INFORMATION REQUEST RESPONSE

EA No: 0809-001 **Information Request No: YKDFN #07**

Date Received:

February 28, 2011

Linkage to Other IRs:

Date of this Draft:

May 31, 2011

Request

Preamble:
The geometric mean and arithmetic mean are both presented in the summary of historical surface and sediment quality data in the Tier II Risk Assessment Report 1 developed for the DAR. The use of geometric mean tends to dampen the effect of very high or low values in the averaging calculation. In general, geometric mean is commonly used in situations where the data range covers several orders of magnitude. The existing monitoring data was collected over a very large sampling period – tens of samples over decades, and produced a dataset that contains wide variance, but because the data is independent, these variations should not be 'processed out'.

The use of the geometric mean to analyze surface water data can produce significantly decreased values. For example in Table 7.1.4 the geometric mean and arithmetic mean are calculated to be 3.8µg/L and 28.1 µg/L for 58 respectively for collected surface water samples. The use of the geometric mean of historical surface and sediment water quality data as the input parameter for the numerical model in the risk assessment could result in the underestimation of arsenic loading into the receiving water bodies.

Question:
It is requested that clarification is provided on what type of mean input values of historical collected data for arsenic surface and sediment were utilized in the numerical modeling simulations used in the risk assessment. It is further requested that the proponent indicate the rational for utilizing a geometric mean for the evaluation of future monitoring data, if that is the intention of the proponent.

Reference to DAR (relevant DAR Sections):

General comment (example of Table 7.14 provided)
S.8.9 (Assessment of Ecological and Human Health Risks)

Summary

The Risk Assessment used geometric mean values but also captured the measured variability in the data through a sensitivity analysis based on the 5^{th} and 95^{th} percentile values. Geometric mean values were applied in the Risk Assessment because they are most representative of long-term exposure scenarios.

Response

Environmental data is often found to be highly variable and to be influenced by any number of factors (e.g., the direction that the wind is blowing, the time of the day when a sample is collected, etc.). In addition, it is often found that the data are highly skewed (i.e., most measurements typically fall within a narrow range while there are a few measurements that are either very low or very high compared to most of the measurements). In analyzing such a dataset, it is often observed that the data when plotted as a frequency histogram are log-normally distributed. The best estimate of the mean of such a dataset is the geometric mean value. In contrast, the arithmetic mean tends to over-emphasize the relatively infrequent occurrence of outliers within a dataset.

In the risk assessment field, where the interest is in evaluating long-term exposures, it is common practice to use geometric mean values and geometric standard deviations to characterize the model input parameter values where log-normal distributions best describe the variability in observed data. In assessing ecological and human health risks on the Giant Mine Remediation Project, it is noted that the assessment was carried out within a probabilistic framework to explicitly take into account uncertainties measured in the environmental data. In other words, the assessment was based not only on geometric mean values but also captured the measured variability in the data. The uncertainties in the exposure estimates were reported as 5th, 50th (mean) and 95th percentile values in the Tier 2 Risk Assessment. So while log-normal distributions were used to describe several of the model input parameter values, the arithmetic mean values were reported for the predicted values (e.g., arsenic exposures).

With regard to the presentation of data from the long-term monitoring program (as described in Chapter 14 of the DAR), it is common practice when reporting environmental data from ongoing monitoring programs to present both arithmetic and geometric mean values as well as minimum and maximum values, total number of measurements and number of measurements less than the method detection limit. A complete compilation of all data is typically included as appendices in summary reports.