Rethinking Poisson-Based Statistics for Ground Water Quality Monitoring

by Jim C. Loftis[a], Hariharan K. Iyer[b], and Heather J. Baker[c]

Abstract

Both the U.S. Environmental Protection Agency (EPA) and the American Society for Testing and Materials (ASTM) provide guidance for selecting statistical procedures for ground water detection monitoring at Resource Conservation and Recovery Act (RCRA) solid and hazardous waste facilities. The procedures recommended for dealing with large numbers of nondetects, as may often be found in data for volatile organic compounds (VOCs), include, but are not limited to, Poisson prediction limits (in both the EPA guidance and ASTM standard) and Poisson tolerance limits (EPA guidance only). However, many of the proposed applications of the Poisson model are inappropriate. The development and application of the Poisson-based methods are explored for two types of data, counts of analytical hits and actual concentration measurements. Each of these two applications is explored along two lines of reasoning, a first-principles argument and a simple empirical fit.

The application of Poisson-based methods to counts of analytical hits, including simultaneous consideration of multiple VOCs, appears to have merit from both a first principles and an empirical standpoint. On the other hand, the Poisson distribution is not appropriate for modeling concentration data, primarily because the variance of the distribution does not scale appropriately with changing units of measurement. Tolerance and prediction limits based on the Poisson distribution are not scale invariant. By changing the units of observation in example problems drawn from EPA guidance, use of the Poisson-based tolerance and prediction limits can result in significant errors. In short, neither the Poisson distribution nor associated tolerance or prediction limits should be used with concentration data. EPA guidance does present, however, other, more appropriate, methods for dealing with concentration data in which the number of nondetects is large. These include nonparametric tolerance and prediction limits and a test of proportions based on the binomial distribution.

Background and Introduction

During the past 10 years, statistical analysis of ground water quality monitoring data from solid and hazardous waste disposal facilities has become routine. The goal of such analyses is most commonly to assess whether or not a contaminant release is likely to have occurred. This is done by objectively comparing water quality observations from two groups: a "compliance" group which would presumably be impacted by a release from the facility; and a "background" group which would be identical, or at least statistically very similar, to the compliance group, but would not be impacted by a release. The compliance group consists of observations collected from wells downgradient of the facility after operation has begun. The background group consists of observations that are collected from wells upgradient of the facility (interwell comparisons) or from the same wells used in the compliance group but prior to facility operation or before any impacts have occurred (intrawell comparisons).

Guidance for selection of statistical procedures appropriate for such monitoring is provided by the EPA (U.S. EPA 1989, 1992) and more recently by ASTM (American Society for Testing and Materials 1996). The EPA and ASTM recommendations have benefited from several research efforts and numerous articles in the scientific literature (which are cited in the aforementioned references). Most certainly, they have come a long way from the early recommendations based on a particular form of the Student's t-test and, in general, provide sound advice, invaluable to practitioners. Current guidance recognizes particularly troublesome characteristics of ground water monitoring data, notably non-normality and the presence of large numbers of nondetects, sometimes comprising the majority of the data. An excellent review of the pertinent literature is presented by Davis and McNichols (1994a, 1994b).

The procedures recommended for dealing with large numbers of nondetects include, but are not limited to, Poisson prediction limits (in both the EPA guidance and ASTM standard) and Poisson tolerance limits (EPA guidance only). Both of these procedures are based on the work of Gibbons (1987). Both tolerance and prediction limits may be constructed from background data for the purpose of determining whether future observations are likely to have come from a different population and might thus suggest possible contamination. Both types of limits may be used in either interwell comparisons or intrawell comparisons.

Tolerance limits define intervals which contain a portion of the population (coverage) greater than or equal to a specified fraction with specified probability (confidence level). Prediction limits define intervals which contain each of k future observations (k may be varied) from a given population with a specified confidence level. For both tolerance and prediction limits, parametric procedures based

[a]Department of Chemical and Bioresource Engineering, Colorado State University, Fort Collins, CO 80523. E-mail: loftis@engr.colostate.edu
[b]Department of Statistics, Colorado State University, Fort Collins, CO 80523. E-mail: hari@stat.colostate.edu
[c]Intelligent Decision Technologies, 203 S. Main St., Longmont, CO 80501. E-mail: heather.baker@idt-ltd.com
Received April 1998, accepted August 1998.

on a normality assumption as well as nonparametric procedures are presented in EPA and ASTM guidance. The parametric procedures are recommended if the data are normally distributed or normality can be achieved by a suitable transformation and/or correction for nondetects. The nonparametric procedures would be recommended for most other cases, including data sets with large numbers of nondetects. However, the nonparametric procedures have the limitation that the upper tolerance or prediction limit can be no greater than the largest observation in the background data set, and one cannot compute limits with large confidence levels and small background sample sizes. The Poisson-based procedures represent a third alternative that does not have this limitation.

The Poisson-based procedures are purported to have several desirable properties, including the ability to deal with high fractions (>90%) of nondetects, to include the magnitudes of observations above the detection limit, and to combine observations for several different VOCs into a single test. However, as demonstrated in this paper, some of the proposed applications of the Poisson model are seriously faulty. The EPA guidance contains alternative methods that are better suited for dealing with the problem of large numbers of nondetects, and these are reviewed at the end of this paper.

There are two possible lines of reasoning to develop arguments in favor of applying the Poisson distribution and associated methods in ground water quality monitoring. The first of these is a first-principles approach in which the events of concern are considered to be rare events that are outcomes of multiple Bernoulli (two possible outcomes) trials. Under certain conditions, the number of outcomes of each type in a given number of trials may be modeled exactly by a binomial distribution, for which the Poisson is a good approximation (again under certain conditions). This is the primary line of reasoning employed by Gibbons (1987). The second is an empirical approach in which the distribution of observed data is compared to a Poisson distribution and a good empirical fit is found in certain cases.

Each of these lines of reasoning may be applied to two types of data. The first data type is simply a count of the number of analytical "hits" or observations above the detection limit out of a given number of analyses. Gibbons (1987) suggests using the Poisson distribution to model "hits/scan," in which a scan is an analysis of a single water sample for multiple (perhaps 20 to 50) VOCs, and the number of hits is the number of compounds for which the measured concentration is above the detection limit. The second type of data is actual concentration measurements for each analysis, i.e., for each VOC in each scan, many of which could be recorded as nondetects or "less-thans." As shown later, there are at least some problems with each line of reasoning and each type of data. Extremely serious problems occur with actual concentration data using either line of reasoning.

Analysis

Applications of the Poisson Distribution

The Poisson distribution is a discrete distribution often used to model rare events. Classic example applications include modeling radiation counts observed at a detector over a fixed interval of time or modeling the number of individuals who contract cancer out of a large study group of fixed size, say 1000. In the latter example, each individual result is the outcome of a Bernoulli trial, meaning that there are two possible outcomes (the individual will contract cancer or will not contract cancer) with probabilities p and (1-p), respectively. If p (the probability than an individual will contract cancer) is constant across all trials, then the number of individuals who contract cancer out of a sample of size n individuals is modeled exactly by a binomial distribution with parameter p.

The Poisson distribution is a useful approximation to the binomial distribution when the number of trials, n, is large and the binomial parameter, p, the probability of the event of interest, is small. The Poisson distribution has a single parameter, usually denoted as λ, which is often called the rate parameter. For a given Poisson distribution, this parameter is both the mean and the variance. Thus the standard deviation of a Poisson distribution is always equal to the square root of the mean. When the Poisson distribution is used to approximate the Binomial distribution, the Poisson parameter is equal to np. Ott (1993) indicates that, when np is less than 5, the Poisson distribution is a good approximation to the binomial distribution. It is easy to show that the Poisson distribution may be a good approximation to the binomial for larger values of np as well, but this criterion will suffice for the present discussion.

Poisson probabilities are given by (Dwass 1970):

$$p(X) = \frac{\lambda^x e^{-\lambda}}{X!} \qquad (1)$$

where X equals 0, 1, 2, 3, ...; p(X) is the probability that the Poisson random variable will take on a given value, X; and λ is the Poisson parameter.

As an example application of the Poisson distribution, suppose that several similar groups of individuals of fixed size (say 1000) are drawn from an infinite population. Let us suppose also that the "true" fraction of individuals in the infinite population who contract cancer is 0.002. In this case, the distribution of the number of individuals who contract cancer in each group is described exactly by a binomial distribution with parameter p = 0.002. For a study group of size 1000, the Poisson distribution should be a good approximation to this "exact" binomial model since n is large, p is small, and np is less than 5. The value of the Poisson parameter, λ, is np = 2 which is the average (mean) number of individuals who contract cancer per study group of 1000.

The sum of m independent Poisson random variables with parameter λ is also Poisson with parameter $m\lambda$ (Dwass 1970). Thus the mean and variance must scale linearly with the size of the observed sample. The standard deviation scales with $\sqrt{m}$.

In the previous example, suppose now that the groups consist of 2000 individuals rather than 1000. This can be viewed either as a single Poisson random variable with n = 2000 or the sum of m = 2 Poisson random variables, each with n = 1000. In the first case, we obtain the mean = λ = np = 2000(0.002) = 4. In the second case, we obtain the new value of λ (for n = 2000) as twice the value of λ for n = 1000 or 2(2) = 4.

The new variance, for n = 2000, would also be equal to 4. Thus, the standard deviation would be 2. The standard deviation does not scale linearly, and the coefficient of variation (standard deviation divided by the mean) gets smaller as the number of individuals in the sample gets larger. This makes intuitive sense in this example. Since the event being observed is rare, we would expect large variability (coefficient of variation) in observed occurrences of cancer when the study group is small and reduced variability when the study group gets large. This line of thought applies equally well to radiation counts. If we double the time of observation, the mean particle count will double, and so will the variance. The coefficient of variation will decrease by a factor of $1/\sqrt{2}$.

Modeling Analytical Hits Using the Poisson Distribution

The same model could be used to describe the number of detects, or "hits," for a given number of analyses when certain conditions apply. As stated previously, one could attempt to justify the use of a Poisson model using either a first-principles argument or an empirical fit of the data.

First let us consider the first-principles argument. Each analysis can be regarded as a Bernoulli trial. The two possible outcomes are detect (or hit) and nondetect. If the trials are independent and the probability of a hit, p, is constant over all trials, then the number of hits out of n trials is modeled exactly by a binomial distribution with parameter p. The Poisson distribution is a good approximation when the number of (independent) analyses is large and the probability of a hit, p, on a given trial is small ($np<5$) and is constant from analysis to analysis. If we are considering a large number of analyses for a single compound by the same laboratory and method, then these conditions probably apply. However, as Davis and McNichols (1988) point out in a comment on Gibbons (1987), both of the assumptions of constant p and independence are likely to be violated in the hits/scan application where multiple compounds are considered simultaneously. Certainly, the probability of a hit is not the same for different compounds, and if one compound is detected, then other detections are more likely. This would seem to invalidate a first-principles argument for application of the Poisson distribution unless we restrict ourselves to one compound at a time. However, it follows from the results of LeCam (1960) (see also Feller 1970, page 286) that a Poisson model provides a good approximation for the sum of n Bernoulli random variables, even with unequal p values, provided that the p values are all small, and n is large, as long as the independence assumption is valid. For modest correlation between trials the approximation may still be satisfactory. Therefore, the first-principles argument might have some validity after all.

Even if the first-principles argument is not particularly strong, one could still make the case for modeling detects or hits/scan based on an empirical fit of the data. In the reply to Davis and McNichols (1988), Gibbons (1988) presents one example data set that fits the Poisson model nicely. The data set includes 127 scans of 32 VOCs each.

Modeling Concentrations Using the Poisson Distribution

As mentioned earlier, the Poisson distribution is purported by Gibbons (1987), U.S. EPA (1992), and ASTM (1996) to be useful for modeling actual concentrations in addition to modeling numbers of hits. Gibbons bases this contention on a molecular argument which is creative but not rigorous. Problems with this argument were correctly pointed out by Davis and McNichols (1988). However, their argument in favor of a continuous (normal) distribution was not fully developed due to the need for brevity in their comment. Using a first-principles argument we shall show that, under reasonable assumptions, the "true concentrations" of a pollutant follow an approximately Poisson distribution. However, the Poisson parameter must be determined with some care. Gibbons' (1987) development is not correct.

Suppose that there are an infinite number of molecules (water molecules, VOC molecules, etc.) in the population of interest and that a fraction p of these molecules are those of a pollutant of interest. Suppose we select a sample of n molecules. If X denotes the number of molecules of the pollutant of interest in this sample of n molecules, then, under reasonable assumptions, X may be modeled using a binomial distribution with parameters n and p. In situations one encounters in practice, n is very large, and p is very small, so a Poisson distribution with $\lambda = np$ may be expected to provide an adequate approximation for this binomial distribution. For the present application, np is many orders of magnitude greater than 5, and as Ott (1993) and Davis and McNichols (1988) point out, a normal distribution actually provides a better approximation than a Poisson. However, we will continue our discussion using a Poisson approximation. Notice that the approximating Poisson distribution has parameter $\lambda = np$. The value of n depends, of course, on the volume of water sampled. For 1 mL samples of water, the value of n will be approximately 3×10^{22} since there are 18 grams/mole and 6.02×10^{23} molecules/mole. If the concentration of the pollutant in the population is 1 ppb, then $p = 10^{-9}$ (neglecting the difference in mass between a VOC molecule and a water molecule), and λ is approximately 3×10^{13}. With this value of λ, which takes into account the actual number of molecules per sample, the true numbers of molecules of the pollutant in a randomly selected sample will follow, approximately, a Poisson distribution. (As pointed out earlier, a normal distribution will provide an even better approximation, but the difference in fit between the two distributions turns out to be negligible for the number of molecules of interest here.)

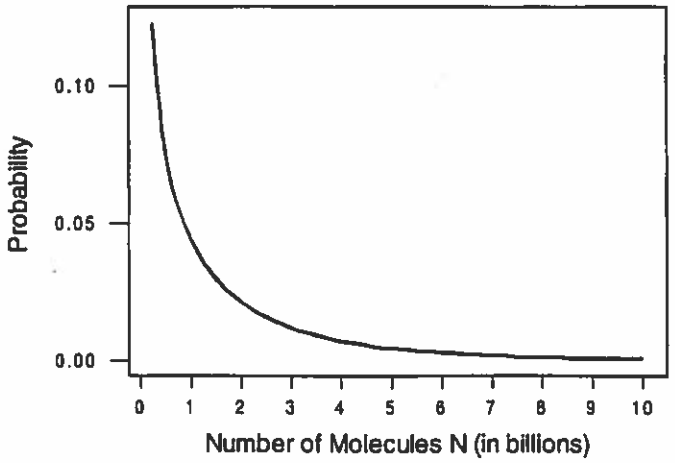

Figure 1. Probability of a sample containing exactly 40 ppb of contaminant vs. the total number of molecules N (in billions) in the water sample when the true population concentration is 34.75 ppb.

This reasoning is precisely what Gibbons (1987) calls the "molecular argument." The critical point in our development of this molecular argument is that the value of λ, and thus the mean and variance of the distribution, depends on the number of molecules present (i.e., the volume) in the water sample. If we double the sample volume, we double the value of λ, thus doubling the mean and variance.

However, Gibbons' (1987) development of the molecular argument missed this critical point, thus failing to correctly identify the parameter λ of the approximating Poisson distribution. In an example illustrating the use of the Poisson model, Gibbons notes that the average concentration of pollutants per sample is 34.75 ppb. He then proceeds to use a Poisson distribution with $\lambda = 34.75$ to model the concentrations in individual samples. For instance, he states that the probability of a sample with a concentration of 40 ppb is equal to

$$\frac{(34.75)^{40} e^{-34.75}}{40!} = 0.04 .$$

Using the correct molecular argument developed here, the value to use for λ is np where n is the total number of molecules per sample, and p is the pollutant fraction of the total population. If the true or population concentration is 34.75 ppb, then $p = 34.75 \times 10\text{–}9$. For simplicity, let N = the number of molecules present in billions. Then $\lambda = 34.75$ N. A sample with a pollutant concentration of 40 ppb will then have 40N molecules of the pollutant. Therefore, the correct Poisson probability is given by $\frac{(34.75N)^{40N} e^{-34.75N}}{(40N)!}$. This can be calculated only when N, the size of the water sample, is known, because the probability definitely depends on N as it should. Figure 1 shows how this probability changes as N takes on different values.

As an aside, notice that as the value of N increases, the probability rapidly but asymptotically approaches zero. In a sample consisting of an extremely large number of molecules, the probability will be miniscule that the number of molecules of the pollutant will equal one specified value. Therefore, it doesn't make sense to talk about the concentration being exactly equal to some specified value. It makes sense only to speak of the concentration being within a specified range of values.

What are the consequences of Gibbons' (1987) error of incorrectly specifying the value of λ in the Poisson approximation? They are quite serious as the next example shows.

Suppose that the average (mean) concentration of lead in a large number of samples is 1 part per million (mg/L). We can permit most of these observations to be nondetects without loss of generality. Following Gibbons' (1987) development, if the observations are really Poisson distributed, then the variance of the observations must also be 1 with units of $(mg/L)^2$, and the standard deviation is 1 mg/L. Suppose now that we do nothing other than change the units of the measurement to parts per billion (micrograms/L or $\mu g/L$). The mean is then 1000 with units of μg /L , and if the Poisson distribution applies, the variance must be 1000 with units of $(\mu g\ /L)^2$. The standard deviation is $\sqrt{1000}$ or about 32 μg /L or 0.032 mg/L in the original system of units. The standard deviation has changed by a factor of 30 simply because we have used a different system of units. Obviously this cannot be. The mean and variance cannot be dependent on the units used. A basic criterion that any model must meet is that it be independent of the system of units, and the Poisson model does not meet that criterion.

One might argue that this problem could be overcome by always using the same system of units such as g/L. In this case, an observation of 1 $\mu g/L$ would be represented as a value of 10^{-6} with units of g/L. However, the Poisson is a discrete distribution, and one cannot have fractions such as 10^{-6} in the Poisson distribution, only integer counts. Thus all Poisson measurements must be represented as integers with appropriate units. Furthermore, there is no system of units which will give the correct results since the molecular argument depends not only on the concentration but also on the size of the sample.

As noted earlier, this scaling problem results from improper specification of the rate parameter λ. If the parameter is specified properly by considering the number of molecules present, then the mean and variance will scale properly (i.e., the mean will remain constant, but the variance will decrease with increasing number of molecules in the sample). Furthermore, the result will not depend on the system of units used—an absolute requirement. In short, concentration measurements are continuous and representing them as discrete counts using the Poisson distribution creates the scaling problem that we have described.

Table 1
Benzene Concentrations (ppb) from Six Background Wells

Month	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6
1	<2	<2	<2	<2	<2	<2
2	<2	<2	<2	15.0	<2	<2
3	<2	<2	<2	<2	<2	<2
4	<2	12.0	<2	<2	<2	<2
5	<2	<2	<2	<2	<2	10.0
6	<2	<2	<2	<2	<2	<2

Role of Measurement Errors

Even if the parameter λ is correctly specified based on the number of molecules actually present in the sample, a correct "molecular" argument based on first principles does not lead to a useful model for observed concentration data. Remember that we are thus far considering only the "true" concentration of VOCs or other pollutants in the water sample. The variance that is predicted by the Poisson model is due only to randomness in the number of pollutant molecules from sample to sample. This variance must be equal to λ for the Poisson model and is vanishingly small for real samples. As discussed previously, if the sample volume is 1 mL, then the number of water molecules present is approximately 3×10^{22}, and λ is of the order of 10^{14} for concentrations in the range of a few ppb. Thus, the variance of the concentration measurements is of the order of 10^{14} $(molecules/mL)^2$, and the standard deviation is of the order of 10^7 (molecules/mL) which is of the order of 10^{-16} molecules/molecule or 10^{-7} ppb.

This demonstrates that the variance associated with the Poisson model and molecular argument is many orders of magnitude smaller than what we actually observe in real data. The reasons for this are fairly obvious. In practice, measured concentrations will differ from the true concentration due to analytical and other errors. If Y is the true concentration of a pollutant in a given sample, then we have the relation

$$Y = X + e \qquad (2)$$

where e denotes measurement error. Even if a Poisson model is a reasonable one for modeling X, it is certainly not the distribution to use to model Y. Furthermore, when one examines the sampling variability of Y, one finds that, for typical concentrations of pollutants and sizes of the sample, most of the variability in Y occurs due to the measurement errors. Given that the distribution of X is adequately modeled using a normal distribution rather than a Poisson, and that measurement errors are often normally distributed, one can see that a normal distribution is likely to provide a reasonable model for Y and that a Poisson model is highly unreasonable.

Errors in Poisson Tolerance and Prediction Limits

We have shown that a first-principles or molecular argument does not lead to a practical model for real concentration data which include analytical errors. We have shown that using this model without considering the number of molecules in the sample creates an obvious scaling problem in which the variance is not constant across different systems of units. Thus, we would expect the results of tolerance and prediction limit calculations to be dependent on the

system of units used, which would be unacceptable. That this, in fact, occurs is easily demonstrated by example.

Example—Errors in Poisson Prediction Limit

We shall consider two sample data sets from the current EPA guidance (U.S. EPA 1992). First let us consider the prediction limit calculation of Example 10 on page 35. Benzene data from six background wells are reproduced in Table 1. The objective is to compute an upper 99% prediction limit for the sum of the next four observations for a single downgradient well assuming the same (Poisson) distribution.

According to U.S. EPA (1992) and Gibbons (1987), an upper prediction limit for the sum of the next k observations from the same population is given by

$$T_k^* = cT_n + \frac{ct^2}{2} + ct\sqrt{T_n\left(1 + \frac{1}{c}\right) + \frac{t^2}{4}} \qquad (3)$$

where

T_k^* = the total Poisson count of the next k observations
T_n = the sum of the Poisson counts of n background samples
c = k/n
t = the upper 1-α quantile point of the Student's t distribution with n-1 degrees of freedom.

If we think carefully about the application of this equation to concentration data, a serious problem becomes apparent. The first term will have concentration units; the second terms has no units; and the third term has concentration units mixed with a term that has no units. So if we change the system of units, we get a different answer.

In the example problem, all of the background wells are assumed to represent the same population, and the data are pooled. Therefore, n = 36, of which 33 are nondetects, k = 4, and c = 4/36 = 1/9. To compute T_n all of the nondetects are set equal to half the detection limit or 1 ppb. Thus T_n= 33(1.0) + 12.0 + 15.0 + 10.0 = 70.0. Using statistical software or a table of the Studentís t distribution we find $t_{99,35}$ = 2.4377. Substituting these values into the above we obtain

$$T_k^* = \frac{1}{9}(70) + \frac{(2.4377)^2}{2(9)} + \frac{2.4377}{9}\sqrt{70(1+9) + \frac{(2.4377)^2}{4}} = 15.3 \text{ ppb}$$

which is the result obtained in U.S. EPA (1992).

If we change the units from ppb to parts per 100 million (pphm), so that 12.0 ppb becomes 1.2 pphm, then the upper 99% prediction limit for the sum of the next k = 4 observations changes from 15.3 ppb to 3.4 pphm or 34 ppb as shown here.

$$T_k^* = \frac{1}{9}(7.0) + \frac{(2.4377)^2}{2(9)} + \frac{2.4377}{9}\sqrt{7.0(1+9) + \frac{(2.4377)^2}{4}} = 3.4 \text{ pphm}$$

The answer has changed by a factor of 2 using the very same data.

Similarly, if we change the units to parts per ten billion (pptb) so that 12.0 ppb becomes 120 pptb, then the upper 99% prediction limit becomes 100.8 pptb or 10.1 parts per billion. After this change in units, all of the observations are still integers and could be considered Poisson "counts." However, the final result is much different from the original. The difference in the result due to units alone is unacceptable and indicates that the method is faulty.

Example—Errors in Poisson Tolerance Limit

Next, let us consider the tolerance limit calculation of Example 12 on page 40 of the EPA guidance document. This example uses the same data as the preceding example. The objective here is to compute an upper Poisson tolerance limit with 95% coverage and a 95% confidence level. In the guidance document the calculations follow the method presented by Gibbons (1987) and proceed as follows.

First calculate the most probable rate parameter using the following equation from Zacks (1970):

$$\lambda_{T_n} = \frac{1}{2n}\chi^2_{\lambda}\left[2T_n + 2\right] \qquad (4)$$

where T_n is the sum of the Poisson counts of n background samples (with nondetects set equal to half of the detection limit); χ^2_{λ} is the λ quantile of the chi-squared distribution with $(2T_n+2)$ degrees of freedom; and λ is the desired confidence level.

For this example, λ = 0.95, and the most probable rate parameter is given by

$$\lambda_{T_n} = \frac{1}{72}\chi^2_{0.95}\left[142\right] = 2.37$$

Next, compute twice the probable occurrence rate as $2\lambda_{T_n}$ = 4.74.

The upper tolerance limit is then given by the least positive integer, k, such that

$$\chi^2_{1-\beta}\left[2k + 2\right] \geq 2\lambda_{T_n} \qquad (5)$$

where [2k+2] = the degrees of freedom of the chi-squared distribution and β = the desired coverage.

Since we wish to obtain 95% coverage, 1-β = 0.05. Therefore, using a chi-squared table or statistical software, one finds the smallest degrees of freedom, (2k+2), such that

$$\chi^2_{0.05}\left[2k + 2\right] \geq 2(2.37) \qquad (6)$$

From the table we find that $\chi^2[11]$ = 4.57 and $\chi^2[12]$ = 5.23. Thus, the smallest degree of freedom which satisfies the inequality (Equation 6) above is 12. Therefore, [2k+2] = 12, so that k = 5, and the upper Poisson tolerance limit with 95% coverage and 95% confidence level is 5 ppb, which is the result obtained in U.S. EPA (1992).

Now, as in the prediction limit problem, let us see what happens if we change the units. It is difficult to carry out the calculations if the order of magnitude of the data is changed, since we need chi-squared statistics with large degrees of freedom. So let us simply change the units to parts per two billion, multiplying each data point by two. In this case, the most probable rate parameter is given by

$$\lambda_{T_n} = \frac{1}{72}\chi^2_{0.95}\left[284\right] = 4.50$$

And the upper tolerance limit is given by the smallest integer k such that

$$\chi^2_{0.05}\left[2k + 2\right] \geq 2(4.5)$$

Since $\chi^2[17] = 8.67$ and $\chi^2[18] = 9.39$, we find that $[2k+2] = 18$. Therefore, $k = 8$, and the calculated upper tolerance limit is 8 parts per two billion or 4 ppb. Thus the calculated upper tolerance limit changed from 5 ppb in the original example to 4 ppb, using the same data and simply changing the units. Of course, this is unacceptable and demonstrates that the method is faulty.

Therefore, we can state conclusively that the methods (as currently recommended) for construction of tolerance and prediction limits based on the Poisson distribution should not be used with concentration data.

Empirical Fit of the Poisson Model to Concentration Data

We have seen that a first-principles argument in favor of a Poisson model for concentration data is seriously flawed. As a result, statistical methods based on the Poisson distribution are not scale invariant when applied to concentration data and should, therefore, not be used. Gibbons (1988) makes an argument that concentration data with many nondetects actually fit a Poisson model empirically, and others might argue that one could rescale data to achieve a Poisson fit. However, Gibbons' own data set (Gibbons 1988), consisting of 3998 nondetects and 66 detects, provides strong evidence that concentration data are not Poisson distributed. Gibbons fits a Poisson distribution to the 66 detects only, ignoring the 3998 nondetects. The expected frequencies are computed assuming that the number of nondetects is zero when, in reality, their number is 3998. The fact is, a Poisson model does not fit Gibbons' (1988) data. His argument is therefore inapplicable to the problem under consideration.

With regard to rescaling observations in order to achieve a Poisson fit, we have shown that the Poisson model and methods are not scale invariant when applied to concentration data. Thus, different analysts will obtain different results unless they all rescale in exactly the same way. In order to achieve a Poisson distribution by rescaling, one must chose a scale factor such that the mean and variance, after rescaling, are equal (at least approximately). Two approaches present themselves in this regard. Suppose X represents the concentration random variable. Suppose also that there exists a positive constant c such that the random variable $Y = c\,X$ has, approximately, a Poisson distribution.

In the first approach for rescaling, we observe that the mean of Y is equal to c times the mean of X. So the mean of Y may be estimated as c times $\bar{x}$ where $\bar{x}$ is the sample mean of the concentration values. Also, the variance of Y may be estimated as $c^2 s_x^2$ where s_x^2 is the sample variance of the concentration values. Since the variance of Y should equal the mean of Y, we get $c = \frac{\bar{x}}{s_x^2}$. When there are nondetects, one could replace the nondetects by half the detection limit and then compute the mean and the variance of the sample concentration values. If this method is applied to the benzene concentrations in Example 10 in the EPA guidance document (U.S. EPA 1992) we get $c = 0.186$. If we round this value and take $c = 0.2$, then, in order for a Poisson model to have a chance of fitting the benzene data adequately, the concentrations must be expressed in parts per 200 million. Unfortunately, it may be easily verified that, with or without such a rescaling, the fit of a Poisson model to data is extremely poor.

The second approach to rescaling involves the use of the method of maximum likelihood to estimate the scale constant c and the Poisson parameter λ simultaneously. Interestingly, it can be shown that the likelihood function does not have a unique maximum value. So the likelihood approach does not lead to a useful solution for the rescaling problem.

Appropriate Alternative Methods

If Poisson-based methods cannot be used with concentration data, what methods are appropriate for cases with large fractions of nondetects? Fortunately, useful alternative methods are currently contained in the EPA guidance. Normal prediction limits and tolerance limits would be the preferred alternative if a normal distribution applies or can be obtained by suitable transformation. Cohen's or Aitcheson's adjustment are described in the EPA guidance for situations in which there are fewer than 50% nondetects. If there are between 50 and 90% nondetects, nonparametric prediction or tolerance limits are recommended. EPA recommendations notwithstanding, these methods can also be applied to data with greater than 90% nondetects. Since the upper limit can be no greater than the largest background observation, one cannot compute limits with larger confidence levels and smaller sample sizes. However, the fact that the nonparametric method does not extrapolate beyond the range of the observed data is a strength when the normality assumption is tenuous.

For cases in which the fraction of nondetects is large, however, the most appropriate alternatives might be those which consider only the number or fraction of detections, not the numerical values of the detected concentrations. These tests are based on the binomial distribution or a Poisson approximation to the binomial distribution. A "test of proportions" described in U.S. EPA (1989) tests a null hypothesis that the proportion of detections in the background data (population) is equal to the proportion of detections in the compliance data (population.) As we have said earlier, Poisson tolerance or prediction limits on a future number of hits could be used when the probability of a hit is small.

Summary and Conclusions

Tolerance and prediction limits based on the Poisson distribution have become standard recommended procedures for analyzing data with large numbers of nondetects (especially VOCs) from RCRA solid and hazardous waste sites. The purported advantages of the Poisson procedures compared to other alternatives seemed almost too good to be true. Unfortunately, things that seem too good to be true usually are, and this is no exception. We have revisited the development and application of these methods for two types of data, counts of analytical hits and actual concentration measurements. Each of these two applications was explored along two lines of reasoning, a first-principles argument and a simple empirical fit.

The application of Poisson-based methods to counts of analytical hits, including simultaneous consideration of multiple VOCs, appears to have merit from both a first-principles and an empirical standpoint. The concerns raised by Davis and McNichols (1988) regarding dependence between VOCs and different probabilities of detection for different VOCs are valid. However, it follows from the results of LeCam (1960) that when the probabilities of the events of concern are small, the Poisson approximation is still valid, even if the assumptions of constant p does not hold. We may expect the approximation to be satisfactory even when the independence assumption fails to hold, provided that the dependence is weak. The Poisson distribution appears to fit at least some data sets of hits/scan.

On the other hand, we have shown that the Poisson distribution is not appropriate for modeling concentration data, primarily

because the variance of the distribution does not scale appropriately with changing units of measurement. Tolerance and prediction limits based on the Poisson distribution are not scale invariant. While the molecular argument of Gibbons appears to make sense, his original argument did not consider the actual number of molecules present in a real water sample and, therefore, did not calculate the rate parameter, λ, properly. When we perform the first principles development correctly, we find that (1) the Poisson distribution is of no value because the normal approximation to the exact binomial distribution is applicable, and (2) the variance explained by the molecular argument is insignificant, by several orders of magnitude, compared to actual sample-to-sample variance since it does not consider measurement and other sources of error.

There are serious practical problems associated with using the Poisson-based limits with concentration data. By changing the units of observation in example problems drawn from EPA guidance, we demonstrated that use of the Poisson-based tolerance and prediction limits can result in significant errors. In short, neither the Poisson distribution nor associated tolerance or prediction limits should be used with concentration data.

Thus, the logical alternatives for situations in which the fraction of nondetects is large appear to be (1) the nonparametric tolerance or prediction limits currently recommended in EPA guidance in which the largest observation establishes the limit for most practical sample sizes, or (2) methods based on the binomial distribution (also described in EPA guidance) such as a test of proportions or Poisson tolerance and prediction limits on the number of hits.

References

American Society for Testing and Materials (ASTM). 1996. Provisional standard guide for developing appropriate statistical approaches for ground-water detection monitoring programs. PS 64-96. West Conshohocken, Pennsylvania: ASTM.

Davis, C.M., and R.J. McNichols. 1988. Discussion of "Statistical models for the analysis of volatile organic compounds in waste disposal sites," by R.D. Gibbons. *Ground Water* 26, no. 5: 649-651.

Davis, C.M., and R.J. McNichols. 1994a. Ground water monitoring statistics update: Part I: Progress since 1988. *Ground Water Monitoring and Remediation* 14, no. 4: 148-158.

Davis, C.M., and R.J. McNichols. 1994b. Ground water monitoring statistics update: Part II: Nonparametric prediction limits. *Ground Water Monitoring and Remediation* 14, no. 4: 159-175.

Dwass, M.D. 1970. *Probability and Statistics: An Undergraduate Course.* Menlo Park, California: W.A. Benjamin Inc.

Feller, W. 1970. *An Introduction to Probability Theory and Its Applications,* Vol. II, 2nd ed. New York: John Wiley & Sons Inc.

Gibbons, R.D. 1987. Statistical models for the analysis of volatile organic compounds in waste disposal sites. *Ground Water* 25, no. 5: 572-580.

Gibbons, R.D. 1988. Reply to Discussion of "Statistical models for the analysis of volatile organic compounds in waste disposal sites," by C.M. Davis and R.J. McNichols. *Ground Water* 26, no. 5: 649-651.

LeCam, L. 1960. An approximation theorem for the Poisson binomial distribution. *Pacific Journal of Mathematics* 10: 1181-1197.

Ott, R.L. 1993. *An Introduction to Statistical Methods and Data Analysis,* 4th ed. Belmont, California: Duxbury Press.

U. S. Environmental Protection Agency (EPA). 1992. Statistical analysis of ground-water monitoring data at RCRA facilities: Addendum to interim final guidance. Washington,D.C.: Office of Solid Waste.

U. S. Environmental Protection Agency (EPA). 1989. Statistical analysis of ground-water monitoring data at RCRA facilities: Interim final guidance. Washington, D.C.: Office of Solid Waste.

Zacks, S. 1970. Uniformly most accurate upper tolerance limits for monotone likelihood ratio families of discrete distributions. *Journal of American Statistical Association* 65, no. 329: 307-316.