STOCKPILING

A guide for data analysts, project managers and reviewers on statistical issues related to stockpiled material

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 contaminated sites, the affected material is not always sampled and classified *in situ*; occasionally it is first stockpiled and then classified on the basis of samples taken directly from the stockpiles. Though this is generally not a good practice, it may be inevitable, unfortunately, when the logistics of the remediation require material to be excavated before chemical analyses are available. This guidance document addresses the issue of stockpiling and discusses why it is regarded as an inefficient and possibly ineffective approach to remediation. It also provides guidance on sampling stockpiles and on using these samples to classify the material.

Many of the issues raised in this guidance document are also discussed in other documents in this series. The reader should also read the documents entitled *ESTIMATING A GLOBAL MEAN*, *COMPOSITE SAMPLES*, *CLASSIFICATION*, and *DESIGNING A SAMPLING PLAN*, each of which contains information that is relevant to the issues discussed in this guidance document.

DRAWBACKS OF STOCKPILING

Loss of spatial context

The most effective and efficient approach to remediation is to use *in situ* samples, along with qualitative information about the site history and usage, to prepare maps or three dimensional models of the contamination as it exists *in situ*. Once material has been excavated and stockpiled, it usually has lost its spatial context and much of the information about site history and usage loses its value. For example, it may become apparent during the remediation of a site that the contamination has been dispersed along old and leaking waste-water lines that run underneath several buildings on the site. This insight could be used to refine models of *in situ* contamination so that the most contaminated material immediately adjacent to the old waste-water lines can be segregated from the less contaminated material on the site. If the affected soil has already been stockpiled, however, then we have likely mixed highly contaminated material with weakly contaminated material and have lost the opportunity to effectively segregate the affected soil into appropriate categories.

Even in studies where the relevant physical and chemical controls are understood, the remediation usually encounters unexpected "hot spots". When these are encountered *in situ*, it is possible through further sampling to delineate their lateral extent and to effectively segregate the "hot spot" in the appropriate contamination category. When a stockpile sample encounters an anomalously high analytical value, it is virtually impossible to tell where the rest of the "hot spot" has ended up; often, the only environmentally prudent approach is to condemn the entire stockpile as belonging to the contamination category of the single anomalous value.

Whether or not the stockpile samples encounter anomalously high values, the loss of spatial context reduces the confidence of any estimates we might need to make with the available sample data. Closely spaced *in situ* samples usually show more similarity in their contaminant concentrations than do closely spaced stockpile samples; the acts of excavating and stockpiling tend to destructure the *in situ* pattern of spatial continuity. The consequence of this for statistical studies of contaminated sites is that our predictions of the contaminant concentrations in material that we have not directly sampled will be more reliable when we use *in situ* samples to estimate nearby *in situ* concentrations than when we use samples from a stockpile to estimate contaminant concentrations elsewhere in the same stockpile.

Dilution

A second drawback of stockpiling is that it can result in highly contaminated soil being diluted with uncontaminated (or less contaminated) soil. If the relative proportion of the highly contaminated soil is low, then this dilution may cause material that should have been classified as contaminated to pass as uncontaminated. Since the adverse effects of many contaminants are directly related to the total mass or quantity of the contaminant, rather than to its concentration, such dilution of contaminated material is not an appropriate practice.

Difficulty of sampling

One of the other major shortcomings of stockpiling is that it is very difficult to obtain samples that fairly represent the entire stockpile. As discussed in *SAMPLING PLANS*, one of the important principles in sampling is that every member of the larger population has the same chance of being selected. Regrettably, stockpile samples are rarely fair; stockpile samples are commonly "grab" samples that are collected where the material is most accessible: usually from the surface of the pile, and often near the base. If the stockpile was homogeneous, then surface samples might be an acceptable basis for determining the average concentration of the entire pile. Stockpiles are rarely homogeneous, however, and there are a variety of reasons why surface samples might be badly biased.

When excavated material is dumped onto a pile, it inevitably segregates according to grain size as it cascades down the slope of the existing pile. If the contaminants are preferentially concentrated in the finer grain sizes — the silts and fine grained sands — then the material that accumulates along the toe of the pile will tend to show higher concentrations than the material that accumulates along the crest. Similarly, if the contamination is concentrated in the coarser fractions — the gravels and coarse sands — then the concentrations will tend to be lower at the toe and higher along the crest. Stockpiled material also segregates according to lithology; sticky lumps of clayey material, for example, will tend to end up in different parts of the stockpile than will loose dry soil.

Another reason that surface samples may be badly biased is that weathering often affects the contaminant concentrations on the exposed surface of a stockpile. With heavy metal contaminants, for example, rainwater may leach contaminants from the surface deeper into the pile. Similar problems arise when the contaminants are volatile organic compounds; the surface of the pile will often show considerably lower concentrations than the interior core since the surface material has had greater exposure to the air for a longer period of time.

STOCKPILE DESIGN

Given the various drawbacks of stockpiles as an intermediary step in remediation, the following principles should be used when designing stockpiles:

1. Wherever possible, stockpiling should always be preceded by *in situ* sampling and mapping of the contaminant concentrations. This *in situ* information should be used to identify areas that are sufficiently homogeneous that the mixing of material from different contamination categories is avoided.
2. To minimize the possibility of misclassification of contaminated material, the size of stockpiles must be kept relatively small, especially when the material is in the vicinity of any very highly contaminated *in situ* samples. Stockpiles should never exceed 50 m^3 when any of the stockpiled material is within 50 m of an *in situ* sample in which the contamination exceeds the concentration for BC Environment's "special waste" category. In no situation should stockpiles ever exceed 250 m^3.

STOCKPILE SAMPLING

Though proper stockpile sampling is difficult, it is not impossible. Three appropriate methods for sampling a stockpile are:

1. If the stockpile is small, create a random sub-sample by shovelling the pile into two separate piles, with one shovelful in every N shovels being randomly selected to go into the smaller pile that will form the sub-sample. With this approach, the selection of N depends on the size of sub-sample we can manage. If we need a small sub-sample, this random splitting of the entire pile may have to be repeated two or more times to obtain an appropriate sub-sample.
2. If the stockpile is too large for the previous procedure to be pragmatic, then collect samples at a regular spacing from vertical borings that completely penetrate the pile. These vertical borings should either be located randomly on the pile or should be located on a regular grid that covers the areal extent of the pile. In this approach, if the discrete samples from a single vertical boring are composited to produce a single analysis for each vertical boring, then the calculation of the average concentration in the entire stockpile should recognize that the borings have different lengths. *ESTIMATING A GLOBAL MEAN* describes how to accommodate different weights in the estimation of a global mean and in the quantification of the uncertainty on such an estimate.
3. As the stockpile is being created, whether by a shovel, by a loader, or by a series of dumped truckloads, the material can be randomly sampled as it accumulates. This approach to stockpile sampling ensures that some samples from the core of the ultimate stockpile will be available when it comes time to classify the material. As an example, if we are building a 200 m^3 stockpile by dumping 10 m^3 truckloads, then we could choose a random sample from each truckload immediately before it is dumped. By the time the entire stockpile has been created, we will have twenty samples that do a much better job of fairly representing the entire pile than any twenty samples we could collect from the surface of the ultimate pile.

Regardless of the method used to collect stockpile samples, the sampling program should be accompanied by a QA/QC program that monitors and documents the reliability and repeatibility of the sample analyses.

CLASSIFICATION OF STOCKPILED MATERIAL

When an entire stockpile is being classified, there are two questions that need to be considered:

1. Is the average contaminant concentration in the entire pile above or below the classification threshold?
2. Is the pile sufficiently homogeneous that classification on the estimated mean is appropriate?

Classification based on the global mean

The most straightforward check of whether the stockpile should be classified as being above or below any specific threshold is simply to check the mean of the available samples. If the sample mean is above the threshold, then the entire stockpile must be classified as being above the same threshold.

Even if the sample mean is lower than the target threshold, this does not ensure that the average concentration in the entire pile is below the threshold. The arithmetic average of the available samples is only an estimate of the true (but unknown) average concentration of the entire stockpile. As discussed in *ESTIMATING A GLOBAL MEAN*, the reliability of this estimate depends largely on two factors: the number of available samples and on the spread of the available sample values, usually measured in terms of the variance or standard deviation.

When the available samples fairly represent the entire stockpile, the uncertainty on the estimate of the global mean can be expressed through a quantity that is usually called the "standard error":

$$\text{Standard error of global mean} = \sigma_{\text{global mean}} = \frac{s}{\sqrt{n}}$$

where s is the standard deviation of the available samples and n is the number of available samples. The standard error can be thought of as the standard deviation of the distribution of the underlying true global mean. Though there is only one true global mean, we don't know what it is and our uncertainty entails that there is some range of possible values; the standard error describes the breadth of this range. If the standard error is very high, then the range of possible values is very broad and we don't know very much about the true underlying mean; this can be caused either by having a large value of s (which means that the available sample values are very erratic) or by having a small value of n (which means that we have only a very few samples). If s is small or if n is large, then the standard error will be small, which signifies that the true underlying mean must fall within a narrow range of possible values.

Since there is always some uncertainty in our estimate of the true average concentration of an entire stockpile, the classification of stockpiled material should accommodate the fact that we don't really know exactly the true average concentration. For an entire stockpile to be classified as being below a specified threshold, we need to be at least 95% certain that the overall average concentration in the pile is below the target threshold. This can be accomplished by calculating a pessimistically high estimate of the average concentration that is often referred to as the "upper 95% confidence limit" of the global mean:

$$\text{Upper 95\% confidence limit} = m + 2 \cdot \sigma_{\text{global mean}}$$

where m is the arithmetic average of the available samples and $\sigma_{\text{global mean}}$ is the standard error described above. If this pessimistically high estimate of the global mean is below the target threshold, we have addressed the first of the questions given above, and can turn our attention to the issue of whether the pile is sufficiently homogeneous based on the global mean alone.

Classification for inhomogeneous piles

Whenever the stockpile samples suggest that even a pessimistically high estimate of the global mean is below the target threshold, it is important to address the issue of whether the material within the pile is sufficiently homogeneous to warrant classifying the entire pile as uncontaminated.

It is not appropriate to assume that a stockpile is homogeneous simply because the process of excavating and piling the material has mixed up the soil — stockpiling is not the same as blending. The blending that is accomplished in the stockpiles used in many industrial processes is not due to the casual mixing that occurs when the material is excavated and piled, but is the result of a carefully engineered stockpile. Blending piles typically are constructed with many thin layers and are reclaimed with specialized equipment that cuts across as many layers as possible to maximize the blending efficiency. The stockpiles used in contaminated site remediation exercises are not engineered as blending piles and homogeneity should not be assumed, but should be explicitly checked with the available sample data.

There are two recommended checks of homogeneity:

1. If any single analysis is more than twice the target threshold, then the entire pile should be classified as being above the threshold regardless of the estimated global mean.
2. If the coefficient of variation is larger than one, then classification of the entire pile should be based on the highest sample concentration regardless of the estimated global mean. Since the coefficient of variation is the ratio of the standard deviation of the samples to their mean, this second check entails that if the standard deviation of the samples is higher than their mean, then we should check to see if the highest sample value is above or below the target threshold.

Examples of use of classification criteria

Table 1 below shows examples of styrene analyses for discrete samples taken from stockpiled material that needs to be classified according to the BC Environment categories: "waste" if the styrene concentration exceeds 50 ug/g, and "industrial quality" if the styrene concentration does not exceed 50 ug/g. For each of the four cases shown in Table 1, the classification of the stockpile according to the criteria presented above is discussed below.

Table 1 Styrene measurements (in ug/g) from stockpiles.

Pile 1	Pile 2		Pile 3		Pile 4	
25	7	1	7	1	7	1
51	3	2	3	2	3	2
19	2	7	2	7	2	7
26	1	7	1	7	1	7
87	6	2	6	2	6	2
42	1	1	1	1	1	1
33	101	2	77	2	29	2
29	1	2	1	2	1	2
39	2	3	2	3	2	3
	4	2	4	2	4	2
	1	1	1	1	1	1
	2	7	2	7	2	7

Pile 1: With the nine available samples having a mean of 39 ug/g and a standard deviation of 18.9 ug/g, the standard error of the global mean is $18.9 \div \sqrt{9} = 6.3$ ug/g. The upper 95% confidence limit of the global mean is $39 + 2 \times 6.3 = 51.6$ ug/g. The chance that the true average concentration of the stockpile is above 50 ug/g is not negligible, and the entire stockpile would have to be classified as waste.

Pile 2: With the 24 available samples having a mean of 7 ug/g and a standard deviation of 19.5 ug/g the standard error of the global mean is $19.5 \div \sqrt{24} = 4.0$ ug/g. The upper 95% confidence limit of the global mean is $7 + 2 \times 4 = 15$ ug/g. Even though the mean of the entire pile is almost certainly below 50 ug/g, there is a single sample that is more than twice the waste threshold of 50 ug/g. This indicates that some "hot spot" material has inadvertently been included in a pile that was only very weakly contaminated, and the entire stockpile would have to be classified as waste.

Pile 3: With the 24 available samples having a mean of 6 ug/g and a standard deviation of 14.7 ug/g the standard error of the global mean is $14.7 \div \sqrt{24} = 3.0$ ug/g. The upper 95% confidence limit of the global mean is $6 + 2 \times 3 = 12$ ug/g. As with Pile 2, the true average concentration of the entire pile is almost certainly below 50 ug/g. This example also has a single anomalous value, the 77 ug/g analysis, that causes the standard deviation, 14.7 ug/g, to be noticeably higher than the mean, 6 ug/g. In this case, the coefficient of variation is bigger than one and the classification should be based on the highest value and the entire pile would have to be classified as waste.

Pile 4: Like the previous two examples, the sample mean (4 ug/g) and standard deviation (4.9 ug/g) are both low enough that the upper 95% confidence limit of the global mean is well below the 50 ug/g target threshold. Like the previous example, the coefficient of variation is above one and the classification should be based on the highest value. With the highest value being only 29 ug/g, the entire pile would not have to be classified as waste, but in the next lower category (industrial waste).

In all of the three last examples, Piles 2 through 4, the fact that the coefficient of variation is above one should cause the stockpiling practice to be suspended until the reasons for the lack of sufficient homogeneity can be documented and corrective action can be taken.

RECOMMENDED PRACTICE

1. Stockpiles should be created only where *in situ* sampling has confirmed that the material being stockpiled is homogeneous, with a coefficient of variation of one or less.
2. Stockpile sampling programs should be designed to ensure a fair representation of the contaminant concentrations in the entire pile. Particular attention should be paid to the possibility that the concentrations in the core of the pile are different from those on the surface.
3. Classification of stockpiled material should be based on at least five separate analyses, some of which may be composite samples, and on the following statistical criteria:
   (a) If the "upper 95% confidence limit of the global mean", as described in *ESTIMATING A GLOBAL MEAN*, is above the classification threshold, then the entire stockpile must be classified as being above the threshold.
   (b) If any single analysis is more than twice the classification threshold, then the entire stockpile must be classified as being above the threshold.
   (c) If the standard deviation of the available analyses is larger than their mean, then the stockpiled material should be classified according to whether the highest analysis is above or below the classification threshold.

   If the stockpiled material is classified as being above the threshold for the reasons (b) or (c), and not for (a) alone, then the stockpiling practice is not accumulating homogeneous material; in this event, the stockpiling practice should not continue until the reasons for lack of homogeneity have been documented and corrective action has been taken.
4. Though the classification of stockpiled material may make use of composite samples, all of the discrete samples should be analyzed separately for at least one in every ten of the stockpiles. If the analyses of the discrete samples have a coefficient of variation greater than one, then the stockpiling practice should not be continued until further *in situ* sampling and data analysis allow more homogeneous regions to be identified.
5. As with any sampling program, stockpile samples should be accompanied by a QA/QC study that allows the quality of the analytical values to be monitored and documented.

REFERENCES AND FURTHER READING

In addition to the other guidance documents referenced on the first page of this document, the following references provide useful supplementary material.

Cochran, W.G., *Sampling Techniques*, 3rd edition, John Wiley & Sons, New York, 1977.

Heuer, H., "Stockpiling and Blending of Bulk Materials", in *Stacking Blending Reclaiming of Bulk Materials*, edited by R.H. Wöhlbier, Series on Bulk Materials Handling, Volume 1, No. 5, Trans Tech Publications, Aedermannsdorf, Switzerland, 1977.

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