UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION 5
77 WEST JACKSON BOULEVARD
CHICAGO, IL 60604-3590

REPLY TO THE ATTENTION OF:

E-19J

April 28, 2011

Jon Ahlness
U.S. Army Corps of Engineers
Regulatory Branch
180 Fifth Street East, Suite 700
St. Paul, Minnesota 55101-1678

Bill Johnson
Minnesota Department of Natural Resources
500 Lafayette Road
St. Paul, Minnesota 55155-4040

Re: Comments on the Technical Memorandum for Characterizing Uncertainty in the Mine Site Background Groundwater Data

Dear Mr. Ahlness and Mr. Johnson:

The U.S. Environmental Protection Agency has reviewed the referenced Technical Memorandum entitled Characterizing Uncertainty in the Mine Site Background Groundwater Data, prepared by Barr Engineering. Our comments are provided pursuant to our authorities under the National Environmental Policy Act (NEPA), Council on Environmental Quality regulations (40 CFR Parts 1500-1508), Section 309 of the Clean Air Act, and Section 404 of the Clean Water Act.

General guidelines for Combining Data from Different Projects

When combining data from different projects, it becomes extremely important to ensure that the data represents the same sampling conditions and procedures. Based on project objectives, there are a number of different methods that can be used in collecting water samples. Some examples of these differences in methods used may include the following:

Filtered or unfiltered samples

Filtering a sample often removes most of the suspended solids that would be present in unfiltered samples, decreasing the concentration of an element.

Acidified or non-acidified sample

Acidification of a sample dissolves any suspended solids, increasing the concentration of an element, whereas in a non-acidified sample the suspended solids may settle out.

Type of sample containers/time between sampling and analysis

Elements can be absorbed into the sample container, especially if there is a significant amount of time between sampling and analysis, thus, consideration of container type may be important.

For groundwater, whether the sample is just a grab sample or whether the well was pump dried (or pumped for a set amount of time) and allowed to fill back up before the sample was taken, the length of time that water sits in a well can change the chemistry of the water. This becomes more important when the groundwater flow is slow. In addition, seasonal changes, upstream disturbances, and antecedent conditions (i.e., the amount, duration and intensity of the precipitation), can influence the chemistry.

Differences in Analytical Analysis

Differences between data sets

Care needs to be taken when combining data sets. Analytical methods used should be comparable. For example, one would not want to combine an analysis that was done by Colorimetry with one done by Atomic Absorption Spectrometry.

Differences within a data set

It is extremely important to insure that there is no change in analytical methods within the data set. This may be the case in the PolyMet data set. There are three different detection limits listed on Table 1 of the Attachment A; <1, <0.52, and <0.5. A change in the detection limits generally indicates a change in analytical method, or possibly in the sample preparation. An explanation for why detection limits are different should be provided. For instance, a possible explanation for the <1.0 and <0.5 detection limits could be the need to dilute the sample (e.g. there was not enough of a sample for analysis (although this should not be the case for water samples). A need to dilute a sample on a 1:1 basis (1ml sample with 1 ml pure water) would change the detection limit from a 0.5 detection limit for the instrument to a 1.0 detection limit for the sample.

Discrepancies between the Memorandum and the Attachment A

There are discrepancies between the number of samples indicated in the cover memo (dated April 13, 2011) and the Attachment A (dated April 12, 2011). For this document, it involves the number of samples with lead analysis. It is presumed that this discrepancy is repeated with the other parameters.

In Table 1 of the Technical Memorandum, the number of samples in the USGS study analyzed for lead is listed as 30, but there are only 22 analyses listed on Table 1 of the Attachment A. EPA notes that 8 samples were not included. We recommend including the 8 missing samples or explain why they were excluded.

In Table 1 of the Technical Memorandum, the number of samples with lead analysis for PolyMet data set is listed as 18, but there are 20 analysis listed on Table 1 of the Attachment A. A brief discussion of these additional samples is recommended.

The large number of non-detects in the PolyMet data (13 out of 20, meaning there are only 7 (or 5 in the case of 13 out of 18) genuine data points) is in contrast to the USGS data (no non-detects out of 22 values) and the MPCA data (three non-detects out of 21 values). It may be true that consistent handling of non-detects is defensible, but the difference in occurrence of non-detects suggests the PolyMet Data is suspect. It appears that the PolyMet detection limit is at least an order of magnitude higher than MPCA's and is also higher than USGS's. The USGS data has seven values lower than the lower PolyMet detection limit.

In Table 1 of the Attachment A, EPA notes that the numbers provided may not be comparable. How comparable are these numbers?

Nothing was presented in the Technical Memorandum about the analytical methods used to collect these data. There were no indications of the temporal and spatial properties of these data, and if these samples were all within the same watershed/aquifer?

Barr Engineering's response to EPA's previously expressed concern that the 18 samples were too small of a data set to capture the rarer and much higher values that would likely to be occurring at the site, the Technical Memo response to that EPA concern is logically flawed. Establishing this inadequate data set as the standard to which all potential data may be compared is, by its nature, limiting the variation allowed in the pooled data set, will in fact limit further expansion of the data distribution. If it can be shown the distribution is representative of the water quality of the surficial aquifer, then use of the pooled data set should be proposed. If not, then it is prudent to expand the data set on the surficial aquifer in a planned manner (new samples) to capture a truer representation.

EPA has confirmed that use of the "*mean*" is inappropriate for determining the water quality results for many parameters. Water quality sample analysis should be done in a manner used by NPDES permitting in the "reasonable potential analysis" (regarding the amount of monitoring, the number of samples and temporal spread, and the 95th percentile of the data used as opposed to just a mean) where estimating the higher percentiles is based on the target percentiles of the criteria of each parameter behind the standard.

Statistical Evaluations and Data Adequacy

EPA has several reservations about the data.

- As stated above, serious concern, is that the sample sizes for lead (listed in Table 1 of the Technical Memorandum for USGS data (n=30) and PolyMet data (n=18) do not match the numbers of measurements used for the analyses (n=22 for USGS data set, thus eight measurements were not used; n=20 for PolyMet). The missing and added values means that if incorrect data were used for the analyses, then presented statistical results may yield incorrect conclusions. These inconsistencies must be explained and data corrections and re-analyses must be done.

- An even more critical concern is the adequacy of the data for the present purposes. Assuming that this data (n = 30 for USGS, n=18 for PolyMet and n=21 for MPCA for a total of 69 values) are sufficiently representative of the range of environmental conditions (this may not be the case as mentioned above); this means that the sample data distribution approximates the true population data distribution. If the data really are log-normally distributed (the sample data support this conclusion), EPA has several concerns about whether or not the data actually reflect the relevant environmental reality. The estimated lognormal shape factor derived from the pooled data is approximately 1.61 (indicative of a highly skewed data set). This is further verified by the estimated standard deviation of about 6.8, which is about 3.5 times the estimated average of 1.93. The estimated variance is approximately 46.2. The relatively high variability suggests that the sample sizes for each of the individual data sets (and the sample size if all of the 69 values are pooled) are too low to get a reasonable margin of error. This means that margins of error attached to any parameter estimates from the sample data make the estimates unusable for decision making.

- Another concern is that given the small sample sizes, the upper percentiles may be substantially underestimated. The lognormal estimates of the 95th and 99th percentiles are greater than the actual sample values (for the 95th percentile the actual sample value is 6.1, but the lognormal estimate is about 7.46. For the 99th percentile, the actual sample value is 13 but the lognormal estimate is about 22.37). This supports the suggestion that the small sample sizes are not yielding true high percentile estimates if the population values are lognormally shaped.

- EPA also has concerns about the procedures used to compare the shapes of the data sets. While they seem to be appropriate; distribution location comparisons were not done. Before two or more data sets can be viewed as having similar results, both the shapes and the actual data values must be similar. Given the heterogeneity of the descriptive statistics estimates shown in Table 2, it is uncertain whether the actual values in the data sets are statistically similar. The data sets range from 5.85 (PolyMet) to 12.9 (USGS). This certainly should be tested before pooling can be assumed to be appropriate. Another more fundamental concern relates to the hypotheses. It should be noted that the null hypotheses are that the data sets being compared are from same population. EPA suggests reversing the directions of the null and alternate hypotheses to reverse the burden of proof. Given the relatively low power of the tests associated with the insufficient sample size of each data set, this comment is crucial.

In summary, due to deficiencies discussed and the concerns raised above, using a limited data distribution will in fact limit expansion of the data distribution and will increase the uncertainties. EPA continues to believe that 18 data points constitute an insufficient sample size from which to draw valid conclusions. If additional data set are to be used to supplement the PolyMet data, those data sets must have much wider range than the PolyMet data in order to represent a true distribution rather than using select data to reinforce the PolyMet data which is limited to only 7 (or 5) data points due to high percentage of non-detect values.

If the available data will not address the inadequacy concerns, then PolyMet should collect new data, using agreed upon sampling method and approved sampling plan to generate a statistically valid data to be used in impact assessment predictive models.

EPA is available to discuss these comments to the Characterizing Uncertainty in the Mine Site Background Groundwater Data Memo at your convenience. Please feel free to contact me at 312-353-2681 to discuss these comments.

Sincerely,

Simon Manoyan
NorthMet Mining Project Manager
NEPA Implementation Section
Office of Enforcement and Compliance Assurance