methods. In this study, data from several large field sampling programs are used to test the performance of 16 methods for calculating UCLs with 11 contaminant distributions, nine of which originated from field studies. The results of these simulations are summarized in this document.

The primary objective of this study was to test the robustness of several UCL methods when underlying distributions cannot be well characterized by normal or lognormal theoretical distributions, or when the true underlying distribution cannot be easily determined.

Methods

Sample analytical data from sediment, soil, and fish from nine empirical and two theoretical distributions were used to test the robustness of 16 UCL estimation methods. Data used for these simulations were from studies conducted at large contaminated sites in the northeastern United States. They included a study of Total DDT in four floodplain areas (Floodplain DDT 1, 2, 4 and 5), a subset of the four floodplain areas (Floodplain DDT 2, 4 and 5), Total Polychlorinated Biphenyls (PCBs) in fish tissue samples (Fish PCBs), manganese concentration in residential soils (Soil Manganese), Total PCBs in sediment from three river reaches (Sediment PCBs 1, 2 and 3), arsenic and chromium concentrations in soil (Soil Arsenic and Soil Chromium, respectively) and two theoretical populations based on the lognormal distribution (Simulated Lognormal and Simulated Trucated Lognormal). The 11 populations and the data used in this study are further described in Table 1. The cumulative distribution functions (CDFs) are plotted on Figure 1 and the population distributions are described in Table 2. These distributions were sub-sampled with varying sample sizes, and UCLs were estimated using 16 methods (Table 3). The results were compared to the actual means of the full population distributions to evaluate precision and accuracy of estimates of the UCL. This random sampling process was repeated 5,000 times and the performance of each method was summarized.

Test Data

Sample analytical data were gathered from contaminated sediment and soil from Superfund sites. As mentioned above, analytical measurements included PCBs in river sediment and fish tissue, Total DDT, manganese in soil from a residential area, and arsenic and chromium in soil. Additionally, lognormal and truncated lognormal distributions were studied. The truncated lognormal distribution was included because contaminant distributions often appear to be lognormally distributed, yet do not have the infinite upper tail that is characteristic of the lognormal and other theoretical probability distributions. The truncated lognormal distribution was generated by simulating a lognormal distribution and dropping all values above a specified cut-off (in this case, 50). Also in this case, parameters were chosen from the maximum likelihood estimation (MLE) estimates for the Fish PCBs data set. The maximum value was also chosen to mimic the maximum observed value in the Fish PCBs data.

Figure 1 illustrates that the distributions tend to be right skewed as is characteristic of environmental contaminant data. This characteristic can also be seen in the summary statistics shown in Table 2 in which mean values consistently exceed medians, and skewness values are generally large and positive, indicative of right-skewed data.

UCL Estimation Methods

Sixteen methods for estimating UCLs for the mean were evaluated in this study (Table 3). The methods were chosen based primarily on those methods proposed in the literature, most of which are available within the ProUCL© software package developed by U.S. EPA. Several of the methods were not expected to perform well for skewed, non-normally distributed data (e.g., the normal theory method), but were included to help provide a basis for comparison with other methods. ProUCL© has become one of the most commonly used software tools at contaminated sites where statistical procedures are used to estimate exposure point concentrations (EPCs). In general, these methods are used in screening level and baseline risk assessments where more sophisticated statistical methods either are not needed or cannot be justified given project size and complexity.

Generally, practitioners rely on ProUCL© outputs to provide guidance that would otherwise be provided by a professional statistician. The ProUCL© method includes a series of distributional tests to develop a recommendation regarding the most appropriate UCL method. We dubbed this approach *Distribution Dependent Method* (DDM) *.* In our simulation study, we embedded the primary features of this ProUCL© logic for method selection into the computer code and tested the efficacy of this approach. U.S. EPA guidance and manuals associated with ProUCL© recommend qualitative assessments of data and plots in this process which could not be programmed into our simulation study, so it is likely that these results are not exactly what one would obtain using ProUCL©, but a reasonable approximation.

The general idea of this approach is to use the statistics of observed data to select a method with assumptions expected to be consistent with the inferred distribution. The selected UCL method is then applied based on the results of the distribution testing. This approach is intuitively appealing and is consistent with recommended approaches in the statistical literature.

A complete list of the methods we tested is provided in Table 3.

Performance Metrics

The purpose of the 95% UCL is to use sample data to provide a bound such that there is 95% confidence that the true population mean is less than that bounding value. This indicates that if many samples of a particular sample size are selected from the sample population, the UCL should exceed the true population mean in 95% of those samples. This proportion of samples in which the estimated UCL exceeds the true mean is called the method coverage rate (MCR). If the method is working well, the MCR should be close to the nominal or “advertised” confidence level, in this case 95%. In addition to coverage rate, it is also desirable for a method to be precise. A method would be considered precise if an estimated UCL varies by a small amount from one sample to the next. If a method provides the correct coverage probability and is also precise, then the estimated UCL will tend to be too high or too low by only a small amount. An imprecise method with the proper coverage-probability would tend to be too high or low at the proper frequency, but errors would be larger in magnitude than those of the more precise method.

Simulation Methods

To evaluate the performance of statistical methods, it is common practice to simulate a number of values from a theoretical distribution with known properties (i.e. known mean and variance) and then apply the proposed statistical procedure to the simulated sample data. Table 3 provides a summary of studies that have done this. The procedure is repeated many times and the estimates are compared with the true population parameters. In this case, the frequency with which the UCL exceeds the true population mean is of primary interest. For the most part, these studies have considered either normal, lognormal or gamma distributions. Studies based on actual distributions from field data have not been available.

In this study, actual data from contaminated sites were treated as if they were true populations and the re-sampling procedure described above was implemented. The arithmetic average of the empirical data was treated as the true population mean, samples were selected randomly with replacement, and UCLs were calculated for each sample. The procedure was repeated 5,000 times for samples of size 9, 15, 20, 25 and 30, and the coverage rate was calculated for each combination of the 16 UCL methods, 11 distributions, and 5 sample sizes.

Results

Compliance demonstrations under Part 201 are often based on relatively small samples of size N=9. At more complex sites, samples of size 15 to 30 are also relatively common. To evaluate preferred methods for Part 201 compliance demonstrations, UCL method performance was studied for samples of size 9, 15, 20, 25, and 30. Results from this study provide an understanding of the robustness of the 16 methods for small sample sizes and populations that are not normally and not lognormally distributed. Some results are most easily interpreted by comparing method performance across distributions, and other results are more easily interpreted by comparing performance within each distribution. Therefore, results are presented graphically in two ways: 1) grouped by method (Figures 2-1 and 2-2) and 2) grouped by distribution (Figures 3 through 13).

Population Distributions

Figure 1 provides a graph of the distributions of each population used in the simulation study. The horizontal axis represents concentration and is scaled in log-scale, so that populations that are close to lognormally distributed plot symmetrically, whereas those that differ from lognormal distributions do not. The PCB concentration distributions from river Sediment PCBs 1, 2 and 3 are close to lognormally distributed, whereas the Floodplain DDT distributions differ substantially from a lognormal distribution. The Soil Manganese data is unusual in that there was a single observation that was much lower than the median concentration, but it was found that this single sample value did not influence performance of UCL estimates.

Results Grouped By Method

For each of the 16 methods, Figures 2-1 and 2-2 provide a graphical comparison of the MCR to the nominal 95% level across the range of sample sizes and each of the 11 distributions. Methods are considered robust to distributional assumptions when the MCR is close to the 95% nominal level within a particular sample size. It can be seen that robustness of methods varied among distributions and with sample size.

Figures 2-1 and 2-2 each contain plots of the coverage probabilities for 8 of the 16 statistical methods evaluated for each of the 11 distributions. Each plot illustrates the coverage provided by the identified method and sample sizes ranging from 9 to 30 selected from each of the 11 populations. As expected, coverage increases for all methods and all populations as sample size increases.

Samples of Size N=9

Samples of size N=9 are of primary focus because many Part201 compliance demonstrations are based on nine samples per exposure unit.

For samples of size N=9, it can be seen that only the sample maximum, the 97.5% and 99% Chebyshev method, and Land's H (lognormal) methods provided adequate coverage for the majority of populations. Of these, Land's H and the 99% Chebyshev methods were overly conservative providing nearly 100% coverage for most distributions. Perhaps surprisingly, the relatively simple sample maximum and the 97.5% Chebyshev methods most closely matched the nominal 95% coverage rates for the majority of distributions.

For other methods, coverage rate is well below the nominal 95% for small sample sizes (marked by the horizontal red line on the Figures).

Samples of size N=15 to 30.

For larger sample sizes, coverage rates improved for all distributions, but only Hall's bootstrap, Bootstrap t, and Cox's t methods consistently provided coverage rates close to the nominal 95% level without being overly conservative. For these methods, coverage rates were less variable across distributions than other methods, and the method did not break down for any particular distribution. The Cox t method performed similarly to the Bootstrap t method for all populations except the Soil Manganese data, for which the method under covered (inadequately represented the mean) regardless of sample size. Cox's t method did not provide proper coverage for any sample size for the Soil Manganese data set.

Results Grouped by Distribution

Distributions

Figures 3 through 13 provide detailed descriptions of each of the distributions, a graphical comparison between the study population and fitted theoretical distributions, and the performance of each UCL method for samples of size 9, 15 and 30. These figures provide an indication of the bias and variability of the methods for each distribution, and also show how method variability and bias vary with increasing sample size.

In general, it can be seen that distributions were all right skewed and would often be treated as if they were lognormally distributed. Performance of UCL methods varied with distributions and number of samples. No single method provided satisfying coverage rates for all distributions, and UCLs for some distributions were approximated poorly by nearly all methods. Coverage rates for Floodplain DDT from areas 1, 2, 4, and 5 were low for nearly all methods, particularly for smaller sample sizes. This may be due to the fact that these data represent a mixture of highly and moderately contaminated areas. When DDT from floodplain areas 2, 4 and 5 - which are all moderately contaminated - were considered as a single population, most methods performed reasonably well, particularly for moderate to large sample sizes.

Figures 3 and 4 summarize the performance of each method for the Soil Chromium and Soil Arsenic populations, respectively. For these populations it can be seen that the Bootstrap t and Cox t methods performed best for samples of size 9. For samples of size 15 and 30, the DDM and the Bootstrap t and Cox's t performed similarly. Chebyshev's method was overly conservative for these data because they are nearly normally distributed.

Figure 5 summarizes method performance for the Floodplain Total DDT population for floodplain areas 2, 4, and 5, which are similarly distributed and can be considered to represent a single statistical population. It can be seen that for this population, Hall's bootstrap method outperformed other methods for samples of size 9, 15 and 30. The other bootstrap methods also performed reasonably well, while the DDM tended to be overly conservative for sample sizes of 15 and 30.

Figure 6 shows that Land's H method provided proper coverage probability for N=9, while all other methods grossly under covered the true mean. At the same time, Land's H also was highly variable, resulting in many estimates that were orders of magnitude larger than the highest observed concentrations. For larger sample sizes, the bootstrap methods performed as well or better than others.

Figures 7, 8 and 9 provide results for three Sediment PCB populations (1, 2, and 3), where it can be seen that the field data are very consistent with both theoretical lognormal and gamma distributions. In spite of this apparent agreement with the theoretical distributions, the gamma method tended to under cover the true mean, whereas Land's H method for lognormal data tended to over cover substantially. For samples of size N=9, the sample maximum provided the best coverage rates, while for N=15 and 30, the Bootstrap t and Cox t methods performed best. The DDM approach tended to under cover substantially.

Figure 10 shows relatively good agreement between the Soil Manganese data and gamma and lognormal distributions. Again, in spite of this apparent similarity to these distributions, the gamma and Land's H methods do not perform well for large or small sample sizes. For N=9 and 15, the sample maximum and the 97.5% Chebyshev's method provide the best coverage rates. For N=30, all methods under cover but the two bootstrap methods perform substantially better than other methods, with 94% coverage rates.

In Figure 11, it can be seen that the Fish PCB concentrations are reasonably approximated by either gamma or lognormal distributions. In spite of this apparent agreement with the theoretical distributions, the sample maximum and Cox's t outperform Land's H and the gamma methods for smaller sample sizes. For larger sample sizes, the Bootstrap t method outperforms all other methods. Because the DDM typically selected parametric methods the DDM also performed poorly.

Figure 12 summarizes the results for methods applied to samples from a true underlying lognormal distribution. As expected, the Land's H method provides the best coverage rate. All other methods, with the exception of the 97.5% Chebyshev method, under covered the true population mean. It is notable that when the underlying population was truly normally distributed, only Land's H provided adequate coverage. Figure 13 provides results for a moderately censored lognormal distribution for which the bootstrap methods perform similarly to the Land's H method for samples of size 15 and 30. For N=9, Land's H method and Chebyshev's method both provided similar coverage rates near to the 95% nominal level. All other methods, including the sample maximum, under covered the true population mean.

Distribution Dependent Method

The performance of the DDM is of particular interest because this approximates the default approach recommended by U.S. EPA for ProUCL© users. The primary determining step in the DDM approach is to identify a statistical distribution to which the sample data are consistent. Figure 14 summarizes the results of these distributional tests and shows that for the 11 populations tested, the gamma distribution was selected in the majority of simulated samples for all but the Soil Arsenic and Soil Chromium populations. This can also be seen in Table 4 where the results of the DDM distributional testing are summarized. Therefore, the Gamma UCL method would be selected as the preferred method the majority of the time for all but these two distributions. Referring back to Figure 2-2, it can be seen that the Gamma and Adjusted Gamma methods both performed poorly for most distributions. Ironically, the Gamma methods performed well for the Soil Arsenic and Soil Chromium populations, although the DDM approach did not identify them as candidates for the Gamma methods.

The DDM method performance also varied with sample size. It can be seen that for N=9, the method resulted in selection of either the normal or gamma distributions; whereas for larger sample sizes, the normal distribution was almost never selected and the gamma model was selected the most frequently in spite of its tendency to under cover the true population mean. This is primarily the consequence of the low power of the tests for normality for small sample sizes. It can be seen that even for highly skewed populations - such as Total Sediment PCBs, Fish PCBs and the Soil Manganese data - the normal distribution was incorrectly selected in approximately 30% of simulations.

Figure 15 and Table 5 show the methods selected by the DDM approach. It can be seen that the testing procedure tended to default toward selection of parametric methods that in general do not perform as well as the nonparametric bootstrap procedures. It can be seen that the bootstrap Hall procedure, which was one of the better performing methods, was rarely selected by the DDM approach

Discussion

Statistical methods have evolved over the period of time in which U.S. EPA has been investigating performance of the large number of UCL methods that are available. Early on, methods based on bootstrapping were perceived as difficult to execute and their reliability was not well understood, particularly for small sample sizes. Since the 1990s, new bootstrap methods, such as Hall's method and the Studentized bootstrap methods, have become increasingly available and, with increasing capacity of computers, more accessible through a host of software packages. The logic deployed as part of ProUCL© reflects earlier assumptions that a well matched parametric method would be preferred and better accepted than less familiar bootstrap procedures. However, these simulation studies indicate that, in most situations, this *default* reliance on sample statistics to select a preferred parametric method may not be the best approach for data from actual field sampling programs. Rather, based on these 11 distributions, it appears that for larger sample sizes from 15 to 30 it may be preferable to apply the Hall's bootstrap or the Bootstrap t approaches irrespective of the apparent distributional characteristics.

It can be seen from Table 2 that 10 of 11 distributions were statistically different from normal, lognormal and gamma distributions; yet, sample statistics based on small sample sizes routinely identified distributions as normal, gamma or lognormal. It is recognized that several of these distributions were close in form to the gamma or lognormal; nonetheless, small deviations from these theoretical distributions caused parametric UCL procedures to perform poorly. Importantly, the gamma distribution under covered the mean significantly, whereas results from Land's H method were far too conservative and even erratic. Confidence limits were commonly more than an order of magnitude higher than the largest values in the populations. In fact, the only situation where Land's H covered adequately was for the simulations based on the true underlying lognormal distribution. It is interesting that no other method provided adequate coverage for the lognormal distribution.

For samples of size 9, only the sample maximum or the 97.5% Chebyshev inequality provided reasonable coverage for most distributions. Other methods tended to under cover the mean, which would tend to understate EPCs and, hence, estimated risks.

For larger sample sizes, the Hall's bootstrap and the Bootstrap t methods most consistently covered the true mean at levels near the nominal 95% rate. Interestingly, the variant of Cox's method (Cox's t) proposed by Remington (2003) was also a strong competitor. An advantage of Cox's t method is that it does not require bootstrapping.

It is recognized that this study is based on a relatively small number of distributions and may not be representative of all situations; however, it is thought that these distributions are not particularly unusual in their statistical characteristics and are likely to be representative of commonly encountered situations. Importantly, most of the 11 distributions were very similar to lognormal or gamma distributions, which are often assumed to be applicable in environmental studies; however, the typical parametric methods did not perform well on the majority of these distributions.

Recommendations

Methods

Based on this simulation study, Chebyshev's method adjusted to a confidence level of 97.5% is recommended for calculating UCLs for the mean when sample size is 15 or lower. For samples of size N=9, the sample maximum is also recommended. For larger sample sizes (N > 15), the Bootstrap t or Hall's bootstrap methods are recommended options for calculating UCLs for the mean. The gamma method and the DDM approach are not recommended for data appearing to be similar to distributions tested in this study. These simulations suggest that the more complex logic of the DDM may be counterproductive because identification of proper data distributions is ineffective for data that are close to lognormal as is frequently the case. This study shows that for modest sample sizes ranging from 9 to 30 it is better to forego the distributional testing and associated ProUC©L logic in favor of application of the sample maximum, Chebyshevs's method or bootstrap procedures, which tend to outperform for the majority of distributions.

Software

The Chebyshev, Hall's bootstrap and Bootstrap t methods are available in ProUCL© Version 4.00.02 (U.S. EPA, 2007). Although the DDM logic provided by ProUCL© is not recommended; however, ProUCL© can be used to perform calculations necessary to implement selected UCL methods. Essentially, any well-documented and tested software package offering Chebyshev's method or Hall's bootstrap or Bootstrap t methods should be considered acceptable.

These evaluations were based on data from seven large and two relatively small field studies. This evaluation was large relative to other studies that have been conducted yet, it is recognized that UCL method performance may vary for data from other field studies. It is recommended that additional simulation studies be conducted with additional contaminant distributions from other large field studies. Ultimately, it would be useful to develop a catalog of field studies and UCL recommendations, which could be used to select UCL methods empirically.

References

Chen, L. 1995. Testing the mean of skewed distributions. *Journal of the American Statistical Association* 90: 767-772.

Efron, B. 1982. *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM, Philadelphia, Pennsylvania.

Efron, B. and Tibshirani, R.J. 1993. *An introduction to the bootstrap*. London: Chapman and Hall.

Grice, J.V., and L.J. Bain. 1980. Inferences concerning the mean of the gamma distribution. *Journal of the American Statistical Association* 75(372):929-933.

Hall, P. 1988. *Theoretical comparison of bootstrap confidence intervals*. Annals of Statistics, 16, 927-953.

Hall, P. 1992. On the removal of skewness by transformation. *Journal of the Royal Statistical Society, Series B*. 54: 221-228.

Hogg, R.V., and Craig, A. 1995. *Introduction to Mathematical Statistics*; 5th edition. Macmillan Publishing Company.

Hogg, R.V., and Craig, A.T. 1978. *Introduction to Mathematical Statistics*, New York: Macmillan Publishing Company.

Johnson, N.J. 1978. Modified t-tests and confidence intervals for asymmetrical populations. *Journal of the American Statistical Association*, 73:536-544.

Land, C.E. 1971. Confidence intervals for linear functions of the normal mean and variance. *Annals of Mathematical Statistics*, 42: 1187-1205.

Land, C. E. 1975. *Tables of Confidence Limits for Linear Functions of the Normal Mean and Variance*. In Selected Tables in Mathematical Statistics, Vol. III, American Mathematical Society, Providence, R.I., pp. 385-419.

Lefante, J.J., Jr., and A.K. Shah. 2002. Robustness properties of lognormal confidence intervals for lognormal and gamma distributed data. *Communications in Statistics: Theory and Methods*, 31(11):1939-1957.

MDEQ, 2002. DEQ Sampling Strategies and Statistics Training Materials for Part 201 Cleanup Criteria. August 2002.

MDEQ, 2004. Part 201 Operational Memorandum No. 1. December 2004.

Manly, B.F.J. 1997. *Randomization, Bootstrap, and Monte Carlo Methods in Biology*. Second Edition. Chapman Hall, London.

McCullagh, P. and J.A. Nelder. 1984. *Generalized Linear Models, Second Edition*. *Monographs on Statistics and Applied Probability 37.* Chapman and Hall/CRC. New York.

Parkin, T.B., S.T. Chester, and J.A. Robinson. 1990. Calculating confidence intervals for the mean of a lognormally distributed variable. *Journal of the Soil Science Society of America*, 54:321-326.

Remington, R.E. 2003. Robust Confidence Interval Estimation on the Lognormal Mean. Master's Thesis. College of Graduate Studies, University of Idaho.

Schulz, T.W. and S. Griffin. 1999. Estimating risk assessment exposure point concentrations when the data are not normal or lognormal. *Risk Analysis* 19:577-584.

Singh, A.K., A. Singh, and M. Engelhardt. 1997. The lognormal distribution in environmental applications. EPA/600/R-97/006.

Singh, A, Singh, A.K., and Laci, R.J. 2002. *Estimation of the Exposure Point Concentration Term Using a Gamma Distribution*. EPA/600/R-02/084.

Singh, A. and Singh, A.K. 2003. Estimation of the Exposure Point Concentration Term (95% UCL) Using Bias-Corrected Accelerated (BCA) Bootstrap Method and Several other methods for Normal, Lognormal, and Gamma Distributions. Draft EPA Internal Report.

Sutton, D.D. 1993. Computer-Intensive Methods for Tests About the Mean of an Asymmetrical Distribution. *Journal of the American Statistical Association*. 88, 423:802-810.

U.S. EPA. 1992. *Supplemental Guidance to RAGS: Calculating the Concentration Term*. OSWER 9285.708I.

U.S. EPA. 2002. *Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites*. OSWER 9285.6-10.

U.S. EPA. 2004. *ProUCL Version 3.1, a Statistical Software*, National Exposure Research Lab, EPA, Las Vegas Nevada. ProUCL 3.0 can be freely downloaded from the EPA website: http://www.epa.gov/nerlesd1/tsc/tsc.htm.

U.S. EPA. 2007. *ProUCL Version 4.00.02 User Guide*. EPA/600/R-07/038.

Wong, A. 1993. *A Note on Inference for the Mean Parameter of the Gamma Distribution*. Statistics Probability Letters, Vol. 17, 61-66.

Zhou, X.H. and Gao S. 1997. Confidence intervals for the lognormal mean. *Statistics in Medicine*. 16:783-790.

Zhou, X.H. and Gao S. 2000. One-sided confidence intervals for means of positively skewed distributions. *American Statistician*. 54:100-104.

Table 1. Study Data Information

Population	Description	Contaminant	Medium
Lognormal	Based on PCB Concentrations in Fish	- -	- -
Truncated Lognormal	Based on PCB Concentrations in Fish	- -	- -
Fish PCB	Concentrations of total PCB in fish tissue	PCB	Fish
Soil Manganese	Maganese from top 3 inches of soil	Manganese	Soil
Sediment PCB 1	Total PCB in sediment from River Section 1	PCB	Surface Sediment
Sediment PCB 2	Total PCB in sediment from River Section 2	PCB	Surface Sediment
Sediment PCB 3	Total PCB in sediment from River Section 3	PCB	Surface Sediment
Soil DDT 1,2,4,5	Total DDT in Floodplain Soil	DDT	Floodplain Surface
Soil DDT 2,4,5	Total DDT in Floodplain Soil	DDT	Floodplain Surface
Soil Arsenic	Arsenic in Soil	Arsenic	Soil
Soil Chromium	Chromium in Soil	Chromium	Soil

- - = Not applicable

Table 2. Summary statistics for test distributions.

Population	Size	Minimum	Median	Maximum	Mean	SD	Skew	Mean Log	SD Log	Shapiro-Wilkes Normal p-value	Siginificantly Non-Normal?	Shapiro-Wilkes Lognormal p-value	Significan ly Non-Lognormal?	K-S Statistic (Gamma)	Approx. K-S Cri ical Value (Level 0.05)	Significantly Non-Gamma?
Theoretical Distributions																
Lognormal	- -	- -	0.9473	- -	2.5208	5.87165	10 96	-0 05	1.40	- -	- -	- -	- -	- -	- -	- -
Truncated Lognormal	- -	- -	0.9455	50	2.5345	6.25311	16.44	-0 06	1.40	- -	- -	- -	- -	- -	- -	- -
Empirical Distributions																
Total PCB Fish	1,842	0 01	1.00	49.54	2 25	3.30	4.13	-0 05	1.45	9.12E-053	YES	4.66E-013	YES	0.0784	0.030	YES
Soil Manganese	262	2 06	550 00	8,449.00	926.51	1,098.20	3.40	6.46	0.83	1.21E-024	YES	8.18E-013	YES	0.1671	0.050	YES
Sediment PCB 1	3,489	0 02	15.14	913.85	31 39	54.84	6.49	2 56	1.50	6.07E-072	YES	5.54E-024	YES	0.0585	0.030	YES
Sediment PCB 2	1,703	0 03	15.68	2,960.82	37.78	91.24	20 39	2.67	1.51	1.72E-062	YES	7.68E-013	YES	0.0777	0.030	YES
Sediment PCB 3	3,853	0 01	5.45	521.11	15.14	25.84	5.28	1.68	1.60	1.60E-071	YES	1.87E-016	YES	0.0897	0.030	YES
Soil DDT 1,2,4,5	111	0 01	0.15	24.20	1 91	4.69	3.11	-1 82	2.36	2.03E-018	YES	5.06E-006	YES	0.1789	0.097	YES
Soil DDT 2,4,5	88	0 01	0.07	1 81	0 29	0.42	1.80	-2 57	1.79	5.77E-012	YES	1.09E-006	YES	0.1762	0.100	YES
Soil Arsenic	32	2 00	4.60	9 00	4 80	1.89	0.45	1.49	0.41	0.180850536	NO	0.53891628	NO	0.0876	0.160	NO
Soil Chromium	32	3 30	11.00	26.00	11 95	6.86	0.50	2 30	0.65	0.039268512	YES	0.03897813	YES	0.1036	0.160	NO

- - = Not applicable
SD - Standard Deviation

Table 3. Methods Summary

	Method	Reference(s)
1	Sample Max	EPA (1992, 2002)
2	Adjusted Student's t	Chen (1995), EPA (2007), Johnson (1978), Sutton (1993)
3	Student's t	EPA (2007), Hogg and Craig (1978,1995)
4	Adjusted Central Limit Theorem (CLT)	Chen (1995), EPA (2007)
5	95% Chebyshev Inequality	EPA (2007), Hogg and Craig (1978,1995)
6	97.5% Chebyshev Inequality	EPA (2007), Hogg and Craig (1978,1995)
7	99% Chebyshev Inequality	EPA (2007), Hogg and Craig (1978,1995)
8	ProUCL© Distribution Dependent Recommendation	EPA (2007, Section 2.5)
9	Bias-Correct Accelerated (BCA) Percentile Bootstrap	Efron and Tibshirani (1993), EPA (2007), Manly (1997)
10	Hall's Bootstrap	EPA (2007), Hall (1992), Manly (1997)
11	Bootstrap t	Efron (1982), EPA (2007), Hall (1988)
12	Cox t	Remington (2003), Zhou and Gao (1997)
13	Land's H	EPA (2007), Land (1971, 1975)
14	Approximate Gamma	EPA (2007), Grice and Bain (1980), Wong (1993)
15	Adjusted Gamma	EPA (2007), Grice and Bain (1980), Wong (1993)
16	Parkin	Parkin et al. (1990)

Methods incorporated in ProUCL© but not evaluated in this study include:
- Central Limit Theorem
- Jackknife
- Chebyshev inequality using lognormal MVUEs
- Standard Bootstrap
- Percentile Bootstrap

Table 4. Distribution Classification Summary

Population	Sample Size	Normal	Gamma	Lognormal	Indeterminate
Simulated Lognormal	9	24.6%	70.4%	2.4%	2.6%
	15	4.2%	85.7%	7.9%	2.2%
	20	1.6%	81.7%	13.5%	3.2%
	25	0.5%	72.7%	24.5%	2.3%
	30	0.0%	65.5%	31.7%	2.8%
Simulated Truncated Lognormal	9	29.4%	66.6%	1.9%	2.1%
	15	6.1%	85.2%	7.3%	1.4%
	20	0.8%	84.3%	13.6%	1.3%
	25	0.3%	78.1%	20.2%	1.4%
	30	0.0%	69.3%	28.8%	1.9%
Sediment PCB 1	9	36.1%	60.8%	1.3%	1.8%
	15	10.3%	84.4%	3.1%	2.2%
	20	2.9%	89.1%	6.3%	1.7%
	25	0.8%	88.0%	9.0%	2.2%
	30	0.1%	86.9%	10.1%	2.9%
Sediment PCB 2	9	30.7%	65.7%	2.0%	1.6%
	15	6.1%	88.6%	4.2%	1.1%
	20	1.2%	88.8%	8.0%	2.0%
	25	0.2%	87.7%	10.2%	1.9%
	30	0.0%	84.9%	12.6%	2.5%
Sediment PCB 3	9	22.8%	73.4%	1.9%	1.9%
	15	2.8%	93.1%	3.2%	0.9%
	20	1.0%	90.5%	7.5%	1.0%
	25	0.0%	88.2%	10.8%	1.0%
	30	0.0%	79.9%	18.7%	1.4%
Total PCB Fish	9	32.5%	65.2%	0.8%	1.5%
	15	6.5%	89.0%	2.8%	1.7%
	20	1.5%	91.9%	5.0%	1.6%
	25	0.1%	90.4%	7.7%	1.8%
	30	0.0%	86.3%	12.4%	1.3%
Soil Arsenic	9	88.1%	9.5%	0.7%	1.7%
	15	80.1%	15.1%	1.7%	3.1%
	20	70.3%	22.4%	2.6%	4.7%
	25	59.3%	30.1%	3.7%	6.9%
	30	49.4%	35.5%	4.6%	10.5%
Soil Chromium	9	83.9%	13.5%	0.1%	2.5%
	15	69.9%	25.3%	0.4%	4.4%
	20	56.8%	34.5%	0.1%	8.6%
	25	40.4%	44.7%	0.9%	14.0%
	30	26.3%	50.9%	0.7%	22.1%
Soil Manganese	9	32.3%	49.8%	2.8%	15.1%
	15	11.6%	50.9%	10.7%	26.8%
	20	4.5%	42.4%	17.3%	35.8%
	25	1.4%	30.2%	20.3%	48.1%
	30	0.2%	20.4%	20.5%	58.9%
Floodplain Total DDT 1,2,4,5	9	5.0%	60.0%	21.0%	14.0%
	15	0.1%	61.3%	14.3%	24.3%
	20	0.0%	44.9%	16.5%	38.6%
	25	0.0%	25.2%	18.8%	56.0%
	30	0.0%	15.2%	13.7%	71.1%
Floodplain Total DDT 2,4,5	9	13.3%	63.5%	4.4%	18.8%
	15	0.5%	69.9%	2.0%	27.6%
	20	0.4%	58.4%	1.0%	40.2%
	25	0.0%	43.8%	0.4%	55.8%
	30	0.0%	30.0%	0.4%	69.6%

Table 5. Methods Classification Summary

Poulation	Sample Size	Chebyshev 0.95	Chebyshev 0.975	Chebyshev 0.99	Adjusted Gamma	Approximate Gamma	Student's t	Adjusted Student's t	Land's H	Bootstrap Hall
Simulated Lognormal	9	1.2%	0.0%	3.5%	35.9%	34.5%	24.6%	0.0%	0.1%	0.2%
	15	5.6%	0.0%	4.4%	25.0%	60.7%	4.2%	0.0%	0.1%	0.0%
	20	9.0%	4.2%	3.2%	15.8%	65.9%	1.6%	0.0%	0.2%	0.1%
	25	0.0%	9.1%	2.2%	8.3%	64.4%	0.5%	0.0%	15.4%	0.1%
	30	0.0%	10.9%	2.8%	5.4%	60.1%	0.0%	0.0%	20.8%	0.0%
Simulated Truncated Lognormal	9	1.8%	0.0%	2.0%	33.4%	33.2%	29.4%	0.0%	0.2%	0.0%
	15	4.9%	0.0%	3.5%	20.3%	64.9%	6.1%	0.0%	0.3%	0.0%
	20	9.2%	3.8%	1.4%	12.9%	71.4%	0.8%	0.0%	0.5%	0.0%
	25	0.1%	4.9%	1.3%	10.0%	68.1%	0.3%	0.0%	15.3%	0.0%
	30	0.0%	7.1%	1.9%	3.9%	65.4%	0.0%	0.0%	21.7%	0.0%
Sediment PCB 1	9	0.9%	0.0%	1.7%	34.3%	26.5%	36.1%	0.0%	0.1%	0.4%
	15	2.2%	0.0%	2.8%	22.1%	62.3%	10.4%	0.0%	0.1%	0.1%
	20	3.9%	1.9%	1.6%	14.3%	74.8%	3.0%	0.0%	0.4%	0.1%
	25	0.0%	3.7%	2.0%	9.2%	78.8%	0.8%	0.0%	5.3%	0.2%
	30	0.1%	4.0%	3.0%	8.6%	78.3%	0.1%	0.0%	5.8%	0.1%
Sediment PCB 2	9	0.8%	0.0%	2.0%	35.7%	30.0%	30.7%	0.0%	0.5%	0.3%
	15	2.8%	0.0%	2.0%	26.2%	62.4%	6.1%	0.0%	0.3%	0.2%
	20	5.3%	2.3%	1.9%	17.4%	71.4%	1.2%	0.0%	0.2%	0.3%
	25	0.0%	3.0%	2.0%	12.2%	75.5%	0.2%	0.0%	7.0%	0.1%
	30	0.0%	4.0%	2.4%	8.3%	76.6%	0.0%	0.0%	8.6%	0.1%
Sediment PCB 3	9	1.2%	0.0%	2.4%	46.6%	26.8%	22.8%	0.0%	0.2%	0.0%
	15	1.2%	0.0%	2.7%	39.5%	53.6%	2.8%	0.0%	0.1%	0.1%
	20	3.5%	3.8%	1.2%	32.8%	57.7%	1.0%	0.0%	0.0%	0.0%
	25	0.0%	7.1%	1.1%	24.5%	63.7%	0.0%	0.0%	3.6%	0.0%
	30	0.0%	9.6%	1.4%	19.4%	60.5%	0.0%	0.0%	9.1%	0.0%
Fish PCB	9	0.9%	0.0%	1.3%	36.0%	29.2%	32.5%	0.0%	0.0%	0.1%
	15	2.3%	0.0%	1.8%	19.7%	69.3%	6.6%	0.0%	0.3%	0.0%
	20	3.3%	1.7%	1.5%	11.8%	80.1%	1.5%	0.0%	0.1%	0.0%
	25	0.0%	1.9%	1.8%	7.3%	83.1%	0.1%	0.0%	5.8%	0.0%
	30	0.1%	2.4%	1.2%	4.5%	81.8%	0.0%	0.0%	10.0%	0.0%
Soil Arsenic	9	0.4%	0.0%	0.0%	0.0%	9.5%	88.1%	2.0%	0.0%	0.0%
	15	0.2%	0.0%	0.0%	0.0%	15.1%	80.6%	4.1%	0.0%	0.0%
	20	0.0%	0.0%	0.0%	0.0%	22.4%	71.4%	6.2%	0.0%	0.0%
	25	0.1%	0.0%	0.0%	0.0%	30.1%	61.5%	8.3%	0.0%	0.0%
	30	0.0%	0.0%	0.0%	0.0%	35.5%	53.7%	10.8%	0.0%	0.0%
Soil Chromium	9	1.6%	0.0%	0.0%	0.0%	13.5%	84.0%	0.9%	0.0%	0.0%
	15	3.4%	0.0%	0.0%	0.0%	25.3%	70.8%	0.1%	0.4%	0.0%
	20	7.2%	0.0%	0.0%	0.0%	34.5%	58.1%	0.1%	0.1%	0.0%
	25	11.0%	0.0%	0.0%	0.0%	44.7%	43.3%	0.2%	0.8%	0.0%
	30	17.5%	0.0%	0.0%	0.0%	50.9%	30.9%	0.0%	0.7%	0.0%
Soil Manganese	9	11.0%	0.0%	2.6%	1.2%	48.6%	32.3%	0.6%	2.3%	1.4%
	15	22.5%	0.0%	4.8%	0.0%	50.9%	11.6%	1.3%	8.8%	0.1%
	20	29.8%	0.0%	6.6%	0.0%	42.4%	4.5%	1.0%	15.7%	0.0%
	25	38.4%	0.0%	9.4%	0.0%	30.2%	1.4%	0.8%	19.8%	0.0%
	30	48.6%	0.0%	10.1%	0.0%	20.4%	0.2%	0.6%	20.1%	0.0%
Floodplain DDT 1,2,4,5	9	0.1%	0.0%	23.3%	57.6%	2.4%	5.0%	0.0%	0.0%	11.6%
	15	0.1%	0.0%	17.2%	60.1%	1.2%	0.1%	0.0%	0.0%	21.3%
	20	0.1%	1.7%	17.4%	44.0%	0.9%	0.0%	0.0%	0.0%	35.9%
	25	0.0%	2.3%	20.3%	24.6%	0.6%	0.0%	0.0%	0.0%	52.2%
	30	0.0%	2.1%	15.5%	15.0%	0.2%	0.0%	0.0%	0.0%	67.2%
Floodplain DDT 2,4,5	9	1.4%	0.0%	16.3%	57.3%	6.2%	13.6%	0.0%	0.0%	5.2%
	15	0.7%	0.0%	24.1%	57.6%	12.3%	0.6%	0.0%	0.0%	4.7%
	20	0.4%	0.6%	35.2%	44.4%	14.0%	0.5%	0.0%	0.0%	4.9%
	25	0.0%	0.3%	52.0%	28.6%	15.2%	0.0%	0.0%	0.1%	3.8%
	30	0.0%	0.3%	65.6%	17.2%	12.8%	0.0%	0.0%	0.1%	4.0%

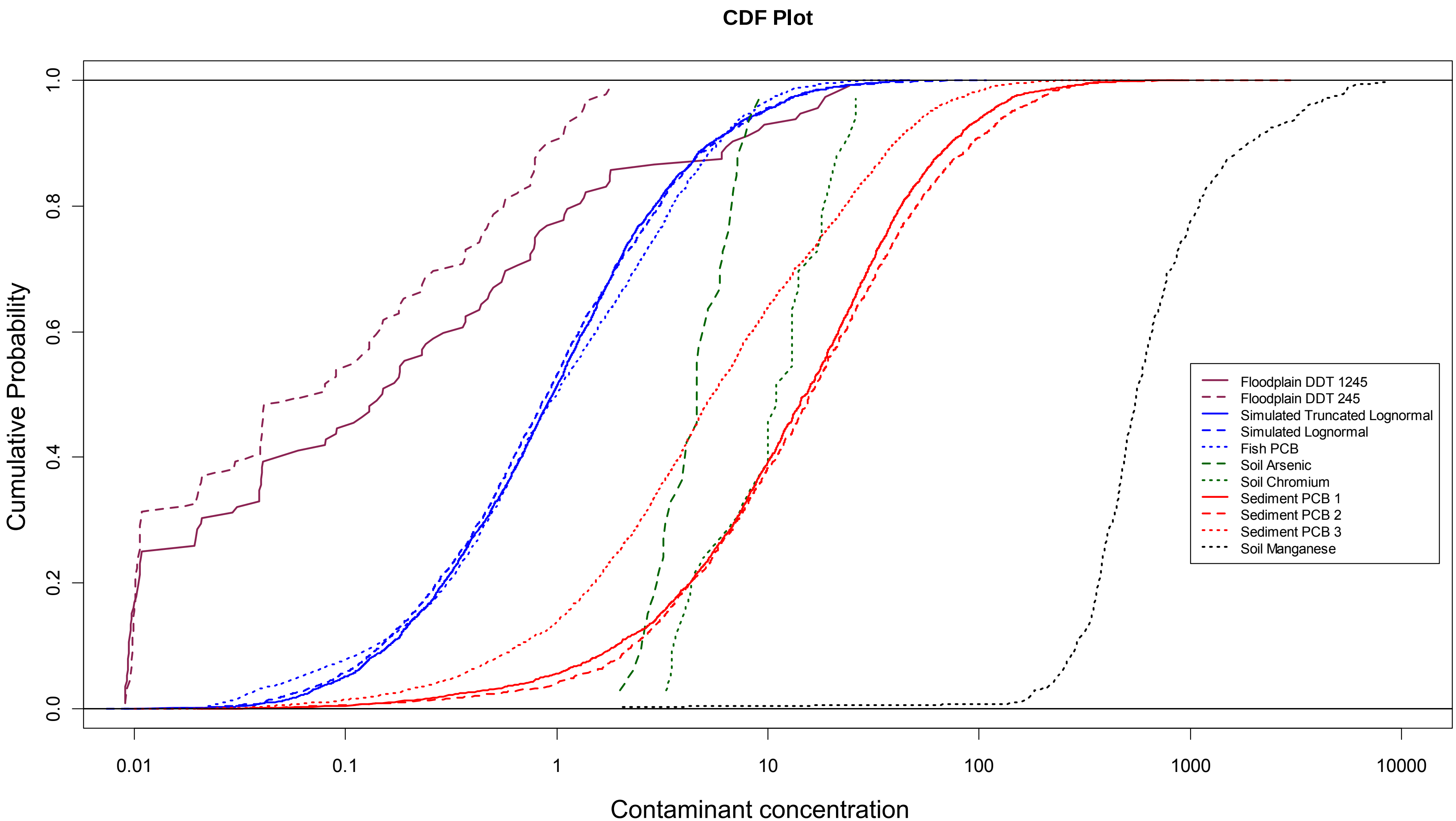

Figure 1. Cumulative distribution function (CDF) plots for each of the 11 populations identified in the legend. The x-axis represents concentration and the y-axis represents the proportion of the population that falls below a given concentration. The solid blue and dashed blue lines represent the theoretical truncated lognormal and lognormal distributions, respectively. The empirical CDFs for the three Sediment PCB populations, the Fish PCB population, and the Soil Manganese population are very similar in shape to the theoretical CDFs, although the Soil Manganese CDF indicates a longer left tail due to the presence of a low outlier. The Floodplain DDT CDFs differ most from the theoretical CDFs.

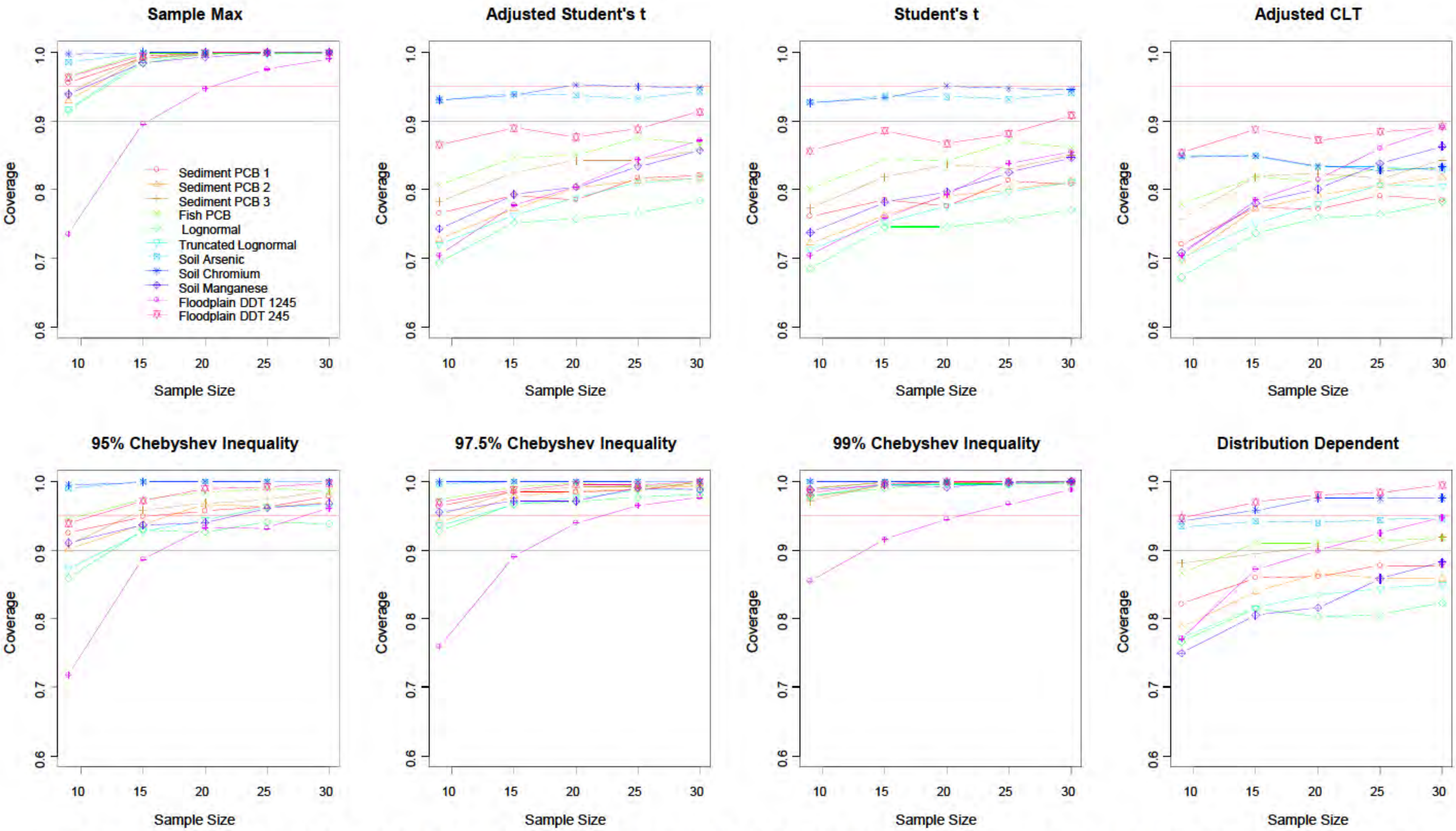

Figure 2-1. Coverage plots for 8 of the 16 statistical methods evaluated. Each plot illustrates the coverage provided by the identified method and sample sizes ranging from 9 to 30 selected from each of the 11 populations. As expected, coverage increases for all methods and all populations as sample size increases. Coverage is well below the nominal 95% (marked by the horizontal red line) for the adjusted CLT method, as is the case for the Student's t, adjusted Student's t, and U.S. EPA's recommended distribution dependent method (ProUCL; 2007) for most populations. Coverage is consistently above 95% for most sample sizes for the 99% Chebyshev and sample maximum. Coverage is generally centered around 95% for the 95% and 97.5% Chebyshev methods, with the 97.5% Chebyshev method performing best at a sample size of 9.

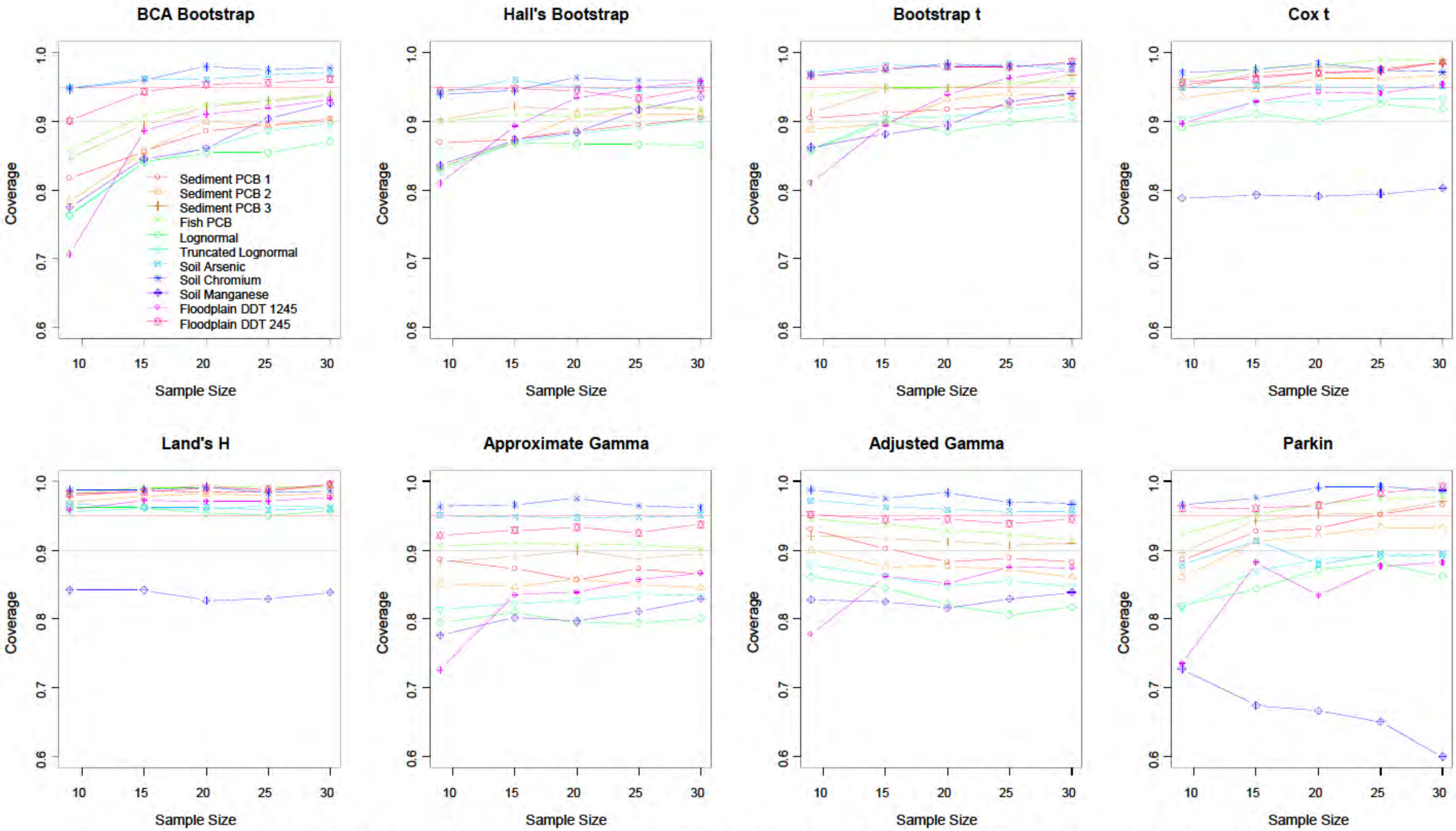

Figure 2-2. Coverage plots for the remaining 8 of the 16 statistical methods evaluated. Each plot illustrates the coverage provided by the identified method and sample sizes ranging from 9 to 30 selected from each of the 11 populations. As expected, coverage generally increases as sample size increases. Coverage is below the nominal 95% (marked by the horizontal red line) for most populations for gamma-based methods, Parkin's method, and BCA and Hall's bootstrap methods (particularly for smaller sample sizes). Coverage for Land's H method is above 95% for all sample sizes and all populations with the exception of the Soil Manganese population. Coverage is generally centered around 95% for Cox's t (also with the exception of the Soil Manganese population).

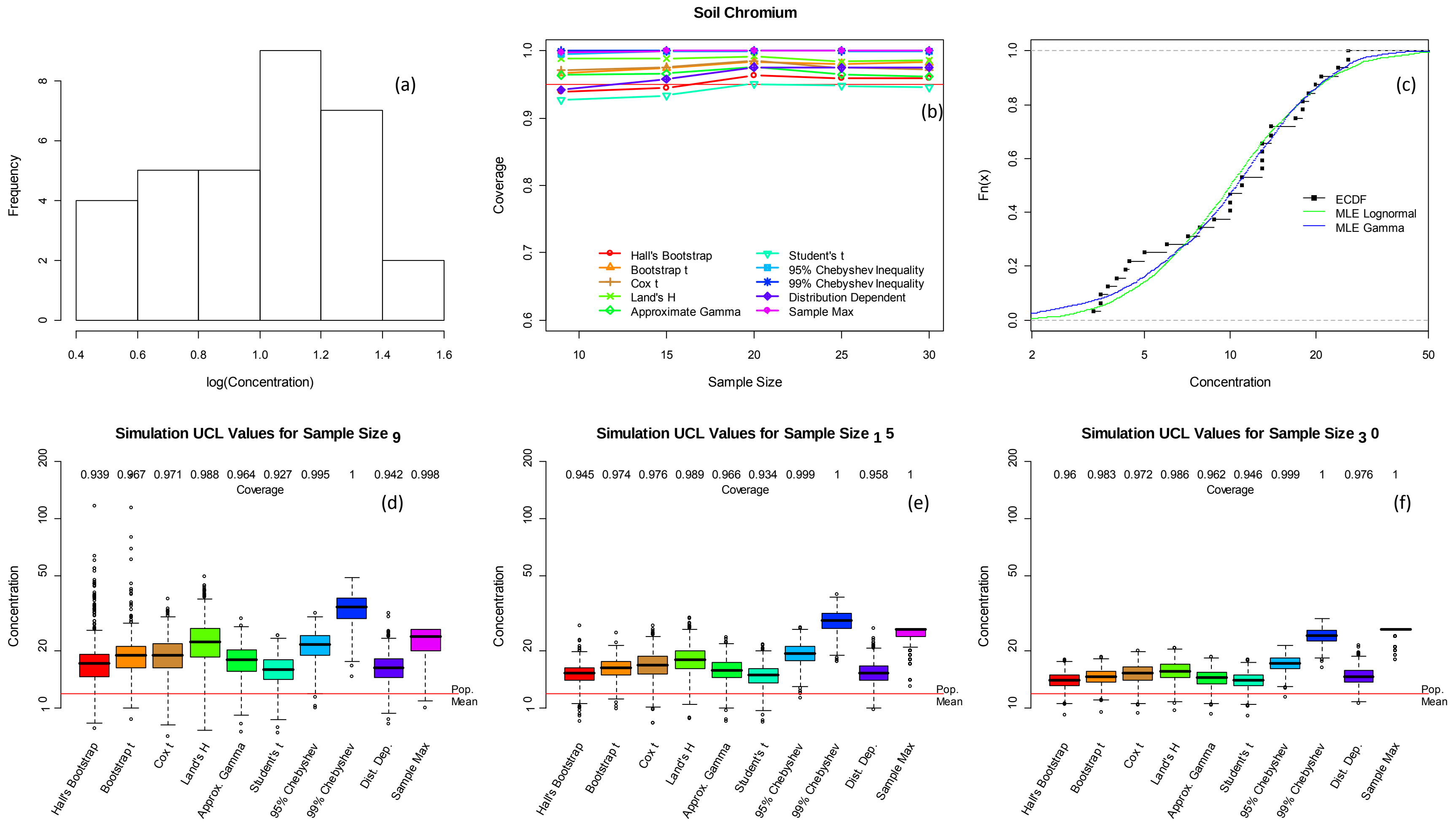

Figure 3. Summary of simulation results for Soil Chromium population. The histogram (a) shows that log(Concentration) is slightly skewed to the left, indicating that log-transforming the data overcorrected for the initial right skewness. Coverage for this population does not appear to be sensitive to method selection: coverage exceeds 95% for most methods at all sample sizes. The empirical CDF (c) is generally closer to the theoretical gamma distribution than the lognormal, but does not approximate either well at lower concentrations. The box plots shown in (d)-(f) summarize the range of simulated upper confidence limits (UCL) for the mean concentration for each of 10 methods and three sample sizes (9, 15, and 30). The true population mean is shown as the horizontal red line and coverage is summarized above the plots for each method. As expected, the box plots tighten up as sample size increases, indicating that simulated UCLs vary less as sample size increases. The box plots shown in (d) illustrate the tendency of some methods (e.g., Hall's bootstrap, Bootstrap t, and Land's H method) to provide relatively extreme UCLs for small sample sizes.

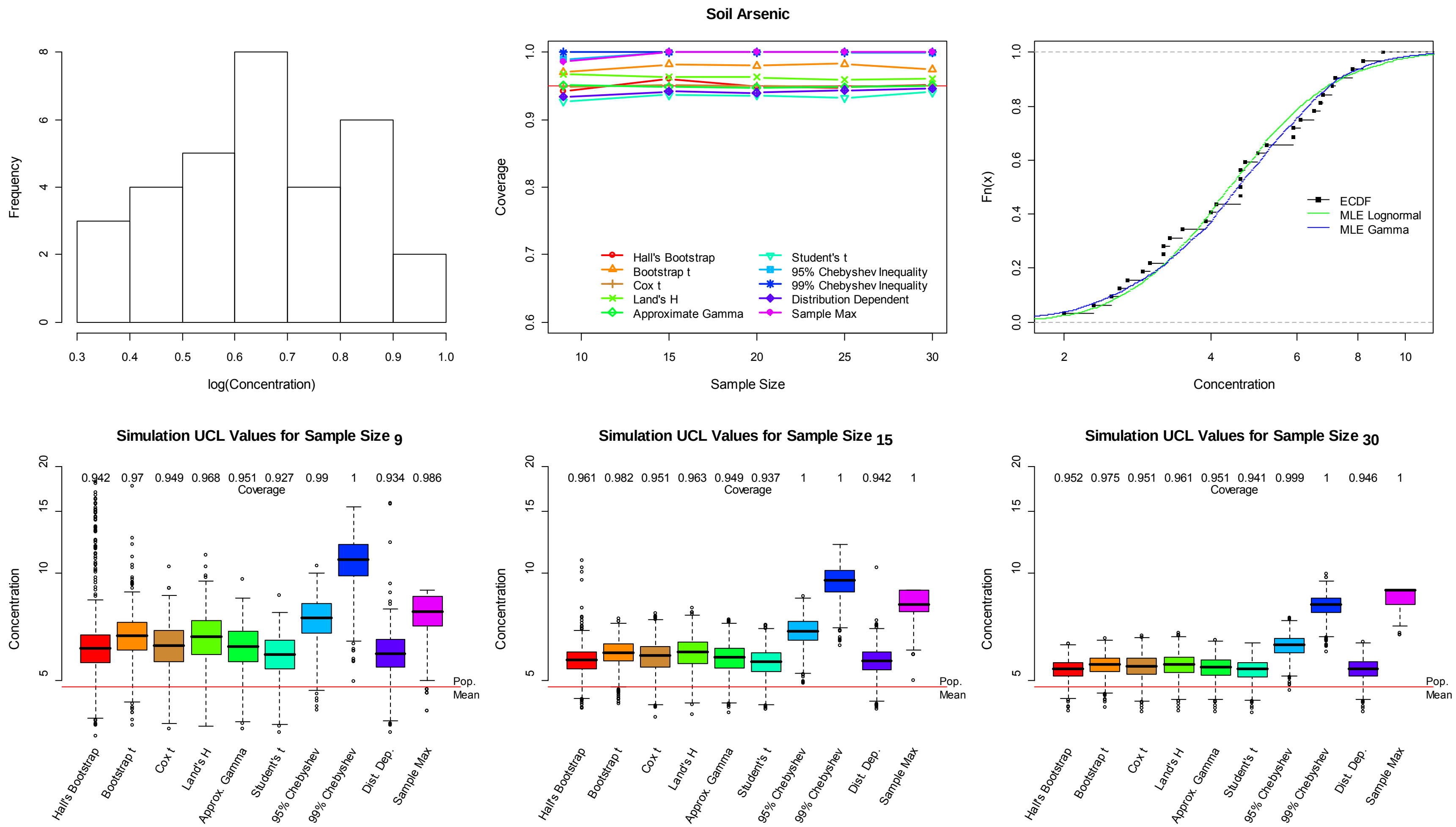

Figure 4. Summary of simulation results for the Soil Arsenic population. The histogram (a) shows that log(Concentration) is relatively symmetric. Coverage for this population does not appear to be sensitive to method selection: coverage meets or exceeds 95% for most methods at all sample sizes. The empirical CDF (c) is generally closer to the theoretical gamma distribution than the lognormal. The box plots shown in (d)-(f) illustrate the tendency of some methods (e.g., Hall's bootstrap, Bootstrap t, Land's H method, and U.S. EPA's recommended distribution dependent method (2007)) to provide relatively extreme UCLs for smaller sample sizes. Methods with a coverage equal to 1 and box plots completely above the population mean indicate that all simulated UCLs exceeded the true population mean.

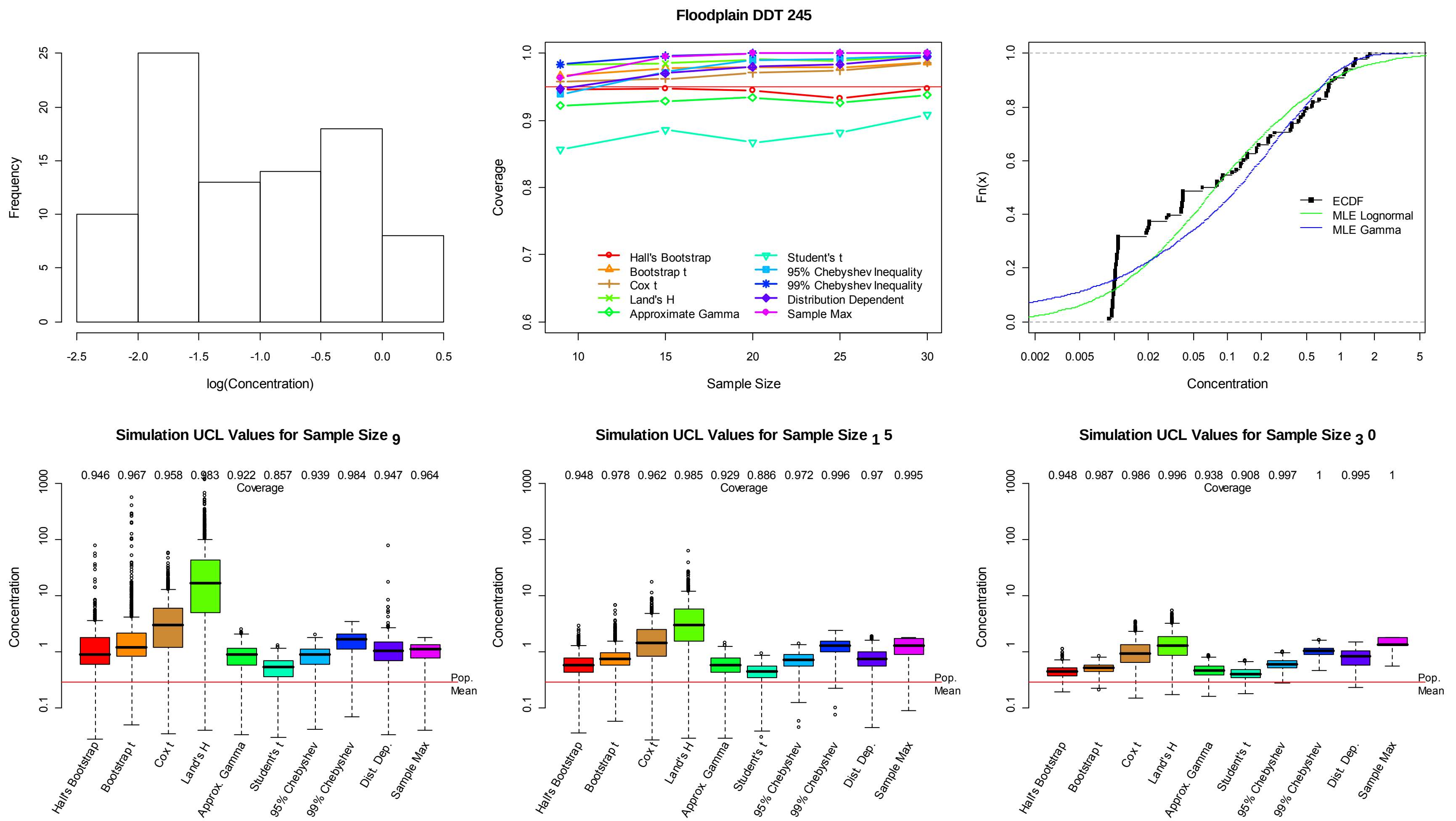

Figure 5. Summary of simulation results for the Floodplain DDT (Sections 2, 4, and 5) population. The histogram (a) shows that log(Concentration) is relatively symmetric. Coverage for this population is not highly sensitive to method selection: coverage meets or exceeds 95% for most methods at all sample sizes. The major exception is the Student's t method. The empirical CDF (c) does not fit either theoretical distribution well. The box plots shown in (d)-(f) illustrate the tendency of some methods (e.g., Hall's bootstrap, Bootstrap t, Land's H method, and U.S. EPA's recommended distribution dependent method (2007)) to provide relatively extreme UCLs for small sample sizes.

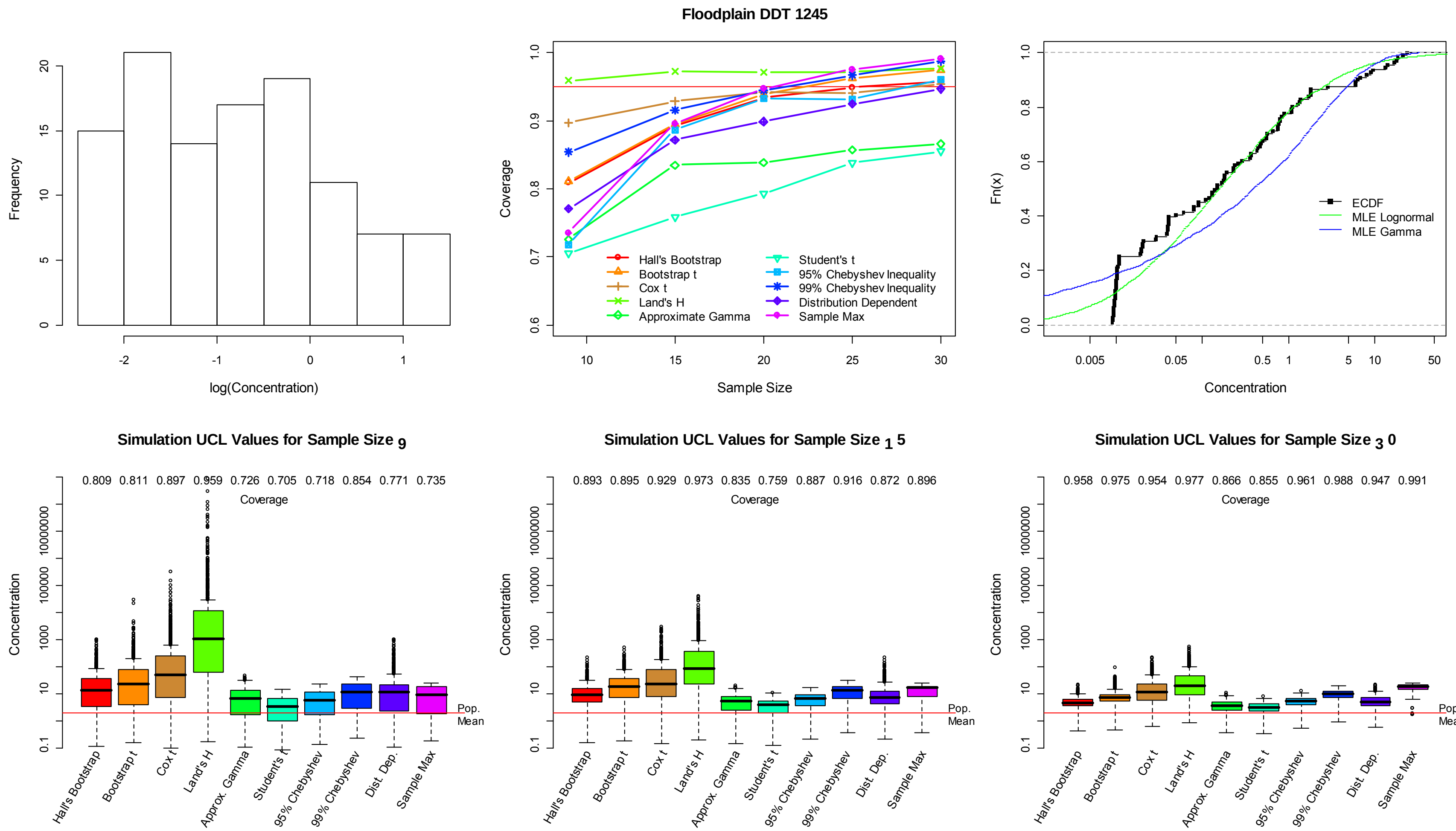

Figure 6. Summary of simulation results for the Floodplain Total DDT (Sections 1, 2, 4, and 5) population. The histogram (a) shows that log(Concentration) is still right skewed when Section 1 is included in the population. Coverage for this population is sensitive to method selection: coverage by method is highly variable and consistently exceeds 95% for Land's H method only. The empirical CDF (c) does not fit either theoretical distribution well at low concentrations, although more closely approximates the lognormal distribution at higher concentrations. The box plots shown in (d)-(f) illustrate the tendency of some methods (e.g., Hall's bootstrap, Bootstrap t, Land's H method, and U.S. EPA's recommended distribution dependent method (2007)) to provide relatively extreme UCLs for small sample sizes.

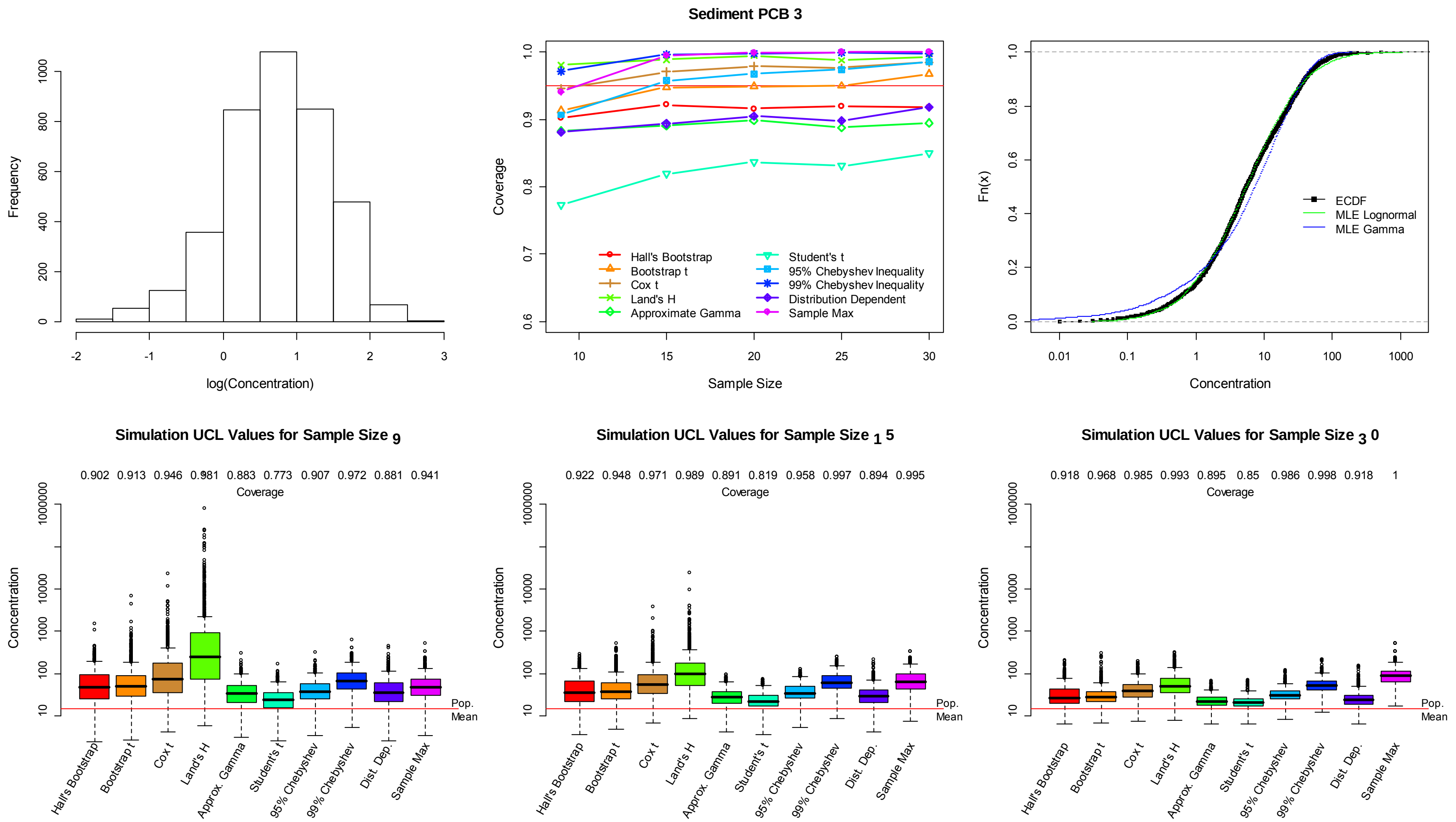

Figure 7. Summary of simulation results for the river Sediment PCB (Section 3) population. The histogram (a) shows that log(Concentration) is symmetric. Coverage for this population is sensitive to method selection: coverage by method is variable and consistently meets or exceeds 95% for four of the 10 methods. The empirical CDF (c) closely approximates the lognormal distribution. Even though this population appears to fit the lognormal distribution well, coverage between methods is variable and the box plots in (d)-(f) still indicate the tendency of some methods (e.g., Hall's bootstrap, Bootstrap t, Cox's t, and Land's H method) to provide relatively extreme UCLs for small sample sizes.

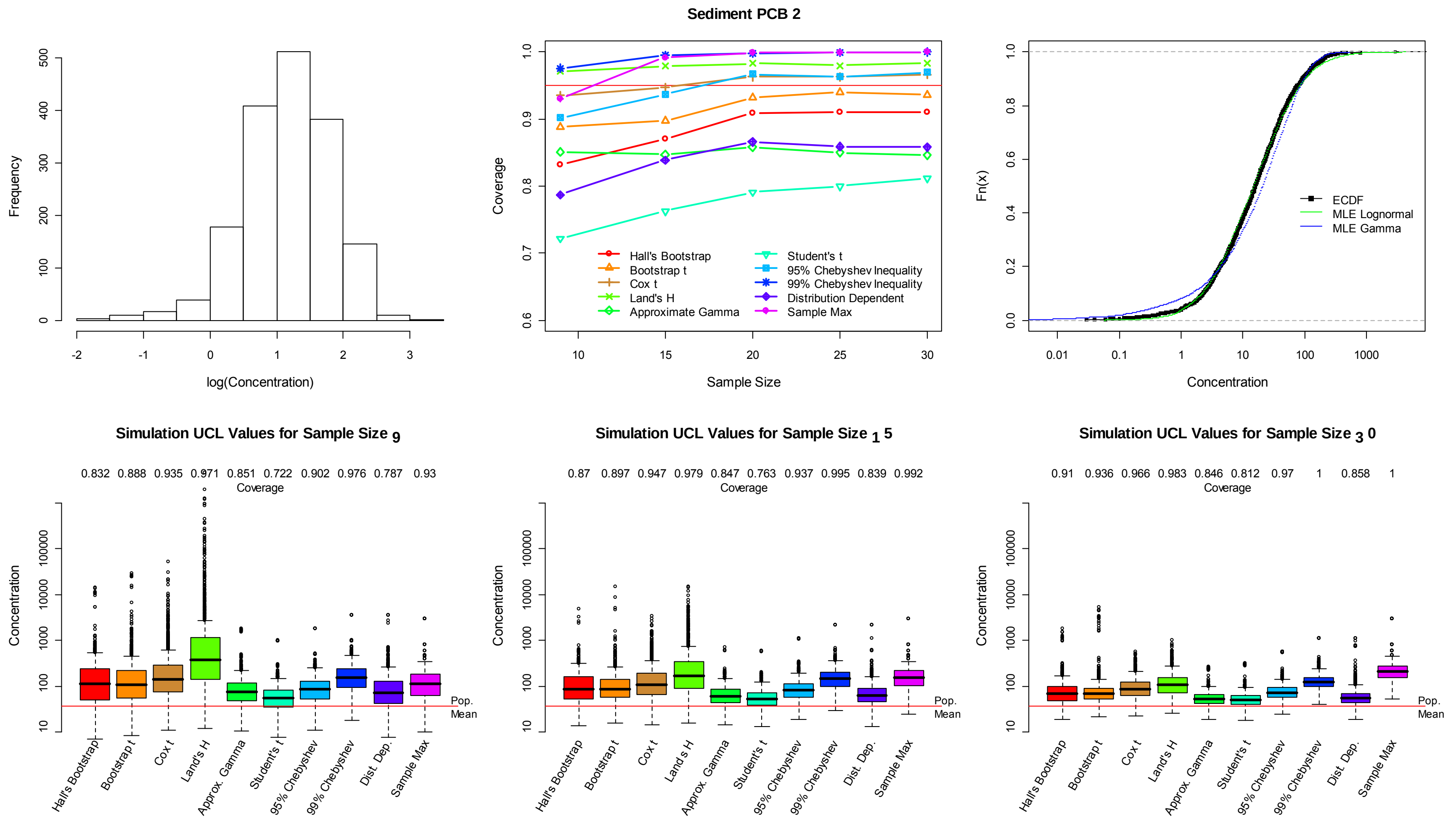

Figure 8. Summary of simulation results for the river Sediment PCB (Section 2) population. The histogram (a) shows that log(Concentration) is symmetric. Coverage for this population is sensitive to method selection: coverage by method is variable and consistently meets or exceeds 95% for four of the 10 methods. The empirical CDF (c) closely approximates the lognormal distribution. Even though this population appears to fit the lognormal distribution well, coverage between methods is variable and the box plots in (d)-(f) indicate the tendency of some methods (e.g., Hall's bootstrap, Bootstrap t, Cox's t, and Land's H method) to provide relatively extreme UCLs for small sample sizes.

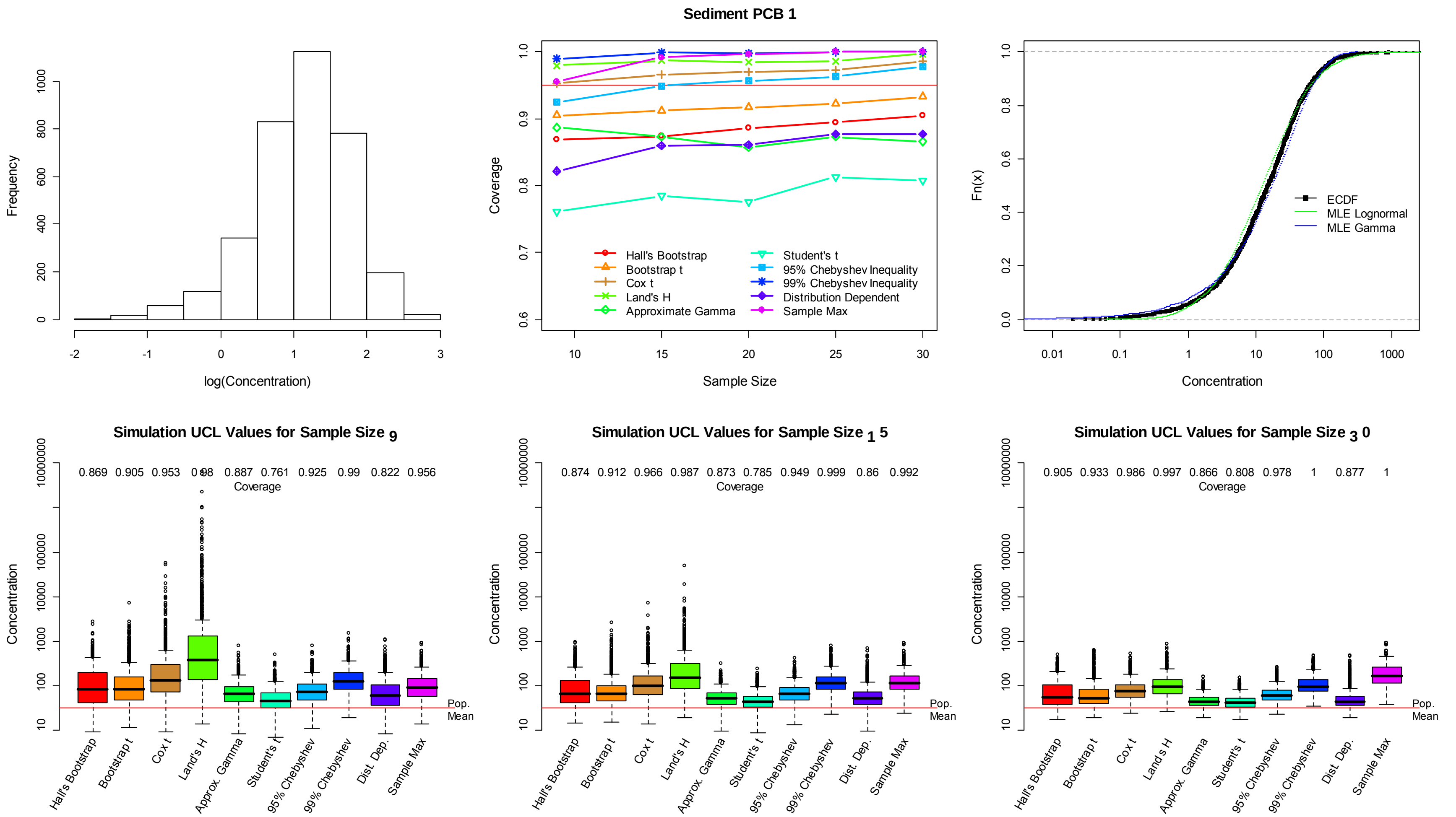

Figure 9. Summary of simulation results for the river Sediment PCB (Section 1) population. The histogram (a) shows that log(Concentration) is slightly left skewed. Coverage for this population is sensitive to method selection: coverage by method is variable and consistently meets or exceeds 95% for four of the 10 methods. The empirical CDF (c) generally approximates the lognormal distribution, but less so than the other Sediment PCB populations; at lower concentrations, the empirical CDF falls between the lognormal and gamma distributions. Even though this population is similar to both theoretical distributions, coverage between methods is variable and the box plots in (d)-(f) still indicate the tendency of some methods (e.g., Hall's bootstrap, Bootstrap t, Cox's t, and Land's H method) to provide relatively extreme UCLs for small sample sizes.

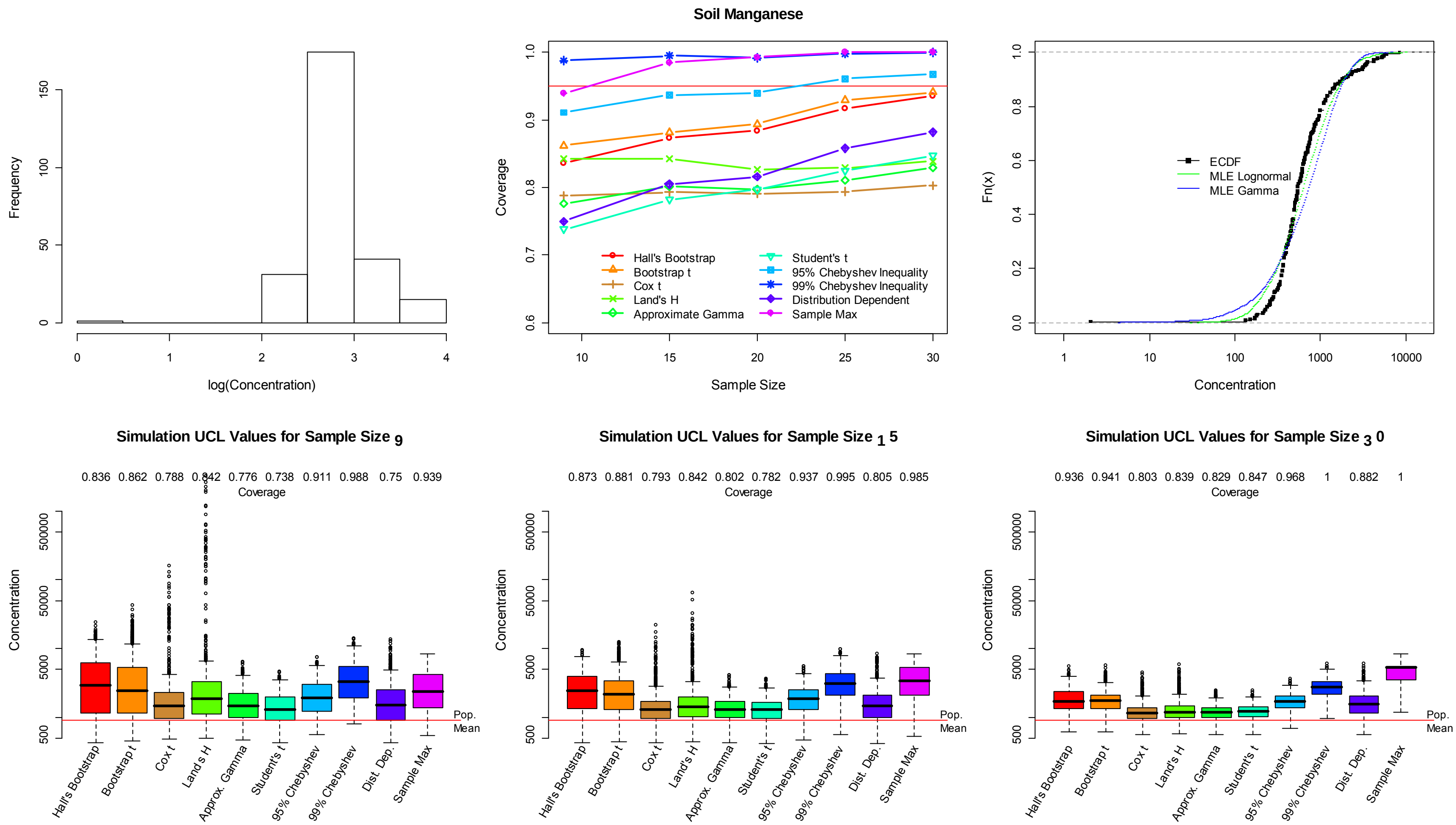

Figure 10. Summary of simulation results for the Soil Manganese population. The histogram (a) shows that log(Concentration) is relatively symmetric with the exception of a low outlier. Coverage for this population is sensitive to method selection: coverage by method is variable and consistently meets or exceeds 95% for only two of the 10 methods. The empirical CDF (c) generally does not approximate either theoretical distribution well. The box plots in (d)-(f) indicate the tendency of some methods (particularly Cox's t and Land's H method) to provide relatively extreme UCLs for smaller sample sizes.

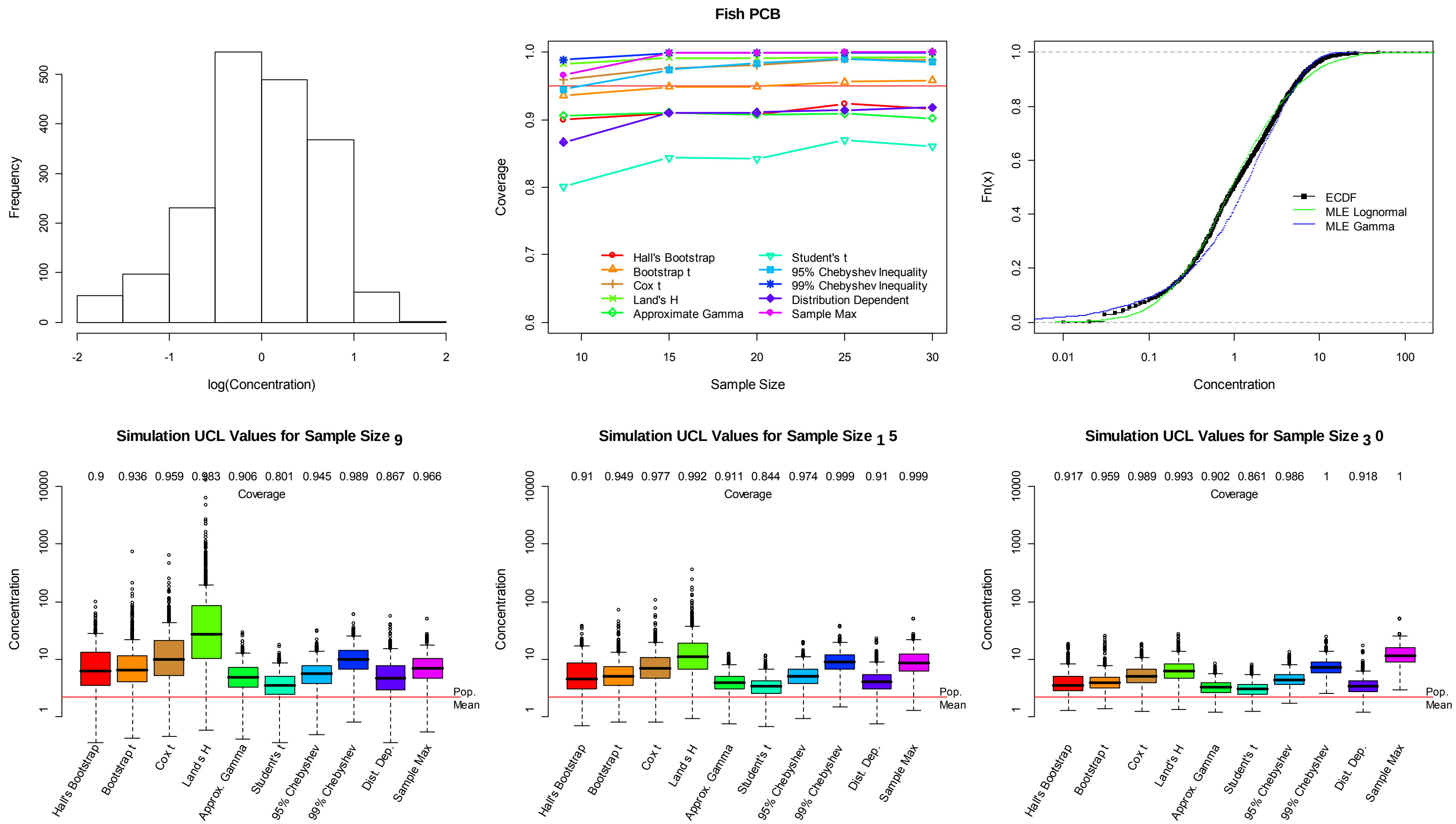

Figure 11. Summary of simulation results for the fish tissue PCB (Fish PCB) population. The histogram (a) shows that log(Concentration) is roughly symmetric. Coverage for this population is somewhat sensitive to method selection: coverage by method is variable, although it consistently approximates 95% for six of the 10 methods. The empirical CDF (c) generally falls between the lognormal and gamma distributions. Even though this population is similar to both theoretical distributions, coverage between methods is variable and the box plots in (d)-(f) indicate the tendency of some methods (e.g., Hall's bootstrap, Bootstrap t, Cox's t, and Land's H method) to provide relatively extreme UCLs for small sample sizes.

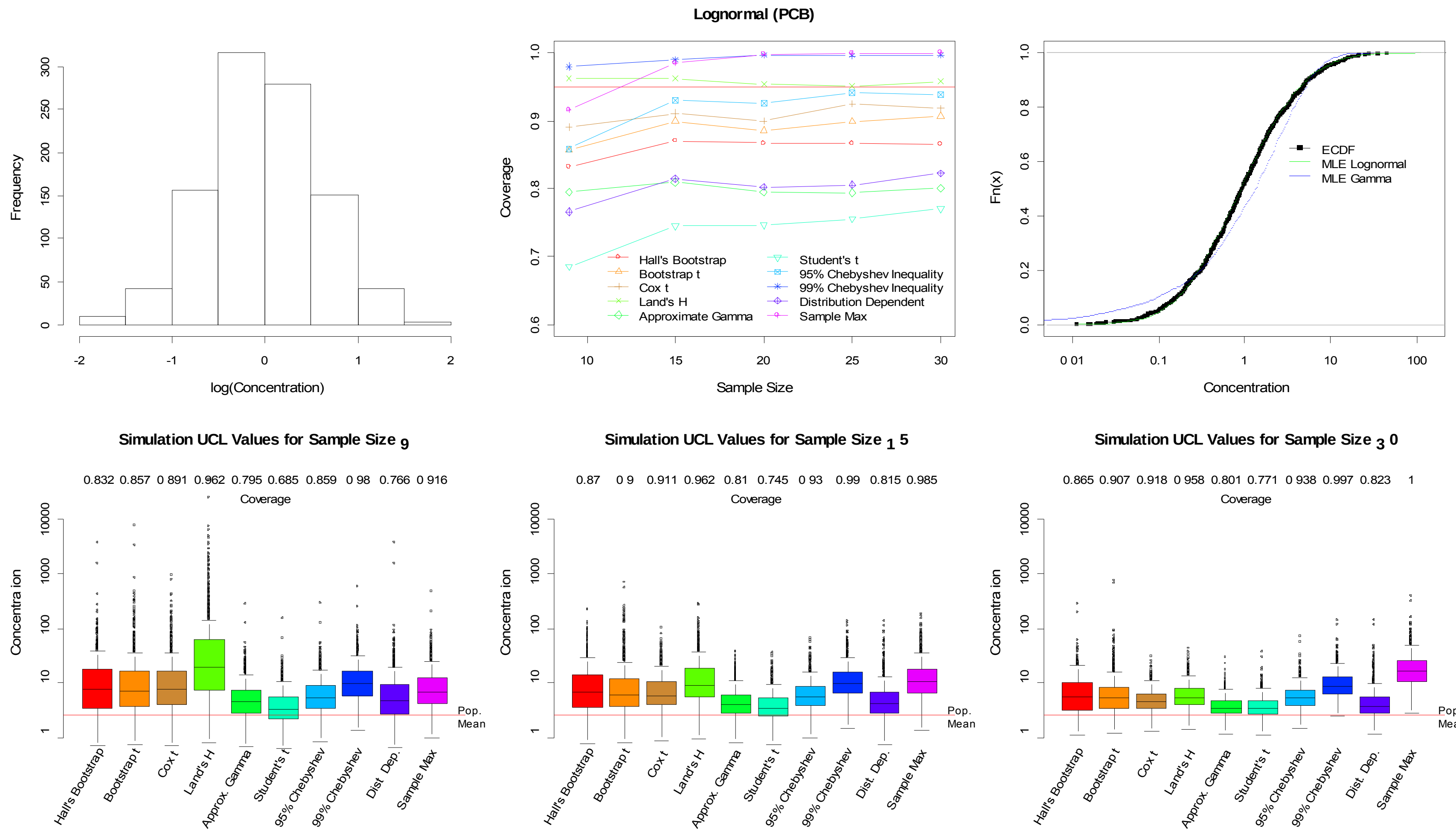

Figure 12. Summary of simulation results for the theoretical lognormal population. The histogram (a) shows that log(Concentration) is symmetric. Even though samples were selected from a lognormal population, coverage for this population is highly sensitive to method selection: coverage by method is highly variable and exceeds the nominal rate of 95% for only two of the 10 methods at a sample size of 9. The worst coverage is provided by the Student's t, approximate gamma, and U.S. EPA's recommended distribution dependent methods (2007). The empirical CDF (c) matches the lognormal CDF, as expected. Several extreme UCL values can be seen across all methods in box plots (d)-(f) for all sample sizes.

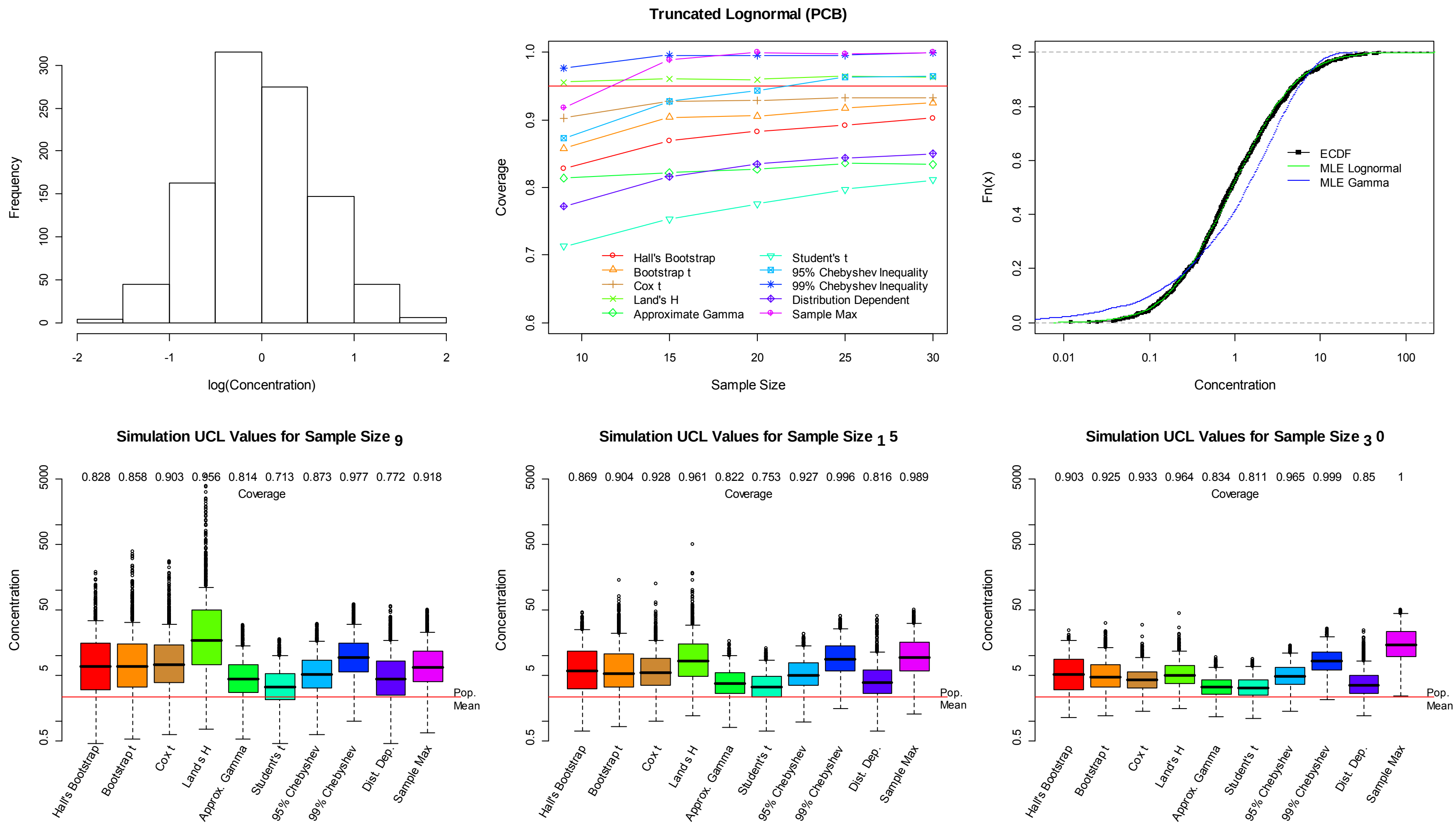

Figure 13. Summary of simulation results for the truncated theoretical lognormal population. The histogram (a) shows that log(Concentration) is symmetric. Coverage for this population is highly sensitive to method selection: coverage by method is highly variable and exceeds the nominal rate of 95% for only two of the 10 methods at a sample size of 9. The worst coverage is provided by the Student's t, approximate gamma, and U.S. EPA's recommended distribution dependent methods (2007). The empirical CDF (c) closely follows the lognormal CDF, as expected. The box plots in (d)-(f) indicate the tendency of some methods (e.g., Hall's bootstrap, Bootstrap t, Cox's t, and Land's H method) to provide relatively extreme UCLs for small sample sizes.

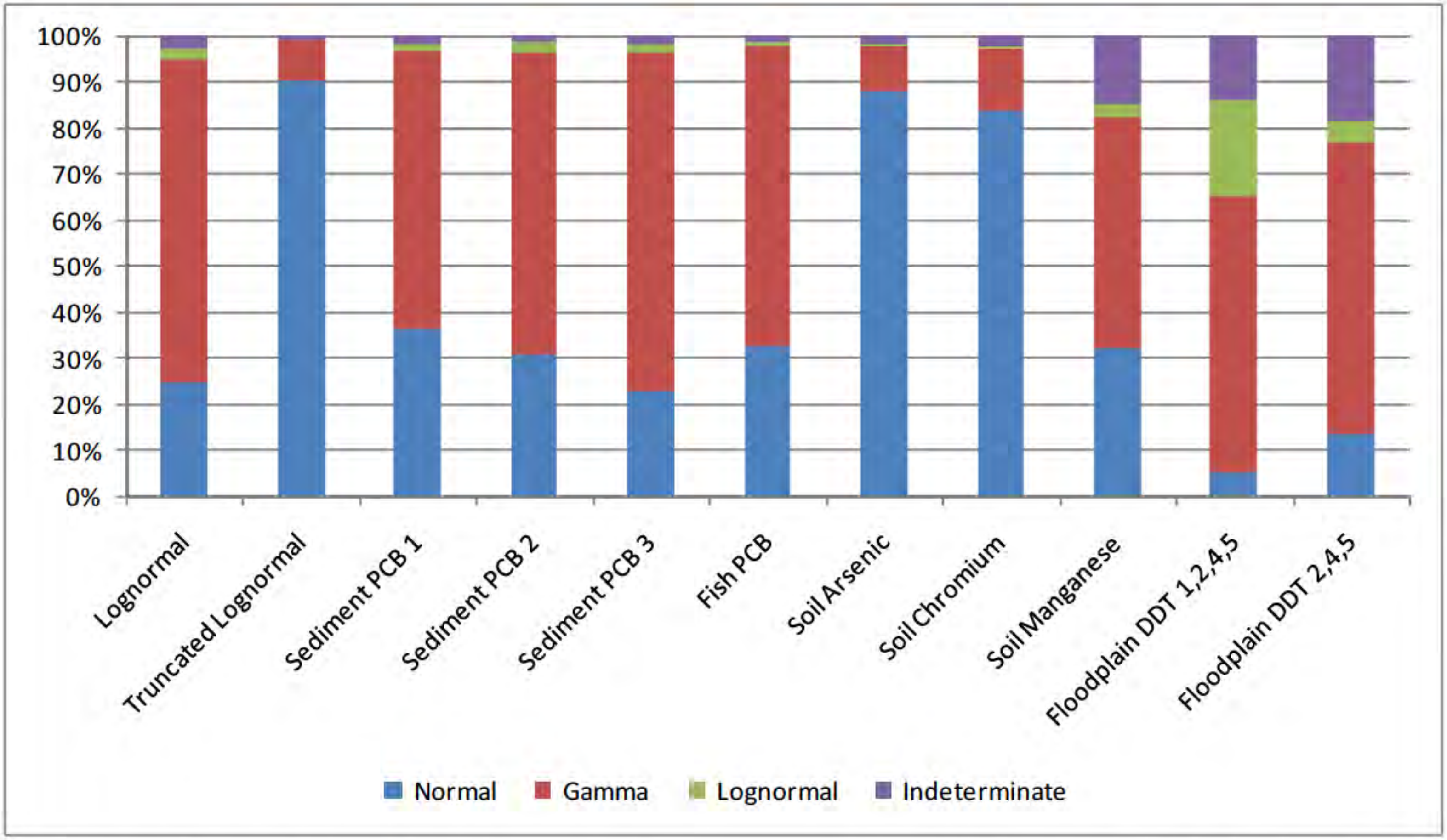

Figure 14. Summary of selected statistical distributions using U.S. EPA's recommended method of selection (2007) for 5,000 simulated data sets of size 9 for each population. U.S. EPA's method involved checking first for a normal distribution, followed by a gamma distribution if significantly non-normal, then a lognormal distribution if neither normal nor gamma. Checking for distribution type in a different order would likely yield substantially different results. The gamma distribution was selected most frequently for all populations (even the theoretical lognormal and truncated lognormal), with the exception of the Soil Arsenic and Soil Chromium populations, which were concluded to be normally distributed in more than 80% of simulations. The normal distribution was identified with the second highest frequency for all but the Soil Arsenic and Soil Chromium populations and the Floodplain DDT population (Sections 1, 2, 4, 5), which was concluded to be lognormal in approximately 20% of the simulations. Lognormal distributions and indeterminate conclusions were identified in less than 5% of simulations for all but the Soil Manganese, Floodplain DDT (Sections 1, 2, 4, 5), and Floodplain DDT (Sections 2, 4, 5) which were indeterminate in 14% or more simulations.

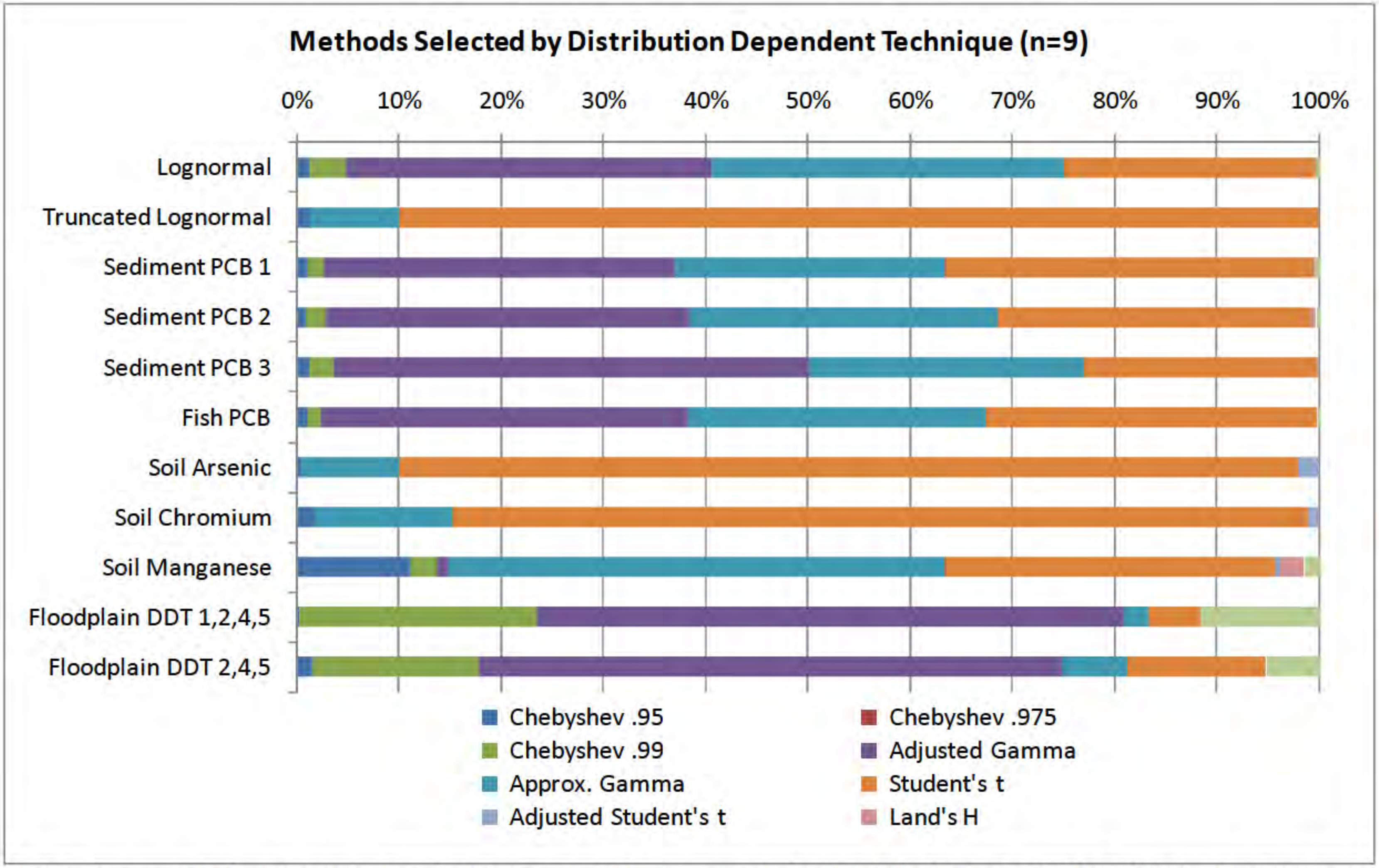

Figure 15. Summary of selected statistical methods using U.S. EPA's recommended method of selection (2007) for 5,000 simulated data sets of size 9 for each population. U.S. EPA's method involved checking first for a normal distribution, followed by a gamma distribution if significantly non-normal, then a lognormal distribution if neither normal nor gamma. Methods for calculating UCLs were selected based on statistical distribution (i.e., distribution dependent). Checking for distribution type in a different order would likely yield substantially different results. The statistical methods for calculating UCLs based on the gamma distribution (adjusted or approximate) were selected most frequently for all populations (including the theoretical lognormal and truncated lognormal), with the exception of the Soil Arsenic and Soil Chromium populations for which the Student's t method was selected in more than 80% of simulations.

RELABLE NEXT TWO SLIDES TO ELIMINATE NAMES OF STUDIES

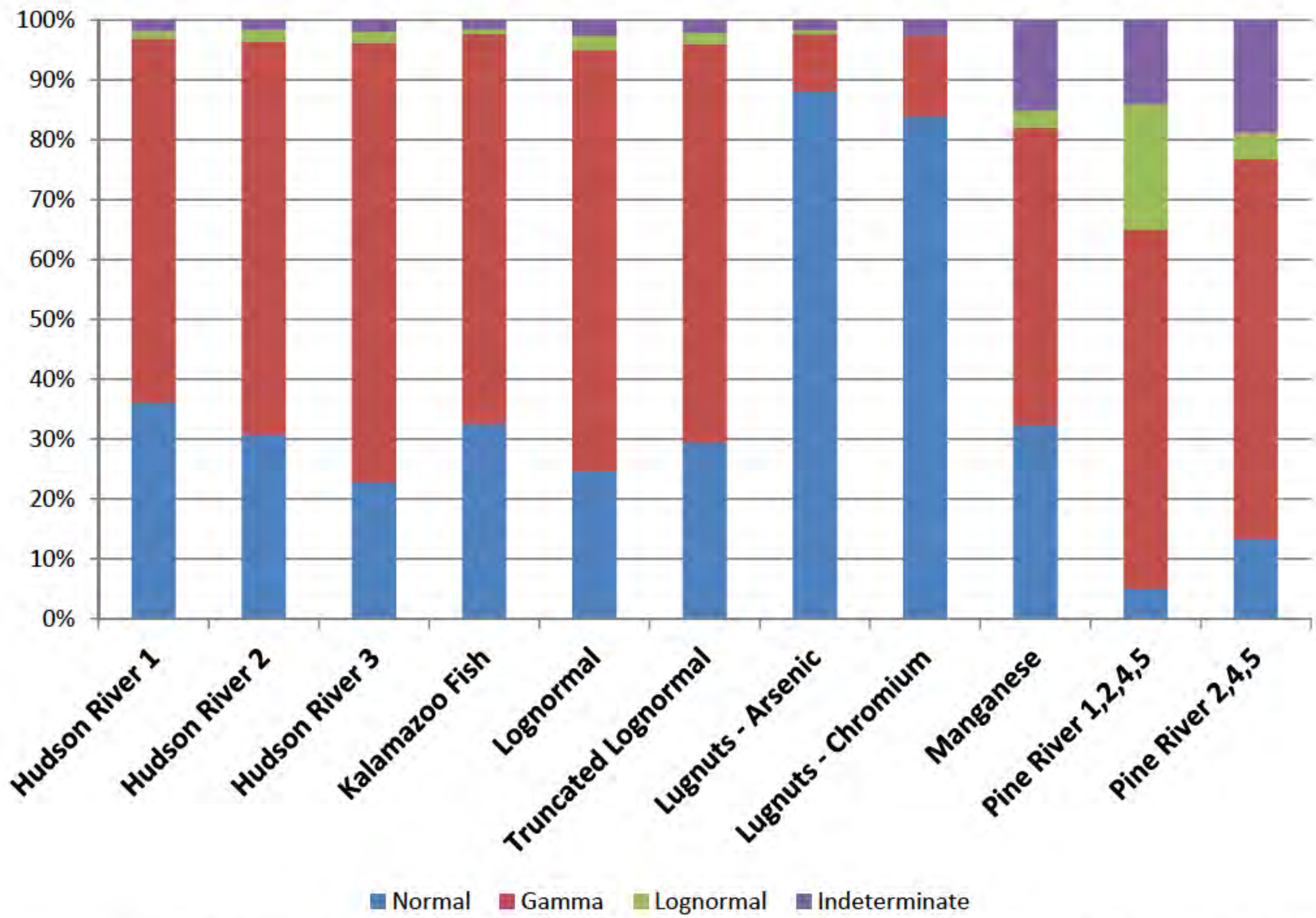

Figure 14. Summary of selected statistical distributions using EPA's recommended method of selection (2007) for 1000 simulated data sets of size 9 for each population. EPA's method involved checking first for a normal distribution, followed by a gamma distribution if significantly non-normal, then a lognormal distribution if neither normal nor gamma. Checking for distribution type in a different order would likely yield substantially different results. The gamma distribution was selected most frequently for all populations (even the theoretical lognormal and truncated lognormal), with the exception of the two Lugnuts populations, which were concluded to be normally distributed in more than 80% of simulations. The normal distribution was identified with the second highest frequency for all but the Lugnuts populations and the Pine River (Sections 1, 2, 4, 5), which was concluded to be lognormal in approximately 20% of the simulations. Lognormal distributions and indeterminate conclusions were identified in less than 5% of simulations for all but the down river manganese, Pine River (Sections 1, 2, 4, 5), and Pine River (Sections 2, 4, 5) which were indeterminate in 14% or more simulations.

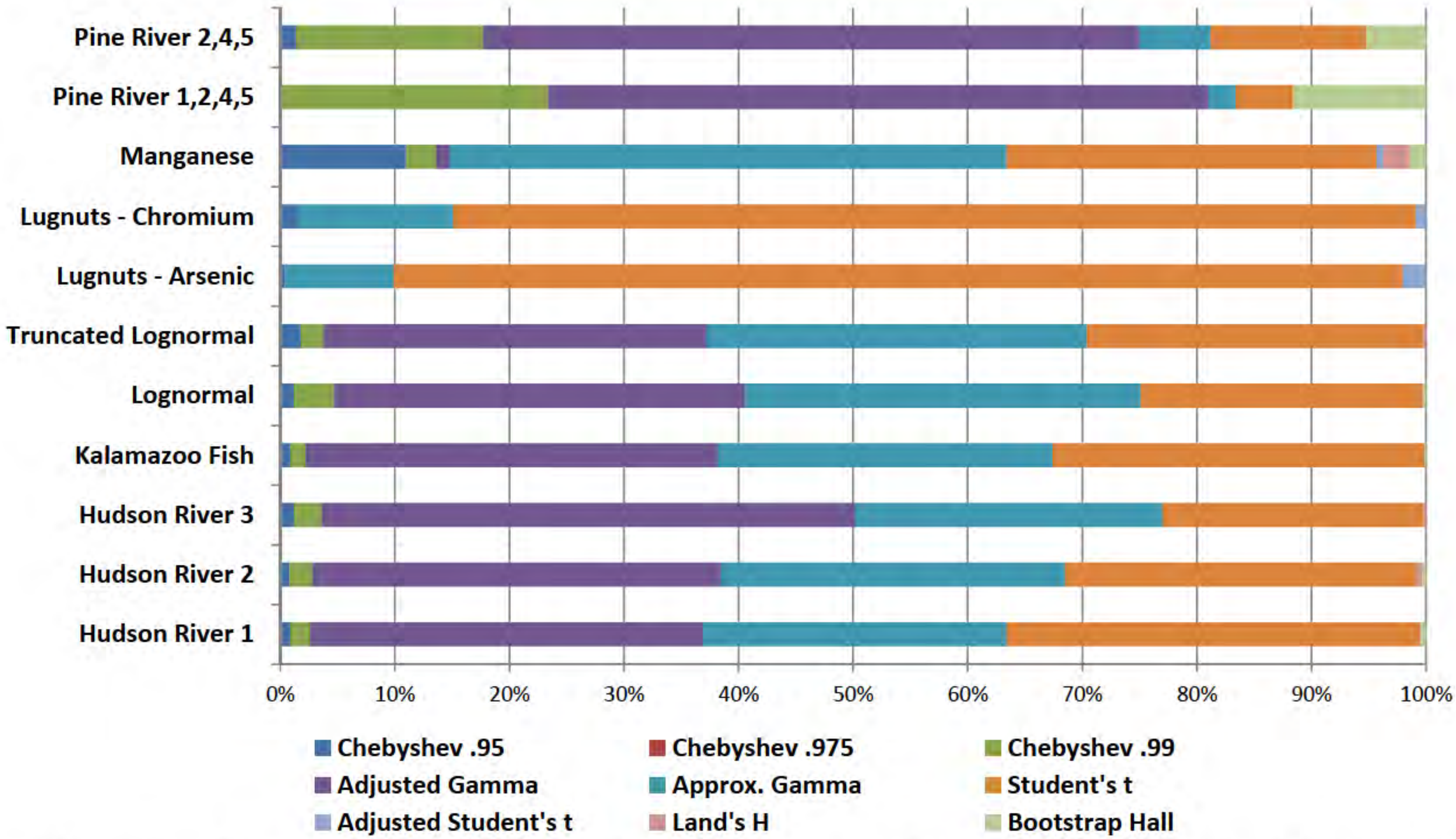

Figure 15. Summary of selected statistical methods using EPA's recommended method of selection (2007) for 1000 simulated data sets of size 9 for each population. EPA's method involved checking first for a normal distribution, followed by a gamma distribution if significantly non-normal, then a lognormal distribution if neither normal nor gamma. Methods for calculating UCLs were selected based on statistical distribution (i.e., distribution dependent). Checking for distribution type in a different order would likely yield substantially different results. The statistical methods for calculating UCLs based on the gamma distribution (adjusted or approximate) were selected most frequently for all populations (including the theoretical lognormal and truncated lognormal), with the exception of the two Lugnuts populations for which the Student's t method was selected in more than 80% of simulations. The 99% Chebyshev method was selected with the next highest frequency for both Pine River populations. Hall's Bootstrap procedure was also selected in at least 5% of simulations for the Pine River populations. The Student's t method was selected with at least 20% frequency for the theoretical lognormal populations, the Kalamazoo fish population, and the 3 Hudson River populations.