Battelle
Pacific Northwest Laboratories
P.O. Box 999
Richland, Washington U.S.A. 99352
Telephone (509) 376-4503
Telex 15-2874

June 23, 1980

Mr. Gregory Eadie
Uranium Recovery Licensing Branch
U.S. Nuclear Regulatory Commission
Mail Stop #48355
Washington, DC 20555

Dear Greg:

This letter and attachments are my response to the seven comments in your letter to me dated June 2, 1980.

Item No. 1 is guidance for further statistical analyses. We will follow your directive and not perform such analysis.

Your letter states that samples were predominantly collected in hot spots of the Rio Puerco. The sampling protocol calls for samples to be collected every 1000 feet along the arroyo, not for the samplers to locate and sample possible hot spots. The three types of pool samples could be considered samples of hot spots but it should be remembered that the pool samples were taken only when a pool was located at a regular sampling site. For all other types of samples, the terraces, etc., samples were taken as prescribed by the protocol.

Item No. 2 requests more information on distance measures. United Nuclear Corporation provided us with Rectified Photobase maps of the Rio Puerco from the mill site through the city of Gallup. These maps show one "sector" of land each and are on a one inch equals two hundred scale. The location of each of the survey stakes used for establishing sample sites is indicated on these maps. The survey stakes were placed at approximately 500 foot intervals along the uncut bank of the arroyo. Using these maps we measured the distance along the bottom of the arroyo, that is, along the watercourse, between successive sampling sites. These distances are reported as distance in meters from stake number zero. Stake zero is placed at the point where the spill entered the Pipeline Arroyo. Because of the extreme meandering of the arroyo these streambed distances are not well correlated with the distances between stakes which are on the uncut banks of the arroyo. The extremes of streambed distances perpendicular to two successive stakes (500 feet apart on the uncut bank) are zero and 362 meters (1188 feet). Such variation is due to the physical geography of the arroyo and does not suggest inaccurate work by the survey or sampling crews. The actual streambed distances are a more meaningful measure for hydrologic, chemical and sedimentation rate considerations than are the survey stake distances. Because these maps are available only for a portion of the arroyo sampled, we had to discontinue

Battelle

these distance measures at stake number 305. We do have copies of the EPA aerial photographs of the entire arroyo, however these photographs do not indicate where stakes are located and thus cannot be used for measurements.

For that portion of the arroyo for which we were able to measure distances, from stake number zero through stake number 305, the data listing contains both stake number and distance in meters from stake zero. For the remaining sampling we can provide only the stake numbers, this holds for the samples collected from stake 307 through stake number 491.

Item No. 3 asks for mathematical equations describing technical terms used. The data was first segregated by type of sample and sampling site so that all the "first terrace" data was analyzed independently of all other types, etc. Then the replicates at each sampling site were averaged. The counting error is a measure of the accuracy of the corresponding level determination, the more accurate ones should be considered more important than the less accurate sample values. The use of a weighted average accomplishes this adjustment for relative accuracy. The statistically standard weighting factors are the inverse of the variances (counting errors squared). Let $X_1, \ldots, X_N$ represent the replicates for one type of sample at one sampling point and $S_1, \ldots, S_N$ represent the corresponding counting errors. Then the weighted mean is:

$$M = \frac{\sum_{i=1}^{N} X_i/S_i^2}{\sum_{i=1}^{N} 1/S_i^2}$$

and the corresponding standard error of the mean is

$$S.E. = \frac{1}{\sum_{i=1}^{N} 1/S_i^2} .$$

Under the assumption that at each sampling point the replicates are sampled from the same statistical population, it is inappropriate to include a between samples term in the standard error. For some uses of the data, it may be desirable to add such a term to the standard error formula.

Plots of the weighted averages were not significantly easier to interpret than plots of the raw data. We therefore applied a simple smoothing or filtering algorithm to the data. Note that the weighted averages were applied to the subsets of the data consisting of each type of sample at each sampling site. The smoothing is an analog of averaging, again for each type of sample, over a few successive sampling sites. One of the simplest low pass filters of time series analyses was used, called a "First Order Lowpass Filter". It is computed by:

Battelle

$$Y_j = (1-a)Z_j + aY_{j-1}$$

where Z_j is the series of data values to be filtered, and a is the filter factor. A filter factor of 0.25 was chosen and used to filter both the weighted means and standard errors of the means. This value of a filtering factor was determined by repeatedly evaluating the data for the first 10,000 meters of arroyo with various filter factors, then choosing a value that just averages out noise but not any patterns. Note that this filtering scheme is very easy to compute. It uses data from a few previous data points with an exponential weighting scheme. Data points more than 5 to 8 prior to the current point are essentially "swamped out" or weighted almost zero so they no longer influence the computations.

Plots of the filtered weighted averages only have already been sent to you and were named "running averages" to distinguish them from another type of plot also sent to you. After the filtering of the weighted means and weighted standard errors, twice the standard error was added and subtracted from the means to give an approximate 95% confidence interval. These means with confidence intervals are identified as smoothed averages.

We realize that much more elegant, and time consuming, methods are available for smoothing or filtering and averaging data. Since we were under the constraint of maximum economy we believe that this averaging and filtering is justified and adequate for the display and visual interpretation of the Rio Puerco data.

Item No. 4 asks for a combination of the plots we have already sent to you. When first developing the plots, we attempted to produce plots such as you desire. We found that such plots contained so much information that they were confusing and very difficult to use. We therefore chose the option of using a large number of simple plots that can be quickly evaluated visually. When one uses complicated plots, errors are frequently made and errors are particularly detrimental to any regulatory or enforcement actions. You mention plots " ... over the entire 50,000 meters of the Rio Puerco." The 50,000 meters is only the distance of the arroyo for which we have maps suitable for measuring distances, this was discussed under item No. 2. Assuming that the average distance between survey stakes is constant, if there are 50,000 meters in 305 stakes then there would be about 77,000 meters in 491 stakes. We have data through stake number 491 which is at the Arizona-New Mexico border, and the arroyo continues into New Mexico for some distance unknown to us. It is our understanding that the Rio Puerco joins drainage systems that eventually drain into the lower Colorado River basin and thus eventually reach the Gulf of Mexico. Conversations we have had with Dr. Thomas Buhl of the State of New Mexico suggest that he has determined that the spill from the Church Rock mill incident was completely absorbed into the streambed by the time the spill reached the region of the Arizona-New Mexico border.

Battelle

Item No. 5 asks for more information on sampling and analytical conditions. Some of these questions concern the field sampling conditions, the sample preparation, and instrumentation procedures; these will be covered in Dr. Weimer's report. The background samples were treated as a separate group in the data plots that you already have; they were not averaged in with anything else. The simple mean value for the thorium levels in the 80 background samples in the entire data set is 0.8 picocuries per gram, with a between samples standard deviation of 11.2 picocuries per gram.

Item No. 6 requests a summary table of all data. An improved version of the table we already sent you is attached. The previous table was simply a listing of the data cards. The new table has the columns of data separated for easy reading, and their order is somewhat rearranged. The headings at the top of the table are somewhat cryptic because of space limitations. "Stake No." refers to the survey stakes discussed in previous sections of this letter, it is the primary number used for sample identification. The "Distance, Meters" column gives the distance from stake zero as described in item No. 2 of this letter. The sample type and replicate columns contain those codes used by the State of New Mexico for coding data; the types are identified in the following table.

Type	Identification
1	first terrace
2	second terrace
3	concentrated data
4	pool center
5	pool discharge point
6	background
7	core near stream
8	core near cut bank
9	associated salts

Sample replicate refers to the replicate number of the sample, first sample, second sample, etc., with two exceptions: for the core samples the replicate refers to the first foot, second foot, etc., of a core.

The isotope numbers are left off the name columns, the full headings should be thorium-230, radium-226, lead-210, and uranium-235/238. The date of sample collection is not included because this information was not provided to us; all samples were collected with two weeks of September 25, 1979, the day sampling commenced. The date of sample analysis was provided to us but has been eliminated as you requested in our telephone conversations.

There seems to be some concern about the negative values of isotope concentrations given in the data listing. Often such values are eliminated since no real concentration can be negative. They are an artifact produced by the necessity of subtracting our instrumentation background levels from sample measurements. We prefer to include the negative values in the data set

Battelle

so that others may work with the data and get unbiased statistical results. Any analyses of data with negative results removed with produce statistically biased, and therefore not very useful, results; however, if the negative values are left in anyone can work with the data in a statistically valid manner.

Item No. 7 requests an evaluation of the newly proposed cleanup verification sampling scheme outlined in Dr. Buhl's letter. This evaluation is attached as an addendum to this letter.

Respectfully submitted,

Robert R. Kinnison

R. R. Kinnison
Senior Research Scientist
Statistics and Materials Safeguards

RRK/sp

Enclosures

cc: RW Perkins
WC Weimer

Evaluation of Revised Rio Puerco Cleanup Criteria

The revised criteria, as proposed by Dr. Thomas E. Buhl in his letter to Mr. Hubert Miller and dated April 23, 1980, are statistically simple and easy to apply. I see no substantial statistical deficiencies in these new criteria. However, since these criteria are stated in a somewhat non-standard manner, I do believe some discussion of their attributes is called for. The following paragraphs present this discussion.

The use of a 67% confidence level rather than the commonly used 95% level should be noted. The choice of confidence level is arbitrary and there is no statistical or mathematical reason for any particular level. With the range of thorium data values and counting errors stated in Dr. Buhl's letter, a

67% confidence level with a limit of 60 picocuries per gram is approximately equivalent to a 95% confidence level with a limit of 80 picocuries per gram. The use of a one sided upper confidence level, as indicated by the addition of the standard error to the mean but no subtraction, is standard and appropriate for this type of situation. It should be noted that dividing the standard error by 2 yields a 69% rather than a 67% confidence level.

The new criteria do not place any control on the sample size to be used. This results in the situation of no control of the second type of statistical error, that of concluding that radioisotope levels are within the limit when in fact they are high. The 67% confidence level controlls the first type of statistical error, that of concluding that radioisotope levels are high when in fact they are within the limits. Because a large sample size results in a small standard error, a large sample allows the mean values to be very close to the limit of 60 picocuries per gram. Another way of looking at this sample size situation is to note that more information (samples) are needed to make a decision when the mean values are close to the limit than when mean levels are much lower (or higher) than the decision point. Small samples, on the other hand, would allow many of the second type of statistical error, that of concluding a mean value is less than the limit when in fact it is slightly higher.

The use of an arbitrary level of radioisotope as the limit or decision point, 60 picocuries per gram for thorium etc., eliminates the need for any information about background levels in the decision process. A more common statistical procedure would be to state some allowable deviation above background levels, thus including the background information in the statistical analysis. There is nothing statistically invalid or strange about not using the background levels, its just not a common way of analysing data. Considering the fact that there was much less background data collected in the Rio Puerco sampling program than radioisotope data, the use of an arbitrary radioisotope level in the decision rule is perhaps a statistically preferred scheme because the lack of background information at all sampling points would require the use of some sort of interpolation scheme.

Our work with the available maps indicates that the survey stakes are not an accurate indication of streambed distances, and we have such maps for only the first 305 of the 491 numbered stakes. Since it would be a major task to identify 1000 foot sections on the maps or to return to the arroyo for measurements, and since no maps with stake location exist for the arroyo below stake number 305, I recommend that the area criteria be stated in terms of stakes rather than feet. I cannot see any reason for not allowing a 20% to 50% error in area included in an average, as would be the case if measurements based on stakes are used, and such would substantially reduce the data analysis effort required by eliminating the need to measure areas. My discussion with the personel involved in this study seems to indicate that 500 feet of the arroyo is conceived to be equivalent to the distance between two stakes. Since this equivalence will not be perceived by others and in fact there is no real equivalence, I recommend that the 1000 foot criteria be changed to the distance between alternate stakes.

The criteria speak of using 1000 foot sections of the arroyo for averaging purposes. There is no control of how this section is to be chosen. Most of the samples were taken only at the odd numbered stakes. This results in two distinctly different possible situations for computing area averages. If the areas are chosen to be between alternate even numbered stakes, as would occur if one starts with stake number zero, then data would be available only at the intermediate odd numbered stake. In this case the averages for that odd numbered stake would be used as the average for the entire area. However if the areas are chosen to be between alternate odd numbered stakes, then data exist only at the boundaries or ends of the area and an average of the two boundary averages would be used. A third alternative is to use overlapping areas starting with each stake, using both kinds of area averages mentioned above. This third alternative would be statistically troublesome since the two kinds of averages have different characteristics and distributions. Using an area criteria of 1000 feet rather than alternate stakes would be equivalent to this third alternative because of the lack of association of stake numbers with actual arroyo streambed distances. The criteria should specify how the areas are to be chosen so that conflict cannot arise from different investigators using different methods. Sampling at the midpoint of a section is the easiest computationally, and a good way from the standpoint of statistical work.

The criteria seem to not use the information contained in the counting error data. The use of a simple standard error of the mean indicates that one is to calculate a standard deviation of the several replicated for each type of sample at each collection point and divide by the square root of the sample size. This is, in statistical terms, a between samples error measurement. When this is used alone the within samples error information contained in the counting error data are not used. It is possible to use both the within and the between samples errors in a rather elegant manner to calculate a total standard error. This total standard error can then be used to determine the 67% confidence level as before. The counting error is a measure of the accuracy of the corresponding sample value, the more accurate ones should be considered more important than the less accurate sample values. A statistical way of accomplishing this weighting is to use the inverse of the variances (counting errors squared) as the weighting factors. For the weighted means and corresponding total standard errors, advanced statistics text give the following formulas:

$$\text{mean} = \frac{\sum y/v}{\sum 1/v}$$

$$\text{standard error} = \sqrt{\frac{1}{1/v} + (\text{B.S.E.})^2}$$

where the y's are the data values, the v's are the corresponding variances (counting errors squared), and B.S.E. (between samples standard error) is the common standard deviation calculated from the data values then divided by the square root of the sample size.

I recommend the use of these formulas to calculate weighted means and standard errors because they make use of the counting error information to adjust for the variations in accuracy between replicate samples, as well as including both between sample and within sample errors. These formulas are simple to use, even though their derivation is complicated.

Respectfully submitted,

Robert Kinnison, Ph.D.

Robert Kinnison

Battelle Northwest Laboratories