Response to NDEP Comments Received February 17, 2009 on the Supplemental Shallow Soil Background Report dated December 2008

General Comments:

1. From the previous round of revisions the first general comment was not entirely addressed. There are still several instances in the text where the phrase "At the direction of NDEP…" still exists (pages 3-7, 3-18, 3-20). This phrase is not necessary. It is not clear why BRC includes this phrase. Please discuss with NDEP if necessary.

Response: *This phrase was retained in previous versions of the report when NDEP requested statistical analyses that were not identified or proposed in existing state or federal guidance. Given BRC agreement to perform these analyses at NDEP's request, the subject phrase has been deleted from this version of the report.*

2. The objectives as stated in Section 1.1 seem on target. The basic goal is to determine if the northern River range geology is different enough that a local background dataset corresponding to that area is different than the background dataset reported in the 2005 BRC/TIMET background report. The final conclusion verifies that this is the case, but there are other ancillary conclusions that do not seem necessary. The focus should be on whether the supplemental background dataset is statistically similar to or different to the 2005 BRC/TIMET background data, while also bearing in mind the differences within the 2005 BRC/TIMET background data. Some specific comments on this issue are also provided below.

Response: *A key objective of this study is to evaluate whether the supplemental shallow soil background dataset is statistically similar to or different to the 2005 BRC/TIMET background data. Text has been modified to focus on this objective.*

3. Although the final conclusion of the statistical analysis is that there are differences, and the final table in Section 4 suggests that sub-sets of the background data that could be used for different sub-areas, more should be made of the overall result of the background studies that a rich background dataset has been assembled that covers several different soil geologies at the site, and that for each sub-area background comparison the appropriate sub-set of the background data should be used. This should also be extended to differences by depth as necessary.

Response: *The overall robustness of the assembled background soil data will be identified and described in the upcoming Background Soil Summary Report. Note that the findings of this study found few statistically significant differences among the 0, 5, and 10 ft bgs depth intervals for the 2008 River background data. As suggested in the report, the 0, 5, and 10 ft bgs data may be pooled and applied as a single dataset, promoting more powerful statistical analyses for future assessments in support of decision-making.*

4. Overall, more emphasis should be placed on the conclusion that the background data differ by geology, with minor differences by depth, and that appropriate sub-sets of the background

data should be identified for sub-area background comparisons. This is not explicitly clear within the report, however, it is expected that this issue can be resolved within the forthcoming report which will encompass all of the background data sets.

Response: *Text in Section 4 has been revised to emphasize that background data differ by geology, with minor differences by depth within the 2008 River dataset. Recommendations for the use of specific datasets is provided in Section 4, Summary and Conclusions.*

5. The results of the semi-volatile organic compound (SVOC) analyses are not discussed in the report until a one line mention in the conclusions of Section 4. There was a purpose to collecting these data, and some discussion of the results is warranted. There is also some discussion under Criterion V in the Data Usability Section, but this is inadequate. The results need to be discussed in more detail in Section 3.

Response: *BRC has expanded the discussion in Section 2.4 in response to NDEP's comment.*

In BRC's opinion, presentation of these results under Criterion V in the Data Usability Section (Section 2.4 - with a table summarizing the results, Table 3), separate and apart from the discussion of the metals and radionuclide results, is appropriate given (1) the purpose of the analyses (i.e., as indications of the potential for impacts to the sampling location that could suggest a certain location should be excluded from the background dataset); (2) the fact that there is no intent to establish background SVOC concentrations for comparison to detections at the site; and (3) the general lack of SVOC detections (only bis(2-ethylhexyl)phthalate, a common laboratory contaminant, was reported).

Furthermore, Section 3 comprises the summary of statistical analyses performed on the background datasets. Because statistical analyses were not performed on the SVOC data, including discussion of those data within Section 3 seems inappropriate. Thus, discussion of the SVOC results will be confined to Section 2.4.

6. Some of the Data Usability sections are inadequate. For example, for Criteria II and II no references are provided demonstrating that these criteria were met. There is discussion, but no references to where the relevant information is presented. Some further comments are made in the specific comments below.

Response: *See responses in Specific Comment #12 below.*

Specific Comments:

7. Page 1-2; last paragraph (after bullets). It is not clear in this document what "Qr1" and "Qr2" refer to. Please clarify.

Response: *The subject sentence has been expanded to provide an explanation of the terms Qr1 and Qr2 (mapped lithologic units representing pediment and fan deposits of the River Mountains).*

8. Page 1-3; last paragraph. In this paragraph reference is made to Figure 3. However, the relationship between designations in Figure 3 and Qr1 and Qr2 mentioned on Page 1-2 is not clear. Please clarify.

Response: *The paragraph has been expanded to define the soil units and clarify their relationship to lithologic units Qr1 and Qr2.*

9. Page 2-1; Section 2.1; second paragraph. Change "and along" to "along".

Response: *The revised text has been modified as suggested.*

10. Page 2-1; Section 2.1; last paragraph. Change "because the" to "because they".

Response: *The revised text has been modified as suggested.*

11. Page 2-5; Section 2.4; first paragraph. Reference should be made to the October 2008 NDEP guidance on Data Usability, rather than the United States Environmental Protection Agency's (USEPA's) 1992 guidance.

Response: *NDEP's 2008 guidance builds on USEPA's 1992 guidance and both are now referenced.*

12. Pages 2-6 and 2-7; Criterion II and III. The purpose of the criterion is described, and a description is provided that various activities were performed appropriately. But, there is no practical way to verify these assertions. References to the available information are needed. Appropriate references might include the data validation summary report (DVSR), laboratory reports, field reports, etc.

Response: *Appropriate references have been added to the subject text as requested in NDEP's comment.*

13. Page 2-7; Criterion IV, last sentence. NDEP suggests that BRC delete "although unfortunate". This is not necessary in the report. (Please note that this occurs in at least two other places in the report.)

Response: *The subject text has been modified as suggested; however, it should be noted that the cited example is the only such occurrence that BRC was able to identify in the report.*

14. Page 2-8; Criterion IV, top of page. NDEP suggests that BRC reword the last two sentences along the lines of "BRC uses GiSdT to conduct non-parametric tests including the Wilcoxon Rank Sum test, the quantile test and the slippage test. The Gehan ranking system is used for

these tests to accommodate multiple detection limits within the same dataset. However, if detection limits are among the largest values in the dataset, then conclusions from the statistical test results should be treated with caution.".

Response: *The revised text has been modified as suggested.*

15. Page 2-8; Criterion V, first line. Change "primarily of" to "primarily on".

Response: *The revised text has been modified as suggested.*

16. Pages 2-8 and 2-9; Criterion V and VI. Reference is made to the DVSR, but reference should also be made to the Tables in Appendix B, since these tables show results of the data usability evaluation.

Response: *The revised text has been modified as suggested.*

17. Page 3-1; Section 3.0. The USEPA references need to updated to the more recent 2006 USEPA guidance.

Response: *The revised text has been modified as suggested to reflect the more current guidance.*

18. Page 3-1; Section 3.0, last line. The following sections do not discuss data usability. The Data Usability section is Section 2. Please revise.

Response: *The comment refers to a section of test that lists topics discussed in Section 3.0. In response to this comment, the term "data usability" has been removed from that list.*

19. Page 3-2; bullet (bottom of page). It would be helpful to list the four metals that are not included in the 2008 data, and to recognize that changes to the site-related chemicals list (SRC list) for radionuclides are the reason why only eight radionuclides are included (and perhaps list those nuclides by their radionuclide chains).

Response: *A footnote has been added to explain the differences between the two datasets in this regard, and the reasons for the changes.*

20. Page 3-3; 1st Bullet. There is a minor error in the response to specific comment 3 in Appendix A, which indicates that 104 data points from the 2005 BRC/TIMET dataset and 15 from the Environ dataset comprise the 2005 background dataset. The 15 should be changed to 16.

Response: *The tallies of sample points have been reviewed and BRC has confirmed that there are 120 total data points in the 2005 BRC/TIMET dataset; 104 from the 2005 background investigation and 16 from the Environ investigation.*

21. Page 3-3; 1st paragraph (after bullets). This paragraph is confusing. The discussion jumps from metals to radionuclides and back to metals again. Some cleanup of this issue would help. Also, it is not clear what this discussion is doing in this section. It appears that this discussion would be more appropriate in the Data Usability section under Criterion IV and/or VI. It is not clear why sample- specific Minimum Detectable Activities (MDAs) should have an effect on detection frequency. Since all radionuclide data are going to be used, it is not clear why this argument is even necessary, except, perhaps, in terms of data usability.

Response: *The original purpose of this paragraph was to discuss the effects of reporting limits on detection frequencies. Because this particular issue has been discussed in greater detail elsewhere in the report, the paragraph that is the subject of this comment has been removed from the report.*

22. Page 3-4; 1st paragraph (top of page). The Gehan ranking method should be described here.

Response: *The text has been expanded to include a discussion of the Gehan ranking method.*

23. Page 3-4, Section 3.1.4. It is not clear why the section on outliers appears before the exploratory analysis (plots) presented in Section 3.2, and summary statistics presented in Section 3.3, especially since the outlier analysis relies on some of these plots (box plots in particular). Outlier analysis is usually one of the last statistical analyses performed, not the first. In addition, the treatment of outliers is over-emphasized. There are no outliers in this dataset. This is demonstrated by the plots and correlation analysis. We recognize that outliers are defined according to the 1.5 x box height measure used to identify more extreme tail data, but this is a definition of statistical outlier, and not of an outlier per se. Outliers should not be identified based on agreement with an underlying statistical distribution, which might not reflect the underlying process anyway (parametric distributions are approximations to reality that are used to support prediction and decision making). In addition, with the number of data points involved, some values outside the 1.5 box height limits should be expected, even if the underlying process is normal. NDEP continues to be concerned about the large emphasis on outlier analysis in this report, given the potential uses of these data for background comparisons.

Response: *The text has been revised and moved to follow Section 3.3. For further details regarding outliers, the reader is referred to Appendix E.*

24. Page 3-8; Box plots, last paragraph. The reference to 6,700 records is unnecessary and not very informative. What is more informative is the number of data points per chemical and the number of chemicals. Please revise.

Response: *The reference to the number of records was included to give perspective to the term "large," which is a subjective term. The subject text has been revised to include a reference to Table 2, which present the number of data points associated with each analyte.*

25. Page 3-10; first paragraph. It is not clear why barium is being discussed here. There is no discussion of any other chemical in this section. NDEP suggests that BRC either delete this discussion from here, or use this as an opportunity to describe more conclusions from the plots and summary statistics.

Response: *The paragraph that is the subject of this comment has been deleted from the revised document.*

26. Page 3-10; Chemical sub-sections under Section 3.4. For cadmium, the median detected concentration for the 2005 BRC/TIMET shallow data set is less than the respective reportable detection limit (RDL) for non-detects. For silver, both the 2005 BRC/TIMET and 2008 Supplemental datasets have median RDLs that are greater than the median detected concentration. For selenium the median RDLs for non-detects differ by a factor of two. For thallium, the median RDL for non-detects are different for the 2005 BRC/TIMET and 2008 Supplemental datasets. All of these issues can compromise statistical analyses in this report and potential future background comparisons. Some further discussion of these issues is needed in the Data Usability section. NDEP recognizes that there are not good options, but some further recognition of the issues would clarify the limitations of the future uses of this data. There is also a discrepancy between the text in the "Assessment of RDL Effects…"section and the 2008 non-detect RDL for zirconium. The text in the assessment portion refers to a 2008 non-detect RDL of 0.3 mg/kg while the value in the table is 0.8 mg/kg. Please clarify.

Response: *The following text has been added to as the last paragraph of Section 3.1.3 "It should be noted that the method detection limit (MDL) is established by the laboratories and represents the minimum concentration of a substance that can be measured and reported with 99 percent confidence that the analyte concentration is greater than zero. MDLs are established using matrices with little or no interfering species using reagent matrices and are considered the lowest possible reporting limit. Often, the MDL is represented as the instrument detection limit. The RDL (also known as the sample quantitation limit [SQL]) is defined as the MDL adjusted to reflect sample-specific actions, such as dilution or use of smaller aliquot sizes, and takes into account sample characteristics, sample preparation, and analytical adjustments. It represents the sample-specific detection limit and all non-detected results are reported to this level. Therefore, because the RDL is a sample-specific detection limit, for the dataset as a whole there may be instances where the maximum non-detect value may be higher than the lowest detected concentration, the median RDL for a chemical in a dataset is greater than the median detected concentration, or median RDL for non-detects are different for different datasets. It is recognized that these limitations may compromise statistical analyses in this report and potential future background comparisons.*

Also, the document has been revised to repair the discrepancy between the zirconium text and table.

27. Page 3-14; Section 3.5. NDEP suggests that BRC reword the first sentence. "Findings ….were used to infer…" does not seem like a good construction. The following sentence states "Specifically, the following were conducted", however, conducting does not follow from because the previous sentence, which refers to findings. It is also not clear what was conducted, although presumably it is some form of statistical procedure. The bullets might also need to be reworded once the introductory paragraph is changed.

Response: *The subject paragraph has been revised to address NDEP's comment; no revisions to the bullets were necessary.*

28. Page 3-15; footnote 8. This footnote should be listed on the previous page.

Response: *The pagination has been adjusted such that the footnote in question now falls on the page in which it is referenced.*

29. Page 3-15, Section 3.5.1. The first sentence is incomplete. Statistical hypotheses are framed in terms of both a null and an alternative hypothesis. Both need to be specified. More description is needed in this introductory paragraph. Reference should also be made to significance testing or classical statistical methods, or the like, since the statement is not true otherwise. In addition, the description of the null hypothesis in each of the 2 cases should also be rewritten. Null hypotheses are not about datasets, they are about population parameters. For example, BRC needs to discuss if mean concentrations are statistically similar for different populations (although a different statistic is used for the Wilcoxon Rank Sum (WRS), quantile, slippage and Kruskal-Wallis (KW) tests).

Response: *The subject paragraph has been revised to address NDEP's comment.*

30. Page 3-15, Section 3.5.2. There is still a mathematical form for non-parametric tests. For example, the WRS test assumes symmetry in the respective distributions. The difference is that a parametric form of statistical distribution is not assumed.

Response: *The subject text has been revised to address NDEP's comment.*

31. Page 3-16, first paragraph. A significance level of 0.05 is indicated here. When many tests are used on the same data, a smaller significance level should be used. Note also that, on the next page, an indication is made that a significance level of 0.025 is used for the set of 2-sample tests. Some clarification is needed.

Response: *It is ERM's understanding that NDEP is referring to the use of a correction when more than one test in a particular study is applied when a single null hypothesis of no effect is* *tested. A Bonferroni correction/adjustment is one of the more basic and common procedure used to adjust the alpha level to account for random chance when using multiple tests to test a single null hypothesis. Text has been revised and a discussion of a Bonferroni correction has been included in the report as Section 3.6.2.4 to provide an added perspective to the findings of multiple tests.*

Note that the use of a Bonferroni correction would not have changed the overall conclusions of the study with regard to significant geochemical differences (i) among 0, 5, and 10 ft bgs depth intervals within the 2008 River background data (Table E-1), (ii) among the four lithologic units (Tables F-2 and F-3), and (iii) between 2008 River and 2005 McCullough by depth interval (Tables F-6 through F-8)

32. Page 3-16, t-test. Reference to large sample sizes is made. This should be accompanied to reference to the Central Limit Theorem, which is the basis for assuming the mean is normal even when the data are not normal.

Response: *The subject text has been revised to identify that parametric tests assume that both datasets are normally distributed and have equal variances.*

33. Page 3-17, Kruskal-Wallis test, 2nd sentence. Change "The Kruskal-Wallis tests" to "The Kruskal-Wallis test is used to test".

Response: *The revised text has been modified as noted in the comment.*

34. Page 3-18, Item 1. Change "conduct test" to "conduct a test".

Response: *The revised text has been modified as noted in the comment.*

35. Page 3-18, paragraph below Item 2. It is not clear what is meant by these paragraphs. It is not clear why a reference to tests involving medians is made here. Please explain why all the tests are not admissible again.

Response: *The subject text has been revised to address NDEP's comment.*

36. Page 3-18, 2nd paragraph below Item 2. The last sentence should be reworded. The intent seems to be that the tests involving full datasets unimpacted by non-detects (NDs) are more reliable. While that might be true as a general statement, it is not a helpful statement for chemicals such as thallium, or silver, or antimony, which are affected by their detection limits (DLs). This same statement appears several times in this report. NDEP suggests that it is reworded everywhere it appears. If BRC does not agree that performing this analysis is productive, then NDEP is willing to discuss the issue. The binomial proportions tests are reasonable if the DLs are approximately the same. Performing comparisons for the detected data if the frequency of detection (FOD) is the same and the DLs are about the same can

perhaps be performed through exploratory data analysis (EDA) as opposed to using statistical significance tests.

Response: *A key reason/objective for this study is to determine whether there is sufficient evidence to suggest that background lithologic units and depth intervals are different to promote/ensure proper future application of the data to different sites of interest. Conclusions of this study are based on the preponderance of the evidence for 46 constituents. Given the relatively few constituents affected by their detection limits and the associated unreliability of statistical analyses for these constituents, study objectives can be met considering the more reliable analyses for the far greater number of the 46 constituents.*

Concerns with regard to the low frequency of detects (FODs) for thallium, silver, antimony are more appropriate and will be addressed when applying background datasets to identify specific constituents at sites that are considered to be elevated above background concentrations.

Text has been revised in Section 3.7 to indicate that study conclusions related to whether differences exists is better served based on the preponderance of the evidence from the more reliable analyses associated with the majority of the 46 constituent with greater frequency of detects.

37. Page 3-18, Footnote 14. The test of proportions is not usually described as a non-parametric test. It is usually described as a binomial test, a proportions test, or as a chi-square test for independence.

Response: *The text has been modified to more accurately describe the Z-test for two proportions.*

38. Page 3-19, Pearson's section, 2nd sentence. Change "The Pearson's" to "Pearson's".

Response: *The revised text has been modified as noted in the comment.*

39. Page 3-19, Footnotes 15 and 16. These footnotes are unnecessary, since the same words are in the text.

Response: *The two footnotes referenced in NDEP's comment have been deleted from the revised text.*

40. Page 3-19, Section 3.5.3, first sentence. This sentence seems strange since the previous analysis of the 2005 data suggest that these data should be sub-setted for background comparisons because of geologic differences.

Response: *The primary conclusion from the 2007 report was that: "The statistical test of background soil sample data, based on location, suggest a number of statistically significant differences; however, because the data represent the range of background conditions at the site,* *there is no rationale for dividing the data into separate datasets based on location, soil origin, or study." Therefore, the sentence in the report is considered appropriate.*

41. Page 3-19, Section 3.5.3, third sentence. NDEP suggests that BRC delete the words "semi-quantitatively" as they are not necessary.

Response: *The revised text has been modified as noted in the comment.*

42. Page 3-20, first paragraph. This paragraph describes differences for arsenic, and then jumps into other differences that have nothing to do with concentration differences. The other differences are issues with the data set that have been described previously. If these paragraphs and bullets are to remain here, then the discussion of arsenic should be moved down in this sub-section. It would also help to include some discussion about other metals for which differences were observed.

Response: *The subject paragraph and bullets have been deleted from the revised document.*

43. Page 3-20, last line. Change "the Test of Proportion" to "a binomial proportions test".

Response: *The revised text has been modified as noted in the comment.*

44. Page 3-21; Table. Detection limits for cadmium and thallium are not similar, so it is difficult to understand why the test of proportion is applicable in these instances. Pages 3-10 and 3-12 show markedly different RDLs for the different background data sets for these metals.

Response: *Tables embedded within text and Tables E-4 and E-5 have been revised.*

45. Page 3-21, paragraph under table. Much like in subsequent sections, more specific results should be detailed here. Also, the statement in the last sentence is unnecessary (see previous comment).

Response: *The subject text has been expanded as noted in the comment.*

46. Section 3.5.3 in general. This section probably summarizes the most important results in the study. However, specific results are not provided in this section in nearly the level of detail provided in subsequent sections. The important results should be described in this section, including identifying metals and radionuclides for which differences are seen.

Response: *Text has been revised to provide specific results, including identifying metals and radionuclides for which differences were observed.*

47. Page 3-25, 2nd paragraph. Change "2008 River differs" to "2007 River data differ" (or some other similar change).

Response: *The revised text has been changed to read "2008 River data differ" in place of "2008 River differs."*

48. Page 3-25, Section 3.5.5, 1st paragraph. Reference is again made to a significance level of 0.05. Clarification is needed considering the comment above.

Response: *Please see response to Specific Comment #31.*

49. Page 3-27; last paragraph, 1st sentence. A comment from the previous round of comments was not addressed. Please change "…were be examined…" to "…were examined…".

Response: *The revised text has been modified as noted in the comment.*

50. Page 3-28; 3rd paragraph. This paragraph does not discuss correlations within the thorium chain. The issues here should be discussed in greater detail.

Response: *The revised text has been expanded as noted in NDEP's comment.*

51. Section 3.5.6 in general. The final conclusions that the correlation analysis together with the EDA suggests that these are background data is not made sufficiently clear in this section. This is the purpose of the section. The outlier analysis should also be included in this section as well, since it is also aimed at whether these data seem to represent background (although both sections could come before the comparisons between data sub-sets). And, mention of the organics results should be made in the same context.

Response: *The revised text has been expanded as noted in NDEP's comment.*

52. Page 4-1; 1st paragraph, last sentence. NDEP believes that this sentence does not fully describe the objective. The objective is to add background data from another geology (to accommodate background comparisons at the Mohawk sub-area and Parcel 4B). The statistical analyses are performed to determine if this is appropriate, or if the data do not represent background conditions, or if they do not represent a geology that is already covered in the 2005 background dataset.

Response: *The revised text has been modified to reflect NDEP's comment.*

53. Page 4-1; 3rd paragraph, 2nd sentence. Change "Several outliers" to "Several statistical outliers". Suggest instead that the focus of this paragraph be changed to one of using the organic data, the correlation analysis, the EDA and outlier analysis to confirm that these are

background data. This can be achieved by merging, and rewording as necessary, this and the next paragraph.

Response: *The subject sentence has been modified as noted in the comment, and the paragraph has been merged with the subsequent paragraph and reorganized.*

54. Page 4-1; 4th paragraph, 2nd sentence. The results of the SVOCs analysis should be described in Section 3.

Response: *See prior response regarding the inappropriateness of including the discussion in Section 3. The discussion of SVOC results has been expanded in Section 2.4 and is summarized in this paragraph.*

55. Page 4-1; 5th paragraph. The purpose of this is not clear. The datasets do not overlap for some metals (e.g., arsenic) in the way described. That is the purpose. That is, these data represent a different geology. NDEP suggests that BRC delete this paragraph and refocus on the objectives.

Response: *The paragraph that is the subject of this comment has been deleted from the revised document.*

56. Page 4-1; 5th paragraph, last sentence. Start a new paragraph here.

Response: *The text has been modified as suggested in NDEP's comment.*

57. Page 4-1; bullets. Suggest moving the 3rd bullet to the 1st.

Response: *The bullet order has been modified as suggested in NDEP's comment.*

58. Table 1. There are still a few instances in the summary statistics table where the maximum non-detect value is greater than the minimum detect value (e.g., lithium and silver). Please clarify.

Response: *See response to comment #26.*

59. Appendix E Tables. Different shading is used for some test results, presumably as a consequence of different nominal significance levels. However, it is not clear in the tables exactly how the shading is used. Please clarify.

Response: *NDEP requested that results be presented for both parametric and nonparamentric statistical tests. Grey text and shading were used to identify results for statistical tests that are* *less preferred given the distribution of the datasets.A footnote has been added to the tables to clarify this.*

2008 SUPPLEMENTAL SHALLOW SOIL BACKGROUND REPORT

BMI COMMON AREAS (EASTSIDE)
CLARK COUNTY, NEVADA

Prepared for:
Basic Remediation Company (BRC)
875 West Warm Springs Road
Henderson, Nevada 89011

Prepared by:

ERM-West, Inc.
2525 Natomas Park Drive, Suite 350
Sacramento, California 95833

MARCH 2009

I hereby certify that I am responsible for the services described in this document and for the preparation of this document. The services described in this document have been provided in a manner consistent with the current standards of the profession and to the best of my knowledge comply with all applicable federal, state and local statutes, regulations and ordinances. I hereby certify that all laboratory analytical data was generated by a laboratory certified by the NDEP for each constituent and media presented herein.

Ranajit Sahu March 16, 2009

Dr. Ranajit Sahu, C.E.M. (No. EM-1699, Exp. 10/07/2009) Date
BRC Project Manager

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES

LIST OF APPENDICES

ABBREVIATION AND ACRONYM LIST

ANOVA	analysis of variance
bgs	below ground surface
BMI	Basic Management, Inc.
BRC	Basic Remediation Company
DOE	U.S. Department of Energy
DQIs	data quality indicators
DQOs	data quality objectives
DVSR	Data Validation Summary Report
FSSOP	Field Sampling and Standard Operating Procedures
GEL	General Engineering Laboratories
GiSdT®	Guided Interactive Statistical Decision Tools
HSD	Honestly Significant Difference
LCS	laboratory control sample
LCSD	laboratory control sample duplicate
MDA	minimum detectable activity
MDL	method detection limit
mg/kg	milligrams per kilogram
MS/MSD	matrix spike/matrix spike duplicate
NBMG	Nevada Bureau of Mines and Geology
NDEP	Nevada Division of Environmental Protection
NRS	Nevada Revised Statutes
PARCC	precision, accuracy, representativeness, comparability, and completeness
pCi/g	pico Curies per gram
PID	photoionization detector
QA/QC	quality assurance/quality control
QAPP	Quality Assurance Project Plan
RDL	reporting detection limit
RPD	relative percent difference
SAP	Sampling and Analysis Plan
SQL	sample quantitation limit
SVOC	semi-volatile organic compound
SSURGO	Soil Survey Geographic
SOP	standard operating procedure
µg/kg	micrograms per kilogram
USDA	U.S. Department of Agriculture
USEPA	U.S. Environmental Protection Agency
WRS	Wilcoxon Rank Sum

1.0 INTRODUCTION

On behalf of Basic Remediation Company (BRC), ERM-West, Inc. (ERM) has prepared this Supplemental Shallow Soil Background Report applicable to the Basic Management, Inc. (BMI) Complex and Common Areas in Clark County, Nevada. The supplemental shallow soil background data were collected in accordance with the *Supplemental Background Shallow Soil Sampling and Analysis Plan* (SAP) dated March 2008, and approved by the Nevada Division of Environmental Protection (NDEP) in March 2008. The general scope of work included the collection of soil samples from background areas upgradient of the Site industrial areas and analysis of these samples for metals and radionuclides that are of interest at sites within the Complex and Common Areas. In addition, since the sample locations were adjacent to Lake Mead Parkway, surface samples were analyzed for semi-volatile organic compounds (SVOCs), as well as field screened using a photoionization detector (PID).

This revision of the report, Revision 4, incorporates (1) comments received from the NDEP, dated August 1, 2008, on Revision 0 of the report, dated July 2008; (2) comments received from the NDEP, dated September 23, 2008, on Revision 1 of the report, dated August 2008; (3) resolution of issues discussed during teleconferences between NDEP and BRC on August 5, 2008 and September 26, 2008; (4) comments received from the NDEP, dated November 13, 2008, on Revision 2 of the report, dated October 2008; and (5) comments received from the NDEP, dated February 17, 2009, on Revision 3 of the report, dated December 2008. The NDEP comments and BRC's responses to these comments are included in Appendix A. Also included in Appendix A is a redline/strikeout version of the text showing the revisions from the December 2008 version of the report. An electronic version of the entire report, as well as original format files (MS Word and MS Excel) of all text and tables are included in Appendix B.

1.1 OBJECTIVES AND PURPOSE

The purpose of this investigation was to collect and analyze data for metals and radionuclides in background shallow soils that are comparable to site soils in geologic units not covered by the existing *Background Shallow Soil Summary Report* (BRC/TIMET 2007) dataset. This supplemental background study was primarily undertaken because background comparisons for arsenic have failed at both the Mohawk and Parcel 4B sub-areas. However, there is no history of arsenic contamination at these sites; therefore, some consideration has been given to the possibility that the eastern part of the site exhibits different background levels of arsenic and, potentially, other metals. The northeastern part of the site is close to the northern part of the

River Mountains range. A mile or two to the northeast of the Mohawk area, in the vicinity of the Henderson Landfill, and still in the River Mountains range, very high concentrations of arsenic have been observed in background samples (see discussion in Section 3.4). Consequently, the reason for collecting these supplemental background samples was so that a specific subset of background conditions could be used for comparison with site concentrations, primarily at the Mohawk and Parcel 4B sub-areas.

At present, insufficient background data exist for alluvial fan materials downgradient of the northern River Mountains to evaluate whether concentrations of site-related chemicals detected in site samples in the eastern portion of the BMI Common Areas statistically exceed concentrations of these chemicals in background soil.[1] Therefore, the specific objectives proposed for the supplemental shallow soil background study included the collection of data:

- From sampled soil units that are representative of Site soils not covered by the existing background shallow soil dataset;
- That form a sufficient sample population that can be used to support statistical comparison of on-site and background datasets;
- That could be used to evaluate the comparability of soil originating from geologic units from the River Mountains; that is, comparison of the northern River Mountains (this 2008 Supplemental dataset) with the southern River Mountains and McCullough Range (2005 BRC/TIMET dataset).

This supplemental shallow soil background sampling event specifically targeted the lithologic units defined as "Pediment and fan deposits of the River Mountains" depicted as being located in the southeastern-most edge of the Common Areas in the Nevada Bureau of Mines and Geology (NBMG) *Las Vegas SE Folio Geologic Map (1977) and the Geologic Map of the Henderson Quadrangle, Nevada* (NBMG 1980) (see Figure 1, Qr_1 and Qr_2 labels).

To support this data collection effort, soils collected from the background borings were analyzed for SVOCs to evaluate potential soil impacts at the background drilling locations. The underlying

1 The existing BRC/TIMET background shallow soil dataset consists of samples collected almost exclusively from soils originating from the McCullough Range. Only background sample location BRC-BKG-12 is considered to be a mixed alluvium location. No samples during the BRC/TIMET background shallow soil investigation were collected exclusively from the alluvial fan materials downgradient of the River Mountains. Although there were several background samples collected by Environ (2003) in this geologic unit, given recent sample results at the site, the Environ data is considered inadequate for characterizing the northern part of the River Mountains.

assumption was that if potential chemical impacts were observed at a given boring location, the designation of that boring as representing background conditions would be suspect.

1.2 SITE LOCATION AND GEOLOGIC SETTING

The Site is located in Clark County, Nevada, and is situated approximately 2 miles west of the River Mountains and 1 mile north of the McCullough Range (Figure 2). For reference, it is noted that the Upper Ponds occupy the southern portion of the BMI Common Areas, and the Lower Ponds occupy the northern part of the BMI Common Areas. The McCullough Range is the primary source of materials upslope of the BMI Complex, the Lower Ponds, and the western and central portions of the Upper Ponds. Both the River Mountains and the McCullough Range are primary sources of materials upslope of the eastern portion of the Upper Ponds. According to NBMG (1980), the River Mountains and McCullough Range consist of volcanic rocks: dacite in the River Mountains and andesite in the McCullough Range. The land surface slopes in a westerly to northwesterly direction from the River Mountains and in a northerly to northeasterly direction from the McCullough Range. Near the Site, the surface topography slopes in a northerly direction towards the Las Vegas Wash.

A soils map reproduced from the U.S. Department of Agriculture (USDA) Soil Survey Geographic (SSURGO) database shows that the soil type classification for the Upper and Lower Ponds area proper is map unit 600, “slickens,” a non-native soil type (artificial fill). This term is presumed to reflect the non-native material observed in those Ponds that were used for waste disposal. The soil type classification for the BMI Complex is map unit 615, “urban land.” Native soils underlying the slickens and urban land are assumed to be consistent with the surrounding map units (*i.e.*, primarily map unit 184, and, to a lesser extent, map units 112, 117, 182, 187 and 326). As seen in the USDA soils map excerpted on Figure 3 that is based on the 1985 USDA Soils Survey (USDA 1985), the area targeted in this investigation falls within the boundaries of mapped soil unit 182 (Caliza-Pittman-Arizo complex), which is the native soil type mapped as being present in the southeastern-most portion of the Common Areas and associated with the Qr_1 and Qr_2 lithologic units.

2.0 SUMMARY OF THE INVESTIGATION

This section identifies the sampling locations, presents the sampling and analytical methods, and summarizes the results of data validation.

2.1 SAMPLING LOCATIONS

Soil samples were collected from three depth intervals at each sampling location, including surface soil (0 to 0.5 feet below ground surface [bgs]), and two subsurface depths (4 to 6 feet and 9 to 11 feet bgs). The background soil study collected data for site-related metals and radionuclides. Data for SVOCs were also collected to evaluate whether the background soil locations are impacted by other anthropogenic sources.

Soil samples were collected from 10 initial sampling locations adjacent to Lake Mead Parkway, on the south side of the roadway away from the Site. These 10 locations are shown on Figure 1, along with sampling locations for the 2005 BRC/TIMET and 2003 Environ studies on Figure 2.

The 10 sampling locations were selected because they exhibited the following characteristics:

- They are off-Site locations, in relatively close proximity to the Site; however, they are upgradient and sufficiently distant from the Site such that impacts from Site operations are not likely;
- They are upwind of the Site (wind direction plots indicate the predominant wind direction is from the south and southwest; see Figure 2) and are thus less likely to have been affected by aerial deposition of wind-borne dusts or vapors from Site operations; and
- They are upslope of the Site and are thus unlikely to have been affected by overland surface-water transport of potentially contaminated site sediments.

Available background sample locations are constrained due to rapid development in the area. Undeveloped areas in close proximity to the site, without access problems, are scarce. Although the 10 locations are adjacent to Lake Mead Parkway, as can be seen from Figure 1 they are within undisturbed areas. Therefore, the 10 sampling locations were chosen because they exhibited the characteristics identified above and are considered adequate for representing undisturbed alluvial material washed down from the northern River Mountains.

2.2 SUMMARY OF SAMPLING PROCEDURES AND ANALYSES

Soil samples were collected from a single boring at each location, drilled using a hollow-stem auger rig. Samples were collected in a split-spoon sampler lined with stainless steel sleeves. Samples collected from each boring are considered independent samples. Sampling and sample handling procedures were consistent with the standard operating procedures (SOPs) developed for the BMI Common Areas as provided in the *BRC Field Sampling and Standard Operating Procedures* (FSSOP; BRC, ERM and MWH 2008). Subsurface soil samples were collected from each two-foot interval of drill core (*i.e.*, 4 to 6 feet bgs and 9 to 11 feet bgs).

For this study, surface soil is defined as the upper 0.5 feet of the soil horizon; subsurface soil is defined as below 0.5 feet bgs. Soil samples were collected from three zones in each boring as follows:

- Surface Soil (soil samples collected from within the depth interval from 0-0.5 ft bgs; hereinafter referred to as “0 ft bgs” interval);
- Shallow Subsurface Soil (soil samples collected from within the depth interval from 4-6 ft bgs; core homogenized; hereinafter referred to as “5 ft bgs” interval); and
- Deeper Subsurface Soil (soil samples collected from within the depth interval from 9-11 ft bgs; core homogenized; hereinafter referred to as “10 ft bgs” interval).

Ten borings were advanced and three samples from each zone were collected for an initial total of 30 soil samples. Field duplicate samples were collected at three locations; from locations BRC-BKG-R01 (0 ft bgs), BRC-BKG-R05 (0 ft bgs), and BRC-BKG-R08 (5 ft bgs) for metals and SVOCs; and from locations BRC-BKG-R01 (5 ft bgs), BRC-BKG-R05 (0 ft bgs), and BRC-BKG-R08 (5 ft bgs) for radionuclides. Inadequate sample volume was collected from location BRC-BKG-R01 (0 ft bgs), the first sample collected, which is why the field duplicate at this location for radionuclides is at a different depth (5 ft bgs) than that for metals and SVOCs. Because these samples are considered field duplicates, and not split samples, each is considered an independent sample. Therefore, there were a total of 33 soil samples collected as part of this investigation. Soil boring logs representing each location are also included in Appendix C.

The soil samples were submitted for analysis to TestAmerica in St. Louis, Missouri. Analyses were conducted at three TestAmerica laboratory locations: St. Louis, Missouri; Burlington, Vermont; and West Sacramento, California. General Engineering Laboratories (GEL), located in Charleston, South Carolina, performed the radionuclide analyses.[2] At the time of analysis, all laboratories were NDEP-certified laboratories for the analyses conducted. Surface and subsurface sample analyses consisted of a full suite of metals, eight radionuclides (radium-226, radium-228, thorium-228, thorium-230, thorium-232, uranium-233/234, uranium-235/236, and uranium-238), SVOCs, and general soil characteristics. The individual analytes, analytical methods, and reporting detection limits (RDLs) are presented in Table 1. These analytes and methods are consistent with the BRC site-related chemicals list and analytical program previously established in the *BRC Quality Assurance Project Plan* (QAPP; BRC and ERM 2008a). All radionuclide analyses underwent full dissolution preparatory methods. All preparatory methods and analyses are consistent with the 2005 BRC/TIMET background dataset.

The detection frequencies for metals and radionuclides evaluated during this supplemental shallow soil background study are presented in Table 2. Detection frequencies observed for these analytes during the 2005 shallow background study are also provided in Table 2 for comparison. As seen in Table 2, most of the metals and radionuclides that are the subject of the supplemental shallow soil background investigation were detected routinely in the 2008 shallow soil samples. Exceptions are:

- Antimony
- Boron
- Chromium (VI)
- Lithium
- Mercury
- Niobium
- Platinum
- Selenium
- Silver
- Thallium
- Tin
- Tungsten
- Uranium 235/236
- Zirconium

These fourteen constituents were detected in fewer than fifty percent of the samples in which they were analyzed during the supplemental shallow soil background investigation. Most of these same compounds were also not detected routinely during the 2005 shallow soil background investigation. Exceptions to this observation consist of lithium, mercury, tin and zirconium, which were routinely detected in the 2005 samples but not in the 2008 samples. Selenium and

[2] GEL labeled all primary samples that required matrix spike/matrix spike duplicates (MS/MSD) with the sample name specified on the chain-of-custody, but included an MS/MSD identification (*e.g.*, BRC-BKG-R02-5-MS/MSD). Due to the unaccustomed labeling, all samples with the MS/MSD were inadvertently regarded as quality control samples and not included with the original sample dataset. GEL was contacted and they confirmed the results for samples labeled as MS/MSD are actual primary sample results.

thallium were also detected at a noticeably lower frequency in the 2008 supplemental shallow samples than in the 2005 samples. In contrast, cadmium, silver, and uranium-233/234 were detected at a noticeably higher frequency in the 2008 supplemental shallow background samples than in those from the 2005 shallow background investigation. It should be noted that variations in detection frequencies are influenced by the associated RDL, and may not reflect trends in actual concentrations; the effect of RDLs on detection frequencies is discussed further in Section 3.5.

2.3 DATA VALIDATION SUMMARY

All of the data were subjected to a Level 3 review. In addition to the Level 3 review, 20 percent of all data collected during the course of the investigation were subjected to full Level 4 data validation. Level 3 and 4 reviews are provided in the *Data Validation Summary Report (DVSR)—2008 Supplemental Shallow Soil Background Sampling Event* (BRC and ERM 2008b; approved by NDEP on June 9, 2008). Stable chemistry sample results (metals) for supplemental shallow soil background samples were validated in accordance with the following U.S. Environmental Protection Agency (USEPA) guidance document *U.S. EPA Contract Laboratory Program National Functional Guidelines for Inorganic Data Review* (USEPA 2004). USEPA has not standardized the validation of radionuclide data. Radionuclide results for supplemental shallow soil background samples were validated in accordance with SOP-40 (BRC, ERM and MWH 2008) and the project QAPP (BRC and ERM 2008a).

Based on data validation and review, data qualifiers were placed in the electronic supplemental shallow soil background database to classify whether the data were acceptable, acceptable with qualification, or rejected. Where applicable, an indication of result bias is presented. In addition, for every data validation qualifier, a secondary comment code was entered to indicate the reason for qualification. The DVSR (BRC and ERM 2008b) provides the definitions for the data validation qualifiers and comment codes used in the supplemental shallow soil background database. Validation qualifiers and definitions are based on those used by USEPA in the current validation guidelines (USEPA 2004) and summarized in the SOP-40 (BRC, ERM, and MWH 2008).

Results that are qualified as estimated may generally be usable for the purposes of establishing background and for comparison to Site-specific sample data. Based on the evaluation of the dataset, 100 percent of the data obtained during the field investigation are valid (that is, not rejected) and acceptable for their intended use. With 100 percent of the dataset validated as usable, the overall objective of the data collection event was met.

2.4 DATA USABILITY EVALUATION

The analytical data were reviewed for applicability and usability following procedures in the *Guidance for Data Usability in Risk Assessment (Part A)* (USEPA 1992) and *Supplemental Guidance for Assessing Data Usability for Environmental Investigations at the BMI Complex and Common Area in Henderson, Nevada* (NDEP 2008a). A quality assurance/quality control (QA/QC) review of the analytical results was conducted during the sampling events. According to both NDEP's and USEPA's Data Usability Guidance, there are six principal evaluation criteria by which data are judged for usability. The six criteria are:

- availability of information associated with site data;
- documentation;
- data sources;
- analytical methods and detection limits;
- data review; and
- data quality indicators, including precision, accuracy, representativeness, comparability, and completeness.

In addition to the six principal evaluation criteria, NDEP's Data Usability Guidance includes a step for data analysis. Items for this step are discussed in Section 3. A summary of these six criteria for determining data usability is provided below. Data usability evaluation tables are provided in Appendix B.

2.4.1 Criterion I – Availability of Information Associated with Supplemental Shallow Soil Background Data

The usability analysis of the supplemental shallow soil background data requires the availability of sufficient data for review. The required information is available from documentation associated with the data collection efforts. Data have been validated per the NDEP-approved DVSR (BRC and ERM 2008b). The following lists the information sources and the availability of such information for the data usability process:

- Background description and objectives provided in the NDEP-approved SAP (BRC 2008) and in Section 1.
- A site map with sample locations is provided on Figure 1.
- Sampling design and procedures were provided in the NDEP-approved SAP (BRC 2008) and discussed in Sections 2.1 and 2.2.
- Analytical methods and detection limits are provided in Table 1.
- A complete dataset is provided in Appendix B.
- Field conditions and physical parameter data as applicable to the background dataset are provided in the field investigation report (GES 2008) and DVSR (BRC and ERM 2008b).
- The laboratory provides a narrative with each analytical data package outlining any problems encountered in the laboratory, control limit exceedances, and rationale for any deviations from protocol. These narratives are included as part of the DVSR (BRC and ERM 2008b).
- QC results are provided by the laboratory, including blanks, replicates, and spikes. The laboratory QC results are included as part of the DVSR (BRC and ERM 2008b).
- Data flags used by the laboratory were defined adequately.
- Electronic files containing the raw data made available by the laboratory are included as part of the DVSR (BRC and ERM 2008b).

2.4.2 Criterion II – Documentation Review

The objective of the documentation review is to confirm that the analytical results provided are associated with a specific sample location and collection procedure, using available documentation. For the purposes of this data usability analysis, the chain-of-custody forms prepared in the field were reviewed and compared to the analytical data results provided by the laboratory to ensure completeness of the dataset as discussed in the DVSR (BRC and ERM 2008b). Based on the documentation review, all samples analyzed by the laboratory correspond to their respective geographic locations as discussed in Section 2 and shown on Figure 1. The samples were collected in accordance with the NDEP-approved SAP (BRC 2008) and SOPs developed for the BMI Common Areas as provided in the FSSOP (BRC, ERM and MWH 2008). Field procedures included documentation of sample times, dates and locations, and other sample-specific information (*e.g.*, sample depth). Information from field forms generated during sample collection activities was imported into the project database.

The analytical data were reported in a format that provides adequate information for evaluation, including appropriate quality control measures and acceptance criteria. Each laboratory report describes the analytical method used, provides results and detection limits on a sample-by-sample basis, and provides the results of appropriate quality control samples (*e.g.*, laboratory control spike samples, sample surrogates and internal standards [organic analyses only], and matrix spike samples). All laboratory reports provided the documentation required by USEPA's Contract Laboratory Program (USEPA 1999, 2001, 2004) which includes chain of custody records, calibration data, QC results for blanks, duplicates, and spike samples from the field and laboratory, and all supporting raw data generated during sample analysis. Reported sample analysis results were imported into the project database.

2.4.3 Criterion III –Data Sources

The review of data sources is performed to determine whether the analytical techniques used in the site characterization process are appropriate for the exposure area and medium of interest and that appropriate analytical methods were used. The data collection activities were developed to characterize a broad spectrum of background metals and radionuclides in soil. As described in the SAP, samples were collected in areas of no known impacts for the target soil lithologies. The State of Nevada is in the process of certifying the laboratories used to generate the analytical data. As such, standards of practice in these laboratories follow the quality program developed by the Nevada Revised Statutes (NRS) and are within the guidelines of the analytical methodologies established by the USEPA. Based on the review of the available information, the data sources for chemical and physical parameter measurements are adequate for use.

2.4.4 Criterion IV – Analytical Methods and Detection Limits

In addition to the appropriateness of the analytical techniques evaluated as part of Criterion III, it is necessary to evaluate whether the detection limits are low enough to allow adequate characterization of the data. At a minimum, this data usability criterion can be met through the determination that routine USEPA reference analytical methods were used in analyzing the samples. Table 1 identifies the USEPA methods that were used in conducting the laboratory analysis of soil samples. Each of the identified USEPA methods is considered the most appropriate method for the respective constituent class and each was approved by NDEP as part of the SAP (BRC 2008).

Laboratory RDLs were based on those outlined in the reference method, the SAP, and the project QAPP (BRC and ERM 2008a). In accordance with respective laboratory SOPs, the analytical processes included instrument calibration, laboratory method blanks, and other verification standards used to ensure quality control during the analyses of collected samples.

Datasets with multiple detection limits are not uncommon in analytical chemistry data. As discussed in Section 2.2, fourteen constituents were detected in fewer than fifty percent of the samples--differences in detection limits is anticipated to have the greatest effect on calculations of descriptive statistics for these constituents. With regard to future statistical analyses, datasets with different detection limits are not anticipated to severely impact proposed statistical comparisons to background. BRC uses the computer statistical software program Guided Interactive Statistical Decision Tools (GiSdT®; Neptune and Company 2007) to conduct non-parametric tests including the Wilcoxon Rank Sum (WRS) test, quantile test, and slippage test. The Gehan ranking system is used for these tests to accommodate multiple detection limits within the same dataset. However, if detection limits are among the largest values in the dataset, then conclusions from the statistical test results should be treated with caution.

2.4.5 Criterion V – Data Review

The data review portion of the data usability process focuses primarily on the quality of the analytical data received from the laboratory. However for this study, the data review also included evaluation of the SVOC data to identify any evidence of impacts that might indicate that these locations are not suitable for consideration as background. Both elements are discussed below.

Data Quality Review. Soil sample data were subject to data validation. The DVSR was prepared as a separate deliverable (BRC and ERM 2008b). The analytical data were validated according to the internal procedures using the principles of USEPA National Functional Guidelines (USEPA 1999, 2001, 2004) and were designed to ensure completeness and adequacy of the dataset. Any analytical errors and/or limitations in the data have been addressed and an explanation for data qualification provided in the respective data tables. The results of ERM's data review for these issues are presented in the DVSR and are summarized as qualifiers in the dataset provided electronically in Appendix B.

For some analytical results, quality criteria were not met and various data qualifiers were added to indicate limitations and/or bias in the data. The definitions for the data qualifiers, or data validation flags, used during validation are those defined in SOP-40 (BRC, ERM and MWH

2008) and the project QAPP (BRC and ERM 2008a). Sample results are rejected based on findings of serious deficiencies in the ability to properly collect or analyze the sample and meet QC criteria. Only rejected data are considered unusable for decision-making purposes. No samples were rejected in the supplemental shallow soil background dataset. Sample results qualified as estimated indicate an elevated uncertainty in the value. A bias flag may have been applied to indicate a direction of the bias. Estimated analytical results are included in the supplemental shallow soil background dataset.

Evaluation for Evidence of Impacts/Background Unsuitability. The surface samples at each boring location[3] were analyzed for SVOCs. As previously noted, the purpose of these analyses was to identify any evidence of impacts that might indicate that these locations are not suitable for consideration as background. As summarized in Table 3, only one SVOC was detected in the samples; bis(2-ethylhexyl)phthalate, a common laboratory contaminant, was detected at low concentrations (56 micrograms per kilogram [µg/kg] and 69 µg/kg[4]) in the two samples collected from location BRC-BKG-R01 (initial and field duplicate). The RDLs for the SVOC analyses were relatively low (*i.e.*, approximately 340 µg/kg for most compounds), and are consistent with the RDLs presented in the project QAPP (BRC and ERM 2008a). Furthermore, the data review performed for the SVOC data did not identify any issues of concern with respect to the SVOC data quality (BRC and ERM, 2008b). Therefore, the SVOC data did not provide any evidence suggesting that use of the samples for determining background conditions would not be appropriate.

2.4.6 Criterion VI – Data Quality Indicators

Data quality indicators (DQIs) are used to verify that sampling and analytical systems used in support of project activities are in control and the quality of the data generated for this project is appropriate for making decisions affecting future activities. The DQIs address the field and analytical data quality aspects as they affect uncertainties in the data collected. The DQIs include precision, accuracy, representativeness, comparability, and completeness (PARCC). The project QAPP provides the definitions and specific criteria for assessing DQIs using field and laboratory QC samples and is the basis for determining the overall quality of the dataset. Data validation activities included the evaluation of PARCC parameters, and all data not meeting the established

[3] There was one exception – the surface soil sample at location BRC-BKG-R09 was not analyzed for SVOCs.
[4] Both results were flagged as estimated (J) due to their low concentrations below the RDLs.

PARCC criteria were qualified during the validation process using the guidelines presented in the National Functional Guidelines (USEPA 1999, 2001, 2004).

Precision is a measure of the degree of agreement between replicate measurements of the same source or sample. Precision is expressed by relative percent difference (RPD) between replicate measurements. Replicate measurements can be made on the same sample or on two samples from the same source. Precision is generally assessed using a subset of the measurements made. The precision of the data was evaluated using several laboratory QA/QC procedures such as field duplicates, laboratory duplicates, laboratory control sample (LCS), laboratory control sample duplicate (LCSD), and MS/MSD results. Based on ERM's review of the results of these procedures, there do not appear to be any wide-spread data usability issues associated with precision.

Accuracy measures the level of bias that an analytical method or measurement exhibits. To measure accuracy, a standard or reference material containing a known concentration is analyzed or measured and the result is compared to the known value. Several QC parameters are used to evaluate the accuracy of reported analytical results:

- Holding times and sample temperatures;
- LCS percent recovery;
- MS/MSD percent recovery (organics);
- Spike sample recovery (inorganics)
- Surrogate spike recovery; and
- Blank sample results.

Detailed discussions of and tables with specific exceedances, with respect to precision and accuracy, are provided in the NDEP-approved DVSR (BRC and ERM 2008b) and data qualified as a result of this evaluation are presented with qualifiers in the dataset provided electronically in Appendix B.

Representativeness is the degree to which data accurately and precisely represent a characteristic of the population at a sampling point or an environmental condition (USEPA 2002). There is no standard method or formula for evaluating representativeness, which is a qualitative term. Representativeness is achieved through selection of sampling locations that are appropriate relative to the objective of the specific sampling task, and by collection of an adequate number of samples from the relevant types of locations.

Completeness is commonly expressed as a percentage of measurements that are valid and usable relative to the total number of measurements made. Analytical completeness is a measure of the number of overall accepted analytical results, including estimated values, compared to the total number of analytical results requested on samples submitted for analysis after review of the analytical data. None of the data were eliminated due to data usability concerns. The percent completeness for the dataset is 100 percent.

Comparability is a qualitative characteristic expressing the confidence with which one dataset can be compared with another. The desire for comparability is the basis for specifying the analytical methods; these methods are consistent with those used in the 2005 BRC/TIMET background dataset. The comparability goal is achieved through using standard techniques to collect and analyze representative samples and reporting analytical results in appropriate units. The ranges of sample results from both the supplemental shallow soil background dataset and the 2005 BRC/TIMET background dataset are provided electronically in Appendix B. As discussed in Section 2.4, differences in detection limits among datasets may affect data comparability for datasets comprised primarily of non-detected values. For these datasets, left-censored data can result in difficulties in differentiating whether datasets are actually different or merely an artifact of detection limits. Note that for constituents with detection limits that meet data quality objectives (DQOs), comparisons between site and background may be less important as these left-censored data are likely to indicate conditions that pose an "acceptable" risk and further analysis is not necessary.

3.0 STATISTICAL METHODS AND FINDINGS

The exploratory data analysis and statistical evaluation of data for background soils generally followed industry-standard guidance documents (USEPA 2006a,b; Navy 1999, 2002) and standards agreed upon with NDEP, including the *Guidance on the Development of Summary Statistics Tables* (NDEP 2008b). These guidance documents discuss the use of statistical plots, calculation of summary statistics (such as the arithmetic mean), treatment of non-detect data, and selection of statistical tests. The following sections discuss data preparation, statistical plots, summary statistics and statistical tests, and the types of comparisons conducted.

3.1 DATA PREPARATION

3.1.1 Spatial Independence Assumptions

There are 10 soil boring locations that were sampled for the supplemental shallow soil background dataset. The 10 soil boring locations are treated as spatially independent in this background soil study. The concentrations of each analyte at each sample location and depth is dependent on the origin of the sediment and the composition of the parent material (with the exception of anthropogenic deposition of analytes such as lead).

Naturally occurring variability is associated with the deposition of sediments, and these variations may never be fully characterized and result in unexplainable data clusters. The naturally occurring variability may be impacted by sediment transport, leaching, weathering, and other geochemical processes within the alluvium; therefore, when statistical tests are performed, it is expected that some spatial correlation may be seen, but the impact of this on the background evaluation is assumed to be negligible, and all sampling locations were therefore treated as independent in the statistical tests and calculations performed for this study. Treating the data points as independent is more conservative since the larger number of samples will result in narrower confidence intervals when comparing the background data to site data.

3.1.2 Data Filtering and Combining Rules

Results from both the 2005 BRC/TIMET (which includes the Environ dataset) and 2008 supplemental shallow soil background (this report) analytical datasets were validated. In order to prepare the datasets for statistical evaluation, results from each dataset were filtered down so that each background soil sample had one result per analyte and the two datasets were combined into one database. The following steps were taken to filter and combine the 2005 BRC/TIMET and 2008 Supplemental shallow soil background datasets into one database.

1) Filtered out all laboratory QC samples from both datasets
2) Filtered out all split sample results from both datasets; retained field duplicate results in the 2008 Supplemental shallow soil background dataset
3) Filtered out all rejected (R-qualified) data in both datasets
4) Aligned chemical names for both datasets so that names are exactly the same for each
5) Aligned units for both datasets so they are exactly the same for each
6) Filtered non-metals/non-radionuclides (*e.g.*, percent moisture) from both datasets
7) Filtered out all metals and radionuclides from the 2005 BRC/TIMET background dataset that were not included in the 2008 Supplemental shallow soil background dataset
8) Added fields to both datasets that include Dataset (2005 BRC/TIMET, 2008 Supplemental), Origin (McCullough, River, or Mixed), and Depth (0, 5, or 10)
9) Aligned field names for both datasets so they can be combined for statistical evaluation
10) Identified final subset of fields that will be required to conduct the data analyses

For direct comparison of the 2005 BRC/TIMET and 2008 Supplemental shallow soil background datasets, any chemical analyzed by one study but not the other was not considered in the comparison.

After filtering and prior to final combination of the two datasets, a comparison table was prepared. Table 2 shows the comparison of analyte lists and detection frequencies between the two datasets for metals and radionuclides.

Based on the information shown in Table 2, the following observations were made:

- The 2005 BRC/TIMET background dataset contains results for 42 metals and anions and 35 radionuclides; while the 2008 Supplemental dataset contains results for 38 metals and eight radionuclides.[5]

- The sample size for the 2005 BRC/TIMET background dataset is generally 120 results for each analyte (with a few exceptions); while the sample size for the 2008 Supplemental dataset is generally 33 results for each analyte.

- In cases where analyte results are available for both datasets, the detection frequencies were compared. As discussed in Section 2.2, detection frequencies were notably different for cadmium, lithium, mercury, selenium, silver, thallium, tin, zirconium, and uranium-233/234.

3.1.3 Treatment of Data Qualified as Non-Detections

When radionuclides were not detected at activities greater than the minimum detectable activity (MDA), the laboratory reported the measured activity. Treatment of radionuclide data qualified as non-detections followed U.S. Department of Energy (DOE) guidance (DOE 1997), which states that, for radionuclide activity data:

> "All of the actual values, including those that are negative, should be included in the statistical analysis. Practices such as assigning a zero, a detect limit value, or some in-between value to the below-detectable data point, or discarding those data points can severely bias the resulting parameter estimates and should be avoided."

Therefore, for radionuclides, the reported activities (in pico Curies per gram [pCi/g]) were used without censoring to calculate all descriptive statistics (Tables 4 through 26), prepare plots (*e.g.*, boxplots), and conduct statistical analyses presented in this report.

For metals, a value of one-half the RDL was used as a replacement value for non-detected data for t-tests, parametric and nonparametric analysis of variance (ANOVA, Kruskal-Wallis tests), and calculation of parametric and nonparametric correlation coefficients. The ½-RDL substitution method was not applied to data analyzed using the WRS test because this test (as

[5] The following five inorganic constituents were included in the 2005 background investigation but were not included in the 2008 investigation: chloride, fluoride, nitrate, nitrite, and sulfate. Phosphorus was included in the 2008 investigation, but was not included in the 2005 analyte list. With NDEP concurrence, the project list of analytes was reduced in 2007 from 35 radionuclides to the following eight: uranium-238, uranium-233/234, thorium-230, and radium-226 (Uranium-238 Decay Chain), thorium-232, radium-228, and thorium-228 (Thorium-232 Decay Chain) and uranium-235/236 (Uranium-235 Decay Chain).

currently supported by GiSdT®) handles non-detected values using the Gehan ranking system (the Gehan test uses a modified ranking of sample results to accommodate non-detected values together with detected values), a method considered to be more robust than the ½-RDL substitution method. The GiSdT®'s WRS test uses the Mantel (1981) approach, which is equivalent to using the Gehan ranking system. The summary statistics (Tables 4 through 26) and plots (boxplots, individual value plots, and probability plots in Appendix D) incorporate the full RDL for non-detects.

It should be noted that the method detection limit (MDL) is established by the laboratories and represents the minimum concentration of a substance that can be measured and reported with 99 percent confidence that the analyte concentration is greater than zero. MDLs are established using matrices with little or no interfering species using reagent matrices and are considered the lowest possible reporting limit. Often, the MDL is represented as the instrument detection limit. The RDL (also known as the sample quantitation limit [SQL]) is defined as the MDL adjusted to reflect sample-specific actions, such as dilution or use of smaller aliquot sizes, and takes into account sample characteristics, sample preparation, and analytical adjustments. It represents the sample-specific detection limit and all non-detected results are reported to this level. Therefore, because the RDL is a sample-specific detection limit, for the dataset as a whole there may be instances where the maximum non-detect value may be higher than the lowest detected concentration, the median RDL for a chemical in a dataset is greater than the median detected concentration, or median RDL for non-detects are different for different datasets. It is recognized that these limitations may compromise statistical analyses in this report and potential future background comparisons.

3.2 STATISTICAL PLOTS

Statistical plots are used in exploratory data analysis to show characteristics and relationships of the data, to evaluate fit to a normal distribution, to identify anomalous data points or outliers, and to provide a general overview of the data. Probability plots, boxplots, and individual value plots were constructed as part of the data evaluation for this investigation. Preliminary evaluation of the data included an assessment of data characteristics through graphical and quantitative analysis. The 2008 Supplemental data were summarized overall and by depth interval, with data plotted for the various groupings. The 2008 Supplemental data were compared with the 2005 BRC/TIMET background data using the probability plots, boxplots, and individual value plots. The graphical analysis of the analytical data is described in the following sections, and Appendix D contains the statistical plots.

Probability Plots. The distribution plots for each chemical include a probability plot that shows how well the dataset for the chemical fits a normal or lognormal distribution. Probability plots are also useful to visually identify outliers and to evaluate the possible presence of multiple populations within a dataset. Potential multiple populations are identified by inflection points on the probability plot. Inflection points are not defined statistically, and should be used with considerable caution.

The probability plots are graphs of values, ordered from lowest to highest and plotted against a standard normal or lognormal distribution function. The vertical axis is scaled in units of concentration (or activity, in the case of radionuclides), and the horizontal axis is scaled in units of the normal/lognormal distribution function. The vertical scale is plotted as a linear scale (concentration versus normal/lognormal quantile) and populations of data that plot as a straight line in a linear scale are referred to as normally distributed (or lognormally distributed).

Boxplots. Boxplots provide a method for comparing data groupings or datasets side by side. The boxplots simultaneously display the full range of data, as well as key summary statistics, such as the median, 25th and 75th percentiles, and minimum and maximum values. The top and bottom of the box are the 75th and 25th percentiles, respectively, of the dataset. The length from the top to the bottom of the box is the interquartile range; therefore, the box represents the middle 50 percent of the data. The width of the box is arbitrary. The horizontal line within the box depicts the median value (the 50th percentile) of the dataset. The upper and lower whiskers are defined as follows:

Upper whisker = 75^{th} percentile + (1.5 • interquartile range)

Lower whisker = 25^{th} percentile – (1.5 • interquartile range)

These plots show the symmetry of the dataset, the range of data, and a measure of central tendency (median).

The boxplots, which group data for each dataset, by chemical, and by depth interval, are provided along with the probability and individual value plots for each analyte in Appendix D for the 2008 Supplemental dataset and the 2005 BRC/TIMET background dataset (including Environ dataset).

Probability and boxplots were used for identifying anomalous data points (outliers) and data clusters in the 2008 Supplemental and 2005 BRC/TIMET datasets. All anomalous data points and clusters were investigated further.

The plots shown in Appendix D summarize a large amount of data. The number of data points associated with each analyte is presented in Table 2. The plots are presented to provide a comprehensive overview of the 2008 Supplemental and 2005 BRC/TIMET background datasets for soils, to compare the 2008 Supplemental background dataset to the 2005 BRC/TIMET background dataset, and to compare the different depth intervals.

Scatterplots. A scatterplot uses a Cartesian coordinate system to display values for two variables from a dataset (*e.g.*, arsenic *vs.* aluminum concentrations for the 2008 dataset). The data are displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis.

Scatterplots were constructed for those constituent pairs with significant correlation coefficients. Scatterplots were visually examined and best professional judgment was used to ascertain whether high-concentration outliers[6] occur "near" the least-square linear trend line. Where high-concentration outliers occur "near" the trend line, one may infer that these concentrations are consistent with background concentrations.

3.3 DESCRIPTIVE SUMMARY STATISTICS

Descriptive summary statistics for metals and radionuclides were calculated for the 2008 Supplemental and 2005 BRC/TIMET datasets (Tables 4 through 26). Descriptive summary statistics for each of the two datasets were also prepared for the following depth intervals, structured around the sampling intervals employed for the 2005 shallow soil background sampling event and the 2008 supplemental shallow soil sampling event (Section 2.2):

- Surface soils (0 ft bgs);
- Shallow subsurface soils (5 ft bgs);
- Deeper subsurface soils (10 ft bgs);
- Subsurface combined (5-10 ft bgs); and
- All depths combined (0-10 ft bgs).

6 High concentration outliers were identified from boxplots (see Section 3.4).

The descriptive summary statistics calculated for each analyte include the sample size, frequency of detections, and, for both censored and detected data, the minimum and maximum concentration, the median, the mean, and the 25th and 75th percentiles (quantiles).

3.4 IDENTIFICATION AND TREATMENT OF OUTLIERS

Statistical outliers are data points that are extremely large or small relative to the rest of the data, and may not, therefore, be representative of the population sampled (USEPA 2000a). Statistical outliers may be identified using statistical methods (*e.g.*, boxplots, probability plots, associations)—however, statistical methods alone should not be the basis for removing these data from the background dataset. Background soil samples were collected in known/suspected unimpacted areas. Accordingly, once statistical outliers are identified using statistical methods, only a weight of evidence based on sound geochemical and other regional-specific knowledge should be used to identify these data as "true" outliers and justify removing them from the background dataset.

For this investigation, boxplots, individual value plots, and probability plots were used to identify statistical outliers for further investigation. Outliers were further evaluated using correlation analyses and examination of scatterplots to further assess whether associations among these relatively few outlier data points were consistent with background concentrations (see Section 3.7.4). If the statistical outlier could not be confirmed to be a transcription or other verifiable error, all statistical plots and tests were performed with the statistical outlier included in the dataset.

As shown on the boxplots[7] in Appendix D, several statistical outliers were found in the dataset,[8] which is not unusual for a dataset of this size. Several of the outliers are artifacts of the RDLs. For example, for constituents with few detections, those detections are often classified as outliers on the boxplots because they are outside the typical range of detection limits. In addition,

---

7 Statistical outliers within the 2008 dataset were defined as those points corresponding to detected metal concentrations or radionuclide activities (*i.e.*, ignoring non-detection report limit artifacts) that were greater than 1.5 times the interquartile range for the (i) combined depth plots and (ii) individual depth plots, and are shown as an asterisk (*) on the boxplots (see Section 3.2).

8 For several constituents (*e.g.*, beryllium), boxplots of the 2008 data identified outliers for the combined dataset (all depths combined), but outliers were not identified in the boxplots for individual depth intervals. In addition, in some cases (*e.g.*, calcium, 5 and 10 ft datasets), a given point that was considered an outlier for a given depth interval was not considered an outlier for the combined 2008 dataset (all depths combined) for that constituent. In these cases, the specific outlier was not considered anomalously high, and the representativeness of those values of background conditions was not questioned further.

elevated RDLs are also classified as outliers in some cases. The probability plots for the constituents identified in Section 2.2 as “not being routinely detected” demonstrate the effect of the RDLs being substituted for non-detected values in the dataset; for those constituents (*i.e.*, antimony, boron, chromium (VI), lithium, mercury, niobium, platinum, selenium, silver, thallium, tin, tungsten, uranium-235/236, and zirconium), two distinct non-linear groupings of data are clearly visible in the probability plots. Other outliers occur sporadically; these outliers were reviewed to confirm that they were not the result of reporting errors;[9] no such errors were identified.

Overall, statistical outliers represent only a small proportion of the entire dataset. In addition, the lack of a consistent pattern related to statistical outliers would suggest that the data are not indicative of naturally occurring background conditions. Finally, the sample design for collection of the supplemental soil background data intentionally focused on suspected unimpacted areas. Given the lack of scientifically defensible reasons to consider these statistical outliers to be incongruous with background conditions (*i.e.*, “true” outliers), these data were considered representative of background and retained in the supplementary background soil dataset (see also Appendix E).

3.5 FREQUENCY OF DETECTION

As noted in Section 2.2, cadmium, silver, and uranium-233/234 were detected at noticeably higher frequencies in the 2008 supplemental shallow background samples than in those from the 2005 shallow background samples, and lithium, mercury, selenium, thallium, tin and zirconium were detected at noticeably lower frequencies in the 2008 deep samples than in the shallow background studies. The statistical summaries in Tables 4 through 26 were evaluated to assess the likely influence of RDLs on these observed detection frequencies. This evaluation determined that variations in RDLs are likely to have had effects on detection frequencies for certain constituents (*i.e.*, cadmium selenium, and silver), as summarized below.

---

9 Reporting or transcription errors are unlikely given the direct electronic data uploads from the laboratory, which were in turn uploaded directly into the spreadsheets used for statistical analysis, with no manual entry of concentration values.

Cadmium	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection[10]	63.6%	13.3%
Median RDLs for Non-Detects (milligrams per kilogram [mg/kg])	0.04	0.1291
Median Detected Concentration (mg/kg)	0.11	0.105
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2005 cadmium FOD is appreciably lower than that for the 2008 data. The detected concentrations are comparable between the two datasets. The range of the 2008 detected values (0.053 to 0.26 mg/kg) is higher than the non-detect RDLs for that event (0.04 mg/kg); however, a large percentage of these data would not have been reported as detections under the higher 2005 RDLs (*i.e.*, the median value of 2008 detections was 0.11 mg/kg– less than the 2005 median RDL for non-detections [0.1291 mg/kg]). It therefore appears likely that the higher RDLs of the 2005 dataset are one cause of the lower frequency of detection in that dataset, although lower cadmium concentrations in the 2005 samples could be another explanation.	

Lithium	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	18.2%	100%
Median RDLs for Non-Detects (mg/kg)	7.314	--
Median Detected Concentration (mg/kg)	32.95	12.75
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2008 lithium FOD is appreciably lower than that for the 2005 data. The range of 2005 detections (7.5 to 26.5 mg/kg) is higher than a large percentage of the 2008 non-detect RDLs, based on the 7.314 mg/kg median 2008 RDL value, and many would have been reported as detections if present at those levels in the 2008 samples. This suggests that the 2008 samples may have generally lower lithium concentrations than the 2005 samples, despite the higher 2008 median detected concentration. However, the elevated 2008 RDLs (*i.e.*, 75th percentile of 14.628 mg/kg and beyond, which are higher than the majority of the 2005 detections [median detect 12.75 mg/kg]), complicate the analysis.	

[10] For all summary tables in this section, the value for Percent Detection reflects the full dataset for each event, as taken from Table 2, and the values provided for the other parameters were taken from Tables 4 and 9.

Mercury	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	0%	77.5%
Median RDLs for Non-Detects (mg/kg)	0.00668	0.0072
Median Detected Concentration (mg/kg)	- -	0.019
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2008 mercury FOD is appreciably lower than that of the 2005 data; the non-detect RDLs of the two events are fairly comparable. The range of 2005 detections (0.0084 to 0.11 mg/kg) is higher than the 2008 non-detect RDLs (0.00668 mg/kg), and would have been reported as detections if present at those levels in the 2008 samples. This suggests that the 2008 samples have generally lower mercury concentrations than the 2005 samples. Differences in RDLs do not appear to have caused the differences in the FODs in this case.	

Selenium	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	0%	43.3%
Median RDLs for Non-Detects (mg/kg)	0.32	0.1579
Median Detected Concentration (mg/kg)	- -	0.29
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2008 FOD for selenium is appreciably lower than for the 2005 data; the RDLs for the 2008 non-detects are about twice as high as those for the 2005 samples. A large percentage of the 2005 data detections (more than 50% based on median detect value 0.29 mg/kg), would not have been reported as detections under the higher 2008 RDLs (0.32 mg/kg). Therefore, it appears likely that the higher RDLs of the 2008 dataset are one cause of the lower frequency of detection in that dataset, although lower selenium concentrations in the 2008 samples could be another explanation.	

Silver	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	42.4%	13.3%
Median RDLs for Non-Detects (mg/kg)	0.11	0.2609
Median Detected Concentration (mg/kg)	0.076	0.0445
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2005 silver FOD is appreciably lower than that for the 2008 data; RDLs for the 2005 non-detects are more than twice as high as those for the 2008 samples. The range of 2008 detections (0.054 to 0.17 mg/kg) is lower than the 2005 non-detect RDLs (0.2609 mg/kg), and would not have been reported as detections if present at those levels in the 2005 samples. Therefore, it appears likely that	

the higher RDLs of the 2005 dataset are one cause of the lower FOD in that dataset, although lower silver concentrations in the 2005 samples could be another explanation.

Thallium	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	18.2%	35%
Median RDLs for Non-Detects (mg/kg)	0.3	0.5428
Median Detected Concentration (mg/kg)	0.46	1.1
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2008 thallium FOD is about 17% less than that for the 2005 data, RDLs for the 2008 non-detects are slightly lower than those for the 2005 samples. The majority of 2005 detections (1.1 mg/kg median value) are higher than the 2008 non-detect RDLs (0.3 mg/kg), and would have been reported as detections if present at those levels in the 2008 samples. This suggests that the 2008 samples have generally lower mercury concentrations than the 2005 samples. Differences in RDLs do not appear to have caused the differences in FODs in this case.	

Tin	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	48.5%	99%
Median RDLs for Non-Detects (mg/kg)	0.3	0.187
Median Detected Concentration (mg/kg)	0.43	0.49
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2008 tin FOD is appreciably less than that for the 2005 data; the non-detect RDLs for the 2008 data are nearly twice as high as those for the 2005 data. The majority of 2005 detections (0.4 mg/kg 1st quartile value) are higher than the 2008 non-detect RDLs (0.3 mg/kg), and would have been reported as detections if present at those levels in the 2008 samples. This suggests that the 2008 samples have generally lower tin concentrations than the 2005 samples. Differences in RDLs do not appear to have caused the differences in FODs in this case.	

Uranium-233/234	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	100%	50.8%
Median MDA for Non-Detects (pCi/g)	Not determined, because all results, including those lower than the MDA, were used in statistical analyses	
Median Detected Activity (pCi/g)	1.17	0.99

Assessment of MDA Effects on Frequency of Detection (FOD)	The 2005 shallow soil frequency of detection for uranium 233/234 is appreciably less than the frequency of detection of the 2008 data. The detected concentrations are comparable between the two datasets. Reported uranium 233/234 detections in both datasets are higher than the 2005 RDLs associated with non-detections. The assessment of RDL effects on the frequency of detection was not completely conclusive, but based on the above, it does not appear likely that the RDLs are contributing appreciably to the frequency of detection differences.

Zirconium	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	39.4%	100%
Mean RDLs for Non-Detects (mg/kg)	0.8	- -
Mean Detected Concentration (mg/kg)	11.5	125
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2008 zirconium FOD is less than that of the 2005 data. The range of 2005 detections (60.1 to 179 mg/kg) is higher than the 2008 non-detect RDLs (0.8 mg/kg), and would have been reported as detections if present at those levels in the 2008 samples. This suggests that the 2008 samples have generally lower tin concentrations than the 2005 samples. Differences in RDLs do not appear to have caused the differences in FODs in this case.	

Datasets with high frequency of detects tend to be better suited to statistical analyses than those with low frequency of detects (*i.e.*, less than 50 percent), because detection limits in the latter tend to drive the analyses. The majority of the elements in this study have comparable frequency of detects near 100 percent, and statistical analyses were performed without concern for the effect of non-detections on the findings. For the other elements with far less than 100 percent frequency of detects, the frequency of detects tended to be comparably low in the two datasets; as discussed in the following section, statistical analyses considering the effects of non-detections were developed for these elements or were omitted altogether if the number of detections was too low. The eight metals discussed above represent the few cases in which frequency of detects were appreciably different between the two datasets; these are of particular concern in this study because this situation complicates statistical comparisons. As discussed above, BRC's evaluation of the associated RDLs and ranges of detected concentrations found that differences in RDLs did not appear to have caused the differences in frequency of detects, with the possible exception of cadmium, selenium, and silver, for which the evaluations were inconclusive. For these three metals, statistical comparisons may not be reliable between the two datasets, or in the future, between the background datasets and BMI Common Areas site data.

3.6 STATISTICAL METHODS

Statistical evaluations were used to infer whether metal concentrations and radionuclide activity in 2008 supplemental background soils were comparable to those in the 2005 BRC/TIMET background soils. The following procedures were conducted as part of the statistical evaluations:

- Data were organized by lithologic unit, constituent, and soil interval;
- Data were viewed using boxplots and scatterplots (Section 3.2);
- Data were characterized using descriptive statistics and tests of normality (Section 3.3 and 3.6);
- 2008 supplemental background data were compared to 2005 BRC/TIMET background data using two- and multiple independent sample tests (Sections 3.7.1 and 3.7.2);[11,12]
- 2008 supplement background data were tested to identify potential differences among 0 ft bgs, 5 ft bgs, and 10 ft bgs depth intervals using multiple independent sample tests (Sections 3.7.3); and
- Inter-element associations were identified using correlation analyses and used to further verify that samples were appropriate for characterizing background conditions (Section 3.7.4).

3.6.1 Hypothesis Testing

A common application of statistics is to test some scientific hypothesis. A statistical test examines a set of sample data and, based on the underlying distribution of the data, leads to a decision whether to (i) accept the hypothesis or (ii) reject the hypothesis and accept an alternative one. Accordingly, statistical hypotheses are framed in terms of a null hypothesis (H_o) and an alternative hypothesis (H_a).

---

[11] 2008 River dataset was compared to the 2005 McCullough, 2005 River, and 2005 Mixed datasets for the following soil intervals: (i) 0 ft bgs, (ii) 5 ft bgs, (iii) 10 ft bgs, (iv) 5-10 ft bgs combined, and (v) 0-10 ft bgs (0, 5, and 10 ft bgs depths combined).

[12] Tests of proportions and comparisons of detected-only data were used when two- and multiple independent sample tests were not recommended—*i.e.*, when sample sizes were greater than four samples and frequency of detections were less than 50 percent.

When comparing the mean or median background concentrations for a constituent, the null hypothesis was that the mean/median background concentration for a specific constituent are comparable (*i.e.*, data populations/datasets are the same); therefore, the rejection of the null hypotheses results in the acceptance of the alternative hypothesis that the means/medians of the data populations/datasets are different.

When comparing the right-tails of two distributions, the null hypothesis was that larger values for background concentrations for a specific constituent are comparable; therefore, the rejection of the null hypotheses results in the acceptance of the alternative hypothesis that the two data populations/datasets are different with regard to larger values (*i.e.*, the values in the right-tail of one distribution are generally larger than the values in the right-tail of the other distribution).

When examining the relationship between the concentration of two constituents, the null hypothesis was that there is no correlation between two constituents (*i.e.*, no inter-element correlation); therefore, should this null hypothesis be rejected, one would accept the alternative hypothesis and infer that there exists a relationship (positive or negative) in concentrations between the two constituents. These hypotheses are also discussed in BRC/TIMET (2007) report.

3.6.2 Statistical Tests

Statistical tests were conducted to infer whether datasets are comparable and whether there exist relationships between two constituents. A key decision is whether a parametric or nonparametric statistical test is to be used. Parametric statistical tests used in this evaluation of supplement background concentrations assume the following:

- Samples are independent and drawn randomly from the population.
- Data are normally distributed for each population.

Nonparametric methods/tests are not dependent on a specific distribution (*e.g.*, normal distribution) for its validity (Singh and Singh 2007; Gilbert 1987; Sokal and Rohlf 1981; Zar 1984).[13] These methods do not require estimates of the population variance or mean.

[13] Accordingly, nonparametric tests are also known as distribution-free tests.

Nonparametric statistical tests assume that samples are independent and drawn randomly from the population.

Methods used to evaluate and compare the data groups for this supplemental background dataset are summarized below. The computer statistical software program GiSdT® (Neptune and Company 2007) was used to perform two-sample statistical comparisons. All parametric and nonparametric multiple independent sample comparisons and correlation analyses were performed using SPSS v. 15.[14] Consistent with previous studies of background concentrations at BRC, a level of significance (α) equal to 0.05 was used (BRC TIMET 2007).[15]

3.6.2.1 Two-Sample Tests

Statistical comparisons between the 2008 Supplemental dataset and the 2005 BRC/TIMET background dataset for each depth interval were performed using the Quantile test, Slippage test, the *t*-test, and the WRS test with Gehan modification. The Quantile test, Slippage test, and WRS test are non-parametric. That is, the tests are distribution free, thus an assumption of whether the data are normally or lognormally distributed is not necessary.

***t*-Test.** The *t*-test is a hypothesis test for two population means to determine whether they are significantly different. To conduct a two-sample *t*-test, the two populations must be independent; in other words, the observations from the first population must not have any bearing on the observations from the second population. Assumptions of the *t*-test are that both datasets are comprised of randomly sampled data, data are normally distributed, and datasets have equal variances[16] (Sokal and Rohlf 1981; Gilbert 1987; Zar 1984).

Wilcoxon Rank Sum (WRS). The WRS test performs a test for a difference between the sum of the ranks for two populations. This is a nonparametric method for assessing differences in the centers of the distributions that relies on the relative rankings of data values. Knowledge of the precise form of the population distributions is not necessary. The two underlying distributions are assumed to have approximately the same shape The WRS test has less power than the two-sample *t*-test when the data are normally distributed, but the assumptions are not as restrictive.

---

[14] Note a Gehan ranking is not supported by SPSS v.15 and was not used to accommodate non-detects in the Kruskal-Wallis and Kendall tau analyses.

[15] Where appropriate, a confidence level (1-α) of 95 percent confidence was used.

[16] Student *t*-test is used when datasets have equal variances. Welch's or Satterthwaite *t*-test may be applied when datasets have unequal variances.

The GiSdT® version of the WRS test uses the Mantel approach which is equivalent to using the Gehan ranking system.

Quantile Test. The Quantile test performs a test for a shift to the right in the right-tail of the site or tested population versus the reference population. This may be regarded as being equivalent to detecting if the values in the right-tail of the tested distribution are generally larger than the values in the right-tail of the reference distribution. This test assumes that the populations have approximately the same shape. The Quantile test is performed using a defined quantile = 0.80.

Slippage Test. The Slippage test looks for a shift to the right in the extreme right-tail of one population versus the extreme right-tail of a reference population. This is equivalent to asking if a set of the largest values of the tested distribution are significantly larger (in a statistical sense) than the maximum value of the reference distribution.

3.6.2.2 Multiple Independent Sample Tests

One-Way Analysis of Variance (ANOVA). The parametric one-way ANOVA tests the hypothesis that multiple (k) population means are equal (Sokal and Rohlf 1981; Gilbert 1987; Zar 1984). Where one-way ANOVAs indicated the existence of significant differences among soil strata, the Tukey Honestly Significant Difference (HSD) test was used to conduct pair-wise *post-hoc* comparisons.[17]

Kruskal-Wallis Test. Kruskal-Wallis test is a non-parametric one-way ANOVA for ranks and is used to test the equality of medians among multiple (k) populations. The Kruskal-Wallis test is used to test the null hypothesis that several populations have the same continuous distribution. If the null hypothesis is rejected, one may infer that measurements tend to be higher in one or more of the populations. Fundamentally, this test is analogous to a parametric one-way ANOVA with the exception that the measured/observed values are replaced by their ranks. Accordingly, it is an extension of the Wilcoxon-Mann-Whitney test for three or more groups. Where Kruskal-Wallis tests indicated the existence of significant differences among soil strata, examinations of boxplots were used to conduct pair-wise *post-hoc* comparisons.[18]

Examination of Constituents with Less than 50 Percent Frequency of Detection. When frequency of detections is less than 50 percent, even the nonparametric tests have little power to

[17] Note that only *post-hoc* (= *a posteriori*) comparisons were conducted.
[18] SPSS v. 15 does not support the nonparametric Behrens-Fisher *post-hoc* comparison test.

detect differences in central values (Smeti *et al*. 2007). For those constituents where the frequency of detection was less than 50 percent, two- or multiple independent sample tests were not conducted. The following approach was conducted:

1. For individual constituent datasets in which RDLs are comparable, a Z-test for two proportions[19] was conducted to identify similarities in datasets based on the proportion of detected concentrations.

2. For individual constituent datasets in which RDLs are comparable and RDLs are higher than detections, where the proportion of detected concentrations was found to be similar and the number of detected concentrations was greater than four for both datasets, two- or multiple independent sample tests were conducted on detected data only.

Note that for constituents with frequency of detections less than 50 percent and RDLs meeting analytical DQOs, one may conclude that these constituents are present at low concentrations in background soils.

3.6.2.3 Correlation Analysis

Correlations or "measures of association" are of interest because they offer another line of evidence to distinguish background and non-background data or multiple populations of data (BRC/TIMET 2007). Inter-element correlation analyses were conducted to identify those constituents that needed further examination (using scatterplots) to ensure that high concentration outliers were congruous with background concentrations.

Pearson's Product-Moment Correlation Coefficient. The Pearson product-moment correlation coefficient (r) is a parametric measure of the correlation between two variables (Sokal and Rohlf 1981; Gilbert 1987; Zar 1984). Pearson's correlation reflects the degree of linear relationship between two variables and ranges from +1 to -1. A correlation of +1 means that there is a perfect positive linear relationship between variables. A correlation of -1 means that there is a perfect negative linear relationship between variables. A correlation of 0 means there is no linear relationship between the two variables.

[19] In this investigation, the Z-test for two proportions (http://www.dimensionresearch.com/resources/calculators/ztest.html) was used to test the null hypothesis that the proportion of detected concentrations is the same among two datasets. If the null hypothesis is rejected, one may infer that the two populations are different with respect to the proportion of detected data.

Kendall Tau Correlation Coefficient. The Kendall tau rank correlation coefficient (or Kendall tau coefficient) is a non-parametric statistic used to measure the degree of correspondence between the ranks of two populations—it measures the strength of association of cross tabulations. As with the Pearson's correlation coefficient, Kendall tau ranges from +1 to -1. A value of +1 means that there is 100 percent positive association between the two variables—*i.e.*, rankings for both variables are identical. A value of -1 means that there is 100 percent negative association between the two variables—*i.e.*, the ranking of one variable is the reverse of the other variable. A value of zero indicates the absence of an association between the two variables—*i.e.*, rankings are independent.

3.6.2.4 Correction for Use of Multiple Tests

A Bonferroni correction concerns the question if, in the case of more than one test in a particular study, the level of significance (α) should be adjusted to account for random chance. As related to the supplemental shallow soil background investigation, the Bonferroni correction may be applied when a single hypothesis of no effect is tested using more than one test (*i.e.*, multiple tests for multiple constituents), and the hypothesis is rejected if one of the tests shows statistical significance. This adjustment is intended to correct for the probability of making a Type I error (*i.e.*, incorrectly concluding there exists a difference when, in fact, there is no difference among datasets) when multiple tests are used. Note that this adjustment for reducing the chance of making a Type I error will increase the probability of a Type II error—*i.e.*, incorrectly concluding there is no difference when, in fact, there is a difference datasets.

The Bonferroni correction is performed by dividing the level of significance (usually set to 0.05 by convention) by the number of tests performed. For the supplemental shallow soil background investigation study, 46 constituents were tested to determine if lithologic units and/or depth intervals are different. Accordingly, the Bonferroni correction would divide an alpha of 0.05 by 46, resulting in an alpha of 0.0011.

When comparing among background datasets, both types of error are relevant and are of interest. For the purposes of this study and to be consistent with previous studies of background concentrations at BRC (BRC TIMET 2007), a level of significance equal to 0.05 was used. The potential effects of a Bonferroni correction on the overall conclusions of the study are also discussed in appropriate sections of this report[20] to address potential consequences of making a Type I error to the overall conclusions.

3.7 RESULTS OF STATISTICAL ANALYSES

A key objective of this investigation is to evaluate whether the supplemental shallow soil background dataset is statistically similar to or different to the 2005 BRC/TIMET background data. The results of the following statistical analyses are provided with the intention of supporting a weight-of-evidence evaluation as part of this investigation.

3.7.1 Comparison of 2008 Supplemental and 2005 BRC/TIMET Datasets (All Depths Combined)

The 2008 Supplemental and 2005 BRC/TIMET datasets were evaluated to determine if they may be combined into one dataset for future consideration. The results of the statistical analyses are included in Appendix F. Probability plots, boxplots, and individual value plots were used to compare the 2008 Supplemental and 2005 BRC/TIMET data. These plots are included in Appendix D. Overall, the samples for the 2005 BRC/TIMET background study appear to have captured a fair range of natural variability and heterogeneity (largely a consequence of the larger sample size); typically showing a wider range of concentrations/activities than samples from the 2008 Supplemental shallow soil background study. Because the 2005 BRC/TIMET background data spanned a broader geographic area and included 120 samples compared with 33 samples collected for the 2008 Supplemental shallow soil background study, this is not an unexpected outcome.

The 2008 dataset was compared to each of following lithologic units: 2005 McCullough, 2005 River, and 2005 Mixed datasets (Table F-2 of Appendix F). Consistent with the Shallow Background Study (BRC/TIMET 2007), if a given dataset had fewer than four detections, it was deemed to lack data sufficient to support a robust statistical analysis and was not included in the statistical comparisons. If no more than two datasets had greater than four detections, no statistical comparisons were performed for that constituent. Accordingly, statistical tests were not performed for chromium VI, niobium, platinum and tungsten—and it was not possible to

[20] A review of tables in Appendix F indicate that the use of this correction would not have changed the overall conclusions of this study with regard to significant geochemical differences (i) among 0, 5, and 10 ft bgs depth intervals within the 2008 River background data (Table F-1), (ii) among the four lithologic units (Tables F-2 and F-3), and (iii) between 2008 River and 2005 McCullough by depth interval (Tables F-6 through F-8).

determine whether significant differences were associated with the 2008 River and the three 2005 soil lithology datasets for these metals.

Overall, statistical comparisons indicated that a number of significant differences existed for 34 of 46 constituents among the four lithologic units: 2005 McCullough, 2005 River, 2005 Mixed, and 2008 River (Table F-2 of Appendix F):

- Antimony
- Arsenic
- Barium
- Beryllium
- Boron
- Cobalt
- Copper
- Iron
- Lead
- Lithium
- Magnesium
- Mercury
- Molybdenum
- Nickel
- Palladium
- Phosphorus
- Potassium
- Silicon
- Silver
- Sodium
- Strontium
- Thallium
- Tin
- Titanium
- Uranium
- Vanadium
- Zirconium
- Radium 226
- Radium 228
- Thorium 228
- Thorium 230
- Thorium 232
- Uranium 233/234
- Uranium 238

The greatest number of significant differences was noted between 2005 McCullough and 2005 River datasets.

Differences between the 2008 River dataset and one of the 2005 datasets were identified for 14 constituents (Table F-2 of Appendix F):

- Arsenic
- Barium
- Boron
- Lithium
- Palladium
- Potassium
- Silicon
- Sodium
- Zirconium
- Radium 228
- Thorium 230
- Uranium 233/234

- Magnesium
- Strontium

With respect to the 2008 River dataset, a greater number of significant differences were noted between (a) 2008 River and 2005 McCullough and (b) 2008 River and 2005 Mixed datasets as compared to other inter-lithologic unit comparisons. As might be expected, the fewest number of significant differences were noted between the 2005 River and 2008 River datasets. Note that higher concentrations of arsenic in the 2008 River soils as compared to the 2005 River soils may be inferred from the Tukey HSD comparison results. For most constituents, the probability (*p*) values for the ANOVA/Kruskal-Wallis were less than 0.001 (Table F-2). Accordingly, the application of a Bonferroni correction to the significance level would not change the overall conclusions that differences exist among the four lithologic units and that the 2008 River dataset is significantly different than the three 2005 dataset for several constituents.

When the frequency of detections is less than 50 percent, even the nonparametric tests have little power to detect differences in central values (Smeti *et al*. 2007). For constituents with frequency of detects less than 50 percent and similar detection limits, a binomial proportions test was conducted to determine if frequency of detects between background datasets were comparable. Where frequency of detects were found to be similar, subsequent comparisons using detected-only data were conducted for infrequently detected constituents to identify potential similarities among background datasets.[21] Differences between the 2008 and the 2005 background datasets may also be inferred from these analyses (Table F-4 of Appendix F) and are summarized:

Constituent	**Sample Size*** (n > 4)	**Z-Test for Two Proportions**	**Additional Analysis Candidate**
Antimony	Yes	Similar frequency of detection	Yes
Boron	Yes	Similar frequency of detection	Yes
Silver	Yes	Dissimilar frequency of detection	No
Tin	Yes	Similar frequency of detection	Yes
Radium-228	Yes	Similar frequency of detection	Yes

* for two or more lithologic units

Comparisons of detected-only values between 2008 River and 2005 lithologic units were mixed for infrequently detected constituents—*i.e.*, differences may be inferred for some infrequently

[21] Only when datasets have comparable detection limits can this analysis be performed as a line of evidence to infer differences between datasets; otherwise, the test will only reflect differences in detection limits.

detected constituents; while no differences may be inferred for other infrequently detected constituents (Table F-9). Note that infrequently detected constituents are, by definition, characterized by a high proportion of censored data. Accordingly, it is both reasonable and defensible that study conclusions related to similarities/dissimilarities among background datasets consider the overall preponderance of the evidence from the more reliable statistical analyses associated with the majority of the 46 constituents with greater frequency of detects.

All in all, from these statistical comparisons, it may be inferred that the 2008 River data differ with respect to metal concentrations and radionuclide activities to the 2005 lithologic units. These findings are consistent with the findings reported in the Shallow Background Study (BRC/TIMET 2007). Therefore, it is recommended that the 2008 Supplemental Background dataset not be pooled with the 2005 BRC/TIMET background dataset for future applications; however, this will be evaluated on a case-by-case basis.

3.7.2 Comparison of 2008 Supplemental and 2005 BRC/TIMET Datasets (Depth-Specific Evaluations)

The 2008 Supplemental and 2005 BRC/TIMET background soil datasets were also evaluated on a depth interval-specific basis to further evaluate potential similarities/dissimilarities. Accordingly, two-sample tests were performed to compare the 2008 River to the 2005 McCullough datasets for 0 ft bgs, 5 ft bgs, and 10 ft bgs depths intervals.[22] ANOVA/Kruskal-Wallis analyses were performed for the 5-10 ft bgs combined dataset for the 2008 River, 2005 McCullough, and 2005 Mixed datasets[23] (Table F-3). The results of the statistical analyses are included in Appendix F. Probability plots, boxplots, and individual value plots were used to semi-quantitatively compare the 2008 Supplemental and 2005 BRC/TIMET data. These plots are included in Appendix D.

22 The sample size for constituents in the 2005 River and 2005 Mixed datasets for 0 ft bgs, 5 ft bgs and 10 ft bgs depth intervals were less than four (4) samples and were considered insufficient to support robust comparisons.

23 The sample size for constituents in the 2005 River dataset (5-10 ft bgs combined depth interval) were less than four (4) samples and were considered insufficient to support robust comparisons.

3.7.2.1 Two Sample Test Results (individual 0, 5 & 10 ft bgs comparisons)

Consistent with the findings of statistical comparisons described in the prior section, a number of differences in metal concentrations were inferred based on statistical comparisons between the 2008 River and the 2005 McCullough datasets (Tables F-6, F-7, and F-8 in Appendix F):

- Arsenic (all depths)
- Barium (all depths)
- Beryllium (5 and 10 ft bgs)
- Boron (all depths)
- Cobalt (all depths)
- Copper (5 and 10 ft bgs)
- Iron (5 ft bgs)
- Lead (5 and 10 ft)
- Lithium (10 ft bgs)
- Magnesium (0 and 10 ft bgs)
- Manganese (5 ft bgs)
- Nickel (all depths)
- Palladium (0 and 5 ft bgs)
- Phosphorus (all depths)
- Potassium (all depths)
- Silicon (5 ft bgs)
- Silver (0 ft bgs)
- Sodium (all depths)
- Strontium (0 and 5 ft bgs)
- Tin (5 ft bgs)
- Titanium (all depths)
- Vanadium (0 and 5 ft)
- Zirconium (all depths)

No differences in radionuclide activities were inferred based on the results of statistical comparisons for any of the three depth intervals (Tables F-6, F-7, and F-8 in Appendix F). For most constituents, the probability (*p*) value for at least one parametric or nonparametric two-sample tests is less than 0.001 (Tables E-6 through E-6). Accordingly, the application of a Bonferroni correction to the significance level would not change the overall conclusion that differences exist between 2008 River and 2005 McCullough on a depth interval basis.

3.7.2.2 ANOVA/Kruskal-Wallis Test Results (5 - 10 ft bgs combined)

Consistent with the Shallow Background Study (BRC/TIMET 2007), the datasets for the 5 ft bgs and 10 ft bgs depth intervals within a lithologic unit were combined to produce a dataset for the 5-to-10 (5-10) ft bgs depth interval. Overall, a number of significant differences in metal concentrations among the three lithologic units (2008 River, 2005 McCullough, and 2005 Mixed) were identified for the 5-10 ft bgs depth interval based on the results of ANOVAs/Kruskal-Wallis tests (Table F-3 in Appendix F). The only constituents for which no significant differences were identified include:

- Calcium
- Zinc
- Thorium-228
- Thorium-232

For most constituents, the probability (*p*) values for the ANOVA/Kruskal-Wallis tests were less than 0.001 (Table F-3). Accordingly, the application of a Bonferroni correction to the significance level would not change the overall conclusions that differences exist among the four lithologic units with respect to the 5-10 ft bgs depth interval.

Consistent with the Shallow Background Study (BRC/TIMET), no statistical tests were conducted for metals that had fewer than four detections in one or more of the unit-specific datasets, specifically:

- Antimony
- Boron
- Cadmium
- Chromium VI
- Mercury
- Niobium
- Platinum
- Selenium
- Silver
- Thallium
- Tungsten

Because these constituents were not subjected to statistical comparisons, it was not possible to determine whether significant differences were associated with the 5-10 ft bgs depth interval among the 2008 River, 2005 McCullough, and 2005 Mixed datasets.

Significant differences were noted between the 2008 River dataset and the datasets for the other two lithologic units (Table F-3 of Appendix F). More significant differences were identified between the 2008 River and 2005 McCullough datasets. However, differences in metal concentrations and radionuclide activities were inconsistent between the units—*i.e.*, one lithologic unit did not have consistently higher concentrations or activities. The 2005 Mixed dataset was nearly always indistinguishable from either one or both of the other two lithologic units. That is, for all elements except uranium-238, the 2005 Mixed dataset was (1) statistically indistinguishable from both the 2005 McCullough and the 2008 River datasets (*e.g.*, arsenic, lead); (2) statistically indistinguishable from the 2005 McCullough dataset but had inferred significant differences from the 2008 River dataset (*e.g.*, magnesium, manganese; or (3) statistically indistinguishable from the 2008 River dataset but had inferred significant differences from the 2005 McCullough dataset (*e.g.*, barium, tin) (Table F-3 of Appendix F)., This observation is consistent with the interpretation of the 2005 Mixed dataset being derived from soils that reflect a mixture of McCullough and River sediments. The 2005 Mixed dataset had significant differences inferred relative to the 2008 River dataset for several common parent elements (*e.g.*, silicon, aluminum, magnesium, potassium), which suggests a closer affinity between the Mixed and McCullough sediments.

The following constituents were considered to be present at higher concentrations in the 2008 River dataset than the other two datasets:

- Arsenic
- Palladium
- Silicon
- Strontium
- Chromium
- Potassium
- Sodium
- Uranium

For infrequently detected constituents (less than 50 percent frequency of detection), differences between the 2008 River and the 2005 datasets may also be inferred from these analyses (Table F-5 of Appendix F) and are summarized:

Constituent	**Sample Size*** (n > 4)	**Z-Test For Two Proportions**	**Additional Analysis Candidate**
Antimony	Yes	Similar frequency of detection	Yes
Radium-226	Yes	Similar frequency of detection	Yes
Radium-228	Yes	Similar frequency of detection	Yes

* for two or more lithologic units

Comparisons of detected-only values between 2008 River and 2005 lithologic units were mixed for infrequently detected constituents—*i.e.*, differences may be inferred for some infrequently detected constituents (antimony, boron); while no differences may be inferred for other infrequently detected constituents (radium-226, radium-228). Note that infrequently detected constituents are, by definition, characterized by a high proportion of censored data. Accordingly, it is both reasonable and defensible that study conclusions related to similarities/dissimilarities among background datasets consider the overall preponderance of the evidence from the more reliable statistical analyses for the vast majority of the 46 constituents with greater frequency of detects.

Again, when results of statistical comparisons are taken as a whole, it may be inferred that the 2008 River data differ with respect to metal concentrations and radionuclide activities to the 2005 lithologic units. These findings support the recommendation not to pool the 2008 Supplemental Background dataset with the 2005 BRC/TIMET background datasets for future applications.

3.7.3 Comparison of 2008 Supplemental Shallow Data by Depth Intervals

Soil samples were collected from three depth intervals from the 2008 Supplemental background soil study: 0 ft bgs, 5 ft bgs, and 10 ft bgs. Data for samples from each depth interval were compared using the statistical tests identified in Section 3.6.2. Multiple population (ANOVA) tests were selected and used to compare data among surface, middle shallow, and deeper shallow soil samples. The results of the statistical analyses are included in Appendix F. Results that are statistically significant at a p-level of 0.05 are indicated in each table (see Section 3.6.2.4 regarding correction for use of multiple tests). Boxplots and individual value plots shown in Appendix D compare the data by depth interval and offer a visual semi-quantitative appraisal of differences for each analyte among the groups of data. Statistical tests provide a quantitative analysis to determine if the differences are statistically significant at a specified significance level.

For the most part, metal concentrations were comparable among the three soil depth intervals (Table F-1 of Appendix F). Statistically significant differences in concentrations or activity among soil depth intervals were found for only seven of 46 constituents examined:

- Cobalt[24]
- Potassium
- Thorium-230
- Uranium-238
- Nickel
- Sodium
- Uranium-233/234

For most constituents, the probability (*p*) values for the ANOVA/Kruskal-Wallis tests were greater than 0.05 (Table F-1). Accordingly, the application of a Bonferroni correction to the significance level would not change the overall conclusions that few differences exist among the 0, 5, and 10 ft bgs depth interval for the 2008 supplemental shallow soil data (Table F-1). In fact, using a Bonferroni correction, differences for only two of 46 constituents would be statistically significant: concentrations of potassium and activities of uranium 233/234 (Table F-1).

---

[24] The ANOVA results for cobalt suggested that there were significant differences between lithologic units; however, the *post-hoc* testing did not identify specific differences.

The statistical comparisons found that statistically significant differences could be inferred primarily between (i) 0 ft bgs and 5 ft bgs and (ii) 0 ft bgs and 10 ft bgs for metals; no significant differences were inferred for metals between the 5 ft bgs and 10 ft bgs datasets. For radionuclides, comparisons found that statistically significant differences could be inferred primarily between the 0 ft bgs and 10 ft bgs datasets only. In addition to those apparent significant differences, only one other significant difference was inferred for radionuclides, for the thorium-230 5 ft bgs and 10 ft bgs datasets.

Differences in metal concentrations and radionuclide activities were inconsistent between the units—*i.e.*, one lithologic unit did not have consistently higher concentrations or activities. Sodium concentrations and radionuclide activities were found to be greater for the 10 ft bgs depth interval as compared to the other depth intervals. Nickel and potassium concentrations were found to be greater in the 0 ft bgs depth interval as compared to deeper intervals.

Although some identified statistically significant differences were observed for the above metals and radionuclides, these differences may not be significant from a geochemical perspective. Nonetheless, the findings of these statistical analyses suggest that the 0 ft bgs, 5 ft bgs, and 10 ft bgs depth intervals may be pooled and applied as a single dataset for future applications.

3.7.4 Inter-Element Correlations

In addition to statistical tests comparing background soils data among lithologic units and depth intervals, 2008 River data were evaluated with respect to inter-element correlations. Correlations or "measures of association" are of interest because they offer another line of evidence to distinguish background and non-background data or multiple populations of data (BRC/TIMET 2007). Correlation analyses[25] were conducted and used to identify those constituent pairs whose scatterplots should be examined to ascertain whether high-concentration outliers should be considered background. Both parametric (Pearson's product-moment) and nonparametric (Kendall tau) correlation coefficients are presented in correlation matrices (Appendix G). Note that statistically significant correlation coefficients (at a significance level of 0.05)[26] are indicated by bold font and are color-coded for parametric and nonparametric coefficients in each table. Scatterplots for constituents with significant correlation coefficients and high-concentration outliers are also presented in Appendix G.

[25] All correlation analyses were performed using SPSS v. 15.
[26] A Bonferroni correction was not applied to the correlation analyses because these analyses were used to identify constituents requiring further analysis and not for distinguishing between datasets using multiple tests.

Statistically significant associations were observed for several elements. The association of aluminum with trace metals was evaluated, and statistically significant associations were found for barium, beryllium, cadmium, cobalt, copper, iron, lead, manganese, nickel, phosphorus, potassium, silicon, silver, tin, titanium, uranium, vanadium, and zirconium (Table G-1 of Appendix G). Strong inter-element correlations are normally expected between alkaline and alkaline-earth metals (BRC/TIMET 2007)—for the supplemental background data, statistically significant correlation coefficients between alkaline and alkaline-earth metals ranged from 0.25 to 0.40 (Table G-3 of Appendix G). These associations may be useful in distinguishing soils derived from different source materials and in distinguishing site-related contamination from natural background. Statistically significant associations among uranium-238 decay chain radionuclides were also observed—correlation coefficients ranged from 0.32 to 0.54 (Table G-5 of Appendix G). Correlation among activities for radionuclides within the decay chain (parents and daughters) is anticipated, unless there are differences in geochemical behavior and mechanisms to separate the species (BRC TIMET 2007).

Note that statistically significant associations were observed for several metals and radionuclides; however these statistical associations should also be evaluated based on known geochemical characteristics.

3.7.5 Scatterplots

In addition to the calculated inter-element correlations, scatterplots with regression lines provide a visual assessment of inter-element associations. Statistically significant associations and high-concentration outliers were identified for several elements within the 2008 dataset (Appendix G):

- Aluminum
- Arsenic
- Barium
- Copper
- Lithium
- Nickel
- Palladium
- Silver
- Strontium

Scatterplots for identified constituent pairs were examined to determine whether high-concentration outliers are consistent with background (Appendix G)—*i.e.*, high-concentration outliers were “near” the linear least-square trend line. To identify potential deviations from trend lines, constituents listed above were plotted against constituents that were correlated and considered ubiquitous and relatively constant for identified lithologic units—*i.e.*, aluminum, iron, and magnesium. In general, no consistent and conspicuous deviations from least-square trend lines were observed for high concentration outliers.

Certain inter-element relationships are expected on the basis of geochemical behavior and expected mineralogical associations. For example, alkaline metals (such as lithium, sodium, and potassium) and alkaline-earth metals (such as barium, calcium, and magnesium) can be expected to behave similarly in solution and may therefore be expected to show an association in certain environmental media. Other metals are found in association in common minerals and show correlations in soils containing these minerals (such as feldspars; metal oxides such as hematite, goethite and pyrolusite; and carbonate minerals such as calcite). These associations are useful in distinguishing soils derived from different source materials and in distinguishing site-related contamination from natural background.

The association of aluminum with trace metals was also evaluated. Trace metals such as chromium, cobalt, copper, nickel, and vanadium may occur as impurities in the common alumino-silicate family of minerals known as feldspars. Clays and other secondary aluminum minerals in soils may host sorption sites for trace metals, thereby associating these metals. In general, these associations are evident.

Scatterplots were also constructed for radionuclides within the thorium-232 and uranium-238 decay chains and are included in Appendix G. Species within the decay chains (parents and daughters) should show statistically significant correlations in most cases unless there are great differences in geochemical behavior and sufficient mechanisms to separate the species. The same generally holds true for radionuclides in the thorium-232 decay chain (radium-228 and thorium-228). In general, most of the radionuclides in the uranium-238 decay chain (radium-226, thorium-230, and uranium-233/234) did show significant associations.

Finally scatterplots were constructed for arsenic and other metals commonly found at high levels in the Upper Ponds (chromium, lead, manganese, and vanadium) as well as radium-226 to support the contention that the 2008 Supplemental dataset is representative of background. Some correlation betweens these elevated levels would be expected in the ponds given the depositional history of the site. In general, most of these contaminants did show varying degrees of visual correlation with arsenic, with the possible exception of manganese. If aerial deposition of wind-borne dusts from Site operations were occurring at the background locations, a similar pattern may be expected. However, these same metals and radium-226 did not show any correlation with arsenic in either the 2008 supplemental or 2005 BRC/TIMET background datasets. Although some correlation appears evident between arsenic and vanadium in the 2008 Supplemental dataset, this is primarily driven by their highest concentrations being found in the same sample (BRC-BKG-R09) in the subsurface (10 ft bgs); likely not a result of contamination from the site.

4.0 SUMMARY AND CONCLUSIONS

The purpose of the 2008 Supplemental shallow soil background study was to collect and analyze data for metals and radionuclides in background shallow soils that are representative of soils in geologic units not covered by the existing 2005 background shallow soil dataset (BRC/TIMET 2007). The objective of this report was to determine whether these data, which are assumed representative of another geology, may be added to the background data pool to accommodate background comparisons at portions of the Common Areas (*i.e.*, the Mohawk sub-area and portions of Parcel 4B).

Soil sampling was conducted in April 2008. Samples were collected from 10 soil boring locations that represent the specific lithologies targeted by this supplemental shallow soil background sampling study and that extend the representative range of soils found in the vicinity of the Site. A total of 30 field and three duplicate soil samples were collected from the 10 borings for analysis. The data validation for the 2008 Supplemental dataset included 20 percent full validation and 100 percent partial validation. Results qualified as estimated based on the data validation are usable for the purposes of establishing background concentrations and for comparison to site-specific sample data. No soil sample results were rejected. One hundred percent of the dataset were validated as usable, indicating that the overall data collection objectives for the study were met. However, as noted in Section 3.5, for a few metals (*e.g.*, cadmium, selenium, and silver), variations in RDLs may have affected the frequency of detection and the validity/applicability of statistical analyses between the 2008 and 2005 background datasets as well as in comparisons of these data to future site data.

Several statistical outliers were found in the dataset, which is a common, anticipated observation for a dataset of this size. Moreover, these potential outliers occur sporadically and there are no apparent geology-based causes for these outliers. Accordingly, these outliers were considered likely due to naturally occurring variability.

Based on sampling location characteristics information obtained from published documentation, site inspection, and sample collection, it is reasonable to conclude that the background samples collected as part of this investigation reflect background soil conditions that may be used to support assessments of soils at the Mohawk sub-area and Parcel 4B. As discussed in Section 2.4, SVOC analyses were used to assess the potential for impacts to the sampling locations from anthropogenic sources. SVOC detections in surface soil samples collected at the background sampling locations are limited to bis(2-ethylhexyl)phthalate, a common lab contaminant. Therefore, the SVOC data did not provide any evidence suggesting that use of the samples for characterizing background conditions would be inappropriate. The results of correlation analyses and scatterplots also corroborate the conclusion that this dataset is appropriate for use as a representative background soil dataset.

Key findings from the analyses of the shallow background soils data include:

- Based on the statistical analyses performed, there appear to be distinct differences between the populations associated with sediments derived primarily from the McCullough and River Mountains, and with sediments representing a mixture of both sources. It is therefore appropriate to perform comparisons of background to Site data using the subset of background data that most closely matches the geologic conditions of that part of the Site as follows:

Portion of Site	**Applicable Background Dataset**
Southeastern portion (*e.g.*, Mohawk)	2008 River dataset
Northeastern portion	2005 McCullough and Mixed datasets
Northwestern portion (*e.g.*, Western Hook)[27]	2005 McCullough dataset
Central portion	2005 McCullough and Mixed datasets

- Because statistical analyses suggest that the 2008 Supplemental and 2005 BRC/TIMET datasets exhibit a number of statistically significant differences, it is recommended not to combine these datasets in support of future comparisons to site data. Potential exceptions to this recommendation will be considered on a case-by-case basis—for example, for areas of the site that may occur at the interface of different geologic units (*e.g.*, Parcel 4B).

- Findings of the ANOVA/Kruskal-Wallis tests found few statistically significant differences among the 0, 5, and 10 ft bgs depth intervals for the 2008 River background data. This findings suggests that data for the 0, 5, and 10 ft bgs depth intervals may be pooled and applied as a single dataset, promoting more powerful statistical analyses for future assessments in support of decision-making.

[27] Note that portions of surface and/or near surface soils in the northwestern portion of the Site may also be associated with the Upper Muddy Creek formation (UMCf).

- Because of the limited inferred differences in the depth-specific sample populations for the 2008 River unit, it is not necessary or appropriate to compare depth-specific Site data to the associated depth-specific background dataset.

Although the various background datasets are all contained within the project database, combining the background dataset by depth and/or lithology for subsequent comparison with Site data will be influenced by potential exposures at varying depth intervals and the location of a particular receptor – in other words, based on data usability and conceptual site model considerations.

These findings suggest that these data are appropriate for supporting future assessments and decision-making with respect to soils at sites within the Complex and Common Areas. Specific decisions regarding how best to use the background soils data for future Site-to-background comparisons will be made on a case-by-case basis in consultation with NDEP.

5.0 REFERENCES

Basic Remediation Company (BRC). 2008. Supplemental Background Shallow Soil Sampling and Analysis Plan, BMI Complex and Common Areas Vicinity, Clark County, Nevada. March.

Basic Remediation Company (BRC), ERM, and MWH. 2008. BRC Field Sampling and Standard Operating Procedures. BMI Common Areas, Clark County, Nevada. December.

Basic Remediation Company (BRC) and Titanium Metals Corporation (TIMET). 2007. Background Shallow Soil Summary Report, BMI Complex and Common Areas Vicinity. March.

Basic Remediation Company (BRC) and ERM. 2008a. BRC Quality Assurance Project Plan. April.

Basic Remediation Company (BRC) and ERM. 2008b. Data Validation Summary Report: Supplemental Shallow Soil Background Sampling Event; April 2008 (Dataset 34b). June.

Environ. 2003. Risk Assessment for the Water Reclamation Facility Expansion Site, Henderson, Nevada. Volume II, Appendix E. October 15.

Geotechnical & Environmental Services, Inc. (GES). 2008. BRC Supplemental Shallow Background Data Set. April 24.

Gilbert, R.O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold Company, New York, NY.

Mantel, N. 1981. Calculation of Scores for the Wilcoxon Generalization Applicable to Data Subject to Arbitrary Censorship. Am. Stat 35(4): 244-.

Nevada Bureau of Mines and Geology (NBMG). 1980. Las Vegas SE Folio Geologic Map (1977) and the Geologic Map of the Henderson Quadrangle, Nevada.

Nevada Division of Environmental Protection (NDEP). 2008a. Supplemental Guidance for Assessing Data Usability for Environmental Investigations at the BMI Complex and Common Area in Henderson, Nevada.

Nevada Division of Environmental Protection (NDEP). 2008b. Guidance on the Development of Summary Statistics Tables. December 10.

Singh, A. and A.K. Singh. 2007. ProUCL Version 4.0. Technical Guide. Prepared for the U.S. Environmental Protection Agency by Lockheed Martin Environmental Services.

Smeti, E.M., L.P. Kousouris, P.C. Tzoumerkas, and S.K. Golfinopoulos. 2007. Trend analysis and variability of microbiological parameters of the Mornos Reservoir. Proc. 10th Internat'l Conf. on Environ. Sci. Tech.

Sokal, R.R. and F.J. Rohlf. 1981. Biometry. Second Edition. W.H. Freeman and Company. San Francisco, CA.

SPSS, Inc. (SPSS). 2006. SPSS Base User's Guide. Chicago, IL.

U.S. Department of Agriculture (USDA). 1985. Soil Survey of Las Vegas Valley Area Nevada-Part of Clark County. Soil Conservation Service. July.

U.S. Department of Energy (DOE). 1997. Evaluation of Radiochemical Data Usability. Office of Environmental Management. Oak Ridge, Tennessee. April.

U.S. Department of Navy (Navy). 1999. Handbook for Statistical Analysis of Environmental Background Data.

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

U.S. Environmental Protection Agency (USEPA). 1992. Guidance for Data Usability in Risk Assessment. Part A. Office of Emergency and Remedial Response, Washington D.C. Publication 9285.7-09A. PB92-963356. April.

U.S. Environmental Protection Agency (USEPA). 1999. National Functional Guidelines for Organic Data Review. USEPA 540/R-99-008. OSWER 9240.1-05A-P. October.

U.S. Environmental Protection Agency (USEPA). 2001. National Functional Guidelines for Low-Concentration Organic Data Review. USEPA 540-R-00-006. OSWER 9240.1-34. June.

U.S. Environmental Protection Agency (EPA). 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, DC. September.

U.S. Environmental Protection Agency (USEPA). 2004. National Functional Guidelines for Inorganic Data Review. USEPA 540-R-04-004. OSWER 9240.1-45. October.

U.S. Environmental Protection Agency (EPA). 2006a. Data Quality Assessment: Statistical Methods for Practitioners. EPA QA/G-9S. Office of Environmental Information, Washington, DC. EPA/240/B-06/003. February.

U.S. Environmental Protection Agency (EPA). 2006b. Guidance on Systematic Planning Using the Data Quality Objectives Process. EPA QA/G-4. Office of Environmental Information, Washington, DC. EPA/240/B-06/001. February.

Zar, J.H. 1984. Biostatistical Analysis. Prentice-Hall, Inc. Englewood Cliffs, N.J.

Appendix E
Discussion of Statistical Outliers

Anomalously high statistical outliers were identified using the criterion identified in Section 3.4 of the report for the following constituents:

Arsenic	BRC-BKG-R02 (5 ft bgs) BRC-BKG-R09 (10 ft bgs)	Silicon	BRC-BKG-R10 (0 ft bgs)
Boron	BRC-BKG-R09 (10 ft bgs)	Sodium	BRC-BKG-R09 (0 ft bgs)
Cadmium	BRC-BKG-R01 (0 ft bgs) BRC-BKG-R10 (5 ft bgs) BRC-BKG-R09 (10 ft bgs)	Thallium	BRC-BKG-R04 (0 ft bgs)
Copper	BRC-BKG-R01 (0 ft bgs)	Tin	BRC-BKG-R01 (0 ft bgs)
Lead	BRC-BKG-R01 (0 ft bgs) BRC-BKG-R04 (0 ft bgs)	Uranium	BRC-BKG-R09 (10 ft bgs)
Magnesium	BRC-BKG-R09 (5 ft bgs)	Thorium-230	BRC-BKG-R08 (10 ft bgs)
Manganese	BRC-BKG-R04 (0 ft bgs) BRC-BKG-R02 (10 ft bgs)	Thorium-232	BRC-BKG-R04 (10 ft bgs)
Molybdenum	BRC-BKG-R01 (0 ft bgs)	Uranium-233/234	BRC-BKG-R08 (10 ft bgs)
Phosphorus	BRC-BKG-R09 (0 ft bgs)	Uranium-235/236	BRC-BKG-R01 (5 ft bgs)
		Uranium-238	BRC-BKG-R08 (10 ft bgs)

As seen above, several samples exhibit statistical outliers for one or more constituents. However, no one sample is routinely anomalously high in a way that suggests the associated detections are not representative of background. That said, the surface samples at locations BRC-BKG-R01 and BRC-BKG-R04 exhibited elevated constituent concentrations relative to the other samples (*i.e.*, BRC-BKG-R01 and BRC-BKG-R04) as follows:

- The surface sample at location BRC-BKG-R01 had the highest detected value for several metals (aluminum, beryllium, cadmium, chromium, cobalt, copper, iron, lead, molybdenum, nickel, potassium, tin, titanium, and zinc), and in several instances it is the highest of either

2005 BRC/TIMET or 2008 Supplemental datasets (aluminum, cadmium, chromium, copper, iron, lead, molybdenum, potassium, and tin).

- The surface sample at location BRC-BKG-R04 also had high detect values for several metals (lead, manganese, potassium, and thallium).

As discussed in Section 3.7.4, these values were further evaluated using correlation analysis/scatter plots to evaluate whether they were statistical outliers. This analysis identified no statistical outliers. Furthermore, there is no consistent pattern to the data that would suggest that the data are not indicative of naturally occurring background conditions. Sample locations BRC-BKG-R01 and BRC-BKG-R04 are not adjacent to each other, and if aerial deposition of wind-borne dusts from Site operations were suspected, then higher levels of metals typically found in soils at the site; for example, arsenic and vanadium would be expected at the surface in these samples. However, this is not the case. As noted above, the highest arsenic concentrations are found in the subsurface (BRC-BKG-R02 at 5 ft bgs and BRC-BKG-R09 at 10 ft bgs).

The supplemental background sample locations are west of the River Mountains. Formations associated with these mountains contain volcanic intrusions that are known to contain elevated concentrations of naturally occurring arsenic (Bevans *et al.*, 1998). The supplemental background locations are geologically similar to the western and central portions of the Henderson Landfill (see Figure 2 for landfill location). The central portion of the landfill relates to the artificial fill area that covers the pediment and fan deposits of the River Mountains and further to the east the Horse Spring Formation (from CH2MHill 2006; approved by NDEP on August 7, 2006). The western portion relates to the uncovered areas of the pediment and fan deposits of the River Mountains and the modern wash deposits (CH2MHill 2006). Arsenic levels found in undisturbed areas from the western and central portions of the landfill ranged from 3.7 to 34 mg/kg. The two highest arsenic concentrations from the supplemental background dataset (sample location BRC-BKG-R02 at 5 ft bgs and sample location BRC-BKG-R09 at 10 ft bgs) are within this range. They are therefore likely due to naturally occurring variability.

REFERENCES

Bevans, H.E.; Lico, M.S.; Lawrence, S.J. 1998. Water Quality in the Las Vegas Valley Area and the Carson and Truckee River Basins, Nevada and California", 1992-96; updated March 19.

CH2MHill. 2006. Technical Memorandum: Henderson Landfill Response Program Site Soils Criteria. June 2.

2008 SUPPLEMENTAL SHALLOW SOIL BACKGROUND REPORT

Deleted: ¶

BMI COMMON AREAS (EASTSIDE)
CLARK COUNTY, NEVADA

Prepared for:
Basic Remediation Company (BRC)
875 West Warm Springs Road
Henderson, Nevada 89011

Prepared by:

ERM-West, Inc.
2525 Natomas Park Drive, Suite 350
Sacramento, California 95833

MARCH 2009

Deleted: DECEMBER 2008¶

Deleted: December 2008

I hereby certify that I am responsible for the services described in this document and for the preparation of this document. The services described in this document have been provided in a manner consistent with the current standards of the profession and to the best of my knowledge comply with all applicable federal, state and local statutes, regulations and ordinances. I hereby certify that all laboratory analytical data was generated by a laboratory certified by the NDEP for each constituent and media presented herein.

March 16, 2009

Dr. Ranajit Sahu, C.E.M. (No. EM-1699, Exp. 10/07/2009) Date
BRC Project Manager

Deleted: <sp>¶
<sp>
December 12, 2008¶

Deleted: *3*

TABLE OF CONTENTS

Field Code Changed (×21)
Deleted: 1
Deleted: 5
Deleted: 5
Deleted: 5
Deleted: 5.3
Deleted: 5.4
Deleted: 5.5
Deleted: 3

Deleted: December 2008

Deleted: 5.6

Field Code Changed

Field Code Changed

Field Code Changed

Deleted: *3*

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES

LIST OF APPENDICES

Deleted: December 2008

ABBREVIATION AND ACRONYM LIST

ANOVA	analysis of variance
bgs	below ground surface
BMI	Basic Management, Inc.
BRC	Basic Remediation Company
DOE	U.S. Department of Energy
DQIs	data quality indicators
DQOs	data quality objectives
DVSR	Data Validation Summary Report
FSSOP	Field Sampling and Standard Operating Procedures
GEL	General Engineering Laboratories
GiSdT®	Guided Interactive Statistical Decision Tools
HSD	Honestly Significant Difference
LCS	laboratory control sample
LCSD	laboratory control sample duplicate
MDA	minimum detectable activity
MDL	method detection limit
mg/kg	milligrams per kilogram
MS/MSD	matrix spike/matrix spike duplicate
NBMG	Nevada Bureau of Mines and Geology
NDEP	Nevada Division of Environmental Protection
NRS	Nevada Revised Statutes
PARCC	precision, accuracy, representativeness, comparability, and completeness
QA/QC	quality assurance/quality control
QAPP	Quality Assurance Project Plan
RDL	reporting detection limit
RPD	relative percent difference
SAP	Sampling and Analysis Plan
SQL	sample quantitation limit
SVOC	semi-volatile organic compound
SSURGO	Soil Survey Geographic
SOP	standard operating procedure
µg/kg	micrograms per kilogram
USDA	U.S. Department of Agriculture

Deleted: MDA minimum detectable activity¶

Deleted: pCi/g pico Curies per gram¶
PID photoionization detector¶
PQL practical quantitation limits¶

Deleted: pCi/g pico Curies per gram¶
PID photoionization detector¶

Deleted: QC quality control¶

Deleted:

Deleted: *3*

USEPA U.S. Environmental Protection Agency

WRS Wilcoxon Rank Sum

Deleted: December 2008

1.0 INTRODUCTION

On behalf of Basic Remediation Company (BRC), ERM-West, Inc. (ERM) has prepared this Supplemental Shallow Soil Background Report applicable to the Basic Management, Inc. (BMI) Complex and Common Areas in Clark County, Nevada. The supplemental shallow soil background data were collected in accordance with the *Supplemental Background Shallow Soil Sampling and Analysis Plan* (SAP) dated March 2008, and approved by the Nevada Division of Environmental Protection (NDEP) in March 2008. The general scope of work included the collection of soil samples from background areas upgradient of the Site industrial areas and analysis of these samples for metals and radionuclides that are of interest at sites within the Complex and Common Areas. In addition, since the sample locations were adjacent to Lake Mead Parkway, surface samples were analyzed for semi-volatile organic compounds (SVOCs), as well as field screened using a photoionization detector (PID).

Deleted:),

Deleted: site-related

This revision of the report, Revision 4, incorporates (1) comments received from the NDEP, dated August 1, 2008, on Revision 0 of the report, dated July 2008; (2) comments received from the NDEP, dated September 23, 2008, on Revision 1 of the report, dated August 2008; (3) resolution of issues discussed during teleconferences between NDEP and BRC on August 5, 2008, and September 26, 2008; (4) comments received from the NDEP, dated November 13, 2008, on Revision 2 of the report, dated October 2008; and (5) comments received from the NDEP, dated February 17, 2009, on Revision 3 of the report, dated December 2008. The NDEP comments and BRC's responses to these comments are included in Appendix A. Also included in Appendix A is a redline/strikeout version of the text showing the revisions from the December 2008 version of the report. An electronic version of the entire report, as well as original format files (MS Word and MS Excel) of all text and tables are included in Appendix B.

Deleted: 3

Deleted:)

Deleted: ,

Deleted: and

Deleted: October

1.1 OBJECTIVES AND PURPOSE

The purpose of this investigation was to collect and analyze data for metals and radionuclides in background shallow soils that are comparable to site soils in geologic units not covered by the existing *Background Shallow Soil Summary Report* (BRC/TIMET 2007) dataset. This supplemental background study was primarily undertaken because background comparisons for arsenic have failed at both the Mohawk and Parcel 4B sub-areas. However, there is no history of arsenic contamination at these sites; therefore, some consideration has been given to the possibility that the eastern part of the site exhibits different background levels of arsenic and, potentially, other metals. The northeastern part of the site is close to the northern part of the

Deleted: 2005

Deleted: north

Deleted: *3*

Deleted: December 2008

River Mountains range. A mile or two to the northeast of the Mohawk area, in the vicinity of the Henderson Landfill, and still in the River Mountains range, very high concentrations of arsenic have been observed in background samples (see discussion in Section 3.4). Consequently, the reason for collecting these supplemental background samples was so that a specific subset of background conditions could be used for comparison with site concentrations, primarily at the Mohawk and Parcel 4B sub-areas.

Deleted: 1.

Deleted: purpose of

At present, insufficient background data exist for alluvial fan materials downgradient of the northern River Mountains to evaluate whether concentrations of site-related chemicals detected in site samples in the eastern portion of the BMI Common Areas statistically exceed concentrations of these chemicals in background soil.[1] Therefore, the specific objectives proposed for the supplemental shallow soil background study included the collection of data:

Deleted: The existing BRC/TIMET background shallow soil dataset consists of samples collected almost exclusively from soils originating from the McCullough Range. Only background sample location BRC-BKG-12 is considered to be a mixed alluvium location. No samples during the BRC/TIMET background shallow soil investigation were collected exclusively from the alluvial fan materials downgradient of the River Mountains. Although there were several background samples collected by Environ (2003) in this geologic unit, given recent sample results at the site, the Environ data is considered inadequate for characterizing the northern part of the River Mountains.¶ Thus, at

Deleted: goals and comparisons

- From sampled soil units that are representative of Site soils not covered by the existing background shallow soil dataset;
- That form a sufficient sample population that can be used to support statistical comparison of on-site and background datasets;
- That could be used to evaluate the comparability of soil originating from geologic units from the River Mountains; that is, comparison of the northern River Mountains (this 2008 Supplemental dataset) with the southern River Mountains and McCullough Range (2005 BRC/TIMET dataset).

This supplemental shallow soil background sampling event specifically targeted the lithologic units defined as "Pediment and fan deposits of the River Mountains" depicted as being located in the southeastern-most edge of the Common Areas in the Nevada Bureau of Mines and Geology (NBMG) *Las Vegas SE Folio Geologic Map (1977) and the Geologic Map of the Henderson Quadrangle, Nevada* (NBMG 1980) (see Figure 1, Qr_1 and Qr_2 labels).

Deleted: Specific lithological units targeted by this

Deleted: are

Deleted: Qr1

Deleted: Qr2

Deleted:).

To support this data collection effort, soils collected from the background borings were analyzed for SVOCs to evaluate potential soil impacts at the background drilling locations. The underlying

[1] The existing BRC/TIMET background shallow soil dataset consists of samples collected almost exclusively from soils originating from the McCullough Range. Only background sample location BRC-BKG-12 is considered to be a mixed alluvium location. No samples during the BRC/TIMET background shallow soil investigation were collected exclusively from the alluvial fan materials downgradient of the River Mountains. Although there were several background samples collected by Environ (2003) in this geologic unit, given recent sample results at the site, the Environ data is considered inadequate for characterizing the northern part of the River Mountains.

Deleted: *3*

assumption was that if potential chemical impacts were observed at a given boring location, the designation of that boring as representing background conditions would be suspect.

1.2 SITE LOCATION AND GEOLOGIC SETTING

The Site is located in Clark County, Nevada, and is situated approximately 2 miles west of the River Mountains and 1 mile north of the McCullough Range (Figure 2). For reference, it is noted that the Upper Ponds occupy the southern portion of the BMI Common Areas, and the Lower Ponds occupy the northern part of the BMI Common Areas. The McCullough Range is the primary source of materials upslope of the BMI Complex, the Lower Ponds, and the western and central portions of the Upper Ponds. Both the River Mountains and the McCullough Range are primary sources of materials upslope of the eastern portion of the Upper Ponds. According to NBMG (1980), the River Mountains and McCullough Range consist of volcanic rocks: dacite in the River Mountains and andesite in the McCullough Range. The land surface slopes in a westerly to northwesterly direction from the River Mountains and in a northerly to northeasterly direction from the McCullough Range. Near the Site, the surface topography slopes in a northerly direction towards the Las Vegas Wash.

Deleted: the Nevada Bureau of Mines and Geology (

Deleted:) *Las Vegas SE Folio Geologic Map (1977) and the Geologic Map of the Henderson Quadrangle, Nevada* (NBMG

A soils map reproduced from the U.S. Department of Agriculture (USDA) Soil Survey Geographic (SSURGO) database shows that the soil type classification for the Upper and Lower Ponds area proper is map unit 600, “slickens,” a non-native soil type (artificial fill). This term is presumed to reflect the non-native material observed in those Ponds that were used for waste disposal. The soil type classification for the BMI Complex is map unit 615, “urban land.” Native soils underlying the slickens and urban land are assumed to be consistent with the surrounding map units (*i.e.*, primarily map unit 184, and, to a lesser extent, map units 112, 117, 182, 187 and 326). As seen in the USDA soils map excerpted on Figure 3 that is based on the 1985 USDA Soils Survey (USDA 1985), the area targeted in this investigation falls within the boundaries of mapped soil unit 182 (Caliza-Pittman-Arizo complex), which is the native soil type mapped as being present in the southeastern-most portion of the Common Areas and associated with the Qr_1 and Qr_2 lithologic units.

Deleted: The

Deleted: is presented in

Deleted: and

Deleted: December 2008

2.0 SUMMARY OF THE INVESTIGATION

This section identifies the sampling locations, presents the sampling and analytical methods, and summarizes the results of data validation.

2.1 SAMPLING LOCATIONS

Soil samples were collected from three depth intervals at each sampling location, including surface soil (0 to 0.5 feet below ground surface [bgs]), and two subsurface depths (4 to 6 feet and 9 to 11 feet bgs). The background soil study collected data for site-related metals and radionuclides. Data for SVOCs were also collected to evaluate whether the background soil locations are impacted by other anthropogenic sources.

Soil samples were collected from 10 initial sampling locations adjacent to Lake Mead Parkway, on the south side of the roadway away from the Site. These 10 locations are shown on Figure 1, along with sampling locations for the 2005 BRC/TIMET and 2003 Environ studies on Figure 2.

Deleted: and

The 10 sampling locations were selected because they exhibited the following characteristics:

- They are off-Site locations, in relatively close proximity to the Site; however, they are upgradient and sufficiently distant from the Site such that impacts from Site operations are not likely;
- They are upwind of the Site (wind direction plots indicate the predominant wind direction is from the south and southwest; see Figure 2) and are thus less likely to have been affected by aerial deposition of wind-borne dusts or vapors from Site operations; and
- They are upslope of the Site and are thus unlikely to have been affected by overland surface-water transport of potentially contaminated site sediments.

Deleted: <#>.¶

Deleted: .

Available background sample locations are constrained due to rapid development in the area. Undeveloped areas in close proximity to the site, without access problems, are scarce. Although the 10 locations are adjacent to Lake Mead Parkway, as can be seen from Figure 1 they are within undisturbed areas. Therefore, the 10 sampling locations were chosen because they exhibited the characteristics identified above and are considered adequate for representing undisturbed alluvial material washed down from the northern River Mountains.

Deleted: the

Deleted: *3*

2.2 SUMMARY OF SAMPLING PROCEDURES AND ANALYSES

Soil samples were collected from a single boring at each location, drilled using a hollow-stem auger rig. Samples were collected in a split-spoon sampler lined with stainless steel sleeves. Samples collected from each boring are considered independent samples. Sampling and sample handling procedures were consistent with the standard operating procedures (SOPs) developed for the BMI Common Areas as provided in the *BRC* Field Sampling and Standard Operating Procedures (FSSOP; BRC, ERM and MWH 2008). Subsurface soil samples were collected from each two-foot interval of drill core (*i.e.*, 4 to 6 feet bgs and 9 to 11 feet bgs).

Deleted: 2007

For this study, surface soil is defined as the upper 0.5 feet of the soil horizon; subsurface soil is defined as below 0.5 feet bgs. Soil samples were collected from three zones in each boring as follows:

- Surface Soil (soil samples collected from within the depth interval from 0-0.5 ft bgs; hereinafter referred to as “0 ft bgs” interval);
- Shallow Subsurface Soil (soil samples collected from within the depth interval from 4-6 ft bgs; core homogenized; hereinafter referred to as “5 ft bgs” interval); and
- Deeper Subsurface Soil (soil samples collected from within the depth interval from 9-11 ft bgs; core homogenized; hereinafter referred to as “10 ft bgs” interval).

Deleted:);

Ten borings were advanced and three samples from each zone were collected for an initial total of 30 soil samples. Field duplicate samples were collected at three locations; from locations BRC-BKG-R01 (0 ft bgs), BRC-BKG-R05 (0 ft bgs), and BRC-BKG-R08 (5 ft bgs) for metals and SVOCs; and from locations BRC-BKG-R01 (5 ft bgs), BRC-BKG-R05 (0 ft bgs), and BRC-BKG-R08 (5 ft bgs) for radionuclides. Inadequate sample volume was collected from location BRC-BKG-R01 (0 ft bgs), the first sample collected, which is why the field duplicate at this location for radionuclides is at a different depth (5 ft bgs) than that for metals and SVOCs. Because these samples are considered field duplicates, and not split samples, each is considered an independent sample. Therefore, there were a total of 33 soil samples collected as part of this investigation. Soil boring logs representing each location are also included in Appendix C.

The soil samples were submitted for analysis to TestAmerica in St. Louis, Missouri. Analyses were conducted at three TestAmerica laboratory locations: St. Louis, Missouri; Burlington, Vermont; and West Sacramento, California. General Engineering Laboratories (GEL), located in Charleston, South Carolina, performed the radionuclide analyses.[2] At the time of analysis, all laboratories were NDEP-certified laboratories for the analyses conducted. Surface and subsurface sample analyses consisted of a full suite of metals, eight radionuclides (radium-226, radium-228, thorium-228, thorium-230, thorium-232, uranium-233/234, uranium-235/236, and uranium-238), SVOCs, and general soil characteristics. The individual analytes, analytical methods, and reporting detection limits (RDLs) are presented in Table 1. These analytes and methods are consistent with the BRC site-related chemicals list and analytical program previously established in the BRC Quality Assurance Project Plan (QAPP; BRC and ERM 2008a). All radionuclide analyses underwent full dissolution preparatory methods. All preparatory methods and analyses are consistent with the 2005 BRC/TIMET background dataset.

Deleted: practical quantitation
Deleted: PQL

The detection frequencies for metals and radionuclides evaluated during this supplemental shallow soil background study are presented in Table 2. Detection frequencies observed for these analytes during the 2005 shallow background study are also provided in Table 2 for comparison. As seen in Table 2, most of the metals and radionuclides that are the subject of the supplemental shallow soil background investigation were detected routinely in the 2008 shallow soil samples. Exceptions are:

Deleted: frequency
Deleted: is
Deleted: on that table

- Antimony
- Boron
- Chromium (VI)
- Lithium
- Mercury
- Niobium
- Platinum
- Selenium
- Silver
- Thallium
- Tin
- Tungsten
- Uranium 235/236
- Zirconium

These fourteen constituents were detected in fewer than fifty percent of the samples in which they were analyzed during the supplemental shallow soil background investigation. Most of these same compounds were also not detected routinely during the 2005 shallow soil background investigation. Exceptions to this observation consist of lithium, mercury, tin and zirconium, which were routinely detected in the 2005 samples but not in the 2008 samples. Selenium and

[2] GEL labeled all primary samples that required matrix spike/matrix spike duplicates (MS/MSD) with the sample name specified on the chain-of-custody, but included an MS/MSD identification (*e.g.*, BRC-BKG-R02-5-MS/MSD). Due to the unaccustomed labeling, all samples with the MS/MSD were inadvertently regarded as quality control samples and not included with the original sample dataset. GEL was contacted and they confirmed the results for samples labeled as MS/MSD are actual primary sample results.

thallium were also detected at a noticeably lower frequency in the 2008 supplemental shallow samples than in the 2005 samples. In contrast, cadmium, silver, and uranium-233/234 were detected at a noticeably higher frequency in the 2008 supplemental shallow background samples than in those from the 2005 shallow background investigation. It should be noted that variations in detection frequencies are influenced by the associated RDL and may not reflect trends in actual concentrations; the effect of RDLs on detection frequencies is discussed further in Section 3.5.

2.3 DATA VALIDATION SUMMARY

All of the data were subjected to a Level 3 review. In addition to the Level 3 review, 20 percent of all data collected during the course of the investigation were subjected to full Level 4 data validation. Level 3 and 4 reviews are provided in the *Data Validation Summary Report (DVSR)—2008 Supplemental Shallow Soil Background Sampling Event* (BRC and ERM 2008b; approved by NDEP on June 9, 2008). Stable chemistry sample results (metals) for supplemental shallow soil background samples were validated in accordance with the following U.S. Environmental Protection Agency (USEPA) guidance document *U.S. EPA Contract Laboratory Program National Functional Guidelines for Inorganic Data Review* (USEPA 2004). USEPA has not standardized the validation of radionuclide data. Radionuclide results for supplemental shallow soil background samples were validated in accordance with SOP-40 (BRC, ERM and MWH 2008) and the project QAPP (BRC and ERM 2008a).

Based on data validation and review, data qualifiers were placed in the electronic supplemental shallow soil background database to classify whether the data were acceptable, acceptable with qualification, or rejected. Where applicable, an indication of result bias is presented. In addition, for every data validation qualifier, a secondary comment code was entered to indicate the reason for qualification. The DVSR (BRC and ERM 2008b) provides the definitions for the data validation qualifiers and comment codes used in the supplemental shallow soil background database. Validation qualifiers and definitions are based on those used by USEPA in the current validation guidelines (USEPA 2004) and summarized in the SOP-40 (BRC, ERM, and MWH 2008).

Results that are qualified as estimated may generally be usable for the purposes of establishing background and for comparison to Site-specific sample data. Based on the evaluation of the dataset, 100 percent of the data obtained during the field investigation are valid (that is, not

Deleted: December 2008

rejected) and acceptable for their intended use. With 100 percent of the dataset validated as usable, the overall objective of the data collection event was met.

2.4 DATA USABILITY EVALUATION

The analytical data were reviewed for applicability and usability following procedures in the *Guidance for Data Usability in Risk Assessment (Part A)* (USEPA 1992) and *Supplemental Guidance for Assessing Data Usability for Environmental Investigations at the BMI Complex and Common Area in Henderson, Nevada* (NDEP 2008a). A quality assurance/quality control (QA/QC) review of the analytical results was conducted during the sampling events. According to both NDEP's and USEPA's Data Usability Guidance, there are six principal evaluation criteria by which data are judged for usability. The six criteria are:

Deleted: the USEPA

- availability of information associated with site data;
- documentation;
- data sources;
- analytical methods and detection limits;
- data review; and
- data quality indicators, including precision, accuracy, representativeness, comparability, and completeness.

In addition to the six principal evaluation criteria, NDEP's Data Usability Guidance includes a step for data analysis. Items for this step are discussed in Section 3. A summary of these six criteria for determining data usability is provided below. Data usability evaluation tables are provided in Appendix B.

Criterion I – Availability of Information Associated with Supplemental Shallow Soil Background Data

The usability analysis of the supplemental shallow soil background data requires the availability of sufficient data for review. The required information is available from documentation associated with the data collection efforts. Data have been validated per the NDEP-approved DVSR (BRC and ERM 2008b). The following lists the information sources and the availability of such information for the data usability process:

Deleted: *3*

Deleted: December 2008

- Background description and objectives provided in the NDEP-approved SAP (BRC 2008) and in Section 1.

Deleted: workplan

- A site map with sample locations is provided on Figure 1.

Deleted: in

- Sampling design and procedures were provided in the NDEP-approved SAP (BRC 2008) and discussed in Sections 2.1 and 2.2.

Deleted: workplan

- Analytical methods and detection limits are provided in Table 1.
- A complete dataset is provided in Appendix B.
- Field conditions and physical parameter data as applicable to the background dataset are provided in the field investigation report (GES 2008) and DVSR (BRC and ERM 2008b).
- The laboratory provides a narrative with each analytical data package outlining any problems encountered in the laboratory, control limit exceedances, and rationale for any deviations from protocol. These narratives are included as part of the DVSR (BRC and ERM 2008b).

Deleted: exceedences

- QC results are provided by the laboratory, including blanks, replicates, and spikes. The laboratory QC results are included as part of the DVSR (BRC and ERM 2008b).
- Data flags used by the laboratory were defined adequately.
- Electronic files containing the raw data made available by the laboratory are included as part of the DVSR (BRC and ERM 2008b).

Criterion II – Documentation Review

The objective of the documentation review is to confirm that the analytical results provided are associated with a specific sample location and collection procedure, using available documentation. For the purposes of this data usability analysis, the chain-of-custody forms prepared in the field were reviewed and compared to the analytical data results provided by the laboratory to ensure completeness of the dataset as discussed in the DVSR (BRC and ERM 2008b). Based on the documentation review, all samples analyzed by the laboratory correspond to their respective geographic locations as discussed in Section 2 and shown on Figure 1. The samples were collected in accordance with the NDEP-approved SAP (BRC 2008) and SOPs developed for the BMI Common Areas as provided in the FSSOP (BRC, ERM and MWH 2008). Field procedures included documentation of sample times, dates and locations, and other sample-

Deleted: 3

Deleted: December 2008

specific information (*e.g.*, sample depth). Information from field forms generated during sample collection activities was imported into the project database.

The analytical data were reported in a format that provides adequate information for evaluation, including appropriate quality control measures and acceptance criteria. Each laboratory report describes the analytical method used, provides results and detection limits on a sample-by-sample basis, and provides the results of appropriate quality control samples (*e.g.*, laboratory control spike samples, sample surrogates and internal standards [organic analyses only], and matrix spike samples). All laboratory reports provided the documentation required by USEPA's Contract Laboratory Program (USEPA 1999, 2001, 2004) which includes chain of custody records, calibration data, QC results for blanks, duplicates, and spike samples from the field and laboratory, and all supporting raw data generated during sample analysis. Reported sample analysis results were imported into the project database.

Criterion III –Data Sources

The review of data sources is performed to determine whether the analytical techniques used in the site characterization process are appropriate for the exposure area and medium of interest and that appropriate analytical methods were used. The data collection activities were developed to characterize a broad spectrum of background metals and radionuclides in soil. As described in the SAP, samples were collected in areas of no known impacts for the target soil lithologies. The State of Nevada is in the process of certifying the laboratories used to generate the analytical data. As such, standards of practice in these laboratories follow the quality program developed by the Nevada Revised Statutes (NRS) and are within the guidelines of the analytical methodologies established by the USEPA. Based on the review of the available information, the data sources for chemical and physical parameter measurements are adequate for use.

Criterion IV – Analytical Methods and Detection Limits

In addition to the appropriateness of the analytical techniques evaluated as part of Criterion III, it is necessary to evaluate whether the detection limits are low enough to allow adequate characterization of the data. At a minimum, this data usability criterion can be met through the determination that routine USEPA reference analytical methods were used in analyzing the samples. Table 1 identifies the USEPA methods that were used in conducting the laboratory analysis of soil samples. Each of the identified USEPA methods is considered the most appropriate method for the respective constituent class and each was approved by NDEP as part of the SAP (BRC 2008).

Deleted: workplan

Deleted: 3

Laboratory RDLs were based on those outlined in the reference method, the SAP, and the project QAPP (BRC and ERM 2008a). In accordance with respective laboratory SOPs, the analytical processes included instrument calibration, laboratory method blanks, and other verification standards used to ensure quality control during the analyses of collected samples.

Datasets with multiple detection limits are not uncommon in analytical chemistry data. As discussed in Section 2.2, fourteen constituents were detected in fewer than fifty percent of the samples--differences in detection limits is anticipated to have the greatest effect on calculations of descriptive statistics for these constituents. With regard to future statistical analyses, datasets with different detection limits are not anticipated to severely impact proposed statistical comparisons to background. BRC uses the computer statistical software program Guided Interactive Statistical Decision Tools (GiSdT®; Neptune and Company 2007) to conduct non-parametric tests including the Wilcoxon Rank Sum (WRS) test, quantile test, and slippage test. The Gehan ranking system is used for these tests to accommodate multiple detection limits within the same dataset. However, if detection limits are among the largest values in the dataset, then conclusions from the statistical test results should be treated with caution.

Criterion V – Data Review

The data review portion of the data usability process focuses primarily on the quality of the analytical data received from the laboratory. However for this study, the data review also included evaluation of the SVOC data to identify any evidence of impacts that might indicate that these locations are not suitable for consideration as background. Both elements are discussed below.

Data Quality Review. Soil sample data were subject to data validation. The DVSR was prepared as a separate deliverable (BRC and ERM 2008b). The analytical data were validated according to the internal procedures using the principles of USEPA National Functional Guidelines (USEPA 1999, 2001, 2004) and were designed to ensure completeness and adequacy of the dataset. Any analytical errors and/or limitations in the data have been addressed and an explanation for data qualification provided in the respective data tables. The results of ERM's data review for these issues are presented in the DVSR and are summarized as qualifiers in the dataset provided electronically in Appendix B.

- **Deleted:** December 2008
- **Deleted:** reporting limits
- **Deleted:** workplan
- **Deleted:** , though unfortunate,
- **Deleted:** the nonparametric test
- **Deleted:**),
- **Deleted:** , and Quantile tests
- **Deleted:** WRS test (as currently supported by GiSdT®) can
- **Deleted:** datasets having
- **Deleted:** and the slippage and Quantile tests are affected when non-detected values
- **Deleted:** (see Section 3.5.2).
- **Deleted:** [3]
- **Deleted:** of
- **Deleted:** *3*

Deleted: December 2008

For some analytical results, quality criteria were not met and various data qualifiers were added to indicate limitations and/or bias in the data. The definitions for the data qualifiers, or data validation flags, used during validation are those defined in SOP-40 (BRC, ERM and MWH 2008) and the project QAPP (BRC and ERM 2008a). Sample results are rejected based on findings of serious deficiencies in the ability to properly collect or analyze the sample and meet QC criteria. Only rejected data are considered unusable for decision-making purposes. No samples were rejected in the supplemental shallow soil background dataset. Sample results qualified as estimated indicate an elevated uncertainty in the value. A bias flag may have been applied to indicate a direction of the bias. Estimated analytical results are included in the supplemental shallow soil background dataset.

Deleted: 2007

Evaluation for Evidence of Impacts/Background Unsuitability. The surface samples at each boring location[4] were analyzed for SVOCs. As previously noted, the purpose of these analyses was to identify any evidence of impacts that might indicate that these locations are not suitable for consideration as background. As summarized in Table 3, only one SVOC was detected in the samples; bis(2-ethylhexyl)phthalate, a common laboratory contaminant, was detected at low concentrations (56 micrograms per kilogram [µg/kg] and 69 µg/kg[5]) in the two samples collected from location BRC-BKG-R01 (initial and field duplicate). The RDLs for the SVOC analyses were relatively low (*i.e.*, approximately 340 µg/kg for most compounds), and are consistent with the RDLs presented in the project QAPP (BRC and ERM 2008a). Furthermore, the data review performed for the SVOC data did not identify any issues of concern with respect to the SVOC data quality (BRC and ERM, 2008b). Therefore, the SVOC data did not provide any evidence suggesting that use of the samples for determining background conditions would not be appropriate.

Deleted: In addition, under this criterion, the SVOC data

Deleted: evaluated

Deleted: detections of these constituents are sporadic and

Deleted: , and no evidence of appreciable impacts was observed

Criterion VI – Data Quality Indicators

Data quality indicators (DQIs) are used to verify that sampling and analytical systems used in support of project activities are in control and the quality of the data generated for this project is appropriate for making decisions affecting future activities. The DQIs address the field and analytical data quality aspects as they affect uncertainties in the data collected. The DQIs include precision, accuracy, representativeness, comparability, and completeness (PARCC). The project QAPP provides the definitions and specific criteria for assessing DQIs using field and laboratory

[4] There was one exception – the surface soil sample at location BRC-BKG-R09 was not analyzed for SVOCs.
[5] Both results were flagged as estimated (J) due to their low concentrations below the RDLs.

Deleted: 3

QC samples and is the basis for determining the overall quality of the dataset. Data validation activities included the evaluation of PARCC parameters, and all data not meeting the established PARCC criteria were qualified during the validation process using the guidelines presented in the National Functional Guidelines (USEPA 1999, 2001, 2004).

Precision is a measure of the degree of agreement between replicate measurements of the same source or sample. Precision is expressed by relative percent difference (RPD) between replicate measurements. Replicate measurements can be made on the same sample or on two samples from the same source. Precision is generally assessed using a subset of the measurements made. The precision of the data was evaluated using several laboratory QA/QC procedures such as field duplicates, laboratory duplicates, laboratory control sample (LCS), laboratory control sample duplicate (LCSD), and MS/MSD results. Based on ERM's review of the results of these procedures, there do not appear to be any wide-spread data usability issues associated with precision.

Accuracy measures the level of bias that an analytical method or measurement exhibits. To measure accuracy, a standard or reference material containing a known concentration is analyzed or measured and the result is compared to the known value. Several QC parameters are used to evaluate the accuracy of reported analytical results:

- Holding times and sample temperatures;
- LCS percent recovery;
- MS/MSD percent recovery (organics);
- Spike sample recovery (inorganics)
- Surrogate spike recovery; and
- Blank sample results.

Detailed discussions of and tables with specific exceedances, with respect to precision and accuracy, are provided in the NDEP-approved DVSR (BRC and ERM 2008b) and data qualified as a result of this evaluation are presented with qualifiers in the dataset provided electronically in Appendix B.

Representativeness is the degree to which data accurately and precisely represent a characteristic of the population at a sampling point or an environmental condition (USEPA 2002). There is no

Deleted: December 2008

standard method or formula for evaluating representativeness, which is a qualitative term. Representativeness is achieved through selection of sampling locations that are appropriate relative to the objective of the specific sampling task, and by collection of an adequate number of samples from the relevant types of locations.

Completeness is commonly expressed as a percentage of measurements that are valid and usable relative to the total number of measurements made. Analytical completeness is a measure of the number of overall accepted analytical results, including estimated values, compared to the total number of analytical results requested on samples submitted for analysis after review of the analytical data. None of the data were eliminated due to data usability concerns. The percent completeness for the dataset is 100 percent.

Comparability is a qualitative characteristic expressing the confidence with which one dataset can be compared with another. The desire for comparability is the basis for specifying the analytical methods; these methods are consistent with those used in the 2005 BRC/TIMET background dataset. The comparability goal is achieved through using standard techniques to collect and analyze representative samples and reporting analytical results in appropriate units. The ranges of sample results from both the supplemental shallow soil background dataset and the 2005 BRC/TIMET background dataset are provided electronically in Appendix B. As discussed in Section 2.4, differences in detection limits among datasets may affect data comparability for datasets comprised primarily of non-detected values. For these datasets, left-censored data can result in difficulties in differentiating whether datasets are actually different or merely an artifact of detection limits. Note that for constituents with detection limits that meet data quality objectives (DQOs), comparisons between site and background may be less important as these left-censored data are likely to indicate conditions that pose an "acceptable" risk and further analysis is not necessary.

Deleted: shallow

Deleted: soil

Deleted: 3

Deleted: December 2008

3.0 STATISTICAL METHODS AND FINDINGS

The exploratory data analysis and statistical evaluation of data for background soils generally followed industry-standard guidance documents (USEPA 2006a,b; Navy 1999, 2002) and standards agreed upon with NDEP, including the *Guidance on the Development of Summary Statistics Tables* (NDEP 2008b). These guidance documents discuss the use of statistical plots, calculation of summary statistics (such as the arithmetic mean), treatment of non-detect data, and selection of statistical tests. The following sections discuss data preparation, statistical plots, summary statistics and statistical tests, and the types of comparisons conducted.

Deleted: 2000a and 2000b and

Deleted: and

Deleted: usability, data

3.1 DATA PREPARATION

3.1.1 Spatial Independence Assumptions

There are 10 soil boring locations that were sampled for the supplemental shallow soil background dataset. The 10 soil boring locations are treated as spatially independent in this background soil study. The concentrations of each analyte at each sample location and depth is dependent on the origin of the sediment and the composition of the parent material (with the exception of anthropogenic deposition of analytes such as lead).

Naturally occurring variability is associated with the deposition of sediments, and these variations may never be fully characterized and result in unexplainable data clusters. The naturally occurring variability may be impacted by sediment transport, leaching, weathering, and other geochemical processes within the alluvium; therefore, when statistical tests are performed, it is expected that some spatial correlation may be seen, but the impact of this on the background evaluation is assumed to be negligible, and all sampling locations were therefore treated as independent in the statistical tests and calculations performed for this study. Treating the data points as independent is more conservative since the larger number of samples will result in narrower confidence intervals when comparing the background data to site data.

3.1.2 Data Filtering and Combining Rules

Results from both the 2005 BRC/TIMET (which includes the Environ dataset) and 2008 supplemental shallow soil background (this report) analytical datasets were validated. In order to prepare the datasets for statistical evaluation, results from each dataset were filtered down so that each background soil sample had one result per analyte and the two datasets were combined into

Deleted: *3*

one database. The following steps were taken to filter and combine the 2005 BRC/TIMET and 2008 Supplemental shallow soil background datasets into one database.

1) Filtered out all laboratory QC samples from both datasets

2) Filtered out all split sample results from both datasets; retained field duplicate results in the 2008 Supplemental shallow soil background dataset

3) Filtered out all rejected (R-qualified) data in both datasets

4) Aligned chemical names for both datasets so that names are exactly the same for each

5) Aligned units for both datasets so they are exactly the same for each

6) Filtered non-metals/non-radionuclides (*e.g.*, percent moisture) from both datasets

7) Filtered out all metals and radionuclides from the 2005 BRC/TIMET background dataset that were not included in the 2008 Supplemental shallow soil background dataset

8) Added fields to both datasets that include Dataset (2005 BRC/TIMET, 2008 Supplemental), Origin (McCullough, River, or Mixed), and Depth (0, 5, or 10)

9) Aligned field names for both datasets so they can be combined for statistical evaluation

10) Identified final subset of fields that will be required to conduct the data analyses

For direct comparison of the 2005 BRC/TIMET and 2008 Supplemental shallow soil background datasets, any chemical analyzed by one study but not the other was not considered in the comparison.

After filtering and prior to final combination of the two datasets, a comparison table was prepared. Table 2 shows the comparison of analyte lists and detection frequencies between the two datasets for metals and radionuclides.

Based on the information shown in Table 2, the following observations were made:

Deleted: December 2008

- The 2005 BRC/TIMET background dataset contains results for 42 metals and anions and 35 radionuclides; while the 2008 Supplemental dataset contains results for 38 metals and eight radionuclides.[6]

- The sample size for the 2005 BRC/TIMET background dataset is generally 120 results for each analyte (with a few exceptions); while the sample size for the 2008 Supplemental dataset is generally 33 results for each analyte.

- In cases where analyte results are available for both datasets, the detection frequencies were compared. As discussed in Section 2.2, detection frequencies were notably different for cadmium, lithium, mercury, selenium, silver, thallium, tin, zirconium, and uranium-233/234.

Deleted: The detection frequency differences were investigated by reviewing the analytical data to see if apparent differences in methodology or level of quality may have produced the differences. The same laboratory analyzed the metals for both background studies. The laboratories used for the radionuclide listed above are different between the background studies. In addition, the minimum detectable activity (MDA) for the identified radionuclides are calculated on a sample-by-sample basis and do not account for the difference in detection frequency. With the exception of cadmium and selenium, for which variations in reporting limits are likely to have contributed to the frequency of detections (see Section 3.4), no apparent analytical differences account for the difference in detection frequencies.¶

Formatted: Bullets and Numbering

3.1.3 Treatment of Data Qualified as Non-Detections

When radionuclides were not detected at activities greater than the minimum detectable activity (MDA), the laboratory reported the measured activity. Treatment of radionuclide data qualified as non-detections followed U.S. Department of Energy (DOE) guidance (DOE 1997), which states that, for radionuclide activity data:

> "All of the actual values, including those that are negative, should be included in the statistical analysis. Practices such as assigning a zero, a detect limit value, or some in-between value to the below-detectable data point, or discarding those data points can severely bias the resulting parameter estimates and should be avoided."

Therefore, for radionuclides, the reported activities (in pico Curies per gram [pCi/g]) were used without censoring to calculate all descriptive statistics (Tables 4 through 26), prepare plots (*e.g.*, boxplots), and conduct statistical analyses presented in this report.

For metals, a value of one-half the RDL was used as a replacement value for non-detected data for t-tests, parametric and nonparametric analysis of variance (ANOVA, Kruskal-Wallis tests), and calculation of parametric and nonparametric correlation coefficients. The ½-RDL substitution method was not applied to data analyzed using the WRS test because this test (as

[6] The following five inorganic constituents were included in the 2005 background investigation but were not included in the 2008 investigation: chloride, fluoride, nitrate, nitrite, and sulfate. Phosphorus was included in the 2008 investigation, but was not included in the 2005 analyte list. With NDEP concurrence, the project list of analytes was reduced in 2007 from 35 radionuclides to the following eight: uranium-238, uranium-233/234, thorium-230, and radium-226 (Uranium-238 Decay Chain), thorium-232, radium-228, and thorium-228 (Thorium-232 Decay Chain) and uranium-235/236 (Uranium-235 Decay Chain).

Deleted: 3

Deleted: December 2008

currently supported by GiSdT®) handles non-detected values using the Gehan ranking system (the Gehan test uses a modified ranking of sample results to accommodate non-detected values together with detected values), a method considered to be more robust than the ½-RDL substitution method. The GiSdT®'s WRS test uses the Mantel (1981) approach, which is equivalent to using the Gehan ranking system. The summary statistics (Tables 4 through 26) and plots (boxplots, individual value plots, and probability plots in Appendix D) incorporate the full RDL for non-detects.

It should be noted that the method detection limit (MDL) is established by the laboratories and represents the minimum concentration of a substance that can be measured and reported with 99 percent confidence that the analyte concentration is greater than zero. MDLs are established using matrices with little or no interfering species using reagent matrices and are considered the lowest possible reporting limit. Often, the MDL is represented as the instrument detection limit. The RDL (also known as the sample quantitation limit [SQL]) is defined as the MDL adjusted to reflect sample-specific actions, such as dilution or use of smaller aliquot sizes, and takes into account sample characteristics, sample preparation, and analytical adjustments. It represents the sample-specific detection limit and all non-detected results are reported to this level. Therefore, because the RDL is a sample-specific detection limit, for the dataset as a whole there may be instances where the maximum non-detect value may be higher than the lowest detected concentration, the median RDL for a chemical in a dataset is greater than the median detected concentration, or median RDL for non-detects are different for different datasets. It is recognized that these limitations may compromise statistical analyses in this report and potential future background comparisons.

Deleted: For metals, a value of one-half the reporting detection limit (RDL) was used as a replacement value for non-detected data for t-tests, parametric and nonparametric analysis of variance (ANOVA, Kruskal-Wallis tests), and calculation of parametric and nonparametric correlation coefficients. The ½-RDL substitution method was not applied to data analyzed using the WRS test because this test (as currently supported by GiSdT) handles non-detected values using a method considered to be more robust than the ½-RDL substitution method. The summary statistics (Tables 4 through 26) and plots (boxplots, individual value plots, and probability plots in Appendix D) incorporate the full RDL for non-detects.¶
<#>Identification and Treatment of Outliers¶
Outliers are data points that are extremely large or small relative to the rest of the data, and may not, therefore, be representative of the population sampled (USEPA 2000a). Outliers may be identified using statistical methods (*e.g.*, boxplots, probability plots, associations)—however, statistical methods alone should not be the basis for removing these data from the background dataset. Background soil samples were collected in known/suspected unimpacted areas. Accordingly, once outliers are identified using statistical methods, only a weight of evidence based on sound geochemical and other regional-specific knowledge should be used to remove them from the background dataset.¶
For this investigation, boxplots, individual value plots, and probability plots were used to identify outliers for further investigation. If the outlier could not be confirmed to be a transcription or other verifiable error, all statistical plots and tests were performed with the outlier included in the dataset. As shown on the boxplots in Appendix D, several o ... [1]

Formatted: Bullets and Numbering

Deleted: [7] no such errors were identified. The boxplots for each metal and radionuclide were reviewed to identify anomalously high outliers that may not be characteristic of background conditions. Anomalously high outliers within the 2008 dataset were identified as those points corresponding to detections (*i.e.*, ignoring non-detection report limit artifacts) on the boxplots that were higher than 1.5 times the interquartile ran ... [2]

Deleted: [8] Anomalously high outliers were identified using this criterion for the following constituents:¶
Arsenic ... [3]

Deleted: *3*

3.2

3.2 STATISTICAL PLOTS

Statistical plots are used in exploratory data analysis to show characteristics and relationships of the data, to evaluate fit to a normal distribution, to identify anomalous data points or outliers, and to provide a general overview of the data. Probability plots, boxplots, and individual value plots were constructed as part of the data evaluation for this investigation. Preliminary evaluation of the data included an assessment of data characteristics through graphical and quantitative analysis. The 2008 Supplemental data were summarized overall and by depth interval, with data plotted for the various groupings. The 2008 Supplemental data were compared with the 2005 BRC/TIMET background data using the probability plots, boxplots, and individual value plots. The graphical analysis of the analytical data is described in the following sections, and Appendix D contains the statistical plots.

Probability Plots. The distribution plots for each chemical include a probability plot that shows how well the dataset for the chemical fits a normal or lognormal distribution. Probability plots are also useful to visually identify outliers and to evaluate the possible presence of multiple populations within a dataset. Potential multiple populations are identified by inflection points on the probability plot. Inflection points are not defined statistically, and should be used with considerable caution.

The probability plots are graphs of values, ordered from lowest to highest and plotted against a standard normal or lognormal distribution function. The vertical axis is scaled in units of concentration (or activity, in the case of radionuclides), and the horizontal axis is scaled in units of the normal/lognormal distribution function. The vertical scale is plotted as a linear scale (concentration versus normal/lognormal quantile) and populations of data that plot as a straight line in a linear scale are referred to as normally distributed (or lognormally distributed).

Boxplots. Boxplots provide a method for comparing data groupings or datasets side by side. The boxplots simultaneously display the full range of data, as well as key summary statistics, such as the median, 25th and 75th percentiles, and minimum and maximum values. The top and bottom of the box are the 75th and 25th percentiles, respectively, of the dataset. The length from the top to the bottom of the box is the interquartile range; therefore, the box represents the middle 50 percent of the data. The width of the box is arbitrary. The horizontal line within the box depicts the median value (the 50th percentile) of the dataset. The upper and lower whiskers are defined as follows:

Upper whisker = 75^{th} percentile + (1.5 • interquartile range)

Lower whisker = 25^{th} percentile – (1.5 • interquartile range)

These plots show the symmetry of the dataset, the range of data, and a measure of central tendency (median).

The boxplots, which group data for each dataset, by chemical, and by depth interval, are provided along with the probability and individual value plots for each analyte in Appendix D

Deleted: December 2008

for the 2008 Supplemental dataset and the 2005 BRC/TIMET background dataset (including Environ dataset).

Probability and boxplots were used for identifying anomalous data points (outliers) and data clusters in the 2008 Supplemental and 2005 BRC/TIMET datasets. All anomalous data points and clusters were investigated further.

The plots shown in Appendix D summarize a large amount of data. The number of data points associated with each analyte is presented in Table 2. The plots are presented to provide a comprehensive overview of the 2008 Supplemental and 2005 BRC/TIMET background datasets for soils, to compare the 2008 Supplemental background dataset to the 2005 BRC/TIMET background dataset, and to compare the different depth intervals.

Deleted: (more than 6,700 records)

Scatterplots. A scatterplot uses a Cartesian coordinate system to display values for two variables from a dataset (*e.g.*, arsenic *vs.* aluminum concentrations for the 2008 dataset). The data are displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis.

Scatterplots were constructed for those constituent pairs with significant correlation coefficients. Scatterplots were visually examined and best professional judgment was used to ascertain whether high-concentration outliers[9] occur "near" the least-square linear trend line. Where high-concentration outliers occur "near" the trend line, one may infer that these concentrations are consistent with background concentrations.

Deleted: As directed by NDEP, where

3.3 DESCRIPTIVE SUMMARY STATISTICS

Descriptive summary statistics for metals and radionuclides were calculated for the 2008 Supplemental and 2005 BRC/TIMET datasets (Tables 4 through 26). Descriptive summary statistics for each of the two datasets were also prepared for the following depth intervals, structured around the sampling intervals employed for the 2005 shallow soil background sampling event and the 2008 supplemental shallow soil sampling event (Section 2.2):

- Surface soils (0 ft bgs);
- Shallow subsurface soils (5 ft bgs);

[9] High concentration outliers were identified from boxplots (see Section 3.4).

Deleted: -

Deleted: *3*

Deleted: December 2008

- Deeper subsurface soils (10 ft bgs);
- Subsurface combined (5-10 ft bgs); and
- All depths combined (0-10 ft bgs).

The descriptive summary statistics calculated for each analyte include the sample size, frequency of detections, and, for both censored and detected data, the minimum and maximum concentration, the median, the mean, and the 25th and 75th percentiles (quantiles).

3.4 IDENTIFICATION AND TREATMENT OF OUTLIERS

Statistical outliers are data points that are extremely large or small relative to the rest of the data, and may not, therefore, be representative of the population sampled (USEPA 2000a). Statistical outliers may be identified using statistical methods (*e.g.*, boxplots, probability plots, associations)—however, statistical methods alone should not be the basis for removing these data from the background dataset. Background soil samples were collected in known/suspected unimpacted areas. Accordingly, once statistical outliers are identified using statistical methods, only a weight of evidence based on sound geochemical and other regional-specific knowledge should be used to identify these data as "true" outliers and justify removing them from the background dataset.

For this investigation, boxplots, individual value plots, and probability plots were used to identify statistical outliers for further investigation. Outliers were further evaluated using correlation analyses and examination of scatterplots to further assess whether associations among these relatively few outlier data points were consistent with background concentrations (see Section 3.7.4). If the statistical outlier could not be confirmed to be a transcription or other verifiable error, all statistical plots and tests were performed with the statistical outlier included in the dataset.

As shown on the boxplots[10] in Appendix D, several statistical outliers were found in the dataset,[11] which is not unusual for a dataset of this size. Several of the outliers are artifacts of the

[10] Statistical outliers within the 2008 dataset were defined as those points corresponding to detected metal concentrations or radionuclide activities (*i.e.*, ignoring non-detection report limit artifacts) that were greater than 1.5 times the interquartile range for the (i) combined depth plots and (ii) individual depth plots, and are shown as an asterisk (*) on the boxplots (see Section 3.2).

[11] For several constituents (*e.g.*, beryllium), boxplots of the 2008 data identified outliers for the combined dataset (all depths combined), but outliers were not identified in the boxplots for individual depth intervals. In addition, in some cases (*e.g.*, calcium, 5 and 10 ft datasets), a given point that was considered an outlier for a given depth interval was not considered an outlier for the combined 2008 dataset (all depths combined) for that constituent.

Deleted: 3

Deleted: December 2008

RDLs. For example, for constituents with few detections, those detections are often classified as outliers on the boxplots because they are outside the typical range of detection limits. In addition, elevated RDLs are also classified as outliers in some cases. The probability plots for the constituents identified in Section 2.2 as "not being routinely detected" demonstrate the effect of the RDLs being substituted for non-detected values in the dataset; for those constituents (*i.e.*, antimony, boron, chromium (VI), lithium, mercury, niobium, platinum, selenium, silver, thallium, tin, tungsten, uranium-235/236, and zirconium), two distinct non-linear groupings of data are clearly visible in the probability plots. Other outliers occur sporadically; these outliers were reviewed to confirm that they were not the result of reporting errors;[12] no such errors were identified.

Overall, statistical outliers represent only a small proportion of the entire dataset. In addition, the lack of a consistent pattern related to statistical outliers would suggest that the data are not indicative of naturally occurring background conditions. Finally, the sample design for collection of the supplemental soil background data intentionally focused on suspected unimpacted areas. Given the lack of scientifically defensible reasons to consider these statistical outliers to be incongruous with background conditions (*i.e.*, "true" outliers), these data were considered representative of background and retained in the supplementary background soil dataset (see also Appendix E).

3.5 FREQUENCY OF DETECTION

As noted in Section 2.2, cadmium, silver, and uranium-233/234 were detected at noticeably higher frequencies in the 2008 supplemental shallow background samples than in those from the 2005 shallow background samples, and lithium, mercury, selenium, thallium, tin and zirconium were detected at noticeably lower frequencies in the 2008 deep samples than in the shallow background studies. The statistical summaries in Tables 4 through 26 were evaluated to assess the likely influence of RDLs on these observed detection frequencies. This evaluation determined that variations in RDLs are likely to have had effects on detection frequencies for certain constituents (*i.e.*, cadmium selenium, and silver), as summarized below.

Deleted: The descriptive summary statistics calculated for each analyte include the sample size, frequency of detections, and, for both censored and detected data, the minimum and maximum concentration, the median, the mean, and the 25th and 75th percentiles (quantiles). ¶
As noted above in Section 3.1.4, review of the statistical plots identified several outliers in the dataset. Of particular note were outliers associated with two sample locations and depth; the surface samples at locations BRC-BKG-R01 and BRC-BKG-R04. However, there is a lack of a consistent pattern to the data that would suggest that the data are not indicative of naturally occurring background conditions; therefore, it is inferred that the data are representative of background.
Barium concentrations are generally higher than in the 2005 BRC/TIMET dataset. It was noted in the 2005 BRC/TIMET report that barium generally had higher concentrations in the Environ data and from BRC/TIMET location BRC-BKG-12. These locations were classified as either 'River' or 'Mixed.' Thus, this supplemental dataset would seem to confirm that these higher barium levels are due to the different geology. Other high detect values occur sporadically both spatially and with depth.¶

Formatted: Bullets and Numbering

Deleted: 3

In these cases, the specific outlier was not considered anomalously high, and the representativeness of those values of background conditions was not questioned further.

12 Reporting or transcription errors are unlikely given the direct electronic data uploads from the laboratory, which were in turn uploaded directly into the spreadsheets used for statistical analysis, with no manual entry of concentration values.

Deleted: 3

Deleted: December 2008

Deleted: ¶

Cadmium	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection[13]	63.6%	13.3%
Median RDLs for Non-Detects (milligrams per kilogram [mg/kg])	0.04	0.1291
Median Detected Concentration (mg/kg)	0.11	0.105
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2005 cadmium FOD is appreciably lower than that for the 2008 data. The detected concentrations are comparable between the two datasets. The range of the 2008 detected values (0.053 to 0.26 mg/kg) is higher than the non-detect RDLs for that event (0.04 mg/kg); however, a large percentage of these data would not have been reported as detections under the higher 2005 RDLs (*i.e.*, the median value of 2008 detections was 0.11 mg/kg– less than the 2005 median RDL for non-detections [0.1291 mg/kg]). It therefore appears likely that the higher RDLs of the 2005 dataset are one cause of the lower frequency of detection in that dataset, although lower cadmium concentrations in the 2005 samples could be another explanation.	

Lithium	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	18.2%	100%
Median RDLs for Non-Detects (mg/kg)	7.314	--
Median Detected Concentration (mg/kg)	32.95	12.75
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2008 lithium FOD is appreciably lower than that for the 2005 data. The range of 2005 detections (7.5 to 26.5 mg/kg) is higher than a large percentage of the 2008 non-detect RDLs, based on the 7.314 mg/kg median 2008 RDL value, and many would have been reported as detections if present at those levels in the 2008 samples. This suggests that the 2008 samples may have generally lower lithium concentrations than the 2005 samples, despite the higher 2008 median detected concentration. However, the elevated 2008 RDLs (*i.e.*, 75th percentile of 14.628 mg/kg and beyond, which are higher than the majority of the 2005 detections [median detect 12.75 mg/kg]), complicate the analysis.	

[13] For all summary tables in this section, the value for Percent Detection reflects the full dataset for each event, as taken from Table 2, and the values provided for the other parameters were taken from Tables 4 and 9.

Deleted: 3

Deleted: December 2008

Mercury	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	0%	77.5%
Median RDLs for Non-Detects (mg/kg)	0.00668	0.0072
Median Detected Concentration (mg/kg)	- -	0.019
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2008 mercury FOD is appreciably lower than that of the 2005 data; the non-detect RDLs of the two events are fairly comparable. The range of 2005 detections (0.0084 to 0.11 mg/kg) is higher than the 2008 non-detect RDLs (0.00668 mg/kg), and would have been reported as detections if present at those levels in the 2008 samples. This suggests that the 2008 samples have generally lower mercury concentrations than the 2005 samples. Differences in RDLs do not appear to have caused the differences in the FODs in this case.	

Selenium	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	0%	43.3%
Median RDLs for Non-Detects (mg/kg)	0.32	0.1579
Median Detected Concentration (mg/kg)	- -	0.29
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2008 FOD for selenium is appreciably lower than for the 2005 data; the RDLs for the 2008 non-detects are about twice as high as those for the 2005 samples. A large percentage of the 2005 data detections (more than 50% based on median detect value 0.29 mg/kg), would not have been reported as detections under the higher 2008 RDLs (0.32 mg/kg). Therefore, it appears likely that the higher RDLs of the 2008 dataset are one cause of the lower frequency of detection in that dataset, although lower selenium concentrations in the 2008 samples could be another explanation.	

Silver	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	42.4%	13.3%
Median RDLs for Non-Detects (mg/kg)	0.11	0.2609
Median Detected Concentration (mg/kg)	0.076	0.0445
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2005 silver FOD is appreciably lower than that for the 2008 data; RDLs for the 2005 non-detects are more than twice as high as those for the 2008 samples. The range of 2008 detections (0.054 to 0.17 mg/kg) is lower than the 2005 non-detect RDLs (0.2609 mg/kg), and would not have been reported as detections if present at those levels in the 2005 samples. Therefore, it appears likely that	

Deleted: 3

the higher RDLs of the 2005 dataset are one cause of the lower FOD in that dataset, although lower silver concentrations in the 2005 samples could be another explanation.

Thallium	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	18.2%	35%
Median RDLs for Non-Detects (mg/kg)	0.3	0.5428
Median Detected Concentration (mg/kg)	0.46	1.1
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2008 thallium FOD is about 17% less than that for the 2005 data, RDLs for the 2008 non-detects are slightly lower than those for the 2005 samples. The majority of 2005 detections (1.1 mg/kg median value) are higher than the 2008 non-detect RDLs (0.3 mg/kg), and would have been reported as detections if present at those levels in the 2008 samples. This suggests that the 2008 samples have generally lower mercury concentrations than the 2005 samples. Differences in RDLs do not appear to have caused the differences in FODs in this case.	

Tin	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	48.5%	99%
Median RDLs for Non-Detects (mg/kg)	0.3	0.187
Median Detected Concentration (mg/kg)	0.43	0.49
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2008 tin FOD is appreciably less than that for the 2005 data; the non-detect RDLs for the 2008 data are nearly twice as high as those for the 2005 data. The majority of 2005 detections (0.4 mg/kg 1st quartile value) are higher than the 2008 non-detect RDLs (0.3 mg/kg), and would have been reported as detections if present at those levels in the 2008 samples. This suggests that the 2008 samples have generally lower tin concentrations than the 2005 samples. Differences in RDLs do not appear to have caused the differences in FODs in this case.	

Uranium-233/234	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	100%	50.8%
Median MDA for Non-Detects (pCi/g)	Not determined, because all results, including those lower than the MDA, were used in statistical analyses	
Median Detected Activity (pCi/g)	1.17	0.99

Deleted: December 2008

Assessment of MDA Effects on Frequency of Detection (FOD)	The 2005 shallow soil frequency of detection for uranium 233/234 is appreciably less than the frequency of detection of the 2008 data. The detected concentrations are comparable between the two datasets. Reported uranium 233/234 detections in both datasets are higher than the 2005 RDLs associated with non-detections. The assessment of RDL effects on the frequency of detection was not completely conclusive, but based on the above, it does not appear likely that the RDLs are contributing appreciably to the frequency of detection differences.

Zirconium	**2008 Supplemental Shallow Data**	**2005 BRC/TIMET Shallow Data**
Percent Detection	39.4%	100%
Mean RDLs for Non-Detects (mg/kg)	0.8	- -
Mean Detected Concentration (mg/kg)	11.5	125
Assessment of RDL Effects on Frequency of Detection (FOD)	The 2008 zirconium FOD is less than that of the 2005 data. The range of 2005 detections (60.1 to 179 mg/kg) is higher than the 2008 non-detect RDLs (0.8 mg/kg), and would have been reported as detections if present at those levels in the 2008 samples. This suggests that the 2008 samples have generally lower tin concentrations than the 2005 samples. Differences in RDLs do not appear to have caused the differences in FODs in this case.	

Deleted: 3

Datasets with high frequency of detects tend to be better suited to statistical analyses than those with low frequency of detects (*i.e.*, less than 50 percent), because detection limits in the latter tend to drive the analyses. The majority of the elements in this study have comparable frequency of detects near 100 percent, and statistical analyses were performed without concern for the effect of non-detections on the findings. For the other elements with far less than 100 percent frequency of detects, the frequency of detects tended to be comparably low in the two datasets; as discussed in the following section, statistical analyses considering the effects of non-detections were developed for these elements or were omitted altogether if the number of detections was too low. The eight metals discussed above represent the few cases in which frequency of detects were appreciably different between the two datasets; these are of particular concern in this study because this situation complicates statistical comparisons. As discussed above, BRC's evaluation of the associated RDLs and ranges of detected concentrations found that differences in RDLs did not appear to have caused the differences in frequency of detects, with the possible exception of cadmium, selenium, and silver, for which the evaluations were inconclusive. For these three metals, statistical comparisons may not be reliable between the two datasets, or in the future, between the background datasets and BMI Common Areas site data.

Deleted: %
Deleted: %
Deleted: %
Deleted: reporting limits
Deleted: reporting limits

Deleted: *3*

Deleted: December 2008

3.6 STATISTICAL METHODS

Statistical evaluations were used to infer whether metal concentrations and radionuclide activity in 2008 supplemental background soils were comparable to those in the 2005 BRC/TIMET background soils. The following procedures were conducted as part of the statistical evaluations:

Deleted: Findings of statistical methods

Deleted: Specifically, the

Deleted: lithological

- Data were organized by lithologic unit, constituent, and soil interval;
- Data were viewed using boxplots and scatterplots (Section 3.2);
- Data were characterized using descriptive statistics and tests of normality (Section 3.3 and 3.6);
- 2008 supplemental background data were compared to 2005 BRC/TIMET background data using two- and multiple independent sample tests (Sections 3.7.1 and 3.7.2);[14,15]
- 2008 supplement background data were tested to identify potential differences among 0 ft bgs, 5 ft bgs, and 10 ft bgs depth intervals using multiple independent sample tests (Sections 3.7.3); and
- Inter-element associations were identified using correlation analyses and used to further verify that samples were appropriate for characterizing background conditions (Section 3.7.4).

Deleted: 5

Deleted: 5.3

Deleted: 5.4

Deleted: 5.5

Deleted: 5.6

3.6.1 Hypothesis Testing

A common application of statistics is to test some scientific hypothesis. A statistical test examines a set of sample data and, based on the underlying distribution of the data, leads to a decision whether to (i) accept the hypothesis or (ii) reject the hypothesis and accept an alternative one. Accordingly, statistical hypotheses are framed in terms of a null hypothesis (H_o) and an alternative hypothesis (H_a).

Deleted: Statistical

[14] 2008 River dataset was compared to the 2005 McCullough, 2005 River, and 2005 Mixed datasets for the following soil intervals: (i) 0 ft bgs, (ii) 5 ft bgs, (iii) 10 ft bgs, (iv) 5-10 ft bgs combined, and (v) 0-10 ft bgs (0, 5, and 10 ft bgs depths combined).

[15] Tests of proportions and comparisons of detected-only data were used when two- and multiple independent sample tests were not recommended—i.e., when sample sizes were greater than four samples and frequency of detections were less than 50 percent.

Deleted:

Deleted:

Deleted: 4

Deleted: 3

When comparing the mean or median background concentrations for a constituent, the null hypothesis was that the mean/median background concentration for a specific constituent are comparable (*i.e.*, data populations/datasets are the same); therefore, the rejection of the null hypotheses results in the acceptance of the alternative hypothesis that the means/medians of the data populations/datasets are different.

When comparing the right-tails of two distributions, the null hypothesis was that larger values for background concentrations for a specific constituent are comparable; therefore, the rejection of the null hypotheses results in the acceptance of the alternative hypothesis that the two data populations/datasets are different with regard to larger values (*i.e.*, the values in the right-tail of one distribution are generally larger than the values in the right-tail of the other distribution).

When examining the relationship between the concentration of two constituents, the null hypothesis was that there is no correlation between two constituents (*i.e.*, no inter-element correlation); therefore, should this null hypothesis be rejected, one would accept the alternative hypothesis and infer that there exists a relationship (positive or negative) in concentrations between the two constituents. These hypotheses are also discussed in BRC/TIMET (2007) report.

3.6.2 Statistical Tests

Statistical tests were conducted to infer whether datasets are comparable and whether there exist relationships between two constituents. A key decision is whether a parametric or nonparametric statistical test is to be used. Parametric statistical tests used in this evaluation of supplement background concentrations assume the following:

- Samples are independent and drawn randomly from the population.
- Data are normally distributed for each population.

Nonparametric methods/tests are not dependent on a specific distribution (*e.g.*, normal distribution) for its validity (Singh and Singh 2007; Gilbert 1987; Sokal and Rohlf 1981; Zar 1984).[16] These methods do not require estimates of the population variance or mean. Nonparametric statistical tests assume that samples are independent and drawn randomly from the population.

[16] Accordingly, nonparametric tests are also known as distribution-free tests.
[18] Note a Gehan ranking is not supported by SPSS v.15 and was not used to accommodate non-detects in the Kruskal-Wallis and Kendall tau analyses.

Deleted:
Deleted: December 2008
Deleted:). For this study, one
Deleted: were derived from
Deleted: population
Deleted: provides evidence
Deleted: not the same. The
Deleted: elements
Deleted: could
Deleted: correlation
Deleted: (inter-element correlations)
Deleted: 2005
Deleted: Tests that
Deleted: specific mathematical form for the underlying distribution
Deleted: data are called nonparametric statistical tests.
Deleted: [17]
Deleted: *3*

Methods used to evaluate and compare the data groups for this supplemental background dataset are summarized below. The computer statistical software program GiSdT® (Neptune and Company 2007) was used to perform two-sample statistical comparisons. All parametric and nonparametric multiple independent sample comparisons and correlation analyses were performed using SPSS v. 15.[18] Consistent with previous studies of background concentrations at BRC, a level of significance (α) equal to 0.05 was used (BRC TIMET 2007).[19]

Two-Sample Tests

Statistical comparisons between the 2008 Supplemental dataset and the 2005 BRC/TIMET background dataset for each depth interval were performed using the Quantile test, Slippage test, the *t*-test, and the WRS test with Gehan modification. The Quantile test, Slippage test, and WRS test are non-parametric. That is, the tests are distribution free, thus an assumption of whether the data are normally or lognormally distributed is not necessary.

***t*-Test.** The *t*-test is a hypothesis test for two population means to determine whether they are significantly different. To conduct a two-sample *t*-test, the two populations must be independent; in other words, the observations from the first population must not have any bearing on the observations from the second population. Assumptions of the t-test are that both datasets are comprised of randomly sampled data, data are normally distributed, and datasets have equal variances[20] (Sokal and Rohlf 1981; Gilbert 1987; Zar 1984).

Wilcoxon Rank Sum (WRS). The WRS test performs a test for a difference between the sum of the ranks for two populations. This is a nonparametric method for assessing differences in the centers of the distributions that relies on the relative rankings of data values. Knowledge of the precise form of the population distributions is not necessary. The two underlying distributions are assumed to have approximately the same shape The WRS test has less power than the two-sample *t*-test when the data are normally distributed, but the assumptions are not as restrictive. The GiSdT® version of the WRS test uses the Mantel approach which is equivalent to using the Gehan ranking system.

Quantile Test. The Quantile test performs a test for a shift to the right in the right-tail of the site or tested population versus the reference population. This may be regarded as being equivalent to

[19] Where appropriate, a confidence level (1-α) of 95 percent confidence was used.
[20] Student *t*-test is used when datasets have equal variances. Welch's or Satterthwaite *t*-test may be applied when datasets have unequal variances.

Deleted: December 2008
Deleted: Guided Interactive Statistical Decision Tools (
Deleted: ®;
Deleted:),
Deleted: 2005
Deleted: Wilcoxon Rank Sum
Deleted: Wilcoxon Rank Sum
Deleted: An assumption
Deleted: is
Deleted: populations are approximately
Deleted: or the sample sizes are large (both at least 30)
Deleted: the populations
Deleted: Wilcoxon Rank Sum
Deleted: Wilcoxon Rank Sum
Deleted: Wilcoxon Rank Sum
Deleted:
Deleted: *3*

Deleted: December 2008

detecting if the values in the right-tail of the tested distribution are generally larger than the values in the right-tail of the reference distribution. This test assumes that the populations have approximately the same shape. The Quantile test is performed using a defined quantile = 0.80.

Slippage Test. The Slippage test looks for a shift to the right in the extreme right-tail of one population versus the extreme right-tail of a reference population. This is equivalent to asking if a set of the largest values of the tested distribution are significantly larger (in a statistical sense) than the maximum value of the reference distribution.

Deleted: Typically an alpha = 0.05 is used to evaluate a statistically significant result for the two-sample tests. Since several correlated tests are conducted, a lower alpha is selected. As more tests are performed, it is more likely that a statistically significant result will be obtained purely by chance. Given the use of multiple statistical tests, an alpha = 0.025 was selected as a reasonable significance level.¶

Multiple Independent Sample Tests

One-Way Analysis of Variance (ANOVA). The parametric one-way ANOVA tests the hypothesis that multiple (*k*) population means are equal (Sokal and Rohlf 1981; Gilbert 1987; Zar 1984). Where one-way ANOVAs indicated the existence of significant differences among soil strata, the Tukey Honestly Significant Difference (HSD) test was used to conduct pair-wise *post-hoc* comparisons.[21]

Kruskal-Wallis Test. Kruskal-Wallis test is a non-parametric one-way ANOVA for ranks and is used to test the equality of medians among multiple (*k*) populations. The Kruskal-Wallis test is used to test the null hypothesis that several populations have the same continuous distribution. If the null hypothesis is rejected, one may infer that measurements tend to be higher in one or more of the populations. Fundamentally, this test is analogous to a parametric one-way ANOVA with the exception that the measured/observed values are replaced by their ranks. Accordingly, it is an extension of the Wilcoxon-Mann-Whitney test for three or more groups. Where Kruskal-Wallis tests indicated the existence of significant differences among soil strata, examinations of boxplots were used to conduct pair-wise *post-hoc* comparisons.[22]

Deleted: tests

Examination of Constituents with Less than 50 Percent Frequency of Detection. When frequency of detections is less than 50 percent, even the nonparametric tests have little power to detect differences in central values (Smeti *et al*. 2007). For those constituents where the frequency of detection was less than 50 percent, two- or multiple independent sample tests were not conducted. The following approach was conducted:

Deleted: At the direction of NDEP, the

[21] Note that only *post-hoc* (= *a posteriori*) comparisons were conducted.

[22] SPSS v. 15 does not support the nonparametric Behrens-Fisher *post-hoc* comparison test.

Deleted:

Deleted: 3

Deleted: December 2008

1. For individual constituent datasets in which RDLs are comparable, a Z-test for two proportions[23] was conducted to identify similarities in datasets based on the proportion of detected concentrations.

Deleted: reporting limits

Deleted: conduct test of

For individual constituent datasets in which RDLs are comparable and RDLs are higher than detections, where the proportion of detected concentrations was found to be similar and the number of detected concentrations was greater than four for both datasets, two- or multiple independent sample tests were conducted on detected data only.

Deleted: reporting limits

Deleted: reporting limits

Deleted: is

Deleted: is

Deleted: (4)

Deleted: conduct two- or multiple independent sample tests on detected data only.¶
Similarities among datasets may be inferred based on results of

Deleted: comparing medians of

Deleted: -only

Deleted: (*i.e.*, when similarities are inferred between medians)

Note that for constituents with frequency of detections less than 50 percent and RDLs meeting analytical DQOs, one may conclude that these constituents are present at low concentrations in background soils.

Deleted: data quality objectives (

Deleted:)

Deleted: Moreover, it is recommended that characterizations of similarities/dissimilarities among background datasets be ascertained based on the more robust statistical analyses of constituents with greater frequency of detections.

Correlation Analysis

Correlations or "measures of association" are of interest because they offer another line of evidence to distinguish background and non-background data or multiple populations of data (BRC/TIMET 2007). Inter-element correlation analyses were conducted to identify those constituents that needed further examination (using scatterplots) to ensure that high concentration outliers were congruous with background concentrations.

Deleted: 2005

Pearson's Product-Moment Correlation Coefficient. The Pearson product-moment correlation coefficient (r) is a parametric measure of the correlation between two variables (Sokal and Rohlf 1981; Gilbert 1987; Zar 1984). Pearson's correlation reflects the degree of linear relationship between two variables and ranges from +1 to -1. A correlation of +1 means that there is a perfect positive linear relationship between variables. A correlation of -1 means that there is a perfect negative linear relationship between variables. A correlation of 0 means there is no linear relationship between the two variables.

Deleted: The

Kendall Tau Correlation Coefficient. The Kendall tau rank correlation coefficient (or Kendall tau coefficient) is a non-parametric statistic used to measure the degree of correspondence between the ranks of two populations—it measures the strength of association of cross tabulations. As with the Pearson's correlation coefficient, Kendall tau ranges from +1 to -1. A

[23] In this investigation, the Z-test for two proportions (http://www.dimensionresearch.com/resources/calculators/ztest.html) was used to test the null hypothesis that the proportion of detected concentrations is the same among two datasets. If the null hypothesis is rejected, one may infer that the two populations are different with respect to the proportion of detected data.

Deleted: non-parametric

Deleted: of

Deleted: *3*

Deleted: December 2008

value of +1 means that there is 100 percent positive association between the two variables—*i.e.*, rankings for both variables are identical. A value of -1 means that there is 100 percent negative association between the two variables—*i.e.*, the ranking of one variable is the reverse of the other variable. A value of zero indicates the absence of an association between the two variables—*i.e.*, rankings are independent.

3.6.2.1 Correction for Use of Multiple Tests

A Bonferroni correction concerns the question if, in the case of more than one test in a particular study, the level of significance (α) should be adjusted to account for random chance. As related to the supplemental shallow soil background investigation, the Bonferroni correction may be applied when a single hypothesis of no effect is tested using more than one test (*i.e.*, multiple tests for multiple constituents), and the hypothesis is rejected if one of the tests shows statistical significance. This adjustment is intended to correct for the probability of making a Type I error (*i.e.*, incorrectly concluding there exists a difference when, in fact, there is no difference among datasets) when multiple tests are used. Note that this adjustment for reducing the chance of making a Type I error will increase the probability of a Type II error—*i.e.*, incorrectly concluding there is no difference when, in fact, there is a difference datasets.

The Bonferroni correction is performed by dividing the level of significance (usually set to 0.05 by convention) by the number of tests performed. For the supplemental shallow soil background investigation study, 46 constituents were tested to determine if lithologic units and/or depth intervals are different. Accordingly, the Bonferroni correction would divide an alpha of 0.05 by 46, resulting in an alpha of 0.0011.

When comparing among background datasets, both types of error are relevant and are of interest. For the purposes of this study and to be consistent with previous studies of background concentrations at BRC (BRC TIMET 2007), a level of significance equal to 0.05 was used. The potential effects of a Bonferroni correction on the overall conclusions of the study are also discussed in appropriate sections of this report[24] to address potential consequences of making a Type I error to the overall conclusions.

[24] A review of tables in Appendix F indicate that the use of this correction would not have changed the overall conclusions of this study with regard to significant geochemical differences (i) among 0, 5, and 10 ft bgs depth intervals within the 2008 River background data (Table F-1), (ii) among the four lithologic units (Tables F-2 and F-3), and (iii) between 2008 River and 2005 McCullough by depth interval (Tables F-6 through F-8).

Deleted: 3

3.7 RESULTS OF STATISTICAL ANALYSES

A key objective of this investigation is to evaluate whether the supplemental shallow soil background dataset is statistically similar to or different to the 2005 BRC/TIMET background data. The results of the following statistical analyses are provided with the intention of supporting a weight-of-evidence evaluation as part of this investigation.

3.7.1 Comparison of 2008 Supplemental and 2005 BRC/TIMET Datasets (All Depths Combined)

The 2008 Supplemental and 2005 BRC/TIMET datasets were evaluated to determine if they may be combined into one dataset for future consideration. The results of the statistical analyses are included in Appendix F. Probability plots, boxplots, and individual value plots were used to compare the 2008 Supplemental and 2005 BRC/TIMET data. These plots are included in Appendix D. Overall, the samples for the 2005 BRC/TIMET background study appear to have captured a fair range of natural variability and heterogeneity (largely a consequence of the larger sample size); typically showing a wider range of concentrations/activities than samples from the 2008 Supplemental shallow soil background study. Because the 2005 BRC/TIMET background data spanned a broader geographic area and included 120 samples compared with 33 samples collected for the 2008 Supplemental shallow soil background study, this is not an unexpected outcome.

The 2008 dataset was compared to each of following lithologic units: 2005 McCullough, 2005 River, and 2005 Mixed datasets (Table F-2 of Appendix F). Consistent with the Shallow Background Study (BRC/TIMET 2007), if a given dataset had fewer than four detections, it was deemed to lack data sufficient to support a robust statistical analysis and was not included in the statistical comparisons. If no more than two datasets had greater than four detections, no statistical comparisons were performed for that constituent. Accordingly, statistical tests were not performed for chromium VI, niobium, platinum and tungsten—and it was not possible to determine whether significant differences were associated with the 2008 River and the three 2005 soil lithology datasets for these metals.

[27] Only when datasets have comparable detection limits can this analysis be performed as a line of evidence to infer differences between datasets; otherwise, the test will only reflect differences in detection limits.

Deleted: December 2008

Deleted: [25]

Deleted: [26]

Deleted: E

Deleted: semi-quantitatively

Deleted: One significant difference between the two datasets is arsenic, which showed generally higher concentrations in 2008 Supplemental samples. Boxplots show that concentrations of arsenic were greater for samples from most 2008 Supplemental locations than in samples from most 2005 BRC/TIMET locations. Other differences between the 2008 Supplemental and 2005 BRC/TIMET datasets include: ¶
1) The size of the dataset (*i.e.*, the BRC/TIMET dataset is considerably larger [n = 120] than the 2008 Supplemental dataset [n = 33]; ¶
2) The analyses conducted were not consistent in all cases between the two events (*e.g.*, 2008 Supplemental did not analyze samples for anions and others); and ¶
3) The reporting limits for some analytes (*e.g.*, antimony, boron, mercury, niobium, platinum, selenium, silver, and tungsten) differed between the two datasets.¶

Deleted: lithological

Deleted: E

Deleted: E

Deleted: 2005

Deleted: *3*

Deleted: December 2008

Overall, statistical comparisons indicated that a number of significant differences existed for 34 of 46 constituents among the four lithologic units: 2005 McCullough, 2005 River, 2005 Mixed, and 2008 River (Table F-2 of Appendix F):

Deleted: 2008 and the three 2005 datasets, with the possible exception of antimony and calcium (Table E

- Antimony
- Arsenic
- Barium
- Beryllium
- Boron
- Cobalt
- Copper
- Iron
- Lead
- Lithium
- Magnesium
- Mercury
- Molybdenum
- Nickel
- Palladium
- Phosphorus
- Potassium
- Silicon
- Silver
- Sodium
- Strontium
- Thallium
- Tin
- Titanium
- Uranium
- Vanadium
- Zirconium
- Radium 226
- Radium 228
- Thorium 228
- Thorium 230
- Thorium 232
- Uranium 233/234
- Uranium 238

The greatest number of significant differences was noted between 2005 McCullough and 2005 River datasets.

Differences between the 2008 River dataset and one of the 2005 datasets were identified for 14 constituents (Table F-2 of Appendix F):

- Arsenic
- Barium
- Boron
- Lithium
- Magnesium
- Palladium
- Potassium
- Silicon
- Sodium
- Strontium
- Zirconium
- Radium 228
- Thorium 230
- Uranium 233/234

Deleted: 3

With respect to the 2008 River dataset, a greater number of significant differences were noted between (a) 2008 River and 2005 McCullough and (b) 2008 River and 2005 Mixed datasets as compared to other inter-lithologic unit comparisons. As might be expected, the fewest number of significant differences were noted between the 2005 River and 2008 River datasets. Note that higher concentrations of arsenic in the 2008 River soils as compared to the 2005 River soils may be inferred from the Tukey HSD comparison results. For most constituents, the probability (*p*) values for the ANOVA/Kruskal-Wallis were less than 0.001 (Table F-2). Accordingly, the application of a Bonferroni correction to the significance level would not change the overall conclusions that differences exist among the four lithologic units and that the 2008 River dataset is significantly different than the three 2005 dataset for several constituents.

When the frequency of detections is less than 50 percent, even the nonparametric tests have little power to detect differences in central values (Smeti *et al*. 2007). For constituents with frequency of detects less than 50 percent and similar detection limits, a binomial proportions test was conducted to determine if frequency of detects between background datasets were comparable. Where frequency of detects were found to be similar, subsequent comparisons using detected-only data were conducted for infrequently detected constituents to identify potential similarities among background datasets.[27] Differences between the 2008 and the 2005 background datasets may also be inferred from these analyses (Table F-4 of Appendix F) and are summarized:

Constituent	Sample Size* (n > 4)	Z-Test for Two Proportions	Additional Analysis Candidate
Antimony	Yes	Similar frequency of detection	Yes
Boron	Yes	Similar frequency of detection	Yes
Silver	Yes	Dissimilar frequency of detection	No
Tin	Yes	Similar frequency of detection	Yes
Radium-228	Yes	Similar frequency of detection	Yes

* for two or more lithologic units

Comparisons of detected-only values between 2008 River and 2005 lithologic units were mixed for infrequently detected constituents—*i.e.*, differences may be inferred for some infrequently detected constituents; while no differences may be inferred for other infrequently detected constituents (Table F-9). Note that infrequently detected constituents are, by definition, characterized by a high proportion of censored data. Accordingly, it is both reasonable and defensible that study conclusions related to similarities/dissimilarities among background datasets consider the overall preponderance of the evidence from the more reliable statistical analyses associated with the majority of the 46 constituents with greater frequency of detects.

Deleted: December 2008
Deleted: E). More
Deleted:)
Deleted: lithological
Deleted: inferred
Deleted: are
Deleted: At the direction of NDEP, for constituents
Deleted: the Test of Proportion
Deleted: E
Deleted: E
Deleted: Constituent ... [4]
Deleted: Cadmium ... [5]
Deleted: Thallium ... [6]
Deleted: Uranium-233/234* ... [7]
Deleted: lithological
Deleted: E
Deleted: characterizations of
Deleted: be largely ascertained based on
Deleted: robust
Deleted: detections
Deleted: *3*

Basic Remediation COMPANY

All in all, from these statistical comparisons, it may be inferred that the 2008 River data differ with respect to metal concentrations and radionuclide activities to the 2005 lithologic units. These findings are consistent with the findings reported in the Shallow Background Study (BRC/TIMET 2007). Therefore, it is recommended that the 2008 Supplemental Background dataset not be pooled with the 2005 BRC/TIMET background dataset for future applications; however, this will be evaluated on a case-by-case basis.

3.7.2 Comparison of 2008 Supplemental and 2005 BRC/TIMET Datasets (Depth-Specific Evaluations)

The 2008 Supplemental and 2005 BRC/TIMET background soil datasets were also evaluated on a depth interval-specific basis to further evaluate potential similarities/dissimilarities. Accordingly, two-sample tests were performed to compare the 2008 River to the 2005 McCullough datasets for 0 ft bgs, 5 ft bgs, and 10 ft bgs depths intervals.[28] ANOVA/Kruskal-Wallis analyses were performed for the 5-10 ft bgs combined dataset for the 2008 River, 2005 McCullough, and 2005 Mixed datasets[29] (Table F-3). The results of the statistical analyses are included in Appendix F. Probability plots, boxplots, and individual value plots were used to semi-quantitatively compare the 2008 Supplemental and 2005 BRC/TIMET data. These plots are included in Appendix D.

Two Sample Test Results (individual 0, 5 & 10 ft bgs comparisons)

Consistent with the findings of statistical comparisons described in the prior section, a number of differences in metal concentrations were inferred based on statistical comparisons between the 2008 River and the 2005 McCullough datasets (Tables F-6, F-7, and F-8 in Appendix F):

- Arsenic (all depths)
- Barium (all depths)
- Beryllium (5 and 10 ft bgs)
- Lithium (10 ft bgs)
- Magnesium (0 and 10 ft bgs)
- Manganese (5 ft bgs)
- Silver (0 ft bgs)
- Sodium (all depths)
- Strontium (0 and 5 ft bgs)

[28] The sample size for constituents in the 2005 River and 2005 Mixed datasets for 0 ft bgs, 5 ft bgs and 10 ft bgs depth intervals were less than four (4) samples and were considered insufficient to support robust comparisons.
[29] The sample size for constituents in the 2005 River dataset (5-10 ft bgs combined depth interval) were less than four (4) samples and were considered insufficient to support robust comparisons.
[31] The ANOVA results for cobalt suggested that there were significant differences between lithologic units; however, the *post-hoc* testing did not identify specific differences.

Deleted: December 2008
Deleted:
Deleted: lithological
Deleted: ,
Deleted: differs
Deleted: lithological
Deleted: 2005
Deleted: datasets
Deleted: .
Deleted: [30]
Deleted: E
Deleted: E
Deleted: E
Deleted: E
Deleted: E
Deleted:
Deleted: *3*

Deleted: December 2008

- Boron (all depths)
- Cobalt (all depths)
- Copper (5 and 10 ft bgs)
- Iron (5 ft bgs)
- Lead (5 and 10 ft)
- Nickel (all depths)
- Palladium (0 and 5 ft bgs)
- Phosphorus (all depths)
- Potassium (all depths)
- Silicon (5 ft bgs)
- Tin (5 ft bgs)
- Titanium (all depths)
- Vanadium (0 and 5 ft)
- Zirconium (all depths)

Deleted: ¶
Lithium (10 ft bgs)

No differences in radionuclide activities were inferred based on the results of statistical comparisons for any of the three depth intervals (Tables F-6, F-7, and F-8 in Appendix F). For most constituents, the probability (*p*) value for at least one parametric or nonparametric two-sample tests is less than 0.001 (Tables E-6 through E-6). Accordingly, the application of a Bonferroni correction to the significance level would not change the overall conclusion that differences exist between 2008 River and 2005 McCullough on a depth interval basis.

Deleted: E
Deleted: E-
Deleted: E
Deleted:)

ANOVA/Kruskal-Wallis Test Results (5 - 10 ft bgs combined)

Consistent with the Shallow Background Study (BRC/TIMET 2007), the datasets for the 5 ft bgs and 10 ft bgs depth intervals within a lithologic unit were combined to produce a dataset for the 5-to-10 (5-10) ft bgs depth interval. Overall, a number of significant differences in metal concentrations among the three lithologic units (2008 River, 2005 McCullough, and 2005 Mixed) were identified for the 5-10 ft bgs depth interval based on the results of ANOVAs/Kruskal-Wallis tests (Table F-3 in Appendix F). The only constituents for which no significant differences were identified include:

Deleted: 2005
Deleted: lithological
Deleted: lithological
Deleted: E
Deleted: E

- Calcium
- Zinc
- Thorium-228
- Thorium-232

For most constituents, the probability (*p*) values for the ANOVA/Kruskal-Wallis tests were less than 0.001 (Table F-3). Accordingly, the application of a Bonferroni correction to the significance level would not change the overall conclusions that differences exist among the four lithologic units with respect to the 5-10 ft bgs depth interval.

Consistent with the Shallow Background Study (BRC/TIMET), no statistical tests were conducted for metals that had fewer than four detections in one or more of the unit-specific datasets, specifically:

Deleted: 3

- Antimony
- Boron
- Cadmium
- Chromium VI
- Mercury
- Niobium
- Platinum
- Selenium
- Silver
- Thallium
- Tungsten

Because these constituents were not subjected to statistical comparisons, it was not possible to determine whether significant differences were associated with the 5-10 ft bgs depth interval among the 2008 River, 2005 McCullough, and 2005 Mixed datasets.

Significant differences were noted between the 2008 River dataset and the datasets for the other two lithologic units (Table F-3 of Appendix F). More significant differences were identified between the 2008 River and 2005 McCullough datasets. However, differences in metal concentrations and radionuclide activities were inconsistent between the units—*i.e.*, one lithologic unit did not have consistently higher concentrations or activities. The 2005 Mixed dataset was nearly always indistinguishable from either one or both of the other two lithologic units. That is, for all elements except uranium-238, the 2005 Mixed dataset was (1) statistically indistinguishable from both the 2005 McCullough and the 2008 River datasets (*e.g.*, arsenic, lead); (2) statistically indistinguishable from the 2005 McCullough dataset but had inferred significant differences from the 2008 River dataset (*e.g.*, magnesium, manganese; or (3) statistically indistinguishable from the 2008 River dataset but had inferred significant differences from the 2005 McCullough dataset (*e.g.*, barium, tin) (Table F-3 of Appendix F)., This observation is consistent with the interpretation of the 2005 Mixed dataset being derived from soils that reflect a mixture of McCullough and River sediments. The 2005 Mixed dataset had significant differences inferred relative to the 2008 River dataset for several common parent elements (*e.g.*, silicon, aluminum, magnesium, potassium), which suggests a closer affinity between the Mixed and McCullough sediments.

The following constituents were considered to be present at higher concentrations in the 2008 River dataset than the other two datasets:

Deleted: December 2008
Deleted: lithological
Deleted: E
Deleted: E
Deleted: lithological
Deleted: lithological
Deleted: E
Deleted: E
Deleted: 3

- Arsenic
- Palladium
- Silicon
- Strontium
- Chromium
- Potassium
- Sodium
- Uranium

For infrequently detected constituents (less than 50 percent frequency of detection), differences between the 2008 River and the 2005 datasets may also be inferred from these analyses (Table F-5 of Appendix F) and are summarized:

Deleted: E
Deleted: E

Constituent	Sample Size* (n > 4)	Z-Test For Two Proportions	Additional Analysis Candidate
Antimony	Yes	Similar frequency of detection	Yes
Radium-226	Yes	Similar frequency of detection	Yes
Radium-228	Yes	Similar frequency of detection	Yes

* for two or more lithologic units

Deleted: Test of Proportion
Deleted: Boron ... [8]
Deleted: Uranium-233/234 ... [9]

Comparisons of detected-only values between 2008 River and 2005 lithologic units were mixed for infrequently detected constituents—*i.e.*, differences may be inferred for some infrequently detected constituents (antimony, boron); while no differences may be inferred for other infrequently detected constituents (radium-226, radium-228). Note that infrequently detected constituents are, by definition, characterized by a high proportion of censored data. Accordingly, it is both reasonable and defensible that study conclusions related to similarities/dissimilarities among background datasets consider the overall preponderance of the evidence from the more reliable statistical analyses for the vast majority of the 46 constituents with greater frequency of detects.

Deleted: lithological
Deleted: characterizations of
Deleted:
Deleted: be largely ascertained based on
Deleted: robust
Deleted: detections.

Again, when results of statistical comparisons are taken as a whole, it may be inferred that the 2008 River data differ with respect to metal concentrations and radionuclide activities to the 2005 lithologic units. These findings support the recommendation not to pool the 2008 Supplemental Background dataset with the 2005 BRC/TIMET background datasets for future applications.

Deleted: differs
Deleted: lithological

3.7.3 Comparison of 2008 Supplemental Shallow Data by Depth Intervals

Soil samples were collected from three depth intervals from the 2008 Supplemental background soil study: 0 ft bgs, 5 ft bgs, and 10 ft bgs. Data for samples from each depth interval were compared using the statistical tests identified in Section 3.6.2. Multiple population (ANOVA) tests were selected and used to compare data among surface, middle shallow, and deeper shallow

Deleted: 5

soil samples. The results of the statistical analyses are included in Appendix F. Results that are statistically significant at a p-level of 0.05 are indicated in each table (see Section 3.6.2.4 regarding correction for use of multiple tests). Boxplots and individual value plots shown in Appendix D compare the data by depth interval and offer a visual semi-quantitative appraisal of differences for each analyte among the groups of data. Statistical tests provide a quantitative analysis to determine if the differences are statistically significant at a specified significance level.

For the most part, metal concentrations were comparable among the three soil depth intervals (Table F-1 of Appendix F). Statistically significant differences in concentrations or activity among soil depth intervals were found for only seven of 46 constituents examined:

- Cobalt[31]
- Potassium
- Thorium-230
- Uranium-238
- Nickel
- Sodium
- Uranium-233/234

For most constituents, the probability (*p*) values for the ANOVA/Kruskal-Wallis tests were greater than 0.05 (Table F-1). Accordingly, the application of a Bonferroni correction to the significance level would not change the overall conclusions that few differences exist among the 0, 5, and 10 ft bgs depth interval for the 2008 supplemental shallow soil data (Table F-1). In fact, using a Bonferroni correction, differences for only two of 46 constituents would be statistically significant: concentrations of potassium and activities of uranium 233/234 (Table F-1).

The statistical comparisons found that statistically significant differences could be inferred primarily between (i) 0 ft bgs and 5 ft bgs and (ii) 0 ft bgs and 10 ft bgs for metals; no significant differences were inferred for metals between the 5 ft bgs and 10 ft bgs datasets. For radionuclides, comparisons found that statistically significant differences could be inferred primarily between the 0 ft bgs and 10 ft bgs datasets only. In addition to those apparent significant differences, only one other significant difference was inferred for radionuclides, for the thorium-230 5 ft bgs and 10 ft bgs datasets.

Differences in metal concentrations and radionuclide activities were inconsistent between the units—*i.e.*, one lithologic unit did not have consistently higher concentrations or activities. Sodium concentrations and radionuclide activities were found to be greater for the 10 ft bgs depth interval as compared to the other depth intervals. Nickel and potassium concentrations were found to be greater in the 0 ft bgs depth interval as compared to deeper intervals.

Deleted: December 2008

Although some identified statistically significant differences were observed for the above metals and radionuclides, these differences may not be significant from a geochemical perspective. Nonetheless, the findings of these statistical analyses suggest that the 0 ft bgs, 5 ft bgs, and 10 ft bgs depth intervals may be pooled and applied as a single dataset for future applications.

Deleted:

3.7.4 Inter-Element Correlations

In addition to statistical tests comparing background soils data among lithologic units and depth intervals, 2008 River data were evaluated with respect to inter-element correlations. Correlations or "measures of association" are of interest because they offer another line of evidence to distinguish background and non-background data or multiple populations of data (BRC/TIMET 2007). Correlation analyses[32] were conducted and used to identify those constituent pairs whose scatterplots should be examined to ascertain whether high-concentration outliers should be considered background. Both parametric (Pearson's product-moment) and nonparametric (Kendall tau) correlation coefficients are presented in correlation matrices (Appendix G). Note that statistically significant correlation coefficients (at a significance level of 0.05)[33] are indicated by bold font and are color-coded for parametric and nonparametric coefficients in each table. Scatterplots for constituents with significant correlation coefficients and high-concentration outliers are also presented in Appendix G.

Deleted: lithological
Deleted: 2005). All correlation analyses were performed using SPSS v. 15.
Deleted: F
Deleted: F.

Statistically significant associations were observed for several elements. The association of aluminum with trace metals was evaluated, and statistically significant associations were found for barium, beryllium, cadmium, cobalt, copper, iron, lead, manganese, nickel, phosphorus, potassium, silicon, silver, tin, titanium, uranium, vanadium, and zirconium (Table G-1 of Appendix G). Strong inter-element correlations are normally expected between alkaline and alkaline-earth metals (BRC/TIMET 2007)—for the supplemental background data, statistically significant correlation coefficients between alkaline and alkaline-earth metals ranged from 0.25 to 0.40 (Table G-3 of Appendix G). These associations may be useful in distinguishing soils derived from different source materials and in distinguishing site-related contamination from natural background. Statistically significant associations among uranium-238 decay chain radionuclides were also observed—correlation coefficients ranged from 0.32 to 0.54 (Table G-5 of Appendix G). Correlation among activities for radionuclides within the decay chain (parents

Deleted: F
Deleted: F
Deleted: 2005
Deleted: F
Deleted: F
Deleted: F
Deleted: F

[32] All correlation analyses were performed using SPSS v. 15.
[33] A Bonferroni correction was not applied to the correlation analyses because these analyses were used to identify constituents requiring further analysis and not for distinguishing between datasets using multiple tests.

Deleted:
Deleted: 3

Deleted: December 2008

and daughters) is anticipated, unless there are differences in geochemical behavior and mechanisms to separate the species (BRC TIMET 2007).

Deleted: 2005

Note that statistically significant associations were observed for several metals and radionuclides; however these statistical associations should also be evaluated based on known geochemical characteristics.

Scatterplots

In addition to the calculated inter-element correlations, scatterplots with regression lines provide a visual assessment of inter-element associations. Statistically significant associations and high-concentration outliers were identified for several elements within the 2008 dataset (Appendix G):

Deleted: F

- Aluminum
- Arsenic
- Barium
- Copper
- Lithium
- Nickel
- Palladium
- Silver
- Strontium

Scatterplots for identified constituent pairs were examined to determine whether high-concentration outliers are consistent with background (Appendix G)—*i.e.*, high-concentration outliers were "near" the linear least-square trend line. To identify potential deviations from trend lines, constituents listed above were plotted against constituents that were correlated and considered ubiquitous and relatively constant for identified lithologic units—*i.e.*, aluminum, iron, and magnesium. In general, no consistent and conspicuous deviations from least-square trend lines were observed for high concentration outliers.

Deleted: be

Deleted: F

Deleted: lithological

Certain inter-element relationships are expected on the basis of geochemical behavior and expected mineralogical associations. For example, alkaline metals (such as lithium, sodium, and potassium) and alkaline-earth metals (such as barium, calcium, and magnesium) can be expected to behave similarly in solution and may therefore be expected to show an association in certain environmental media. Other metals are found in association in common minerals and show correlations in soils containing these minerals (such as feldspars; metal oxides such as hematite, goethite and pyrolusite; and carbonate minerals such as calcite). These associations are useful in distinguishing soils derived from different source materials and in distinguishing site-related contamination from natural background.

The association of aluminum with trace metals was also evaluated. Trace metals such as chromium, cobalt, copper, nickel, and vanadium may occur as impurities in the common

Deleted: *3*

alumino-silicate family of minerals known as feldspars. Clays and other secondary aluminum minerals in soils may host sorption sites for trace metals, thereby associating these metals. In general, these associations are evident.

Scatterplots were also constructed for radionuclides within the thorium-232 and uranium-238 decay chains and are included in Appendix G. Species within the decay chains (parents and daughters) should show statistically significant correlations in most cases unless there are great differences in geochemical behavior and sufficient mechanisms to separate the species. The same generally holds true for radionuclides in the thorium-232 decay chain (radium-228 and thorium-228). In general, most of the radionuclides in the uranium-238 decay chain (radium-226, thorium-230, and uranium-233/234) did show significant associations.

Finally scatterplots were constructed for arsenic and other metals commonly found at high levels in the Upper Ponds (chromium, lead, manganese, and vanadium) as well as radium-226 to support the contention that the 2008 Supplemental dataset is representative of background. Some correlation betweens these elevated levels would be expected in the ponds given the depositional history of the site. In general, most of these contaminants did show varying degrees of visual correlation with arsenic, with the possible exception of manganese. If aerial deposition of wind-borne dusts from Site operations were occurring at the background locations, a similar pattern may be expected. However, these same metals and radium-226 did not show any correlation with arsenic in either the 2008 supplemental or 2005 BRC/TIMET background datasets. Although some correlation appears evident between arsenic and vanadium in the 2008 Supplemental dataset, this is primarily driven by their highest concentrations being found in the same sample (BRC-BKG-R09) in the subsurface (10 ft bgs); likely not a result of contamination from the site.

Deleted: December 2008

4.0 SUMMARY AND CONCLUSIONS

The purpose of the 2008 Supplemental shallow soil background study was to collect and analyze data for metals and radionuclides in background shallow soils that are representative of soils in geologic units not covered by the existing 2005 background shallow soil dataset (BRC/TIMET 2007). The objective of this report was to determine whether these data, which are assumed representative of another geology, may be added to the background data pool to accommodate background comparisons at portions of the Common Areas (*i.e.*, the Mohawk sub-area and portions of Parcel 4B).

Deleted: 2005
Deleted: can be used to supplement the existing
Deleted: soil dataset.

Soil sampling was conducted in April 2008. Samples were collected from 10 soil boring locations that represent the specific lithologies targeted by this supplemental shallow soil background sampling study and that extend the representative range of soils found in the vicinity of the Site. A total of 30 field and three duplicate soil samples were collected from the 10 borings for analysis. The data validation for the 2008 Supplemental dataset included 20 percent full validation and 100 percent partial validation. Results qualified as estimated based on the data validation are usable for the purposes of establishing background concentrations and for comparison to site-specific sample data. No soil sample results were rejected. One hundred percent of the dataset were validated as usable, indicating that the overall data collection objectives for the study were met. However, as noted in Section 3.5, for a few metals (*e.g.*, cadmium, selenium, and silver), variations in RDLs may have affected the frequency of detection and the validity/applicability of statistical analyses between the 2008 and 2005 background datasets as well as in comparisons of these data to future site data.

Deleted: 4
Deleted:
Deleted: reporting limits

Several statistical outliers were found in the dataset, which is a common, anticipated observation for a dataset of this size. Moreover, these potential outliers occur sporadically and there are no apparent geology-based causes for these outliers. Accordingly, these outliers were considered likely due to naturally occurring variability.

Deleted: The validated 2008 Supplemental data were evaluated statistically using
Deleted: plots, calculation of summary statistics, and statistical comparison tests. Several
Deleted: not unusual
Deleted: These
Deleted: They are

Based on sampling location characteristics information obtained from published documentation, site inspection, and sample collection, it is reasonable to conclude that the background samples collected as part of this investigation reflect background soil conditions that may be used to support assessments of soils at the Mohawk sub-area and Parcel 4B. As discussed in Section 2.4, SVOC analyses were used to assess the potential for impacts to the sampling locations from anthropogenic sources. SVOC detections in surface soil samples collected at the background sampling locations are limited to bis(2-ethylhexyl)phthalate, a common lab contaminant. Therefore, the SVOC data did not provide any evidence suggesting that use of the samples for

Deleted: It is reasonable to conclude that the background samples collected reflect background conditions for Site soils based
Deleted: . Data for SVOCs do not indicate
Deleted: locations are impacted by other

Deleted: *3*

Deleted: December 2008

characterizing background conditions would be inappropriate. The results of correlation analyses and scatterplots also corroborate the conclusion that this dataset is appropriate for use as a representative background soil dataset.

Deleted: The

Key findings from the analyses of the shallow background soils data include:

Based on the statistical analyses performed, there appear to be distinct differences between the populations associated with sediments derived primarily from the McCullough and River Mountains, and with sediments representing a mixture of both sources. It is therefore appropriate to perform comparisons of background to Site data using the subset of background data that most closely matches the geologic conditions of that part of the Site as follows:

Portion of Site	Applicable Background Dataset
Southeastern portion (*e.g.*, Mohawk)	2008 River dataset
Northeastern portion	2005 McCullough and Mixed datasets
Northwestern portion (*e.g.*, Western Hook)[34]	2005 McCullough dataset
Central portion	2005 McCullough and Mixed datasets

- Because statistical analyses suggest that the 2008 Supplemental and 2005 BRC/TIMET datasets exhibit a number of statistically significant differences, it is recommended not to combine these datasets in support of future comparisons to site data. Potential exceptions to this recommendation will be considered on a case-by-case basis—for example, for areas of the site that may occur at the interface of different geologic units (*e.g.*, Parcel 4B).

- Findings of the ANOVA/Kruskal-Wallis tests found few statistically significant differences among the 0, 5, and 10 ft bgs depth intervals for the 2008 River background data. This findings suggests that data for the 0, 5, and 10 ft bgs depth intervals may be pooled and applied as a single dataset, promoting more powerful statistical analyses for future assessments in support of decision-making.

- Because of the limited inferred differences in the depth-specific sample populations for the 2008 River unit, it is not necessary or appropriate to compare depth-specific Site data to the associated depth-specific background dataset.

Deleted: soil samples for the 2005 BRC/TIMET study have a fair range of natural variability

Deleted: heterogeneity; including as wide or wider range

Deleted: concentrations as those found for the 2008 Supplemental

Deleted: study. As such, the two datasets are not inconsistent with each other. Because the 2005 BRC/TIMET data span a broader geographic area and include 120 samples compared to 33 samples collected for the 2008 Supplemental study, this outcome is not unexpected. The results of this analysis applicable to future use of these datasets for determination of cleanup adequacy are

Deleted: <#>Based on the statistical analyses performed, there appear to be distinct differences between the populations associated with sediments derived primarily from the McCullough and River Mountains, and with sediments representing a mixture of both sources. It is therefore appropriate to perform comparisons of background to Site data using the subset of background data that most closely matches the geologic conditions of that part of the Site as follows:¶
Portion of Site ... [10]

Deleted: 3

Although the various background datasets are all contained within the project database, combining the background dataset by depth and/or lithology for subsequent comparison with Site data will be influenced by potential exposures at varying depth intervals and the location of a particular receptor – in other words, based on data usability and conceptual site model considerations.

These findings suggest that these data are appropriate for supporting future assessments and decision-making with respect to soils at sites within the Complex and Common Areas. Specific decisions regarding how best to use the background soils data for future Site-to-background comparisons will be made on a case-by-case basis in consultation with NDEP.

Deleted: Decisions on

Deleted: December 2008

5.0 REFERENCES

Basic Remediation Company (BRC). 2008. Supplemental Background Shallow Soil Sampling and Analysis Plan, BMI Complex and Common Areas Vicinity, Clark County, Nevada. March.

Basic Remediation Company (BRC), ERM, and MWH. 2008. BRC Field Sampling and Standard Operating Procedures. BMI Common Areas, Clark County, Nevada. December.

Basic Remediation Company (BRC) and Titanium Metals Corporation (TIMET). 2007. Background Shallow Soil Summary Report, BMI Complex and Common Areas Vicinity. March.

Basic Remediation Company (BRC) and ERM. 2008a. BRC Quality Assurance Project Plan. April.

Basic Remediation Company (BRC) and ERM. 2008b. Data Validation Summary Report: Supplemental Shallow Soil Background Sampling Event; April 2008 (Dataset 34b). June.

Bevans, H.E.; Lico, M.S.; Lawrence, S.J. 1998. Water Quality in the Las Vegas Valley Area and the Carson and Truckee River Basins, Nevada and California", 1992-96; updated March 19.

Geotechnical & Environmental Services, Inc. (GES). 2008. BRC Supplemental Shallow Background Data Set. April 24.

Gilbert, R.O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold Company, New York, NY.

Mantel, N. 1981. Calculation of Scores for the Wilcoxon Generalization Applicable to Data Subject to Arbitrary Censorship. Am. Stat 35(4): 244-.

Nevada Bureau of Mines and Geology (NBMG). 1980. Las Vegas SE Folio Geologic Map (1977) and the Geologic Map of the Henderson Quadrangle, Nevada.

Nevada Division of Environmental Protection (NDEP). 2008a. Supplemental Guidance for Assessing Data Usability for Environmental Investigations at the BMI Complex and Common Area in Henderson, Nevada.

Deleted: 2007
Deleted: operating
Deleted: August
Deleted: 2005
Deleted: CH2MHill. 2006. Technical Memorandum: Henderson Landfill Response Program Site Soils Criteria. June 2.¶
Environ. 2003. Risk Assessment for the Water Reclamation Facility Expansion Site, Henderson, Nevada. Volume II, Appendix E. October 15.¶

Deleted: 3

Deleted: December 2008

Nevada Division of Environmental Protection (NDEP). 2008b. Guidance on the Development of Summary Statistics Tables. December 10.

Singh, A. and A.K. Singh. 2007. ProUCL Version 4.0. Technical Guide. Prepared for the U.S. Environmental Protection Agency by Lockheed Martin Environmental Services.

Smeti, E.M., L.P. Kousouris, P.C. Tzoumerkas, and S.K. Golfinopoulos. 2007. Trend analysis and variability of microbiological parameters of the Mornos Reservoir. Proc. 10th Internat'l Conf. on Environ. Sci. Tech.

Sokal, R.R. and F.J. Rohlf. 1981. Biometry. Second Edition. W.H. Freeman and Company. San Francisco, CA.

SPSS, Inc. (SPSS). 2006. SPSS Base User's Guide. Chicago, IL.

U.S. Department of Agriculture (USDA). 1985. Soil Survey of Las Vegas Valley Area Nevada-Part of Clark County. Soil Conservation Service. July.

U.S. Department of Energy (DOE). 1997. Evaluation of Radiochemical Data Usability. Office of Environmental Management. Oak Ridge, Tennessee. April.

U.S. Department of Navy (Navy). 1999. Handbook for Statistical Analysis of Environmental Background Data.

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

U.S. Environmental Protection Agency (USEPA). 1992. Guidance for Data Usability in Risk Assessment. Part A. Office of Emergency and Remedial Response, Washington D.C. Publication 9285.7-09A. PB92-963356. April.

U.S. Environmental Protection Agency (USEPA). 1999. National Functional Guidelines for Organic Data Review. USEPA 540/R-99-008. OSWER 9240.1-05A-P. October.

U.S. Environmental Protection Agency (USEPA). 2001. National Functional Guidelines for Low-Concentration Organic Data Review. USEPA 540-R-00-006. OSWER 9240.1-34. June.

Deleted: U.S. Environmental Protection Agency (EPA). 2000a. Guidance for Data Quality Assessment: Practical Methods for Data Analysis. EPA QA/G-9. July. EPA/600/R-96/084.¶
U.S. Environmental Protection Agency (EPA). 2000b. Guidance for the Data Quality Objectives Process. EPA QA/G-4. August. EPA/600/R-96/055.¶

Deleted: 3

U.S. Environmental Protection Agency (EPA). 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, DC. September.

U.S. Environmental Protection Agency (USEPA). 2004. National Functional Guidelines for Inorganic Data Review. USEPA 540-R-04-004. OSWER 9240.1-45. October.

U.S. Environmental Protection Agency (EPA). 2006a. Data Quality Assessment: Statistical Methods for Practitioners. EPA QA/G-9S. Office of Environmental Information, Washington, DC. EPA/240/B-06/003. February.

U.S. Environmental Protection Agency (EPA). 2006b. Guidance on Systematic Planning Using the Data Quality Objectives Process. EPA QA/G-4. Office of Environmental Information, Washington, DC. EPA/240/B-06/001. February.

Zar, J.H. 1984. Biostatistical Analysis. Prentice-Hall, Inc. Englewood Cliffs, N.J.

For metals, a value of one-half the reporting detection limit (RDL) was used as a replacement value for non-detected data for t-tests, parametric and nonparametric analysis of variance (ANOVA, Kruskal-Wallis tests), and calculation of parametric and nonparametric correlation coefficients. The ½-RDL substitution method was not applied to data analyzed using the WRS test because this test (as currently supported by GiSdT) handles non-detected values using a method considered to be more robust than the ½-RDL substitution method. The summary statistics (Tables 4 through 26) and plots (boxplots, individual value plots, and probability plots in Appendix D) incorporate the full RDL for non-detects.

Identification and Treatment of Outliers

Outliers are data points that are extremely large or small relative to the rest of the data, and may not, therefore, be representative of the population sampled (USEPA 2000a). Outliers may be identified using statistical methods (*e.g.*, boxplots, probability plots, associations)—however, statistical methods alone should not be the basis for removing these data from the background dataset. Background soil samples were collected in known/suspected unimpacted areas. Accordingly, once outliers are identified using statistical methods, only a weight of evidence based on sound geochemical and other regional-specific knowledge should be used to remove them from the background dataset.

For this investigation, boxplots, individual value plots, and probability plots were used to identify outliers for further investigation. If the outlier could not be confirmed to be a transcription or other verifiable error, all statistical plots and tests were performed with the outlier included in the dataset. As shown on the boxplots in Appendix D, several outliers were found in the dataset, which is not unusual for a dataset of this size. The outliers shown on the boxplots (indicated with a * symbol) are defined as observations that are beyond the upper or lower whiskers; with the whiskers extending 1.5 box heights (also known as the interquartile range) from the bottom and top of the box within (see Section 3.2). Overall, outliers represent only a small proportion of the entire dataset.

Several of the outliers are artifacts of reporting limits. For example, for constituents with few detections, those detections are often classified as outliers on the boxplots because they are outside the typical range of detection limits. In addition, elevated reporting limits are also classified as outliers in some cases. The probability plots for the constituents identified in Section 2.2 as "not being routinely detected" demonstrate the effect of the

RDLs being substituted for non-detected values in the dataset; for those constituents (*i.e.*, antimony, boron, chromium (VI), lithium, mercury, niobium, platinum, selenium, silver, thallium, tin, tungsten, uranium 235/236, and zirconium), two distinct non-linear groupings of data are clearly visible in the probability plots.

OTHER OUTLIERS OCCUR SPORADICALLY; THESE OUTLIERS WERE REVIEWED TO CONFIRM THAT THEY WERE NOT THE RESULT OF REPORTING ERRORS;

Page 3-4: [2] Deleted	ERM	3/16/2009 5:40:00 PM

[1] NO SUCH ERRORS WERE IDENTIFIED. THE BOXPLOTS FOR EACH METAL AND RADIONUCLIDE WERE REVIEWED TO IDENTIFY ANOMALOUSLY HIGH OUTLIERS THAT MAY NOT BE CHARACTERISTIC OF BACKGROUND CONDITIONS. ANOMALOUSLY HIGH OUTLIERS WITHIN THE 2008 DATASET WERE IDENTIFIED AS THOSE POINTS CORRESPONDING TO DETECTIONS (*I.E.*, IGNORING NON-DETECTION REPORT LIMIT ARTIFACTS) ON THE BOXPLOTS THAT WERE HIGHER THAN 1.5 TIMES THE INTERQUARTILE RANGE FOR THE (I) COMBINED DEPTH PLOTS AND (II) INDIVIDUAL DEPTH PLOTS FOR THE 2008 DATA.

Page 3-4: [3] Deleted	ERM	3/16/2009 5:40:00 PM

[2] Anomalously high outliers were identified using this criterion for the following constituents:

---

[1] Reporting or transcription errors are unlikely given the direct electronic data uploads from the laboratory, which were in turn uploaded directly into the spreadsheets used for statistical analysis, with no manual entry of concentration values.

[2] For several constituents (e.g., beryllium), boxplots of the 2008 data identified outliers for the combined dataset (all depths combined), but outliers were not identified in the boxplots for individual depth intervals. In addition, in some cases (e.g., calcium, 5 and 10 ft datasets), a given point that was considered an outlier for a given depth interval was not considered an outlier for the combined 2008 dataset (all depths combined) for that constituent. In these cases, the specific outlier was not considered anomalously high, and the representativeness of those values of background conditions was not questioned further.

Arsenic	BRC-BKG-R02 (5 ft bgs) BRC-BKG-R09 (10 ft bgs)	Silicon	BRC-BKG-R10 (0 ft bgs)
Boron	BRC-BKG-R09 (10 ft bgs)	Sodium	BRC-BKG-R09 (0 ft bgs)
Cadmium	BRC-BKG-R01 (0 ft bgs) BRC-BKG-R10 (5 ft bgs) BRC-BKG-R09 (10 ft bgs)	Thallium	BRC-BKG-R04 (0 ft bgs)
Copper	BRC-BKG-R01 (0 ft bgs)	Tin	BRC-BKG-R01 (0 ft bgs)
Lead	BRC-BKG-R01 (0 ft bgs) BRC-BKG-R04 (0 ft bgs)	Uranium	BRC-BKG-R09 (10 ft bgs)
Magnesium	BRC-BKG-R09 (5 ft bgs)	Thorium-230	BRC-BKG-R08 (10 ft bgs)
Manganese	BRC-BKG-R04 (0 ft bgs) BRC-BKG-R02 (10 ft bgs)	Thorium-232	BRC-BKG-R04 (10 ft bgs)
Molybdenum	BRC-BKG-R01 (0 ft bgs)	Uranium-233/234	BRC-BKG-R08 (10 ft bgs)
Phosphorus	BRC-BKG-R09 (0 ft bgs)	Uranium-235/236	BRC-BKG-R01 (5 ft bgs)
		Uranium-238	BRC-BKG-R08 (10 ft bgs)

As seen above, several samples exhibit outliers for one or more constituents. However, no one sample is routinely anomalously high in a way that suggests the associated detections are not representative of background. That said, a few surface soil samples exhibited routinely elevated constituent concentrations relative to the other samples (*i.e.*, BRC-BKG-R01 and BRC-BKG-R04) as follows:

> The surface sample at location BRC-BKG-R01 had the highest detected value for several metals (aluminum, beryllium, cadmium, chromium, cobalt, copper, iron, lead, molybdenum, nickel, potassium, tin, titanium, and zinc), and in several instances it is the highest of either 2005 BRC/TIMET or 2008 Supplemental datasets (aluminum, cadmium, chromium, copper, iron, lead, molybdenum, potassium, and tin).
>
> The surface sample at location BRC-BKG-R04 also had high detect values for several metals (lead, manganese, potassium, and thallium).

As discussed in Section 3.5.6, these values were further evaluated using correlation analysis/scatter plots to evaluate whether they were true outliers. This analysis identified no true outliers. Furthermore, there is no consistent pattern to the data that would suggest that the data are not indicative of naturally occurring background conditions. Sample locations BRC-BKG-R01 and BRC-BKG-R04 are not adjacent to each other, and if aerial deposition of wind-borne dusts from Site operations were suspected, then higher levels of metals typically found in soils at the site; for example, arsenic and vanadium would be expected at the surface in these samples. However, this is not the case. As noted above, the highest arsenic concentrations are found in the subsurface (BRC-BKG-R02 at 5 ft bgs and BRC-BKG-R09 at 10 ft bgs).

The supplemental background sample locations are west of the River Mountains. Formations associated with these mountains contain volcanic intrusions that are known to contain elevated concentrations of naturally occurring arsenic (Bevans *et al.*, 1998). The supplemental background locations are geologically similar to the western and central portions of the Henderson Landfill (see Figure 2 for landfill location). The central portion of the landfill relates to the artificial fill area that covers the pediment and fan deposits of the River Mountains and further to the east the Horse Spring Formation (from CH2MHill 2006; approved by NDEP on August 7, 2006). The western portion relates to the uncovered areas of the pediment and fan deposits of the River Mountains and the modern wash deposits (CH2MHill 2006). Arsenic levels found in undisturbed areas from the western and central portions of the landfill ranged from 3.7 to 34 mg/kg. The two highest arsenic concentrations from the supplemental background dataset (sample location BRC-BKG-R02 at 5 ft bgs and sample location BRC-BKG-R09 at 10 ft bgs) are within this range. They are therefore likely due to naturally occurring variability.

Note that because the sample design for collection of the supplemental soil background data intentionally focused on suspected unimpacted areas, and in the absence of evidence to the contrary, the outliers are assumed to represent background conditions. Therefore, there are no scientifically defensible reasons to consider these samples to be incongruous with background conditions, and identified outliers were retained in the supplementary background soil dataset. At the direction of NDEP, outliers were further evaluated using correlation analyses and examination of scatterplots to further assess whether associations among these relatively few outlier data points were consistent with background concentrations (see Section 3.5.6).

Page 3-21: [4] Deleted **ERM** **3/16/2009 5:40:00 PM**

Constituent	Sample Size* (n > 4)	Test of Proportion	Additional Analysis Candidate

Page 3-21: [5] Deleted ERM 3/16/2009 5:40:00 PM

Cadmium	Yes	Similar frequency of detection	Yes
Lithium	Yes	Dissimilar frequency of detection	No

Page 3-21: [6] Deleted ERM 3/16/2009 5:40:00 PM

Thallium	Yes	Similar frequency of detection	Yes

Page 3-21: [7] Deleted ERM 3/16/2009 5:40:00 PM

Uranium-233/234*	Yes	Similar frequency of detection	Yes
Uranium-235/236	Yes	Similar frequency of detection	Yes

* for two or more lithological units

Page 3-25: [8] Deleted ERM 3/16/2009 5:40:00 PM

Boron	Yes	Similar frequency of detection	Yes
Lithium	Yes	Dissimilar frequency of detection	No
Tin	Yes	Dissimilar frequency of detection	No
Zirconium	Yes	Dissimilar frequency of detection	No

Page 3-25: [9] Deleted ERM 3/16/2009 5:40:00 PM

Uranium-233/234	Yes	Dissimilar frequency of detection	No
Uranium-235/236	Yes	Similar frequency of detection	Yes

* for two or more lithological units

Page 4-2: [10] Deleted ERM 3/16/2009 5:40:00 PM

Based on the statistical analyses performed, there appear to be distinct differences between the populations associated with sediments derived primarily from the McCullough and River Mountains, and with sediments representing a mixture of both sources. It is therefore appropriate to perform comparisons of background to Site data using the subset of background data that most closely matches the geologic conditions of that part of the Site as follows:

<u>**Portion of Site**</u>	<u>**Applicable Background Dataset**</u>
Southeastern portion (*e.g.*, Mohawk)	2005 River dataset
Northeastern portion	2008 River dataset

Western portion (*e.g.*, Western Hook)	2005 McCullough dataset
Central portion	2005 McCullough and Mixed datasets