OUTLINE
COMMENTS ON ROCKY FLATS BACKGROUND STUDY
July 30, 1991

COMMENTS ON "BACKGROUND GEOCHEMICAL CHARACTERIZATION REPORT FOR 1989- ROCKY FLATS PLANT"
Mike Gansecki, EPA Region VIII
August 1, 1991

GENERAL COMMENTS

1. Organizing and interpreting the background data collected for the various media at Rocky Flats Plant (RFP) is a complex and difficult task; the authors of this report should be complimented for their efforts. In particular, a number of issues raised by EPA in earlier comments were given thorough consideration. The following are the major points worthy of notice:

a. Use of General Statistical Principles and Development of a Protocol. A very comprehensive protocol was developed to address a number of issues including the possibility of spatial variability effects on background data distributions. Given the wide array of media and issues involved, the selection of the statistical tests and protocol represents a reasonable approach to developing basic background statistical definitions for sediments, soils and parent geologic material, ground water and surface water.

a. Spatial Considerations- The initial effort to determine whether MANOVA tests could be used for determining areal differences was particularly instructive. For a number of reasons, including limitations on statistical assumptions, it was not generally feasible to make such a distinction (north vs south). However, we do think such an investigation had merit, and perhaps as better definitions of media distributions are obtained, the method can be reapplied.

b. General Water Quality Geochemistry- the use of ion balance analyses, gravimetric versus calculated TDS and other water quality comparisons, and Stiff diagrams to compliment the discussions of basic water geochemistry is particularly well done. The major aqueous chemistry results are reasonable for site conditions, and shows a great deal of continuity between media and geologic strata.

c. Interpretations of Unusual Geochemistry at Surface Water Quality Springs SW-104 and SW-80. The information provided in the report provides substantial credence to the idea that the anomalous data reported from these locations can be correlated with the high total suspended solids collected in the samples. Our own analyses in this review support the report's general conclusions.

e. Presentation of Information. A much more readable document was prepared for this effort than in the original report. The data was substantially consolidated and the report organized more coherently to make following the progress of analysis considerably easier.

2. There are still some substantial problems with some of the data, statistical analyses and interpretations. One reason for EPA Region VIII's focus on this background study development is that we are involved in a number of similar activities at other RCRA and CERCLA sites. In particular, a number of closures of RCRA hazardous waste management facilities deal extensively with soil-based contamination and attainment of appropriate cleanup levels (in some cases, background). Thus, a number of our comments are directed prospectively towards the use of these background data in RFP cleanup or monitoring situations. Hopefully, some of our experience may be of use in the present setting, since many of the issues are similar. All further references to the 1989 background study report are given as "the RFP report".

It needs to be made clear that some of the statistical issues in particular, have not been fully worked out by EPA. Trial-and-error is also a part of the process of developing workable methods for ordering and comparing data. Many of our comments need to be taken as suggestions that may be helpful for the Rocky Flats Plant background definitions.

This review focuses on major statistical concerns, as a number of separate issues. These issues deal simultaneously with different media and sections of the report. Select data reviews of the RFP report and more detailed evaluations developed over time are found in the Appendices. From time to time, references to these evaluations in the Appendices will be found in the major issues review.

MAJOR ISSUES

1. Organization and Presentation of Information

In order to understand and review this RFP report, it is necessary to be able to track the following progression: data acquisition, data validation and evaluation, statistical summaries and further evaluations such as tolerance interval development, outlier analysis, etc. It is still somewhat difficult to move from the raw data tables in Appendix A of the RFP report to understand how the statistical summaries in Section 5 were compiled, and how tolerance intervals and outlier analysis was done. A major problem is the confusing enumeration system. The report should have provided individual maps showing specific sampling locations for each media sampling group. Secondly, an organized keying system should have been used in the raw analytical data tables; some media data were numbered (e.g. borehole data), while others were not (surface water quality data). Since north/south areal comparisons were prominently featured in this report, the keying system should have identified which groups the data belonged with. Without this information, pulling together various grouped data to make independent evaluations was a major effort.

As one example, the report evaluates areal differences between north and south Rocky Flats background for borehole soil samples and ground water monitoring well data. In Section 3 of the report, only the narrative identifies the various groups of ground water well samples. On page 3-5 of the report, it is indicated that the ground water monitoring wells for the north Rocky Flats alluvium were B200589, B200689, B200789 and B200889, and the south Rocky Flats alluvium: B405586, B400189, B400289, B400389, and B400489. This in itself is difficult enough to follow in the raw data tables; however, in comparing chloride statistical calculations for these samples (Tables 5-12 and 13), it was apparent from sample sizes that more data values were in the statistics than just from the above wells. By checking all the data, it appeared that data from wells B405689 and B405789 were also added to the second data set. The text narrative does not explain why.

Borehole locations formed only a subset of holes dug for ground water monitoring wells. In the text, 18 borehole locations are mentioned; Table 3-1 and the raw data only contained 17 borehole locations. The enumeration system using only numbers is extremely confusing, particularly when other locations are added to samples. Surface water quality station and sediment numbers were also supposed to be matched by location; however, the numbers bear no relationship to one another. Some alphanumeric system in the raw data tables and the report which specifically keys the raw data to the various analytical sets which are statistically compared, would be immeasurably helpful.

The study is intended to define overall background conditions at the Rocky Flats plant. While present statistical analysis suggests that occasional north/south or geologic strata form distinct areal populations, it is still useful to include grand statistical summary tables for each analyte of the media sampled. It is somewhat easier to visualize ANOVA decisions by comparison with grand mean and other overall statistics.

The tables in Section 5 of the RFP report are limited in the statistical information they contain. For example, where data are considered follow a lognormal distribution, the arithmetic mean and standard deviation, mean and standard deviation of the logarithms, perhaps other statistics justifying one distribution over another (skewness, kurtosis, results of one or more tests) should also be included. As we will discuss in a further issue, such information could have identified some irregularities in statistical manipulations of lognormal data. An example of a more comprehensive statistical information is found in the discussion of total aluminum in sediments, Appendix B.

At this point, major efforts should concentrate on developing reliable basic soils, ground water and surface water data sets for future statistical analysis. Basic statistical information is probably more important than sophisticated analyses at present. A good geochemical understanding of the data is paramount, especially in evaluating problematic outliers discussed elsewhere in this review, developing a consistent and workable approach to below detection limit data, and perhaps supplementing with improved sampling data.

2. Grouping Data for Statistical Comparisons

The identification of appropriate classification "groups" for statistical analysis is a fairly complex matter, since a varying combination of analytes (individual compounds, analyte groups, total versus dissolved, etc.), spatial locations, certain temporal considerations, and the four principal media are involved. The table below suggests a number of possible spatial/temporal classifications for each medium (assuming all analytes are handled similarly):

MEDIUM	POTENTIAL CLASSIFICATIONS
Sediments	by Entire RFP Site by Stream Drainage by Surface Water/Spring Stations
Boreholes	by Entire RFP Site by Individual Boreholes (all depths) sub: by Depth (Top 4' vs Lower Depths) by Geochemical Stratum sub: by Geographical area (N vs S) by Groundwater Well Association
Surface Water	by Entire RFP Site by Station sub: by Month sub: by Season by Stream Drainage by Surface Stations vs Springs
Ground Water Wells	by Entire Site by Individual Well sub: by Season by Geochemical Stratum sub: by Geographical area (north/south)

These are only the more logical potential classifications for the raw data. Each classification needs to be evaluated for each analyte (or analyte group) to determine whether there are distinct parent population distributions. In this light, the effort that went into distinguishing north versus south areal distributions at the Rocky Flats Plant site, may have overshadowed the need to consider these other potential distinctions. Fairly common statistical definitions suggested above (totals for site, individual stations and wells), etc., were not presented in the report. We believe that it is still somewhat premature to make firm judgments about the north/south classifications, given other problems with the data and statistical analyses and assumptions discussed below.

Geochemical units make sense for some media, so long as it is entirely clear that the samples arise only from these definable units. With borehole data, we are unsure if total depth samples are always entirely within a single stratum. Are wells consistently screened entirely within a given stratum, to exclude waters from other units?

Future versions of this background RFP report should consider addition of some or all of these possible data classification distinctions.

3. Detection Limit Issues

Two major considerations involving detection limits are discussed here: the first deals with how statistical analyses consider the present data below detection limits particularly where variable detection limits are involved, and detection limits in relation to data quality objectives for various sampling media.

a. The RFP report uses the assumed value of 1/2 the reported detection limit for many analyses. In other situations, Cohen's method is used with differing detection limits in a data set by using the average of detection limits.

Specifying half the detection limit will work reasonably well for many statistical analyses, particularly if the normality assumption is being checked and the percentage of non-detect data is fairly low. In other situations, particularly when lognormal distributions are considered, the values assumed for the detection limits can play a major role in the apparent shape of the distribution.

However, variable detection limits in two situations need further consideration: a) when all values are below detection (e.g.: antimony [Sb] in the Rocky Flats alluvial borehole data set); and b) when variable detection limits exceed quantifiable data (e.g.: lithium [Li] for north/south Rocky Flats alluvium and colluvium borehole samples).

In the first case, the specification of a mean and standard deviation based on assumed detection limit values is virtually meaningless. Future comparisons should be made with the most restrictive criteria (i.e., the lowest detection limits). A detectable value (and especially repeat detections) should be considered significant. Hence, some of the highest detection limit values may need to be disregarded.

In the second case, cross-comparisons of "significant differences" are driven almost entirely by the assumed non-detect data. We do not believe that these latter data show significant differences, only uncontrolled differences in detectability during the sampling/analytical process. A further outlier analysis should investigate why detection limits for lithium were in the <30 mg/kg range, while they were in the <3 mg/kg range for the other three lithium data sets.

This problem is general to most of the media data sets. A consistent set of detection limits (or at least an acceptable upper limit) needs to be specified, if cross-comparisons are to be made. We suggest as a start that non-detect data be acceptable only if they are less than 2x the smallest detectable value or smallest detection limit; somewhat more liberal criteria might be used with radiochemical data, if the error terms are used as effective detection limits. In this way, the assumed values of 1/2 the detection limit will be less than the minimally observed quantified data. There may be more sophisticated methods in the statistical literature for dealing with variable non-detects, but for gross, initial comparisons, some acceptable cutoff needs to be made.

After screening out the unacceptably high below detection limit (BDL) data, methods like that of Cohen used in the RFP study or Helsel's[1,2] method of using linear regression on

---

1 Dennis R. Helsel, "Less Than Obvious: Statistical Treatment of Data Below the Detection Limit", Environmental Science & Technology, Vol. 24, #12, 1990, pp.1766-1774

2 Robert J. Gilliom & Dennis R. Helsel, "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data: I. Estimation Techniques", Water

above detection limit values with "fitting" of non-detect data can be used. Appendix G of this review contains such a definition of lognormal data detection limit criteria for three sediment and borehole trace element data sets using Helsel's approach. Certain non-detect values had to be discarded. A critical requirement in these "fitting" techniques is that of assuming an appropriate distribution for the data under evaluation. This is discussed further in the next issue. However, if such an assumption can be made, the fit of the entire data set including BDL values to the assumed distribution can be dramatically improved.

b. Since this is an investigatory study to develop background data for future comparisons, analytical methods should have been chosen which could provide detection limits at low enough levels to characterize most analytes of interest. As Newman et. al.[3] have indicated:

> "... an ultimate goal in analytical method selection and experimental design selection should be the generation of a data set with all values in the region of quantitation."

Table 1 attached to these comments contains a comparison of detection limits identified in this report for total trace element and major cation soil values (sediments and boreholes) with detection limits believed attainable using standard EPA RCRA SW-846 methods. Principal SW-846 methods used are ICP for most metals and special atomic absorption methods for arsenic, lead, mercury and selenium. The fifth column shows the expected mean concentrations of total metals in background soils for this region from USGS and EPA studies.

Quite a wide range of detection limits is evidenced even within each RFP media group. Generally, the detection limits are similar between sediments and boreholes. For the major cations, results for calcium, aluminum and iron are satisfactory. However, detection limits for magnesium, potassium, and sodium in soil are often too high to detect lower values.

Resources Research, Vol 22, #2, February 1986, pp. 135-146

3 Newman, Michael C., Dixon, Philip M., Looney Brian B., & Pinder, John E. III, "Estimating Means and Variance for Environmental Samples with Below Detection Limit Observations", Water Resources Bulletin, Vol. 25, #4, p. 908

TABLE 1. DETECTION LIMITS COMPARISON FOR ROCKY FLATS BACKGROUND SOILS ANALYSES
(All Values in mg/kg)

ANALYTE	ROCKY FLATS DETECTION LIMITS SEDIMENTS	BOREHOLES	OTHER[1] SOIL DLs	ANALYTICAL METHOD	TYPICAL AREA[2] BACKGROUND MEANS	COMMENTS
Aluminum [Al]	<500	<500-<5000	<5-<10	ICP	8000-50000	OK
Antimony [Sb]	<13-<40	<13-<17	<.2-<30	ICP	<10	OK at lower limits
Arsenic [As]	<1.9-<13	<2.1	<.2-<.3	GFAA	2-4	Misses low values; control variability
Barium [Ba]	<43-<140	<40-<45	<.5	ICP	30-500	Misses low values; control variability
Beryllium[Be]	<1-<3.5	<1.1	<.1-<.3	ICP	.6-1.5	Misses low values; lower DL possible
Cadmium [Cd]	<1-<3.5	<1.1-<1.5	<.1-<1	FAA/ICP	<1	OK, very rarely detected
Calcium [Ca]	<1000-3000	<1000	<10	ICP	1000-20000	OK
Cesium [Cs]	<200-<700	<200-<240	ND	-	ND	Misses all values; no comparative data
Chromium [Cr]	<2.2	<3	<1-<3	ICP	10-30	OK
Cobalt [Co]	<10-<30	<10-<12	<.3-<2	ICP	2-8	Misses low values; control variability
Copper [Cu]	<5-<15	<5.7	<.6-<2	ICP	10-20	Misses low values; lower DL possible
Iron [Fe]	<1000	<8000	5	ICP	15000+	OK
Lead [Pb]	<2	<2-<4	<2-<4	GFAA	10-40	OK
Lithium [Li]	<20-<70	<22	ND	-	5-24	Misses low values; lower DL needed
Magnesium[Mg]	<1000-3000	<1100	7	ICP	500-10000	Misses some low values
Manganese[Mn]	<50	<90	<100-500	ICP	100-500	OK
Mercury [Hg]	<.1-<.35	<.1-<.24	<.05	CVAA	.01-.10	OK
Molybdenum[Mo]	<20-<70	<2.3-<2.5	ND	-	<3-<8	OK
Nickel [Ni]	<8-<30	<9-<10	<1-<4	ICP	2-20	Misses low values; lower DL possible
Potassium [K]	<1000-3000	<1000	<4	ICP	500-20000	Misses low values; lower DL possible
Selenium [Se]	<1-<2.5	<1-<5	<.2	GFAA	.1-.5	Misses low values; lower DL possible
Silver [Ag]	<2-<7	<2	<.5-<1	ICP	<.5	OK, very rarely detected
Sodium [Na]	<1000-3000	<1100	<8	ICP	800-10000	Misses low values; lower DL possible
Strontium[Sr]	<25-<220	<200-<500	ND	-	20-200	Misses low values; control variability
Thallium [Tl]	<1-<50	<2.1-<5	<.6	ICP	<.6	OK; control variability
Tin [Sn]	<20-<70	<20-<50	ND	-	1.3	Misses all values; lower DL needed
Vanadium [V]	<10-<33	<10-<11	<.2	ICP	10-100	Misses low values; lower DL possible
Zinc [Zn]	<9	<10	<1-<6	ICP	25-75	OK

[1] Other Soil DLs/Methods refers to analyses done for EPA RCRA background soils studies

[2] Typical Means are values from USGS and EPA background soils studies for Rocky Mountain plains region.

For minor trace elements, only chromium, lead, manganese, and zinc have low enough detection limits to provide quantifiable data for most or all soils. Although antimony, cadmium, mercury, molybdenum, silver, and thallium detection limits were greater than any detectable levels, these are at least comparable to other efforts, so long as detection limits are controlled. The remainder of trace elements have methods and lower detection limits which could provide more quantifiable data for analysis of background soils.

By contrast, RFP study analytical methods and detection limits for aqueous species follow the CLP protocol closely and are adequate for most purposes. A separate table for these comparisons was not presented. Detection limits for arsenic, barium, cadmium, chromium, lead, mercury, selenium, and silver are below present drinking water standards. It is only for the more commonly present cations-- magnesium, potassium, and sodium, that detection limits are high enough to miss low values. These are not considered critical, but it does cause the data analyses to be less reliable. There are routine analytical methods available to reach to at least a 1 mg/l level for these analytes.

4. Outlier Evaluations/Exploratory Data Analysis

The RFP report indicates that "before MANOVA is invoked, the data are examined for outliers, rejected data are excluded and treated as missing values, and results below the detection limit are transformed to half the detection limit for the analyte." The specification of one-half the detection limit may be satisfactory for normal outlier evaluations, but lognormal distributions are sensitive to specified BDL values, since they can take a range from the logarithm of the detection limit to minus infinity. Even one postulated value can change the outcome of a distributional test, as shown for total aluminum in borehole data in Appendix G.

Figure 2-1 of the RFP report shows that an outlier test is run prior to deciding upon the form of the population distribution for a data set. Unfortunately, the form of an outlier test for individual values is highly sensitive to the assumed distribution. So too, is the test for homogeneity of variances. If a data set conforms to a lognormal distribution, the use of ASTM E178-75 on the raw data seems inappropriate. One would expect an increase in variance as the means of the data increased.

In conclusion, we are not sure what function the outlier test provides at this early point in the statistical protocol. There are also more fundamental "outlier" analyses that need to be done for all the raw data, which cannot be simply identified through a statistical test. Gilbert[4] lists a number of data validation procedures, which form an outlier analysis of basic data:

> "1. Routine checks made during the processing of data. Examples include looking for errors in identification codes......
>
> 2. Tests for internal consistency of a data set. These include plotting data for visual examination by an experienced analyst and testing for outliers.
>
> 3. Comparing the current data set with historical data to check for consistency over time....
>
> 4. Tests to check for consistency with parallel data sets, that is, data sets obtained presumably from the same population...."

These approaches need further evaluation in this background study. Gilbert suggests a number of statistical tests which can be done, but it is important to at least have a sense of which data make sense, and which ones may not. The following are a number of examples drawn from the Rocky Flats study which we consider problematic outliers based on Gilbert's criteria:

1. A dissolved Ra-226 value of 170 ± 240 pCi/l in a total data set of 24 values with a range otherwise from -.1 ± .5 to 2.8 ± .5;

2. The surface water total metals and radiochemical species in SW-80 and SW-104, orders of magnitude above other surface water values;

3. Anomalously high values for most natural radiochemical species in Well B405289; and

4. Duplicate analyses for TDS in ground water inorganic analyses showing 290 and 2900 mg/l, and the entire set of trace element duplicates in samples #82 and #83 in the borehole data set.

Such an identification of potential outliers need not exclude data automatically. However, it would be very useful to try to compare these data with past studies either at the site or done under similar circumstances in the region. Conclusions may differ as well.

---

4 Gilbert, Richard O., Statistical Methods for Environmental Pollution Monitoring, Van Nostrand Reinhold Publishing Company, 1987, pp.186-7

The first value described above may need to be deleted from the data set because it appears spurious and belongs under the tritium values (see discussion in Appendix F). The second set of data may need to be further analyzed and perhaps distinguished as a typical of the springs only; the data are unique and unlike other more representative site data (see discussion in Issue #9 below). In the third case, there may be a need to distinguish this individual well from other classifications, since it does indeed appear to be naturally "hot"; the data may be valid but confuse otherwise more predictable comparisons (see Issue #9 discussion, Appendices D and F). Further sampling may also be able to confirm tentative conclusions at this point. The final duplicate data sets show anomalous values; the first TDS value of 290 mg/l is comparable on geochemical terms with other cation/anion data; the second is not. The last set of duplicate values cannot possibly be independent data since they are identical.

These are only typical examples, but are suggestive of the need to further refine raw data sets prior to sophisticated statistical analysis. Other examples of potential outliers based on general mathematical and geochemical principles are found in the Appendices, particularly Appendix F. The RFP protocol identifies typically extreme values for more thorough investigation (p. 4-6); this protocol needs to include approaches as suggested above and in some of the discussions which follow.

5. <u>Selecting Distributions</u>

We believe that for certain fundamental media data sets, where the form of the distribution has been reasonably well identified from past studies, it is more appropriate to assume a distribution before an outlier test is run. For trace elements in background soils, for example, the USGS has indicated that historically lognormal distributions provide a good fit of the data. For individual dissolved species within a given well, a normal distribution is probably a good assumption. Distributional studies on small data sets can prove misleading, since the variability of these sets is great. Other biases such as a lack of complete random sampling might affect these specific data sets; numerous studies involving large samples over time such as the USGS efforts allow a measure of confidence in such a choice of distributions.

EPA can provide other examples of background trace element soils data and ground water data set which appear to conform to these distributional assumptions. We are not stating here that a lognormal distribution is preferable to any other distribution such as a Weibull, which can describe skewed data. However, the lognormal distribution offers some features of being able to work within certain well described statistics based on the normal distribution.

However, for combinations of well and other data into the desired classifications, the question is somewhat less clear. The lognormal distribution will probably better characterize such data. In these assumptions, we also include the radiochemical parameters (discussed later in this review).

We would like to better understand how the criteria were used to decide upon normality versus lognormality. For example, arsenic in sediments was determined to follow a normal distribution. Region VIII experience has been that arsenic tends to be lognormally distributed in background soils, in similar fashion to other total trace elements. We ran a Pearson correlation coefficient test using a method by Devore[5] comparing expected Z-values versus the ordered raw data and logarithms. A higher correlation was observed for the lognormal data, although not significant. Given the large number of below detection data, the results are not unexpected. The non-detect data appears to have had a preponderant influence in making the distribution determination. However, based on similar measured data, we would expect the distributions for trace element soil analytes to be lognormal. A rough indication of anormality is if the correlation coefficient is greater than .4-.5 for reasonably sized data sets; most of the trace element metals had values higher than this. If lognormality were assumed, a robust fit of the data can be obtained using methods akin to Helsel's[6] regression order statistics, even with a substantial number of non-detects.

From the classification groups mentioned in Issue #2 above, we feel the following distributional assumptions can be used unless good evidence is given to the contrary:

---

5 Jay L. Devore, Probability and Statistics for Engineering and the Sciences, Brooks Cole Publishing Company, Second Edition, 1987, pp. 574-5

6 Dennis R. Helsel, "Less than Obvious: Statistical Treatment of Data Below the Detection Limit", op. cit., pp.1766-1774

ANALYTE GROUP	SEDIMENTS (Site)	BOREHOLES (Hole)	BOREHOLES (Site)	SURFACE WATER (Station)	SURFACE WATER (Site)	GROUNDWATER (Well)	GROUNDWATER (Site
Trace Elem(Tot)	Log	?	Log	Normal	Log		
Trace Elem(Dis)	Log	?	Log	Normal	Log	Norm	Log
Inorganics	-	-	-	Normal	Log	Norm	Log
Radiochem(Tot)	Log	?	Log	Normal	Log		
Radiochem(Tot)	Log	?	Log	Normal	Log	Norm	Log

Individual borehole data was not evaluated over depth in the present RFP effort, but, if done, could be used to develop a sense of short-scale spatial variation in soil materials. A limited look at total chromium in borehole samples at the 0-4' versus deeper levels is found in Appendix C.

It should be recognized that these statistical distribution assumptions are only approximations. The independence assumption for soils and ground water data probably is potentially complicated by a non-random component (spatial and/or temporal); nearby samples are more likely to resemble each other than those farther apart in space and time. However, such a determination is complicated and outside the realm of practical assessment with limited sample sizes and locations for most media. More intensive ground water well monitoring on a monthly basis could provide the opportunity for temporal non-random considerations.

The site background data are more or less randomly drawn (even with some selectivity involved in sample locations), and overall characterizations are desired. Some discriminatory power is lost in not precisely identifying such a non-random component, but for most comparisons, the use of random statistics is probably acceptable. EPA has had to admit similar simplifying assumptions in evaluating background/site comparisons at other facilities.

6. Tolerance Interval and Other Statistical Comparisons

A very important consideration in defining the use of tolerance intervals or other statistics is how these background data are to be used in future comparisons with downgradient wells or soil-based cleanups. Individual test samples of the same type as drawn for background, mean value comparisons of limited sets of samples, or composited samples from a number of separate samples, may be compared to the background distributions. The kinds of statistics involved will differ with the kind of comparison desired.

At this point, the RFP background study offers only the use of tolerance intervals defined for either normal or lognormal distributions (a test of proportions for data with less than 50%

Table D-1. STATISTICAL PROPERTIES OF RADIOCHEMICAL PARAMETERS-- Gross Alpha, U-238, Pu-239, & Ra-226-- IN OVERALL GROUNDWATER WELL DATA: ROCKY FLATS BACKGROUND STUDY

STATISTICAL PARAMETERS	GROSS ALPHA		U-238		Pu-239		Ra-226	
	NORMAL	HELSEL	NORMAL	HELSEL	NORMAL	HELSEL	NORMAL	HELSEL
With Well #205589								
Sample Size	96	96 (44)[1]	99	99 (60)	90	90 (15)	21	21 (11)
Arithmetic Mean, pCi/l	7.10		2.61		.0027		.507	
Standard Dev., pCi/l	21.3		9.96		.0056		.570	
Log Mean[2], pCi/l		.101		-.873		-6.455		-1.27
Log Std Dev[2], pCi/l		1.97		1.87		1.239		.966
Corr Coeff, r	.513/.941*	.988(.927)[1]	.472/.883*	.996(.986)	.745/.998*	.999(.973)	.703/.952*	.986(.930)
95% SIGNIF?[3]	NO	YES	NO	YES	NO	YES	NO	YES
95% Toler Interv[4] pCi/l	48.3	50.54	21.9	15.7	.014	.0174	1.61	2.77
95% Data Value, pCi/l	24(est)		15.0(est)		.014(est)		2.8(max- 97%)	
99.0% X+ t*s Lev:[5], pCi/l	57.5	117.9	26.2	41.5	.016	.030	2.13	4.385
Max Data Value, pCi/l	200		16.9		.03		No Value	
Without Well #205589								
Sample Size	94	94 (42)	97	97 (58)[1]	89	89 (15)	20	20 (10)
Arithmetic Mean, pCi/l	4.64		1.225				.387	
Standard Dev., pCi/l	6.30		1.987		(NO CHANGE)		.184	
Log Mean[2], pCi/l		.157		-.813				-1.123
Log Std Dev[2], pCi/l		1.71		1.53				.527
Corr Coeff, r	.897/.994*	.986(.912)	.787/.995*	.997(.988)			.978/.996*	.993(.958)
95% SIGNIF?[3]	NO	YES	NO	YES			NO	YES
95% Toler Interv[4] pCi/l	16.8	32.28	5.08	8.63			.832	1.17
95% Data Value, pCi/l	16(est)		5.65(est)				.80(max)	
99.0% X + t*s Lev:[5], pCi/l	19.6	67.3	5.93	16.7			.854	1.24
Max Data Value, pCi/l	38.9		11.6				No Value	

1 Detectable (ADL) values in parentheses used for HELSEL regression analysis; 2 Log Mean from intercept, Log Std Dev from slope of HELSEL regression equations; 3 Critical values for correlation coefficient test, $c_{.05,+100} = .985$; $c_{.05,21} = .95$;

4 Tolerance interval values: for N=100, K=1.94; for N=21, K=2.371

5 Student-t one-sided value for N=125, at the .01 level of significance: t = 2.37; For N=21, t = 2.53.

* Net correlation coefficient results using Helsel's approximation on raw data

Table D-2. MEAN VALUE CALCULATIONS FOR SELECT RADIOCHEMICAL PARAMETERS
GROUNDWATER WELL DATA: ROCKY FLATS BACKGROUND STUDY

STATISTICAL PARAMETERS	GROSS ALPHA				Ra-226	
	ARITHMETIC	LOG-3 PARAMETER[1]	LOG DET LIM	HELSEL	ARITHMETIC	HELSEL
Background Statistics						
Sample Size	96	96	96	96	20	20(10)
Arithmetic Mean, pCi/l	7.10				.387	
Standard Dev., pCi/l	21.3				.184	
Log Mean, pCi/l	.101	2.0955	.980	.101		-1.123
Log Std Dev, pCi/l	1.97	.697	1.269	1.97		.527
Corr Coeff, r	.513/.941*	.928	.989	.988(.927)	.996	.993(.958)
95% SIGNIF?[3]	NO	NO	YES	YES	YES	YES
Test Statistics						
Sample Size	9	9			9	9
Mean	62.4	3.22			1.82	-.426
Standard Deviation	85.1	1.51			3.196	1.46
Data (synthetic)	250, 150, 80, 30, 18, 16, 10, 5, 3				10, 3, 1.2, .8, .5 .4, <.5, <.3, <.2	
Tolerance Interval Calculations:						
95% Toler Interv[4] pCi/l	48.4	31.46	31.24	50.5	.832	1.17
95% Data Value, pCi/l		24(est)				.80(max)
99.5% X+ t*s Lev, pCi/l	63.1	50.9	74.99	196.8	-	-
Max Data Value, pCi/l		200				.80
Confidence Interval Calculations:						
t-Exponential[2]	20.3	16.0	18.13	26.12	.501	.518
Test:	SIG	SIG	SIG	SIG	SIG	SIG
Land[3]	-	19.9	35.1	428.5	-	.579
Test:	-	SIG	SIG	NOT SIGN	-	SIG
Student-t Test						
Test Statistic:	+5.03	+4.05	+4.95	+4.59	+2.04	1.91
Result	SIG	SIG	SIG	SIG	SIG	SIG

1 +4.5 added to raw data values; 2 Exponential confidence interval; 3 Land confidence interval (see Appendix D discussion)

APPENDIX E. <u>BOREHOLE SAMPLES- TOTAL METALS RF & CO</u>

<u>Table 5-60 and 67 Statistics for Total Metal Concentrations in Rocky Flats Alluvial and Colluvial Borehole Samples</u>

1. The tables indicate that all mean and standard deviation values are "untransformed" logarithmic values. It is unclear what this means. The tables indicate that all values are given as logarithmic statistics. In this case, the untransformed mean for aluminum [Al] should be around 9.0 to 10. The untransformed standard deviation would be somewhere around .1-3. If the antilog of the log-normal values are given, then the means are geometric means. This looks to be the case for aluminum, arsenic, barium, etc. However, the corresponding geometric standard deviations would always be greater than 1.0 and probably in the 1-3.0 range. The barium value in Table 5-60 is 4.6256, which is enormous for the range of data shown. Standard deviations for aluminum, iron [Fe], magnesium [Mg}, manganese {Mn], and sodium [Na] look more like normal statistics. This statistical presentation does not make sense.

2. Statistical values are given for antimony [Sb], where the largest value is a non-detect. This so compromises the data, which only shows a 6.7% detection rate anyway, that the data seem useless. The detection limits are too high and variable for extracting useful information. A comparable situation exists for cadmium [Cd], cesium [Cs], selenium [Se], and probably sodium [Na] and cobalt [Co]. No statistics will work well with these data, other than to state detection limits.

<u>Tables 5-61, 62, 68, and 69. Statistics for Total Metal Concentrations in Rocky Flats Alluvial/Colluvial North and South Sectors</u>

1. These are comparable tables, since they are distinguished by common areal sectors. Many of the same problems with the presentations of logarithmic statistics, described above for Tables 5-60 and 67 are also found here. Calcium [Ca] standard deviations are not untransformed lognormal data, nor are they transformed. It is not clear if the value of 560.02 given in Table 5-68, for example, is even an arithmetic standard deviation. Chromium [Cr] standard deviation is less than one, which looks like an untransformed lognormal standard deviation. The standard deviation for strontium [Sr], however, looks like a transformed value.

2. The lithium [Li] values in Table 5-61 are anomalous compared to the other three tables. In this data set, only 5.6% of the values were above detection, while the other three data sets ranged between 81-100%. The mean value for lithium in the first table is significantly greater than the other three. Something looks amiss here. Other metals like chromium, lead, zinc and

copper, were much more consistent in their rates of detection. A similar situation exists for the metals molybdenum [Mo] and strontium [Sr], with occasional drastically different rates of detection.

3. Whatever the mean values are (arithmetic, geometric, etc.), one can see a great deal of commonality between north and south sectors for both the Rocky Flats alluvial and colluvial borehole data. While there may be minor statistical differences, values for calcium, chromium, copper, lead, nickel, vanadium, and zinc. All of these mean metal values generally agree within a range of 2x [calcium may be an exception because of calcite precipitation/dissolution]. The most significant determinant of whether there will be larger mean trace element differences seems to be the rate of detection; low rates of detection suggest high variability between comparisons. This also suggest that there is a very major problem in deciding how to use less than detection limit data. Even with these limitations, one can see that most of the borehole data conforms to typical background concentrations.

Table 3 to the main EPA comments contains a summary of average values (arithmetic and geometric means from various Rocky Flats and other regional soil backgrounds. Comparing the data in the first column (Rocky Flats Borehole ALL) with other sites (the Denver-Arapahoe Chemical Waste Processing Facility subsurface clays and USGS information for Longmont, CO), it can be seen that most of the more commonly measured trace elements-- arsenic, barium, cadmium, chromium, cobalt, copper, lead, manganese, mercury, nickel, selenium, vanadium, and zinc-- are all within two to three times of each other within the same order of magnitude. Not surprisingly, it is the major cations-- calcium, iron, magnesium and sodium-- which show the greatest variability. The column labeled TSS surface will be explained further on in this review, but is still within similar ranges for most metals.

This table also contains a comparison of a subset of the borehole data with the full set. 0-3' interval data were calculated independently and compared with the full set. The present statistics in the first column suffer from the problems discussed above, regarding the confusion of the statistics. Even with these problems, it can be seen that among the commonly measured trace elements, very similar patterns hold. There is relatively little difference, except for the major cations. Arsenic, chromium, and vanadium show great enough arithmetic differences that there may be distinctive values with depth. Chromium is analyzed separately in Appendix C.

Only beryllium is present at Rocky Flats in somewhat higher concentrations than typical backgrounds (approximately 3-5 mg/kg in borehole data versus .74 to 1.4 for other regional sites). The values are still within reasonable ranges, considering that

Longmont also shows somewhat higher values for a wide area of study.

APPENDIX F. RADIOCHEMICAL ANALYSES

This discussion covers a number of topics, some of which will use results from above. Ground water wells are at the heart of RCRA hazardous waste monitoring. Because RCRA rules have generally presumed a comparison of single point-in-time and space data points against background, the issues concerning the appropriate tolerance intervals are relevant here.

Questions about the appropriate level of statistical development and the use of normal versus lognormal statistics are also considered.

1. Use of Tolerance Intervals

Once a set of background wells have been identified for the appropriate comparisons, downgradient well data will be compared against these background. The form of the statistical test depends on a number of factors, including how the regulations themselves are interpreted.

Historically, EPA 40 CFR 264 Subpart F regulations have specified single well data comparisons against background. Perhaps realizing the weakness of such an approach, EPA's recent modification of the statistical approaches to this section, allows for some variation in approach. The alternatives suggested generally make use of a set of samples collected at given individual monitoring wells, and aggregated for comparison against background through the use of techniques such as ANOVA.

Such comparisons are possible, but also run into some practical difficulties. If ground water data are temporally variable, which most naturally occurring species are, samples taken together close in time (days to a month), are more likely to resemble one another, than represent the full annual variation. We agree that multi-year data collection will be necessary to factor out seasonal effects in the data. This presumes that the data are reasonably orderly. Such may not be the case with radionuclides, which vary over a number of orders of magnitude in background at the site. Rocky Flats' should be prepared to accept the eventuality that single point-in-time comparisons may be necessary, because of the data complexity.

As mentioned earlier, data in the RFP report which are lognormally distributed appear to have problems in their development. While the background report has presently characterized dissolved ground water radionuclides as normally distributed over the site, we believe that such an approach is inadequate. The chief reason for positing normality seems to be the need to avoid transformation of negative numbers (page 2-12). It does not appear that the actual shapes of the distribution for these radionuclides was taken into account. We believe that lognormal statistics suitably applied will provide a much better characterization of the occurrence of these radiochemical species.

2. Level of Statistical Development- Radionuclides

At present, the background report provides detailed statistical calculations in Tables 5-14, 19, 24, 29, 33, 38, and 48, for the various geochemical units-- Rocky Flats alluvium, colluvium, valley fill, weathered claystone, weathered sandstone, unweathered sandstone, and a "lowermost aquifer flow" system. We never did find the rationale for the latter breakout, but the data were not used any further. An explanation of this distinction would be helpful.

A quick look at these tables suggests that an a priori decision was made to define these data as normal for statistical purposes. Yet the coefficients of variation alone, for species known to have above detection values (gross alpha, U-238) suggest that there is considerable skewness to many of these radionuclides. We doubt that these data are well described by normal statistics. We emphasize that it is for ground water monitoring wells above all, that tolerance intervals may be most appropriate. Unfortunately, the form of the distribution is critical in defining the appropriate tolerance levels.

Using the normality test described in Devore (pp. 574-5) on data for the following four radiochemical species: gross alpha, U-238, Pu-239, and Ra-226, the table below shows the results. The comparison includes the raw data versus normal expected Z-scores, and the logarithms of the data:

CORRELATION TEST FOR NORMALITY ON RADIOCHEMICAL SPECIES

LOCATION	ANALYTE											
	GROSS ALPHA			U-238			Pu-239			Ra-226		
	N	Norm	Log	N	Norm	Log	N	Norm	Log	N	Norm	Log
RF Alluv	28	.883	.963*	28	.927	.929	27	.715	.746	-	-	-
Colluv.	9	.750	.964*	9	.810	.964*	7	.651	.651	-	-	-
Valley F	19	.916	.968*	19	.929	.972*	18	.745	.827	-	-	-
Unw Clay	13	.935*	.970*	15	.974*	.975*	11	.737	.783	-	-	-
Unw SS	19	.833	.880	20	.601	.950*	19	.887	.899	-	-	-
Weat SS	8	.968*	.968*	8	.939*	.915*	8	.886	.892**	-	-	-
Totals	96	.816	.903	99	.472	.975**	90	.745	.832	23	.74	.83
										22	.98*	.98

* Significant at 95% level (fits distribution)
**Significant at 99% level (fits distribution)

In general, the overall superiority of using a lognormal distribution can be seen from this table. Logarithms were generate by adding a fixed value to the raw data to ensure that the minimum value of each data set was slightly above zero. This method is described in Gilbert, 1987, as a three-parameter lognormal distribution. While there can be some interpretation problems with this method, the greater accuracy of the lognormal distribution for radiochemical data seems obvious.

This table affords a number of other observations. The overal best data fits occur with gross alpha and U-238, two radiochemical parameters which have considerable positive values in a natural background environment, and particularly in the uranium rich sediments off the Front Range granites. In particular, the lognormal distribution almost accurately describes the entire data set for Rocky Flats background. By contrast, the occurrence of plutonium is not nearly as well described. As will be seen, there i ample reason for this, since plutonium does not occur naturally in the environment.

Finally, the distribution of Ra-226 is not evaluated for individual geochemical strata, since not enough data were available Evaluation of all the Rocky Flats data, however, shows that the accuracy of the distribution is very strongly affected by a single value. Without that value, either a normal or lognormal distribution well describes the present data.

One point further needs to be made regarding both Ra-226 and Ra-228. In the raw data, it became clear that radium data were onl a fraction of the total number of samples taken for other radiochemical parameters. By cross comparison with other data, it is apparent that the radium analytes were measured only with the highest gross alpha or uranium occurrences. This introduces a systematic upward bias into the sample. It should be quite clear that any description provided by these radium data **are not typical for the entire site and should not be used as background parameters independently.** The report should identify which criteria were used to determine when the radium analytes were measured.

The report did not provide details on how it was decided that radiochemical data were significantly different enough to break out into separate geochemical statistics. While we can agree that it makes some sense to provide data for these distinguishable geochemical units, it is unclear whether the differences are always significant.

In addition to the geochemical units, there are two other logical groupings of ground water data: individual wells and data aggregated over the entire site. We provided an independent analysis of the basis for these decisions for a number of parameters. The four parameters used above were carried through fo the remainder of the analysis.

3. Analysis of Outliers in Radiochemical Parameters

Given that there is some basis for preferring lognormal statistical descriptions for radiochemical parameters versus normality, the first step was a direct inspection of the raw data for the ground water wells (lower aquifer data excepted). An early inspection of the data concentrated on two types of anomalous data: very high isolated values, and systematically high values across a number of parameters which might be related.

This kind of screening can point out obvious outliers, or certainly ones which need much further screening. Since these data have not yet been fully screened for quality control, these early identifications may prove helpful. Looking at data within wells might suggest anomalies of a given well itself, irrespective of its probable geochemical location. For this analysis, only the four parameters--gross alpha, U-238, Pu-239, and Ra-226-- are evaluated. However, some of these results might be extrapolated to the remainder of the data set:

* <u>Ra-226, Well B200689, 7/27/89</u>: A value of 170 ± 240 pCi/l is given. This value is at least 70 times higher than the next highest value, and about 425 times the geometric mean value. The value is not matched by any unusual values in other parameters. It also is supposed to have occurred when the alpha value was -1 and the U-238 value 0. This looks to be an obvious mistake. We will guess that this value actually belongs in the Tritium column and was mistakenly moved. The ± 240 error term also looks consistent with tritium analyses. W did not use this value in any further calculations.

* <u>Gross alpha, U-238 and Ra-226, Well B205589, 7/21/89 and 10/25/89</u> : All these values for both dates look very high in comparison to other well data. Plutonium, which is not expected to occur naturally, is not elevated. We believe that these are potentially real occurrences of very high values. However, they appear to be confined to a single well in the colluvial material. This poses some problems for statistical analyses, since it is only one well of 41, yet obviously will have inordinate influence in mathematical terms.

Such an occurrence in the well might be a hydrochemical depositional pocket of uraniferous materials, known to occur i the eastern Rockies. As such, it might be considered representative of the background. However, from a statistical standpoint, it might be more tractable to keep this well independent from the other data, and compare potentially comparable downgradient well data on an individual basis, if a similar regime is suspected. In further evaluations of data for these four parameters, these outliers are included and identified.

No other individual wells or data seemed obvious. Within the a priori geochemical groups, wells within each groups seemed reasonably consistent. For example, U-238 values in the valley fil group consistently averaged in the 1-3 pCi/l range, while Rocky Flats alluvial wells were generally below 1.0. This kind of consistency makes more plausible a radiochemical characterization o these geochemical strata. Incidentally, colluvial data for naturally occurring species were quite high in two other wells, so the incidence of very high levels in well B205589 may be consistent

For descriptions of radiochemical analytes within each well, w believe that normal statistics apply unless shown otherwise. At present, there are too few data to make such a decision. But we believe that they will behave in very similar fashion to other dissolved trace element analytes. The data groupings however, appear to be better characterized by lognormal statistics.

ANOVAs were run on raw data for the six definable geochemical groups on both the raw data and the logarithms of the data. Summar data from these evaluations is found in the table below:

ANOVA EVALUTIONS FOR RADIOCHEMICAL SPECIES/GEOCHEMICAL GROUPINGS

RADIOCHEMICAL ANALYTE	N	OVERALL SIGNIFICANCE F	p	?	GEOCHEMICAL GROUPS DISTINGUISHABLE MEANS
Gross Alpha*					
Raw Data	94	3.85	.003	SI	RF(2.2)+VF(3.1) vs CO(10.4) RF(2.2) vs WC(8.9)
Log Data	94	4.95	.000	SI	RF vs CO+WC+US VF vs WC
U-238*					
Raw Data	97	7.23	.000	SI	RF(.26) vs CO(4.3)+WC(2.0) VF(1.5) vs CO WC(2.0) vs CO US(1.0) vs CO WS(.35) vs CO
Log Data	97	10.34	.000	SI	RF vs CO+VF+WC US vs CO+VF+WC WS vs CO+VF+WC
Pu-239					
Raw Data	91	.55	.739	Not Significant	
Log Data	91	1.07	.384	Not Significant	
Ra-226 (Not compared because of sample size)					

In general, the naturally present species Gross Alpha and U-238 do show some separation by geochemical unit. However, the separation is not completely distinguishable. The colluvial material shows the highest values for both Gross Alpha and U-238. These calculations did not include the two highest values, which would increase the separation from the other means but would not change the overall relationships. Logarithmic data tended to show greater separation and some slightly different relationships. Thus there is some basis for geochemical separation, but not consistentl for all parameters.

Plutonium did not show a significant difference of means among the different geochemical strata. This is excellent confirmation that there should not be any significant differences for a non-naturally occurring chemical. Not enough radium-226 data were available for comparison. As mentioned earlier, radium samples are also not representative of site conditions, because of selectivity in sampling.

Use of Negative and Non-Significant Data

The radium-226 data set for ground water is an interesting example of some of the problems involved with simply using the firs number of an X ± Y value, irrespective of its relationship to zero. In essence, the present method of computing statistics totally ignores the second value. Presumably, the error term has been calculated for a reason.

The following are the 24 data for dissolved Ra-226 in ground water wells:

RA-226 DATA FOR GROUND WATER MONITORING WELLS

WELL NO	DATE	VALUE	WELL NO	DATE	VALUE
B200689	6/7/89	.1 ± .5	B200689	7/27/89	170 ± 240*
B201189	7/21/89	.5 ± .3	B201189	5/5/89	.4 ± .2
B200889	6/5/89	.5 ± .5	B201289	7/25/89	.8 ± .4
B203289	6/20/89	.5 ± .7	B203489	9/27/89	.29 ± .09*
B204189	8/25/89	.2 ± .3	B205589	7/21/89	2.8 ± .5
B302989	7/19/89	.3 ± .3	B302989	4/27/89	.2 ± .2
B302889	4/27/89	-.1 ± .2	B304989	8/23/89	.6 ± .2
B305389	5/31/89	.6 ± .3	B305389	8/23/89	.7 ± .5
B402189	8/24/89	.3 ± .3	B402189	6/5/89	.4 ± 1.0*
B405489	8/25/89	.4 ± .2	B405489	6/21/89	.4 ± .5
B400389	6/16/89	.4 ± .5	B405586	6/8/89	.025 ± 25*
B405889	8/23/89	.4 ± .2	B405589	10/18/89	.4 ± .2

Values shown in bold are two samples from the same well. The four values with asterisks following create some problems of interpretation and need close inspection under quality control and quality assurance of the data. Some explanation is in order why th error terms of these four values vary so significantly from the other data. The value 170 ± 240 simply looks like a mistake; it bears no relationship to the other data, and more importantly, to the first sample from the same well taken on 6/7/89.

In the background report, the 170 pCi/l value is used alone as the maximum value in the data set for Rocky Flats alluvial ground water. This results in a bizarre "average" of 28/5 pCi/l, with a standard deviation of 69.2 pCi/l. Yet the median value of the data set is about .4 pCi/l, with the next highest value only .5. Instea of the Rocky Flats alluvium being relatively low in Ra-226, the dat shows a relatively "hot" situation. Even if the 170 ± 240 value were a radium measurement, the result should be interpreted to mean that the value cannot be distinguished from zero for the specific background count. One could more readily read this value as a zero

The value of .025 ± 25 also needs checking. Why should the counts vary by a factor of 50 for the dissolved well medium? Looking at other radiochemical data fails to show any elevated levels of other analytes. Could this error term actually be .25?

On general analytical principles, similar measurements with th same equipment on similar samples should achieve similar results.

APPENDIX G.

USE OF SELECT METHODS IN TRANSFORMING BACKGROUND SOILS DATA

In developing statistics for background study data, it is necessary to define the distributional properties of the data. In this analysis, soils data from background sediments and boreholes from the Rocky Flats Plant in Colorado were used.

The effort here is intended to improve the use of lognormal statistics, where they are felt to be appropriate. Others such as the U.S. Geological Survey[1] routinely use lognormal statistics in defining trace element distributions in soils.

A recurrent problem in organizing and collating data occurs with data below detection limits. Since trace elements are considered here, significant numbers of the data are below detection. In most cases, the detection limits are themselves variable, although it is assumed that the same sampling and analytical procedures are used. The problem affects how statistics are developed for given databases. For the present study, an initial assumption of 1/2 the detection limit was used for calculation purposes. Limitations to this procedure will be discussed.

Background statistics are being developed at the Rocky Flats plant for making future comparisons with other situations, such as site monitoring and cleanups. The type of statistics generated should reflect the kinds of testing desired. In hazardous waste cleanups and routine monitoring, comparison of individual sample data and grouped mean data to background is desirable. The manner in which the data are collected, however, can affect the type of statistics as well. In cases, where insufficient or unquantifiable data are all that is available, the kinds of statistical testing are accordingly restricted.

The approach here looks at three selected trace elements from two RFP soil databases-- total aluminum, arsenic, and vanadium. They were selected from a suite of 28 measured trace elements to represent varying detection limit conditions. Sediment data for total aluminum were all above detection limits (ADL); about 1/2 of the vanadium values were ADL; arsenic had only 21% ADL. The distributional properties of the three elements were evaluated for both the available background sediment data (17-19 samples) and soil coring boreholes (117-122). Summary statistics for both arithmetic and geometric

---

1 Background Geochemistry of Some Rocks, Soils, PLants, and Vegetables in the Conterminous United States, US. Geological Survey Professional Paper 574-F, 1975, p.F12ff

properties were calculated. Distributional testing for normality and lognormality was performed; the lognormal data were tested utilizing both the 1/2 detection limit values for below detection data, and a second method for deriving synthetic below detection limit values from linear regression described by Helsel[2]. The extrapolated statistics were used to compare estimated extreme values with observed data.

APPROACH

Trace element data were obtained from the Rocky Flats background study[3] for the two soil areas-- surface stream and spring sediments and coring boreholes. Here, data were considered in the aggregate for the Rocky Flats plant background environs as a whole. Sediment samples were taken at select locations near surface streams and springs; borehole data were collected in conjunction with ground water monitoring wells developed at the site.

The RFP data were not purely randomly sampled; selectivity was involved in both the sediment and borehole cores, although there is a fairly wide spatial distribution of boreholes across the background of the site. Borehole data were obtained from often varying total borehole depths at somewhat irregular intervals. These sampling methods could affect the overall distributional properties somewhat; however, experience at other sites suggests that the differences with depth and location can be outweighed by local short-scale variation which cannot be predicted in advance. Also the average properties of a site are typical of the larger geochemical environment of the Colorado Front Range. So long as no obvious sources of contamination or man-influenced activities were used to identify locations, it can still be expected that the data will follow typical soil distributional properties.

Once the basic data were entered into a MINITAB program, summary statistics for both raw and log-transformed data were obtained. These raw and log-transformed data were tested against

---

2 Environmental Science & Technology, Vol. 24, #12, 1990, "Less Than Obvious: Statistical Treatment of Data Below the Detection Limit", Dennis R. Helsel, p.1769

3 Background Geochemical Characterization Report for 1989-- Rocky Flats Plant, Golden, Colorado, EG&G Rocky Flats, Inc., December 21, 1990

assumptions of normality as follows, using a MINITAB method described by Devore[4]:

> "A quantitative measure of the extent to which points cluster about a straight line is the sample correlation coefficient r.... Consider calculating r for the n pairs (x_1,y_1)......(x_n,y_n)..[where x_i are concentrations and $y_i = \text{phi}^{-1}\{(i-.5)/n\}$ phi is the cumulative normal frequency distribution]...the more r deviates from 1, the less the probability plot resembles a straight line. This idea can be used to yield a formal test procedure; reject the hypothesis of population normality if $r \leq c_a$, where c_a is a critical value chosen to yield the desired significance level a. That is, the critical value is chosen so that when the population distribution is actually normal, the probability of obtaining an r value that is at most c_a (and thus incorrectly rejecting H_o) is the desired a. The developers of the MINITAB statistical computer package give critical values for a = .10, .05, and .01 in combination with different sample sizes. These critical values are based on a slightly different definition of the y_i's:...
>
> Let $y_i = [Z_i] = \text{phi}^{-1}\{(i-.375)/(n + .25)\}$, and compute the sample correlation coefficient r for the n pairs $(x_1,y_1$.....(x_n,y_n). A test of
>
> H_o: the population distribution is normal
> versus
> H_a: the population distribution is not normal
>
> consists of rejecting H_o when $r \leq c_a$. Critical values c_a are given {below}:

Appendix Table A.14

n	a = .10	.05	.01
5	.9033	.8804	.8320
10	.9347	.9180	.8804
15	.9506	.9383	.9110
20	.9600	.9503	.9290
25	.9662	.9582	.9408
30	.9707	.9639	.9490
40	.9767	.9715	.9597
50	.9807	.9764	.9664
60	.9835	.9799	.9710
75	.9865	.9835	.9757"

---

4 Probability and Statistics for Engineering and the Sciences, Second Edition, Brooks/Cole Publishing Company, Monterey, California, 1987, pp. 574-5

First the data were sorted in increasing order; then the calculated Z-values were obtained from the y_i above for given sample sizes. Both raw and lognormal data were compared with corresponding Z-values in this way. Critical 95% c_as were extrapolated from the table. For the larger borehole data sets, a 95% c_a value of .985 was used for samples larger than 100.

The MINITAB linear regression model was used to estimate the mean and standard deviations for the lognormal data; raw data were uniformly so far below the critical level that it was not felt reasonable to evaluate the raw data any further.

Linear regression was first run on the log-transformed data having 1/2 detection limit values (LOGREGR). Outputs including information on values having excessive residuals or influence were noted. A second linear regression was run on the subset of these data above detection with the appropriate Z-values (Helsel method). Z-values were calculated from the full data set, and then truncated in similar fashion to the concentration data. For the Helsel approach, the Z-values for the below detection limit data were used to generate synthetic BDL values from the regression equation for the ADL values. Overall correlation coefficient test results were obtained from the resulting regression on the combined synthetic BDL and ADL values.

In order to be able to utilize the Helsel method, a set of rules were developed for BDL data. So long as a sample detection limit was no more than twice the lowest recorded value, the BDL values were used both in the LOGREGR and HELSEL approach. Detection limit values above this criterion were rejected; for example, out of 19 identified vanadium sediment values, 17 were used. The two rejected values had 1/2 detection limit values in the range of recorded values. Secondly, the data sets for the HELSEL method were truncated at the point above the largest BDL value used.

Mean and standard deviation data were used to estimate 95% tolerance intervals, following the EPA guidance recommendations.[5] In order to compare the derived statistics from the larger data sets with maximum observed values, a t-statistic was used since tolerance levels were not available for lower occurring frequencies than 95%.

RESULTS

Tables 1 and 2 present the results of the calculations for sediment and borehole data respectively. Table 1 results showed that the normality assumption for the trace element raw data was

---

5 Statistical Analysis of Groundwater at RCRA Facilities, Draft Guidance, EPA, October 1988

untenable. Standard deviations were close to or larger than arithmetic means; further the correlation coefficient test showed wide divergences from normality for aluminum, arsenic, and vanadium data.

The transformed logarithmic values shown under LOGREGR with the 1/2 detection limit gave mixed results. All aluminum values were above detection, hence no comparison with the HELSEL method is possible. The vanadium data were fit by the LOGREGR method suitably, although the HELSEL method showed even better conformance with the lognormal distribution. Arsenic data did not meet the criterion of the test using LOGREGR, but did so with the HELSEL method. It should be recognized that with so few positive values available for arsenic (21% above detection) in a small sample and a very high maximum value, the results were considerably different from the LOGREGR approach. The slope of the concentration (log) vs Z-value line is much steeper with the HELSEL method, and a correspondingly lower mean. This would be expected to result in much higher predicted extreme values.

Using these derived logarithmic means and standard deviations, 95% tolerance intervals were generated and compared with maximum observed values. Since there were 17-19 points in the original data sets, the maximum observed frequency was around 5% or less, so that the two numbers could be compared. The aluminum upper tolerance level was 44487 ppm, well above the observed maximum of 21600 ppm. A tolerance interval of 33.3 ppm for arsenic was well above the observed 13.0 ppm value, and the maximum tolerance level of 84.87 ppm was above the 50.2 ppm observed level for vanadium. By contrast, the use of normal statistics resulted in underestimates for all three trace elements. The LOGREGR approach also worked reasonably well.

Soil sediments can be expected to be geochemically similar to surrounding soils, except that where water is working continuously on the soils, additional leaching of trace elements can be anticipated. This would mean that sediments should have maxima typical of background soils, but have a wider range of values at the lower end, with a consequent lowering of the average (logarithmic mean) property.

In fact, Table 2 shows this to be the case. Again, the distributions of these three trace elements in soils are seen to follow a logarithmic pattern. The raw data do not conform to the normality assumption, as shown by the correlation coefficient test. The standard deviations of raw data as well as lognormal data are somewhat smaller than for the sediments. However, the mean values are almost twice as high. The maximum observed values are close to the 95% maxima predicted by the sediment data; however, these values are much less frequently occurring (less than .8% of the time) and are not directly comparable.

The borehole data can be inspected more closely for distribution patterns, since the sample size is much larger (greater than 100). The correlation coefficient test value (.985) is correspondingly much tighter than for sediments (.945). The HELSEL method again shows its superiority in generating distribution patterns much more regularly lognormal; the LOGREGR methods do not meet the test conditions although they are shown to be a better fit than the normal distribution.

The reasons for the discrepancies between the LOGREGR and HELSEL methods can be seen in looking at histograms of the raw and log-transformed data. The accompanying figure presents the histograms for borehole aluminum raw (1) and logarithmic data (2), arsenic raw (3) and logarithmic (4) data, vanadium raw (5) and logarithmic (6) data. (7) shows the histogram for the HELSEL logarithmic arsenic data, while (8) is a similar histogram for vanadium HELSEL data. The initial logarithmic histograms identify the 1/2 detection limit data (shown as BDL Values). Trace element histograms (1), (3), and (5) clearly show the skewed character of the raw data.

It becomes apparent from both linear regression analysis of residuals and these figures, how seriously assumptions about BDL values can affect the shape of the logarithmic distributions. For untransformed or raw data, BDL values fall within a very small limit between zero and the highest BDL, which is a minor part of the overall data ranges. Consequently, assumptions about their exact value are not that important and do not seriously affect mean and standard deviations unless the percentage of BDL values is high.

However, in the logarithmic case, the situation is reversed. In the logarithmic domain, BDL values can take a range from the logarithm of the highest DL to negative infinity. With a large number of non-detects, the distributions can become highly skewed negatively, or as in the case of the arsenic, form almost a bimodal distribution.

Since the use of 1/2 the detection limit is an arbitrary approach which works reasonably well with the raw data, another approach which better distributes the BDL values is just as appropriate. The HELSEL method improves the distributional characteristics because it generates values to conform to the lognormal distribution of the above detection values. The smooth shape of the resulting distributions can be seen by comparing (4) with (7) for arsenic, and (6) with (8) for vanadium.

Aluminum borehole LOGREGR and HELSEL results were not compared with histograms, since only the change of a single data point occurred. Yet, even the single change in 1 of 122 points was found to be very significant. In Table 2, the LOGREGR method for aluminum fails the correlation coefficient test using the 1/2 detection limit value (.983 versus the .985 identified level); a very high residual was noted. When this data point is transformed via the HELSEL method, the resulting coefficient test (.994) does meet the lognormal distribution criterion.

Table 2 derived statistical means and standard deviations are used to compare extreme observed values. To compare the predicted 95% tolerance levels, the extrapolated observed values from the original data sets at the 95% level are generated. Results show that the HELSEL and LOGREGR methods generate tolerance levels which include the observed 95% concentrations within their limits for all three trace elements.

To see whether the observed maxima in the three data sets are also included, a comparable tolerance level would need to be generated for the 99.2% range. However, such tabular information is not currently available. As a substitute, a t-statistic for the 99.5% level was used to estimate this concentration. It must be recognized that the predicted t-level is an <u>average</u> value for that confidence level. By contrast, the tolerance interval predicts that 95% of the time, 95% of the values will be included within the tolerance interval. Thus we should expect that the t-statistic generated should be near the observed value, but not necessarily higher than it.

This is shown to be exactly the case for the aluminum data. The predicted t-value is between 39145 (HELSEL) and 41357 ppm for the LOGREGR method. The actual value is 40800 ppm. The t-statistic slightly overpredicts the vanadium concentration (82.58-96.0 ppm) versus the observed 70.0 ppm. It is with the arsenic maximum that the use of the logarithmic distribution may show some limitations. The predicted 99.5% maxima (18.0-18.4) ppm are not close to the observed level of 41.7 ppm.

It is possible that the arsenic value is an analytical error. However, the sediment data also showed a similar pattern, with many low values and a single very high value. It is alternatively possible that arsenic may infrequently occur in much higher concentrations in soil than can be adequately described by a logarithmic distribution. Data sets with extremely long tails may not be best described by a logarithmic function for extreme values. However, the results in the borehole table do show that the 95% confidence levels are reasonably well predicted.

The preferred approach might be to utilize the lognormal distribution, but recognize that occasional extreme values might occur. It is noted that the extreme arsenic value in the borehole data did not occur in conjunction with any other maxima of 27 trace elements simultaneously measured. In this situation, one could disregard occasionally high values, but still test for average or 95% level values using lognormal statistics.

The results here show the general superiority of using the HELSEL approach in generating lognormal statistics. A leap of faith is first necessary, in making the assumption that soils trace element background data will be best described by the lognormal distribution. Although other skewed distributions are available, the lognormal statistics have the virtue that many of the well-developed tests based on the normal curve can be applied to lognormal data.

For example, comparisons of background to cleanup soils using composite samples can be made. Composite samples average the mass of materials from separate locations, and hence are best described by the arithmetic mean; confidence intervals would be uneven, since the original soils background data are shown to generally follow the logarithmic pattern. The arithmetic mean can be estimated from lognormal data, using the short formula:

$$\bar{X} = e^{(\bar{y} + .5s_Y^2)},$$

where $\bar{X}$ is the estimated arithmetic mean, $\bar{y}$ is the logarithmic mean, and s_Y is the logarithmic standard deviation. Gilbert[6] describes a method for generating a confidence interval for an arithmetic mean of size N for logarithmic data that uses the mean and standard deviation of the logarithms. As a quick check of the statistics, the estimated arithmetic mean was derived from the data and compared with the original mean calculations. In all cases using either LOGREGR or HELSEL statistics, results were within $\pm$5% for the borehole data set and $\pm$13 % for the smaller sediment sets. The data are well behaved and able to be used effectively.

Helsel's method is shown to be useful even for data sets with only limited numbers above detection. The arsenic sediment data was able to be converted to logarithmic data with only 21% of its data above detection. It is recognized that this estimated distribution has a very wide variance; however, this is due in part to the small sample size. The arsenic data for sediments were shown to compare reasonably well with the much larger borehole data set. Even using an alternative test of proportions as suggested in the Rocky Flats study, comparable testing sample sizes would be required to make a comparison. Also, individual data could not be compared.

Attachments

---

6 Statistical Methods for Environmental Pollution Monitoring, Richard O. Gilbert, Van Nostrand Reinhold Company, New York, 1987, pp. 169-170

HISTOGRAMS FOR BOREHOLE DATA-- ROCKY FLATS BACKGROUND STUDY

1. ALUMINUM- Raw Data

```
Histogram of Al-Cor   N = 122

Midpoint   Count
      0        3  ***
   5000       39  ***************************************
  10000       50  **************************************************
  15000       13  *************
  20000       11  ***********
  25000        3  ***
  30000        2  **
  35000        0
  40000        1  *
```

2. ALUMINUM-- Logarithmic Data

```
Histogram of lAl-Cor   N = 122
Each * represents 2 obs.

Midpoint   Count
    6.5        1  *  } BDL VALUE
    7.0        0
    7.5        1  *
    8.0        6  ***
    8.5       18  *********
    9.0       53  ***************************
    9.5       25  *************
   10.0       15  ********
   10.5        3  **
```

3. ARSENIC- Raw Data

```
Histogram of As-Cor   N = 122
Each * represents 2 obs.

Midpoint   Count
      0       52  **************************
      5       61  *******************************
     10        8  ****
     15        0
     20        0
     25        0
     30        0
     35        0
     40        1  *
```

4. ARSENIC- Logarithmic Data

```
Histogram of lAs-Cor   N = 122

Midpoint   Count
   -0.5        1  *
    0.0       40  ****************************************} BDL VALUES
    0.5        3  ***
    1.0       34  **********************************
    1.5       26  **************************
    2.0       14  **************
    2.5        3  ***
    3.0        0
    3.5        1  *
```

5. VANADIUM- Raw Data

```
Midpoint   Count                              N = 117
      5        7  *******
     10        3  ***
     15       14  **************
     20       24  ************************
     25       27  ***************************
     30       10  **********
     35        8  ********
     40        6  ******
     45        8  ********
     50        2  **
     55        3  ***
     60        3  ***
     65        1  *
     70        1  *
```

6. VANADIUM- Logarithmic Data

```
Midpoint   Count                              N = 117
    1.6        3  ***   } BDL
    1.8        4  ****  } VALUES
    2.0        0
    2.2        0
    2.4        3  ***
    2.6        3  ***
    2.8       14  **************
    3.0       18  ******************
    3.2       29  *****************************
    3.4       11  ***********
    3.6       12  ************
    3.8       11  ***********
    4.0        6  ******
    4.2        3  ***
```

7. ARSENIC- Fitted Helsel Log Data

```
                 N = 122

Midpoint   Count
   -1.0        1  *
   -0.5        4  ****
    0.0       13  *************
    0.5       26  **************************
    1.0       34  **********************************
    1.5       26  **************************
    2.0       14  **************
    2.5        3  ***
    3.0        0
    3.5
```

8. VANADIUM- Fitted Helsel Logarithmic Data

```
Midpoint   Count                              N = 117
    2.0        1  *
    2.2        2  **
    2.4        7  *******
    2.6        3  ***
    2.8       14  **************
    3.0       18  ******************
    3.2       29  *****************************
    3.4       11  ***********
    3.6       12  ************
    3.8       11  ***********
    4.0        6  ******
    4.2        3  ***
```

Table 1. COMPARISON OF STATISTICAL PROPERTIES OF AL, AS, & V IN OVERALL SEDIMENT DATA-- ROCKY FLATS BACKGROUND STUDY

STATISTICAL PARAMETERS	ALUMINUM			ARSENIC			VANADIUM		
	BASIC STATS	LOGREGR	HELSEL	BASIC STATS	LOGREGR	HELSEL	BASIC STATS	LOGREGR	HELSEL
Sample Size	18	18	All ADL[1]	19	19[2]	19 (4)	17	17	17 (9)
Arithmetic Mean Standard Dev.	5204 5099			2.376 2.913			15.42 11.70		
Log Mean Log Std Dev	8.123 1.017	8.12 1.039		.513 .732	.513 .658	-.316 1.550	2.506 .683	2.51 .686	2.53 .719
Geometric Mean Geom Std Dev	3371 2.765	3361 2.83		1.67 2.079	1.67 1.931	.729 4.71	12.26 1.980	12.30 1.986	12.55 2.052
Regression Equation Ln(C)=	-	8.12+1.039*Z		-	.513+.658*Z	-.316+1.55*Z	-	2.51+.686*Z	2.53+.719*Z
Correlation Coefficient 95% SIGNIF?[3]	.878(N)[4] NO	.982 YES		.718(N) NO	.865 NO	.99(.984)[5] YES	.888(N) NO	.962 YES	.99(.951) YES
95% Tolerance Interval[6] X + K*s	17880(N)	44487		9.43(N)	13.97	33.3	44.52(N)	76.51	84.87
Max Data Value	21600			13.0			50.2		

1 The HELSEL method uses regression analysis only on detectable (ADL) values; where all values ADL, the results are identical to LOGREGR

2 The LOGREGR (logarithmic regression) uses below detection limit data as 1/2 the DL

3 Correlation coefficient significance test between ordered concentration values (C) and Z-values: $c_{.05,\ 18} = .945$

4 (N) indicates correlation coefficient and tolerance tests on raw data (normality assumption)

5 Correlation coefficient test after fitted BDL values added; value in parentheses are for ADL values only

6 95% Tolerance Interval levels taken from "Statistical Analysis of Groundwater at RCRA Facilities", EPA, October 1988: $K_{18} = 2.543$; $K_{19} = 2.423$; $K_{17} = 2.486$

Table 2. COMPARISON OF STATISTICAL PROPERTIES OF AL, AS, & V IN OVERALL BOREHOLE DATA— ROCKY FLATS BACKGROUND STUDY

STATISTICAL PARAMETERS	ALUMINUM			ARSENIC			VANADIUM		
	BASIC STATS	LOGREGR	HELSEL	BASIC STATS	LOGREGR	HELSEL	BASIC STATS	LOGREGR	HELSEL
Sample Size	122	122	122 (121)[1]	122	122[2]	122 (78)	117	117	117 (110)
Arithmetic Mean	10658 ppm			3.56			27.25		
Standard Dev.	6326 ppm			4.17			13.38		
Log Mean	9.116 ppm	9.12	9.12	.951	.951	1.01	3.176	3.18	3.21
Log Std Dev	.581 ppm	.577	.556	.759	.742	.727	.540	.529	.460
Geometric Mean	9100 ppm	9136	9136	2.59	2.59	2.75	23.95	24.1	24.8
Geom Std Dev	1.79 ppm	1.78	1.74	2.14	2.10	2.07	1.72	1.70	1.58
Regression Equation Ln(C)=	-	9.12+.577*Z	(9.12+.555*Z) 9.12+.556*Z	-	.951+.742*Z	(1.01+.728*Z) 1.01+.727*Z	-	3.18+.529*Z	(3.21+.46*Z) 3.21+.46*Z
Correlation Coefficient	.925(N)[4]	.983	.994(.994)[5]	.693(N)	.969	.993(.975)	.966(N)	.970	.995(.994)
95% SIGNIF?[3]	NO	NO	YES	NO	NO	YES	NO	NO	YES
95% Tolerance Interval[6]									
X + K*s	22677(N)	27345	26275	11.48(N)	10.60	10.9	52.67(N)	65.7	59.38
95% Data Value	22500			8.05			55.5		
99.5% Student-t Level[7]									
X + t*s	27213(N)	41357	39145	14.5(N)	18.0	18.4	62.26(N)	96.0	82.58
Max Data Value	40800			41.7			70.0		

1 The HELSEL method uses regression analysis only on detectable (ADL) values; where all values ADL, the results are identical to LOGREGR
2 The LOGREGR (logarithmic regression) uses below detection limit data as 1/2 the DL
3 Correlation coefficient significance test between ordered concentration values (C) and Z-values: $c_{.05,+100} = .985$
4 (N) indicates correlation coefficient and tolerance tests on raw data (normality assumption)
5 Correlation coefficient test after fitted BDL values added; value in parentheses are for ADL values only
6 95% Tolerance Interval levels taken from "Statistical Analysis of Groundwater at RCRA Facilities", EPA, October 1988: K= 1.9
7 Student-t one-sided value for N=125, at the .005 level of significance: t = 2.617