2. NATURE OF SOIL SAMPLING AND INCREMENTAL SAMPLING PRINCIPLES

2.1 Introduction

At their most basic level, the purpose of most environmental investigations is to make decisions about volumes of media which may contain contaminants at concentrations above some level of concern. The concentration of contaminants must be measured to determine whether remediation or other action is required. Such decisions are often made based on an estimate of the mean concentration of contaminants within the identified volume of media. Risk management decisions based on contaminant concentration estimates often involve large volumes of soil at individual sites. The totality of soil management actions throughout the nation each year has enormous public health and economic consequences.

Because it is impractical to collect and analyze the entire volume of soil for which decisions must be made, samples are collected and the results used to represent that entire volume of soil. The industry of environmental investigation, regulation, and laboratory analysis has, to a large extent, developed around the practice of using discrete samples to meet all decision goals, including estimating mean contaminant concentrations. There are many reasons why a mean concentration may be of interest for decision-making purposes, as discussed in Section 3 and Hyperlink 1.

Estimates of the mean may be based on arithmetic or geometric means of discrete sampling data or on upper confidence limits (UCLs). Since the costs of sample analysis can be high, the number of discrete samples collected is often driven down by project budgetary constraints. Collective experience, statistical simulation, empirical data, and sampling theory indicate that in many situations estimates of mean contaminant concentrations in soil made from small numbers of discrete samples are unlikely to be accurate or precise, and are, therefore, more likely to result in decision errors. These decision errors can go both ways. An erroneous decision of “clean” can lead to unacceptable exposure to contaminants. On the other hand, an erroneous decision of “dirty” can lead to a waste of resources “cleaning up” soil unnecessarily.

By its very nature, soil is a highly heterogeneous solid with many components. Sampling soil for the purpose of obtaining an estimate of the mean contaminant concentration is highly susceptible to sampling errors from a variety of sources. One goal of a sampling design should be to minimize the errors that can occur in each step of the sampling and analytical process. Historically, the focus has been on controlling errors associated with the analytical part of the process. A great deal of effort is invested in ensuring good data by requiring strict adherence to analytical methodologies and laboratory QA/QC procedures. But all this attention addresses only the tail end of the process. There are many more steps to the data quality chain that require attention for the output to be good data. According to USEPA’s soil screening guidance (USEPA 1996b):
Data users often look at a concentration obtained from a laboratory as being “the concentration” in the soil, without realizing that the number generated by the laboratory is the end point of an entire process, extending from design of the sampling, through collecting, handling, processing, analysis, quality evaluation, and reporting.

Steps usually overlooked when evaluating data quality include sampling design, sample collection techniques, sample processing, and field and laboratory subsampling. However, there is a growing body of evidence that the predominant source of error in the “entire process” to which USEPA refers is sampling error, which occurs because contaminant concentrations in soil are highly heterogeneous. Heterogeneity makes representative sampling difficult. Sampling errors are manifested as variability (i.e., imprecision observed as large differences in results between replicate samples) and/or bias in the data set (i.e., data results significantly over or under the true concentrations). Data variability is easily measured to evaluate the effects of sampling error on data quality. If concentrations are close to a decision threshold and sampling errors are not controlled, data variability can lead to highly uncertain estimates of mean concentrations, which in turn lead to considerable uncertainty about whether the mean is above or below a decision threshold. Hyperlink 2 provides an example illustrating the importance of considering data variability in decision making. Poorly thought out sampling procedures produce misleading data that can cause decision errors, as illustrated in Figure 2.1, no matter how good the analytical step is.

![Figure 2-1. Heterogeneous nature of contaminants in soils may lead to decision errors.](image)

This document focuses on how to obtain an unbiased and precise estimate of the mean concentration, including UCLs, in heterogeneous bulk volumes of soil with a relatively small number of laboratory analyses using a process called “incremental sampling methodology.” ISM is a suite of planning, sampling, sample preparation, and subsampling techniques that address heterogeneous soil contamination and thereby control sampling errors that may otherwise lead to incorrect decisions.

The sampling theory of Pierre Gy and his procedures for sampling bulk particulate materials have been used and validated for many years in the mining industry. However, only in the last few years has the environmental industry at large become familiar with this set of methods. ISM is based on many of the principles of Pierre Gy’s sampling theory and is intended to address the
problem of making decisions about highly heterogeneous bulk volumes of particulate material (e.g., soil) based on estimates of the mean derived from a relatively small number of samples of that material. Note, however, that many of the principles discussed in this section are also applicable to collecting and processing discrete samples. More attention to Gy theory and management of heterogeneity could reduce sampling error and improve data quality for discrete samples as well.

2.2 Soil Heterogeneity and Variation in Contaminant Concentrations

Taking a scoop of soil to collect and analyze as a soil sample may seem like a simple task. The critical question is whether that scoop of soil will produce meaningful data on the scale at which a decision is to be made. In other words, will results from a tiny sample provide the “right” answer for a volume of soil millions of times larger? Complications arise because soil is made of a variety of different materials which interact with contaminants in different ways. These materials generally take the form of particles of various sizes, which are composed of various mineral and organic substances. Many different kinds of soils exist, as defined by the types of minerals present and their ratios to each other and to organic carbon content. Different kinds of soils can differ widely in their physical and geochemical properties.

As a consequence of the physical and chemical properties of contaminants combined with differences in individual soil particles, contaminant atoms and molecules bind to some particles loosely, but more tightly to others. Further description of the interactions of soil with contaminants is provided in Hyperlink 3. Therefore, a sample of contaminated soil is a heterogeneous mixture of particles that are carrying different amounts of contaminant. This phenomenon is described by terms such as compositional heterogeneity (CH), microscale heterogeneity, or within-sample heterogeneity, and it creates a “nugget effect.” “Nuggets” form when contaminants preferentially attach to certain particles rather than others, such that contaminant-laden nuggets may be present in a matrix of other particles having less or no contaminant loading. Consider the effect of nuggets on concentration. Even if only one or two of these concentrated nuggets happen to be included in a very small sample when it is analyzed, a high concentration will be reported. If those same one or two nuggets were captured in a larger sample, a moderate concentration will be reported. If by chance no nuggets are present in the analyzed sample, then a low or nondetect concentration is reported. Hyperlink 4 provides an example of the “nugget effect” and how it may lead to decision errors. This is closely related to the concept of sample support, which is further discussed in Hyperlink 5.

In contrast to this microscale heterogeneity, which occurs within a single sample, large-scale heterogeneity refers to differences in concentration from location to location across an area, in other words, differences in how contaminants are spatially distributed throughout the DU. For example, contaminants may be released from leaking drums, creating distinct but rather small contaminated areas. Or contaminants may be released by single or multiple large-volume spills, which might create large patterns that are mostly uniform in concentration within the spill area but demarcated by a fairly sharp boundary. Some contaminants, such as pesticides, might have been sprayed only along the edges or in garden pockets of a residential yard. Or pesticide leftovers might be poured out in a single spot. Atmospheric deposition is a common release mechanism with the resulting spatial pattern affected by wind strength and direction and by distance from the source.
Short-scale heterogeneity refers to concentration differences observed at the scale of colocated samples. Colocated samples are taken from the “same” location in the field generally a few inches to a few feet apart and are traditionally considered to be equivalent, meaning that their concentrations are expected to be approximately equal. However, field experience shows that colocated samples often differ in concentration, sometimes quite drastically.

Heterogeneity at each of these different scales poses challenges for the collection of representative samples. Each calls for different sampling strategies, techniques, and QC measures to assess and improve sampling representativeness. ISM recognizes these various scales of heterogeneity and conscientiously attempts to control their effects. The following sections further discuss the differing scales of heterogeneity, and Section 2.4 provides approaches for collecting representative samples in the face of these heterogeneities.

2.2.1 Microscale Heterogeneity

Within-sample matrix heterogeneity is due to the particulate nature of soil. Grain size variation within each sample is a major contributor to microheterogeneity affecting concentration measurements. As shown in Figure 2-2, soil particle sizes span several orders of magnitude, from less than 0.002 mm for fine clay-sized particles to 1–2 mm for very coarse sands (USDA 2010). Commonly, the maximum grain size considered to still qualify as part of soil is 2 mm. This section devotes a great deal of attention to grain size and its effect on concentration heterogeneity because (a) unless specific field and laboratory sampling and subsampling procedures are followed, routine sampling can lead to concentration estimates that are biased high or low due to grain size effects and (b) ISM sampling guidance detailed in later sections offers techniques to reduce the error due to grain size effects.

2.2.1.1 The smallest particles often have the highest contaminant concentrations

Clay-sized particles are of particular note because of their tiny size and mineral makeup. These particles play a large role in how contaminants interact with soil. Due to their small size, clays have a large surface area per unit mass to which all types of molecules, organic and inorganic, can adhere. The chemical makeup of clay minerals gives them a strong negative charge as well as a weaker positive charge, enabling adsorption of both positive and negative ions. Clay minerals take the form of thin sheets or plates,
as depicted in Figure 2-3. This plate structure greatly adds to the surface area of these tiny particles. The layered plates of clay particles also provide spaces for contaminants to absorb into clay particles, as illustrated in Figure 2-4, where the partial negative charges carried by the oxygen atoms on a dibenzo-p-dioxin molecule are attracted to a cation (such as Ca^{+2}) nestled between two clay plates.
The propensity for smaller particles to attract contaminants was dramatically shown in a study of lead-contaminated soil, as presented in Table 2-1. The pattern is clear: the smaller the particle size (i.e., the larger the mesh size number), the larger the concentration of lead associated with that particle size. Although the smallest particle size, that less than 200-mesh (0.074 mm), made up only one-third of the whole sample mass, it carried nearly three-fourths of the lead mass in the sample. Many experimental studies have documented the finding that for most contaminants, soil fractions composed of smaller particle sizes have higher loadings than fractions composed of larger particles. Important exceptions include metal fragments at firing ranges, explosive/propellant fragments, and ore particles at some mining sites (Walsh et al. 2007; Pavlowsky, Owens, and Martin 2009). Compounds bound to soil particles seldom migrate independently of the particle; they migrate with the particles. The very small particles carrying most of the contaminant load in soil are able to migrate with the wind or be carried by water flow in streams and storm water runoff.

Table 2-1. Relationship between particle size and lead concentration for a firing range site
(Adapted from ITRC 2003)

<table>
<thead>
<tr>
<th>Soil grain size (standard sieve mesh size)</th>
<th>Particle size (mm)</th>
<th>Soil fractionization (%)</th>
<th>Lead concentration in fraction by AA (mg/kg)</th>
<th>Lead distribution (% of total lead)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greater than 3/8 (0.375) inch</td>
<td>&gt;9.53</td>
<td>18.85</td>
<td>10</td>
<td>0.20</td>
</tr>
<tr>
<td>Between 3/8 inch and 4-mesh</td>
<td>9.53–4.76</td>
<td>4.53</td>
<td>50</td>
<td>0.24</td>
</tr>
<tr>
<td>Between 4- and 10-mesh</td>
<td>4.76–2.00</td>
<td>3.65</td>
<td>108</td>
<td>0.43</td>
</tr>
<tr>
<td>Between 10- and 50-mesh</td>
<td>2.00–0.297</td>
<td>11.25</td>
<td>165</td>
<td>2.00</td>
</tr>
<tr>
<td>Between 50- and 200-mesh</td>
<td>0.297–0.074</td>
<td>27.8</td>
<td>836</td>
<td>25.06</td>
</tr>
<tr>
<td>Less than 200-mesh</td>
<td>&lt;0.074</td>
<td>33.92</td>
<td>1970</td>
<td>72.07</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>100</td>
<td>927 (wt. averaged)</td>
<td>100</td>
</tr>
</tbody>
</table>
2.2.1.2 Why laboratory duplicates often fail to match

The implication of this microscale heterogeneity is that the concentration of any soil sample analysis depends on the ratio between the small and large particles in the analytical subsample. Unless measures are taken to prevent it, two analytical subsamples from the same sample (e.g., laboratory duplicates) will vary in their proportions of larger particles carrying lower contaminant loadings vs. smaller particles with higher loadings, causing different subsamples to have different concentration results. The situation is exacerbated by what happens to soil samples during collection, shipment to the laboratory, laboratory handling, and subsampling. All these activities promote segregation or stratification of soil samples by particle size and density, making it likely that subsamples are biased for or against certain particles, biasing the concentration results away from the true mean for the sample.

2.2.2 Short-Scale Spatial Heterogeneity

“Short-scale heterogeneity” refers to differences in contaminant concentrations between colocated samples separated by short distances. These distances can be on the order of inches to a few feet. Some causes of short-scale heterogeneity are discussed in Hyperlink 6. Short-scale heterogeneity determines whether the same result is obtained if one happens to take the sample at placement A or placement B, which is close to A. Placement points A and B are equivalent in the sense that the probabilities of choosing one over the other are equal for a given sampling location. Consider the case where the sampling location is designated as the center of a 100 ft\(^2\) grid cell. Suppose that samples are collected with a 2-inch-diameter coring device. Within a 1 ft\(^2\) area at the center of the grid cell, there are 36 nonoverlapping placement points for a 2-inch corer, any of which might be sampled, as shown in Figure 2-5.

Colocated samples are considered equivalent in that roughly the same concentration would be expected from both placements because they are so close spatially. However, colocated samples often do not meet precision expectations.

High variability in colocated samples is illustrated in Figure 2-6, which presents data from a field investigation for uranium. The original sampling design called for one discrete sample per 270 ft\(^2\). In other words, the result from a single sample would be extrapolated to represent the concentration for an area centered on the sample and encompassing 270 ft\(^2\). Prior to the main investigation, a small pilot study was done to see how much short-scale heterogeneity was present at the 1-foot scale. Figure 2-6 displays the results from the pilot study. It is apparent that the concentration assigned to this 270 ft\(^2\) area could vary by an order of magnitude depending on where the technician happened to place the corer. If the sampler collected from sample...
placement #1, the concentration would be 30 mg/kg; if from placement #2 about 8 inches away, the result would be 496 mg/kg. In other words, if a single discrete sample were used to represent the sampling area, a conclusion of “contamination is low” vs. “contamination is high” is purely a matter of chance.

Results from a similar study involving arsenic variability along a transect covering just a few feet are presented in Hyperlink 7. The takeaway point is that one should not place a great deal of confidence in a single discrete sample result when little is known about the magnitude of short-scale heterogeneity.

2.2.3 Large-Scale Spatial Heterogeneity

The highest level of matrix heterogeneity is large-scale heterogeneity. Its spatial scale is usually on the order of tens of meters and larger, and it is the type of heterogeneity that practitioners expect. This is the heterogeneity caused by common release and transport mechanisms, such as spills, dumping of contaminated soil or sediment, atmospheric deposition downwind of a source, or contamination carried downstream via overland flow or a stream. Large-scale heterogeneity is reflected in the difference between soil areas that are, for example, highly contaminated, moderately contaminated, lightly contaminated, and not contaminated. This is the spatial scale often assumed for discrete samples in traditional sampling designs looking for contamination, estimating the volume of contaminated media, and delineating areas for cleanup.

2.3 Foundational Concepts of Sampling

As previously mentioned, the fundamental purpose of sampling is to obtain data that will support decision making about an area or volume of material that is impractical or impossible to analyze in its entirety. For example, consider a volume of soil that has been defined as an exposure area for risk assessment, such as the top 2 inches in a residential yard. A decision is to be made about this volume of soil based on the mean contaminant concentration. The ideal way to estimate the mean concentration would be to collect and analyze the top 2 inches of soil from the entire yard. This method would provide an excellent estimate of the mean contaminant concentration in the yard but clearly this is impractical. Therefore, samples must be collected and conclusions drawn about the yard from the results of those samples.

2.3.1 All Concentrations Are Means

At the most basic level, an analytical result represents the overall mean of all the thousands of individual particles in the 1, 10, or 30 g analytical subsample. As explained earlier, different particles carry different amounts of contaminants. By means of the analytical digestion or extraction process, there is a physical averaging of the various concentrations of contaminant particles within an analytical subsample into a well-mixed liquid extract, as depicted in Figure 2-7.
The laboratory result provides an estimate of the mean concentration of those particles making up the analytical subsample. Note that, as shown in Hyperlink 8, the laboratory measures contaminant mass and then derives a concentration. Laboratory results are then extrapolated to represent larger and larger volumes of soil culminating with the volume of the DU.

**Figure 2-7. Process of extrapolating analytical sample results to soil concentrations.**

As illustrated in Figure 2-7, the first step of the extrapolation series assumes that the result of the analytical subsample (the mean of the particles in the subsample) is representative of the mean of all the particles in the original field subsample jar. If this assumption is correct, then the results of laboratory duplicates (i.e., two samples taken from the same jar) should agree. If they do not, and commonly they do not, it is an indication that microscale heterogeneity is at work, causing within-sample data variability at the level of the sample jar.

The second step of the extrapolation series assumes that the jarred sample concentration is representative of the mean concentration of the discrete field sample taken from the DU. A prepared field sample is depicted by the pan in Figure 2-7. Note that if the field sample is small, the jarred sample may be the same as the field sample. If the assumptions of both the first and second steps are correct, then colocated samples collected from approximately the same field location (i.e., two “identical” jarred samples) should agree. When they do not, also quite common, the culprit is short-scale heterogeneity.

Finally, following the pattern above but scaled up, the assumption is that the concentration measured in the analytical subsample provides a precise and unbiased estimate of the true mean
concentration for some volume of soil surrounding the location where the sample was taken. This volume of soil is seldom overtly specified but is implied by the way that data are collected and used.

In contrast, ISM targets a volume of soil (a.k.a., the DU) that is deliberately identified up front during systematic planning. Increments of soil are collected at a high density across the entire DU and combined together. In this way there can be more confidence in the assumption that the field sample represents the DU. Steps are then taken during sample processing and subsampling to ensure that the aliquot of soil analyzed by the lab represents the field sample and thus the DU. Hyperlink 9 provides an example of how an ISM approach can better represent a DU than discrete sampling.

2.3.2 Representative Soil Samples

The best laboratory cannot produce good data if the sample is not representative of the soil being assessed or of the intended decision (e.g., assessment, exposure, or remedial decisions). A representative sample is one that contains a subset of all the contaminants of a population in exactly the same proportion as they are present in the target population. In other words, the contaminant concentration in a representative sample provides an accurate and precise estimate the true contaminant concentration in the target population. The population is the “whole” from which samples are taken to measure properties of interest. Hyperlink 10 provides further discussion of the concept of “representativeness” as it has been discussed in existing USEPA and ASTM International (ASTM) guidance.

For most soil sampling scenarios, a single sample or even several discrete samples do not well represent the population of interest because soil populations are too heterogeneous. As discussed in Hyperlink 11, even testing a lawn for nutrient status requires more than one sample. If using discrete samples, a set of them is needed to capture the diversity of the population so that a mean can be estimated mathematically for the population. This is not the case for incremental samples because the sample is composed of increments from across the entire population. A well-designed incremental sampling plan can yield a single sample for analysis that has physically captured the population diversity such that it is representative of the mean of the target population.

If a sample or set of samples intended to represent the population does not properly do so, a “sampling error” is said to have occurred. This is why systematic planning must be done before developing the sampling design. Otherwise, it is impossible to know what a sample is supposed to represent and how to collect it so that it is “representative.” Unfortunately, it is common for sampling designs to be developed without a clear picture of how the data will be used. Inadequate sampling designs commonly indicate that “representative samples” will be collected, but often there is no indication what the samples are supposed to be representing. On the other hand, a statement such as the following provides an unambiguous statement about the population of interest: “Samples will provide estimates of the
true mean concentration of arsenic within the <2 mm soil fraction of the upper 6 inches of soil for each residential lot.”

The most representative soil sample is one that captures the characteristic(s) of interest for the targeted population with the least amount of error. Procedures must be in place to manage the various types of heterogeneity and the errors they cause. Interestingly, USEPA’s *Applicability of Superfund Data Categories to the Removal Program* (USEPA 2006a) emphasizes that documenting total measurement error, which includes sampling errors, is a feature of definitive data. For data to be definitive, either analytical or total measurement error must be determined. Traditional QA/QC programs ignore sampling error in favor of analytical error only. But, as discussed previously, analytical error is often only a small fraction of the total measurement error. Obtaining a representative sample is the first requirement, and determining sampling error is a quantitative measure of representativeness. No data can be truly definitive without knowing that the sample was selected, collected, and processed properly.

### 2.4 Scale-Specific Sampling Considerations

Soil-contaminant interactions contribute to concentration heterogeneity and data variability, which operate at progressively larger spatial scales.

#### 2.4.1 Sampling Considerations—Microscale Heterogeneity

Heterogeneity at various scales can lead to large variability in data sets from areas that have traditionally been expected to be fairly uniform. Heterogeneities at very small, apparently inconsequential, spatial scales can create the impression that large hot spots are present when discrete sampling is used. However, it is just as likely that heterogeneity can cause true hot spots to be missed, even though a sample was taken from within the boundaries of a hot spot. Taking a sample from within a hot spot is no guarantee that the few grams actually analyzed will reflect the hot spot’s true average concentration. Both micro- and short-scale heterogeneity complicate detecting and delineating hot spots. See Section 3.5 for a further discussion of hot spots.

#### 2.4.1.1 Sampling error as a consequence of particle size and sample handling

Decision errors can occur because very small amounts of soil (sometimes as little as 0.25–0.5 g) are actually analyzed from the jar that is sent to the laboratory. Differential contaminant loading of small vs. large soil particles has already been discussed, and further examples are provided in Hyperlink 12. The effect on laboratory subsampling shows up as data variability in the sampling results.

Microscale heterogeneity exerts its effects as soon as soil is placed into a container. The settling of soil that occurs during container movement and sample shipment is governed by particle size and density. Settling stratifies a soil sample such that the larger particles usually end up at the top of the jar as smaller particles work their way to the bottom. If the subsampling procedure involves simply opening the jar and scooping from the top, very few small particles will end up in the analytical subsample, which may bias the concentrations low. On the other hand, the type of scoop used to take the subsample may discriminate against larger particle sizes if the surface is flat or very small so that larger particles can roll off. The very process of weighing out the
analytical subsample can select for small particles if they are preferentially tapped onto the balance to slowly bring the subsample up to the desired weight. Laboratories seldom have standard operating procedures (SOPs) for obtaining a representative analytical subsample. Each laboratory, and each technician in the same laboratory, is likely to handle samples somewhat differently. As a result, the analytical subsample may not be representative of the bulk average in the sample container but may over- or underrepresent certain particle sizes from one subsample to the next (Gerlach et al. 2002).

Unfortunately, typical sampling and analysis procedures make little or no effort to control for particle and microscopic effects. In fact, common mixing techniques, such as cone-and-quartering, can even exacerbate the problem (Gerlach et al. 2002). Therefore, it is not surprising that analyses of subsamples repeatedly taken from a single jar of soil can have widely varying results, as reflected in high relative percent difference (RPD) between field or laboratory duplicates. Most of this difference is not due to analytical issues, as is commonly assumed, but is primarily caused by heterogeneity between replicate subsamples.

In summary, though soil sampling seems like a simple process, it is actually quite complex and subject to many kinds of errors. For example, errors occur when the ratio of large to small particles in the subsample do not match the ratio present in the sample container. Taking a representative sample from a heterogeneous bulk particulate material like soil requires careful planning at each stage of sample collection and analysis. Planning to avoid errors requires an understanding of all types of heterogeneity and the spatial scales at which they occur.

2.4.1.2 Measuring the error caused by within-sample heterogeneity

The amount of error caused by within-sample heterogeneity can be measured using replicate subsamples in the field and/or in the laboratory. When each of the subsamples is analyzed, the difference between their respective results is calculated as indicated in Figure 2-8.

A large difference between results indicates that within-sample heterogeneity is present and is causing sampling error. Field splits and laboratory duplicates for soils are common QC checks that often fail to meet QC acceptance criteria. Unfortunately, nothing is typically done to correct the problem(s) indicated by the failed QC. The data may be qualified as estimated, but in practice they are simply used “as is.” Laboratory duplicate results should not be ignored, for they provide very important information about the quality of sample handling and the magnitude of sampling error.
Duplicate results may vary so widely that a different decision about “clean” or “dirty” may be indicated, depending on which result is used. The question is often asked, “Which result is right?” The answer is that they are probably both right and both wrong. Both are right in the sense that the analysis of both subsamples was probably correct unless other QC samples indicate otherwise. It is just that the laboratory subsamples are fundamentally different. Both may be wrong in the sense that neither result adequately represents the true concentration for the jar of soil, and by extrapolation, for the concentration in the DU. Highly variable field and/or laboratory duplicates should be an indication to decision makers that the data generation process is excessively imprecise and could lead to decision errors. Hyperlink 13 provides a discussion of approaches for dealing with within-sample heterogeneity.

2.4.1.3 The effect of subsample mass on data variability

Figure 2-9 presents the result of a study performed by the U.S. Department of Energy (DOE) in the mid-1970s. A large (~4 kg) soil field sample was milled to <10-mesh (2 mm) particle size. Twenty replicate aliquots of various masses were taken from the prepared sample and analyzed. Despite the homogenization efforts, the <10-mesh particle size allowed particle size effects and heterogeneity to persist. The concentration units in this figure are in nanoCi/gram (nCi/g), and the vertical line at 2 nCi/g approximately represents the true concentration for the 4 kg sample. This experiment demonstrated the relationship between analytical sample mass, data variability, and potential decision errors. The results show that data from subsamples of smaller mass, such as the ≤1 g mass commonly used for metals analysis, show more data variability than analytical subsamples of larger mass.

The data variability caused by heterogeneity affects the statistical distribution of the data, as seen in the three curves in the diagram. Data from smaller subsample masses form more lognormal-like statistical distributions. For example, notice how the right side of the 1 g sample mass curve (blue curve) is pulled out or “skewed” to the right much more than the left side. Because it is easy for small subsamples to miss contaminated particles, many small subsamples have low
Figure 2-9. Smaller analytical masses contribute to high data variability. *Source:* Data from an experimental study on radioactively contaminated soil (Gilbert and Doctor 1985).

In summary, Figure 2-9 illustrates clearly how matrix heterogeneity and particle size effects manifest as data variability and nonnormal, skewed statistical data distributions. These effects increase the possibility of decision errors.

### 2.5 Gy Theory and the Source of Sampling Error

Much has been written about sampling over the years; however, the sampling theory of Pierre Gy may be the most comprehensive and mathematically developed. Pierre Gy formed his theory for the sampling of particulate materials beginning in the 1950s, originally focused on the mining and mineral exploration industries, culminating in his final theory in the late 1980s and early 1990s (Gy 1998). Pitard (1993) summarized Gy’s sampling theory for the English-speaking audience and extended it to environmental applications. While all of Gy theory applies to environmental sampling, this section focuses on minimizing sampling errors during extraction of samples from a parent matrix.

#### 2.5.1 Pierre Gy’s Sampling Theory and the Seven Basic Sampling Errors

As discussed in Sections 2.2 and 2.3, the purpose of sampling is to obtain data of sufficient quality to support decisions that will be made about volumes of soil. To support these decisions, the sample(s) collected should represent the volume of soil, i.e., have the same types and distribution of particles, area, or volume of interest.

To achieve a representative sample, potential errors that result from collecting small volumes of material meant to represent a much larger volume need to be addressed in the design of the sampling plan. This requires an understanding of all the potential sampling errors that can bias the result. Pierre Gy describes seven basic sampling errors associated with collecting samples from particulate materials such as soil. The following sections introduce each error.

#### 2.5.2 Compositional Heterogeneity

Before sampling errors can be discussed, the Gy theory concepts of constitutional and distributional heterogeneity must be introduced. Constitutional (or compositional) heterogeneity is a measure of the differences in composition between individual fragments or particles of the population being sampled with respect to a given parameter of interest. It refers to the fact that soil is made of many different types of particles that interact with contaminants in different ways. CH is a direct cause of a sampling error termed “fundamental

**Gy’s Seven Sampling Errors**

1. Fundamental
2. Grouping and segregation
3. Long-range heterogeneity
4. Periodic heterogeneity
5. Increment delimitation
6. Increment extraction
7. Sample preparation

---

**Fundamental error** is controlled by collecting samples of sufficient mass.
error” (FE). A way to control FE is to have large enough samples (or subsamples) so that the probability is high that the composition of the sample will match the composition of the population. Figure 2-10 presents a population with two samples of different masses. Although both samples were collected from the same population, they are not equally representative of the parent population. The larger of the two samples (Sample B) better represents the composition of the population and reduces FE relative to the smaller sample (Sample A). Also, the larger the particles, the larger the sample mass must be to minimize FE. More illustrations of this concept can be found in Hyperlink 12.

2.5.3 Distributional Heterogeneity

Gy’s distributional heterogeneity (DH) is a measure of particle distributions that can take the form of grouping or segregation. Particles may segregate, that is, separate into layers. Segregation is often a result of gravity. The most common example is jiggling a jar of dry soil, causing finer particles to migrate toward the bottom, while the larger particles end up at the top of the soil mass. Figure 2-11 illustrates the two types of DH that Gy described.

These distributional heterogeneities cause grouping and segregation error (GSE). GSE can occur at all spatial scales: within a sample (e.g., within a jar of soil) or within a field population. Note that a jar of soil is both a sample and a population. It is a sample of the field population, but it itself becomes a population when it arrives in the laboratory. That jar is the population from which a representative analytical subsample needs to be taken. If segregation has occurred (e.g., fines at the bottom and the coarser particles at the top of the soil sample jar) a sampling error is committed if the analytical subsample is taken by scooping off the top. This all-too-common sampling error easily leads to decision errors that a site is “clean” when it actually may not be.

It might appear that just mixing the sample solves the problem. Unfortunately, for soil samples, common forms of mixing such as cone and quartering methods can be ineffective and may actually increase GSE (Gerlach and Nocerino 2003). Likewise, attempting to “mix” the parent matrix, such as with a backhoe, is ineffective. A good way to reduce the effects of DH is an incremental sampling approach, where enough increments are collected so that the resulting recombined large-volume sample contains the particle ratios present in the volume of matrix that was sampled.
2.5.4 Long-Range and Periodic Heterogeneities

In the Gy paradigm, “long-range heterogeneity” refers to the same contamination pattern as the term “large-scale heterogeneity,” as discussed in Section 2.4.3. This heterogeneity involves the nonrandom, nonperiodic distribution of contaminant across the site. Identifying this heterogeneity is often an objective of sampling programs, such as mapping site-specific concentration trends. The question is: what is the volume over which knowledge of this heterogeneity is desired vs. what is the volume over which such heterogeneity is a distraction because the mean is the parameter of interest? In Gy’s theory, this heterogeneity is considered the cause of long-range heterogeneity fluctuation error (CE$_2$). This Gy-defined error may or may not be a relevant error for a sampling design, depending on whether knowledge of contaminant distribution or mean is desired and the spatial dimensions of both have been defined. Gy theory assumes that the parameter of interest is the mean, not contaminant distribution.

In the same way, periodic heterogeneity and its corresponding periodic heterogeneity fluctuation error (CE$_3$) is the result of cyclical changes in space or time over a site. An example of a cyclical change in time is measuring nitrogen concentration in agricultural fields over several growing seasons. If sampling were always performed at the start of the growing season when nitrogen levels were highest, a misleadingly high value would be obtained if the average over the entire year were desired. Just as for the long-range error above, it is only a true error if it causes an inaccurate estimate of a mean for some defined area, and in this case, for a defined period of time.

The heterogeneities discussed above can lead to additional sampling errors. Four of Gy’s seven sampling errors are described above; the last three are covered below.

2.5.5 Device and Preparation Errors

Delimitation error (DE) is a result of using an incorrect shape for the sampling device that removes each increment from the population or the incorrect use of a correct sampling device. For example, an incorrectly shaped sampling tool biases the grain sizes included in that sample. A sampling tool should be of a shape and size so that every fragment of the population of interest has an equal probability of being included in the sample. This error is a common source of bias in environmental samples, both in the field and in the laboratory. Figure 2-12 illustrates that, depending on the sample device, some particles have a greater chance of being included in the sample/subsample than others. The sampling interval depicted in Figure 2-12 has a higher proportion of larger particles at the bottom of the interval. This might be the case, for example, in an in situ soil scenario. On the other hand, this particle distribution pattern might be reversed, for example, in the case of soil jars in the laboratory. Subsampling with the rounded scoop preferentially gathers particles from the top, which tend to be the larger particles when stratification occurs as the sample is arranged in a “slabcake” shape in preparation for subsampling. With its narrower bottom width, a rounded scoop discriminates against the particles at the bottom of the sampling interval.

![Figure 2-12. Illustration of the effects of sampling device design on particle sizes in a sample.](image-url)
interval, which tends to be the smaller sizes in many if not all subsampling scenarios. By design, the rectangular scoop tool is a more inclusive tool and gathers particles of various grain sizes consistently throughout the sampling interval. In Gy theory, a sampling tool that promotes DE is termed incorrect”; one that reduces DE is called “correct.”

Extraction error (EE) also results from the use of incorrect sampling devices. Unlike DE, which is only a function of the shape of the sampling device, EE is a function of the sizes of both the tool and the soil particles and the correct use of the sampling device. This error occurs because an inappropriate sampling device can bias the fragments that are included or excluded from being captured by the device. This scenario often plays out when the sampling device is too small and the cutting edge of the tool pushes all or certain particles (e.g., larger sized particles) aside rather than including them in the sample. EE is also a common source of bias in both the field and in the laboratory. An EE that commonly occurs in the field is when full recovery of the core is not attained when using a split-spoon or direct push-sampler. Figure 2-13 shows a sampling device that gives all particles an equal chance of being included in the sample, depending on where the center of gravity lies with respect to the cutting edge of the device. As the sampling device is used, particles are included or excluded with equal probability, thus reducing EE. To reduce EE, a correct sampling device should have a “mouth” size at least three times the size of the largest particle (Gerlach and Nocerino 2003).

Figure 2-14 illustrates both DD and EE. A volume of soil is depicted in a two-dimensional (2-D) plane with larger particles concentrated at the bottom. Coring Device A minimizes the DD and EE because it samples the full thickness of the material and does not discriminate against the larger particle sizes. Coring Device B demonstrates EE and is an incorrect device for this matrix. Its mouth is too small to include larger particles. Coring Device C (the shovel) demonstrates DE because the sample profile it delimits cannot sample the full thickness of the DU. Coring Device D also illustrates DE because the delimited sample profile does not encompass the full shape of the DU in the vertical plane.
Preparation error (PE) is the sum of errors introduced by analyte loss, cross-contamination, or chemical or physical alteration of the sample that biases sample results relative to the true mean. Some of these errors are controlled by traditional QA/QC procedures such as sample preservation, holding times, and blanks.

### 2.5.6 Controlling Gy Errors

To correctly collect samples as defined by Pitard (1993), all these errors should be addressed. Table 2-2 provides a summary of the various errors described by Gy together with measures that might be taken to control each.

#### Table 2-2. Summary of sampling errors described by Gy and control measures
(These apply to both field sampling and subsequent subsampling.)

<table>
<thead>
<tr>
<th>Factor leading to error</th>
<th>Sampling error</th>
<th>Error results from</th>
<th>How to control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compositional heterogeneity (CH)</td>
<td>Fundamental error (FE)</td>
<td>Size and compositional distribution of the particles</td>
<td>Increase the sample mass and/or reduce the size of the particles</td>
</tr>
<tr>
<td>Distributional heterogeneity (DH)</td>
<td>Grouping and segregation error (GSE)</td>
<td>Heterogeneous distribution of particles within the population</td>
<td>Increase the mass of the sample or increase the number of increments</td>
</tr>
<tr>
<td>Large-scale heterogeneity</td>
<td>Long-range heterogeneity fluctuation error (CE&lt;sub&gt;2&lt;/sub&gt;)</td>
<td>Changes in concentration across space or over time</td>
<td>Reduce the spatial interval between samples</td>
</tr>
<tr>
<td>Periodic heterogeneity</td>
<td>Periodic heterogeneity fluctuation error (CE&lt;sub&gt;3&lt;/sub&gt;)</td>
<td>Periodic changes in concentration over time</td>
<td>Change the spatial and/or temporal interval between samples</td>
</tr>
<tr>
<td>Identifying the correct increment geometry</td>
<td>Increment delimitation error (DE)</td>
<td>Incorrect shape (in all three dimensions) of the sample or increment selected for extraction from the population</td>
<td>Use correct sampling plan design and correct sampling equipment that can sample the entire thickness of the population</td>
</tr>
<tr>
<td>Shape of the sample extraction device and nature of the soil</td>
<td>Increment extraction error (EE)</td>
<td>Incorrect extraction of the sample or increment because the sampling device is too small</td>
<td>Use correct sampling equipment that does not push larger particles aside, and use correct sampling protocols</td>
</tr>
<tr>
<td>Loss or gain of contaminants during sample handling</td>
<td>Preparation error (PE)</td>
<td>Contamination loss or gain due to alteration, evaporation, degradation, cross-contamination, mistake, or fraud</td>
<td>Use appropriate sample handling, preservation, transport, and preparation measures</td>
</tr>
</tbody>
</table>
In practice, the focus is usually on FE and GSE; however, the other errors can be important if correct sampling procedures are not used. As illustrated above, the FE can be minimized by collecting sufficient mass of sample, and the GSE can be minimized by collecting numerous increments.

The mass of a sample necessary to minimize FE is primarily related to the largest particle size of the population being sampled. Hyperlink 14 provides more information concerning the calculations for determining sample mass to minimize the FE.

The number of increments needed in the field depends on a number of factors, including heterogeneity within the DU, the difference between the mean concentration and the level of interest (e.g., action level), and project DQOs. It is theoretically possible to determine the number of increments necessary. For example, at a large site or a site with many DUs, a pilot study could be conducted on a portion of the site to provide initial estimates of heterogeneity and mean concentration. That data could then be used to determine the number of increments needed to manage decision error sufficiently. However, this process is often not practical due to cost and time constraints. Then, how many increments are sufficient?

One approach is to use a sufficiently conservative default number of increments—one that is high enough to result in a representative sample for the majority of cases even when the DU is heterogeneous. Based on simulation studies discussed in Section 4 and empirical evidence gathered from using ISM at a variety of sites, a default range of 30–50 increments is adequate for most sites. However, as many as 100 increments may be necessary for larger DUs where the CSM indicates that high heterogeneity is anticipated. One indication of how well the increment density is capturing the heterogeneity within the DU is variation between ISM replicates. If all other sources of error are held constant, the degree to which the number of increments collected are capable of capturing the heterogeneity present in the DU is reflected in how well replicate ISM sample results agree. One should use caution, however, when interpreting results between ISM replicates since this measure of variability integrates all of the sampling errors described above.

### 2.6 Three Sampling Approaches

The total error associated with an estimate may be considered in the following simple equation that relates the true but unknown value of the parameter of interest (in this case the mean concentration) to the estimate of that parameter:

\[
\text{true mean concentration} = \text{estimate of that concentration} \pm \text{total error}
\]

This equation emphasizes several important concepts:

- There is a true mean concentration in any volume of soil.
- Any type of sampling and analysis is capable of providing only estimates of the true mean concentration.
- The best estimate is the one with the least total error.
These concepts provide a basis from which to compare different methods for estimating the mean through different sampling approaches: discrete, composite, and ISM sampling. Any sampling design must include consideration of these questions:

- Which parameter of the population is being estimated (e.g., mean, maximum, proportion)?
- To what soil volume does that estimate apply?
- How will total error be controlled, measured, and assessed?

Although these questions should be addressed at the beginning of any sampling effort, they commonly are not. One of the strengths of the ISM process is that it necessitates a thoughtful consideration of these topics as well as an assessment of the strengths and weaknesses of the various sampling approaches prior to sampling.

The characteristics of these three sampling approaches relevant to providing an estimate of the mean are discussed in the following sections.

2.6.1 Discrete Sampling

Discrete, or grab, soil sampling has a long history of use within the environmental industry. Analytical results from a number of discrete samples collected from a site are typically used to make environmental decisions regarding the site. For example, they may be used to provide an estimate of the mean in some meaningfully sized volume of soil.

A number of factors influence the ability of such a discrete sampling plan to provide an unbiased estimate of mean concentration. The primary factor is the number of discrete samples collected, but sample location, collection method, sample support, and lab handling are also important. For discussion purposes, two types of discrete sampling plans of different sample numbers are identified below: high and low density. Of course, this is a gross oversimplification, but it is useful here for the purpose of highlighting some important concepts.

At face value, low numbers of discrete samples are tempting in terms of cost, ease of implementation, and simplicity. However, simulation studies, empirical evidence, and sampling theory suggest that low numbers of discrete samples do not produce very accurate or precise estimates of the mean because such an approach does not account for heterogeneity. When only costs are considered, discrete sampling plans have historically been preferred. However, comparisons between different sampling approaches must be evaluated not only in terms of their costs but also in terms of the total error and resulting decision quality. The two types of discrete sampling plans discussed below result in dramatically different costs, but they also result in dramatically different decision qualities. Collecting the number of discrete samples sufficient to make a defensible decision at a site may at times be precluded by cost considerations.

2.6.1.1 Heterogeneity and discrete samples

Small- and Micro-Scale Heterogeneity. Discrete samples typically consist of about 200–300 g of soil, of which perhaps 1–30 g is processed and analyzed in the laboratory. Therefore, a discrete
sample contains between about 7 and 300 possible analytical subsamples, only one of which is actually analyzed. The assumption is that every subsample taken from the discrete sample will result in the same concentration estimate if analyzed. As discussed in Section 2.4.1.1, this is often a poor assumption.

**Large-Scale Heterogeneity.** As discussed in Section 2.3.3, variation in contaminant is expected at relatively large scales (i.e., on the order of residential yards and larger). This is the scale at which concentration trends, hot spots, and clean volumes of media are often of most interest. However, when low numbers of discrete samples are used and microscale and short-scale heterogeneity are present, data from discrete samples can miss the presence of large-scale contaminant trends. In other cases, they can misidentify the effects of microscale and short-scale heterogeneity as contaminant trends or hot spots that are not actually present. When a single discrete sample is found to have a concentration that is “hot,” it may mean that some meaningfully sized volume of contaminated soil is actually present at a site. But it may just as easily simply reflect the reality that a few “hot” samples are to be expected when collecting discrete samples from heterogeneous particulate materials like soil. Without additional corroborating evidence or additional discrete samples, these two situations are indistinguishable.

### 2.6.1.2 Discrete sampling plans

**Relatively Low Density.** Often only a few discrete samples are collected, and the results are used to make decisions about relatively large volumes of soil. In these situations, the number of samples collected may be determined by negotiation, budget, professional judgment, convention, or happenstance. The number of samples is often not based on statistical or other scientific rationale, and the location of the samples is often judgmental. Judgmental sampling plans can be used effectively with low numbers of discrete samples if the basis for determining the sample location and the volume of soil it applies to is appropriate. For instance, judgmental sampling plans may be useful when obvious source areas of high concentrations are present.

While low-density discrete sampling plans are tempting in terms of familiarity, relative low cost of collection and analysis, ease of implementation, and simplicity, the performance of these approaches generally is not adequately tested in terms of precision, accuracy, and decision error. However, there is a large body of work in classical statistics, Gy sampling theory, industry experience, and empirical evidence (e.g., results from duplicate samples) which suggests that (a) soil is highly heterogeneous even on extremely small scales and (b) smalls numbers of discrete samples are not likely to provide accurate or precise estimates of mean concentrations. Low-density discrete sampling plans therefore cannot be relied on to consistently produce high-quality decisions.

This is not to say that a low-density discrete sampling approach is insufficient for all cases. If, for example, the true mean in a DU is orders of magnitude above or below the action level for a contaminant of interest, it is possible that a correct decision could be made from very few (or even one) discrete samples. The key factors are the degree of heterogeneity present at the various scales, the action level, and the magnitude of the true mean. Since, as is often the case, knowledge about heterogeneity or the magnitude of the true mean is seldom available (which is why sampling is being conducted), relying on data from low-density discrete sampling plans is more likely to result in decision errors.
Relatively High Density. The second type of discrete sampling plan can be called relatively high density discrete sampling plans. In this context the number of discrete samples approaches the number of increments typically collected with ISM (i.e., 30–50). The number of samples may or may not have been statistically derived based on (among other things) an estimate of the heterogeneity of the soil or the anticipated magnitude of the true mean concentration. There is a large body of guidance and reference material that describes how various discrete sampling plans of this sort can be effectively used to investigate soil contamination and make appropriate environmental decisions. However, cost limitations frequently limit the number of discrete samples employed for environmental investigation, and as discussed below, even relatively high-density discrete sampling plans may produce certain characteristics in the data set which are not ideal. The decision quality of relatively high-density discrete sampling plans, especially those derived through statistical methodology, can compare favorably with ISM sampling plans. However, the analytical costs associated with such plans will likely be considerably greater than those of a comparable ISM approach.

2.6.1.3 Interpreting results of discrete sampling

Action levels are usually derived from risk assessment models that are based on average exposures over time. Use of mean soil concentrations to estimate exposure within a given area of contamination assumes that (a) the estimated mean soil concentration represents the true mean concentration in the exposure area, (b) the receptor is equally likely to be exposed to the soil at any location in the exposure area, and (c) soil concentrations will not change significantly over time. Based on these assumptions, risk assessments and risk management decisions often focus on estimates of the mean soil concentration in each exposure area.

Concentration data obtained from discrete soil samples typically fit frequency distributions that are skewed to the right (i.e., lognormal, gamma, and some nonparametric distributions). Figure 2-15 provides a graphical display of a normal distribution (A) and a right-skewed distribution (B). Notice that the “long tail” extending to the right in Distribution B reflects the higher concentration results that occur at lower frequencies.

![Figure 2-15. Examples of distributions generated by plotting concentration data vs. frequency (i.e., probability) of observation.](image)
Discrete sample data tend to be clustered around the most frequently observed concentration, which is called the “mode.” Because Distribution B in Figure 2-15 is skewed to the right, the mode is less than the mean concentration of the distribution. The tail of such distributions can easily contain concentrations one to two orders of magnitude greater than the value at the mode. In contrast, ISM samples can be expected to fit a distribution closer in shape to Distribution A in Figure 2-15, with less tailing and a mode closer to the mean. This fact can have important implications for making decisions based on discrete sampling data.

Discussion of an idealized spill area scenario is provided in Hyperlink 15 to illustrate the important implications of making decisions based on discrete sampling data for volumes of soil with various levels of contamination.

2.6.1.4 When discrete sampling may be successful

The problems with making decisions about large bulk volumes of soil using discrete sample data have been discussed throughout this section. The particulate nature of soil and its interaction with contaminants, as well as the sheer volume disparity between the amount of soil analyzed and on which decisions are made, means that heterogeneity is the primary factor affecting the sampling error and thereby affecting the quality of environmental decisions.

One is most likely to make correct environmental decisions using discrete sampling in the following circumstances:

- Low-density discrete sampling may be sufficient when the impacts of heterogeneity and sampling error on the decision are expected to be low:
  - previously collected discrete sampling data indicate that the mean (or range) of soil concentrations is well below the action level,
  - previously collected discrete sampling data demonstrate that heterogeneity is very low, or
  - the sampling goal is to obtain qualitative data, for example, when conducting in situ X-ray fluorescence (XRF) soil screening to gain initial estimates of the nature and extent of metal contamination.
- High-density discrete sampling (roughly equivalent to the number of increments collected with ISM) can be useful when sample locations and sampling and subsampling techniques are appropriate for obtaining an unbiased estimate of the mean.
- The volume of soil represented by the discrete sample or samples can be adequately identified. Note that that volume of soil to which discrete samples apply is often determined after the samples are collected and the data apportioned in a variety of different ways, as further discussed in Hyperlink 16.

There are other situations where discrete sampling may be preferred, even though the above conditions are not met, for instance, when (a) discrete sampling is required by regulation, (b) sample collection and/or processing may change the concentration of the sample (e.g., reactive chemicals are investigated), or (c) ISM is cost-prohibitive.
2.6.2 Composite Sampling

A discussion of composite sampling goals and sample collection techniques may seem out of context in a description of ISM principles. However, as is noted in Section 8.2, there is a general misunderstanding that ISM is simply a new term for what many may already be familiar with as composite sampling. Therefore, some background on composite sampling from existing federal guidance is provided here together with potential beneficial and common misuses of this sampling strategy.

A number of guidance documents generated by USEPA and other organizations address the compositing of soil and other environmental media (USEPA 1985, 1986, 1989a, 1995b, 1996a, 1996b, 2002d, 2002e; Gerlach and Nocerino 2003). These documents provide many details for how to use compositing for different project purposes; however, important details on how to collect and process composite samples are generally not discussed in great detail in the existing USEPA guidance. A composite sample is defined by USEPA as a sample created by combining several distinct increments (Gerlach and Nocerino 2003). USEPA guidance frequently acknowledges that sampling error far outweighs analytical error and that soil sample “homogenization” is critical (USEPA 1995b). However, specific guidance for how to achieve relatively even distribution of contaminants throughout the sample via field and laboratory subsampling and processing procedures is not provided. An exception is the RCRA Waste Sampling Technical Guidance (USEPA 2002e). This document explains Gy theory and discusses various applications of composite sampling. USEPA guidance describes several compositing designs, each with a different purpose. One of those purposes is determining the mean over a DU.

2.6.2.1 Beneficial uses of compositing

Several USEPA guidance documents, as referenced above, describe composite sampling designs that can be used for various purposes, such as finding areas of high concentration and estimating a population proportion. Only two simple and beneficial uses of compositing will be discussed in this section and contrasted with ISM. For additional information on composite sampling design options, consult the USEPA documents.

With discrete sampling designs there is often the implicit assumption that each discrete sample represents some volume of soil surrounding the area where it was collected. As shown in Figure 2-6, this may be a faulty assumption when short-scale heterogeneity is significant. Recall that short-scale heterogeneity is what occurs at the scale of colocated samples, where kneeling down in one location can give a radically different concentration than kneeling down in a location one foot away. Composite sampling can be used to reduce errors due to short-scale heterogeneity. For example, in Figure 2-6, instead of collecting a single discrete sample from only one of the five placements, an increment of soil could be taken from each of the five placements and composited into a single sample. This method results in a sample more representative of the 1 ft² area shown in the figure as compared with any single discrete sample. This process could be repeated in other 1 ft² areas, resulting in a number of composite samples. Note that, as with any sampling design, subsequent steps still need to be taken with each sample to address microscale (within sample) heterogeneity to reduce this source of sampling error as discussed in Section 2.4.1. This process is different from an ISM sampling design since it may not include the prior establishment of a specific DU and the goals may not necessarily be limited
to estimating the mean concentration. Also note that the analytical cost of such a composite sample design will typically be much larger than with ISM since many samples will be submitted for separate lab analysis.

A second similar application of compositing is for grid sampling. Instead of taking a single discrete sample from the center of a number of grid cells laid out across a site, a series of composite samples could be taken. The result of a composite sample consisting of several increments of soil collected from across the grid cell is likely to produce a better estimate of the true concentration for that grid cell than will a single discrete result. Note that analytical costs are approximately the same for either the discrete sample/grid center or the composite sample/grid cell design. Again, this approach differs from ISM in that one may have goals in addition to estimating a mean concentration within a predefined volume of soil.

The actual dimensions and number of increments to composite depends, of course, on the spatial scale(s) of the decisions and the degree of short-scale heterogeneity. These can be derived judgmentally or statistically. In either case, it is a good idea to verify that the design is accomplishing its intended goals.

2.6.2.2 Poorly designed composite sampling

Unfortunately, as commonly practiced, composite sampling seldom considers the spectrum of sampling errors or requests that laboratory subsampling be done in a way that addresses microscale heterogeneity. Also, techniques long used to “homogenize” soil samples such as cone-and-quartering have been shown by experiment to be ineffective and are no longer recommended (Gerlach and Nocerino 2003). In summary, composite sampling as conventionally implemented is characterized by unspecified sample collection and analysis procedures that do not adequately consider the following:

- the number of soil increments to be collected
- the intended “area of inference” for the composite samples
- the size and boundaries of the DU
- particle size selection or reduction measures
- bulk sample mass requirements
- field and laboratory subsampling techniques

As routinely applied, composite sampling is viewed primarily as a way to reduce analytical costs, without taking more important sampling goals into account. It is not surprising that over the years composite sampling has developed an unfavorable reputation. It is important to understand that ISM differs greatly from the practices common to poorly designed composite sampling applications. It is worth noting that routine applications of composite sampling also differ significantly from the composite sampling designs recommended in USEPA guidance. Yet, ISM transcends even most USEPA compositing guidance because ISM prominently calls out specific error-controlling steps. These were not yet well researched when most USEPA statistical sampling design documents were written.
However, the primary reason that ISM and composite sampling as typically practiced cannot be considered equivalent is that typical composite sampling rarely involves enough aliquots of soil to manage contaminant heterogeneity over an entire DU. Therefore, even when the goal of compositing is to determine an average over some area, it is less likely to estimate the mean concentration with the precision needed by data users. Empirical studies and sampling theory suggest that composite sampling (with inadequate consideration of the steps listed above) simply does not perform as well as ISM sampling. Indications are that low-increment number composite samples combined with insufficient mixing and processing procedures perform about as well as discrete samples. However, it is acceptable to use the composite sampling approach if it meets the user-defined goals for precision and accuracy. Composite sampling approaches should include a methodology for estimating the total precision and take measures to ensure that an unbiased mean is obtained.

2.6.3 Incremental Sampling Methodology

Although composite samples are not typically considered to be ISM samples, by definition, all ISM samples are considered to be composite samples. It should be noted that a number of organizations, including regulatory agencies, are still in the process of defining what characteristics must be present to be considered an incremental sample vs. a traditional composite sample. However, ISM is a specialized type of composite sampling with specific structure and requirements that stand apart from common compositing practices. ISM is designed to provide more precise and less biased estimates of the mean concentration in soil by addressing specific sampling errors. Consequently, ISM can result in better performance in terms of decision error reduction than other sampling methodologies. The following are primary advantages to the use of ISM sampling approaches:

- requires designation of a targeted population (the DU) prior to sampling
- provides less biased and more precise estimates of the mean than low-density discrete sampling plans
- is more cost-effective than moderate- to high-density discrete sampling plans with a comparable level of decision quality
- tends to produce normal rather than lognormal or nonparametric data distributions
- specifies protocols for laboratory and field procedures to control sampling error

Gy theory is designed to minimize sources of error in the sampling and subsampling of heterogeneous bulk volumes of particulate material. ISM is consistent with the principles of Gy theory and provides a structured sampling protocol intended to reduce the sampling error associated with heterogeneity through the implementation of the following steps:

- collection of a large number of increments
- reduction of particle size reduction
- collection of a large bulk sample mass
- implementation of field and laboratory subsampling techniques

These steps control the FE and the GSE. The long-range and periodic fluctuation heterogeneity errors are controlled through project planning, during which appropriately sized DUs are
identified. The increment DE, the increment EE, and the PE can be controlled through correct sampling and subsampling, aspects of which are discussed in Sections 5 and 6.

ISM sampling produces an estimate of the mean contaminant concentration in soil within a specified volume (i.e., a DU). As with any estimate derived from sampling, ISM results are subject to errors, the components of which were described in Section 2.5. Statistical analysis can provide an understanding of error introduced by sampling. Rigorous statistical analysis regarding the extent to which various ISM sampling strategies provide accurate estimates of the mean contaminant concentration have not yet been published, but Section 4 includes an in-depth discussion of the statistics for ISM, and Appendix A includes relevant simulation studies. This information is necessary to understand how factors such as number of increments, number of replicates, and contaminant distributions across the site influence the reliability of ISM estimates of mean contaminant concentration. The reliability of ISM based on statistical principles is vital to widespread regulatory acceptance of this sampling method.