

TECHNICAL SUPPLEMENT #1
Calculating Background Using MTCASat

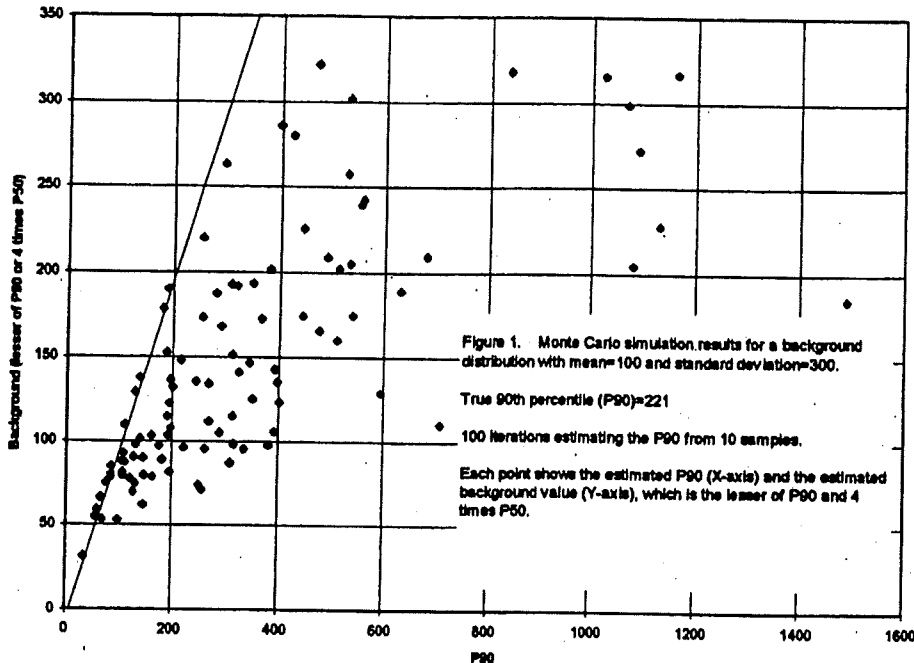
by Nigel Blakley

During beta-testing of the forthcoming upgrade of MTCASat for Excel 5.0, we performed additional evaluations of the background calculation method used in MTCASat. This report describes a potential problem we discovered for estimating background from sampling data, and explains how the problem is minimized by correct use of the decision rule it is handled by the decision rule for determining background as discussed in the *Statistical Guidance for Ecology Site Managers*.

For most calculations, the Background module of MTCASat uses a probability plot method, similar to the procedure described in Gilbert (1987, p. 168), to estimate parameters for a normal or lognormal background distribution fitted to the sampling data.

Estimating parameters of an underlying statistical distribution from a sample of values always introduces uncertainty; the sample is not a perfect representation of the true population distribution. Monte Carlo simulations used in testing MTCASat showed that the probability plot method can produce highly variable estimates of some lognormal parameters, including the 90th percentile. Furthermore, those estimated values were not distributed evenly around the true value, but were skewed towards the higher values and thus were an issue for overestimates of background parameters. (The estimates based on the probability plot method are "biased high" in a statistical sense). The variability and overestimation of 90th percentile values calculated by MTCASat are greatest in cases where (1) the estimate is based on a small number of samples (e.g., 10), and (2) the true distribution of the background data is a highly skewed lognormal curve. A lognormal distribution becomes more skewed (i.e., has a longer tail to the right) as the coefficient of variation (standard deviation divided by the mean) increases. For this evaluation, lognormal distributions with a coefficient of variation (CV) of greater than about 1 can be considered highly skewed.

Results from one of the Monte Carlo simulations (Figure 1) illustrate the problem and show how it is handled by the decision rule for calculating background. In this example, the MTCASat Background module was used to analyze 10 samples drawn at random from a lognormal distribution with mean 100 and a standard deviation of 300 (CV=3). The sampling procedure was repeated to give data for a total of 100 trials (each consisting of 10 random samples).



In this plot, the 90th percentile for background (P90) estimated from the 10 samples in each trial is shown on the X-axis. The plot shows that these estimates do not cluster symmetrically or closely around the true value for P90 (=221) but instead show a large amount of scatter, especially to the right (overestimates). In one trial the estimated P90 was as high as about 1500, approximately seven times the true value.

These simulation results show that estimates of P90 can vary significantly from the true value, particularly as overestimates, in cases of a highly skewed lognormal background distribution and only 10 background samples. However, the simulation results in Figure 1 also show that more reasonable values for background are produced after applying the correct MTCA decision rule for calculating background. This rule states that, for a lognormal distribution, background is set at the estimated P90 or 4 times the estimated 50th percentile (4P50), whichever value is smaller (see section 4.3.3.2 in the *Statistical Guidance for Ecology Site Managers*). Note that for a true lognormal distribution with mean of 100 and CV=3, the 4P50 value of 126.5 is less than the P90 value of 221 and would therefore be the calculated background value.

Figure 1 shows the background value calculated with *MTCASat* in each of the 100 trials on the Y-axis, after applying the decision rule. Points lying on the line $y=x$ indicate that background was set at P90 in these trials (i.e., trials where $P90 < 4P50$, as estimated from the samples, even though this is not the case for the true background distribution). Points to the right of this line indicate that after applying the decision rule, background was set at 4P50 instead (i.e., $4P50 < P90$, based on estimates from the samples).

Despite the extremely high estimates for P90 in some trials, the final background values in this simulation were much less variable, ranging from about 50 to 350. This amount of variation in

background values between trials does not seem unreasonable, given the small sample size (N=10) used for the statistical analysis in each trial.

In summary, estimates of P90 from *MTCASat* can differ considerably from true values when the number of samples collected is small and background has a highly skewed lognormal distribution. However, from this and other Monte Carlo simulations not reviewed here, it appears that the decision rule for calculating background effectively constrains the value to be used under MTCA to a more reasonable range.

Most background data analyzed in Ecology's publication, *Natural Background Soil Metals Concentrations in Washington State*, were reported to have lognormal CVs in the range of 0.5-1.0. In the Monte Carlo simulations with N=10 sampled values, the normal probability plot method appears to provide a relatively narrow and minimally skewed distribution of estimates for P90 for a lognormal CV of 0.5, and background was set at P90 in every trial. However, at CV values of 1 or higher, background was set at 4P50 in some simulation trials where the estimated P90 was high. These results show the importance of always applying the MTCA decision rule for determining background (the lesser of P90 or 4P50) in reducing unwarranted variability in background estimates.

The variability and bias in background estimates is greatest for cases of small sample sizes and a relatively large lognormal CV. Particularly in those cases, increases in sample size should be considered to reduce the variability in background estimates.

Reference: Gilbert, R.O. 1987. *Statistical methods for environmental pollution monitoring*. Van Nostrand Reinhold, New York.