N.A. WATER SYSTEMS

This Submittal Delivered by Email Only

October 17, 2008

Ref. No: 56007746 GE Church Rock Project

Mr. Mark Purcell
Remedial Project Manager
U.S. Environmental Protection Agency
1445 Ross Ave., Suite 1200 (6SF-LP)
Dallas, TX 75202-2733

Re: Revised Submittal
Calculation of Background Statistics with Comparison Values
UNC Church Rock Mill & Tailings Site, Church Rock, New Mexico

Dear Mr. Mark Purcell:

N.A. Water Systems (N.A.WS) is pleased to provide this revised report regarding the calculation of background water statistics with comparison values for the UNC Church Rock Mill & Tailings Site in Church Rock, New Mexico. This report includes revisions to the August 26, 2008 submittal based on comments received from Dennis Beal of Science Applications International Corp. (Beal, SAIC, email communication, Sept. 19, 2008), and other reviewers (Mark Purcell, EPA, and Earle Dixon, NMED) during a teleconference of September 30, 2008.

Introduction

Calculations of background statistics have been completed for the Church Rock project. These calculations were made using results for COPCs (Constituents of Potential Concern) in samples collected from July 1989 through October 2007, inclusive. Similar calculations were made for trace and major metal results obtained from samples collected from May 1988 through April 1989, inclusive. Methods used to calculate the statistics were consistent with those discussed by and agreed to in the teleconference on June 27, 2008. The teleconference participants included representatives of U.S. Environmental Protection Agency (EPA), New Mexico Environment Department (NMED), and N.A.WS. The methodology agreed to in that meeting is summarized by the following steps:

1. Use ProUCL software to estimate the upper confidence limits (UCL95) for the means of background populations of COPC concentrations from samples determined to be representative of background groundwater quality. (Background sample sets for the Southwest Alluvium and Zone 1 were established in the February 2006 license amendment request for changing the Groundwater Protection Standard (GWPS) for combined radium. Determination of background sample sets for Zone 3 and for the older trace metal data are presented below).
2. Determine which COPCs have higher median concentrations in background groundwater than the comparison values (these are presented below). The method of testing recommended during the June 27 teleconference by the expert consultant to the EPA (Dennis Beal of SAIC) was the single sample hypothesis test. Of the three nonparametric methods available in the ProUCL software, he recommended that the Wilcoxon Signed Rank test be used, rather than the Sign Test or the Test of Proportions.
3. Select for consideration as potential modifications to cleanup levels those background UCL95 statistics associated with COPCs that are determined (from single sample hypothesis tests) to have median concentrations equal or exceeding the comparison values. The selected UCL95 statistics (if adopted) would be single-valued standards that will be representative of background UCL95 (i.e., upper confidence limit on the mean at the 95% confidence level). Note that the UCL95 statistics presented in this document as candidates for consideration as modifications to cleanup levels are based solely on statistical calculations.

One of the conclusions of the June 27 teleconference was that the preferred method of comparing site samples to revised background-based cleanup levels is two-sample hypothesis testing (e.g., of a compliance data set against the background data set from which the revised cleanup level was derived). Therefore, one of the objectives of the current work is to define appropriate background data sets for those future comparisons.

Identification of Samples Representative of Background Groundwater Quality

The process used to identify samples representative of background groundwater quality was identical to that described in the license amendment request for changing the GWPS for radium (N.A. Water Systems, February 2006, *Technical Analysis Report in Support of License Amendment Request for Changing the Method of Determining Exceedances of the Combined Radium Groundwater Protection Standard in Source*

Materials License SUA-1475 (TAC LU0092), Groundwater Corrective Action Program, Church Rock Site, Church Rock, New Mexico, pp. 3-6). As such, the wells selected for the purposes of this report as having samples representative of background quality in Zone 1 and in the Southwest Alluvium are the same as those identified in the February 2006 report. One difference is that the data sets used in calculations made for this report are from the period July 1989 through October 2007 inclusive, whereas the February 2006 submittal only included samples collected through October 2005. The methods used to identify wells having background water quality for the February 2006 submittal were used to verify that the designation remained valid for samples collected through October 2007. Table 1 lists wells and sample dates representative of background.

A second difference from the February 2006 report is that the current calculations have been applied to all COPCs, as well as a group of trace metals (plus iron) that had previously been dismissed as COPCs (EPA, *August 1988, Draft Final Remedial Investigation, United Nuclear Church Rock Site*). The inclusion of former COPCs, which had not been a part of the site's approved Sampling and Analysis Plan, required that the much smaller pool of pre-plan (pre-July 1989) sample results be included in the statistical calculations for these metals. Well samples collected from May 1985 through 1989 (including those from the Remedial Investigation sampling) were evaluated for evidence of background water quality. This resulted in the addition of sample results for metals from wells GW-4 and 623 for the Southwest Alluvium and from well EPA-5 for Zone 1. Time series graphs of indicator parameters for wells GW-4, 623, and EPA-5 are included in Appendix A. Metals results from other background wells (identified in the February 2006 report) were also verified to have come from samples representative of background water quality. In other words, those wells identified as having background water quality subsequent to July 1989 (see Table 1) were found, as expected, to have had background water quality prior to July 1989.

Zone 3 groundwater was not a subject of the February 2006 report; therefore, the identification of samples representative of background water quality in Zone 3 is new to this report. Table 2 lists wells and sample dates representative of background in Zone 3. The following criteria have been used to distinguish background versus impacted groundwater quality in Zone 3:

- pH < 5 and bicarbonate < 100 and > 500 mg/L are useful (but not always definitive) indicators of seepage impact (e.g., see discussion of these empirically derived criteria in the 2007 annual review report). See Figure 1 for box-and-whiskers plots of bicarbonate and pH for the background wells.

- Time-series of these two indicator parameters are very helpful, sometimes essential. See Appendix A for time-series of pH and bicarbonate for the background wells.
- Well locations within the overall area impacted by seepage (e.g., see Figure 35 in the 2007 annual report).
- Time trends in the concentrations of major ions; in particular, decreasing ratios of Ca:Mg are associated with degrading groundwater quality (see Appendix B in the 2007 annual report; e.g. well EPA-14).
- Time trends in the concentrations of many metals and radionuclides will usually increase as the water quality degrades in Zone 3 (see Appendix B in the 2007 annual report; e.g. well EPA-15).

Invariably, some wells (or certain time spans at some wells) are difficult to classify because their groundwater chemistry tends to be gradational. For example, during the period of time of relevance for present purposes (1989 to October 2007), the geochemistry associated with well 420 is "borderline" – therefore, we have excluded it from the dataset associated with background water quality.

The time-series included in Appendix A show the inferred dates of the onset of seepage impact for those wells whose sampling regime spanned such a transition. Also shown on time series spanning the date is the May 3, 2000 transition to low-flow, unfiltered sampling from multiple-well-volume, purge-and-filter sampling. This transition date does not coincide with any of the inferred onset dates of seepage impact. However, May 2000 appears to coincide with changes of indicator parameter trends at two wells, EPA-5 and EPA-14. It is unlikely that the change of sampling method initiated the multi-year concentration trends noted at these two wells. Other groundwater parameter changes, post-filtration, can be gleaned from a review of the tabulated historic water quality data in the appendices of the 2007 annual review report (N.A. Water Systems, 2008), and these changes cannot be ascribed to the absence of field filtration.

The background sample sets used to make the current calculations have been revised by the removal of small numbers of sample results having unusually high reporting limits. This culling of data affected the sample data sets for each hydrostratigraphic zone. However, it involved less than approximately one percent of the sample data (typically no more than three data points for a particular COPC) and a relatively small number of COPCs. These data were removed because they were discovered to have undesirable consequences on the results of the single sample hypothesis tests (particularly with the recommended Wicoxon Signed Rank Tests). The causes of these problems and the rationale for removing these data were discussed with Dennis Beal by James Ewart. These problems and our solution for them are described below in the discussion of the results.

Results

Basic Statistics and Upper Confidence Limits for Means

Tables 3 through 5 list basic statistics for all COPCs and additional metals calculated from the background data sets from wells in the Southwest Alluvium, Zone 1, and Zone 3. The data sets include only primary samples (i.e. no QA/QC samples). Also listed are upper confidence limits at the 95% confidence level for the means (UCL95). All of the statistics were calculated using ProUCL software (Singh et al., April 2007, *ProUCL Version 4.00.02 User Guide*, EPA/600/R-07/038). The UCL95 estimates were selected from values recommended by the ProUCL software. Summary tables of the output of UCL95 estimates are provided in Appendix B. In cases where two alternative estimates of UCL95 statistics are provided by ProUCL, the higher value was selected and is listed in Tables 3 through 5. The higher values were selected as conservative estimates, consistent with the use of these same statistics as estimators of exposure point concentrations (EPCs).

Comparisons of Background COPC Concentrations with Comparison Values

Table 6 is a compilation of site cleanup goals (EPA, *September 1988, Record of Decision for the Church Rock Site*) at Church Rock and other more recently developed information sources and standards for COPCs and metals. Comparison values were selected from Table 6 (see the green highlighted values) in consultation with Mark Purcell (EPA). Tables 7 through 9 summarize the background concentrations versus comparison values for COPCs and metals in each of the three hydrostratigraphic zones. The results, as presented, deviate in one significant way from the methodology described in the three steps outlined in the Introduction. The selection of candidates for consideration as new background-based cleanup levels (shown in the last column of each table) was based solely on the estimated UCL95 statistics and comparison values (see column 6, titled UCL95>=CV?, meaning "is the UCL95 value greater than or equal to the comparison value?"). The results of single sample hypothesis testing, which are shown in the tables for information purposes, were not used in this determination. This methodological difference was based on an evaluation of the algorithms employed by the single sample hypothesis test methods (as implemented in ProUCL), and particularly how the accuracy of these methods are affected by the characteristics of the Church Rock datasets. Note that in the following discussion the term nondetect is used as a catchall for censored data, which in the case of the Church Rock data represents a result less than its reporting limit.

Datasets having the following characteristics tended to result in adverse consequences for the single sample hypothesis tests:

1. High percentages of results below reporting limits (nondetects).
2. Multiple values of reporting limits within datasets (i.e. for a particular COPC).
3. Nondetect results having values greater than other results reported as detected within a dataset.
4. Highly skewed distributions.

The Church Rock datasets for several COPCs commonly have one or more of these characteristics. The adverse consequences from characteristics 1 through 3 arise largely because of the handling of nondetect values by the algorithms employed by the single sample hypothesis tests in ProUCL. These consequences typically affected the results of the Wilcoxon Signed Rank Tests more than those of the Sign Tests. The reason for this is that the Wilcoxon Signed Rank Test replaces all nondetect results with a value equal to half the detection limit (or reporting limit). Furthermore, the method treats any result (detected or not) less than the highest nondetect result in a dataset as a nondetect result, and accordingly reduces its value by one half. Therefore, having even a single highly valued nondetect result can have a profound (and undesirable) affect on the outcome of the test. This is the reason that large nondetect results were removed from the datasets for these calculations.

In contrast to the Wilcoxon Signed Rank Test, the Sign Test retains nondetect values at their reported (reporting limit) value. The Sign Test also discards any nondetect result that exceeds the value of the comparison value, thereby typically avoiding the problems created by the handling of large nondetect values by the Wilcoxon Signed Rank Test. However, the Sign Test also resulted in questionable "Do Not Reject" outcomes in some cases where 100% of the data were nondetect results. (Note that the null hypothesis used in all the tests is that the median of the background dataset equals or exceeds the comparison value.)

For example, in cases where more than 50% of the results are nondetects and the reporting limit equals the comparison value, the Sign Test will fail to reject the null hypothesis even though the majority of the data are clearly less than the reporting limit (and the comparison value). This occurs because the Sign Test records a nondetect equaling the comparison value as a tie. The very different handling of nondetects by these two methods is illustrated by the significant differences of outcomes for the two tests shown in Tables 7 through 9. Finally, highly skewed distributions, a characteristic common to many of the Church Rock datasets, is described by the ProUCL documentation as a factor reducing the accuracy of the Wilcoxon Signed Rank Test.

The methodologies used by ProUCL for the Sign Test and Wilcoxon Signed Rank Test were tested by hand calculations. This was done using algorithms published in EPA statistical guidance (EPA, February 2006, *Data Quality Assessment: Statistical Methods for Practitioners*, EPA QA/G-9S, pp. 60-61). The same document is referenced by the ProUCL documentation as a source of its algorithms. The handling of nondetect results by ProUCL was determined to be faithful to the published algorithm for the Wilcoxon Signed Rank Test and numerically accurate. This procedure includes the substitution of values equal to one half of the detection limit (DL/2) for nondetects. The description of the Sign Test in EPA (February, 2006) does not explicitly mention any substitutions for nondetect results. However, the example calculation provided for the Sign Test (EPA, February 2006, Box 3-17, p. 63) indicates the use of the same DL/2 substitution used for the Wilcoxon Signed Rank Test. In this respect the authors of the ProUCL software may have misinterpreted the intentions of their source for the Sign Test algorithm, resulting in an inappropriate treatment of nondetects having the same value as the standard of comparison.

The solution of ignoring the single sample hypothesis tests in favor of direct comparisons of the estimated UCL95 statistics with comparison values avoids the problems described above, and has other additional advantages. Direct comparison of two values has the intuitive advantage of simplicity, and it also avoids the logical inconsistency of concluding (as would have been the case in some instances) that a UCL95 statistic should be considered lower than a comparison value when it obviously is not.

Equally important is the observation that the methodologies employed by the ProUCL software for estimating UCL95 statistics are highly advanced relative to the algorithms used by the single sample hypothesis tests. For example, a battery of more than 20 independent algorithms is employed by ProUCL to estimate UCL95 statistics. The software automatically sifts through these methods to recommend the better one or two estimates according to a variety of dataset characteristics, including number of samples, numbers and values of nondetect results, and shape of the distribution (including skewness). In particular, the handling of nondetect values by the Kaplan-Meier methods (for estimating UCL95 statistics) is more sophisticated than the methods used by the single sample hypothesis tests.

In the teleconference of September 30, 2008, it was agreed that this use of UCL95 statistics was an acceptable alternative to the single sample hypothesis test, for evaluating the background data sets versus the comparison values.

Note that direct evaluation of background UCL95 statistics versus comparison values are being made solely for the purpose of determining whether those statistics are numerically greater than the respective COPC comparison values. It would be inappropriate to use the same methodology for comparisons of compliance samples to cleanup levels, because of the much smaller size of compliance well sample sets (relative to background sample sets). For such comparisons single- or two-sample hypothesis testing is preferable.

Conclusion

We have concluded that direct evaluation of UCL95 statistics versus comparison values is the preferable method of determining which UCL95 statistics should be selected as candidates for consideration for modifying cleanup levels to reflect background concentrations. The UCL95 statistics presented in this document as candidates for consideration as modified cleanup levels are based solely on statistical calculations. One of the conclusions of the June 27, 2008, teleconference was that the preferred method of comparing compliance samples to background-based cleanup levels is two-sample hypothesis testing (e.g. of a compliance data set against the defined background data set). Therefore, one of the objectives of the current work has been to define appropriate background data sets for those future comparisons.

Very Truly Yours,

James Ewart, Ph.D., P.G.
Technical Consultant

JE: abc-220-mj

cc: Roy Blickwedel, GE
Larry Bush, UNC
Earle C. Dixon, NMED

Attachments