

Review of "[Draft] Statistical Evaluation of Additional
Hydrochemical Data from the Saddle Mountains, Wanapum and
Grande Ronde Basalts, Basalt Waste Isolation Project,
Hanford Site"

At the request of Matthew Gordon (Hydrology Section) I have reviewed the subject document for content and form. This draft report is the second of two reports regarding evaluations of hydrochemical data from the Basalt Waste Isolation Project (BWIP). The first report, entitled "Statistical Evaluation of Hydrochemical Data from the Saddle Mountains, Wanapum and Grande Ronde Basalts, Basalt Waste Isolation Project, Hanford Site", was based on data acquired during the DOE/NRC Hydrology Workshop held in Richland, Washington, in July 1982. Additional data have been collected and were provided to the NRC during 1984. The draft document under review incorporates this recent information in updating the previous statistical analyses.

The stated objective of both this and the previous report is to evaluate DOE's assertion that groundwaters in the major geologic units (Saddle Mountains, Wanapum and Grande Ronde Formations) are distinguishable on the basis of hydrochemistry (U. S. Dept. of Energy, 1982, p. 5.1-184 & 202). The report states that the DOE has reached a preliminary conclusion that available data imply the relative absence of vertical mixing of groundwaters among basalt formations at the BWIP site. The potential for vertical mixing must be carefully evaluated because it has great significance with regard to the radionuclide isolation potential of the proposed candidate horizons.

Overall, the approach used in Williams & Associates' report appears to be a very good one both in terms of the scope and depth of the applied statistics. Summary statistics, histograms and box plots were calculated on each numerical response variable, but the results were not included as figures or tables in the draft. It would be helpful in a purely descriptive sense to include these in the final report. The raw data were evaluated for skewness and logarithmically transformed where appropriate. The inclusion of a table of Pearsonian Skewness Coefficients (before and after normalization) would be informative and is recommended. Following Z-score transformation, a correlation matrix (pooled within-groups) was generated. The highest positive correlation shown in the table on page 22 was 0.69 for the parameters Na(+) and Specific Conductivity.

According to the text on page 20, a preliminary cluster analysis was performed based on calculated Euclidian distance coefficients. The resulting cluster diagram is shown on page 23 of the draft. On page 22, opposite the cluster

diagram, is shown the matrix of pooled within-groups correlations. Some readers might thereby conclude that the cluster diagram is based on correlations, not on distance coefficients. It is recommended that the distance coefficients be included in the report in tabular form. They could conveniently fill the upper half of the symmetric correlation matrix shown on page 22 of the draft. Also, the figure description for the cluster diagram on page 23 should state that it is based on distance coefficients. In the present application the use of the Euclidian distance coefficient is fully appropriate as this measure is not constrained within the range from +1 to -1 as is the correlation coefficient and may provide more effective cluster diagrams if some of the variables are very dissimilar from the others.

Following the documentation of the cluster analyses, a confirmatory analysis was performed based on a stepwise MANOVA in order to identify those hydrochemical constituents most responsible for differences among the clusters. Then a canonical analysis was used to identify those factor rotations most aligned with differences among groups. It is noted that the canonical methods are based on the body of theory known most commonly in the literature as Principal Components Analysis and Factor Theory. Results of this preliminary analysis are shown on page 27 of the draft. It is noted that the six clusters mapped in canonical space appear to show spatial relationships as a function of depth. The remainder of the draft document details the simplification of the original functional model and presents interpretations of the hydrogeologic significance of the analyses.

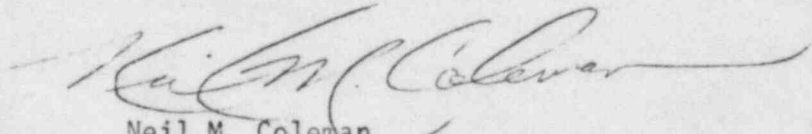
In reviewing the draft document, perhaps the most significant point to be addressed regarding this study is the apparently significant influence of the geothermal gradient on the suite of hydrochemical variables. My own analyses of a similar set of BWIP hydrochemistry data reveal that a number of measured parameters apparently exhibit significant correlations with depth. These parameters include Specific Capacity, Boron(+3), Na(+), F(-), Cl(-), and SO₄(-2). These correlations might simply be the result of increasing mineral solubilities as temperatures rise with depth. As reported on page 7 of the draft report, the geothermal gradient at the Hanford Site has been measured at between 35 and 45 degrees Centigrade per 1000 meters. In the statistical analyses, the net result appears to be a "super-cluster" of geochemical parameters which exhibit reasonably high correlations to each other, primarily as a function of depth/temperature. The term "depth/temperature" is used here because the observed geothermal gradient simplifies the task of estimating temperatures within sampled geologic horizons as a function of depth BLS. The significance of these statements is that mutual correlations of parameters that jointly correlate highly with depth may not be indicative of actual relationships independent of depth. The geothermal gradient seems thereby to impose a vertical spatial relationship on the data matrix. This relationship

may account for the sequential pattern of the clusters mapped in canonical space, as shown in Figure 4.4.1 on page 27 of the draft. One conclusion that can be drawn is that the statistical results should not necessarily be interpreted as distinguishing among basalt formations, as referred to on page 47 of the draft document. Rather, the results appear to be more sensitive to the relative depths of sample groups. My analyses also demonstrated that a cluster of three variables (Ca^{2+} , Mg^{2+} , and Sr^{2+}) exhibited significant joint correlations independent of depth. As pointed out on page 7, para. 1 of the draft document, calcite and dolomite are exceptions to the general trend of significantly increased mineral solubilities under elevated temperatures. Concentrations of the three cations referred to could well be associated with carbonate minerals, which would account for their apparent tendency to cluster away from the depth/temperature controlled "supercluster".

One possible way of dealing with this temperature effect would be to include depths of sampling in the analyses and to obtain correlations between depth and the other variables. The data set provided by DOE included the depths to tops and bottoms of isolated intervals but not the actual levels of sampling. A possibly useful depth estimator would be the midpoint of the sampling interval. Subsequently, for those variables that highly correlate with depth, fitted least-squares lines may be obtained and used to predict isothermal concentrations at a specified temperature. It should be noted that this method would probably not give acceptable results for those parameters that correlate only weakly with depth, but such parameters may not in fact require transformation. Although it may appear to be a somewhat "loose" approach, this technique should provide a reasonably valid means for removing some of the depth/temperature effects. It is probably a better method than evaluating individual mineral solubilities because the latter would be plagued by the complexity of the system, common-ion effects, presence of electrolytes, etc.

My own efforts at evaluating the Hanford hydrochemistry data have apparently been confounded by the same depth/temperature "superfactor". I would be very interested in your comments regarding the relative significance of these proposed temperature effects and would welcome any suggestions you may have for alternatively dealing with this phenomenon. It is very important to note that the effects of increased depth/temperature are an integral part of the data matrix and are not removed by merely excluding depth as a variable from the analysis. The fact that several of the parameters (Ca and Mg) are relatively insensitive to depth due to the low solubilities of carbonates at the prevailing temperatures is encouraging and may provide a starting point for future hydrochemical investigations. We may also discover that a number of parameters which presently seem not to significantly correlate with others may do so if the temperature/depth effects can successfully be removed or reduced. In any event, the draft document appears to thoroughly document, within the

scope of measured parameters, the prevailing hydrochemical regime. In the realization of what appear to be significant thermally-induced phenomena, results of the present study should be quite valuable in evaluating hydrochemical signatures in areas of suspected groundwater upwellings.

A handwritten signature in dark ink, appearing to read "Neil M. Coleman", with a long, sweeping horizontal line extending to the right.

Neil M. Coleman
Hydrology Section
Geotechnical Branch
Division of Waste Management

D RASEY

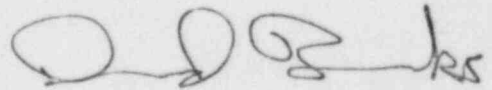
Geochemistry Review of Willimas et al., Draft Report

"Statistical Evaluation of Additional Hydrochemical Data..."

In general, this report concludes that "...statistical procedures can distinguish among basalt formations at the BWIP site when applied to hydrochemical data." However, the following points should be noted:

1. p. 7 and tables 2.1 and 2.2: The geochemistry of the basalt rocks (as detailed in the tables) is filler. Unless these data are going to be used, they should be deleted. If possible, some work should be done to see if the solubility of rock minerals (under the geothermal gradient at Hanford) accounts for the distinctive groundwater characteristics between flows.
2. p. 15 and p. 46: The inability to perform an ion balance is a major short coming of the data base and thus the conclusions reached in the report. It is my understanding that an ion balance provides an approximation of the quality of the groundwater-chemistry analysis. If the balance is not within 5 to 10%, then there could be significant analytical problems which could lead to false conclusions.
3. a. pp. 47 and 48/Discussion: The discussion section of the report should begin with an analysis of each borehole and indicate whether mixing is suggested on a borehole by borehole basis (as I remember, the borehole data is ambiguous-i.e., see pp. 38-29).

- b. Next, the statistical analysis of the complete hydrochemistry data set should be summarized, and
- c. a discussion should be included on the sources of uncertainty and its impact on the conclusion. My reading of this report suggests that uncertainty is significant.

A handwritten signature in dark ink, consisting of several loops and a final vertical stroke, followed by the letters 'RS'.