APPENDIX M — DEVELOPMENT OF 2007 BACKGORUND METALS DATASET AND STATISTICAL COMPARISON TO PGOU METALS DATA

M1.0 INTRODUCTION

Comparisons to background are one method for focusing on constituents that require further quantitative analyses related to site characterization and risk assessments. Accordingly, establishment of an acceptable background data set is a determining step in the Remedial Investigation/Feasibility Study (RI/FS) process.

The purpose of this informal memorandum is to summarize efforts to date and describe the approach for establishing and applying a background soils data set (for metals) for the Boundary Operable Unit (BOU).

M2.0 UNDERSTANDING TO DATE

The results of pre-1991 investigations suggested that large areas, in and around potential source sites, had elevated levels of trace metals such as cadmium and thallium, even though historic site use information indicated that these metals had never been used or released at the site.

In 1994, Aerojet established background (non-anthropogenic) trace metals concentrations in plant-wide and adjacent area surface soils. The procedure used in this investigation differed in some respects from United States Environmental Protection Agency's (USEPA) total digestion procedure (USEPA Method 3052), but the results were considered to be comparable.

In 2004, the Agencies questioned the comparison of the 1994 Aerojet background data set to site data set. The Agencies noted that the analytical procedure used in the 1994 investigation differed significantly from the method used on samples collected during the RI/FS.

In 2006-2007, Aerojet re-assessed background soil concentrations. Background soil samples were meticulously processed and analyzed for 19 metals using USEPA Method 6010 (Aerojet, 2007). Upon review of 2006 results, archived 2006 background soil samples were re-analyzed using

Method 6020 to achieve a more accurate characterization of background concentrations for the following nine metals (Aerojet 2007):

- Antimony;
- Arsenic;
- Cadmium;
- Chromium;
- Lead;
- Molybdenum;
- Selenium;
- Silver; and
- Thallium.

The *Background Metals in Xerorthents and Redding-Corning-Red Bluff Surface Soils at the Aerojet Superfund Site Main Plant , Sacramento, California* (hereafter referred to as the 2007 Background Report) provides a description of Aerojet's efforts to date in establishing background (non-anthropogenic) metal concentrations to support site characterization effort at the facility. Background data reported in the 2007 Background Report have been review and approved by the Agencies.

M2.1 BACKGROUND DATA: USEPA 6010 AND USEPA 6020 DATA

Re-analysis of background soil samples resulted in two results (USEPA 6010 and USEPA 6020) for the following nine metals (Aerojet 2007):

- Antimony;
- Arsenic;
- Cadmium;
- Chromium;
- Lead;
- Molybdenum;
- Selenium;
- Silver; and
- Thallium.

Consequently, at this time, there are potentially four background data sets:

1. Xerothent 6010 data;
2. Xerothent 6020 data (for nine metals only);
3. Redding-Corning-Red Bluff (RCRB) soil 6010 data; and
4. RCRB soil 6020 data (for nine metals only).

Use of all four data sets to identify "elevated" metals will lead to a complicated and confusing application and interpretation of background comparisons. For example, there is the possibility that some site soil samples may be found to be comparable to background, while other site soil samples with similar concentrations may be found to be "elevated" simply due to differences in analytical methods.

To date, background concentration results and descriptive statistics have merely been reported. No direction has been proffered as how to interpret or use the data generated by these two methods to establish background concentrations and identify "elevated" metals for the Aerojet Facility.

M3.0 ESTABLISHING A BACKGROUND DATA SET

The proposed resolution for establishing a background data set and using this data to identify elevated metals is consistent with protocols currently being used at Aerojet and with USEPA guidance.

To promote a straightforward establishment of background data, results from background soil samples analyzed using both analytical methods were considered to be "duplicate" results. Consistent with field duplicate protocols, the following decision rule was applied (see also Figure C-1):

1. When both duplicates are non-detects, the lesser of the detection limits is used;
2. When 1 of 2 duplicates is a non-detect, the detected value is used;
3. When both duplicates are detects, the greater of the detect values is used; and
4. Duplicates include field duplicates as well as method-related duplicates.

Figure M-1. Duplicate Decision Rule

		Duplicate 1	
		ND	Detect 1
Duplicate 2	ND	lesser DL	Detect 1
	Detect 2	Detect 2	greatest detect

This decision rule promotes a clear, conservative approach for handling duplicate samples (both field duplicates and method-related duplicates). Moreover, the decision rule is consistent with the intent of using the best data—i.e., data derived from analytical methods with lower detection limits, rendering more accurate results.

Note that the 2007 Background report eliminated outliers based on statistical analyses alone. Guidance recommends that outliers be identified using statistical tests, but that elimination of data should be based on a weight-of-evidence approach that includes geological and geochemical information. Given a review of the data, ERM geologists could not defend "trimming" the background data set.

Background descriptive statistics for the two soil types are provided in Table C-1.

M4.0 APPROACH FOR EVALUATING WHETHER CONCENTRATIONS ARE "ELEVATED" AS COMPARED TO BACKGROUND

The approach for evaluating whether concentrations are elevated as compared to background will test the following null hypothesis:

> *Concentrations at the management area are less than or comparable to concentrations at background locations.*

Methods for identifying "elevated" concentrations will use (i) visualizations[1] (i.e., box plots), (ii) two-sample statistical tests (i.e., Gehan and Quantile tests), and (iii) best geological understanding of the Site and is summarized in Figure C-2. This approach is consistent with guidance (USEPA, 2002) and the 2007 Background Report.

[1] Note that box plots could not be generated for numerous data sets due to inexplicable plotting program errors reported by ProUCL 4.1.01.

Table C-1. Descriptive Statistics for Background Surface Soils

Soil/Metal	N	6020 vs 6010		Non-Detect				Detect		All Data								$BTV_{95/95}$
		6020 Freq	6020 %Freq	#ND	%ND	Min ND	Max ND	Min	Max	Min	Median	UTL 95/95	Max	Mean	Std Dev	CV	Distrib	
Xerothents																		
Aluminum	39	–	--	0	0%	--	--	12000	54000	12000	20000	54000	54000	24026	11275	47%	Non-param	54000
Antimony	39	39	100%	13	33%	0.13	0.29	0.087	0.64	0.087	0.21	0.46	0.64	0.23	0.11	46%	Lognormal	0.46
Arsenic	39	13	33%	0	0%	--	--	2.5	37	2.5	5.6	37	37	7.3	6.2	85%	Non-param	37
Barium	39	–	--	0	0%	--	--	70	320	70	130	333	320	150	67	45%	Lognormal	320
Beryllium	39	–	--	0	0%	--	--	0.24	1.1	0.24	0.46	1.1	1.1	0.54	0.24	45%	Non-param	1.1
Cadmium	39	17	44%	0	0%	--	--	0.05	1.2	0.05	0.36	1.5	1.2	0.38	0.26	69%	Lognormal	1.2
Chromium	39	7	18%	0	0%	--	--	36	130	36	59	118	130	67	21	32%	Lognormal	118
Cobalt	39	–	--	0	0%	--	--	8.9	33	8.9	13	33	33	16	6	41%	Non-param	33
Copper	39	–	--	0	0%	--	--	23	76	23	42	81	76	46	14	30%	Lognormal	76
Lead	39	22	56%	0	0%	--	--	4.5	27	4.5	9.4	23	27	10	5	45%	Lognormal	23
Mercury	39	–	--	4	10%	0.0031	0.0031	0.0057	0.23	0.0031	0.018	0.13	0.23	0.036	0.044	122%	Lognormal	0.13
Manganese	39	–	–	0	0%	--	--	260	1500	260	490	1500	1500	627	361	57%	Non-param	1500
Molybdenum	39	21	54%	11	28%	0.17	0.21	0.11	0.74	0.11	0.22	0.63	0.74	0.28	0.16	59%	Non-param	0.63
Nickel	39	–	--	0	0%	--	--	23	140	23	40	140	140	47	23	49%	Non-param	140
Selenium	39	37	95%	26	67%	0.20	0.27	0.23	57	0.20	0.23	26	57	3.6	11	298%	Lognormal	26
Silver	39	38	97%	0	0%	--	--	0.023	2.4	0.023	0.09	2.4	2.4	0.18	0.39	221%	Non-param	2.4
Thallium	39	18	46%	0	0%	--	--	0.1	3.3	0.10	1.2	3.3	3.3	1.1	1.0	91%	Non-param	3.3
Vanadium	39	–	--	0	0%	--	--	43	130	43	75	141	130	80	23	29%	Lognormal	130
Zinc	38	–	--	0	0%	--	--	35	89	35	48	89	89	53	15	29%	Non-param	89
RCRB																		
Aluminum	37	–	--	0	0%	--	--	7000	43000	7000	15000	40976	43000	17227	8491	49%	Lognormal	40976
Antimony	37	31	84%	0	0%	--	--	0.065	1.3	0.065	0.19	1.3	1.3	0.27	0.25	90%	Non-param	1.3
Arsenic	37	1	3%	0	0%	--	--	0.89	11	0.89	2.8	11	11	3.7	2.5	67%	Lognormal	11
Barium	37	–	--	0	0%	--	--	36	310	36	130	274	310	139	63	45%	Normal	274
Beryllium	37	–	--	0	0%	--	--	0.17	1	0.17	0.44	1.0	1.0	0.47	0.18	39%	Lognormal	1.0
Cadmium	37	5	14%	0	0%	--	--	0.036	0.89	0.036	0.29	0.72	0.89	0.33	0.18	56%	Normal	0.72
Chromium	37	3	8%	0	0%	--	--	13	110	13	33	111	110	41	23	56%	Lognormal	110
Cobalt	37	–	--	0	0%	--	--	3.7	26	3.7	11	24	26	13	5.5	44%	Normal	24
Copper	37	–	--	0	0%	--	--	11	63	11	23	63	63	27	14	53%	Non-param	63
Lead	37	18	49%	0	0%	--	--	4.5	42	4.5	10	42	42	11	7	59%	Non-param	42
Mercury	37	–	--	0	0%	--	--	0.0037	1.4	0.0037	0.028	0.87	1.4	0.11	0.24	210%	Lognormal	0.87
Manganese	37	–	–	0	0%	--	--	110	1100	110	500	1112	1100	548	263	48%	Normal	1100
Molybdenum	37	27	73%	0	0%	--	--	0.1	0.55	0.10	0.28	0.52	0.55	0.30	0.11	35%	Normal	0.52
Nickel	37	–	--	0	0%	--	--	6.2	75	6.2	17	79	75	24	17	74%	Lognormal	75
Selenium	37	28	76%	32	86%	0.20	0.29	0.22	2.1	0.20	0.22	1.1	2.1	0.33	0.38	115%	Normal	1.1
Silver	37	31	84%	1	3%	0.022	0.022	0.020	0.14	0.020	0.038	0.13	0.14	0.052	0.037	71%	Non-param	0.13
Thallium	37	20	54%	0	0%	--	--	0.1	2.5	0.10	0.13	2.5	2.5	0.52	0.64	124%	Non-param	2.5
Vanadium	37	–	--	0	0%	--	--	27	130	27	65	119	130	68	24	35%	Normal	119
Zinc	37	–	--	0	0%	--	--	18	100	18	39	105	100	42	20	48%	Lognormal	100

Note:

- Consistent with the background report (Aerojet 2007), the 95UCL on the 95th percentile were used to characterize an upper threshold value for background.
- Background threshold value (BTV) is the 95UCL on the 95th percentile or the maximum concentration, whichever is less
- Note that two-population tests (that use the entire data set) will be used to determine whether metal concentrations at management units are elevated as compared to background concentrations

M4.1 VISUALIZATIONS

Histograms and/or box plots provide a visual picture of the symmetry or asymmetry of the data, communicate differences in the distribution of site and background data, and are useful when initially exploring the similarity between data sets (Figures C-2 and C-3).

Note that challenges were encountered when plotting the data using ProUCL v.4.1.01. Graphs of data are provided for MAs only when ProUCL was able to plot data.

Figure M-2. Example Histogram

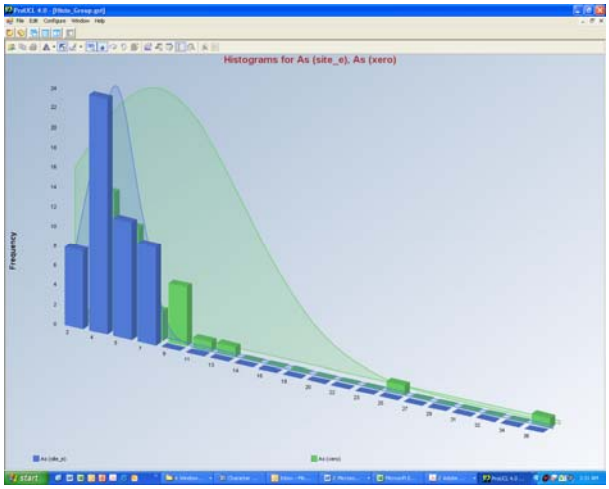
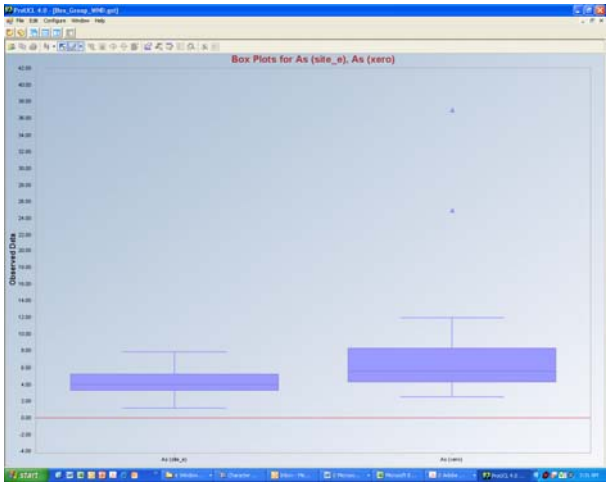


Figure M-3. Example Box Plot



M4.2 BACKGROUND THRESHOLD VALUES

Use of background threshold values (BTVs) is recommended when the number of samples is low (USEPA 2002, 2007). For management areas with numbers of site soil samples less than eight, the maximum detected concentration will be compared to a 95/95 BTV[2]. Site characterization effort will also use BTVs to assess the adequacy of sampling at management areas.

Note that the use of comparisons to background percentiles often results in erroneously identifying elevated concentrations of metals (DON, 2002). When the site and background distributions are identical, it can be shown

[2] The 95 upper confidence limits on the 95th percentile or the maximum detected concentration, whichever is less. Use of the 95/95 BTV is consistent with the 2007 Background Report.

that the probability that one or more of *n* site measurements will exceed the background threshold defined as the *p*th percentile is equal to $1 - (p)^n$, where *n* is the sample size and *p* is the *p*th percentile of the background distribution (DON, 1999, 2002).

The probability of obtaining at least one value greater than the threshold = $1\text{-}p^n$
(DON 2002)

N	80%	90%
1	20%	10%
2	36%	19%
3	49%	27%
4	59%	34%
5	67%	41%
6	74%	47%
7	79%	52%
8	83%	57%
9	87%	61%
10	89%	65%
11	91%	69%
12	93%	72%
15	96%	79%
20	99%	88%
25	100%	93%

For example, when sample size is five (5) and comparisons to the 90th percentile are made, there is a 41 percent chance of erroneously identifying a metal that is elevated—the error is 67 percent for comparisons to the 80th percentile.

M4.3 TWO-SAMPLE TESTS

Given the errors associated with comparisons to percentiles, two-sample tests were considered for comparisons to background. The use of two-sample tests is common in many environmental applications to evaluate

whether soil concentrations at an AOI are less than or comparable to concentrations in soils at background locations[3]. In addition, the use of two-sample statistical tests is consistent with USEPA guidance (2002); and supported by USEPA's (2007) ProUCL version 4.0.1. Several two-sample statistical tests are supported by USEPA's ProUCL 4.0.1, including:

- *t*-test,
- Wilcoxon-Mann-Whitney test,
- Gehan test, and
- Quantile test.

In accordance with USEPA guidance (2002, 2007), the following factors for both the AOI and background data sets were considered when selecting the appropriate two-sample tests:

- Sample size;
- Distribution of the data;
- Equal or unequal variances;
- Percent non-detected (censored) values; and
- Multiple values or a single value for censored data.

Note that natural variance, as characterized by the coefficient of variance (CV), is commonly observed to have values of 150 percent or higher. Although not a requirement of nonparametric tests, the magnitude of variation associated with the data sets will be characterized because large variances can affect the ability of the WMW and Gehan tests to detect differences.

To conduct two-sample tests, it is preferable that each of the data sets being compared contain at least eight observations (USEPA 2007). A decision tree for the selection of the appropriate method for comparing AOI to background data sets is provided in Figure C-4. Note that data sets are considered normally distributed if (i) raw data are normally distributed or (ii) the same function can be used to transform both data sets to be normally distributed. USEPA (2007) recommends that the Quantile Test be conducted in conjunction with either the Wilcoxon-Mann-Whitney or Gehan Test.

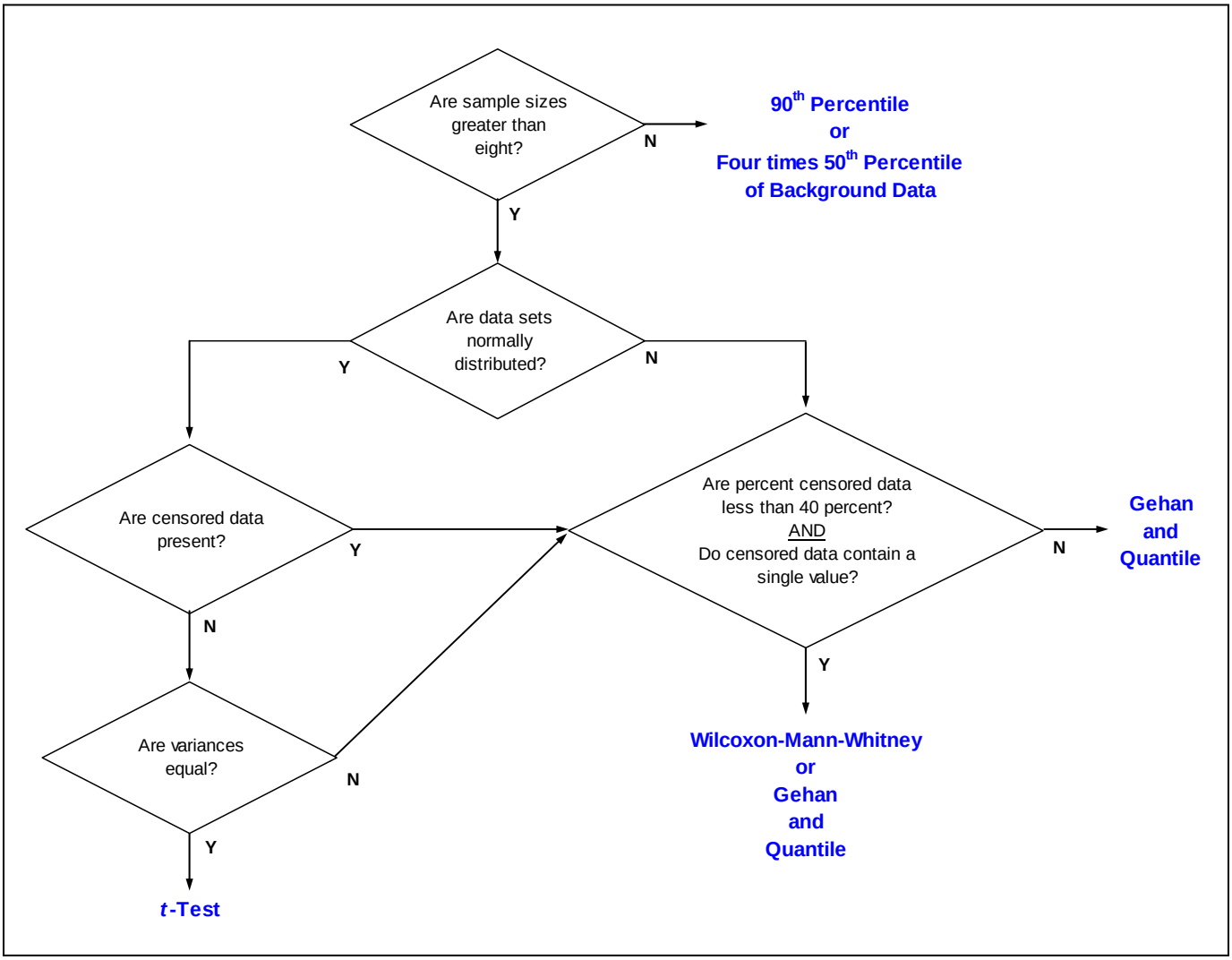
---

[3] Specifically, two-sample tests evaluate the null hypothesis that concentrations of metals at MAs are less than or comparable to concentrations at background locations.

M4.3.1 t-Test

The t-test is a parametric two-sample test that evaluates the differences in the means of the site and background data sets. The t-test requires that both data sets are normally distributed and have comparable variances[4]. This test cannot accommodate data sets with large percent non-detects or data sets with multiple detection limit values—in fact, USEPA (2007) recommends that non-parametric test be used for data sets with non-detected (censored) values. Given its requirements, the t-test is not often applicable for comparisons of environmental data. However, it is among the most well-known statistical tests (DON 2002).

Figure M-4. Approach for Selecting Comparison to Background Method



[4] USEPA (2007) recommends that the Satterthwaite *t*-test or a non-parametric two-sample test be used when unequal variances are present.

M4.3.2 Wilcoxon-Mann-Whitney Test

The Wilcoxon-Mann-Whitney (WMW) test is a nonparametric two-sample test that evaluates the differences in the medians of the site and background data sets. The WMW test can accommodate data sets with moderate percent non-detects, but is not applicable for data sets with multiple detection limit values. The test assumes the same censoring mechanisms apply to the site and background data sets. Where appropriate, the WMW test is used because it is easy to explain and is a relatively well-known test to environmental professionals (DON, 2002).

M4.3.3 Gehan Test

The Gehan test is a nonparametric two-sample test[5] that evaluates the differences in the medians of the site and background data sets. The Gehan test can accommodate data sets with large percent non-detects with multiple detection limit values. The test assumes the same censoring mechanisms apply to the site and background data sets. The Gehan test is commonly selected because it is easy to explain and because the Gehan test reduces to the Wilcoxon Rank Sum (WRS) test, which is a relatively well-known test to environmental professionals (DON, 2002) (Table C-2). Note that the Gehan test was used to perform comparisons to background for all MAs because:

- The distribution types for either or both the background and MA data sets were nonparametric;
- The distribution types for the background and MA required different transformations to produce normally distributed data;
- When a nonparametric test was required, the percent of non-detects was greater than 40 percent in either the background or MA data set;
- When a nonparametric test was required, multiple detection limits were observed in either the background or MA data set; or
- When both the background and MA data sets were normally distributed, ProUCL reported unequal variances and recommended a nonparametric test be conducted.

The use of the Gehan test rather than the WRS test is common practice given the Gehan test reduces to the WRS test for single detection limits and less than 40 percent nondetected values (i.e., left-censored data)..

---

[5] Parametric tests assume normal distributions for site and background data sets. Nonparametric tests make no assumptions with regard to the shape of the site and background concentration data sets

Table M-2. Features of the Gehan Test

Feature	WRS	Gehan
Test type	distribution-free	distribution-free
Handling NDs	< 40% ND	> 40% ND
Handling DLs	single DL	multiple DLs
Calculations	simple	complex
Performance	comparable to t-test	not as well known

M4.3.4 Quantile Test

The Quantile test is a nonparametric test that examines only the largest few background and site measurements (i.e., the right tail of the site and background concentration distributions). Accordingly, this test is considered to be appropriate when knowledge suggests that potential releases were not spread across the site. The Quantile test cannot be applied if less-than values are among the largest background and site data. The Quantile test is commonly selected because it is relatively easily to perform and complements the Gehan test (USEPA 2002; DON 2002) (Table C-3).

Table C-3. Features of the Quantile Test

Limitation	Quantile	Slippage
Cannot compute or test is inconclusive	If ND values are among largest data values	If maximum ND value is greater than maximum detected background value

Both the Gehan and the Quantile tests will be applied as appropriate and when background and site sample sizes are eight (8) or greater.

M4.4 GEOLOGICAL AND GEOCHEMICAL KNOWLEDGE

Use of the best understanding of site conditions is critical in providing perspective for the findings of statistical analyses. Geological and geochemical knowledge will be used in concert with statistical analyses to identify constituents with elevated concentrations and/or to evaluate whether there is a geochemical basis for removing outliers.

M4.5 ADVANTAGES OF APPROACH

A significant advantage in this approach is that there is a single background data set for each soil type. The use of a single data simplifies comparisons to background and is unlikely to lead to a complicated and potentially confusing characterization of management areas and assessment of potential risks.

Given that soils at most management areas are comprised of a single soil type, identification of elevated metals is further simplified. Note that both site characterization and risk assessment will identify "elevated" concentrations using the same two-sample tests – i.e., the Gehan test and Quantile test. Given the anticipated sample sizes and percent non-detected values in the data, these statistical methods provide a defensible and robust test of the following hypothesis:

> *Concentrations at the Management Area are less than or comparable to concentrations at background locations.*

Use of the Gehan and Quantile tests are consistent with guidance (USEPA 2002, 2007) and the 2007 Background Report.

M6.0 LITERATURE CITED

Aerojet. 2007. *Background Metals in Xerorthents and Redding-Corning-Red Bluff Surface Soils at the Aerojet Superfund Site Main Plant, Sacramento, California*. Prepared by Tetra Tech, GeoSyntec, and Aerojet.

U.S. Department of the Navy (DON). 2002. *Guidance for Environmental Background Analysis. Volume I: Soil. NFESC User's Guide*. UG-2049-ENV, NAVFAC. Washington, D.C.

United States Environmental Protection Agency (USEPA). 2002. *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites*. EPA 540-R-01-003. Office of Emergency and Remedial Response. Washington, D.C.

USEPA. 2007. *ProUCL Technical Guidance*. EPA/600/R-07/041. Office of Research and Development. Las Vegas, NV.