Guidance for
Comparing Background and Site Chemical Concentrations in Soil

Florida Department of Environmental Protection
Division of Waste Management
District & Business Support Program
Tallahassee, FL

March 2019

Disclaimer

This is a guidance document intended to provide information and assistance in common methods of comparing site and background chemical concentrations in soil. As guidance, the techniques and procedures in this document are not enforceable by any party. The methods described in this guidance document are not intended to include all methods for evaluating background in soil and alternative methods may always be proposed. Nothing in this guidance supersedes any Federal, State, or Local requirements.

Overview

The purpose of this guidance is to describe procedures the Florida Department of Environmental Protection (FDEP) has found acceptable for the comparison of contaminant concentrations in soil between site and background. An evaluation of local background concentrations is appropriate at a cleanup site whenever it is suspected that certain contaminants detected above applicable cleanup criteria may be equal to, or less than, natural or anthropogenic background concentrations. Some chemicals, such as inorganics and radionuclides, are present naturally in soils or may be introduced as contaminants. If chemical contaminants are the result of a discharge or release and exceed risk-based criteria, cleanup or other risk management measures are required. If chemicals are present due to natural soil conditions, or unrelated anthropogenic impacts, then cleanup is not needed under current rules even if the concentrations exceed risk-based criteria. Consequently, it is important to determine whether or not the presence of a chemical in the detected concentration(s) is representative of a background condition.

Some chemicals, both manmade and natural, are ubiquitous in the urban environment due to human activities. Examples include polycyclic aromatic hydrocarbons (PAHs) and dioxins. Low levels that exist in the environment due to dispersion of these chemicals are often representative of an anthropogenic background condition. The definition for "background concentrations" specific to corrective action conducted under Chapter 62-780.200(3), Florida Administrative Code (F.A.C.), is:

> *"Background concentrations" means concentrations of contaminants that are naturally occurring or resulting from anthropogenic impacts unrelated to the discharge of pollutants or hazardous substances at a contaminated site undergoing site rehabilitation, in the groundwater, surface water, soil or sediment in the vicinity of the site.*

When delineating the boundaries of contamination attributable to a discharge or release, natural and anthropogenic background concentrations become important. They are used to help establish the area where site impacts above background are found, and thus define the limits to where liability for cleanup exists. The limits to which site assessment, possible corrective action and risk management must extend do not need to go beyond the point at which the concentrations of discharge related contaminants become indistinguishable from local background.

Background chemicals fall into the following two categories:

<u>Naturally Occurring Background Chemicals</u>- These are defined as chemicals present as a result of geochemical processes that have not been influenced by human activity. Naturally occurring organic and inorganic background chemicals in soil and in

groundwater are attributable to the natural geological and hydrogeological characteristics of the area.

Anthropogenic Background Chemicals- Constituents that are synthetic or natural substances that have entered the environment as a result of human activities, but are not related to specific activities conducted at the site. Anthropogenic background chemicals typically are widely distributed in the environment due to human activities, not related to site sources or releases, and attributable to past and present legal applications or sources. In some cases, it is not clearly evident whether or not a constituent is naturally occurring or anthropogenic in origin, but this does not prevent the data from being used to establish site-specific cleanup levels based on background concentrations.

Where to Obtain Background Concentration Information

Background concentration information is derived on a site-specific basis using samples from nearby "background" locations. The basic principle in identifying background sampling locations is to find areas that resemble as closely as possible soil conditions at the site had a discharge or release not occurred. The selection of background sampling locations is a matter of professional judgment, but the following points should be considered:

- The background sampling area must be clearly unaffected by releases from the subject site, or any other site. When characterizing natural background conditions, samples are best taken from areas with minimal anthropogenic impact (e.g., natural areas and parks). In establishing anthropogenic background, sampling in areas where contaminants may accumulate should be avoided unless data are needed specifically for comparison with similar features found on a site. These data should be evaluated separately from other anthropogenic background samples. Because selection of background sampling locations is a matter of professional judgment, concurrence from FDEP staff before obtaining background samples is recommended. The following areas are usually inappropriate to sample when determining soil background:

    1. Fill areas;
    2. Areas where known or suspected hazardous substances, petroleum, solid or hazardous wastes or waste waters are managed, treated, handled, stored or disposed;
    3. Areas affected by runoff from a roadway;
    4. Parking lots and areas affected by runoff from parking lots or other paved areas;
    5. Railroad tracts or railway areas or other areas affected by their runoff;
    6. Areas of concentrated air pollutant depositions or areas affected by their runoff;
    7. Storm drains or ditches presently or historically receiving industrial or urban runoff.

- Natural concentrations of inorganics can vary with soil type. When determining natural background, the soil type for the site and background locations should be the same, if possible. The data on contaminant concentrations from different soil types can be pooled if it is determined that there is no significant difference between the contaminant concentration by soil type and/or interval.

- Both natural and anthropogenic chemical concentrations can vary with soil depth. Consequently, background samples should be taken from the same soil horizon(s) as the site soil samples.

- Concentrations from background studies published in the literature cannot be used as the exclusive basis of comparison with site concentrations. Published background studies may be of value in determining whether a site-specific background data set lies within the range of observations by others. If not, the validity of the site-specific background data set may need to be evaluated.

  Prior to using background literature values, they should be reviewed to ensure they are appropriate for the site. It should be understood how the background data were collected, what the data represent (e.g., whether anthropogenic background contributions were considered) and how the literature values need to be compared with site contaminant concentrations. Literature values should not be used for decision making if there is inadequate information to determine how these values were derived. It should be independently determined whether the background data are reasonably representative of site-specific background conditions.

- In measuring chemical concentrations in background samples, the same analytical methods used for site samples should be employed. For example, if ISM samples are collected at the site, ISM samples should be collected in the background area.

- The background data set should be examined carefully for the presence of outliers, i.e., data that may not in fact represent background conditions. Formal outlier tests as well as professional judgment can be used in evaluating the background data set. If using an outlier test, data points may be identified that do not appear similar to the rest of the data set. However, background data points should not be discarded on the sole basis of statistical outlier tests. Outliers should not be discarded without justification.

Non-Statistical Approaches for Comparing Site and Background Data

For many sites, a determination of whether site concentrations represent background conditions can be made without using statistical tests. Two recognized non-statistical approaches are described below. Other approaches may be acceptable pending review by the FDEP.

1. Direct comparison of site concentrations with background

In this approach, the upper end of the range of background concentrations is defined as the lower of:

1) The maximum background concentration, or
2) Twice the mean background concentration.

The maximum concentration on site is compared with this upper limit on background. If the maximum concentration found on site is less than or equal to this upper background limit, the chemical can be considered to be background and removed from further consideration in any risk assessment or site remediation decisions.

This approach has been used for decades and has gained regulatory acceptance. It is simple and works with a limited number of background samples. When conducting this test, the following points apply:

- A minimum of seven background samples is needed (i.e., data from seven different background locations).

- Discrete samples are needed to identify the maximum background and site concentrations, which are critical for this test.

- As noted above, comparisons should be made with equivalent soil horizons. In general, data from different soil horizons should not be combined unless the absence of concentration change with depth can be clearly demonstrated.

- For “non-detect” background samples, one-half the detection limit should be used in calculating the mean background concentration.

If site concentrations are above background, cleanup to background levels is warranted. If background concentrations are above the risk-based criteria, the site-specific upper limit on background (i.e., the lower of the maximum or twice the mean background concentration) can be used as a not-to-exceed cleanup criterion. That is, removal or management of all concentrations above this value will be considered to have restored the site to background conditions with respect to this contaminant.

2. Weight of Evidence demonstration that contaminants are not site related

Contaminants found during a site assessment or cleanup can be excluded from the cleanup if the weight of evidence clearly demonstrates that the contaminant is not related to the discharge(s) for which the responsible party has liability or has accepted responsibility. Upon site closure, the Site Rehabilitation Completion Order (SRCO) should include a statement that the particular contaminant was determined to be unrelated to the discharge for which the SRCO is being issued.

Note that a contaminant need not have been a known constituent of a known release to be discharge related. Contaminants that are present due to degradation, weathering or the effects of the discharge in the environment are considered to be discharge related contaminants. Examples of such contaminants include breakdown products of the parent compounds, reaction products from the interaction of the release with native materials, and mobilization of native metals due to altered pH or redox conditions. Discharge related contaminants also includes those released or introduced from any corrective actions and their corresponding breakdown, reaction or mobilization products, or alterations in site geochemistry.

The following lines-of-evidence can be considered when evaluating site assessment data for making a determination of whether a contaminant is discharge related. Other lines of evidence are possible if supported by sufficient information, such as “chemical fingerprinting” performed as part of an environmental forensics evaluation. In general, no single line of evidence will provide sufficient justification to determine that a particular contaminant is not related to the discharge in question.

- There is no record of the contaminant in question having ever been released to the environment at the site and it is not a breakdown or reaction product of a contaminant that is known to have been released.

- The assessment data do not show any well-defined pattern of concentrations indicative of a release for the contaminant. Typically, a contaminant concentration gradient will lead back to the location of a release.

- The contaminant concentrations show an increasing trend away from possible source areas or towards the property boundary suggesting that there may be an off-property source. However, it must also be demonstrated that the trend away from the known source area is not a trend towards another on-property source area.

- The contaminant exceedances are located only within the confines of a permitted stormwater management feature (i.e. treatment pond, swale or conveyance ditch) and that such exceedances are not due to improper disposal to that feature.

- The contaminant is associated solely with surface soil (i.e., zero to two feet) near site features such as fences, sidewalks or buildings and is known to be a constituent of pesticides/herbicides that were likely used for their intended purpose and properly applied.

- The contamination appears to be the result of natural weathering processes acting on site features constructed with materials known to contain the contaminant (i.e. asphalt or shell road/parking area surfaces, treated wood landscaping products, etc.). Note that contaminants that are the result of the natural weathering of a known discharge are considered discharge related contaminants.

- Soil contaminant(s) detected above CTLs is/are associated with dredge spoil placed at locations authorized by a permitting authority.

An example of the above using a line-of-evidence approach would be the detection of PAHs in soil along the edge of an asphalt parking lot. PAHs are often found close to such asphalt surfaces with concentrations decreasing significantly within a short distance outward. This localized occurrence of PAHs from weathering of the asphalt surface would not be considered a discharge subject to assessment and cleanup under FDEP rules.

Statistical Methods for Comparing Site and Background Data

If sufficient data are available, the combination of statistical methods offer a stronger, more robust method of comparing site and background data.

A complete enumeration and review of the appropriate statistical tests that could be used to compare background against site data are beyond the scope of this guidance document. A number of statistical methods for the comparison of site and background datasets can be found in the USEPA guidance (USEPA, 2002).

Before proceeding to statistical comparisons, preliminary data analysis should occur. This includes calculating summary statistics such as the mean and standard deviation for both the site and background datasets and a visual examination of distribution(s) of these datasets. Any statistical software may be used for these analyses. ProUCL software (available from the USEPA) is capable of performing all of the preliminary analyses and statistical approaches discussed in this guidance.

1. Summary statistics
   Summary statistics should be performed for both site and background datasets. Several important descriptive statistics include:
   - The number of samples
   - The number of non-detects
   - The minimum detected concentration
   - The median detected concentration
   - The maximum detected concentration
   - The mean concentration
   - The standard deviation
   - The minimum and maximum reporting limits

   Estimates of the mean and standard deviation should be appropriate for the distribution of the dataset. Lognormal datasets should use the geometric mean and standard deviation since they are more representative of the central tendency than arithmetic statistics for right skewed distributions. Kaplan Meier mean and standard deviation are more appropriate for datasets with non-detects.

2. Outlier tests
   Outlier tests help identify samples for further evaluation of whether the data points all belong to the same data set. Outlier tests should not be used as standalone tests to exclude or include data points. Decisions regarding inclusion or exclusion of individual data points should be made based upon a weight of evidence approach. Outliers have different implications for site and background datasets. An outlier in the background dataset may indicate that background was collected from a location that was not truly background. An outlier in the site dataset can indicate an area of higher concentration that may not be indicative of background (USEPA, 2002). In both cases, the outlier tests identify samples that may not be representative of the population from which they were drawn. The test provides statistical evidence that a sample is not from the same population as the rest of the data. To show the effect of outliers on the dataset, the statistical analyses should be conducted with and without the statistical outliers (USEPA, 2002). Data points should not be discarded on the sole basis of statistical outlier tests, as an outlier could be indicative of an erroneous distribution assumption and higher than expected variability. Additionally, possible values for the outlier test can depend on variables such as the sample size and range of the data. If samples are identified as outliers, additional investigation may be warranted to determine whether outliers represent areas of contamination or upper percentiles of the background data set. If it can be shown using a weight of evidence approach that an outlier does not represent background, it should be excluded from the background data set.

   There are several parametric tests for outliers. ProUCL currently includes Dixon's test and Rosner's test. Dixon's test evaluates one outlier and is used when there are less than 25 samples in the dataset. Rosner's test evaluates up to 10 outliers and is applicable for datasets with greater than 25 samples. For the Rosner's outlier test, the number of outliers to test for must be specified in advance. This number is subjective and is based on a visual assessment of the dataset and the Q-Q plots.

The following points should be considered in applying statistical tests to environmental data:

- Most environmental data sets do not have characteristics that allow the effective use of parametric tests (e.g., normally distributed data, absence of non-detects, minimum number of samples). Consequently, unless a compelling case can be made for a parametric test, non-parametric approaches such as the Wilcoxon Rank Sum (WRS) or Wilcoxon-Mann-Whitney (WMW) test may be better choices. These tests can be used to determine if site concentrations are greater than background concentrations when there is only one detection limit and no more than 40% of the measurements are non-detects. The Gehan or Tarone-Ware test can be used if there is more than one detection limit and/or if non-detects exceed detected values. These tests can also be used if there are a large number (>40%) of non-detected measurements.

- Statistical testing requires the development of a null hypothesis that contains an assumption about the relationship between the two data sets. In Background Test Form 1, the null hypothesis is that the mean chemical concentration in site samples is less than or equal to the mean concentration in background samples; i.e., the site chemical is not a contaminant. In Background Test Form 2, the null hypothesis is that the mean chemical concentration of site samples exceeds background by more than a specified concentration level; i.e., the concentrations on site reflect contamination. In general, rejection of the null hypothesis in favor of the alternative requires stronger evidence than failing to reject the null hypothesis. Thus, Form 1 requires substantial evidence before concluding that a site is contaminated, and Form 2 requires substantial evidence before concluding that a site is **not** contaminated. The FDEP recommends at a minimum the use of Form 2 of the statistical test described in the USEPA guidance cited above. This form tests the null hypothesis that the site distribution of chemical concentrations exceeds the background distribution by more than a specified difference in concentration levels. Form 1 may also be included as additional information. If the conclusions of Form 1 and Form 2 are contradictory, additional lines of evidence should be provided to support that the site data is consistent with the background data set.

- Note that Form 2 of the test requires specification of a “substantial difference” (S). The substantial difference is the value above background that represents a substantial risk from contamination. There are several ways to derive S, as summarized in Appendix A of the USEPA guidance. At present, any of the methods described in Appendix A may be used to derive S. The most commonly used method is to set S equal to one standard deviation of the background concentrations.

- In general, a *minimum* of 10 samples for both the background and site data sets is required. Greater numbers of samples may be needed, depending in part upon the confidence and power desired in the analysis. Default confidence and power specifications can be found in the USEPA guidance cited above.

- Tests should use the two sample hypothesis testing approach.

- For "non-detect" background samples, one-half the detection limit can be used as a surrogate value[1]. If the analysis is being performed in ProUCL, the detection limit should be used along with a non-detect qualifier.

- As with non-statistical approaches, comparisons should be made between site and background soil from the same soil horizon.

Two important features of a dataset include the central tendency and the dispersion (range, variance, standard deviation). Even if the average site concentration is not statistically different from background, areas of contamination may exist as small isolated areas of elevated concentrations. While the WRS and the Gehan test determine whether the means of site and background are significantly different, they are not very sensitive to identifying outliers that may be indicative of contamination.

In addition to a test of the means, the upper tails of the distributions should be compared to identify any contamination on-site. An example of an upper tail test is the derivation of an upper tolerance limit (UTL) on the background concentration. A UTL is a confidence limit on a percentile of the data, rather than the mean. The USEPA (2002) recommends the calculation of a 95% UTL with 95% coverage. Only 5% of the samples are expected to exceed the 95% UTL. The site dataset is then compared with the 95% UTL to determine the percentage of site samples that exceed this value. If more than 5% of the site samples exceed the 95% UTL, it suggests the site has a greater number of elevated concentrations than the background population. Therefore, the site population is not equivalent to background. As stated in the section below, the UTL approach should not be used in isolation to show similarities or differences between background and site concentrations. It is a test of the upper tails of the distribution and must be used with a test of the means to determine if site and background concentrations are equivalent.

Additionally, the presence of a single large value may not be identified by the test of the means or an upper tail test. However, it may be indicative of contamination in a portion of the region. Because non-parametric ranking tests and upper tail tests use only the relative rank or percentile of the concentration and not the actual value, the magnitude of the highest concentration may be masked. The use of outlier tests and visualization with Q-Q plots, box plots, and probability plots are helpful to identify single samples that are not from the same population as the rest of the site.

Graphing Procedures for Comparing Site and Background Data

Graphing procedures are useful for visualizing differences in the means, variances, and distributions of site and background datasets. They can help identify the range of concentrations in both site and background datasets and determine whether the dataset is distributed lognormally, normally, or fits another type of distribution (NAVFAC, 2002). Commonly used graphing procedures to compare site data and background data are briefly discussed in this section.

[1] The USEPA recommends using zero as a surrogate for "non-detect" values. This guidance suggests the use of one-half the detection limit to be consistent with FDEP convention. Substitution of non-detects with surrogate values instead of interpolating the values may raise some statistical issues. However, substitution is suggested here for simplicity.

1. Quantile-quantile (Q-Q) plots

   Q-Q plots are utilized as a direct visual comparison of site and background distributions. To help visualize whether site and background are from the same population, a Q-Q plot should include both site and background distributions on the same plot. A distribution of background plus the substantial difference (S; discussed below) can also be included on the plot. If the background and site populations were identical, the solid blue "best fit" lines for both datasets would lie directly on top of one another. In a real-world scenario, this is unlikely to occur. However, the lines should be similar to each other. Evaluation of a Q-Q plot to determine whether site and background populations are identical is a subjective comparison and can be used as one line of evidence to support the conclusion. It should not be used in isolation to show that site concentrations are equivalent to or exceed background concentrations. An example of a Q-Q plot displaying site and background distributions that are from the same population is included below. The "best fit" lines for these two plots appear similar and the data points are near these lines.

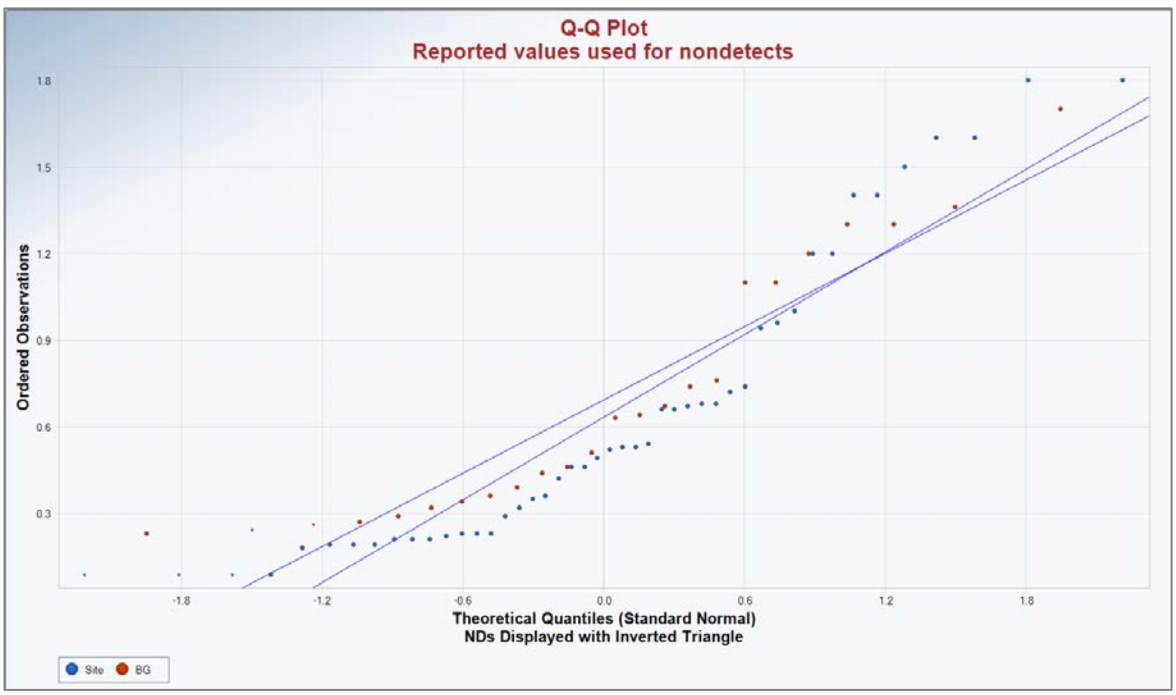

Figure 1. Quantile-Quantile plot displaying site and background distributions that are from the same population.

When site concentrations differ from background, the "best fit" lines are not similar and data points tend to be farther away from the "best fit" lines. The Q-Q plot below shows a site distribution that differs from background in the upper percentiles. While some of the concentrations on-site are similar to background, this plot demonstrates that the site distribution is shifted higher than the background distribution. It suggests some of the concentrations on site exceed background.

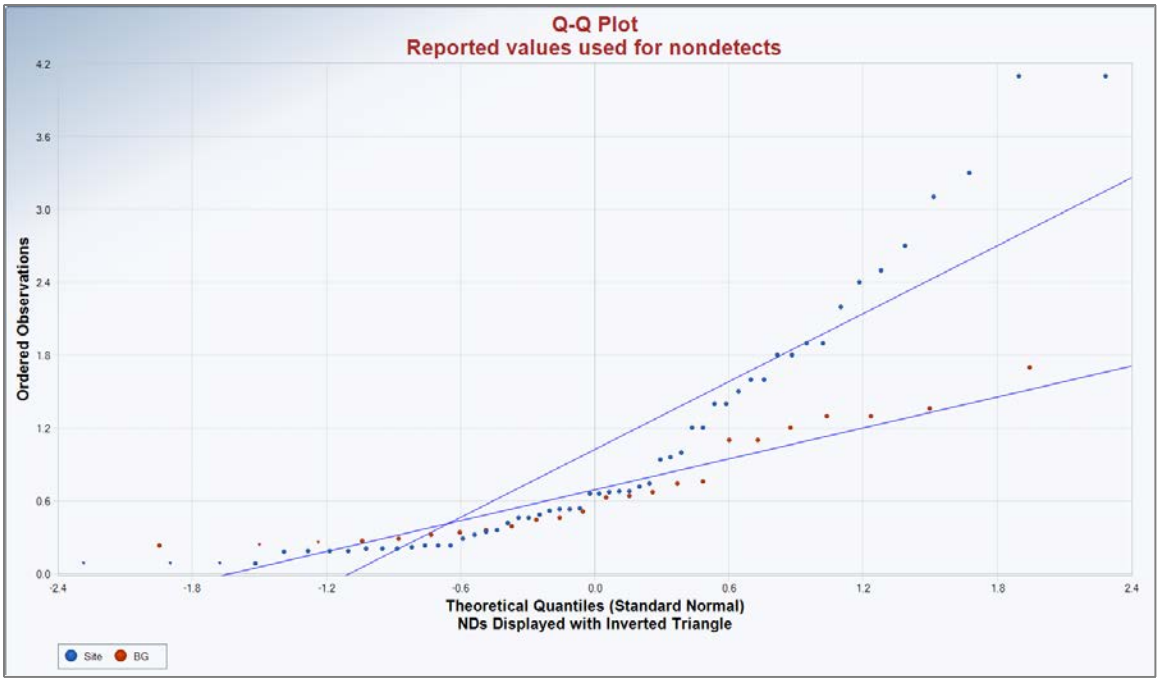

Figure 2. Quantile-Quantile plot displaying site and background distributions that are from different populations.

2. <u>Box Plots</u>
Box plots show the central tendency, shape of the distribution, variability, and potential outliers of a data set. It also includes the maximum and minimum concentrations in the data set. The data set is shown as a rectangular box that represents the middle 50 percent of the data known as the interquartile range. The upper value of the box represents the 75th percentile and the lower value of the box represents the 25th percentile. The median is represented by the middle line in the box. The mean is represented using a plus sign (+). Two vertical lines called whiskers extend out from the box. The upper whisker represents the 75th percentile plus 1.5 times the length of the 50 percent box. The lower whisker represents the 25th percentile minus 1.5 times the length of the 50th percentile box. Any data value that falls outside these ranges are plotted as asterisks. The box plot serves as a visual representation of the symmetry of the dataset and displays both the full range of data as well as summary statistics. The box plot is not intended to be the only graphic used to evaluate site and background concentrations. It should be used in addition to other statistical or graphical procedures (e.g., Q-Q plots). An example of a box plot displaying site and background distributions that are from the same population, and an example of a box plot displaying site and background distributions from different populations are included below.

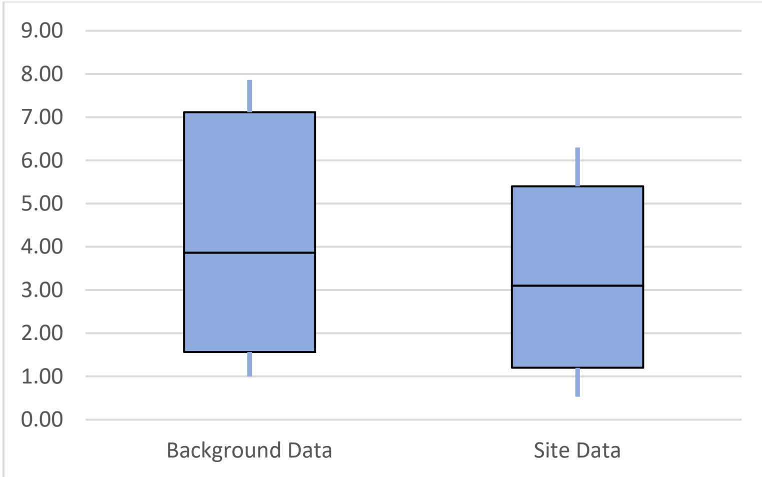

Figure 3. Box plot displaying site and background data sets that are from similar populations.

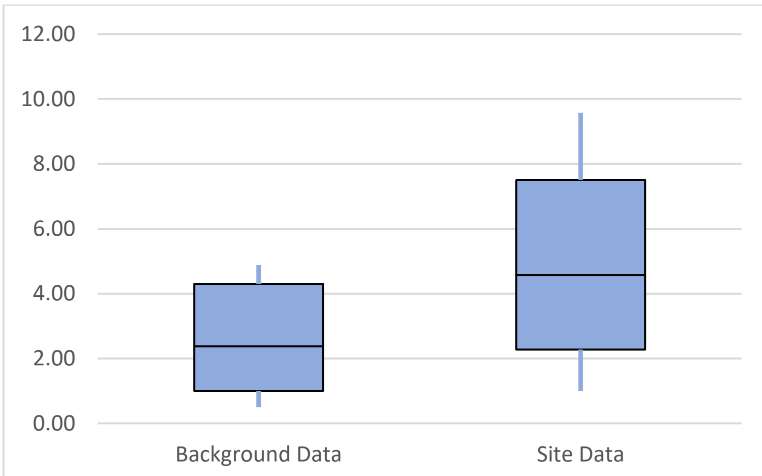

Figure 4. Box plot displaying site and background data sets that are from similar populations.

Approaches that Should Not be Used

The previous sections describe some of the approaches that are acceptable to the FDEP. Other approaches may be proposed; however, some other approaches have been evaluated and have been found not to be acceptable. The following is a partial listing of these approaches with a brief explanation of the reason(s).

1. Comparing the average (or 95% UCL) concentration on site with twice the average background. In this approach, the bases for comparison are not equivalent: one (the 95% UCL) is an expression of the average concentration on site and the other (twice the mean) is a protective estimate for the upper range of background values. Risks from soils at the site could be up to twice the background levels and still pass this test.

2. Comparison of the maximum concentration on site with an upper tolerance limit (UTL) from the background data set. The UTL is an upper confidence limit on an upper percentile of the data distribution. The UTL is discussed in US EPA guidance (USEPA, 2002) as useful in identifying outliers. There are at least two problems with using a

UTL in making comparisons with background: 1) It is sensitive to the choice of distribution to represent the data. The wrong choice of distribution can lead to significant errors in the UTL value; and 2) It is un-protective. As an upper confidence limit value on an upper percentile, the UTL is a function of uncertainty in the data. The greater the uncertainty in the data (e.g., because of limited sample size), the higher the upper confidence limit, and therefore the higher the UTL. As uncertainty in the background data set increases, it becomes easier to dismiss site contamination as representing background conditions using a UTL, when the opposite should be the case.

Under extraordinary circumstances, there are alternative approaches that may be of value (e.g., the use of geostatistical techniques). Before using any alternative approaches in comparing site and background data sets, it is advisable to consult the FDEP and gain approval in advance.

References

NAVFAC (2002) *Guidance for Environmental Background Analysis, Volume I: Soil.* Naval Facilities Engineering Command, Washington, DC.

USEPA (1989) *Risk Assessment Guidance for Superfund, Volume I, Human Health Evaluation Manual, Part A,* Office of Emergency and Remedial Response, Washington, D.C., EPA 540/1-89/002, December 1989.

USEPA (2002) *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites*, Office of Emergency and Remedial Response, Washington, D.C., EPA 540-R-01-00, September 2002.

USEPA (2015) *ProUCL Version 5.1.002. Statistical Software for Environmental Applications for Data Sets with and without Nondetect Observations*. EPA/600/R-07/041, October 2015.

Appendix

Example:

The site study team was interested in determining whether detected arsenic concentrations in surface soil were due to historical site activities or were indicative of local background. Fifteen background and fifteen site samples were available (see below). In a preliminary data analysis, descriptive summary statistics were calculated for both background and site samples (Table A-1). The Kaplan-Meier mean and standard deviation were used for both datasets because the datasets have nondetected values. Detected concentrations are labeled with a "1" and nondetected concentrations are labeled with a "0" per ProUCL format (Tables A-2 & A-3). The detection limit in this example is 1.0 mg/kg.

Table A-1. Statistics for background and site arsenic samples.

Statistic	Background	Site
Number	15	15
Nondetects	3	1
Maximum	3.8	4.0
KM Mean	2.59	2.67
KM SD	0.95	0.87

KM – Kaplan-Meier; SD – Standard deviation.

Table A-2. Arsenic samples for background. Detect = 1.

Background Arsenic (mg/kg)	Detect
1.0	0
1.0	0
1.0	0
3.5	1
2.2	1
2.9	1
2.5	1
3.4	1
2.6	1
3.5	1
3.8	1
2.4	1
3.3	1
1.9	1
2.9	1

Table A-3. Arsenic samples for site. Detect = 1.

Site Arsenic (mg/kg)	Detect
1.0	0
3.6	1
2.4	1
3.1	1
1.7	1
2.2	1
3.4	1
1.4	1
4.0	1
3.7	1
2.2	1
2.1	1
2.8	1
3.2	1
3.3	1

Next, a Q-Q plot was produced to visualize the data (Figure A-1). Based upon the Q-Q plot, site and background concentrations appear to be from the same population.

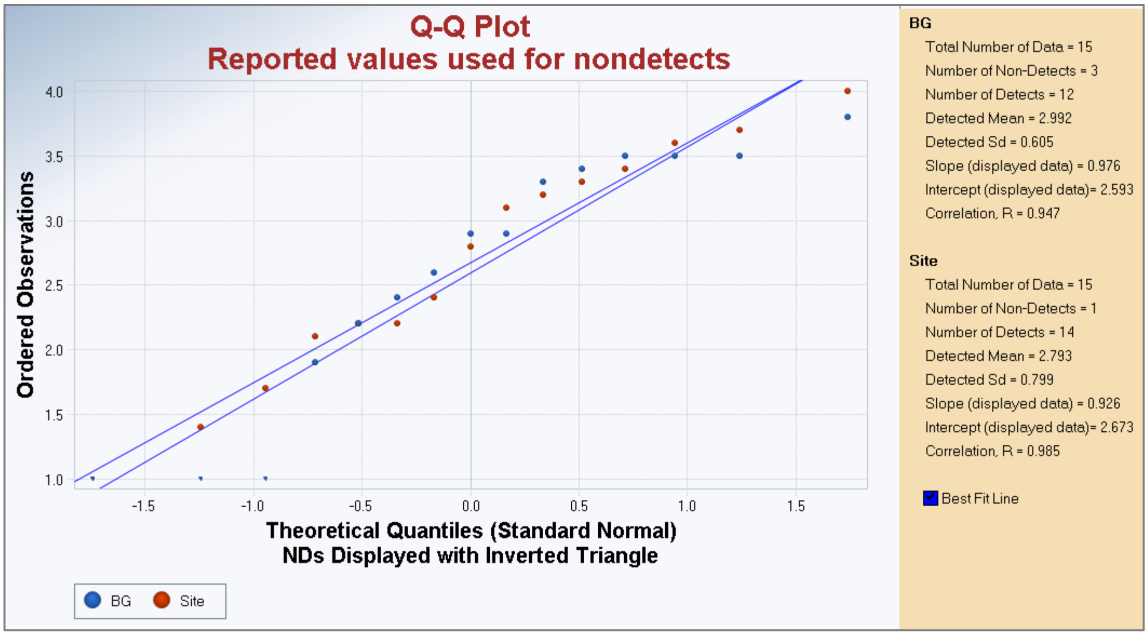

Figure A-1. Quantile-Quantile plot displaying reported values used for nondetects.

As a secondary line of evidence, outlier tests were performed for the site and background populations to determine whether the samples originated from the same population (Table A-4). The Dixon's outlier test was used because less than 25 samples are available. The outlier test for the background samples show no outliers are present. This suggests all background samples are from the same population and do not include isolated samples with low level contamination.

Table A-4. Dixon's Outlier Test parameters for background samples.

Dixon's Outlier Test for BG
Total N = 15 Number NDs = 3 Number Detects = 12 Number Data (n) = 15 10% critical value: 0.472 5% critical value: 0.525 1% critical value: 0.616 Note: NDs replaced by DL/2 in Outlier Test
1. Data Value 3.8 is a Potential Outlier (Upper Tail)?
Test Statistic: 0.091
For 10% significance level, 3.8 is not an outlier. For 5% significance level, 3.8 is not an outlier. For 1% significance level, 3.8 is not an outlier.

BG – Background; ND – Nondetect; DL – Detection limit.

An outlier test was also performed on the site samples (Table A-5). No outliers were identified. This suggests that all site samples are from the same population and there are no areas of elevated concentrations.

Table A-5. Dixon's Outlier Test parameters for site samples.

Dixon's Outlier Test for Site
Total N = 15
Number NDs = 1
Number Detects = 14
Number Data (n) = 15
10% critical value: 0.472
5% critical value: 0.525
1% critical value: 0.616
Note: NDs replaced by DL/2 in Outlier Test
1. Data Value 4 is a Potential Outlier (Upper Tail)?
Test Statistic: 0.174
For 10% significance level, 4 is not an outlier.
For 5% significance level, 4 is not an outlier.
For 1% significance level, 4 is not an outlier.

ND – Nondetect; DL – Detection limit.

Because there is only one detection limit and less than 40% of the values are non-detect, the two-sample hypothesis Wilcoxon-Mann-Whitney test was used. The study team used Background Test Form 2. In Background Test Form 2, a substantial difference between background and site distributions must be specified. In this case, the study team chose the standard deviation of the background concentrations (0.95 mg/kg) as the substantial difference (Tables A-6 & A-7). The background values are adjusted by adding the substantial difference to each value. The null hypothesis stated, "The mean arsenic concentrations at the site exceed mean background concentrations by more than 0.95 mg/kg". A confidence level of 95% (α = 0.05) was utilized in the calculation (Table A-8).

Table A- 6. Arsenic samples for background. Detect = 1.

Background Arsenic (mg/kg)	Detect
1.95	0
1.95	0
1.95	0
4.45	1
3.15	1
3.85	1
3.45	1
4.35	1
3.55	1
4.45	1
4.75	1
3.35	1
4.25	1
2.85	1
3.85	1

* – Original observation plus 0.95 mg/kg

Table A-7. Arsenic samples for site. Detect = 1.

Site Arsenic (mg/kg)	Detect
1.0	0
3.6	1
2.4	1
3.1	1
1.7	1
2.2	1
3.4	1
1.4	1
4.0	1
3.7	1
2.2	1
2.1	1
2.8	1
3.2	1
3.3	1

Table A-8. WMW comparison test ProUCL 5.15 output.

Wilcoxon-Mann-Whitney (WMW) Sample 1 vs Sample 2 Comparison Test for Data Sets with Nondetects

User Selected Options	
Date/Time of Computation	ProUCL 5.15/7/2018 10:43:53 AM
From File	background guidance data_b.xls
Full Precision	OFF
Confidence Coefficient	95%
Selected Null Hypothesis	Sample 1 Mean/Median >= Sample 2 Mean/Median (Form 2)
Alternative Hypothesis	Sample 1 Mean/Median < Sample 2 Mean/Median

Sample 1 Data: Site
Sample 2 Data: BG+S

Raw Statistics

	Sample 1	Sample 2
Number of Valid Data	15	15
Number of Nondetects	1	3
Number of Detects	14	12
Minimum Nondetect	1	1.95
Maximum Nondetect	1	1.95
Percent Nondetects	6.67%	20.00%
Minimum Detect	1.4	2.85
Maximum Detect	4	4.75
Mean of Detects	2.793	3.942
Median of Detects	2.95	4.05
SD of Detects	0.799	0.605

WMW test is meant for a Single Detection Limit Case
Use of Gehan or Tarone-Ware (T-W) test is suggested when multiple detection limits are present
All observations <= 1.95 (Max DL) are ranked the same

Wilcoxon-Mann-Whitney (WMW) Test

H0: Mean/Median of Sample 1 >= Mean/Median of Sample 2

Sample 1 Rank Sum W-Stat	182.5
WMW U-Stat	62.5
Mean (U)	112.5
SD(U) - Adj ties	24.09
WMW U-Stat Critical Value (0.05)	73
Standardized WMW U-Stat	-2.104
Approximate P-Value	0.0177

Conclusion with Alpha = 0.05
Reject H0, Conclude Sample 1 < Sample 2

BG – Background; S – Site; SD – Standard deviation; DL – Detection limit.

The test rejects the null hypothesis that mean site concentrations are greater than mean background concentrations plus a substantial difference of 0.95 mg/kg. Therefore, the test of means concludes site concentrations are representative of background. In addition to a comparison of the means, an upper tail test should be performed. To compare the upper tails of the two distributions, a 95% UTL was calculated on the original background dataset. It is important to note that a UTL should not be calculated for the adjusted background dataset. The output from ProUCL is provided in Table A-9.

Table A-9. Nonparametric upper tail test ProUCL 5.15 output.

Nonparametric Upper Limits for BTVs (no distinction made between detects and nondetects)			
Order of Statistic, r	15	95% UTL with 95% Coverage	3.8
Approx., f used to compute achieved CC	0.789	Approx. Actual CC achieved by UTL	0.537
Approx. Sample Size needed to achieve specified CC	59	95% UPL	3.8
95% USL	3.8	95% KM Chebyshev UPL	6.872

BTV – Background threshold value; CC – Confidence coefficient; UTL – Upper tolerance limit; UPL – Upper prediction limit; USL – Upper simultaneous limit.

The 95% UTL with 95% coverage for the background dataset is 3.8 mg/kg. Only 5% of site samples (one or less in a 15-sample dataset) should be above 3.8 mg/kg. At this hypothetical site, only one sample was above 3.8 mg/kg arsenic (4.0 mg/kg). Therefore, the upper tails for the site and background concentrations are also similar. Both the test of the means and the upper tail test conclude that site concentrations are representative of background. Additionally, the Q-Q plot and outlier test support there are no isolated areas of higher concentrations and that site and background appear to come from the same population. Therefore, it can be concluded that arsenic concentrations on-site are representative of background.