Nature of Soil Sampling

2.1 Introduction

Contaminants in soil and other particulate media are often distributed unevenly at the scales of interest to decision-makers. Conventional soil sampling approaches failing to address this heterogeneity can result in over- or underestimating contaminant concentrations, leading to decision errors. This section of the guidance document provides the reader with a detailed understanding of contaminant heterogeneity in soil, the consequences of heterogeneity for soil sampling and decision-making, and how ISM is systematically designed to address this heterogeneity. Section 2 further details how ISM addresses variability in sampling data resulting from the inherent heterogeneity in soil.

The proper implementation of ISM practices will improve the reliability of soil data, allowing users to avoid many of the problems stemming from decision errors. ISM practitioners have found that ISM improves the reliability of soil data and avoids many problems stemming from the decision errors caused by misleading data results. Examples of these problems include cost and schedule overruns, later discovery of unnecessary cleanup or overlooked contamination, regulatory and stakeholder confusion, and legal challenges. ISM, when implemented properly, results in highly representative soil data (Hadley, Crapps, and Hewitt 2011), (USACE 2013a, b, c) (USACE 2007). This is because ISM incorporates key concepts from soil science, QA, analytical chemistry, and statistics. The ISM approach interweaves the concepts from these disciplines and aids the practitioner in collecting the most representative data possible for decision-making. To this end, Section 2 will introduce the reader to several concepts:

- soil heterogeneity and the impacts of that heterogeneity on contaminant concentrations in soil
- representativeness and why it is critical to understand and incorporate this concept into planning and decision-making
- the relationship among the mass analyzed, the size of soil particles, and the resulting data variability

2.2 ISM Accommodates the Complexities of Soil Testing

Unlike discrete sampling practices, ISM explicitly addresses soil and sediment as heterogeneous media at all scales. Heterogeneity causes sample concentrations to fluctuate dramatically depending on factors sometimes governed by simple chance:

- Where the sample is collected in the field. A few inches in any direction can have an impact on the sample result, thereby changing the decision.
- Ineffective blending. Stirring of a strongly aggregated soil (such as compacted clumps of clay) cannot mix or homogenize the soil, but rather just moves the clumps of soil around. Cohesive clumps must be mechanically broken apart, which is referred to as *disaggregation*.
- Particle segregation and the cohesiveness of the field sample. Even gentle disturbance of a free-flowing soil sample will cause *segregation*, meaning smaller or denser particles work their way to the bottom, leaving larger particles on top. "[M]any samples cannot be made homogeneous enough for [sub]sampling by mixing… Segregation of particles by [gravity] usually occurs at the moment that the mixing has stopped. Some samples will remain segregated even during the mixing process" (USEPA 2003).
- How the laboratory technician takes the analytical subsample. The common practice of scooping off the top is unlikely to obtain soil with a concentration representing the whole soil sample, especially if the sample is segregated. "When [segregation] happens, sampling techniques such as grab [sub]sampling end up underestimating the concentration, which could result in decision errors" (USEPA 2003).
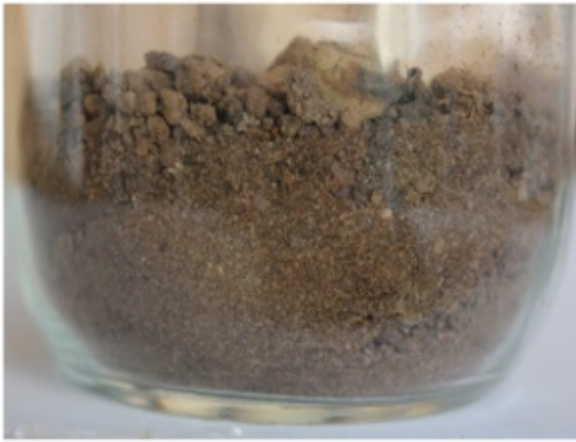
- How much soil mass is actually analyzed. Depending on the laboratory, typical subsamples for metals analysis are in the 0.25- to 2-g range, which is just a few chunks on the displayed spoon. For organic compounds, analytical masses range from less than 1 g to up to 30 g. Experimental studies have documented concentrations.
- Results that indicate they are highly dependent on the mass of soil tested. Smaller analytical masses result in greater variability than larger ones (Doctor and Gilbert 1978) (Pitard 2019).

- soil heterogeneity and the impacts of that heterogeneity on contaminant concentrations in soil
- representativeness and why it is critical to understand and incorporate this concept into planning and decision-making
- the relationship among the mass analyzed, the size of soil particles, and the resulting data variability

2.2.1 Soil data is the outcome of many steps

Figure 2-1 illustrates the process behind ISM, which theoretically intends to provide every soil particle in a DU with an equal probability of being incorporated into the incremental sample. The number of increments controls variability due to distributional heterogeneity (that is, nonuniform distribution of constituents across the SU/DU). The total sample mass controls the variability resulting from particle-to-particle compositional heterogeneity (CH), with is also called constitution or micro-scale heterogeneity. Most of the sample processing steps, starting with drying, are best conducted under controlled conditions in the laboratory. If sieving or grinding are to be done, they must be done to the entire field sample before any sample mass reduction (splitting or subsampling) to preserve the same sample mass to particle size relationship of the field sample in the analytical subsample. It is important to note that even flawless analytical procedures will produce unreliable data results if the upstream steps do not produce a representative subsample for analysis.
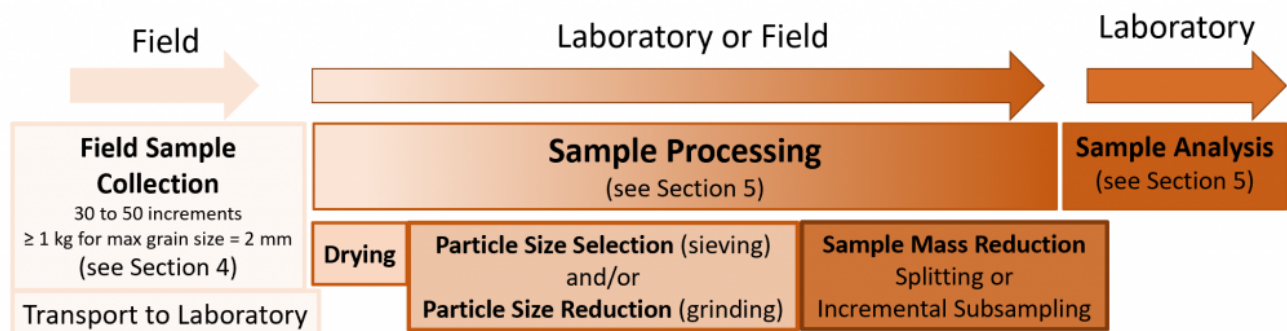

Figure 2-1. Three-stage ISM sampling process. Some processing steps may be optional, depending on soil characteristics, analytes of interest, and project objectives.
Source: ITRC ISM Update Team, 2020.

Project-specific planning must first define the soil information needed to address project decisions. This step helps to identify information gaps and guide the selection of the type(s) of soil data needed to fill those gaps, which in turn drives the development of a sampling design detailing where and how to collect soil samples to represent the needed information. Each subsequent link in the sampling and analysis chain must be tailored to ensure that the generated data accurately represent field conditions relevant to decision-making.

As illustrated in Figure 2-2, ensuring that an analytical subsample is representative of the volume of the DU is not an easy task. Here are some questions to ask when developing the sampling design for collecting representative samples:

Figure 2-2. Subsample sizes and representation. A single arbitrary clump of soil will not have the same concentration, and therefore properly represent, the volume of soil being excavated.
Source: Roger Brewer, HDOH. Used with permission.

- What is the chance that a 1- or 10-g analytical mass scooped from a sample jar has the same concentration as the entire field sample?
- What is the chance that a 200-g scoop of field soil will have the same concentration as the 1 or 2 tons it is supposed to represent?
- Will analytical data support a correct decision if the representativeness of the field sample and the analytical sample are both left to chance?

The following sections help to address these questions.

2.2.2 Soil heterogeneity affects contaminant concentration heterogeneity

The term *heterogeneity* refers to the condition where components of a matrix differ from each other. There are two types of heterogeneity of concern for soil sampling. One is distributional heterogeneity, which refers to the nonuniform distribution of constituents within a volume of soil. Field samples are collected at this scale. The other is CH, which refers to differences between particles of different composition that make up a handful of soil and cause them to have very different loadings of contaminant atoms or molecules. Chemical concentration heterogeneity is also affected by the source matrix (such as Pb shot, paint, PCBs in transformer oil, or ash with dioxins/furans) and the chemical and physical properties of each chemical (for example, ionic charge, lipophilicity, or chemical reactivity). Specific measures for addressing heterogeneity are provided in Section 2.6. Further discussions on the use of ISM to manage soil heterogeneity are covered in Section 2.4.

The following subsections demonstrate with empirical data how heterogeneity can cause decision errors.

2.2.2.1 Scales of heterogeneity

At the micro-scale, soil is heterogeneous in both physical properties and chemical properties. A look at a handful of garden soil makes this obvious. If the soil is dry so that it crumbles, you might see dust-sized clay particles, tiny particles of glittering silt, small sandy particles, little stones, and bits of plant matter in varying stages of decay. Each of these components has different physical and chemical properties causing the particles to interact differently with contaminants. Interacting at the atomic and molecular scales, contaminants bind weakly to certain types of particles but strongly to others,

based on a particle's mineral composition, size, surface area, molecular porosity, and chemistry (especially positive versus negative electrical charges).
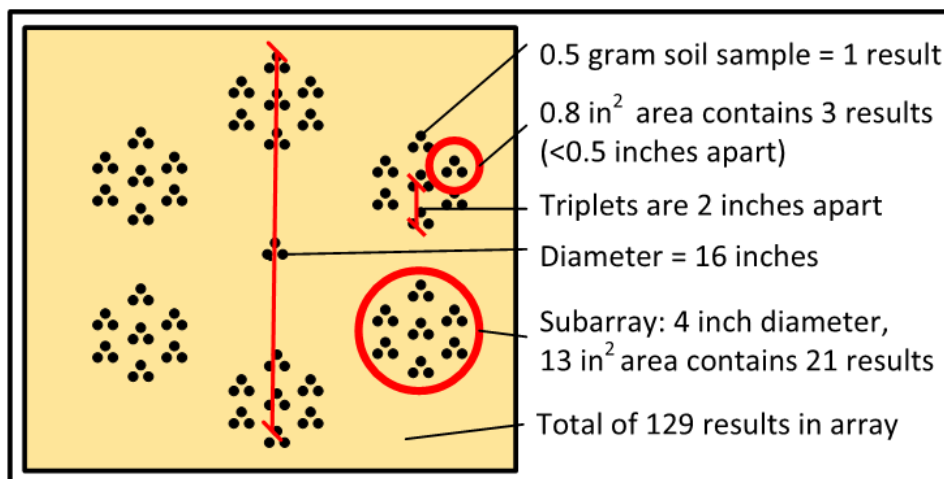

Figure 2-3a. Array pattern for 0.5-g soil samples, which encompass a circular area of 1.4 ft2 and contain 129 individual samples.
Source: Developed by Deana Crumbling, 2014. Used with permission.

Contaminant concentrations can vary greatly over small distances, too. Generally, this occurs because contaminant release and deposition mechanisms never lay down a uniform amount of contaminant into the soil volume. Variation can also occur because of disturbance of the soil after contamination has occurred. For example, human and animal activities can bring cleaner soil up from the subsurface. Landscaping and gardening can move soil around or bring in foreign soil from an uncontaminated location. Wind, rain, and other weathering events can also create stratification and/or cause selective mixing to occur.

The research, discussed below in Section 2.2.2.2, measured the variation in Pb concentrations at the centimeter scale over soil surfaces contaminated by the disposal of Pb-containing paint about 50 years prior to the study. The paint was no longer visible, but the Pb impacts remained.

2.2.2.2 A real-world example

Figure 2-3a shows the sample collection pattern used in a study that examined the effect of Pb concentrations on soil microbes (Becker 2005). Each dot in Figure 2-3a represents a single 0.5-g surface soil sample analyzed in its entirety. The

0.5-g samples were collected in the 1.4 ft^2 array pattern shown. The members of a triplet group (three adjacent samples, shown within the smallest red circle and referred to as Becker triplets) were only 0.5 in apart. Seven triplets were grouped into a subarray of 21 samples (within the larger red circle).
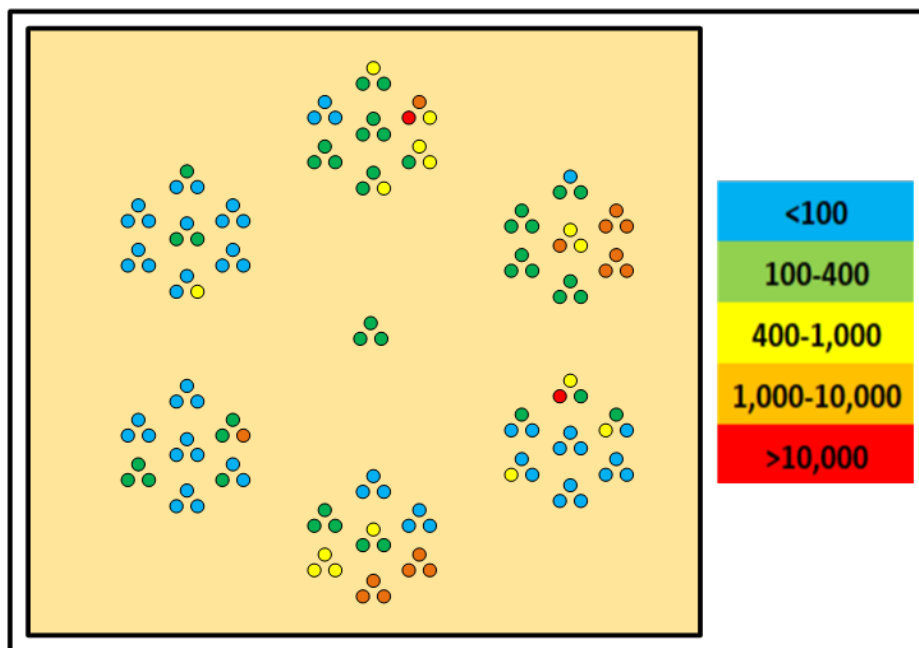
Figure 2-3b. Becker triplets with color-coded sample concentrations. Markedly different concentrations can occur in soil scoops only 1 cm apart. From this dataset, basing a decision for a 400-mg/kg lead threshold on a single random result would underestimate contamination.
Source: Developed by Deana Crumbling on data from Becker, 2017. Used with permission.

The color-coded dots of Figure 2-3b approximate the concentration of each 0.5-g grab sample in the array. The color key shows the range of Pb concentrations in bins. The number of samples in each concentration bin are shown in Table 2-1.

Table 2-1. Number of 0.5-g samples in concentration bins.
Source: Deana Crumbling, 2019. Used with permission.

Concentration Bins (mg/kg Pb)	Number of Samples
10 to 100	55
100 to 400	42
400 to 1000	15
1000 to 10,000	15
10,000 to 29,000	2
Average concentration = 833 mg/kg Pb	Total n = 129
Median concentration = 120 mg/kg Pb	

The 0.5-g scale of heterogeneity demonstrated in the Becker study helps explain the high data variability often observed in soil data. The 0.5-g analytical mass of this study is on par with the mass of analytical subsamples used for routine soil metals analysis, and the spatial dimension of Becker triplets (Figure 2-3b) is on par with the spatial scale of duplicate grab subsamples taken from a field sample jar in the laboratory. These are called *laboratory replicates* because this a laboratory subsampling QC check. Laboratory replicates measure micro-scale heterogeneity within the field sample.

Particle heterogeneity creates data variability. In effect, each analytical subsample has its own concentration, which can be markedly different from other subsamples. The true concentration of all soil in the sample jar is the mean of all subsamples

contained within the jar. However, the chance of any random subsample having the same concentration as the overall sample can be low. The mean Pb concentration for all 129 samples in the 1.4 ft^2 Becker array is 883 mg/kg, but individual concentrations range from 13 to 29,185 mg/kg. In the lower right subarray in Figure 2-3b, a sample with a Pb concentration >10,000 ppm (red dot) is 1 cm away from a sample having a concentration <400 ppm (green dot).

Note that the true concentration for this 1.4 ft^2 area is 883 mg/kg, double the project soil screening level of 400 mg/kg Pb. However, unless action is taken to control for the problem of short-scale distributional heterogeneity, mere chance can cause the correct "positive" decision (exposure risk from Pb could be a problem) to erroneously be a "negative" decision

> ***There is a severe mismatch between the scale of decision-making (tons of soil in a DU) versus the scale of data from analyzing a few grams of soil (see Figure 2-2).***

(there is no risk from Pb exposure). The goal of soil sampling is to provide information representing the true mean concentration of soil within the DU, but soil samples are likely to be non-representative when soil heterogeneity has not been managed. The failure of sample collection and handling procedures to provide a representative result is called *sampling error*.

It is important to understand that there may be nothing wrong with the quality of the laboratory analysis itself. The problem tends to occur in the field sample collection and sample processing steps (refer to Figure 2-1). Without careful, active management of sampling procedures, the analytical mass will not represent the field unit. Data representativeness means a sample result (or group of sample results) must provide an accurate estimate of the overall contaminant concentration in the tons of soil making up the soil volume. For correct decisions about soil volume, the penny-sized mass that is analyzed must somehow be handled in a way that ensures it has the same mean concentration as the tons making up the original sampled volume. An analytical subsample created by sampling only a small volume of soil cannot accurately represent a large volume of soil. Fortunately, a Frenchman named Pierre Gy developed a sampling theory from which procedures were derived that are much more effective at producing a representative sample. For that reason, Gy's theory and its techniques have been recommended in guidance documents from USEPA and other agencies (USEPA 1992c, 1999a, 2002d, 2003, HDOH 2017a, USACE 2009, ITRC 2012). Additional information on Gy's theory of sampling (TOS) are discussed in Section 2.6.

2.3 Mean ISM QC

One way to determine whether a particular sampling design is providing sufficiently accurate estimates of DU concentrations is to run replicates to see if it meets the DQOs. These replicates are used to evaluate if the repeated independent sampling events result in similar concentrations being detected. If not, the design needs to be changed to one that is reproducible. This is why a key QC feature of ISM is the collection of field replicates on all or a portion of the DUs in the project. This topic is discussed in more detail throughout Section 3 and Section 4.

2.4 Physical Causes of Soil Data Variability

Soil heterogeneity is the most important cause of soil data variability. Soil heterogeneity occurs at all spatial scales, from microscopic soil particles to landscapes, but the only relevant scale is the one that matches the scale of decision-making. Heterogeneity below this level can be considered noise.
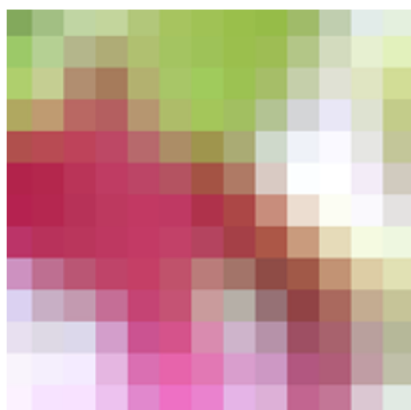
Ideally, the scale of measurement or observation would directly match the decision scale. For example, consider the scale of cleanup for a given project is the mass of soil scooped up by a backhoe bucket. If cleanup decisions were made one bucketful at a time, the concentration of each bucketful would decide if the soil is destined for the landfill or will remain onsite. If the mean concentration of a contaminant in a 100-kg scoop is greater than the cleanup threshold, it does not matter whether there are some 10-g soil clumps scattered within it that have low concentrations of the contaminant. Even if analysis of one 10-g clump finds the contaminant to be non-detect, that concentration is irrelevant because cleanup cannot be done 10 g at a time.

2.4.1 Soil particles create micro-scale heterogeneity and data variability

Soil particles are the tiniest of soil pixels. When examined microscopically, we find that soil particles carry different contaminant loadings depending on their size and composition (type of mineral or organic content).
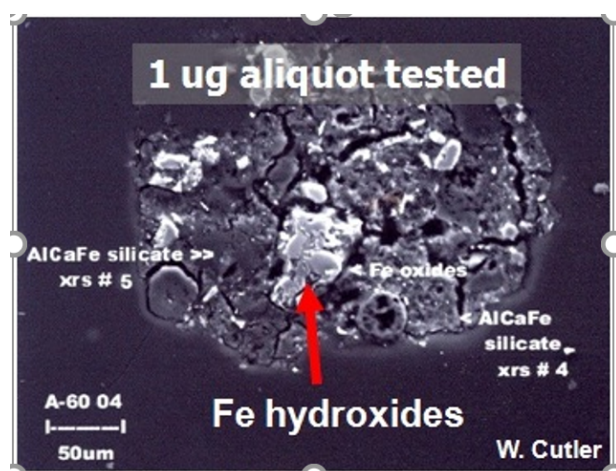

Figure 2-4. Microscopic iron hydroxide grains coated with arsenic, which appears as a light-colored deposit (red arrow). Arsenic-laden iron mineral nuggets are scattered among "clean" particles made of silicate minerals.
Source: Roger Brewer, HDOH, 2012. Used with permission.

Within the same clump of soil, some particles will carry very high loadings and others nearly none. Extremely small particles of certain minerals, such as clays and iron oxides, are well-known to bind high concentrations of contaminants via ionic interactions and large surface areas (USEPA 1992b, Hassellov and von der Kammer 2008, Fye et al. 2002, Weng, Temminghoff, and Van Riemsdijk 2001, Xiong et al. 2018). Particles heavily burdened with contaminants are sometimes referred to as "nuggets" (USEPA 2002g). Special spectral microscopy techniques allow some types of metal contamination to be visualized as isolated nuggets amid many clean particles, such as the example shown in Figure 2-4.

If soil particles could be analyzed individually, a clay nugget would be a pixel with an extremely high concentration, while a particle of pure silica or feldspar would be a non-detect pixel. An analytical mass contains hundreds to thousands of particles, so the reported concentration is the mean of all those individual particle concentrations. The digestion or extraction of an analytical mass averages out the micro-scale heterogeneity within that mass.

There is an extreme mismatch between the scale of measurement (that is, the scale of observation) and the scale of decision-making. The challenge is to generate soil data in a way that makes it representative of the decision, rather than representative of the scale of observation.

At trace concentrations (such as mg/kg and smaller), nuggets make up a very small percentage of all the particles in a sample. Yet those nuggets often carry an outsized portion of a sample's contaminant load. Even when the small particle fraction makes up a minor part of the bulk sample, most of the contaminant load is concentrated in the small particle fraction. An example of this is illustrated in Figure 2-5, where a Pb-contaminated soil sample from a firing range was sieved into six particle sizes. The smallest particle size is less than 0.074 mm (74 microns), which means a pile of these dust-sized particles has the look and feel of sifted flour. This fraction mass made up only 34% of the total sample mass, yet it carried 75% of the sample's Pb content (green bar in Figure 2-5). The two smallest sized fractions have grains only visible under a microscope, yet together they carried 97% of the sample's Pb.
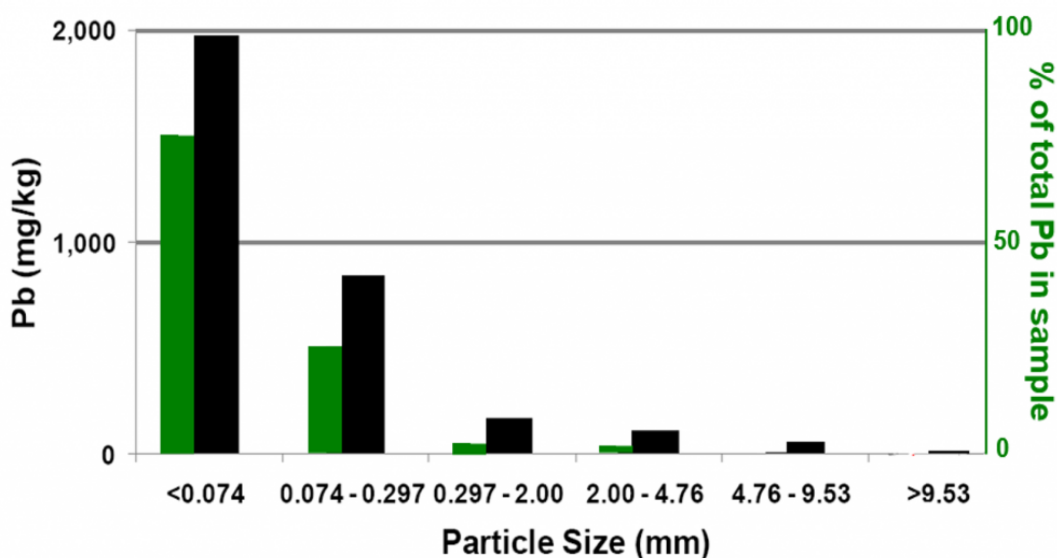

Figure 2-5. The relationship between particle size and Pb concentration for a soil sample collected from a firing range. The black bars show the Pb concentration (in mg/kg) for each of six particle size fractions. The green bars show the percent of the total sample Pb contained by each particle size fraction.
Source: Developed by Deana Crumbling from data presented in ITRC, 2003. Used with permission.

The black bars in Figure 2-5 show the measured concentrations for each particle size fraction. Analyzing the entire sample (stones, sand grains, and dust) gives a concentration of about 900 mg/kg. But the concentration of the smallest fraction by itself (black bar in the graph) was nearly 2,000 mg/kg. When particles become large enough to be visible to the naked eye (the third largest fraction, 0.297 mm to 2 mm), the Pb concentration dropped to 165 mg/kg. For reference, Table 2-2 illustrates particle size as it relates to sieve size.

Table 2-2. Particle size and sieve size.
Source: ITRC ISM Update Team, 2020.

Sieve Mesh No.	Particle Size in mm	Sieve Mesh No.	Particle Size in mm
400	0.037	28	0.7
325	0.044	25	0.707
270	0.053	20	0.841
230	0.063	18	1.0
200	0.074	16	1.19
170	0.088	14	1.410

Sieve Mesh No.	Particle Size in mm	Sieve Mesh No.	Particle Size in mm
140	0.105	12	1.68
120	0.125	10	2.0
100	0.149	8	2.38
80	0.177	7	2.83
70	0.21	6	3.36
60	0.25	5	4.0
50	0.297	4	4.76
45	0.354	3	6.73
40	0.42	2	12.7
35	0.5	1.3	19.0
30	0.595	0	25.4

This concentration of contaminant mass onto small particle fractions creates a problem for laboratory subsampling. Flowing particles readily segregate under the influence of gravity, meaning smaller, denser particles settle to the bottom, leaving coarser sized fractions near the top. This makes it difficult to obtain representative analytical subsamples from a field sample as it is impossible to collect a subsample representing the particle distribution of the entire jar by simply scooping off the top of the sample. Stirring cannot solve the problem either (Figure 2-6). The problem gets worse as the mass of the analytical subsample gets smaller. This phenomenon, and how to address it, is described in detail in Section 2.6.

Freshly collected soil sample in jar: particles of many sizes intermixed

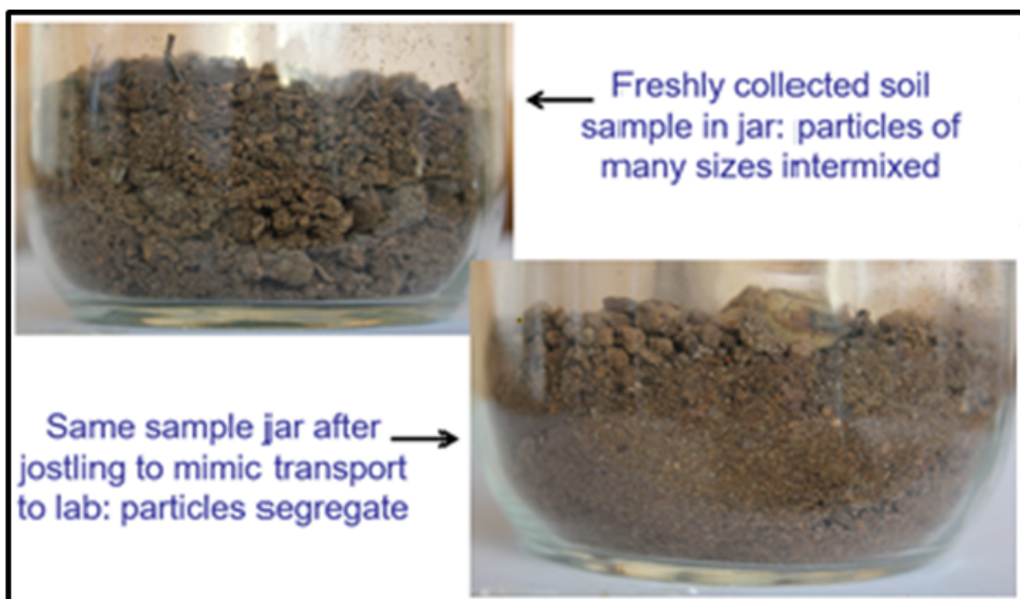Same sample jar after jostling to mimic transport to lab: particles segregate

Figure 2-6. Jostling of a sample jar causing smaller particles to settle downward while larger particles remain at the top. Stirring to mix is ineffectual to redistribute particles evenly and often makes segregation worse.
Source: Deana Crumbling, 2013. Used with permission.

Without procedures to actively manage micro-scale heterogeneity, nuggets contribute to high data variability in another way, as shown in Figure 2-7. The overall concentration of a field sample (the large container) reflects the overall ratio of nuggets (black dots) to cleaner particles (white dots). Analytical subsamples (the three smaller colored containers) will have the same concentration as the whole field sample if the subsample holds the same particle ratio. The chance that a grab subsample will have the same ratio depends on the physical size of the analytical subsample and on random chance. A larger subsample (the red jar in Figure 2-7) has a better chance of capturing the same ratio as the overall sample.

Even if nuggets are evenly distributed throughout the sample mass, smaller subsamples have a higher chance of missing nuggets, giving a lower ratio, and having a falsely low subsample concentration, as illustrated with the blue subsample illustrated on Figure 2-7.

Occasionally, a small subsample will capture more than its share of nuggets (the green subsample on the right side of Figure 2-7). This inflates the subsample concentration since concentration is calculated as . In other words, if the same analyte mass is contained in a smaller soil mass, the concentration gets higher. For example, if 1,000 g of soil contained 1 mg of Pb in the form of nuggets, the Pb concentration would be 1 mg/kg. However, if the same 1 mg of nuggets were picked up in only 100 g of soil (that is, a soil mass 10 times smaller), the concentration would be 10 mg/kg (10 times larger).
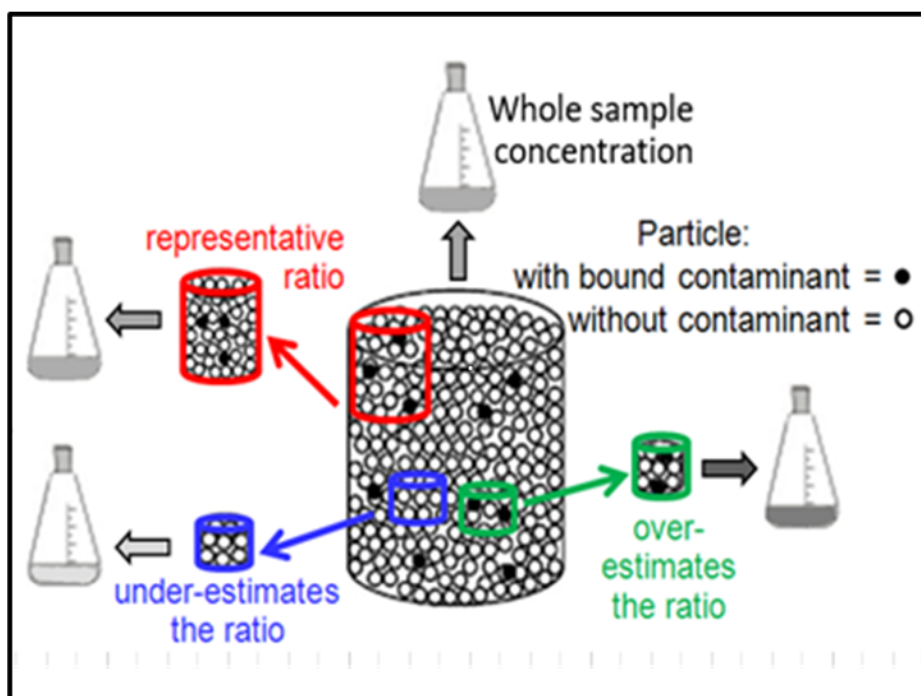

Figure 2-7. A subsample obtaining a representative ratio of nuggets to cleaner particles depends on subsample mass. A sampling error occurs when the subsample does not mirror the ratio of the field sample, producing a significantly higher or lower concentration result.
Source: Developed by Deana Crumbling on data from EPA 2002a, p. 92. Used with permission.

The relationship between contaminant loadings on certain types of soil particles and analytical mass is an important cause of data variability, and it explains the observation that soil data commonly take a lognormal data distribution. This relationship was studied in 1978 by the U.S. Department of Energy (Doctor and Gilbert 1978). The results of the experiment are shown in Figure 2-8. The experiment involved preparing a very large soil sample containing a radioactive analyte. After the large sample was sieved and milled, analytical subsamples of different masses were collected from the parent sample. Twenty replicate subsamples were collected each for analytical masses that ranged from 100 g down to 1 g.
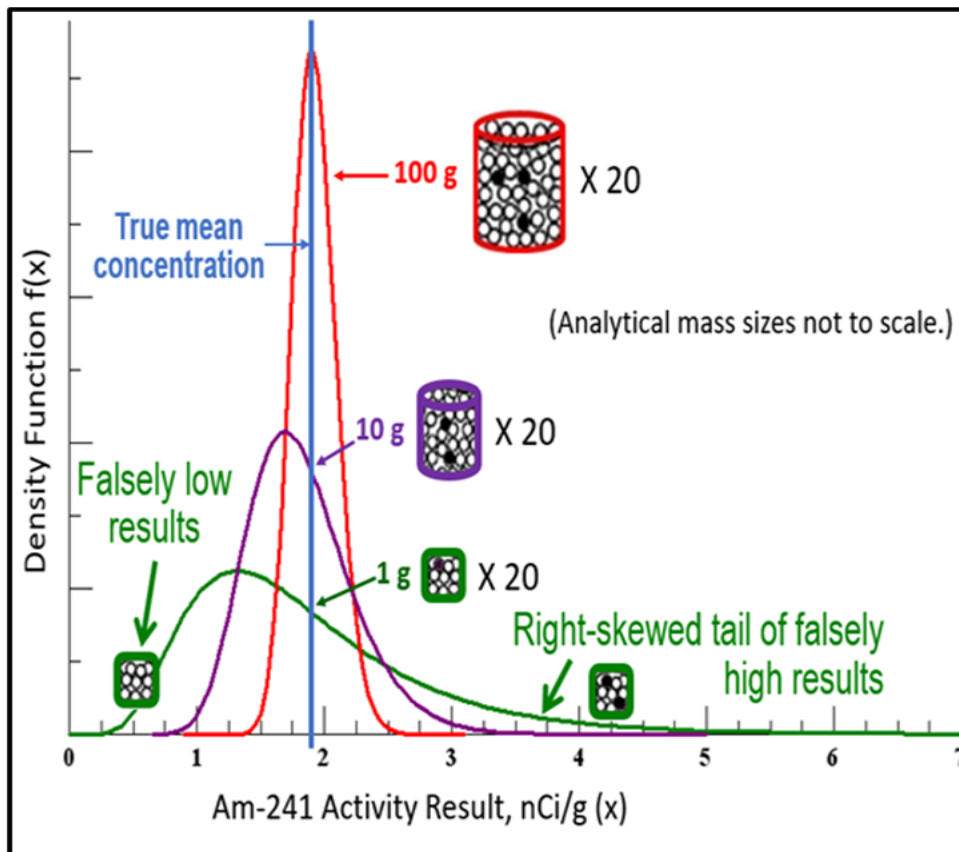
Figure 2-8. Analytical mass influencing the degree of data variability and statistical distribution. Data from 100-g analytical subsample masses (tall, narrow, red peak) were very precise with a normal data distribution. Data from the 10-g analytical mass (purple peak) were less precise, as shown by the broadening of the peak and the right-side tail is becoming skewed. The data distribution from the 1-g mass (green) is quite wide, with a marked right skew. The true concentration of the entire parent sample is indicated by the vertical blue line, which agrees perfectly with the mean of the 100-g subsamples.
Source: Developed by Deana Crumbling from Doctor and Gilbert, 1978. Used with permission.

The use of larger analytical masses in ISM samples results in data that are more precise and more likely to be representative of the parent sample because it normalizes out the constitutional micro-scale heterogeneity. As the analytical mass decreases, not only do data become more variable (less precise), but the data distribution goes from normal to more skewed (that is, more "lognormal"). As shown in Figure 2-8, the skewed data distribution observed for small analytical masses is the direct consequence of in-sample heterogeneity, whereby some subsamples are missing analyte nuggets (so their concentrations are low or non-detect) and some are capturing too many analyte nuggets (so their concentrations are very high). One of the advantages of working with normal data distributions is that they facilitate the use of statistical analyses that are much simpler, more informative, and require fewer data points.

> *The use of larger masses in analytical subsamples results in data that are more precise and more likely to be representative of the parent sample because it normalizes out the compositional heterogeneity.*

2.4.2 Field heterogeneity and data variability

Figure 2-9. Arsenic concentrations for in situ XRF readings spaced about 2 ft apart in a residential yard.
Source: Unpublished data from Deana Crumbling, 2007. Used with permission.

Short-scale distributional heterogeneity is the difference in concentration between field locations that are relatively close together (Figure 2-9). Co-located discrete samples measure short-scale distributional heterogeneity often at a distance of a

foot or less. The assumption is that concentrations in close locations will be nearly the same, but this is often not the case, as exemplified in Figures 2-10 through 2-13.

Short-range variability is concentration variation on a spatial scale that is inconsequential to the decision for a designated volume of soil. One of the central tenants of ISM is that the scale of evaluation is tied to the volume of soil about which decisions are being made. For example, it is not feasible to make different individual decisions for samples located 1 foot apart, such as those samples shown in Figure 2-10.
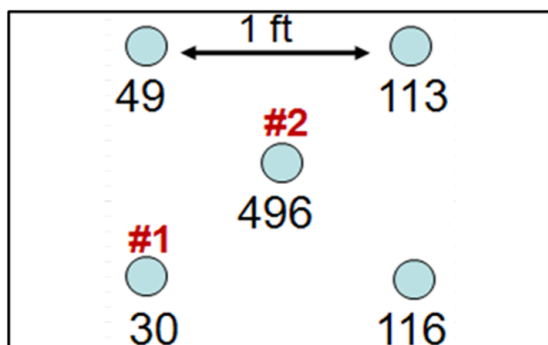

Figure 2-10. Uranium concentrations for five co-located samples located 1 ft or less apart. A decision might be based on location number.
Source: USEPA and DOE, 2008.

A receptor does not live or work solely on the same 496-mg/kg uranium spot for 10 to 30 years. Furthermore, even if a receptor *did* reside in this one location for 30 years, the uranium would be transferred to the receptor over time, thereby depleting the source area of uranium and reducing the exposure concentration in the soil over time. Figure 2-11 and Figure 2-12 give additional examples of short-range variability.

For such reasons, USEPA's risk assessment guidance states that, "in most situations, assuming long-term contact with the maximum concentration is not reasonable." Therefore, "[t]he concentration term in the intake equation is the arithmetic average of the concentration that is contacted over the exposure period." (USEPA 1989a).
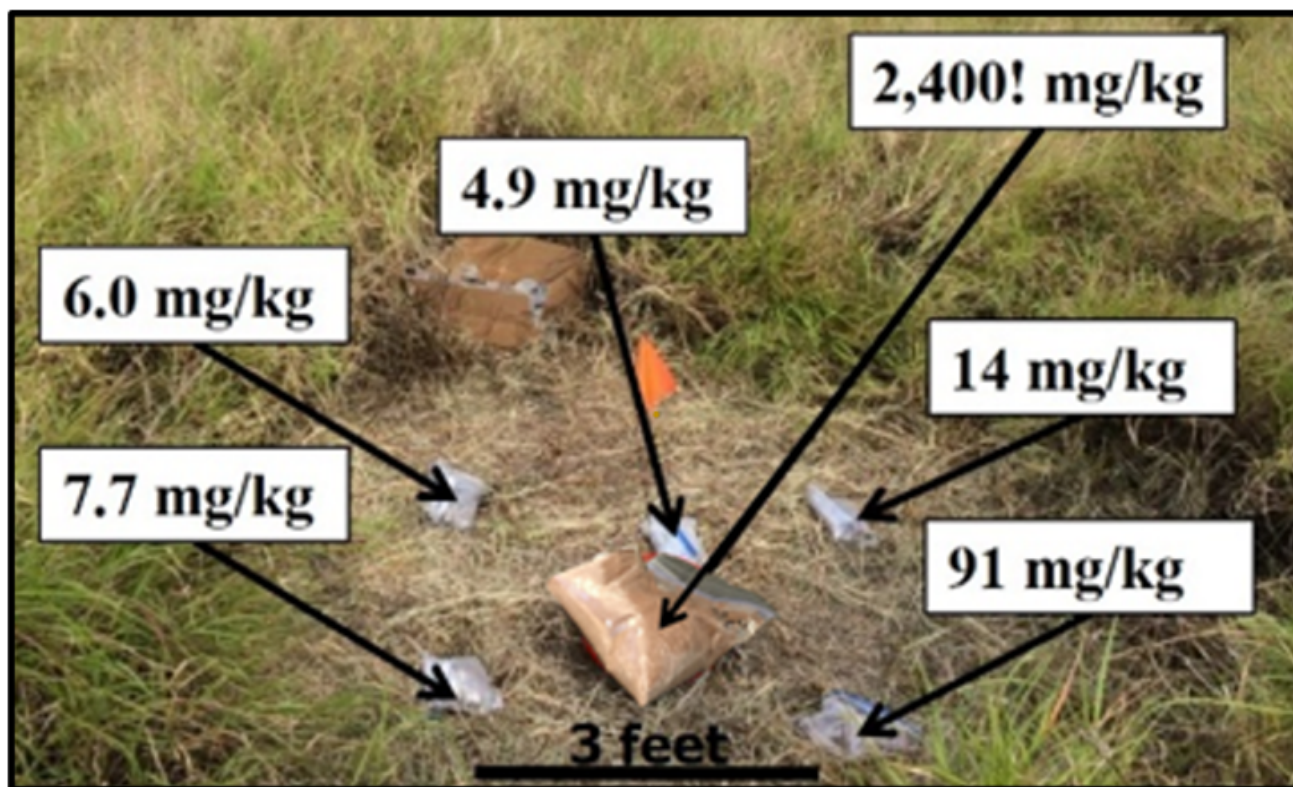

Figure 2-11. PCB concentrations resulting from five small samples and one large sample arranged in an area of only 9 ft2. The samples were all collected immediately adjacent to each other.
Source: Developed by Deana Crumbling from data by HDOH, 2005. Used with permission.
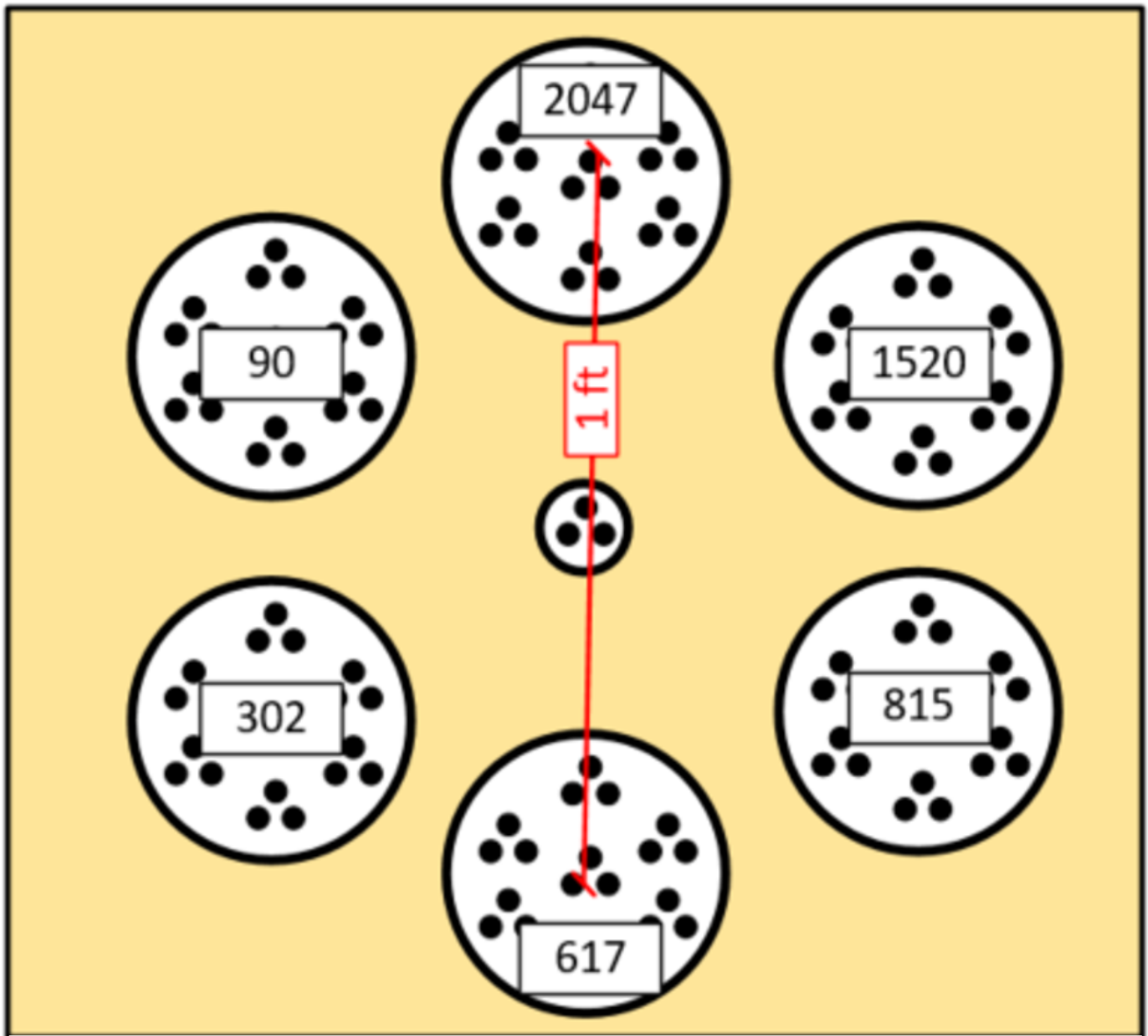
Figure 2-12. Mean Pb concentrations grouped into six 10.5-g subarrays. Results range from 90 to 2,047 mg/kg at a scale of less than 1 ft apart.
Source: Deana Crumbling, 2019, Used with permission.

Figure 2-13. Spilled milk from an overturned trailer.
Source: Adapted by Deana Crumbling from HDOH, 2005. Used with permission.

One question practioners may ask is why heterogeneity occurs on such small spatial scales. To address this, consider the following: if a spill occurred 10 years ago from a drum, you might expect the area to be uniformly contaminated, but this is likely not the case. As an example, Figure 2-13 shows a liquid tanker spill, and since a soil surface is rarely perfectly flat, we would expect the spill to flow around bumps and follow the topography. Years later, these bumps may be gone, but the soil where they formerly were might have non-detect concentrations (the yellow dot in Figure 2-13), whereas the soil under what were once tiny swales around the bumps may have high concentrations (the red dot). A handful of discrete samples will have hit or miss contamination by chance, providing an inaccurate CSM. However, ISM will capture the overall concentration for the area, proving that a significant release occurred. Similarly, a simple model of wind-blown contamination does not deposit as evenly as a conceptual model might depict. Buildings and trees may serve as windbreaks, creating eddies that swirl winds and suspended dust into unpredictable patterns. By the time sampling occurs, structures may be gone, but the heterogeneous contaminant patterns they created remain stamped into the soil.

2.5 Sample and Data Representativeness

The idea of representativeness can be murky for practitioners. In QA project plans (QAPPs), sample representativeness is often stated in terms indicating representativeness will be achieved by following the sampling procedures and using the designated analytical methods. However, QAPPs rarely explain how the intended use of the data guided the choice of those procedures and methods. Well-designed QAPPs will positively state what the data are intended to represent. For example, a QAPP should state the data will represent the hand-to-mouth exposure route for surface soils in a residential yard for the purpose of estimating risk from soil Pb. Once it is clear what the data will represent (that is, how the data will be used), the next sequence of specifics can be addressed (that is, what depth for surface soils, what soil particle size mediates that exposure route, and so on).

2.5.1 Terminology associated with sample representativeness

In 1996, the American Society for Testing and Materials (ASTM) defined a representative sample as "a sample collected in such a manner that it reflects one or more characteristics of interest (as defined by the project objectives) of a population from which it was collected" (ASTM 1996).The property of interest is determined by the decision the data are to support; for ISM, the characteristic of interest is the population mean. In 2002, USEPA further defined representative sample based on the Resource Conservation and Recovery Act's (RCRA's) definition of it as "a sample of a universe or whole (e.g., waste pile, lagoon, ground water) which can be expected to exhibit the average properties of the universe or whole" (USEPA 2002g, 2012a); (USEPA 2012a) (USEPA 2002a). The importance of tying sample representativeness to a project decision was reinforced in an article in *Environmental Forensics*: "A representative sample is one that answers a question about a population with a given confidence," and, "A sample that is representative for a specific question is most likely not representative for a different question" (Ramsey and Hewitt 2005).

Within ISM, the term DU is used to refer to the "whole" or the "population," meaning that which is subjected to a decision about risk, attribution, or cleanup.

2.5.1.1 Understanding DUs

DU originated from USEPA guidance and is defined as "a volume or mass of material (such as waste or soil) about which a decision will be made" (USEPA 2002g).Some USEPA DQO guidance documents also use the terms *action support* and *scale of inference* to refer to the same

> *Some practitioners of ISM use the term DU to describe the volume of soil represented by an incremental sample regardless of what type of decision is associated with that sample. Other practitioners prefer to apply the term DU to all volumes of soil represented by a sample, even if ISM results may not support the site's final endpoint decision for that volume of soil. For example, results from smaller DUs could be used to determine the concentration of a waste pile or sediment catch basin, or to evaluate worker safety within a specific area. In this document, DU is defined as the smallest volume of soil for which a decision will be made.*

concept. As USEPA's 2006 DQO guidance (also known as G-4) states on page 32, "The scale of inference is the area or volume, from which the data will be aggregated to support a specific decision or estimate" (USEPA 2006b).

> *DUs are a core concept of ISM and the pivot point around which all ISM designs revolve.*

Although there are other terms, the term DU is favored by ISM practitioners for its intuitive simplicity. DUs are a core concept of ISM and the pivot point around which all ISM designs revolve. A DU can be thought of as the full area and volume of soil that you would send to the laboratory as a single sample for analysis if you could. This might include, for example, all the soil in a backyard where children routinely play. This is not practical, however, so a representative sample of the targeted area and volume of soil must instead be collected. There are many strategies for constructing DUs, thereby producing different types of DUs. The type most familiar to many practitioners is the EU, which is used to support decisions about risk (USEPA 2002g). Section 3.1 introduces this topic in further detail.

2.5.1.2 Introducing the sampling unit

SU also appears in USEPA guidance documents, where the term refers to the volume of soil from which a sample is collected:

- The 1992 Sampling Techniques and Strategies guidance states the goal of sampling is to "estimate the concentration in a sampling unit or a specific volume of soil" (USEPA 1992b).
- The same 1992 document indicates a "sampling unit [is a] unit of soil, waste, or other particulate matter that is to be sampled...such as the remedial management unit or exposure unit."
- USEPA DQO guidance refers to "a half-acre area...for the sampling unit," and, "...the probability of making an incorrect decision for a particular sampling unit" (USEPA 1992b).

The term SU is therefore best used generically to mean any defined volume of soil intended to be represented by a sample or samples, or "the area and depth of soil (the sampled population) to be represented by the sampling process" (USACE 2009). In other words, an SU is the smallest volume of soil for which a concentration value will be obtained, no matter the intended purpose of the data.

Within ISM, SUs are commonly used to support a specific decision, such as a risk-based decision for a known exposure or source area. However, SUs can also support other kinds of sampling objectives. For example, small SUs may be used to gather spatial information or statistical data in ways that do not fit neatly into the definition of a DU (see Section 3.1.5) or SUs that illustrate the other sampling objectives in Section 3.1.6. The same project may also use both true DUs and SUs. To avoid confusion in planning sessions, some practitioners choose to apply the generic term SU to these smaller units to distinguish them from standard DUs. However, some practitioners may exclusively use DU for all types of SUs. The choice of terminology is the prerogative of the project team, but the terms selected should be clearly defined by the project team during the planning stages.

> *One type of atypical unit involves a series of small areas (sometimes only a few square feet each) sampled to gather information about contaminant spatial patterns or concentration trends. Information from these small SUs is used to guide placement and sizing of DUs for specific decision-making. The pattern created by the group of SUs influences subsequent DU design.*

2.5.1.3 Sample support and its role in sample representativeness

A representative sample depends on proper field collection across a DU. The reasoning behind constructing a DU is to obtain an estimate of the DU's mean contaminant concentration for the purpose of supporting a decision on that whole unit (USEPA 2002d). A DU can be several tons of soil (the big picture), but our technology is limited to measuring grams. Therefore, everything about the sampling strategy must be tailored so the few grams of soil analyzed to best represent the DU in the context of the decision.

The concept of *sample support* helps with this. Practitioners intuitively use this concept without giving it a name, but naming it makes discussion easier. USEPA's RCRA waste sampling guidance defines it this way: "The size, shape, and orientation of a sample are known as the sample support... [f]or heterogeneous media, the sample support will have a substantial effect on the reported measurement values" (USEPA 2002g). The sample support describes the characteristics of an increment (that is, two ISM samples may both contain 30 increments and yet have different sample supports).

Figure 2-14 illustrates the various aspects of sample support and how it affects sample results. The figure depicts a waste material rather than a soil, with the DU extending through the entire depth of the waste pile. It is important to know the concentration of the blue particles in the whole DU volume, even though they have settled to the bottom of the DU.
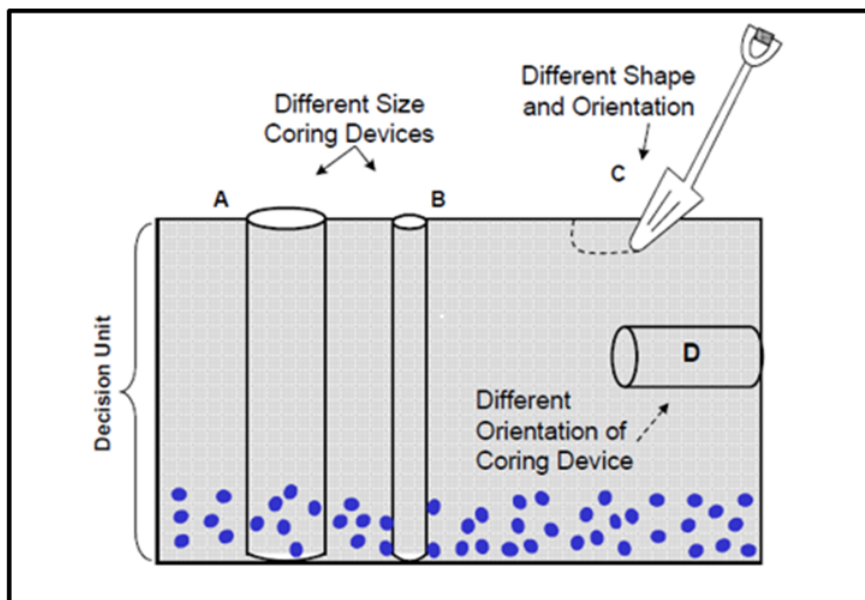
Figure 2-14. Sample support. Only device A has the correct sample support (size, shape, and orientation) to extract a representative sample, which includes the blue dots from the DU.
Source: USEPA, 2002b.

One aspect of sample support is the size of the sample, which is determined by the tool used to extract the waste material. Figure 2-14 conceptualizes the effect of sample support size. Corer B is too small: it pushes aside the blue particles, so samples collected by Corer B will be non-detect or biased low for blue particles. Corer A correctly picks up blue particles in the proper ratio.

Another aspect of sample is shape. A shovel (Device C in Figure 2-14) is the wrong shape because it cannot sample through the full thickness of the DU. Corer D is the correct width, but its orientation is wrong. Coring horizontally from the side of the pile will not obtain a sample representative of the full DU thickness.

The sample support used for the analytical subsample is also critical, a topic that will be covered in more detail in Section 5.

2.5.2 Achieving sample representativeness despite DU heterogeneity

The main objective of Gy's TOS is to allow the sample collected to directly represent the entire mass of soil within the targeted DU. This is accomplished by physically collecting a relatively large mass of soil (such as 1 to 3 kg) and collecting the sample from as many points as practical. A large mass helps to address CH, while a large number of increments helps to address what is referred to as *distributional heterogeneity* (Minnitt, Rice, and Spangenberg 2007, Pitard 2019).In combination, both help to address errors in physical collection of the sample. These relatively simple concepts are the essence of sampling theory for the testing of particulate matter such as soil.

If all the soil within a DU's volume were collected and analyzed so that every gram of the DU soil is included in the measurement process, the result obtained would be the DU's true mean concentration, which is the intended basis of DU decisions. Since the entire DU cannot be sampled, two options remain: the practitioner may either collect discrete samples to be analyzed individually and then mathematically average the results, or the practitioner may sample increments, which will be pooled together into one sample for a single analysis (such as ISM). Either way, when selecting either the first or the second option, the intent is to obtain a value that is close to the true DU mean.
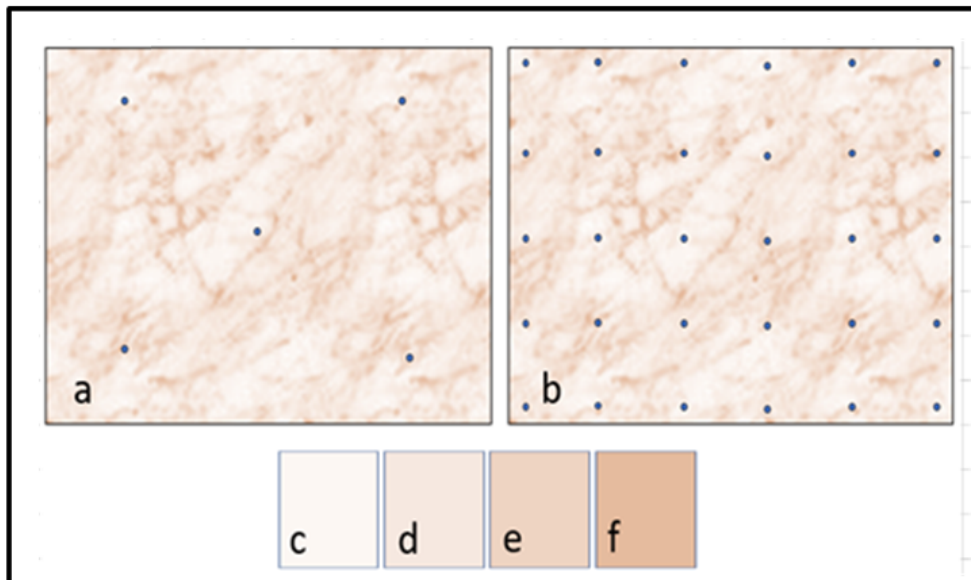
Figure 2-15. Contaminant concentration. Panels "a" and "b" show a heterogeneous DU sampled by five and 30 increments, respectively. Panels "c" through "f" depict concentration levels that range from very low (near white) to high (darkest tan).
Source: Deana Crumbling, 2019. Used with permission.

The previously referenced Becker dataset demonstrated that contaminant concentrations can differ markedly from location to location throughout a DU's volume when measurements are made at the scale of an increment. This can be conceptualized by the variations in color shading within the DU depicted in Figure 2-15. Panels "a" and "b" are of the same DU, the only difference is the number of soil increments as represented by dots in the figure. The tan color in represents a concentration bin (a narrow range of concentration values), four of which are depicted in Panels "c" through "f." The lowest bin is the lightest color (panel "c"), and its area dominates the DU. Higher bins progress through the darker tan shades (panels "d" through "f"), which take the form of clumps and strings that permeate throughout. The milk spill in Figure 2-13 illustrates how patterns resembling strings might be created. Clumps of higher contamination might occur through soil disturbance by animal or human activity. The only possible strategy is to define a specific area (a DU) and obtain the mean over that area. It is important to define the area in a way that provides the information needed to support project decisions. In Figure 2-15, if the four tan shades were averaged out over the entire DU, panel "d" would represent the true concentration of the DU. If the overall DU concentration exceeds a cleanup threshold, the entire DU would be cleaned up.

Obtaining an estimate of mean concentration equivalent to panel "d" is unlikely when only a few increments are used (as shown in panel "a") because it would be difficult for areas of varying concentrations to be sampled in their correct proportions. With more increments (panel "b"), a practitioner is more likely to capture the varying concentrations in their correct proportions, producing a more accurate estimate of the overall DU concentration. The statistical foundation for this concept is further discussed in Section 3.2.4.

The number of increments within the DU can be thought of as *sampling density*. The spacing of increments is a spatial consideration that classical statistics does not accommodate, yet areas of elevated concentration are a spatial attribute that can be important to the overall concentration of the DU. When utilizing ISM, a default of at least 30 increments for up to 0.5 acres is used. One reason why ISM uses so many increments is to have sufficiently dense coverage of the DU area. This improves the chance that the field sample will pick up significant areas of elevated concentration in the same proportion as present across the DU.

Choosing the number of increments under conditions of randomly dispersed contamination can be done by using classical statistics (see Section 3.2.4.2). However, a limitation of classical statistics is that it does not recognize spatial considerations, such as the area of a DU. Whether a DU is 0.05 acres or 5 acres, classical statistics will give the same number because it is driven primarily by the degree of data variability (as standard deviation [SD]) provided as an input parameter. Using statistics to calculate an exact number of increments per DU requires various inputs that are seldom available. These inputs can be determined from a pilot study that collects the data needed to calculate them. Pilot studies may be cost-effective if the site involves hundreds of DUs, or if increment collection is difficult, but most practitioners find it easier to go with the default number of 30 increments unless there are site-specific reasons to increase that number.

As supported by statistical simulations (see Section 3.2.4) and over a decade of practitioner experience, a default of 30 increments per field sample is sufficient to control random contaminant heterogeneity in most cases. Obviously, DUs and contamination do have spatial dimensions. For many reasons, contamination is never purely random. However, the default

of 30 increments will be successful for non-random contamination in most cases where heterogeneity within the DU is not too large. Section 3.1.6.2 offers considerations about when not to use the default of 30 increments, such as known heterogeneity of data due to chemicals of potential concern (COPCs) or source type (for example, paint chips with lead, lead shot, transformer oil with PCBs, or lipophilic compounds such as organochlorines).

> ***The actual number of increments needed to represent a DU's true mean concentration depends on three things, all of which should be key components of CSM and DQO development:***
> * ***the degree of within-DU heterogeneity when variations are more or less random across the DU***
> * ***the presence of significantly large sections within the DU that have higher or lower concentrations***
> * ***the presence and size of small pockets of higher or lower concentrations within the DU***

A reason for doing replicate incremental field samples is to make sure that the current number of increments per sample and the total mass of the bulk sample are sufficient. If they are not, there is a strong chance that replicate results will not agree. An example can be found in the PCB field study performed by the Hawaii Department of Health (HDOH) (Brewer, Peard, and Heskett 2017). Although the sample area was small (6,000 ft^2, or about an eighth of an acre) and 60 increments per field sample were used, the three replicate sample results were 19, 24, and 270 mg/kg for total PCBs. Analytical results were corroborated through confirmation analysis, and the concentration of PCBs was so heterogeneous that 60 increments was not enough to address the distributional heterogeneity, even in a very small DU.

2.5.3 Spatial correlation within a DU

Spatial correlation means that the soil samples or increments near each other are more likely to be similar in concentration than those farther away. Spatially correlated areas within a DU can be large or small, and classical statistics can be an imperfect tool to determine the number of samples when spatial correlation is present. However, there are mechanisms within ISM to handle sampling depending on the scale of the spatially correlated areas.

2.5.3.1 Large areas of spatial correlation

Figure 2-16 displays a scenario in which the DU has been defined as encompassing an entire residential property (pink outline), and the yard as a whole is considered to represent a single, exposure area DU. The objective of the investigation is to assess the risk posed by contaminants in the yard by comparing the DU mean to a pre-established, risk-based screening level.
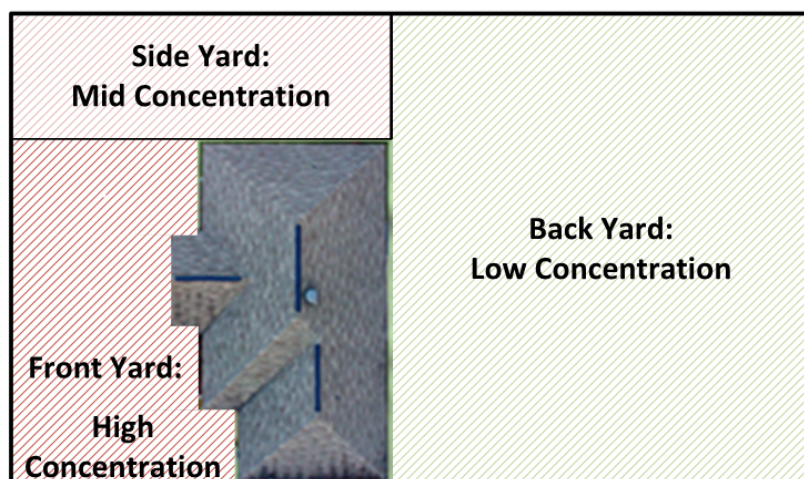

Figure 2-16. Schematic of a DU encompassing a residential property with distinctly different concentrations in the front, side, and back yards. Darker shades represent higher concentrations. *Source: ITRC ISM Update Team, 2020.*

Unknown to the sampling team, different sections of the yard have significantly different concentrations. (Section 3.1.6 has examples of how to design ISM studies with SUs that must consider spatial correlation for exposures and known or suspected spatial differences in contaminant distribution.) The concentration of soil within a section is spatially correlated because the concentrations within a section are more similar to each other than to concentrations in the other yard sections. It is also likely that the heterogeneity of contaminant distribution within each section is also different from that in the other sections.

If the default 30 increments are systematically collected across the entire yard DU, there may be only four or five increments from the front yard where the concentration (and likely heterogeneity) is higher. The results from three

independent replicate DU-ISM field samples using this design would not be likely to agree and could lead to a decision error about the yard. This is because collecting only four or five increments in the front yard is unlikely to capture the variability (that is, the highs and lows of contaminant concentrations) in the correct proportions to the rest of the yard. There are various options for addressing this problem. One is to enlarge the mass of each increment to reduce the degree of micro-scale variability and allow each increment to represent the SU mean. Another is to increase the number of DU increments so even the smaller SUs will have enough increments. The latter is the preferred option because the final mass of the ISM sample needs to meet limits for management by the laboratory (typically 2 to 3 kg maximum).

USEPA indicates that areas of elevated concentration could be relevant if "located near an area which, because of site or population characteristics, is visited or used more frequently, exposure to that area should be assessed separately" (USEPA 1989a). The classic examples are regularly tended garden beds or children's play areas. For this reason, it is common practice for residential ISM designs to split out such areas into their own DUs apart from the rest of the yard.

In cases where large-scale distributional heterogeneity is known or suspected ahead of time (for example, Pb contamination from a roadway source adjacent to the front yard), a better approach would be to stratify the yard into three smaller DUs. Samples consisting of 30 increments are then collected from each of the smaller DUs, along with replicates. Stratification splits a population into subgroups that are "internally consistent with respect to a target constituent or property of interest and different from adjacent portions of the population" (USEPA 2002d). An estimate of the mean concentration for the overall yard is determined by area-weighting the section means (see Sections 3.1.6.2 and 3.1.6.3 for examples of designing ISM studies for area-weighted mean concentrations). Stratification as a sampling strategy is extensively discussed in USEPA guidance in the context of discrete sampling (USEPA 2002c).

2.5.3.2 Small areas of spatial correlation

In some cases, areas of non-random contamination, smaller in scale than the DU in which they occur, may be present. The exact locations of these impacts are often unknown or are too small and scattered for successful stratification of the DU. However, a stratification approach may be possible if the CSM predicts – or if field data suggest – there is a zone of elevated concentrations within the DU identifiable during the DU planning process.

Figure 2-17 illustrates a DU where high-density in situ x-ray fluorescence (XRF) analyses reveal areas of elevated Pb contamination scattered non-randomly within a residential yard. Prior digging related to utilities or other such disturbances (such as gardening) are examples of activities that can cause non-random contaminant distribution. The result creates irregular patches of overall higher concentrations. Furthermore, areas of clean soil are also likely present, such as from soil/dirt fill purchased for landscaping activities. None of these features may be obvious during field observation, and therefore the likelihood of collecting proper representative discrete samples is very low for providing representative ratios of the overall DU.
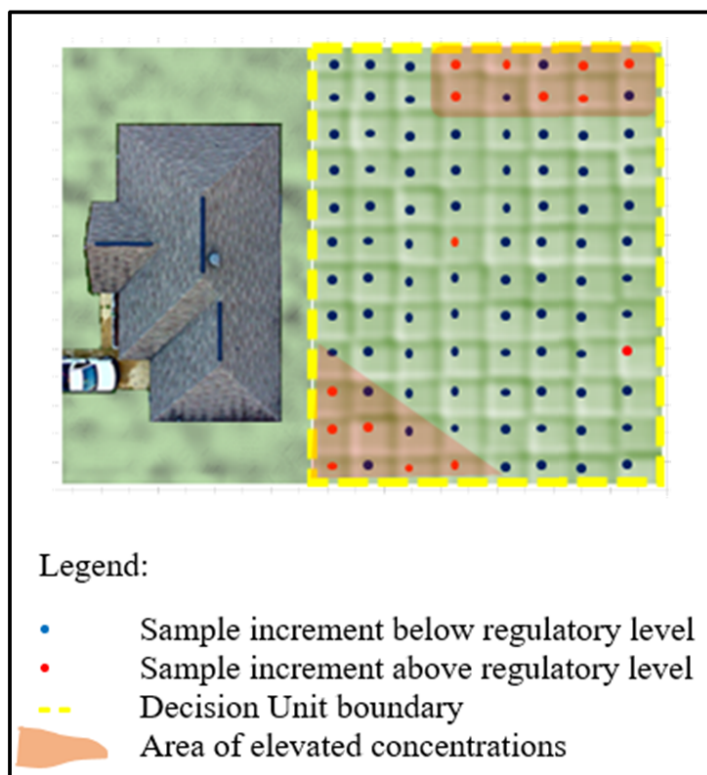

Legend:

- •   Sample increment below regulatory level
- •   Sample increment above regulatory level
- - -   Decision Unit boundary
       Area of elevated concentrations

Figure 2-17. Information from real-time in situ analyses. The yellow areas depict zones.
Source: Deana Crumbling, 2020. Used with permission.

Some practitioners may decide to pinpoint areas of elevated concentrations within a DU in order to initiate a targeted removal. However, in practice, this is not the best approach. Samples collected via ISM are meant to be physically representative of the entire DU and the exposures associated with it. If project DQOs are expressed in terms of the DU mean, then areas of elevated concentration within the DU do not necessarily need to be removed. As a reminder, the DU mean is the basis for a decision, not the isolated areas of impacted soil. This is particularly true for EUs and risk-based regulatory thresholds (see Section 3.1.5.1).

If areas of elevated concentrations are present and significant, it is vital for these concentrations to be represented proportionally in the ISM sample. Statistically based and other sampling designs can be developed to determine whether localized areas of higher soil concentrations exist (see Section 3.1.5) but require sufficient increment mass and density. The goal is to sample these concentrations in the same proportion as the representative area within the DU. If consistently missed during sampling, these concentrations will not be incorporated into the physical mean of the ISM field sample, and the true DU concentration will be underestimated. This is why having a current and detailed CSM is critical. Although the default of 30 increments is sufficient in many cases, it is important to note that, in certain chemical classes (such as munition residues, metals at small-arms firing ranges, and PCBs), more than 30 increments may be necessary due to the highly heterogeneous way these contaminants can be distributed in soil (see Section 3.1).

2.5.3.3 Managing areas of elevated concentration with ISM
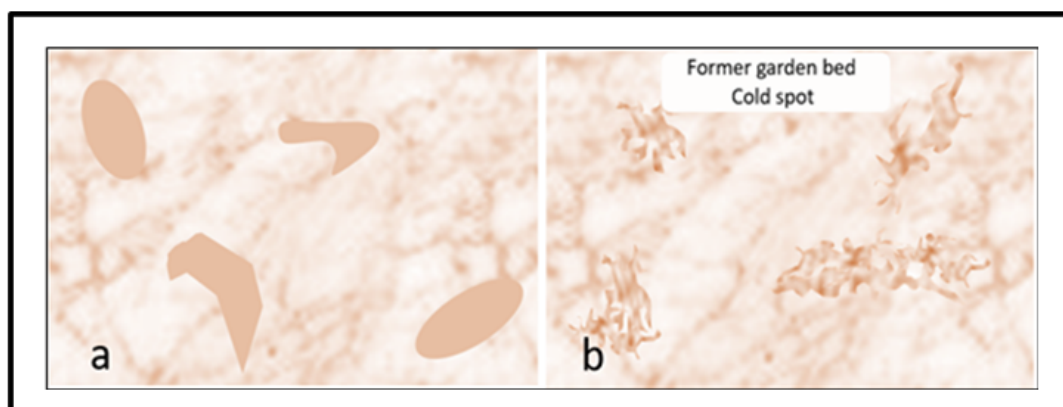

Figure 2-18. Concentration misconceptions. Panel "a" depicts a common misunderstanding about areas of elevated concentration, showing homogenous groupings with defined margins. Panel "b" presents a more realistic representation of such areas along with adjacent areas of lower concentration.
Source: Deana Crumbling, 2019. Used with permission.

Contrary to how some may conceptualize areas of elevated concentrations, actual areas of elevated concentrations within a DU do not appear in homogenous groups with well-defined margins. In Figure 2-18, panel "a" illustrates this misconception. Real-world areas of elevated concentrations will look more like panel "b," with irregular edges blending into the surrounding soil. Such irregularity is expected to be pervasive throughout an entire area, such that a single soil increment (that is, a discrete sample) is not guaranteed to be representative of the elevated constituent concentrations present in soil throughout the DU.

As such, with the expected variation in concentrations throughout the DU, the sampling design must be developed to be as representative of the DU concentrations as possible. Figure 2-19 provides depictions of the likelihood of a suitably representative outcome based on three separate designs.
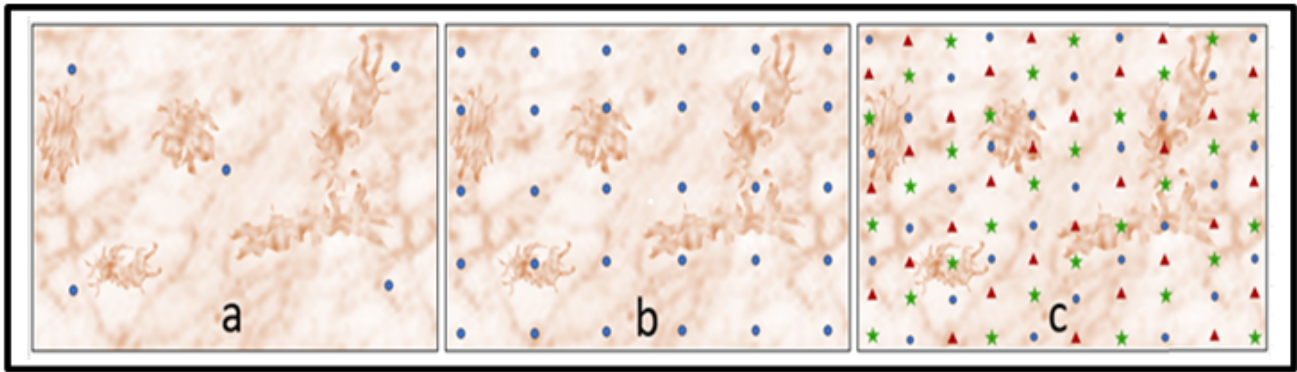
Figure 2-19. Possible design outcomes. Panel "a" illustrates a pattern of five increments, panel "b" shows a pattern of 30 increments, and panel "c" shows 90 increments (three independent replicate samples). The more increments, the better the chance the ISM sample will pick up high concentration areas. Three replicate field samples further reduce the chance that areas of elevated concentration will remain unknown and provide a measure of precision.
Source: Deana Crumbling, 2019. Used with permission.

ISM includes several field and laboratory processes and techniques for ensuring that all of a DU is represented proportionally in a sample. These field and laboratory processes are based on sampling theory and concepts established by Pierre Gy, as discussed in Section 2.6. ISM also features QC procedures designed to measure overall sampling and analysis precision, including the collection of field replicates. One way to assess whether a particular sampling design is providing sufficient estimates of DU concentrations is to replicate the design. Since the goal is to have a sampling protocol that can be replicated multiple times to provide similar, repeatable results, it is important for a practitioner to understand the CSM and implement the proper techniques to provide for the desired representativeness. This design is then implemented in the field. If the replicate results for the samples do not show the desired sampling precision, the design may be flawed. For example, panel "c" of Figure 2-19 shows a sampling design that includes collecting three replicate 30-increment samples from a DU. If the results for the samples do not show the desired measure of precision, the CSM may need to be reevaluated. One reason for data imprecision may be because the number of increments collected was not adequate to represent non-random higher concentrations scattered throughout the DU. As such, the collection of field replicate ISM samples is a key QC feature of ISM.

2.5.4 Errors in data interpretation due to non-representative samples

Decisions based on non-representative data can have a costly impact on projects. The data validator examines the quality of the analytical process itself, but rarely does the data validator have the information needed to judge the representativeness of the data. That task should have been accomplished early on, by the project team. Once the data move into the validation phase, time and money have already been expended, which is why planning is so crucial. Decision errors about the extent of contamination are common when decisions are based on a handful of discrete samples. These errors are further exacerbated when there is a moderate to high degree of heterogeneity in the area being sampled (Pitard 2019), yet it is not uncommon for many practitioners to base decisions on a single sample result, such as one sidewall sample in an excavation determining if the lateral extent of contamination has been addressed. In these instances, the assumption is made that the result represents the actual concentration of an ill-defined volume of soil in the field, when, in fact, the result may stem from an isolated area of contamination being sampled or a small amount of mass randomly taken from a soil sample jar

> *Replicate QC sampling and data quality evaluation are not typically available in composite or discrete sampling designs. This sets ISM designs apart from the usual concept of how conventional soil sampling works.*

for analysis. The resulting high laboratory detections may lead to false assumptions, such as the presence of widespread contamination where there is actually very little. Alternatively, low laboratory results may indicate that a site is less contaminated, thereby leading to erroneous decision-making for the project.

Project managers regularly experience the frustration of returning to delineate an area of elevated concentration identified by a single discrete sample, only to find the area has "disappeared," or more accurately, the sample result cannot be replicated. In addition, in many cases, confirmation sampling may identify contamination where none was identified previously because of these same issues.

> *When soil heterogeneity is not anticipated and controlled by a soil sampling design, an endless cycle of decision errors perpetrated by non-representative data can inflate cleanup cost and project lifecycles.*

The decision to assign a grab field sample to a grid cell (such as illustrated in Figure 2-11) in the geostatistical analysis can also be compromised if apparent concentration differences are due only to random variability and do not reflect real trends in field concentrations.

2.6 Managing Heterogeneity to Ensure Sample Representativeness

As early as 1989, USEPA's Office of Research and Development (ORD) was issuing guidance to advise practitioners that managing soil heterogeneity was as important as analytical quality for obtaining reliable results (USEPA 1989c).

The same 1989 document briefly referenced the work of a leading thinker in the field of sampling particulate materials, Pierre Gy, who worked in the European mining and industrial sector. Gy spent 25 years developing procedures for taking representative samples from particulate materials such as ores and cereal grains. Understanding how to obtain representative samples was vital to the mining industry because non-representative exploration samples could lead mining companies to spend considerable sums of money trying to extract valuable materials that were not actually there in sufficient quantity. Later, Francis F. Pitard made Gy's work accessible to a U.S. audience that included USEPA's ORD (Pitard 2019).

2.6.1 Gy's TOS

Gy's TOS was first discussed in an USEPA guidance document in 1992 (USEPA 1992a). In 1999, ORD's technology support staff developed a two-page flyer discussing Gy's the theory and emphasizing its value to USEPA for soil sampling (USEPA 1999a, 2003):

- "The inherent heterogeneity of soils presents a particular challenge to field personnel...It affects the manner in which analytical chemists subsample in the laboratory....heterogeneity influences the interpretation of data and the decisions made about the actions taken to remediate contamination at a site."
- "[Gy's] theories...present practical sampling and subsampling methods that can be applied for little or no added expense. [They] can result in samples that better represent the site and data that more truly represent the sample."

ORD subsequently hosted training sessions on Gy's TOS developed and delivered by Pitard. Descriptions of the theory and its application to controlling soil sampling errors were included in additional USEPA technical guidance documents (USEPA 2002e, 2003, 2004a).

2.6.2 Gy's TOS

Gy's work explains why the analysis of heterogeneous particles produces highly variable data, provides the language to discuss the various related sampling problems caused by heterogeneity, and most importantly, explains what to do about it. The theory covers at least seven distinct ways that heterogeneous particulate materials affect sampling integrity. The resulting variability, bias, and non-representativeness of data are collectively termed *sampling errors*. Gy's sampling errors originate from three general sources: the material being sampled, the effectiveness of the sampling equipment, and whether the sampling procedures use that equipment correctly (Minkkinen 2004).

Only those concepts and sampling errors that directly affect data variability in contaminated soil are discussed in this section. The specific procedures used to control sampling errors will be discussed thoroughly in the sections that deal with field sampling (Section 4) and laboratory subsampling (Section 5). *RCRA Waste Sampling Draft Technical Guidance* (USEPA 2002g) is written at the beginner level and omits complex equations. In contrast, USEPA's subsampling guidance is very thorough, but the reader should be prepared for technical language and equations (USEPA 2003). Section 12.3.1.4 and Appendix F in Volume 2 of a multi-agency guidance referred to as MARLAP coherently and thoroughly covers Gy's theory and implementation tools pertaining to subsampling and is easy to understand (USEPA 2004b). Gy wrote a series of technical papers detailing the mathematical aspects of sampling theory that appeared in a single issue of the journal *Chemometrics and Intelligent Laboratory Systems* (Gy 1988). Many other researchers have published discourses on Gy's work as it applies to contaminated soil, including (Petersen, Minkkinen, and Esbensen 2005) and (Minkkinen 2004).

2.6.2.1 Fundamental error

Fundamental error (FE) is the first error in Gy's list. It is fundamental because it is a consequence of the fundamental

makeup of the different particles comprising the material to be sampled. FE cannot be reduced by any degree of blending because mixing the particles around does not alter the nature of the particles. FE is a consequence of compositional micro-scale heterogeneity, where particle composition influences particle density, size, and shape, the properties that affect particle movement under the influence of gravity (such as the segregation problem already described Figure 2-6). Chemical composition influences the loading of contaminants onto the individual particles that create nuggets of certain particles but not others (see Section 2.4.1). Gy's theory predicts that as particles get larger or as the range of particle sizes increases, FE grows larger as the cube of the particle diameter (USEPA 2002g).
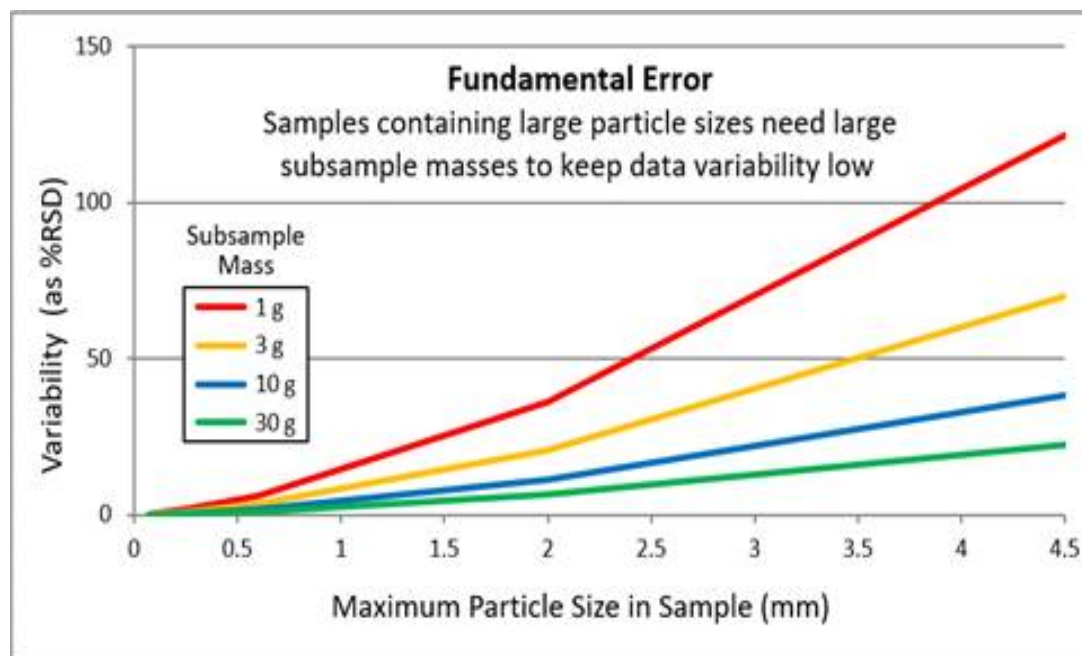

Figure 2-20. Fundamental error as illustrated from equations from ASTM.
Source: Mark Bruce, Eurofins. Based on information provided in ASTM, 2012.

ISM reduces the particle size aspect of FE either by sieving samples to a small particle size (when this is appropriate to the decision; see Section 2.4.1), or by milling (or crushing) particles to a fine powder (Figure 2-21). Milling is vital if larger particle sizes are the target population, meaning that sieving them out to obtain a smaller particle size is not an option (see Section 2.5.2.3). Reducing FE in the field may involve taking more field increments and/or making increment mass larger (such as using a larger diameter core; see Figure 2-14 and Section 4.2.3). For subsampling, digesting or extracting a larger soil mass can also reduce FE. This is especially important if larger particle sizes need to be analyzed, and milling is not recommended. Gy's theory predicts that if the maximum particle size in the material is 2 mm (that is, what goes through a 10-mesh sieve), an analytical subsample should be *at least* 8 g (see Section 5.3.5). The relationship between analytical mass and particle size is explained in Appendix D of "A Quantitative Approach for Controlling Fundamental Error"(USEPA 2002g).

2.6.2.2 Grouping and segregation error

Grouping and segregation error (GSE) can occur during field sample collection or subsampling conducted in the laboratory. At the microscopic scale of individual particles present in a soil sample, *segregation* occurs when particles of different sizes or densities separate (Figure 2-21,left panel). GSE occurs when the subsampler fails to adjust the sampling procedure for sample conditions, and the subsample is collected as if the sample is uniform throughout. If the sample is not uniform, the data will be biased due to segregation error.
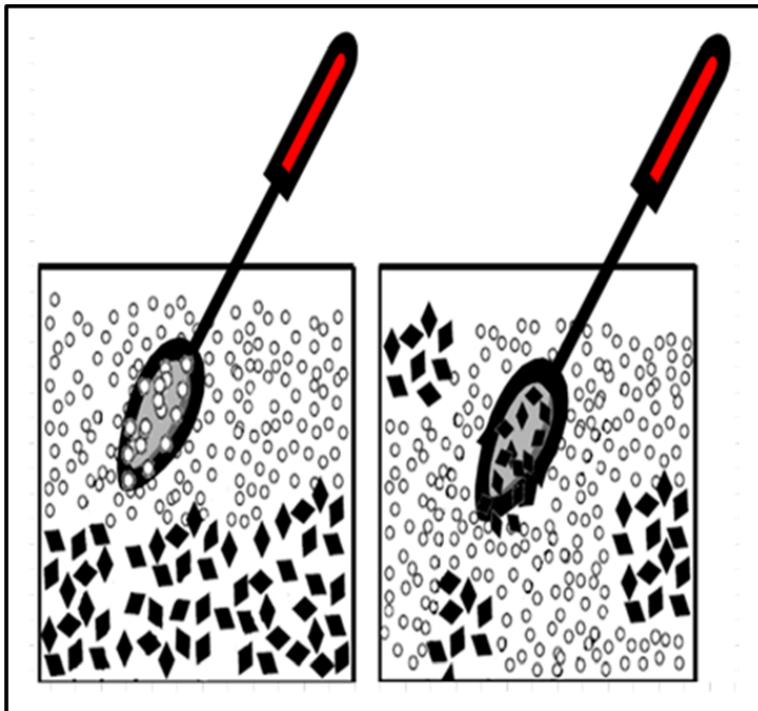
Figure 2-21. Segregation. Non-representative subsampling following the incomplete stirring of a segregated sample creates grouping error.
Source: Developed by Deana Crumbling from data by HDOH, 2015. Used with permission.

Particle grouping can occur for various reasons, for example, when there is a perfunctory attempt to stir a segregated sample (Figure 2-21,right panel). The particles grouped together by a spoon's bowl during stirring can stay grouped together and not be uniformly dispersed. USEPA's subsampling guidance warns that mixing alone is ineffective for homogenization: "Many analysts rely on mixing (or blending) as a preliminary 'homogenization' step before taking a grab [sub]sample. Unfortunately, many samples cannot be made homogeneous enough for representative grab [sub]sampling by mixing, and such a procedure should not be relied upon in the laboratory to reduce [grouping and segregation error]. Segregation of particles by gravitational effects usually occurs at the moment that the mixing has stopped" (USEPA 2002g, 2003). Grouping error is also manifested by cohesive clayey chunks and will not disperse by simple stirring.

GSEs are both caused by distributional micro-heterogeneity (the state where particle types are not uniformly distributed throughout the sample volume; see Section 2.5.3). Subjecting a sample to vigorous blending is not enough to reduce distributional heterogeneity sufficiently to enable representative grab subsamples. "Grab [sub]sampling has been shown to be an unacceptable [sub]sampling method and should not be used with particulate samples" (USEPA 2003). "At least 30 increments are recommended as a rule of thumb to reduce [grouping and segregation error]" (USEPA 2003). This is for laboratory subsamples of 30 increments; Section 5 describes the specifics of these procedures.

> *GSE is addressed by ISM procedures in the laboratory that use special techniques and tools to take many random increments that collectively will form the analytical subsample.*

Figure 2-22. Grouping error. Contaminants grouped by deposition mechanisms (red dots) may bias the sample estimate of the DU mean if sample collection is biased.
Source: Developed by Deana Crumbling from HDOH, 2015. Used with permission.

Grouping error can also occur at the scale of field sampling. Imagine the former low-lying areas in Figure 2-22 (red dots) are filled in with finer sediments over time and are then covered by vegetation. The surface soil outside those filled areas tends to be rocky and more difficult to sample (yellow dots). Grouping error occurs if field collection staff gravitate toward "softer" locations, where it is easier to sample to the intended depth. If the ISM field sample is intended to represent the DU mean, that estimate could be biased high by over-representing areas where fluid flow and contaminants were concentrated or grouped (that is, where red dots outnumber yellow dots). Field samplers need to be warned against inadvertently biasing samples by frequent deferral to easier sampling locations if the CSM suggests contaminant grouping could be a factor. ISM uses systematic random field sample locations for increment collection to avoid this bias.

Delimitation error occurs when improperly shaped tools are used to collect samples. An example is using a scoop with a curved shape instead of a scoop with a square shape. Figure 2-23 shows why the round-shaped scoop causes a sampling error. The square-shaped scoop (bottom panel) correctly delimits, or cuts, a subsample shape that represents the full thickness of the sample and all particle sizes. In contrast, the rounded scoop incorrectly delimits a shape that discriminates against the small particles at the bottom of the pile. The smaller particles are therefore under-represented in the scooped soil. The goal of avoiding delimitation error is to obtain the different particle sizes in the same proportions as present in the sample. ISM controls delimitation error by specifying the appropriate field and laboratory sampling tools.

Delimitation error is also closely connected to the sample support, as shown in Figure 2-14.
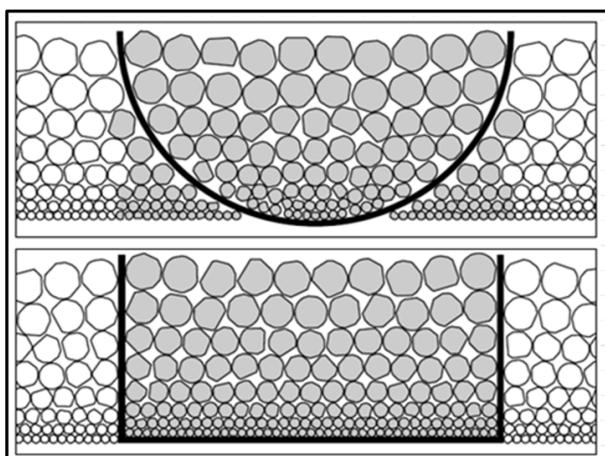
2.6.2.4 Extraction error

Figure 2-23. Delimitation error. A scoop with a round bottom (top panel) discriminates against the fine particles settled at the bottom of the pile, in contrast to the rectangular shape of the lower panel.
Source: USEPA Multi-Agency Radiological Laboratory Analytical Protocols Manual (MARLAP), 2004.

Extraction error is exemplified by sampling dry, sandy soils with an open-bottom corer. As the corer is withdrawn from the ground, the loose material in the bottom part of it can fall out and back into the hole. The sample would then over-represent the upper part of the DU volume because the portion of the core representing the lower part of the DU is lost.

This problem is not unique to ISM, but ISM practitioners are more aware of the opportunity for biased data. Because of this awareness, they attempt to try to correct or minimize the problem as much as possible.

The terms *correct sampling* and *correct sampling devices* were used by Gy to emphasize that proper techniques and tools are needed. If incorrect methods or tools are used to collect samples, field samples and analytical subsamples may not be representative of the parent material, which can lead to incorrect decisions. If the data user does not know how a sample was collected and handled, the term *specimen* is sometimes used to convey the concern that the representativeness of the sample/subsample is unknown.

2.6.3 Application of Gy's theory principles in ISM

Application of the principles of Gy's Theory to ISM are thoroughly discussed in Section 4 (Field Implementation) and Section 5 (Laboratory Processing). For the purpose of this introduction, the following summarizes how ISM integrates the Gy's TOS into practice.

2.6.3.1 Managing micro-scale within-sample heterogeneity

Obtaining a representative subsample from a field sample involves the following steps, although not all may be necessary for any particular sample or project (see Section 2.4 and Section 2.4.1).

Sample processing may include drying, disaggregating (breaking apart particles adhering to each other), sieving (to isolate a target particle size population), and/or milling (crushing particles to the uniform consistency of sifted flour) the field sample. Subsampling involves collecting increments from the processed sample that have been laid out in the form of a *slabcake*. The maximum particle size in the sample determines how much mass is required for a representative subsample (see Section 5.3.5).

QC in the form of independent replicate subsamples is vital to show that processing and subsampling procedures are working as intended. A number of corrective actions may be taken if QC finds that the processing and subsampling procedures are inadequate to provide reproducible results (see Section 5.4).

2.6.3.2 Managing distributional heterogeneity within a DU

Short-scale distributional heterogeneity can lead to decision errors. Compared to traditional sampling approaches, ISM manages, controls, and reduces the effects of short-scale distributional heterogeneity through the larger sample support size. Distributional heterogeneity within a DU is managed by collecting increments across the entire DU. Two factors help determine the number of field increments per DU:

- If the spatial distribution of contaminants is assumed to be random, then classical statistics can be used to determine the appropriate number of increments based on the degree of heterogeneity.
- If the spatial distribution is thought to have a pattern (such as areas of elevated concentration), the spacing of increments (and therefore the number of increments needed to fill the DU area) can be set to have the desired statistical probability of increments "hitting" the smallest area of elevated concentration for incorporation into the field sample (see Section 2.5.2).

The sample support (shape, orientation, and mass; see Section 2.5.1.3) of the increments must be chosen so that the field sample will represent the entire DU and be appropriate for the data use.

QC, in the form of independent field replicate samples, ensures that data are reproducible. Reproducibility is a strong indicator – but not a guarantee – of an unbiased estimate of the DU mean. Many variables affect the reproducibility of soil data, and only strict control over these variables will allow for three or more replicate results to closely agree, ideally approximating the true mean DU concentration (see Section 4).

2.6.3.3 Managing large-scale heterogeneity

"Large-scale heterogeneity reflects local trends and plays an important role in deciding whether to divide the population into smaller internally homogenous decision units" (USEPA 2002d). This is the spatial scale of concentration differences that project managers are trying to find. Large-scale distributional heterogeneity governs three areas:

- the nature and extent of contamination

- the degree of receptor exposures
- the selection of EUs for estimating receptors' exposures and risks

This phenomenon in environmental sampling and analysis investigations does produce heterogeneity in the site dataset that is used for decision-making, hence it is a key factor in designing DUs for ISM. The degree of large-scale distributional heterogeneity dictates decisions about the number of DUs identified, the location of the DUs, and their dimensions.

Click Here to download the entire document.