

SAMPLING PLANS

A guide for data analysts, project managers and reviewers on the design of sampling plans

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

The design of an appropriate sampling plan is a recurring concern throughout the course of a contaminated site project. The issue first arises when a site is suspected of being contaminated and preliminary reconnaissance is required. At this early stage, some thought needs to be given to samples that will provide useful information for more detailed studies that may follow. If the site is deemed to be contaminated, the sampling plan will frequently resurface as an important issue: when the site needs to be characterized so that a remediation strategy can be developed, when global estimates of contaminated volumes are needed for remediation planning, when "hot spots" are encountered during remediation and need to be delineated. At any of these stages, a poorly designed sampling plan can cause the remediation to be inefficient or, worse, ineffective. Sampling plan issues arise even after the remediation, when samples are needed to confirm whether or not the remediation was successful. A poorly designed sampling plan at this final stage can cause residual contamination to go undetected.

This document presents information and advice on designing an appropriate sampling plan. It begins with a discussion of what it means for a sample to be fair and then discusses the volume that a single sample can be assumed to represent. The remaining sections discuss sampling plans according to the goal of the study, and deal with statistical considerations that arise in designing sampling plans for initial reconnaissance, for global estimates, for local estimates and for detection of "hot spots".

All of the guidance given in this document, including the various formulas for calculating the number of samples and their spacing, assume that the sample information is accurate and precise. In addition to defining the number, location and spacing of samples, a sample design should also identify QA/QC procedures that will ensure the reliability and repeatability of the sample information. These QA/QC procedures should not focus solely on the analytical values, but should also address the quality of the sample location information. Another document in this series, *STATISTICAL QA/QC*, discusses the documentation and monitoring of the reliability of sample information.

HOW FAIR ARE THE SAMPLES?

When we are designing a sampling plan, it often helps to think of the sampling plan in terms of an exercise in democracy in which we have to select samples that fairly represent some larger population. The samples themselves should not be the main focus of attention; they are interesting only insofar as they provide insight into the much larger remainder of the population that has not been directly sampled.

A chronic concern of statisticians is that sampling be fair in the sense that every member of the larger population has the same chance of being selected. If sampling is not fair then some members of the population have a greater chance of being selected than others; statistical studies based on such biased samples often reach erroneous conclusions. One of the best historical examples of an unfair sampling is the poll that was conducted by the newspaper that earned a lasting place in the photographic record of the 20th century with its "Dewey Defeats Truman" banner headline. This embarrassing proclamation was based on a public opinion poll in which voters were randomly selected from the phone book. Though this method of sampling a population has now become conventional for many public opinion polls, it was not a fair method in the 1940's. At that time, owning a telephone was enough of a sign of affluence that wealthier voters had a slightly higher chance of being sampled than poorer voters. With the poll being inadvertently stacked with wealthier voters, the preference of the wealthy for Dewey, the Republican candidate, skewed the results enough that the pollsters reached a very wrong conclusion.

For many contaminated sites, the earliest samples are not fair since their selection is based on visual observation. A preliminary reconnaissance is more likely to sample material that looks "interesting" than material that is not visually distinctive. Though such samples may be very useful for establishing an understanding of the nature of the contamination, they usually impart awkward biases to statistical studies that aim at characterizing the entire site. For example, on a landfill site contaminated with glazing sludge from the manufacturing of ceramic products, visible layers of glazing sludge exposed in trenches may be preferentially sampled during a preliminary reconnaissance of the site since these layers are most likely to confirm the severity of lead, zinc and cadmium contamination. Later, when a remediation plan is being prepared for the entire site, which includes many landfill materials other than glazing sludge, the preponderance of highly contaminated samples from preliminary reconnaissance makes it difficult to develop an accurate 3D model of the contamination throughout the site. In this case, though the early samples may fairly represent the contamination within layers of glazing sludge, they do not likely fairly represent contamination in other layers.

Another example of unfair sampling is the siting of additional samples near anomalously high sample values from earlier sampling campaigns. As discussed later in this guidance document, this targetting of suspected "hot spots" does provide valuable information; nevertheless, it also compromises any statistical method that assumes the underlying population has been fairly

sampled. When additional samples have been preferentially located in areas that are suspected of being highly contaminated, sample statistics should not be used as estimates of the corresponding parameters of the underlying population. The sample mean, for example, is usually a biased estimate of the mean of the underlying population if samples are preferentially clustered.

Dealing with preferential sampling

Samples are often collected from contaminated sites before anyone has started thinking about statistical issues, and those who are later responsible for statistical data analysis and interpretation typically have to cope with an unfair sampling that is preferentially biased towards certain regions. In such situations, statistical studies that address the remediation of the entire site should find some appropriate method for mitigating the effect of preferential sampling.

One solution for dealing with samples that preferentially target certain regions is to separate the site into subpopulations. The case described in the previous section, for example, might best be handled by building a model of the locations of the layers of glazing sludge. The many samples from these glazing sludge layers can be used to model the spatial distribution of contaminants within these layers, while the spatial distribution in other layers can be modelled using any samples from layers other than those that were visually recognizable as glazing sludge.

It is not always possible to split a statistical study into separate subpopulations, especially when the preferential sampling is the result of successive infill samples in anomalously high areas rather than the result of an intentional preference for visually distinct material. In such situations, statistical procedures may have to accommodate the effects of preferential sampling by assigning each sample a declustering weight. Samples from regions that have been densely sampled, are given low declustering weights to reduce their influence, while those from sparsely sampled regions are given higher declustering weights. The guidance document entitled *ESTIMATING A GLOBAL MEAN* discusses declustering and shows how declustering weights can be used to develop more accurate estimates of the mean of the underlying population from spatially clustered samples.

The least desirable way of dealing with preferential samples is to ignore or discard them; it is always better to try to limit their spatial influence, either by delineating the spatial extent of the population to which they belong, or by assigning them a declustering weight based on their proximity to neighbouring samples. Though discarding samples is not a good final solution to the problem of preferential sampling, it is often an expedient way to check the sensitivity of a statistical procedure to clustered sampling. If there is a concern that preferential sampling may be leading to erroneous conclusions, the procedure can be repeated with a regularly spaced subset of the available samples. If the conclusions based on all available samples are different from those based on a regularly spaced subset, then the effect of spatial clustering is severe enough to warrant specific attention. A subset that is more regularly spaced than the entire data set can be selected by overlaying a rectangular grid over the site and randomly choosing one sample from each rectangular cell; the document entitled *RANDOMIZATION* provides guidance on procedures for such a random subsampling.

WHAT VOLUME DOES A SAMPLE REPRESENT?

When samples are needed for local remediation planning, the sampling plan should consider the “range of influence” of a single sample. For spatially erratic contaminants, the range of influence is short and an individual sample represents only a small region in its immediately vicinity. For contaminants that are very spatially continuous, the range of influence is longer and an individual sample is representative of a larger region. If the range of influence is short, detailed local remediation planning will require more closely spaced samples than if the range of influence is long. The design of an appropriate sampling plan for detailed local remediation planning therefore requires a good understanding of spatial variation.

The intuitive notion of the “range of influence” can be expressed quantitatively by the “range of correlation”, which is the distance at which pairs of sample values are no longer correlated. The range of correlation is usually determined by grouping pairs of samples according to their separation distance and plotting the correlation coefficient between the sample pairs as a function of the separation distance. Figure 1 shows such a plot, which is usually called a “correlogram”, for a PCB contaminated site. In this example, the PCB concentrations have a range of correlation of 50m in the N–S direction and 30m in the E–W direction. Up to this distance, a single PCB analysis will have some correlation with the unsampled PCB concentrations in its immediate vicinity. Isaaks and Srivastava (1989) present a practical introduction to a variety of tools for analyzing and interpreting the pattern of spatial variation.

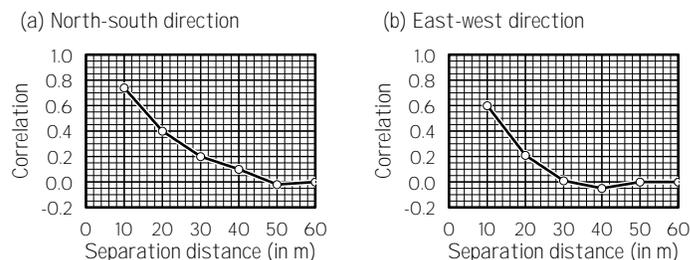


Figure 1 N–S and E–W correlograms of PCB concentration.

The example in Figure 1 shows one of the specific practical benefits of statistically analyzing the spatial variation: it may reveal that the sample spacing does not need to be the same in all directions. For many contaminated sites, the physical and chemical processes that influence the spatial distribution of the contamination are not isotropic — the contamination is likely to be more continuous in certain directions than in others. Without a study of spatial variation in different directions, we are not likely to recognize the opportunity for an efficient sample plan that takes advantage of the directional changes in the spatial continuity of the contamination.

Like all statistical characteristics, the correlogram may change locally; the range of correlation may not be the same throughout the site. On any site where erratic “hot spots” are suspected, we can check the possibility that high values have a shorter range of influence by calculating separate correlograms for high sample values and for low ones. Such information will be very useful for planning an efficient strategy to deal with any “hot spots” that may be encountered during remediation.

STATISTICAL CONSIDERATIONS IN SAMPLING

Preliminary reconnaissance samples

When a site is first suspected of being contaminated, reconnaissance is often performed to provide information on site history and usage. It is during this initial sampling that statistical rigour most frequently collides with other practical issues. Many statisticians disdain the haphazard approach to sampling that often characterizes reconnaissance, dismissing the samples as “grab samples” or mere “specimens” whose selection bias makes them of questionable value for statistical studies.

Unfortunately, preliminary site reconnaissance needs to address issues that are pre-statistical in the sense that statistical studies make little sense until these issues are resolved. For example, we should not initiate a statistical study until we know what we need to study; even without a statistically defensible sampling plan, preliminary reconnaissance can establish which contaminants exist in sufficiently high concentrations to warrant remedial action. Even when the contaminants of concern are already well known, a statistical study may be premature if the relevant populations are poorly understood. Reconnaissance can provide critical insights into the relevant statistical populations by gathering information on the physical and chemical factors that affect the spatial distribution of each contaminant.

Site reconnaissance need not be constrained by statistical concerns but should focus instead on establishing the contaminants of concern and their physical and chemical controls. For contaminants that occur naturally, such as most metals, the initial sampling of the site should also attend to the issue of establishing the statistical characteristics of background concentrations. By identifying and sampling material that should be uncontaminated, a site reconnaissance can document the distribution of background concentrations, information that can be used in subsequent statistical studies that try to split the data into “natural” and “affected” subpopulations.

Even though statistical issues need not be paramount during preliminary reconnaissance, subsequent statistical data analysis and interpretation will find the reconnaissance samples more useable if each sample is accompanied by documentation of the reasons for its collection. This may be a description of the visual appearance of the material, or may simply be a summary of the information that led to the belief that a particular location was contaminated. Reconnaissance samples will also be more useable in later studies if their locations are accurately recorded. If it is not possible at the time of the reconnaissance to survey the sample locations, they should be flagged or marked in the field so that they can be properly surveyed later.

Samples for global estimates

Once contamination on a site has been confirmed, the next statistical issues that arise usually pertain to global estimates, such as the average concentration of a contaminant over the site, or the overall proportion of the site that is contaminated. As discussed in *ESTIMATING A GLOBAL MEAN*, the uncertainty on the estimate of the mean of the underlying population is related to the number of samples used in the estimate. Uncertainty in an estimate is usually expressed in terms of its relative error, as a plus/minus interval around the estimate. The number of independent samples needed to ensure that the relative standard

deviation on the estimate of the global mean is less than $\pm R\%$ can be expressed in terms of the coefficient of variation of the individual samples:

$$\text{Number of samples} = [100 \times CV \div R]^2$$

As an example, if the available samples from a contaminated site show that the CV of individual lead samples is 1.5, and if we need a global estimate of the average lead concentration over the entire site that has a relative standard deviation of $\pm 20\%$ or less, then the number of independent samples we require is:

$$\text{Number of samples} = [100 \times 1.5 \div 20]^2 \approx 56$$

This calculation assumes the samples to be independent; this is rarely the case in contaminated site studies since there is usually some spatial correlation in the contaminant concentrations. In practice, the best we can do for a sampling grid that is intended for global estimates is to use a regular grid whose origin has been randomly selected.

Samples for local estimates

In addition to global estimates, statistical studies of contaminated sites also often call for local predictions. For example, remediation planning may require a map that shows the boundary between contaminated material that requires remediation and uncontaminated material that may be left *in situ*. Such a map is often constructed by contouring the available samples and then designating the remediation limit to be the contour line that corresponds to the remediation threshold.

Sampling plans for local estimation should take into account the range of correlation of the contaminant concentrations. For every location at which we need a local estimate, we should have at least one sample within the range of correlation. For any location where no samples fall within the range of correlation, local estimation will be fruitless since we do not have any information that directly correlates (even weakly) with the unsampled (and unknown) concentration we are trying to predict.

Once the range of correlation has been studied and documented through correlograms like the ones shown in Figure 1, this information can be used to design a sampling grid that will be appropriate for local estimation. In the same way that the formula given in the previous section allowed us to choose the number of samples that would limit the relative error in global estimates, the following formula provides insight into the sample spacing needed to limit the relative error in point estimates:

$$\text{Sample spacing} = \text{Range of correlation} \times [R \div (100 \times CV)]^2$$

As with the corresponding equation given earlier, R is the relative error, expressed in percent and CV is the coefficient of variation of the individual sample values. The range of correlation is the distance at which the correlogram shows sample pairs to be uncorrelated; as noted earlier, this distance may change with direction. As an example of the use of this formula, suppose that we are dealing with lead contamination that has a CV of 1.2 and a range of correlation of 135 m; furthermore, we would like each of our point estimates to have a relative error of less than $\pm 40\%$. The sample spacing should be

$$\text{Sample spacing} = 135 \times [40 \div (100 \times 1.2)]^2 \approx 12\text{m}$$

Though this formula provides a sample spacing that limits the relative error for point estimates, it is rare that we depend on

point estimates in remediation planning. In practice, we do not segregate contaminated from uncontaminated material point by point; the equipment used to implement the remediation strategy limits the volume of material we can effectively segregate. We are not interested in whether the contaminant concentration at a specific point exceeds the remedial action threshold; instead, we usually need to know whether the *average* concentration over a small area is above the threshold. An estimate of the concentration at a single point is less certain than an estimate of the average concentration over some larger area. For this reason, the preceding formula gives a sample spacing that is usually smaller than we actually need in practice.

The following formula provides a sample spacing that should limit the relative error on estimates of the average concentration over a small square whose side is B:

$$\text{Sample spacing} = \text{Range of correlation} \times [R \div (100 \times CV)]^2 + 0.75 \times B$$

As an example of the use of this formula, let us reconsider the previous problem, in which lead contamination had a CV of 1.2 and a range of correlation of 135 m. As before, we would like the relative error to be below $\pm 40\%$ but this relative error now refers to estimates of the average lead concentration over a 10m \times 10m square that represents the smallest amount of soil that can be practically segregated as clean or contaminated. For this situation, an appropriate sample spacing would be

$$\text{Sample spacing} = 135 \times [40 \div (100 \times 1.2)]^2 + 0.75 \times 10 \approx 20\text{m}$$

This formula will give an appropriate sample spacing for sample design problems in which the sample spacing is smaller than the range of correlation but larger than the size of the minimum volume of selective remediation. If the spacing calculated by this formula is larger than the range of correlation or smaller than B, then the result may not be appropriate; Isaaks and Srivastava (1989) present a more detailed discussion of estimation error and provide more general formulas that can be used to assist with the selection of an appropriate sample spacing.

Sampling for “hot spots”

Attempts to design a single sampling grid to delineate “hot spots” usually lead to a sample spacing that is so tight that the total number of samples becomes prohibitively costly and time consuming. The range of correlation is rarely constant throughout a contaminated site, but tends instead to be longer in areas with moderate and low contaminant concentrations and shorter in areas with high concentrations; similarly, the coefficient of variation is rarely constant throughout a contaminated site. As a result, the sample spacing needed to achieve a specified level of confidence in local estimates will vary throughout the site, with more samples usually being needed in anomalously high areas and fewer samples in moderate and low areas. Unfortunately, we usually can't take advantage of this fact in practice because we do not know where the anomalously high areas are located until we collect and analyze the samples. A practical and effective way around this problem is to design the sampling program so that the samples are collected in several stages rather than in a single campaign. With a multi-stage approach to sampling, the initial stage provides sample information on a relatively coarse grid, and each successive stage

adds infill or step-out samples near the locations of the high sample values from earlier stages.

Another approach to identifying and delineating “hot spots” is to use a less costly and more rapid analytical procedure, such as X-ray fluorescence, to supplement the more costly and time consuming chemical analyses. Rapid and cost-effective analytical procedures are typically less reliable and can not be used without a careful QA/QC program that constantly monitors the reliability of the information they generate. If a rapid and cost-effective analytical procedure has been properly calibrated through statistical QA/QC, it can provide a dense sampling of the site that can then serve as the basis for selecting optimal locations for more reliable (and costly) chemical samples.

RECOMMENDED PRACTICE

1. Initial reconnaissance sampling may be based on visual inspection and should be designed to provide data on the contaminants that exist and their maximum concentrations. For any contaminant that requires remedial action, the initial samples should document its physical and chemical controls and, if the contaminant occurs naturally, should also serve to establish the distribution of background concentrations. Though the initial sampling need not be statistically based, a complete description of the location of each sample and the rationale for its collection should be compiled so that subsequent statistical analysis can make appropriate use of the sample information.
2. For sampling plans that aim to provide information for global estimates, the samples should fairly represent the underlying population; this can be accomplished with a regular grid whose origin has been randomized. The total number of samples should take into consideration the level of confidence that global estimates will need to have to meet the study objectives and this, in turn, requires that the coefficient of variation be taken into account.
3. For sampling plans that aim to provide information for local estimates, their design should take spatial variation into account by ensuring that the spacing between samples is smaller than the range of correlation. If closely spaced samples are not already available from earlier sampling campaigns, they should be added at this stage to provide data for quantifying spatial variation.
4. For sampling plans that are intended to detect “hot spots” or to check for residual contamination following a remediation exercise, a multi-stage sampling plan should be used.

REFERENCES AND FURTHER READING

The guidance documents entitled *ESTIMATING A GLOBAL MEAN*, *RANDOMIZATION* and *STATISTICAL QA/QC* provide more information on topics related to the sampling plan issues discussed in this document. In addition to these, the following references also provide useful supplementary material.

- Cochran, W.G., *Sampling Techniques*, 3rd edition, John Wiley & Sons, New York, 1977.
- Isaaks, E.H. and Srivastava, R.M., *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 1989.