Guidance for
Comparing Background and Site Chemical Concentrations in Soil

Florida Department of Environmental Protection
Bureau of Waste Cleanup
Program & Technical Support Section
Tallahassee, FL

March 2008

Overview

The purpose of this guidance is to describe procedures acceptable to the Florida Department of Environmental Protection (FDEP) for the comparison of site contaminant levels to background concentrations in soil. 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 background concentrations. Some chemicals, such as inorganics and radionuclides, are present naturally in soils or may be introduced as contaminants. If they are present from a chemical release and exceed risk-based criteria, cleanup or other risk management measures are typically required. If the chemical is present due to natural soil conditions, cleanup is not needed under current rules, even if the concentrations exceed risk-based criteria. Consequently, it is important in the management of a number of sites to determine whether or not the presence of a chemical represents natural background conditions.

Some chemicals, both man-made and natural, are ubiquitous in the environment due to human activities. Examples include polycyclic aromatic hydrocarbons and dioxins. Low levels that exist in the environment due to dispersion of these chemicals are termed anthropogenic background. Current statutes and rules in Florida do not recognize comparisons with anthropogenic background concentrations as a basis for determining a chemical is not a concern for a site. However, when delineating the boundaries of contamination attributable to a release, anthropogenic background concentrations become important. They are used to help establish the area where liability for cleanup exists – where the chemical concentrations from the release become indistinguishable from concentrations present from other, non-specific sources. The procedures described in this guidance are also useful in this context (i.e., for comparing site with anthropogenic background), despite the somewhat different objective from comparisons with natural background.

According to the United States Environmental Protection Agency, background chemicals fall into the following two categories:

Naturally Occurring Chemicals- 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 Chemicals- Constituents that are synthetic or natural substances that have been released to the environment as a result of human activities, but are not related to specific activities conducted at the site. US EPA (1989b) considers the following sources anthropogenic background chemicals: agricultural runoff, septic systems, agricultural and residential application of pesticides, air pollution, industrial discharges, landfills, and urban pollution (lead and PAHs from automobiles and combustion process). 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 possible to determine whether or not a constituent is naturally occurring or anthropogenic in origin, and thus

the background data **cannot** be 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, it is best to obtain concurrence from FDEP staff before obtaining background samples. The following areas are 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.

- 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 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.

- In measuring chemical concentrations in background samples, the same analytical methods used for site samples should be employed.

- 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.

Non-Statistical Approaches for Comparing Site and Background Data

For most sites, a determination of whether site concentrations represent background conditions can be made without using statistical tests. The basic approach is to define the upper end of the range of background concentrations 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 widespread regulatory acceptance. It is simple, conservative, 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).

- Both background and site samples should be discrete rather than composite samples. 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, and background concentrations are above risk-based criteria, cleanup to background levels may be warranted.   In this situation, 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.

Another non-statistical approach involves a comparison of the 95% UCL of site samples with the 95% UCL for background. This method has technical limitations that could, in theory, lead to misclassification of contaminants as background, or not background. The approach is accepted, however, while the Department evaluates its performance in practice. Consistent with requirements for using a 95% UCL in Chapter 62-780, F.A.C., a minimum of 10 samples is needed, of which seven must be above detection limits (or three above detection limits to use the Bounding Method). This minimum sample requirement applies for both site and background samples at each soil horizon. Also consistent with Chapter 62-780, F.A.C., the maximum concentration on the site should be less than, or equal to, 3-times the default CTL. The 95% UCL approach should not be used for chemicals with CTLs based on acute toxicity when exposure scenarios with children are plausible (e.g., residential land use, parks, schools).

Statistical Approaches for Comparing Site and Background Data

If sufficient data are available, statistical methods offer a stronger, more robust method of comparing site and background data. Statistical methods for this purpose can be found in US EPA guidance (*Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites*, EPA 540-R-01-003, 2002). The following points must be considered in applying this guidance:

- 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) test should be used. The WRS test 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 nondetects. When different detection or reporting limits are present, the Gehan Test should be used. The Gehan Test also requires that no more than 40% of the measurements are nondetects. If more than 40% of the data are nondetects, a two-sample test of proportions is an appropriate statistical test. These statistical tests are discussed further in *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites*, EPA 540-R-01-003, 2002 and *Handbook for the Statistical Analysis of Environmental Background Data,* Naval Facilities Engineering Command, SWDIV and EFA WEST, 1999.

- FDEP requires at a minimum the use of Form 2 of the statistical test described in US EPA 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.

- In general, a *minimum* of 15 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 US EPA guidance cited above.

- 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 US EPA guidance. At present, S derived using any of the methods described in Appendix A will be accepted.

- Tests should be conducted as "one-tailed". Critical values for a one-tailed WRS test are calculated using the equation given in Example 1.

- For "non-detect" background samples, one-half the detection limit should be used as a surrogate value.[1]

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

- When one on-site concentration is much larger than the other values, non-parametric tests (such as the WRS test) may reject the null hypothesis that the site distribution exceeds the background distribution by a substantial difference and conclude there is no substantial difference between site and background distributions. However, the presence of this large value may be indicative of contamination in a portion of the region. Because non-parametric ranking tests use only the relative rank of the concentration and not the actual value, the magnitude of the highest concentration may be masked. Therefore, it is important to use caution when interpreting the results in identifying contamination attributable to the site.

Approaches that Should Not be Used

The approaches described in previous sections are acceptable to the Department. The following is a partial list of approaches that are **not** acceptable and a brief explanation for the reason(s).

1. <u>Comparing the average (or 95% UCL) concentration on site with twice the average background</u>. 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

---

[1] The US EPA 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.

the mean) is a conservative 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. <u>Comparison of the maximum concentration on site with an upper tolerance limit (UTL) from the background data set</u>. The UTL is an upper confidence limit on an upper percentile of the data distribution. The UTL is discussed in US EPA guidance (*Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites*, EPA 540-R-01-003, 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-conservative. 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 Department and gain approval in advance.

Appendix

Wilcoxon Rank Sum Example:

The site study team was interested in determining whether the 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. The detection limit in this example is 1.0 mg/kg. Non-detects were assigned a value of one-half the detection limit (0.5 mg/kg).

Data (mg/kg)	Source		Data (mg/kg)	Source
0.5	Background		3.4	Site
0.5	Background		3.6	Site
0.5	Background		2.4	Site
3.5	Background		3.1	Site
2.2	Background		4.3	Site
2.9	Background		2.2	Site
2.5	Background		3.4	Site
3.4	Background		1.4	Site
2.6	Background		4.0	Site
3.5	Background		3.7	Site
2.8	Background		2.2	Site
2.4	Background		4.1	Site
3.3	Background		2.8	Site
1.9	Background		3.2	Site
2.9	Background		3.3	Site

Statistic	Background	Site
Mean	2.4	3.1
Median	2.6	3.3
Range	0.5 – 3.5	1.4 – 4.3
Variance	1.15	0.65
Standard Deviation	1.07	0.81

Initially, a non-statistical test was used to assess whether site concentrations exceed the upper limit of the range of background concentrations. The upper limit of the range of background concentrations was set to the lower of twice the mean background concentration and the maximum background concentration. The maximum background concentration of 3.5 mg/kg was lower than twice the mean (4.8 mg/kg) and was therefore used as the upper limit of the range of background concentrations. The maximum arsenic concentration in soil detected on-site, 4.3 mg/kg, exceeded the 3.5 mg/kg background limit, suggesting that concentrations found on-site were not indicative of background conditions.

The site study team decided to use a statistical test to more rigorously compare site concentrations to background concentrations. The Wilcoxon Rank Sum test, a non-parametric statistical test, was chosen for this analysis. This test is used when there is only one reporting limit and the reporting limit is smaller than the smallest reported concentration.

Two different null hypotheses were tested, one using Background Test Form 1 and the other using Background Test Form 2. In Background Test Form 1, the null hypothesis states that the mean chemical concentration in site samples is less than or equal to the mean concentration in background samples; the chemical is not a contaminant. In Background Test Form 2, the null hypothesis states the mean chemical concentration of site samples exceeds background by more than a specified concentration level; 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 study team began by using the less conservative Background Test Form 1. The null hypothesis stated, "The mean arsenic concentration on-site is less than or equal to the mean arsenic concentration in background areas." The study team used a confidence level of 95% ($\alpha = 0.05$). To calculate a Wilcoxon Rank Sum (WRS) Test, the background and site data sets were combined in ascending order (non-detects were assigned a value of one-half the detection limit). Rank was assigned to the values in ascending order with an average rank used for equal values. The ranks were then summed separately over background and site data sets:

		Site Rank	Background Rank
0.5	Background		2
0.5	Background		2
0.5	Background		2
1.4	Site	4	
1.9	Background		5
2.2	Background		7
2.2	Site	7	
2.2	Site	7	
2.4	Background		9.5
2.4	Site	9.5	
2.5	Background		11
2.6	Background		12
2.8	Background		13.5
2.8	Site	13.5	
2.9	Background		15.5
2.9	Background		15.5
3.1	Site	17	
3.2	Site	18	
3.3	Background		19.5
3.3	Site	19.5	

3.4	Background		22
3.4	Site	22	
3.4	Site	22	
3.5	Background		24.5
3.5	Background		24.5
3.6	Site	26	
3.7	Site	27	
4.0	Site	28	
4.1	Site	29	
4.3	Site	30	
	Rank Sum:	279.5	185.5

For Background Test Form 1, the sum of the ranks of the site values is the test statistic. Tables of WRS critical values are available for small samples. For samples of at least 10 for both the background and test sites, the WRS critical value for a one-sided test may be approximated using the following equation:

$$W_{crit} = \frac{n_S(N+1)}{2} + z_\alpha \left[\frac{n_S n_B (N+1)}{12} \right]^{1/2}$$

where W_{crit} is the critical value for the WRS test, n_S is the number of measurements in the site sample, n_B is the number of measurements in the background sample, $N = n_S + n_B$, and z_α is the $100(1-\alpha)^{th}$ percentile of the standard normal distribution. A table of common z_α values is included below:

Confidence Level (α)	z_α value
0.20	0.842
0.15	1.039
0.10	1.282
0.05	1.645
0.01	2.326

In this example $n_B = 15$, $n_s = 15$, $N = 30$, and $z_\alpha = 1.645$ (since the study team used a confidence level of 95% ($\alpha = 0.05$)). The calculated critical value using these site-specific numbers is 272. Because the sum of site ranks, 279.5, is greater than the critical value of 272, the null hypothesis that mean site concentrations of arsenic are less than or equal to mean background concentrations is rejected at the 95% confidence level. Therefore, the site values are not representative of background and indicate contamination.

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 (1.1 mg/kg) as the substantial difference. 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 1.1 mg/kg". A confidence level of 95% (α = 0.05) was utilized in the calculation:

Data* (mg/kg)	Source	Data (mg/kg)	Source
1.6	Adjusted Background	3.4	Site
1.6	Adjusted Background	3.6	Site
1.6	Adjusted Background	2.4	Site
4.6	Adjusted Background	3.1	Site
3.3	Adjusted Background	4.3	Site
4.0	Adjusted Background	2.2	Site
3.6	Adjusted Background	3.4	Site
4.5	Adjusted Background	1.4	Site
3.7	Adjusted Background	4.0	Site
4.6	Adjusted Background	3.7	Site
3.9	Adjusted Background	2.2	Site
3.5	Adjusted Background	4.1	Site
4.4	Adjusted Background	2.8	Site
3.0	Adjusted Background	3.2	Site
4.0	Adjusted Background	3.3	Site

* - Original observation plus 1.1 mg/kg

After adding the substantial difference to background, the adjusted background and site data sets were combined in ascending order. Rank was assigned to the values in ascending order with an average rank used for equal values. The ranks were then summed separately over adjusted background and site data sets:

		Site Rank	Background Rank
1.4	Site	1	
1.6	Adjusted Background		3
1.6	Adjusted Background		3
1.6	Adjusted Background		3
2.2	Site	5.5	
2.2	Site	5.5	
2.4	Site	7	
2.8	Site	8	
3.0	Adjusted Background		9
3.1	Site	10	
3.2	Site	11	
3.3	Site	12.5	
3.3	Adjusted Background		12.5
3.4	Site	14.5	
3.4	Site	14.5	
3.5	Adjusted Background		16
3.6	Site	17.5	
3.6	Adjusted Background		17.5
3.7	Site	19.5	

		Site	Adjusted Background
3.7	Adjusted Background		19.5
3.9	Adjusted Background		21
4.0	Site	23	
4.0	Adjusted Background		23
4.0	Adjusted Background		23
4.1	Site	25	
4.3	Site	26	
4.4	Adjusted Background		27
4.5	Adjusted Background		28
4.6	Adjusted Background		29.5
4.6	Adjusted Background		29.5
Rank Sum:		200.5	264.5

For Background Test Form 2, the sum of the ranks of the adjusted background measurements is the test statistic. For large samples, the WRS critical value for a one-sided test may be approximated by

$$W_{crit} = \frac{n_B(N+1)}{2} + z_\alpha \left[\frac{n_S n_B(N+1)}{12} \right]^{1/2}$$

Note that n_B appears in the first term of the computation for W_{crit} for Background Test Form 2 because the sum of the background ranks is the test statistic whereas n_S appears in the first term of the computation for W_{crit} for Background Test Form 1 because the sum of the site ranks is the test statistic. For $\alpha=0.05$ and $n_B = n_S = 15$ is 272. Because the sum of adjusted background ranks, 264.5, is less than the critical value of 272, the null hypothesis that mean site concentrations of arsenic exceed mean background concentrations by more than 1.1 mg/kg cannot be rejected at the 95% confidence level. Therefore, there is insufficient evidence to reject the claim that the mean of the site values is more than 1.1 mg/kg greater than the mean of the background values. The failure to reject the null hypothesis for Background Test Form 2 results in the same conclusion as rejecting the null hypothesis for Background Test Form 1. In both cases, the conclusion is that the site is contaminated.

In this example, statistical and non-statistical approaches yielded the same answer – that the site concentrations are significantly greater than background and thus do not represent background conditions.