Has your discharged storm water really degraded the receiving stream? (Newsletter)*

February 7, 2019

Most industrial operations discharge storm water into a receiving water body from a single point of discharge from the permitted area. The water leaving the site will have many chemical constituents, some of which are considered to be pollutants (or contaminants) by statute. States differ in their water quality standards, but all require permit compliance monitoring. How the reported results are analyzed can make a huge difference to the permit holder's operations. This newsletter uses a real example to explain how wrong analytical results can result in unwarranted financial penalties.

Discharge permit compliance is based on comparing analytical chemical laboratory concentrations with arbitrary maxima. Exceeding these maxima can result in fines, or worse penalties such as costly remedial actions or cease-work orders.

One state's reporting form for industrial storm water discharge tells the permit holder, "For non-detect sampling results, use 1/2 the detection limit to calculate the geom mean." Do NOT do this. There are two very serious errors in this sentence: imputing an arbitrary value and calculating the wrong most expected value.

Addressing nondetected geochemical values

When a geochemical concentration is below the analytical chemical laboratory's method detection limit it is reported as 'nondetected' (censored in data analysis language). In other words, the concentration value (if the constituent is even present) is unknown. Substituting (imputing) *any* value is arbitrary and wrong. There are three robust statistical methods that can describe and summarize a set of geochemical data containing censored values:

Maximum Likelihood Estimation (MLE) selects the probability distribution whose parameters best fit the existing data set.

Regression on Ordered Statistics (ROS) summarizes the data set when it is sorted from lowest to highest value. It is the sequence that matters, not the values themselves.

Kaplan-Meier Survival Analysis (K-M) has been used since at least the middle of the last century in medicine and by manufacturers who report, for example, mean time between failures. This is the preferred method for most geochemical analyses with as many as 80-90% unknown values.

Estimating the most expected value

When we speak of an "average" we mean the expected value of any random observation or event. The mathematical term for average is a *mean* and there are several ways of calculating it.

The *arithmetic mean* is the sum of the observations divided by the number of observations. It is the most common calculation of most expected value in common use. Frequently applied to geochemical data summaries it has a characteristic that makes it undesirable for these data: very high outlying values can skew the mean value.

The *geometric mean* is the n^{th} root of the product of the observations. An example of when a geometric mean should be used is to calculate the average rate of return of an investment. Consider investment which earns 10% the first year, 50% the second year, and 30% the third year. Its average rate of return is not the arithmetic mean, because what these numbers mean is that on the first year the investment was multiplied (not added to) by 1.10, on the second year it was multiplied by 1.60, and the third year it was multiplied by 1.20. The relevant quantity is the geometric mean of these three numbers. Geochemical data should not be multiplied (see example below).

The most accurate and useful measure of geochemical concentration expected value is the *median*, the value that has the same number of observations less than it and greater than it. It is the 50th percentile in data analysis language.

To illustrate the results of calculating the most expected value using the above three methods I use 11 measurements of total arsenic (in mg/L) at a surface water sampling location near a metal mine:

```
0.0006, 0.0006, 0.0006, 0.0008, 0.0008, 0.0011,
0.0015, 0.0022, 0.0024, 0.0099, 0.0113
```

The arithmetic mean concentration is 0.0029 mg/L. The geometric mean concentration is 0.1783 mg/L. The median concentration is 0.0011 mg/L.

If the regulatory threshold is 0.02 mg/L and the permit holder reports the arithmetic mean the operation is well within the accepted range. If the regulator recalculates using the geometric mean the permittee is badly out of compliance. Using the median value demonstrates that the site is not degrading the receiving water because the most expected value is not artificially raised by two very high values.

Analyzing permit compliance monitoring data using technically sound and legally defensible methods benefits environmental, economic, and societal ecosystems.

—