(2") 53953

APPLICATION OF LANL BACKGROUND DATA TO ER PROJECT DECISION-MAKING PART I: INORGANICS

Prepared by
Randall Ryti, Patrick Longmire and Eric McDonald for the
Los Alamos National Laboratory
Environmental Restoration Decision Support Council

March 29, 1996

LA-UR-96-

Received by ER-RPF
APR 19 1996
Dic

10448

INTRODUCTION

The purpose of this policy paper is to provide guidance to Los Alamos National Laboratory (LANL or the Laboratory) Environmental Restoration (ER) Project personnel on the Laboratory's approach to conducting background comparisons for inorganics. Part II, presented as a separate document, will present the background comparison approach for radionuclides.

The background comparison approach consists of two steps. The first step is the assembly of a defensible set of background data. This document summarizes the Laboratory-wide set of background analytical data from samples of soils, sediment and tuff collected by Longmire and others (Longmire et al. 1995, 1266), and presents a simple decision logic to select geologically defensible subsets of these data. All ER Project reports that evaluate differences from background will justify the use of Laboratory-wide background concentration data or present the rationale for using site-specific background concentration data. The second step is the selection of the statistical method(s) used to compare site data with background data. Two statistical methods are presented. The first compares the site concentration data with a statistic representing the upper percentile of background concentrations. The second is a group of methods designed to detect a distributional shift between site data and background data. Although guidelines for the application of these methods are presented in this document, each ER Project report that includes background comparisons should briefly describe the statistical analysis method chosen and justify its application to the data in question. Lastly, particular attention should be paid to background comparisons of arsenic, beryllium and manganese, because natural background concentrations of these elements exceed EPA risk-based screening values.

SUMMARY OF REGULATIONS AND GUIDANCE GOVERNING STATISTICAL COMPARISONS TO BACKGROUND

The Environmental Protection Agency (EPA) guidance documents supporting the Comprehensive Environmental Response Compensation and Liability Act (CERCLA) and Resource Conservation Recovery Act (RCRA) programs provide specific information on how to design background studies and how to statistically compare site data with background data.

The CERCLA document, *Guidance on Data Useability in Risk Assessment (Part A)* (EPA 1992, 1166), recommends collecting background data prior to collecting site data. If the comparison of background data with site-derived data for a given constituent does not show a difference statistically, that constituent is eliminated from further evaluation. The CERCLA guidance also suggests that the number of background samples collected from a site be based on the "minimum detectable difference" procedure (EPA 1989, 0303). Data analysts unfamiliar with this approach should contact the statistical specialists within the Decision Support Council.

Background comparisons for groundwater monitoring data are addressed in the RCRA document, *The RFI Guidance* (EPA 1989, 0088). Methods for comparing data derived from upgradient wells with data from downgradient wells are presented in the RCRA groundwater statistical analysis document (EPA 1989, 1141). These statistical methods are codified in 40 CFR Part 264, *Statistical Methods for Evaluating Ground-Water Monitoring from Hazardous Waste Facilities: Final Rule* Federal Register Tues. Oct. 11, 1988.

Statistical methods used for background comparisons of groundwater can be applied to background comparisons for data from other media as stated in the preface of the RCRA groundwater statistical analysis document (EPA 1989, 1141):

> "This scenario can be applied to other non-RCRA situations involving the same spatial relationships and the same null hypothesis. The explicit null hypothesis for testing contrasts between means, or where appropriate between medians, is that the means between groups (here monitoring wells) are equal (i.e., no release has been detected), or that the group means are below a prespecified action level (e.g., the ground-water protection standard). Statistical methods that can be used to evaluate these conditions are described in Section 5.2 (Analysis of Variance), 5.3 (Tolerance Intervals), and 5.4 (Prediction Intervals)."

The RCRA groundwater monitoring guidance states that the specific approach proposed by the owner/operator should be submitted to EPA for approval, especially where methods other than those presented in the guidance are used. Statistical methods presented below are consistent with those found in the analysis of variance and tolerance interval sections of the RCRA groundwater statistical analysis document (EPA 1989, 1141).

LABORATORY BACKGROUND COMPARISON APPROACH

The Laboratory background comparison approach is summarized in Figures 1 and 2. Figure 1 shows the overall background comparison strategy that supports ER Project decision-making. The decision objectives for the background comparison are an important factor in selecting both the appropriate subset of Laboratory background data and the statistical method used to make the comparison.

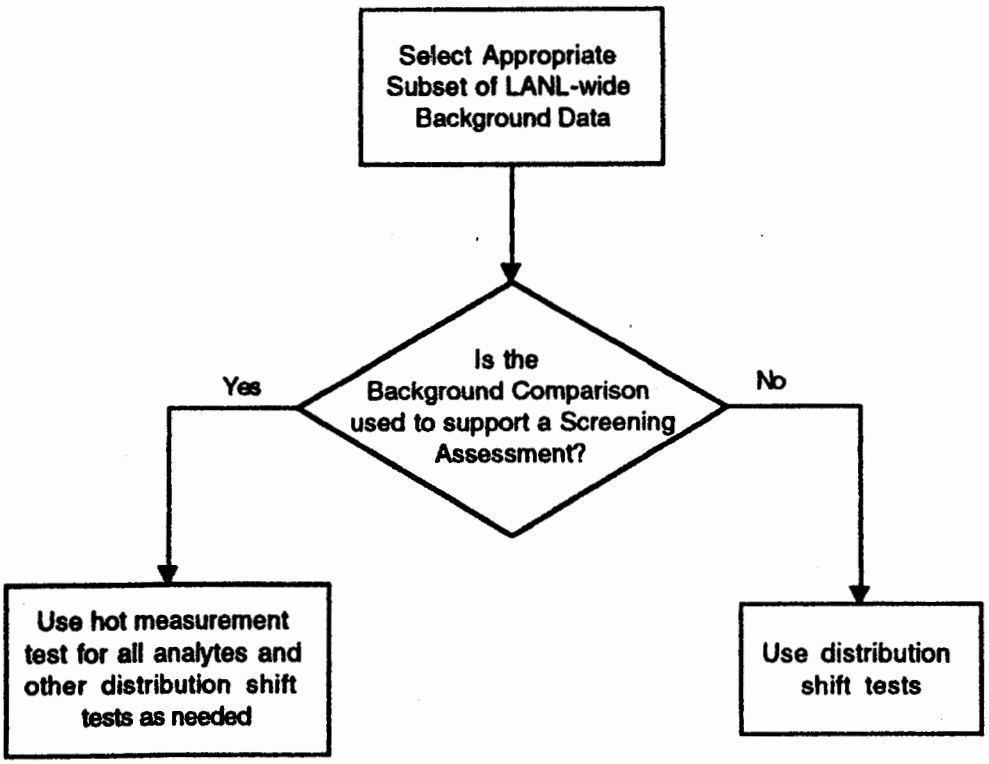


Figure 1. BACKGROUND COMPARISON STRATEGY.

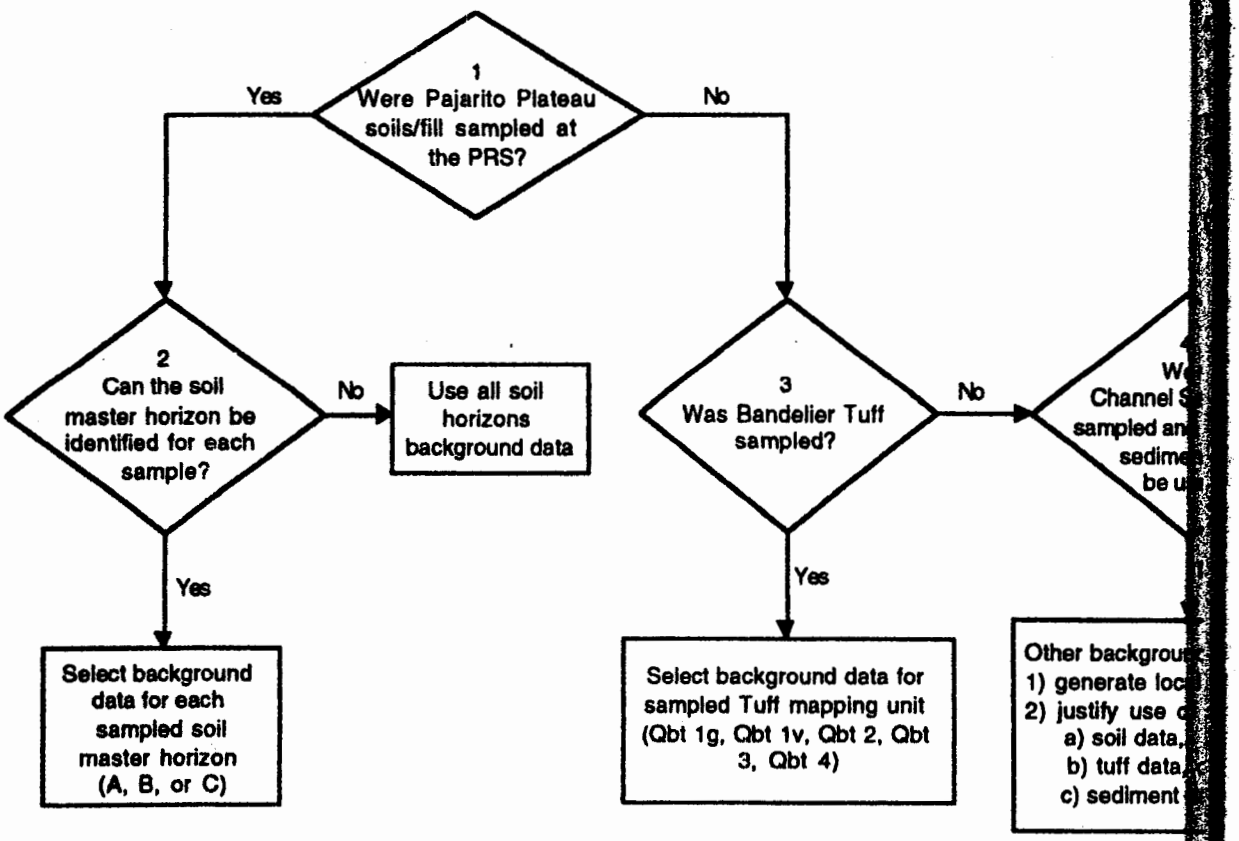


Figure 2. BACKGROUND DATA SELECTION PROCESS.

Selecting Background Data

Selection of the appropriate Laboratory background data set(s) for making statistical background comparisons is essential for PRS decision-making. PRS decisions are ultimately based on samples collected from a number of primary geomorphic units, including: mesa top, hill or canyon slope, and canyon bottoms. In addition there are subdivisions within the primary geomorphic units. For example, the geomorphic unit designated as canyon slopes is a mixture of mesa top soils and Bandelier Tuff. Although not inclusive of all Laboratory geomorphic units, existing Laboratory-wide background data do include samples of mesa top soils, Bandelier Tuff, and preliminary data on channel sediments. The purpose of this section is to guide the selection of appropriate subsets of these background data.

To support RFI screening assessments or other data analyses, background data is often selected after characterization samples have been collected. However, background data selection should also be considered in planning for sampling. The planning team should consider what quality of background data are needed to meet their specific sampling objectives. Whether considered before or after data collection, the basic considerations are similar and members of the Decision Support (chemistry, risk analysis, and statistics) and Earth Science (pedology, geology, geochemistry, geomorphology, stratigraphy) Councils should be consulted to provide guidance on the selection and uses of background data.

The process for selecting the most appropriate Laboratory background data set is summarized in Figure 2. In addition to the considerations discussed below, it is essential that comparable sample preparation and analytical methods be used for background and PRS samples. For example, XRF (X-Ray Florescence) may be comparable to whole sample analysis (e.g., hydrofluoric acid digestion) for some inorganic analytes (e.g., iron), but not for other inorganic analytes (e.g., barium).

A discussion of the decision points shown in Figure 2 follows.

Decision 1. Were Pajarito Plateau soils[1] and/or fill material sampled at the potential release site (PRS)?

"Yes" Decision. Soils across the Laboratory are highly variable spatially and in complexity due to the complicated history of landscape evolution and variation in ages of soil parent material. Most PRSs on mesa tops and within canyon bottoms consist of a mixture of native soils and fill material. The amount of fill material can vary (0 to 100%). Fill material typically consists of disturbed soils with crushed Bandelier Tuff, but other rock types may be also be present. Soil consists of layers or horizons of mineral and/or organic matter of variable thickness that parallel the land surface and differ from their parent material in morphological, physical, chemical, and mineralogical properties and in biological characteristics. Soil horizons are identified by a master horizon designation (see Appendix for a more detailed discussion). Most soil at the Laboratory consists of A, B or C master horizons.

"No" Decision. If no Pajarito Plateau soils and/or fill material were sampled, move on to decision 3, which pertains to Bandelier Tuff.

Decision 2. Can a soil master horizon be identified for each sample?

"Yes" Decision. When PRS samples are not representative of all three Laboratory soil master horizons, (A, B, and C), horizon-specific background data is the most applicable for statistical background comparisons. Horizon-specific information is useful where variability between horizons is significant relative to sampling objectives. For example, a sampling objective may be to determine the volume of soil above a risk-based threshold for beryllium. In this case, differences in beryllium concentrations between the A and B horizons may be greater than a risk-based cleanup threshold (residential scenario). Therefore, determining the volume elevated above background would determine the effect above background given specific soil horizons.

"No" Decision. When sites have no well-defined soil horizons (such as sites with fill material) or when Field Unit personnel have determined that variability between soil horizons was not relevant to the sampling objectives, the Laboratory-wide soils background data are the appropriate data set for statistical background comparisons. An example of soil horizon variability not relevant to sampling objectives would be a sampling plan designed to estimate

[1] Most Laboratory PRSs are located on the Pajarito Plateau. Therefore, Pajarito Soil samples form the bulk of the soil samples included in the Laboratory-wide background soil database. One exception is Fenton Hill, which is located in the Jemez Mountains.

the volume of barium above cleanup levels. Such a sampling plan would not require horizon specific background data because the cleanup level is at least ten times greater than soil background levels.

Decision 3. Was Bandelier Tuff sampled?

"Yes" Decision. Use the appropriate background data set for specific rock units of the Bandelier Tuff (Longmire et al. 1995, 1266). Bandelier Tuff (Tshirege Member) consists of four rock units (Longmire et al. 1995, 1266) that can be identified in the field by mapping and/or by evaluating core samples. These include in ascending order: Units 1 (Qbt 1g and Qbt 1v), 2 (Qbt 2), 3 (Qbt 3), and 4 (Qbt 4). Inorganic background data for Qbt 1g, Qbt 1v, Qbt 2, Qbt 3, and Qbt 4 are available (Longmire et al. 1995, 1266; Broxton et al. 1996, 1305). Readers should note that limited background data are available for other stratigraphic units (Cerro Toledo, Otowi member, and Tschicoma), these data are summarized in the ER Project background report *Natural Background Geochemistry and Statistical Analysis of Selected Soil Profiles, Sediments, and Bandelier Tuff, Los Alamos, New Mexico* (Longmire et al. 1995, 1266).

"No" Decision. If the Bandelier Tuff was not sampled, move on to decision 4 that addresses Laboratory background channel sediment data.

Decision Number 4. Were channel sediments sampled and can Laboratory sediment data be used?

"Yes" Decision. If channel sediments were sampled and the Laboratory sediment data can be used, the Laboratory sediment background data are the appropriate data set for statistical comparisons. An initial background data set for channel sediments has been provided by Longmire and others (Longmire et al. 1995, 1266) for Ancho Canyon and Indio Canyon. Channel sediments within the two canyons are derived entirely from Bandelier Tuff. Currently, there are no channel sediment data for canyons with sediments derived from the Tschicoma Formation (e.g., Los Alamos and Guaje Canyons). These preliminary data will be supplemented by additional samples collected by the ER Project on an as needed basis.

"No" Decision. A "no" decision indicates that none of the existing subsets of Laboratory-wide background data (soil, Bandelier Tuff, and channel sediments) are obviously applicable. Other background data options should be considered, including: evaluating background data

needs relative to sampling objectives, evaluation of data through interelement correlations, or generating site-specific (local) background.

The risk management and decision objectives for the background comparison are equally important during planning for background data needs and during the post sampling background comparison. The data analyst or planning team must determine if the variation between subsets of the Laboratory-wide background is significant relative to the decision to be made with the data. For example, Laboratory-wide soil background data may be adequate to support a screening assessment at a site having no history of inorganics release. Whereas, a screening assessment for a firing site, where significant quantities of inorganics would have been released, would require use of the appropriate subset of Laboratory-wide data.

One way to justify the use of Laboratory-wide background data is to evaluate the data through interelement correlations. Typically, there are significant correlations between major (aluminum, iron, and potassium) and trace elements (arsenic, beryllium, copper, nickel, vanadium, and zinc). The correlations are presented and the geochemical basis is detailed in *Natural Background Geochemistry and Statistical Analysis of Selected Soil Profiles, Sediments, and Bandelier Tuff, Los Alamos, New Mexico* (Longmire et al 1995, 1266). For most inorganic chemicals, these strong correlations result in a consistent ratio of trace to major elements. A significantly elevated ratio of a given trace to a major element can be used to document a release of that trace element. Bivariate plots of trace elements to major elements are one way to visually display the ratios for background and PRS data. An example data display is presented in Figure 3. This plot shows the bivariate relationship between beryllium and iron for TA-10 surface samples and Laboratory-wide soil background data. Another example is the strong correlation between concentrations of thorium and uranium in the Bandelier Tuff, presented in Figure 4. The bivariate plot shows that each rock unit has similar ratios of thorium to uranium (the uranium concentration is roughly 30% of the thorium concentration).

Generating appropriate subsets of background data can be performed very cost-effectively by using interelement correlations to statistically subsample Laboratory-wide data to create a conditional set of site-specific background data. At a minimum, this statistical subsampling requires that the concentration of one or more of the major inorganic elements (aluminum, iron, or potassium) can be shown, through archival information, to have never been released at a PRS, and that other inorganics are highly correlated to at least one major element. The

concentration range and statistical distribution of the major element results at a PRS are used to subsample the expected concentration of a trace element in the Laboratory-wide background data. For example, if a PRS had uniform concentration of iron between 5000 and 10000 mg/kg, the expected range of beryllium concentrations would be predicted to be between 0.3 and 1.1 mg/kg. PRS beryllium concentrations greater than 1.1 mg/kg would be outside the range of a statistically-based subsample of the Laboratory-wide data. This approach is more completely discussed in Campbell (1994, 1294), and data analysts unfamiliar with this statistical subsampling approach should contact the Decision Support Council for more information.

If site-specific background data are needed, statistical guidance can be used to help determine an appropriate number of background samples. One such approach, the minimum detectable difference procedure (EPA 1989, 0303), is mentioned in the Regulatory Background section of this paper (p. 2). This procedure requires three types of input: 1) the difference between the mean concentration of site data and background data that is desired to be detected (e.g., 50% of the background mean), 2) the desired probability of detecting that difference (e.g., 20%), and 3) the expected variability in the concentration data (usually expressed as the relative variability or coefficient of variation, e.g., 100% is typical). Given these inputs, 20 samples per background media are typically considered adequate for making background comparisons. As stated above, it is critical that consistent sample digestion and analytical methods are used for the background data and the PRS data. Before collecting site-specific background data, the potential use of the existing Laboratory data should be fully explored and advice on a recommended background data collection design, including sample digestion and analysis procedures should be sought from subject matter experts on the Decision Support and Earth Science Councils.

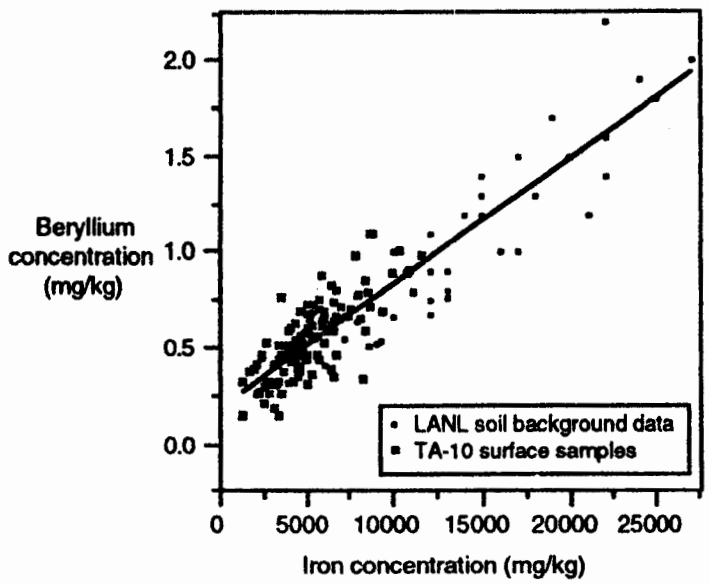


Figure 3. **BIVARIATE PLOT OF BERYLLIUM AND IRON (NITRIC ACID FRACTION) FROM THE LABORATORY SOIL BACKGROUND DATA AND TA-10 SURFACE SAMPLES. CORRELATION COEFFICIENT IS 0.916 FOR 174 BACKGROUND SAMPLES.**

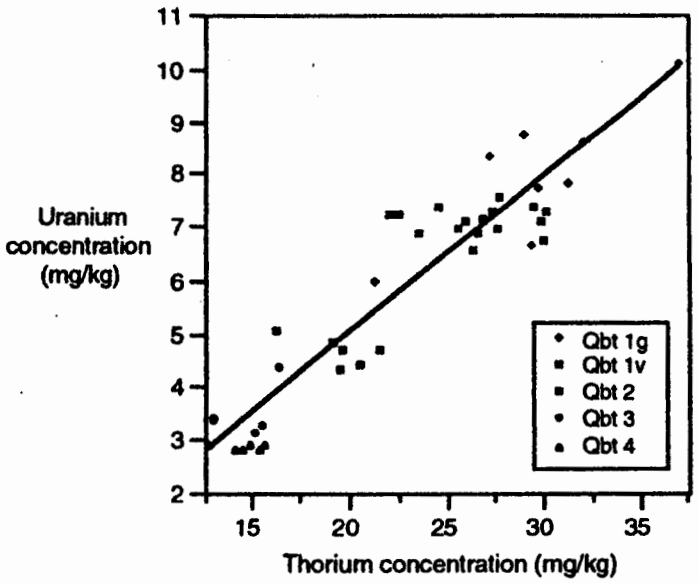


Figure 4. **BIVARIATE PLOT OF URANIUM AND THORIUM (WHOLE ROCK ANALYSIS) FROM BANDELIER TUFF SAMPLES IDENTIFIED BY BANDELIER TUFF UNIT. CORRELATION COEFFICIENT IS 0.933 FOR 44 SAMPLES.**

Recommendations for Collection of Data to Support Background Data Selection

Laboratory soil background data has been collected according to soil horizons (A, B, C) and analysis of these data indicates that naturally occurring levels of inorganic chemicals will vary as a function of certain soil properties (e.g., clay and iron content, see Longmire et al., 1995 1266). In addition, the soil for many sites where data will be collected may have been disturbed and may now consist of heterogeneous fill material. Careful documentation of sampled material is necessary to ensure that data analysts will have sufficient information to select the most appropriate Laboratory background data set for comparison of site-specific data. Background data sets can be chosen with greater confidence when several key soil properties (color, structure, presence of cutans[2], approximate gravel content, and presence of organic matter) are recorded during collection of PRS samples. Recording appropriate key soil information during sample collection is easily achievable by following a simple check-list that your Earth Science Council representative can provide.

The Decision Support and Earth Science Councils can provide further guidance and technical support for sampling and analysis plan development and to support sampling teams in the field.

[2] Cutans are concentration of a particular soil constituent (e.g. clay, organic matter, iron oxides) along soil surface features (e.g. ped faces, pores) and coating the exterior of soil particles (e.g. sand, gravel).

RECOMMENDED STATISTICAL METHODS FOR BACKGROUND COMPARISONS

Because background comparisons are used to make decisions throughout the RCRA process, from site screening to corrective measures implementation, data analysts must use statistical methods that can be applied to a broad range of decisions. This guidance defines two statistical methods for background comparisons, which meet the requirements for RCRA decision making[3]. In the first method, the Hot Measurement comparison, site concentration data are compared with a statistic representing the upper percentile of background concentrations[4]. In the second method, the distributional shift test, the mean (or mean rank, quantile) of site data is compared with the mean (or mean rank, quantile) of background data to determine whether the former is statistically greater than the latter. Used together or separately, these tests help demonstrate whether a release has occurred at a PRS and help define what risk consequence the release may have. Figure 5 illustrates the differences between site data and background data detected by the two methods.

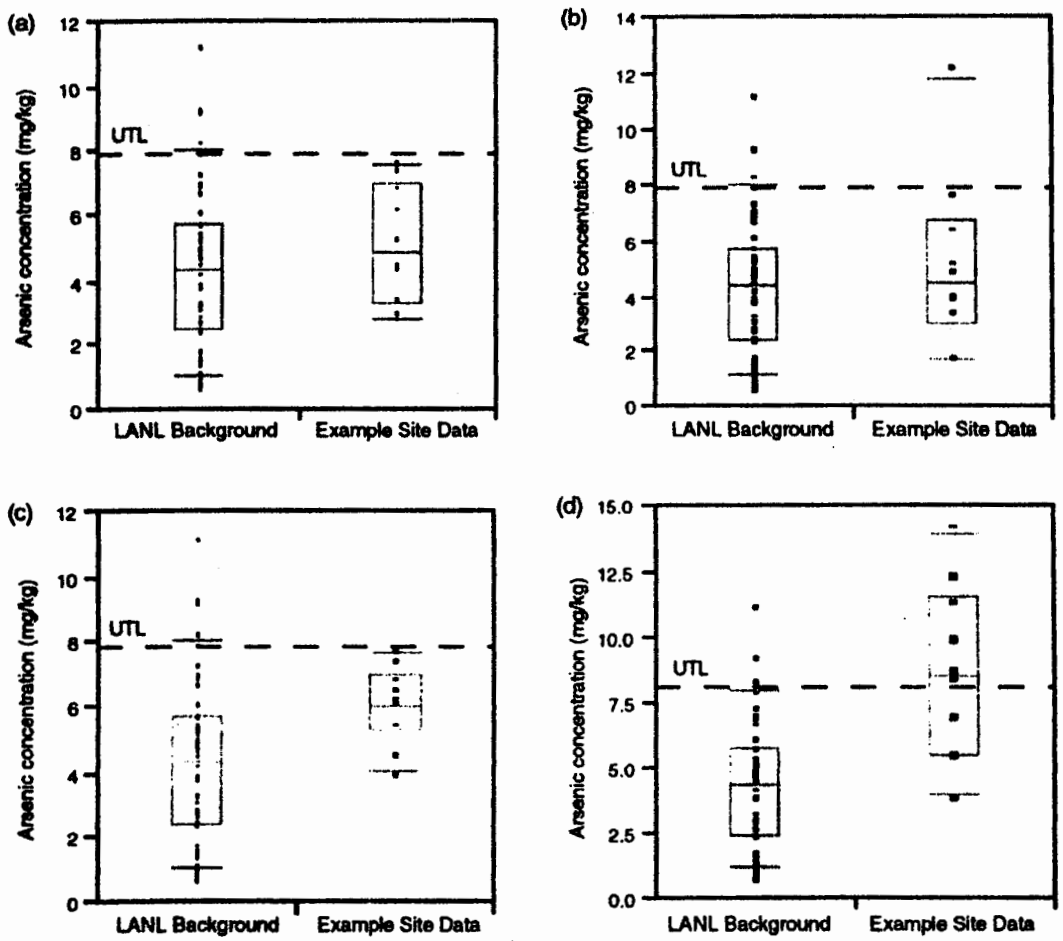
The decision to be supported by the background comparison determines which test is appropriate (see Figure 1). When performing screening assessment, a single high value requires further analysis, and the hot measurement comparison is recommended. Additional statistical tests may be needed to support screening assessments in those cases in which the hot measurement comparison provides inconclusive results. When extensive data are collected to support a risk assessment or corrective action and a shift in the distribution could lead to further action at the site, the distributional shift test is more appropriate. The rationale for selecting a statistical method that differs from those presented in this guidance will be clearly indicated in the ER Project report that summarizes the background comparison.

Because the selection of a particular statistical method depends on the statistical distributions of site and background data, data analysts are encouraged to prepare graphical data displays to communicate the results of data comparisons. Box plots (see Figure 5), in which background data and site data can be compared side-by-side, are most useful. The box plots show actual values (as filled circles) for each data group (Laboratory background and example PRS data in Figure 5). The ends of each box represent the "inter-quartile" range which is specified by the 25th and 75th percentiles of the data distribution. The line within the box represents the median (50th percentile) of the data distribution. Thus the box indicates concentration values for the central half of the data, and concentration shifts can be readily assessed by comparing boxes. If the majority of the data is represented by a

[3] The methods are among those discussed in the RCRA groundwater monitoring guidance document.
[4] The Slippage test is briefly discussed below, and is an alternative to the Hot Measurement test.

single concentration value (usually the detection limit), the box is reduced to a single line. The solid line spanning the series of box plots is the mean value for the entire data set, and the lines above and below the boxes represent the 10th and 90th percentiles of the data. In addition to box plots, data analysts should also consider using histograms and probability plots to provide tangible evidence of similarities or differences between site and background data.

The level of effort spent to evaluate potential differences between PRS data and background data should be related to the site-specific information available. For example, if historical information indicates that beryllium was released at a firing site, the potential differences between beryllium concentration data from firing site activities and Laboratory-wide or site-specific background data should be carefully evaluated to determine the levels of anthropogenic beryllium added to the environment. In all cases, data comparisons will be documented in the appropriate ER Project report.



(a) Site data are within the range of background: no distributional shift or hot measurements (i.e., no values are greater than the UTL).

(b) Site data fail hot measurement comparison: one of eleven arsenic concentrations exceeds the UTL of 7.82 mg/kg.

(c) Site data show a distributional shift: the Wilcoxon rank sum test shows that site data tend to be greater than the background data.

(d) Site data show both a distributional shift and a failure of the hot measurement comparison: seven of ten arsenic concentrations exceed the UTL of 7.82 mg/kg and the site data tend to be greater than the background data.

Figure 5. BOX PLOT COMPARISONS OF EXAMPLE SITE DATA WITH LABORATORY BACKGROUND DATA.

Hot Measurement Comparison

The Hot Measurement comparison uses a threshold value that represents high natural background concentrations. No matter what parameters are chosen to define the threshold, there exists a probability that a natural background measurement will exceed the hot measurement threshold. The frequency of false positive results is minimized by using a threshold statistically related to higher background concentrations. The confidence limit on a percentile of the distribution, termed the tolerance limit, is such a value and is one of the background comparison methods recommended by EPA (1989, 1141). The Laboratory has selected the 95th percentile for calculating the UTL (upper tolerance limit), based on the general guidance in the RCRA groundwater document. EPA recommends calculating an upper 95% confidence limit for the target percentile (EPA 1989, 1141). The UTL for the 95th percentile at 95% confidence can be calculated using Equation (1).

$$\text{UTL} = \text{mean} + \text{standard deviation} * k_{0.95,0.95} \qquad (1)$$

The k-factor depends on the number of background samples; complete tables of k-factors are published in the RCRA groundwater statistical analysis document (EPA 1989, 1141) and in Gilbert (1987, 0312). Table 1 presents k-factors selected to represent the range of values used to compute UTLs for Laboratory background soil samples. To apply Equation (1), the background data must be normally distributed or transformed to normality (e.g., by using a square-root or log-transformation). If data deviate sufficiently from normality, nonparametric methods for calculating tolerance limits should be considered (e.g., Gilbert [1987, 0312]). Alternatively, when appropriate, the data analyst may trim outliers from the distribution and calculate the UTL based on the trimmed mean and standard deviation.

TABLE 1
SELECTED K-FACTORS USED TO CALCULATE UTL*

Number of Background Samples	$k_{0.95,0.95}$
10	2.911
20	2.396
30	2.220
40	2.126
50	2.065
100	1.927

* reprinted from Gilbert (1987, 0312)

An alternative to the UTL is the maximum reported concentration. It should be noted that when few background samples are available, using the maximum concentration will result in

an underestimation of the upper background percentile. In general, the sample maximum concentration (for "n" samples) is an estimate of the $100 \bullet \left[\frac{n-0.5}{n}\right]$th percentile. Thus, if 10 samples are collected, the sample maximum concentration is an estimate of the 95% percentile. Because the observed maximum is extremely sensitive to background sample size, it is not recommended for use as a hot measurement threshold. Rarely detected analytes, which include antimony, cadmium, mercury, selenium and thallium, are an exception to this general recommendation. For this limited subset of rarely detected analytes, the maximum detected background concentration is used as the hot measurement threshold.

Background screening values[5] for inorganics in Laboratory-wide soil, sediment, and Bandelier Tuff are summarized in Tables 2, 3, and 4. As discussed earlier (see Laboratory Background Comparison Approach, pages 4-8), use of the appropriate subset of Laboratory-wide background data is essential, and the rationale for selecting background data should be fully documented. The background screening values found in Tables 2 through 4 should be used only if comparable sample preparation and analytical methods are used for background and PRS samples. In addition, the background screening values are available on the FIMAD (Facility for Information Management and Display) home page and data analysts are encouraged to use the most recent values for background comparisons. Readers interested in details on the calculation of the background screening values are referred to Longmire and others (Longmire et al. 1995, 1266).

[5] Because both maximums and UTLs can be used as hot measurement thresholds, the more generic term "background screening value" is used to describe the values presented in Tables 2 through 4.

TABLE 2
SUMMARY OF MEASURED INORGANIC CHEMICAL BACKGROUND SCREENING VALUES BY SOIL HORIZON

Chemical[1]	Soil Master Horizon A	B	C	All Soil Horizon Data
Aluminum (2)	26600	43600	38700	38700
Antimony (2)	0.5	1	<5	1
Arsenic (3)	6.99	8.12	6.58	7.82
Barium (2)	263	321	286	315
Beryllium (2)	1.41	1.91	1.95	1.95
Cadmium (2)	1.4	2.7	<0.4	2.7
Calcium (2)	4030	6480	5930	6120
Chloride (4)	25.0	78.2	170	75.9
Chromium (2)	19.3	19.0	17.0	19.3
Cobalt (2)	31.0	14.8	41.2	19.2
Copper (2)	15.5	14.3	13.4	15.5
Iron (2)	18100	21800	18500	21300
Lead (2)	28.4	22.3	21.9	23.3
Magnesium (2)	3460	4480	4610	4610
Manganese (2)	1000	673	463	714
Mercury (6)	<0.1	0.1	0.1	0.1
Nickel (2)	12.2	16.0	13.3	15.2
Potassium (2)	3070	3420	3410	3410
Selenium (3)	0.7	1.3	1.7	1.7
Sodium (2)	602	798	2680	915
Sulfate (4)	42.7	249	712	317
Tantalum (5)	<0.9	<0.9	<0.9	<0.9
Thallium (5)	0.4	1	0.6	1
Thorium (5)	13.3	15.0	12.3	14.6
Uranium (5)	1.87	1.72	1.36	1.87
Vanadium (2)	42.8	42.0	32.0	41.9
Zinc (2)	47.1	51.5	50.8	50.8

(1) - All values are in mg/kg, and sample preparation was by EPA method 3050, except for chloride and sulfide, which were extracted by distilled water
(2) - Inductively Coupled Plasma Emission Spectroscopy
(3) - Electrothermal Vapor Atomic Absorbance Spectroscopy
(4) - Ion Chromatography
(5) - Inductively Coupled Plasma Mass Spectroscopy
(6) - Cold Vapor Atomic Absorption Spectroscopy

TABLE 3
SUMMARY OF MEASURED INORGANIC
BACKGROUND SCREENING VALUES BY ROCK UNIT

Chemical[1]	Tt	Qbo	Qct	Qbt 1g	Qbt 1v	Qbt 2	Qbt 3	Qbt 4[8]
Aluminum (2)	4500	1800	3400	3700	8170	3700	3700	15700
Antimony (5)	<0.3	<0.3	0.2	<0.3	0.3	0.3	0.4	<0.3
Arsenic (3)	<0.5	<0.5	0.5	0.7	2	2	5	2.18
Barium (2)	69	23	18	28.0	28.0	28.0	28.0	56.6
Beryllium (2)	0.21	1.2	0.95	1.53	1.53	1.53	1.53	1.82
Calcium (2)	2700	890	1500	4140	4140	1520	1520	2770
Chloride (4)	67	7.7	379	405	405	107	64.8	465
Chromium (2)	10	2.3	1.8	0.94	1.7	1.6	2.1	10.9
Cobalt (7)	NA	8.88	NA	1.27	1.78	1.34	1.39	3.14
Copper (2)	16	2.6	2.2	2.4	2.6	2	2	6.43
Iron (2)	13000	3700	2400	3250	9040	9040	9040	19500
Lead (5)	6.7	5	7.1	16.2	21.9	16.2	16.2	11.0
Magnesium (2)	950	510	510	548	628	548	628	2950
Manganese (2)	280	170	90	273	533	533	426	656
Nickel (2)	15	2.8	<2	<2	2	<2	2.6	8.72
Potassium (2)	1100	960	1600	2730	5540	2730	735	4540
Silver (2)	<1	<1	<1	<2	<2	<1	1.9	<1
Sodium (2)	610	1900	3500	4290	4290	1940	1940	3290
Sulfate (4)	38.6	12.7	548	815	815	815	815	1430
Tantalum (5)	0.2	0.9	0.3	0.3	0.5	2	0.8	0.5
Thallium (5)	<0.3	0.9	<0.2	0.7	1.7	1.3	1.7	0.49
Thorium (5)	6.4	1.4	4.2	7.69	22.1	11.5	9.29	6.1
Uranium (5)	0.6	0.2	0.4	1.39	5.93	2.48	1.64	0.9
Vanadium (2)	29	2.8	3.8	7.67	4.01	4.01	4.01	20.2
Zinc (2)	41	21	17	56.3	84.6	59.0	59.0	75.4

(1) - All values are in mg/kg, and sample preparation was by EPA method 3050, except for chloride and sulfide, which were extracted by distilled water
(2) - Inductively Coupled Plasma Emission Spectroscopy
(3) - Electrothermal Vapor Atomic Absorbance Spectroscopy
(4) - Ion Chromatography
(5) - Inductively Coupled Plasma Mass Spectroscopy
(6) - Cold Vapor Atomic Absorption Spectroscopy
(7) - Maximum detect of Instrumental Neutron Activation Analysis
(8) - Background screening values are from Broxton et al. (1996, 1305)
NA - Not available

TABLE 4
SUMMARY OF BACKGROUND SCREENING VALUES FOR SEDIMENT SAMPLES

Chemical[1]	Screening Value ($UTL_{95,95}$)
Aluminum (2)	11800
Arsenic (3)	3.94
Barium (2)	141.
Beryllium (2)	1.40
Calcium (2)	3340
Chloride (4)	NA
Chromium (2)	8.77
Cobalt (2)	5.16
Copper (2)	9.85
Iron (2)	14400
Lead (2)	13.8
Magnesium (2)	2310
Manganese (2)	490
Nickel (2)	10.0
Potassium (2)	2850
Sodium (2)	195
Sulfate (4)	NA
Thorium (5)	11.1
Uranium (5)	1.29
Vanadium (2)	21.3
Zinc (2)	62.1

(1) - All values are in mg/kg, and sample preparation was by EPA method 3050, except for chloride and sulfide, which were extracted by distilled water
(2) - Inductively Coupled Plasma Emission Spectroscopy
(3) - Electrothermal Vapor Atomic Absorbance Spectroscopy
(4) - Ion Chromatography
(5) - Inductively Coupled Plasma Mass Spectroscopy
(6) - Cold Vapor Atomic Absorption Spectroscopy
NA - Not available.

The hot measurement comparison is made between the maximum detected site sample and the background screening value (UTL or maximum). Exceeding the UTL as a background screening value is not definitive evidence that a release has occurred at a PRS. Assuming the PRS is at background and the statistical model is correct, there is a 5% probability that the 95th percentile will be exceeded by each sample collected from the PRS. Furthermore, a typical inorganic chemical suite requires comparison of 23 analytes with background. If the concentrations of the 23 inorganic analytes vary independently, the 5% probability that each PRS sample exceeds the 95th percentile increases to a 69% probability that at least one of the 23 ninety-fifth percentiles will be exceeded in a single sample. Additionally, given that the probability values for these multiple comparisons have not been adjusted, the overall confidence level for 23 analytes will be substantially less than 95%. In addition to the strictly probability-based discussion presented above, the possibility of exceeding a UTL due to an unusual, but naturally occurring, soil matrix is a further consideration. Consequently, the results of a hot measurement comparison must be carefully evaluated.

The results of the UTL or maximum comparison should also be evaluated with respect to potential human health or ecological effect concentrations. Some inorganic chemicals (arsenic, beryllium and manganese in particular) represent a special case. Because soil and tuff background levels for arsenic, beryllium or manganese at the Laboratory exceed risk-based screening levels, no screening action levels (SALs) are being used for these chemicals for the Laboratory ER Project. Two inorganics (aluminum and thallium) have background UTLs that represent a significant fraction of the SAL (10 to 50%). If, in a comparison similar to a multiple chemical evaluation, all inorganic chemicals had a concentration equal to the UTL, the total of the inorganic UTLs divided by the corresponding SAL (in effect normalizing the UTL-to-SAL ratio) would equal 116%. Thus, concentrations of most naturally occurring inorganic chemicals are significantly lower than their respective SALs, except aluminum, arsenic, beryllium, manganese and thallium. Thus, background comparisons are more important where releases of aluminum, arsenic, beryllium, manganese and thallium are suspected.

Both the multiple chemical evaluation and the UTL-to-SAL comparison help determine what level of effort should be expended to evaluate deviations from background. When only a single statistical comparison with background is performed, the UTL comparison is adequate for most naturally occurring inorganic chemicals because probability levels are not compromised. Under this circumstance, the UTL is the simplest comparison and is functionally most similar to comparisons of site data to target risk levels or SALs.

Slippage Test

The Slippage test is an alternative to the Hot Measurement or UTL comparison. It is based on the maximum observed background value and the number of site concentration values ("n") that exceed the maximum of the background data (Gilbert and Simpson 1990, 0972). The result of the Slippage test is a probability that "n" site samples exceed the maximum background concentration by chance alone. The Slippage test is potentially more useful than the UTL comparison because it is based on a statistical hypothesis test and not simply a statistic of a distribution. However, because the Slippage test is similar to the UTL comparison, performing it will not usually provide any additional information beyond the UTL comparison. Thus, performing one or more of the following distribution shift test is recommended if additional statistical tests are warranted.

Distributional Shift Tests

A distributional shift test is used to determine whether site data are systematically greater than background data. Several types of distributional shift tests are available, and these tests are presented in two groups below. The preferred statistical method in each group is indicated where there are multiple options.

For detecting distribution shifts between all PRS data and the appropriate subset of LANL-wide background data, the following statistical tests can be used:

- The Student t-test is a parametric, statistical, two-sample test that determines whether the mean concentration of site data is statistically greater than the mean concentration of background data (Gilbert 1987, 0312). Data analysts should be aware that the t-test performs well for some deviations from normality, but in general, the t-test is not recommended, because it assumes that the data being compared are normally distributed.

- The Wilcoxon rank-sum test is the nonparametric equivalent to the t-test (Gilbert 1987, 0312; Gilbert and Simpson 1992, 0974). The Wilcoxon test pools site and background data into one aggregate set and determines whether the average rank of site data is greater than that of the background data. The Wilcoxon test is recommended when site data consist of few samples or when nondetects are relatively infrequent (<10%). The Wilcoxon rank-sum test is recommended as the default statistical test to determine distribution shifts between all data, because most environmental data are rarely fit by a normal distribution and frequency of detection for most inorganics is greater than 90%.

- The Gehan test, recommended when non-detects are relatively frequent (>10% and <50%), handles a single detection limit in a statistically robust manner (Gehan 1965, 1296). It is identical to the Wilcoxon rank-sum test when applied at sites for which no non-detects occur.

- The Peto-Prentice test is another variation on the Wilcoxon rank-sum test (Millard and Deverel 1988, 1295). The Peto-Prentice test, also recommended when non-detects are relatively frequent (>10% and <50%), handles multiple

detection limits in a statistically robust manner. It is identical to the Wilcoxon rank-sum test when applied at sites for which no non-detects occur.

For detecting distribution shifts between the upper ranges of PRS data and the appropriate subset of LANL-wide background data, the following statistical test can be used:

- The Quantile test (Gilbert and Simpson 1992, 0974), which compares the upper quantile of background data with that of PRS data, is capable of detecting a statistical difference when only a small number of PRS concentrations are elevated. Because it does not artificially reduce statistical significance, the Quantile test is the most useful distributional shift test for PRSs at which samples from a release represent a small fraction of the overall data collected. For example, to detect contamination from historical spills at unknown locations, an RFI work plan may call for samples to be collected from a grid. Most sample results show no contamination, but those in or near spill locations show elevated concentrations. The Quantile test can be used when the frequency of non-detects is approximately the same as the quantile being tested. For example, in a case having 75% non-detects in the combined background and PRS data set, application of a quantile test comparing 80th percentiles would be appropriate.

The ability to use either of types of distributional shift tests is dependent on the number of samples available for comparison. In general, at least 10 sample concentrations for comparison with background data are needed to provide adequate confidence for detecting a shift. Frequently, during Phase I of an RFI, inadequate numbers of samples are collected to warrant a distributional shift comparison.

To infer a significant result in a distributional shift test, a 95% confidence level is recommended. Given that multiple comparisons will be performed with the distributional shift test, the same statistical interpretation issues cited above for the hot measurement test are also relevant (see p. 19). In addition, the human health and ecological consequences of PRS concentration data above background must be considered along with differences in inorganic chemical concentrations between soil horizons.

REFERENCES

EPA (US Environmental Protection Agency), 1989. "Ecological Assessment of Hazardous Waste Sites: A Field and Laboratory Reference," EPA/600 3-89/013, Environmental Research Laboratory, Corvallis, OR. (EPA 1989, 0303)

EPA (US Environmental Protection Agency), 1989. "Interim Final RCRA Facility Investigation (RFI) Guidance, Volume I of IV, Development of an RFI Work Plan and General Considerations for RCRA Facility Investigations," EPA/530-SW-89-031, OSWER General Directive 9502.00-6D, Office of Solid Waste, Washington, DC. (EPA 1989, 0088)

EPA (US Environmental Protection Agency), April 1989. "Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities. Interim Final Guidance," Office of Solid Waste, Waste Management Division, US Environmental Protection Agency, Washington DC. (EPA 1989, 1141)

EPA (US Environmental Protection Agency), April 1992. "Guidance for Data Useability in Risk Assessment (Part A)," Office of Emergency Remedial Response, US Environmental Protection Agency, Washington DC. (EPA 1992, 1166)

Broxton, D.E., R.T. Ryti, D. Carlson, R.G. **Warren, E. Kluk, and S. Chipera.** March 20, **1996.** "Natural **Background Geochemistry of the Bandelier Tuff at MDA P,** Los Alamos National Laboratory," Los Alamos National Laboratory Report LA-UR-96-1151, Los Alamos, New Mexico. **(Broxton et al. 1996,** 1305)

Campbell, K., 1994. "Estimation of Local Background Concentrations for Identification of Environmental Releases," Los Alamos National Laboratory Report LA-UR-94-2274, Los Alamos, New Mexico. (Campbell 1994, 1294)

Gehan, E. A., 1965. "A Generalized Wilcoxan Test for Comparing Arbitrarily Singly-Sensored Samples," *Biometrika*, Vol. 52, Nos. 1 and 2, pp. 203-223. (Gehan 1965, 1296)

Gilbert, R.O., 1987. Statistical Methods for Environmental Pollution Monitoring. Von Nostrand Reinhold Company Inc. New York. 320 pp. (Gilbert 1987, 0312)

Gilbert , R. O., and J. C. Simpson 1990. "Statistical Sampling and Analysis Issues and Needs for Testing Attainment of Background-Based Cleanup Standards at Superfund Sites," in *Proceedings of The Workshop on Superfund Hazardous Waste: Statistical Issues in Characterizing a Site: Protocols, Tools, and Research Needs,* Environmental Protection Agency, Arlington, Virginia. (Gilbert and Simpson 1990, 0972)

Gilbert, R. O., and J. C. Simpson, 1992. "Statistical Methods for Evaluating the Attainment of Cleanup Standards, Volume 3: Reference-Based Standards for Soils and Solid Media," Pacific Northwest Laboratory, Richland, Washington 99352. (Gilbert and Simpson 1992, 0974)

Longmire, P. A., D. E. Broxton, and S. L Reneau (Eds.), October 1995. "Natural Background Geochemistry and Statistical Analysis of Selected Soil Profiles, Sediments, and Bandelier Tuff, Los Alamos, New Mexico, Los Alamos National Laboratory Report LA-UR-95-3486, Los Alamos, New Mexico. (Longmire et al. 1995, 1266)

Millard, W. P., and S. J. Deverel, 1988. "Nonparametric Statistical Methods for Comparing Two Sites Based on Data with Multiple Nondetect Limits," *Water Resources Research,* Vol. 24, No. 12, pp. 2087-2098. (Millard and Deverel 1988, 1295)

APPENDIX

Soil Horizon Primer:

Soil horizons at the Laboratory generally are unconsolidated and largely consist of A, B, and C horizons. The A horizon occurs at the land surface, and therefore, accumulates humified organic matter typically mixed with a much larger fraction of minerals than found in the B or C horizons. A horizons can also occur as thin, disturbed soils horizons with only minimal amounts of humified organic matter. The latter are common where surface activity (e.g., grazing, overland traffic) has compacted or partially stripped off original upper soil horizons.

The B horizon underlies the A horizon and shows little or no evidence of the parent rock structure. The B horizon often contains increased geochemically reactive mineral phases (phyllosilicate clay minerals, iron oxyhydroxides, organic coatings) that may concentrate major and trace elements. B horizons in soils at the Laboratory display a wide range of features and degrees of development. B horizons largely display characteristics resulting from several primary soil-forming processes and types of Laboratory illuvial[6] material. Laboratory B horizon soils consist of three subgroups: 1) weakly developed B horizons (B_w) that have minimal changes in physical and chemical properties relative to the parent material; 2) clay-rich B horizons that have increased in clay-sized material over time B_t); and 3) B horizons that have been influenced by the accumulation of calcium carbonate (B_k). Transitional soil horizons may also occur (e.g., AB, BA, BC) sharing physical and chemical properties with another soil horizon. The first letter for transitional horizons indicates the dominant soil horizon property, and the appropriate background data for such transitional soils is the dominant soil horizon.

C horizons, wide-spread at the Laboratory, consist of slightly altered- and non-altered parent materials. Examples of minimally altered parent material would include accumulation of silica and calcium carbonate, mineral alteration through oxidation and reduction processes, and gleying[7].

All soil profiles across the Laboratory are underlain by R horizons that consist of consolidated bedrock that is usually highly fractured, but has undergone minimal chemical alteration. Laboratory background data include some R horizon soil samples, but no statistical evaluation of these samples has been made because of the small number of R horizon

---

[6] Illuvial is the process of deposition of colloidal material (e.g. clay, iron oxides) within a horizon. Colloidal material is usually from an overlying horizon.
[7] Gleying is the reduction of iron (Fe+3 to Fe+2) under anaerobic (i.e., saturated or 'waterlogged') soil conditions.

samples collected. Parent materials of the R horizons include alluvium, colluvium, Bandelier Tuff, El Cajete pumice, and other geological material. In most mesa top cases, Bandelier Tuff is the primary parent material. Thus, comparison to Bandelier Tuff background data should be evaluated for PRS samples that have been designated as R soil horizon samples.