CONTAMINATED SITES STATISTICAL APPLICATIONS GUIDANCE DOCUMENT NO. 12-13

CLASSIFICATION

A guide for data analysts, project managers and reviewers on the classification of contaminated soil

This guidance document is one of a series that outlines important basic statistical concepts and procedures that are useful in contaminated sites studies. BC Environment recommends that these suggestions be followed where applicable, but is open to other techniques provided that these alternatives are technically sound. Before a different methodology is adopted it should be discussed with BC Environment.

April 2001

THE GENERAL IDEA

During the remediation of most contaminated sites, the affected material throughout the site eventually has to be classified into one of several contamination categories, usually according to whether the contaminant concentrations are above or below specified regulatory thresholds. In the system of classification used by the BC Ministry of the Environment, for example, the most contaminated material is assigned to the "special waste" and "waste" categories, and less contaminated material is classified from industrial quality to agricultural quality. The criteria for these categories are defined by the *BC Waste Management Act*, and the *Contaminated Sites* and *Special Waste Regulations*. When it comes time to assign contaminated material to one of these categories, we very rarely know the exact contaminant concentration of the material in question. Instead, we base our classification on samples that represent only a tiny fraction of the total amount of material that needs to be classified.

This guidance document discusses a variety of topics that all relate to the issue of classification. It begins with a brief presentation of some of the terminology that is commonly used when discussing misclassification, and moves on to a discussion of contouring, with a brief summary of the techniques that are commonly used to interpolate between the available sample data. It addresses the issue of quantifying the uncertainty on local estimates and presents a procedure for developing maps that directly show the probability of encountering contamination rather than showing estimates of the contaminant concentration. The volume–variance effect is then discussed, along with the related issue of selectivity and the concept of the volume of selective remediation.

This document focuses on the issue of classification of *in situ* material based on *in situ* samples. The document entitled *STOCKPILES* discusses classification of material that has already been excavated and is awaiting classification in stockpiles. For some small sites, the affected material does not need to be locally segregated into different categories. For such situations, where all the material on the site is assigned to a single category, the reader should consult the documents entitled *STOCKPILES* and *ESTIMATING A GLOBAL MEAN*.

MISCLASSIFICATION

Whenever we attempt to classify material as being above or below some specified threshold, there are two kinds of errors that may occur. The first, which is often called a "false negative error", occurs when we mistakenly classify material that is actually above the threshold as being below the threshold. The second, which is often called a "false positive error", occurs when we mistakenly classify material that is actually below threshold as being above the threshold.

For the remediation of contaminated sites, where classification is usually accomplished by comparing local estimates of the contaminant concentration to some regulatory threshold, a false positive error results in the remediation of material that did not, in fact, have to be remediated; a false negative error results in the failure to remediate material that should, in fact, have been remediated. These two types of errors have different impacts. False positive errors cost us the money that it requires to excavate and treat soil that could have been left unexcavated and untreated; residual contamination that results from false negative errors has the potential for damaging human health and the environment.

Though remediation should, ideally, minimize both types of misclassification, it is usually difficult to minimize both simultaneously. Decreases in the probability of a false negative error usually entail increases in the probability of a false positive error. Damage to human health and the environment is usually regarded as a much higher cost than the money spent on additional remediation. As a result, the focus of most statistical guidance on classification is to keep the false negative rate below some acceptable minimum.

CONTOURING

The most common approach to classification is to use the available sample data to contour or interpolate the contaminant concentrations into areas that have not been directly sampled. Figure 1 shows lead data from a site affected by airborne contamination from a lead smelter located roughly in the center of the map area. Using these data, a contour map, such as the one shown in Figure 2 can be constructed and used as the basis for classification. For this particular site, the target threshold for remediation was 500 ug/g, the first of the thicker contour lines shown in Figure 2.

Contouring is not a unique exercise: there are many different algorithms that can interpolate sample data into unsampled areas. The contour map shown in Figure 2 was created using a geostatistical procedure known as "kriging", which uses statistical information on the pattern of spatial continuity to calculate appropriate weights for the nearby samples in the vicinity of any point at which we need a local estimate. If there are enough samples on which to base an analysis of spatial continuity, as described in *SAMPLING PLANS*, then kriging will usually produce an excellent interpolation of the available data.

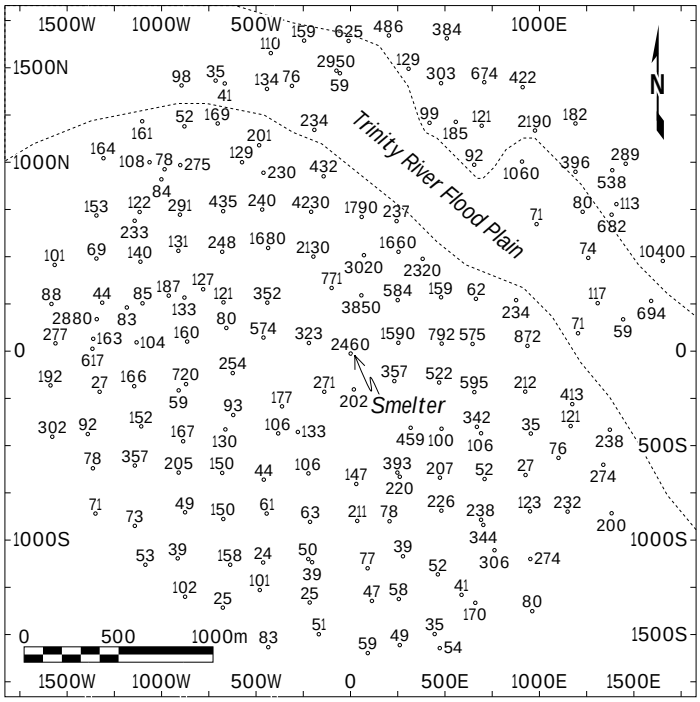

Figure 1 Lead sample data in the vicinity of a smelter.

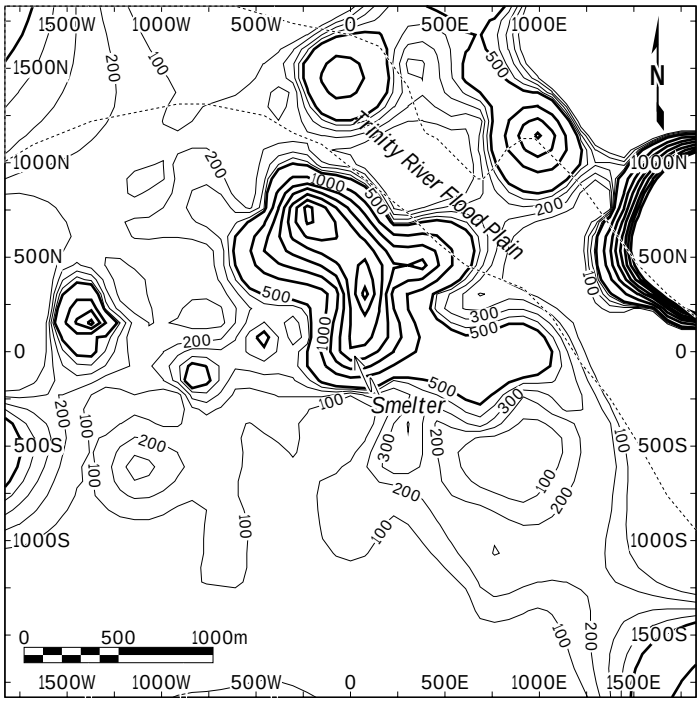

Figure 2 A kriged contour map of the lead data in Figure 1.

Other contour methods that are widely available in commercial software packages include the "inverse squared distance" method and "spline" or "minimum tension" interpolation. With an inverse distance procedure, each sample is given a weight that is inversely proportional to the squared distance from that sample location to the location of the point at which we need a local estimate. The "spline" or "minimum tension" approach to interpolation finds a surface that passes through all of the available sample data values and that has variations that are as smooth and gentle as possible.

Isaaks and Srivastava (1989) provide a practical introduction to kriging and to some of the more traditional and simpler alternatives, such as the inverse squared distance approach.

QUANTIFYING UNCERTAINTY

Whenever we make an estimate, regardless of the interpolation procedure we choose, one of the few things that we know about our estimate is that it is very likely wrong — we would be unimaginably lucky to hit the nail right on the head and predict the exact value of the unknown concentration. Since error is inevitable in any estimation procedure, our classification should not be based merely on estimated concentrations, but should also try to take into account the uncertainty on these estimates. In the same way that we consider a pessimistically high estimate of the global mean when we classify stockpiled material (see the document entitled *STOCKPILES*), so too should we consider the effect of uncertainty on any local estimates that we calculate.

Geostatistics provides one approach to incorporating estimation uncertainty into a classification procedure. The geostatistical interpolation known as kriging produces, along with an estimate of the concentration, a quantity called the "estimation variance" that is often used to build confidence intervals around local estimates. Unfortunately, this procedure assumes that the distribution of error is normal, an assumption that is rarely justified when the underlying population is skewed (as are the contaminant concentrations for most contaminated sites). Furthermore, the traditional calculation of the estimation variance does not take into account the sample data values themselves, but considers only their location. In many instances, the uncertainty in our predictions is linked as much to the actual data values as to their location, and the fact that the estimation variance does not depend on the data values makes it inappropriate as a measure of local uncertainty in such situations.

Probability maps

Geostatistics also offers an alternate method for quantifying uncertainty, a procedure known as "indicator kriging" that results in a map that directly displays the probability that the contaminant concentration exceeds a specified threshold.

Figure 3 shows the 50 ug/g "indicators" of the lead data shown earlier in Figure 1. These indicators are simply 0's and 1's that record whether each individual sample is above or below the 500 ug/g threshold. At any location where the lead concentration is known to be below 500, the corresponding indicator is 0, and at any location where the lead concentration is known to be above 500, the corresponding indicator is 1.

Figure 4 shows a kriged contour map of these indicators. By interpolating between a set of 0's and 1's, we end up with a map of intermediate values between 0 and 1; this contour map can be interpreted directly as a probability map. For example, in the region immediately north of the smelter, where the contour lines show values above 0.8, we can interpret this as a 80% probability that the contamination in this area is above the 500 ug/g threshold. Similarly, in the southern third of the map area, where the interpolated values drop below 0.1, there is less than a 10% chance of encountering lead contamination in excess of 500 ug/g.

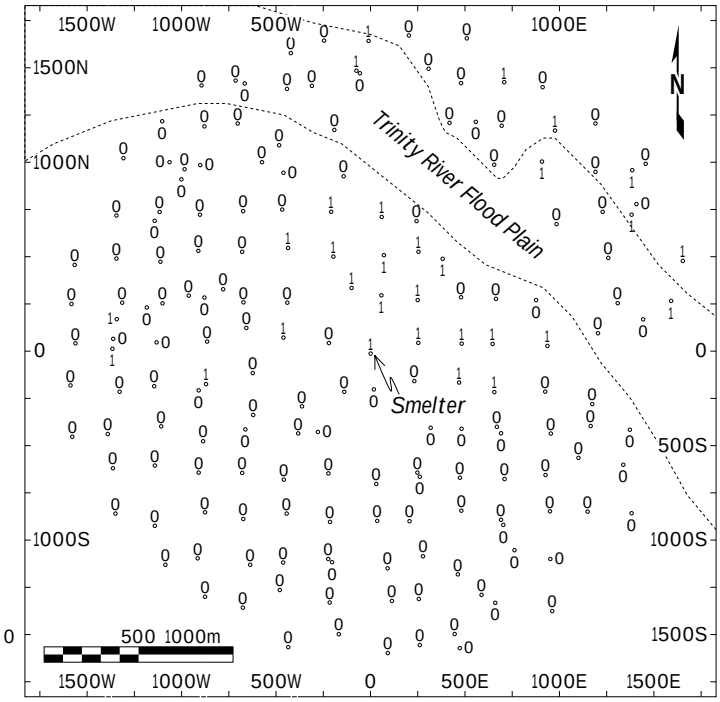

Figure 3 500 ug/g indicators for the sample data in Figure 1.

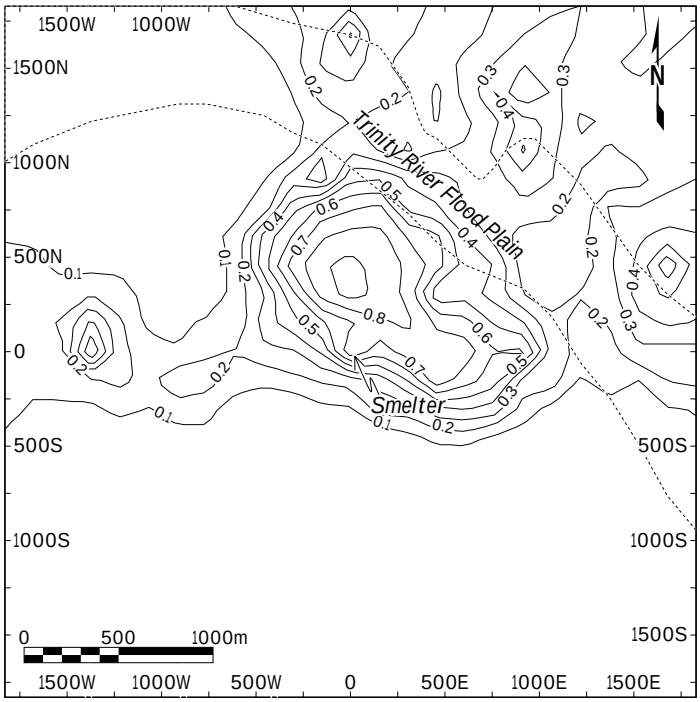

Figure 4 A probability map based on indicator data in Figure 3.

The reason that interpolated indicators can be viewed as probabilities is that individual indicators are themselves probabilities. In Figure 3, those locations where the indicator is 0 are locations where the chance of encountering lead in excess of 500 ug/g is 0. Similarly, those locations where the indicator is 1 are locations where the probability of encountering contamination in excess of 500 ug/g is 1, or certainty. With the indicator data being local probabilities, an interpolation of these data leads to an estimation of the probability of encountering contamination in excess of the threshold used to define the indicators.

Though the use of indicators was first presented in a geostatistical setting, with kriging as the interpolation procedure, the same approach can also be used with the more traditional and simpler interpolation procedures such as inverse squared distance or splines.

THE VOLUME–VARIANCE EFFECT

A histogram of contaminant concentrations for very small volumes, such as those typically collected by a split-spoon sampling device, will show a greater spread and more skewness than histograms of the same contaminant concentration measured on larger volumes, such as truckloads. The tendency for values based on larger volumes to be less variable is often called the "volume-variance" effect or the "support" effect ("support" being the word that statisticians often use for the size and geometry of a sample). This support effect is due to the fact that we are less likely to observe extreme high or low values in larger volumes of material since there is more opportunity for mixing high and low values together. The gradual disappearance of the extreme values causes the spread of the distribution to decrease and the symmetry of the distribution to increase.

The support effect has important implications for making predictions about remediable volumes. The problem in most contaminated site studies is that the remediable volume is quite different from the volume of the available samples. With the available samples coming from a population that is based on small volumes and is highly variable, and the remediation plan calling for estimates for a different population that is based on much larger volumes that should be much less variable, we have to be careful about how we use the sample information to make predictions. Parker (1979) and Isaaks and Srivastava (1989) provide an overview of the volume-variance problem, on its implications and on how to deal with it.

THE VOLUME OF SELECTIVE REMEDIATION

When we say that the goal of our classification is to correctly identify *"all material for which the lead contamination exceeds 500* ug/g", what exactly does this mean? By using a concentration-based criterion, such a statement carries some implicit assumption about the volume; without a volume of material, the concept of a concentration would have no meaning. What is the volume at which the remediation is intended to succeed? Do we clean up every spoonful of material whose lead concentration exceeds 500 ug/g? Or do we clean up every shovelful whose lead concentration exceeds 500 ug/g? Or does the statement pertain to entire truckloads? Or to even larger volumes? If we intend to go after every spoonful of material in excess of 500 ug/g then we'll have to do a lot of sampling or excavate virtually all of the soil as contaminated (or both). If we intend only that the remediation clean up truckload sized volumes, then fewer samples will be required and less soil will probably need to be excavated.

Without consensus on the volume to which a regulatory threshold applies, detailed remediation planning is usually premature. If the engineering plan is designed to correctly classify large truckloads of material while the regulatory agency intends that much smaller volumes be correctly classified, then the reme-

diation, even if successful according to those who designed it, may not satisfy the regulatory agency. This is typically the case when remediation with large equipment is thought to be complete, only to discover that verification samples still encounter residual contamination. Even if the original remediation did, in fact, remove all contaminated material at the scale of large trucks and no entire truckload of contamined soil remains on the site, it is still possible that small surface samples will encounter contamination in excess of the regulatory threshold.

The "volume of selective remediation", or VSR, is the smallest volume of material that the remediation exercise intends to segregate into one category or another. When deciding what the VSR should be, the first consideration should be the basis for the regulatory thresholds. The adverse effects of a particular contaminant often depend on a variety of factors, including the exposure pathway, the chemical form of the contaminant and its physical form. The selection of a target threshold for a remediation should address the minimum volume for which the adverse effects can manifest themselves. For example, with lead contamination in the soil in a residential area, the primary focus of the remediation may be to minimize the risk of a child picking up a handful of dirt and eating it. For this situation, the VSR is usually considered to be very small, about a handful of dirt. As another example, the most adverse effect of mercury contamination in marine sediments may be the ingestion of mercury by bottom-feeding organisms that eventually become food themselves for larger aquatic life and, eventually, for humans. For this situation, the VSR may be the area over which a bottom-feeder is likely to browse in its lifetime.

In addition to considering the adverse effects of the various contaminants, the selection of an appropriate VSR will often also take into account the practical realities of the remediation exercise. Even though we would like to segregate each and every handful of lead contaminated soil, there is no equipment that can economically achieve such a fine level of selectivity. If the smallest loaders we can use have 3 m^3 buckets, the VSR might reflect this minimum loader size.

The VSR we achieve in actual practice may be much larger than the intended level of selectivity if we do not address the issue of selectivity in all of the different aspects of the remediation and verification design. For example, if the remediation plan is to excavate the upper one metre of soil wherever the available samples suggest contamination, and if the samples are spaced at 10 m, then the minimum volume that we actually end up segregating is roughly $10 \times 10 \times 1 = 100$ m^3 regardless of how small our equipment might be.

The volume of selective remediation and probability maps

When using a probability map, such as the one shown earlier in Figure 4, as the basis for classification, it should be recognized that the probabilities being shown on this type of map refer to the volume of the samples used to define the 0/1 indicators. For example, in areas where the probability map in Figure 4 shows a 50% probability of encountering contamination, single samples from these areas will have a 50% chance of being contaminated. This does not mean that there is a 50% chance that truckloads of material from the same areas will be contaminated.

The selection of an appropriate probability threshold for classification depends on the relative concentrations of the contaminated and uncontaminated material. For the lead example shown in Figure 4, the contaminated samples above 500 ug/g have an average lead concentration of 1850 ug/g, while the uncontaminated samples below 500 ug/g have an average lead concentration of 158 ug/g. With these relative contaminant concentrations, any area that has even a 20% chance of being contaminated will likely have VSRs whose average contaminant concentration exceeds 500 ug/g. The 0.2 contour line from the probability map in Figure 4 is therefore a reasonable basis on which to classify material as contaminated or not. Regions inside this 0.2 contour line have a good chance of containing VSRs whose average contaminant concentration is above 500 ug/g and which should, therefore, be classified as being above the threshold even if individual samples in these areas show lead concentrations less than the 500 ug/g threshold.

RECOMMENDED PRACTICE

1. Classification should be based on interpolation of *in situ* sample data and on the uncertainty in these estimates.

2. When affected material is being classified, the classification procedure should recognize that the regulatory threshold usually pertains to a volume of material that is different from the volume of material that is typically sampled. The "volume of selective remediation" should be explicitly stated during the design of an appropriate remediation strategy. The entire remediation strategy, including equipment selection and the sampling plan that will be used for local refinement of preliminary excavation limits, should be designed so that material can be effectively segregated at the intended level of selectivity.

3. The goal of a classification exercise should be to ensure that there is a less than 5% chance of making a false negative error on a volume of material equal in size to the intended volume of selective remediation. This goal can be met with the use of probability maps and the selection of an appropriate probability threshold based on the relative contaminant concentrations of material above and below the threshold of interest.

REFERENCES AND FURTHER READING

In addition to the other guidance documents in this series, the following references provide useful supplementary material.

Davis, J.C., *Statistics and Data Analysis in Geology*, 2nd edition, John Wiley & Sons, New York, 1986.

Isaaks, E.H. and Srivastava, R.M., *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 1989.

Jones, T., Hamilton, D. and Johnson, C., *Contouring of Geological Surfaces with the Computer*, Van Nostrand Reinhold, New York, 1986.

Parker, H.M., "The volume–variance relationship: a useful tool for mine planning," in *Geostatistics*, edited by P. Mousset-Jones, McGraw Hill, New York, 1980.