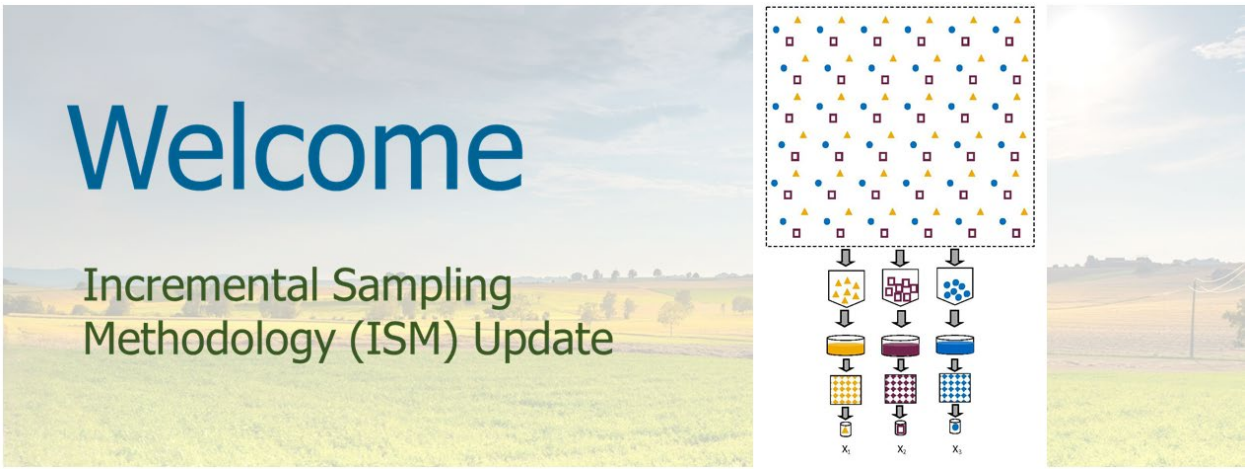
Technical/Regulatory Guidance

Incremental Sampling Methodology (ISM) Update



IMPORTANT: This guidance document was developed for use on the web. Please review the following disclaimers before using the PDF version of the web-based document:

- Web-based formatting may not be optimal when using this document as a PDF.
- Hyperlinks will send users to the web-based document.
- Glossary terms are not hyperlinked.
- The majority of the references are not hyperlinked.

October 2020

Prepared by

The Interstate Technology & Regulatory Council (ITRC)

Executive Summary:

Incremental Sampling Methodology (ISM) is a structured composite sampling and processing protocol that reduces data variability and provides a reasonably unbiased estimate of mean contaminant concentrations in a volume of soil targeted for sampling. ISM provides representative samples of specific soil volumes defined as decision units by collecting numerous increments of soil (typically, 30 to 100 increments) that are combined, processed, and subsampled according to specific protocols.

ISM is increasingly used in the environmental field to sample contaminants in soil. Proponents have found that the sampling density afforded by collecting many increments, together with the disciplined processing and subsampling of the combined increments, in most cases yields more consistent and reproducible results than those obtained by more traditional (that is, discrete) sampling approaches.

Overall, members of the ISM Team have found that ISM provides reliable and reproducible sampling results and leads to less uncertainty better and more defensible decisions than have typically been achieved with many traditional sampling approaches. Such improvements result from the inherent attributes of ISM and the details of its implementation, including a clearer connection between sampling objectives and sampling approach. ISM works to address and overcome the sampling errors associated with soil sampling, integrates attention to detail in planning and field work, and requires attention to quality assurance/quality control measures throughout the sampling effort, not just in the laboratory. ISM also affords an economy of effort and resources. Generally, it would take dozens of discrete samples from any particular area to approach the reliability in an estimate of the mean provided by a well-designed incremental sampling approach. As a result of the advantages and improvements inherent in ISM over traditional methods, ISM is finding increased use in the field, as well as acceptance and endorsement by an increasing number of state and federal regulatory organizations.

How to Use This Guidance

This document is arranged into the following sections:

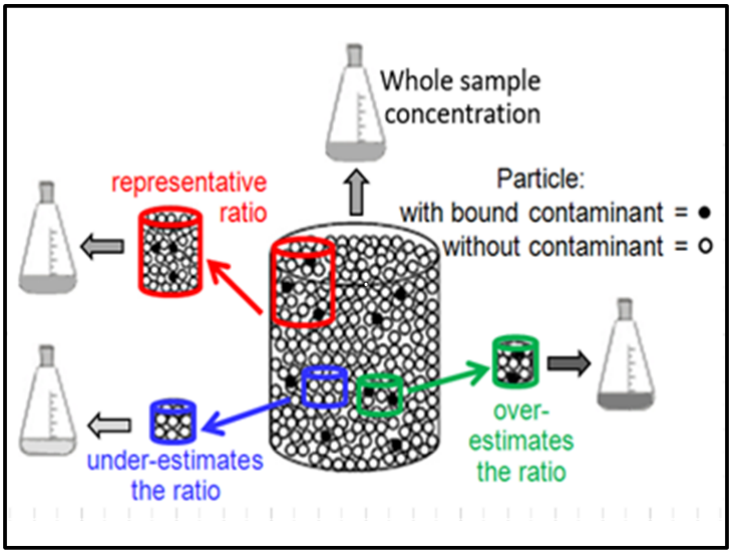
- Section 1 – Introduction. This updated ITRC Incremental Sampling Methodology guidance document (ISM-2) was developed to build on the 2012 version (ISM-1) and to reflect advancements in technology and share case studies that provide insight into the potential applications, benefits, and challenges of the approach. The ISM-2 Update Team also determined that clarification of the ISM-1 guidance was necessary. That clarification statement is now appended at the beginning of the ISM-1 document. ISM-1 continues to be a useful tool for those interested in learning the concepts of ISM.

- Section 2 – Nature of Soil Sampling provides the reader with an understanding of the unique challenges associated with sampling soil for the purpose of obtaining representative contaminant concentrations and how ISM is specifically designed to address these challenges. Contaminants in soil and other particulate media are often distributed unevenly at the scales of interest to decision-makers. Conventional soil sampling approaches that fail to address this heterogeneity can result in over- or underestimating contaminant concentrations, leading to decision errors. Section 2 provides the reader with a detailed understanding of why contaminants are heterogeneously distributed in soil, the consequences of that heterogeneity for soil sampling and decision-making, and how ISM is systematically designed to address this heterogeneity through the use of replicates of multiple increments. This section also discusses how data variability caused by soil heterogeneity is measured, as well as provides a simplified introduction to Gy theory, which is the basis for ISM procedures to increase the representativeness of soil data without breaking the budget.
- Section 3.1 – Systematic Planning and Decision Unit Designation provides a summary of the key aspects of systematic planning and DU design in relation to the collection of soil and sediment samples that have unique applicability or challenges using ISM sample collection. As with any sampling event, characterization must generate data in three dimensions so that data needs are met for a range of technical users who participate in the site investigation process. This means collecting data that inform each step of problem formulation: source area identification, fate and transport, and exposure/risk. Examples illustrate the key aspects of systematic planning and DU design.
- Section 3.2 – Statistical Concepts and Calculations for ISM answers the question, “Why use statistics?” Statistics can be used to answer important questions bearing on decision confidence:

- Is the sampling and analysis design giving accurate information?
    - Are the data good enough to support confident decisions?
    - Are there enough data points to make decisions?

- Section 3.3 – Planning for the Use of ISM Data describes the application of statistical concepts to ISM work plans (WPs) for use in decision-making. Specifically, this section will discuss DQOs steps 5 and 6 as they apply to ISM. The intent is to provide a link between the DQOs, the sample plan, and the data quality evaluation (Section 6).
- Section 3.4 – Cost-Benefit Analysis provides a cost-benefit analysis of ISM sampling relative to more traditional sampling methodologies, including factors such as time to project completion, and shares example case studies to assist in the determination of how ISM may be appropriate for a specific site.
- Section 4 – Field Implementation, Sample Collection, and Processing describes practical methods for collecting consistently sized increments for surface soil, subsurface soil, and sediment from various environments. The sampling method includes guidance on field planning, locating samples, sampling tools, collection and field processing procedures, decontamination, sample handling, and sample shipping.
- Section 5 – Laboratory Sample Processing and Analysis presents current practices and options available to process and subsequently analyze field samples obtained via ISM. Incremental sampling has been successfully implemented at numerous sites for a variety of contaminants, and multiple options are available depending on contaminants such as metals, pesticides, dioxins/furans, semi-volatile organic compounds (SVOCs), VOCs, PCBs, perchlorate, white phosphorus, and energetics (propellants, explosives, and pyrotechnics).
- Section 6 – Data Quality Evaluation demonstrates how to determine if your ISM data are sufficient for your purpose(s) and provides guidance on using appropriate statistical methods for data evaluation and confident decision-making.
- Section 7 – Regulatory Acceptance describes the regulatory environment surrounding the use of ISM, including typical regulatory concerns, problems, or incentives that may apply. This section also presents the state of regulatory acceptance for ISM by comparing its use since 2009 as well as practical guidance for working with or within a regulatory agency to gain consensus for using ISM in investigations, risk assessments, and confirmation sampling.
- Section 8 – ISM for Risk Assessment provides guidance for use of ISM data in risk assessment and risk-based decision-making. Key concepts in this go-to resource include the importance of understanding the nature and extent of contamination when designing EUs, the necessity of ISM replicates for calculating EPCs (and how to calculate EPCs), background comparisons using ISM data, and communicating ISM-based risk assessment results.
- Section 9 – Stakeholder Input discusses applicable stakeholder concerns, points of view, and interaction with the remediation process or other issues discussed.
- Appendix A – Case Study Summaries provides information regarding ISM design, implementation, and assessment methodologies. The case studies presented were selected based on their relevance to the use and application of ISM.
- Appendix B – Statistical Simulations – updated from ISM-1.

ISM Manages Micro-scale Heterogeneity:



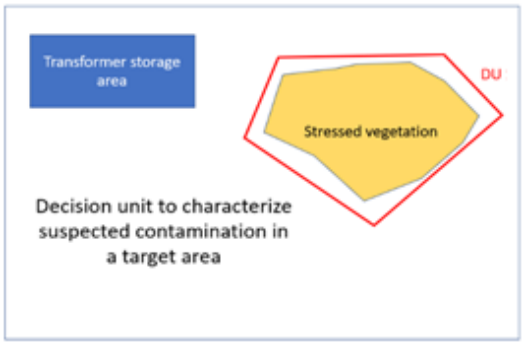
A subsample obtaining a representative ratio of nuggets to cleaner particles depends on subsample mass. A sampling error occurs when the subsample does not mirror the ratio of the field sample, producing a significantly higher or lower concentration result.) See Fig 2-7 in Sect 2.4.1 for more information on how ISM actively manages micro-scale heterogeneity in soil where nuggets contribute to high data variability.

ISM Addresses Grouping Errors:

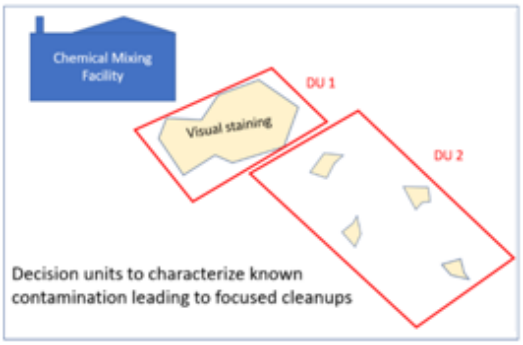
Contaminants grouped by deposition mechanisms [red dots] may bias the sample estimate of the DU mean if sample collection is biased.

ISM helps effectively account for heterogeneity of contaminant distribution across the study area which is prevalent at investigation sites. ISM minimizes decision errors as would occur if individual samples are collected solely from either red or yellow sample locations. See Fig 2.22 in Section 2.6.2.2 for information on how ISM uses systematic random field sample locations for increment collection to avoid this bias.

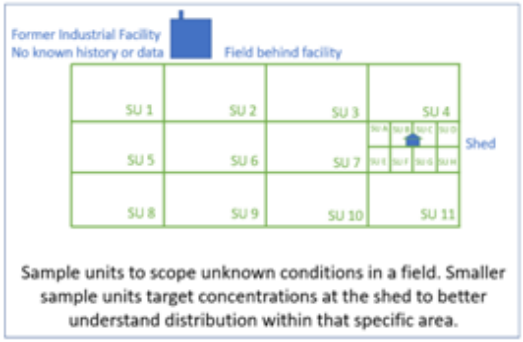
Decision Units Examples:



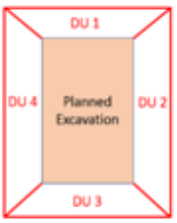
Decision unit to characterize suspected contamination in a target area



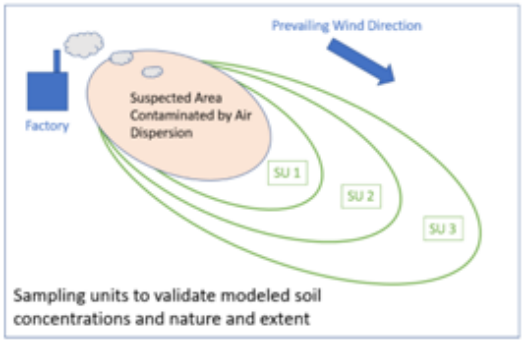
Decision units to characterize known contamination leading to focused cleanups



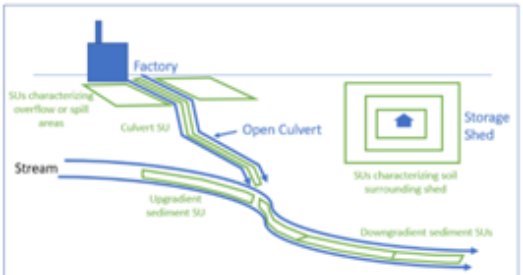
Sample units to scope unknown conditions in a field. Smaller sample units target concentrations at the shed to better understand distribution within that specific area.



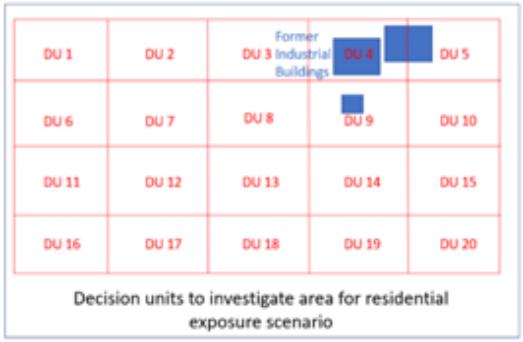
Decision units to characterize soil adjacent to cleanup action to determine if additional soil should be removed



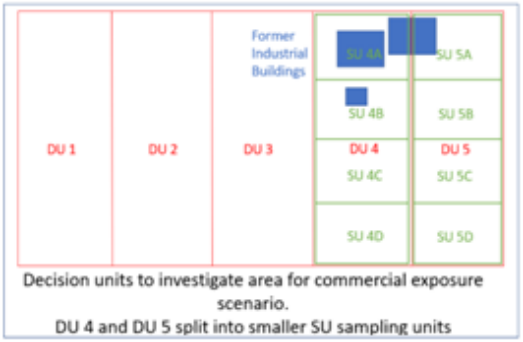
Sampling units to validate modeled soil concentrations and nature and extent



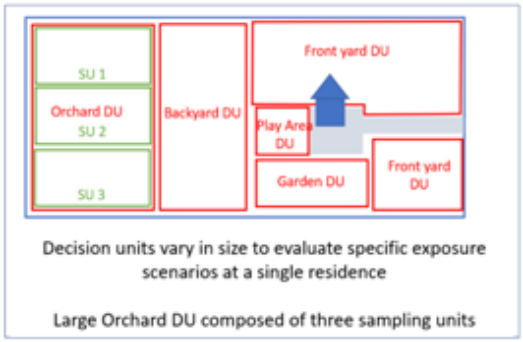
Sample units to define nature and extent, and refine site conceptual model



Decision units to investigate area for residential exposure scenario



Decision units to investigate area for commercial exposure scenario.
DU 4 and DU 5 split into smaller SU sampling units



Decision units vary in size to evaluate specific exposure scenarios at a single residence

Large Orchard DU composed of three sampling units

The determination of sizes and arrangement of DUs or SUs is always project-specific and dependent upon the conceptual site model (CSM) and end use of the data. Figure 1-1 from Sect 1 demonstrates just a few simplified examples of the

diversity of DU and SU sizes and uses. Examples for designing DUs are presented in Section 3.1.

ISM Tools Available:

Checklist: Prior to Field Implementation:

Before proceeding with field implementation, the following checklist can be used to help ensure that adequate planning unique to ISM has occurred. Have you:

Topic:	**Question**:
DQOs	identified the problem/decision to be made?
	identified objectives/goals?
	ensured that the inputs I have designed meet these objectives?
	made sure that I understand the CSM (see Section 3.1.2 through 3.1.4)?
DUs/SUs (see Section 3.1.5.1 and Section 3.1.5.2)	evaluated size and depth relative to decisions to be made?
	evaluated if the DU design can be adjusted to serve all data needs (statistics, nature and extent, background, risk assessment, and so on)?
	mapped out shape and considered site's physical constraints?
Increments and Samples	identified the appropriate number of increments per DU/SU (see Section 2.5.2 and Section 3.1.5.5)
	considered if pilot or early replicates for unassessed areas are warranted prior to full-scale implementation?
	designed an increment sampling path that ensures unbiased locating of increments (see Section 3.1.5.4)?
	considered resulting ISM sample size relative to scale of decision-making?
Laboratory	confirmed the processing steps to be conducted by the laboratory (see Section 3.1.5.3 and Section 5)?
	confirmed the sample analysis procedures to be used by the laboratory (see Section 3.1.5.3 and Section 5)?
	considered QC samples and frequency needed for data validation and statistical analysis (see Section 3.2)?

After reviewing Sect 4, you should be able to complete the following checklist to ensure proper field preparation and implantation. Have you:

Question:
allocated appropriately trained staff to execute the task?
identified site-specific means of marking out a DU calculated the increment size based on the sample design total volume?
identified the appropriate tool(s) to obtain each increment from the required depth?
assessed if a mass reduction technique will be needed during increment collection?
requested/obtained sample containers specific to ISM (large volume, special considerations for VOCs for ISM)?
considered the added resources for sample storage (added ice and coolers)?
communicated to the laboratory the required processing and QC requirements for the chain of custody

These checklists are available in Section 4's Regulatory Context Field Implementation, Sample Collection, and Preparation section.

Cost Comparison Calculator:

Costs for ISM may be much lower than the costs of discrete sampling, and the data used to make decisions are less likely to mis-characterize remaining contamination and lower the potential for remobilization and continued excavation and soil treatment.

Using the cost comparison calculator presented in Section 3.4.1.1 the field and laboratory costs can be estimated as in Table 3-10.

Parameter	...Using ISM	...Using Discrete Sampling Methodology
Step 1: Determine number of samples for laboratory analysis		
# of DUs	5	n/a
Number of Investigative Samples	5	40
Number of Replicate Samples per DU	2	20%
Number of Increments per DU	30	n/a
Number of Increments for Project	450	n/a
Total Number of Samples for Laboratory Analysis	15	48
Parameter	...Using ISM	...Using Discrete Sampling Methodology
Step 2: Determine field labor, equipment, and sample shipping costs		
Est. Labor Hours per Sample (in whole hours)	1	0.5
Total Labor Hours (for two people, in whole hours)	15	24
Hourly Rate	$100	$100
Subtotal Labor Costs	$1,500	$2,400
Sampling Equipment	$500	$500
Days of Sampling	2	3
Subtotal Sampling Equipment Costs	$1,000	$1,500
Sample Shipping	1	5
	15	10
	$50	$50
Subtotal Shipping Costs	$750	$500
Total Estimated Field Labor, Equipment, and Sample Shipping Costs	$3,250	$4,400
Parameter	...Using ISM	...Using Discrete Sampling Methodology
Step 3: Determine laboratory and data validation costs		
Sample Processing	$100	$0
Number of Samples to be Processed	15	48
Total Processing Cost	$1,500	$0
Per Sample Analytical Costs	$250	$250
Total Analytical Costs	$3,750	$12,000
Subtotal Sample Preparation and Analytical Costs	$5,250	$12,000
Laboratory Data Package (Stage 4)	10%	10%
	2	5
	100%	100%
Subtotal Additional Laboratory Costs for Laboratory Data Packages	$500	$1,250
Data Validation	$10	$10
	$20	$20
Subtotal Validation Costs	$170	$530
Total Estimated Laboratory and Data Validation Costs	$5,920	$13,780
Estimated Total Costs	**$12,420**	**$22,580**
Notes:		
All costs are in U.S. dollars		
Assumptions:		
The time taken to collect sample will depend on a number of factors. The values used here are based on the experience. The rate of 0.5 hours for a team of two people to collect 30 increments in triplicate for 1/4 acre site is based on case studies reviewed to date. Calculation provided in this table is adjusted for when more than 30 increments are required.		

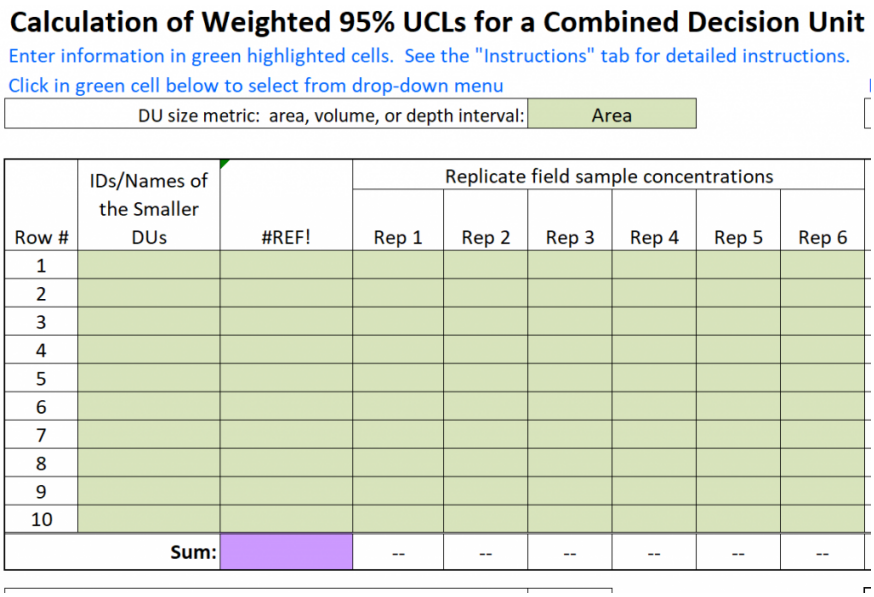
ISM 95% UCL Calculator:

Equations for ISM Student's-t and Chebyshev 95% UCLs were programmed into an Excel spreadsheet file that is in Section

3.2.4.1 (ISM 95% UCL Calculator):

The ISM 95% UCL calculator built by the first ISM Team has been updated since then with an improved modeling procedure. This calculator has several benefits:

- The user only has to enter the results of three to six replicate field samples, as well as the number of increments per sample (not shown on the image below).
- The ISM 95% UCL spreadsheet calculates both the Student's-t and Chebyshev 95% UCLs.
- The spreadsheet recommends which 95% UCL should be used.



Calculation of Weighted 95% UCLs for a Combined Decision Unit

Enter information in green highlighted cells. See the "Instructions" tab for detailed instructions.

Click in green cell below to select from drop-down menu

DU size metric: area, volume, or depth interval:	Area

Row #	IDs/Names of the Smaller DUs	#REF!	Replicate field sample concentrations					
			Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Rep 6
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
	Sum:		--	--	--	--	--	--

Introduction

Incremental Sampling Methodology Update

Incremental Sampling Methodology (ISM) is a statistically supported technique for assessing the mean contaminant concentration in soil, sediment, and other environmental media. Environmental investigators have demonstrated that the methodology can be a useful tool to accurately represent site conditions when applied to bulk particulate materials such as soil, sediment, or waste. For clarity, this document uses the term soil, understanding that other solid particulate media can also be assessed using this methodology. This guidance, developed by the Interstate Technology Regulatory Council (ITRC), reflects recent advances in the technique, shares case studies that provide insight into potential applications, and helps those considering ISM for their sites determine whether to pursue an ISM approach, and if so, how to implement it. It serves as an update to the ISM guidance released in 2012 and represents the current state of ISM that should be adapted for use at contaminated and potentially contaminated sites.

1.1 Sampling to Accurately Inform Environmental Cleanup

The assessment of contaminated sites and areas of potential concern relies on the collection and analysis of samples of environmental media. Analytical data from samples provide a *representation* of site conditions, showing what contaminants are present and at what concentrations. Decision-makers can then utilize this data to assess risk, determine the need for cleanup action, and guide remediation to protect human health and the environment. As part of this process, representative sample data are either compared with decision criteria, such as numerical regulatory cleanup standards for a given contaminant or are used in a baseline risk assessment. An example of such decision criteria would be a state-specific cleanup goal for a contaminant in soil.

To ensure effective decision-making, it is critical that the data compared with the criteria or used to calculate risks accurately represent the conditions at a site in terms of a mean contaminant concentration that meets the data quality objectives (DQOs). Both risk-based screening levels and cleanup levels apply to the mean concentration of a contaminant for a designated volume/mass of soil.

In particular, the characterization of contamination in soil and sediment poses unique challenges due to the heterogeneity of contaminant distribution in these media. As discussed in Section 2, *heterogeneity* refers to the condition where components of a matrix differ from each other between individual soil particles and in the nonuniform distribution of the constituents across the site. Heterogeneity is present at all scales and results in spatial variation of contaminant concentrations.

Historically, soil sampling approaches have relied on the collection of a number of discrete samples at randomly selected or biased locations and depths, providing a patchwork of data across a given site. Composite sampling, one such historically used method, is the collection of multiple increments (MIs) mixed together and then analyzed to provide a sample that is intended to represent a larger area or multiple depths at one location. While this can provide an improvement over discrete sampling methods, the basis behind the selection of a number of increments, total bulk sample mass, and processing and subsampling requirements can limit the reliability of resulting data.

Moreover, contaminant data can vary substantially between locations, even in near proximity to each other. Attempts to confirm data quality through the use of field duplicate samples may not provide the desired level of confidence that data are reproducible.

ISM addresses the problems associated with the reproducibility and reliability of discrete sample data by presenting a structured approach to historically termed “composite” sampling modified with the collection of many increments (30 to 100) and standardized processing. The methodology has gained acceptance and attention due to its robust statistical basis, the reproducibility of data gained from ISM, and the growth in understanding of its applications in the regulatory and environmental practice communities.

1.2 What Is ISM?

ISM is a structured sampling and processing protocol that reduces data variability and provides an estimate of mean

contaminant concentrations in a defined volume of soil. Essentially, the methodology provides representative contaminant concentrations in samples from specific soil volumes, defined as decision units (DUs) or sampling units (SUs), by collecting numerous increments of soil that are combined, processed, and subsampled for laboratory analysis according to specific field and laboratory protocols. Analytical data for small subsamples of soil tested by a laboratory represent the mean for the volume of soil included in the subsample. ISM includes specific planning, sampling, and processing aspects that have historically been underutilized or not performed, including:

- establishing DU or SU boundaries that define the scale of decision-making and/or scale of data to spatially structure the assessment
- verifying the laboratory's ability to perform ISM processing and analysis
- reviewing ISM procedures with the sampling team
- reviewing field sampling protocols for ISM
    - collecting a sufficient quantity of increments for each DU (typically 30 to 100)
    - collecting equal mass per each increment
    - collecting increments throughout DUs in an unbiased manner
    - compensating for media heterogeneity by collecting a sufficient mass of sample (typically 1 to 2 kg dry weight)
- verifying laboratory sample processing techniques
    - for non-volatile compounds, air-drying the entire field sample and sieving and/or milling
    - subsampling from the entire processed sample

The goal of ISM is to obtain and analyze a sample that contains analytes in the same proportions as the soil throughout a given DU. To achieve this goal, many increments are obtained from a single DU. Another key element of the sampling approach is that replicate samples (typically three, each with the same coverage and representativeness) are collected from the same DU. Traditional discrete soil sampling approaches often employ the collection and analysis of duplicate samples to obtain a percentage of the samples and evaluate data reproducibility. However, in practice, the results for soil duplicate samples, which are intended to represent the concentration at a single location, can often vary widely, even when proper soil sampling procedures are followed. Due to the challenges of addressing unsatisfactory duplicate soil sample results, duplicates are not even collected for some traditional applications. With ISM, replicates are typically collected for each DU, providing much greater insight into the reproducibility of the data, as well as the degree of heterogeneity.

The statistical basis and terminology for ISM is attributed to Pierre Gy, who used the term *incremental sampling* to describe an improved method for obtaining representative samples from heterogeneous media. His work, based in the mining industry, was prompted by the highly variable distribution of minerals within a rock formation. Misleading data were often obtained from a limited number of samples and yielded poor decisions about mine development. His analysis led to a statistically defensible method for sample collection and sample processing (Gy 1953, 1988) The U.S. Environmental Protection Agency (USEPA) documented improvements in site soil data that could be achieved by applying Gy's sampling theory (USEPA 1999a). USEPA continued to refine the description of ISM procedures in SW846 Method 8330B, published in 2006 (USEPA 2006c), and in 2019, issued "Incremental Sampling Methodology (ISM) at Polychlorinated Biphenyl (PCB) Cleanup Sites" (USEPA 2019). The U.S. Army Corps of Engineers (USACE) issued guidance in 2009 for munitions sampling in surface soil using ISM (USACE 2009). As more environmental practitioners grappling with how to obtain better representative soil data became aware of this theory, ISM began to be used in multiple states. It gained early acceptance in Hawaii and Alaska, which issued guidance in 2008 and 2009, respectively, as well as in Ohio and Michigan, and has now been used at sites in many states across the U.S. (see Section 7).

1.3 Why Should ISM Be Considered for My Site?

The overall efficiency and effectiveness of environmental investigation and cleanup actions can be assessed based on three criteria: time, cost, and reliability of results. ISM investigation methods can improve each of these criteria. Environmental cleanup activities in the U.S. has cost approximately $2 billion annually over the past two decades, just for Superfund sites alone (*Washington Post*, "Taxpayer dollars fund most oversight and cleanup costs at Superfund sites,") (Anderson 2017). Cleanup costs at privately funded industrial facilities are typically not shared with the public but are likely many times higher than costs incurred under the Superfund program. Even modest improvements in the accuracy of delineating target areas for cleanup can have significant cost savings. On the other hand, if areas that pose significant risk are not properly identified, contamination may be allowed to persist, potentially migrate further, and impact human and environmental receptors, posing risks that are not acceptable. One key goal of ISM is to reduce the potential for such "decision errors,"

such as when deciding that cleanup is needed when it is not or missing cleanup that is warranted.

Significant attention to quality assurance (QA) and quality control (QC) for laboratory analytical methods has resulted in accurate, reproducible methods for analyzing small subsamples, or *aliquots*, of media such as soil. While laboratory methods continue to improve, much of the technological advances in recent decades have resulted in lowering detection limits or in detecting more contaminants, rather than in large improvements in accuracy. However, the limitations for obtaining a representative value for a soil concentration are driven far more by sample collection and processing protocols than by subsample analysis techniques. It is estimated that the errors introduced during sample collection and subsampling are at least 10 times greater than errors resulting from laboratory analytical inaccuracies. To improve decision-making for site cleanup, there is much more opportunity for gains in the sample collection and processing realm than in analytical improvements.

Because of the fundamental heterogeneity of contaminant distribution in soil and sediment at most sites, the collection of a small number of traditional discrete or composite samples can underestimate or overestimate the actual mean concentration. ISM, through preparation of a sample by combining increments (small masses) of soil from a large number of points within a well-defined area/volume, provides a more robust picture of the mean contaminant concentration for that volume. This allows a site to be reliably characterized with the collection of a relatively small number of analytical samples. Overall site characterization/remediation costs – and in many cases, time to project completion – can be reduced while providing data that are more statistically robust and reliable than discrete sampling approaches. To the extent that ISM can reduce the need for multiple mobilizations, "step out" the removal of contaminated media, or reduce unnecessary cleanup costs, even greater cost reductions may be recognized.

The potential costs and time efficiencies that can be attained using ISM are demonstrated in the case studies summarized in Section 3.4 and Appendix A. To help evaluate whether ISM may be appropriate for a given site, many key questions should be addressed:

- How variable is contaminant distribution at the site? The more variable the distribution, the greater the potential to obtain a better estimate of the mean using ISM.
- Is the goal of sampling to obtain an accurate mean or upper bound estimate of the mean concentration for a given DU? If so, ISM will yield superior results, but it will not provide information about the distribution of contaminants within a DU. If that is the goal, the DU may need to be defined on a smaller scale.
- How easy is it to collect samples? Because more increments are collected from a given DU with ISM than with discrete or traditional composite approaches, the ease with which samples can be collected affects the time and cost necessary to conduct soil assessment. For subsurface samples at deeper depths in formations that require slower drilling techniques, more time may be required to collect the necessary increments for a given DU. The user will have to balance potential higher costs in deeper DUs with the much lower confidence in decision-making using other methods.
- Is ISM restricted to specific contaminants? Initially, ISM was used for munitions and metals, and procedures have since been developed for all categories of contaminants. Specific refinements to procedures are required for some contaminants, such as not drying or milling samples in the laboratory for volatile organic compounds (VOCs), or based on state or territory regulatory agencies' guidance/requirements.
- How quickly are analytical results available? Does the additional laboratory processing time for ISM samples meet the schedule? Laboratory processing of ISM samples requires additional time (typically one to three days) compared to traditional approaches.
- Do site conditions limit the ability to safely or efficiently collect a number of increments from a given DU (for example, are there hazards such as subsurface utilities throughout the target area or a developed, paved property where access to underlying soil is restricted)?

The determination of sizes and arrangement of DUs or SUs is always project-specific and dependent upon the conceptual site model (CSM) and end use of the data. Figure 1-1 demonstrates just a few simplified examples of the diversity of DU and SU sizes and uses (see Section 3.1.6).

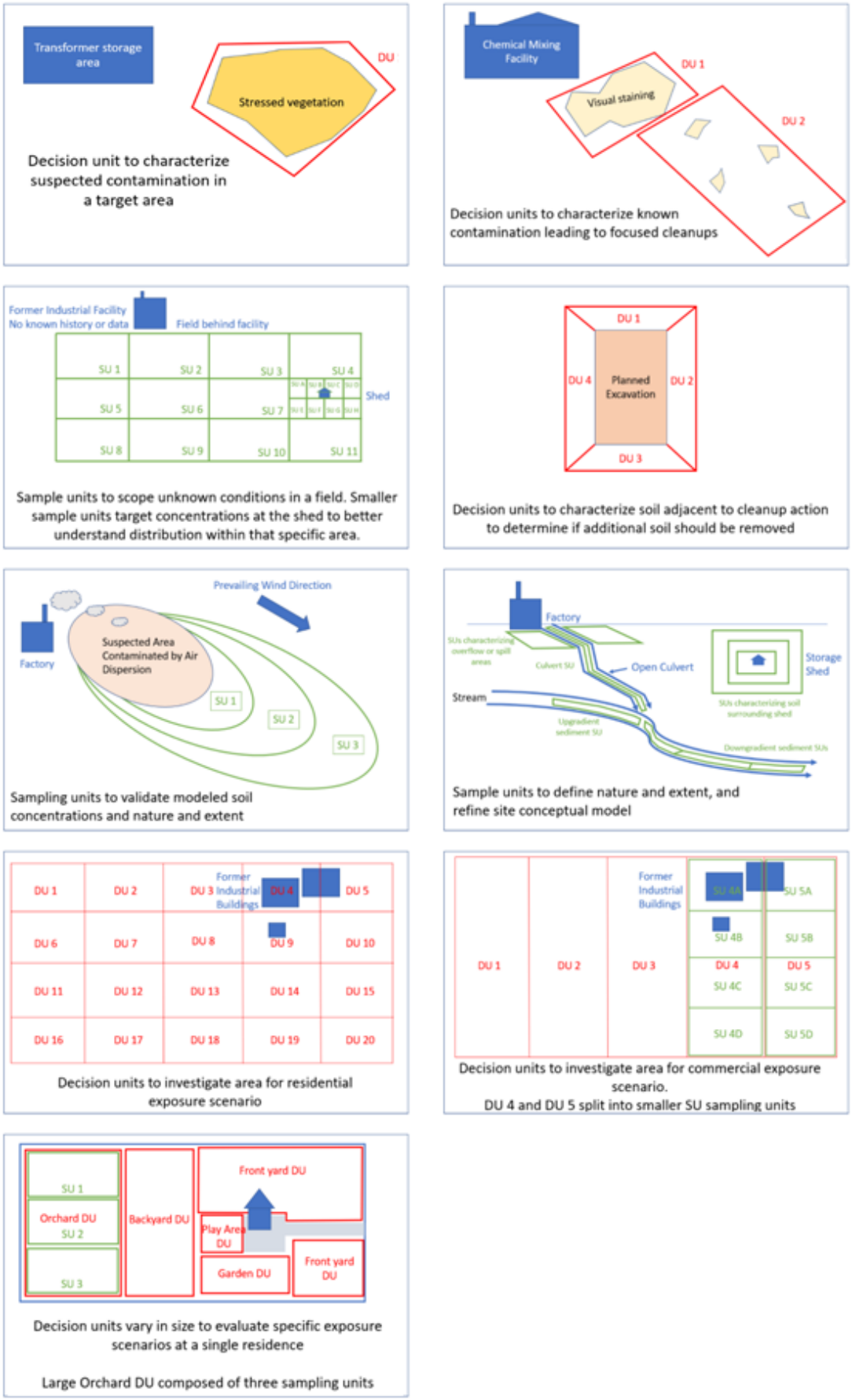


Figure 1-1. Examples of diversity of DU and SU sizes and uses.

Source: Jason Brodersen, Tetra Tech, Inc., 2020. Used with permission.

1.4 How Is ISM Implemented for Environmental Sampling?

The application of ISM in environmental assessment has expanded from initial use for munitions and metals to a much broader range of contaminants and a wide range of site applications. While this guidance is not intended to be a detailed recipe for every application, it does provide key considerations to help decision-makers determine whether ISM makes sense for their site and, if so, how to tailor ISM to the conditions at a given site.

To begin, practitioners need to be familiar with several key concepts and terms associated with ISM. While many of these terms are described in more detail in subsequent sections, a general understanding is important for even starting a consideration of whether to use ISM:

- The CSM serves to define the relationship between contaminant sources, contaminated media, and receptors through consideration of potential or actual migration and exposure pathways. It presents the current understanding of the site, helps to identify data gaps, and helps to focus data collection efforts.
- A DU is the smallest volume of soil for which a decision will be made. It is the area and depth of soil from which mean analyte concentrations are obtained and is representative of a specifically defined population. A DU can be as small as a storm drain outlet or as large as a commercial parcel or agricultural field.
- An SU is a subdivision of a DU, or the volume of soil from which increments are collected to determine an estimate of the mean concentration. A DU may consist of one or more SUs. The use of SUs is project-specific and not always necessary.
- An increment is a specified volume of soil collected from a specific point within a DU. Multiple increments (typically 30 or more) are collected from a specified DU and combined into a single sample.
- The combined increments are referred to as a **multiple-increment**, **incremental soil sample**, or **ISM sample** that is representative of the mean contaminant level with in specified DU.
- An exposure point concentration (EPC) is a conservative estimate of the mean chemical concentration in an environmental medium. The EPC can be determined for an entire site or for an individual exposure unit (EU) – a location within a site where exposure potential may vary from the overall site, often where regular exposure is currently or anticipated to occur.
- A 95% upper confidence limit (95% UCL) of the arithmetic mean is a calculated value to ensure that the mean concentration is not underestimated, so a user can be 95% confident that the true mean (average) concentration of the population is below this value. Three or more ISM samples or replicates are required to calculate a 95% UCL, provided the appropriate statistical methods are used to calculate the 95% UCL.

The process for ISM is illustrated in Figure 1-2, which shows key steps through the three phases: planning, implementation, and data analysis and decision-making.

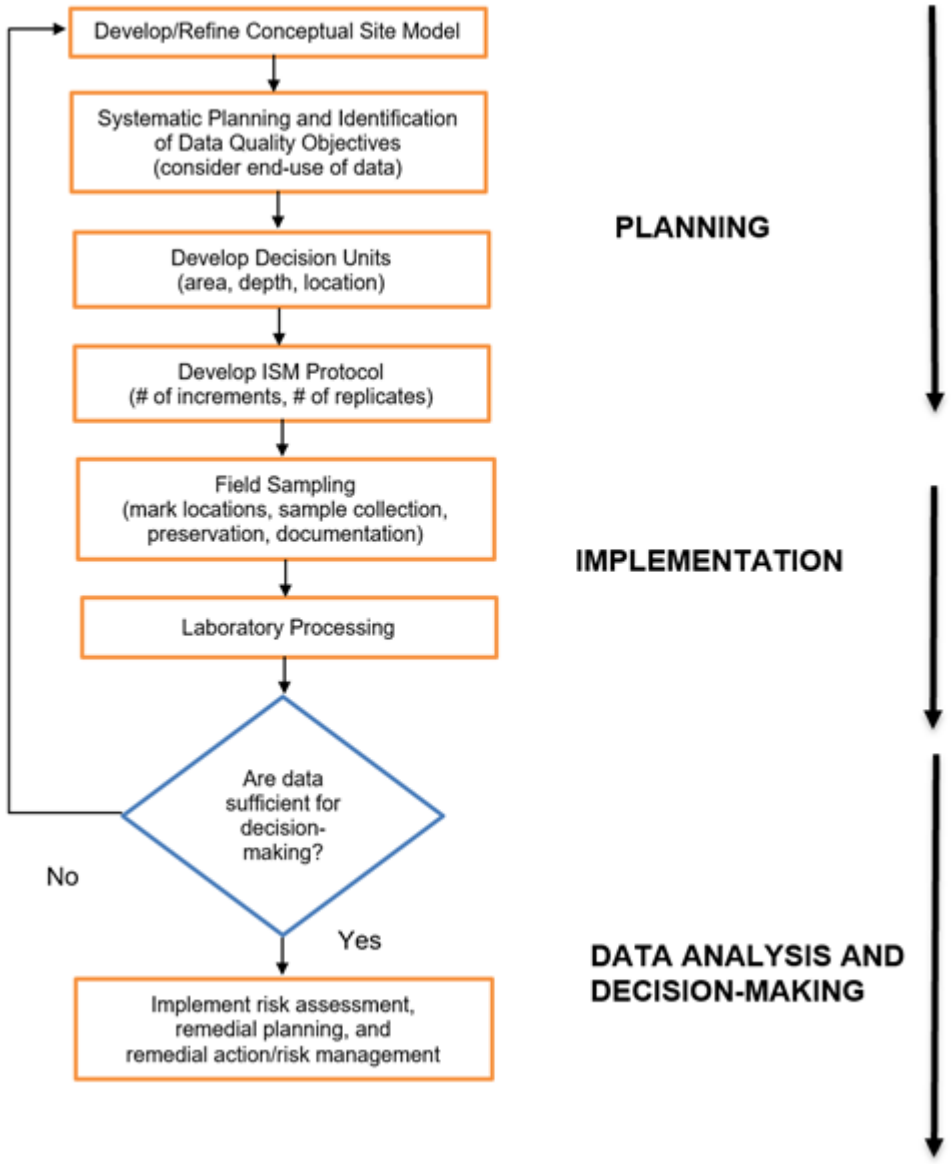


Figure 1-2. ISM flow chart.
Source: ITRC ISM Team, 2012.

Note that this guidance is intended to provide a high-level overview to help chart a path to plan and implement ISM – it is not a step-by-step detailed sampling and analysis plan or recipe that can be applied to any given site. The assessment of contaminated sites is a complex process that requires advanced planning, an understanding of contaminant fate and transport, and clear objectives for final site disposition. Any soil sampling approach, including ISM, is much more likely to be successful if there is an accurate CSM to guide planning (critical for knowing where to look and what to look for), the DQOs are clearly defined (so we understand what data is needed and how we will use it), and DUs are enumerated and geographically delineated prior to sample collection (to develop the scope and cost of assessment and reduce the potential need for multiple field mobilizations). A multi-disciplinary team experienced in planning, field implementation, laboratory analysis, risk assessment, and ultimate site cleanup is critical to that success.

1.5 Document Content – How to Use This Guidance

In 2009, ITRC created a working team to develop guidance for practitioners interested in utilizing ISM. The ISM Team published its first guidance document in 2012 to provide an overview of the concepts and principles of the methodology, emphasize the importance of clear objectives, and provide a basis for adapting ISM to meet project- and site-specific objectives. This 2020 ISM Guidance Document builds on the 2012 version to reflect advancements in technology and to share case studies that provide insight into potential applications, benefits, and challenges of the approach.

This document is arranged into the following sections:

- Section 2 – Nature of Soil Sampling provides the reader with an understanding of the unique challenges associated with sampling soil for the purpose of obtaining representative contaminant concentrations and how ISM is specifically designed to address these challenges. Contaminants in soil and other particulate media are often distributed unevenly at the scales of interest to decision-makers. Conventional soil sampling approaches that fail to address this heterogeneity can result in over- or underestimating contaminant concentrations, leading to decision errors. Section 2 provides the reader with a detailed understanding of why contaminants are heterogeneously distributed in soil, the consequences of that heterogeneity for soil sampling and decision-making, and how ISM is systematically designed to address this heterogeneity through the use of replicates of MIs. This section also discusses how data variability caused by soil heterogeneity is measured, as well as provides a simplified introduction to Gy theory, which is the basis for ISM procedures to increase the representativeness of soil data without breaking the budget.
- Section 3.1 – Systematic Planning and Decision Unit Designation provides a summary of the key aspects of systematic planning and DU design in relation to the collection of soil and sediment samples that have unique applicability or challenges using ISM sample collection. As with any sampling event, characterization must generate data in three dimensions so that data needs are met for a range of technical users who participate in the site investigation process. This means collecting data that inform each step of problem formulation: source area identification, fate and transport, and exposure/risk. Examples illustrate the key aspects of systematic planning and DU design.
- Section 3.2 – Statistical Concepts and Calculations for ISM answers the question, "Why use statistics?" Statistics can be used to answer important questions bearing on decision confidence:
    - Is the sampling and analysis design giving accurate information?
    - Are the data good enough to support confident decisions?
    - Are there enough data points to make decisions?

- Section 3.3 – Planning for the Use of ISM Data describes the application of statistical concepts to ISM work plans (WPs) for use in decision-making. Specifically, this section will discuss DQOs steps 5 and 6 as they apply to ISM. The intent is to provide a link between the DQOs, the sample plan, and the data quality evaluation (Section 6).
- Section 3.4 – Cost-Benefit Analysis provides a cost-benefit analysis of ISM sampling relative to more traditional sampling methodologies, including factors such as time to project completion, and shares example case studies to assist in the determination of how ISM may be appropriate for a specific site.
- Section 4 – Field Implementation, Sample Collection, and Processing describes practical methods for collecting consistently sized increments for surface soil, subsurface soil, and sediment from various environments. The sampling method includes guidance on field planning, locating samples, sampling tools, collection and field processing procedures, decontamination, sample handling, and sample shipping.
- Section 5 – Laboratory Sample Processing and Analysis presents current practices and options available to process and subsequently analyze field samples obtained via ISM. Incremental sampling has been successfully implemented at numerous sites for a variety of contaminants, and multiple options are available depending on contaminants such as metals, pesticides, dioxins/furans, semi-volatile organic compounds (SVOCs), VOCs, PCBs, perchlorate, white phosphorus, and energetics (propellants, explosives, and pyrotechnics).
- Section 6 – Data Quality Evaluation demonstrates how to determine if your ISM data are sufficient for your purpose(s) and provides guidance on using appropriate statistical methods for data evaluation and confident decision-making.
- Section 7 – Regulatory Acceptance describes the regulatory environment surrounding the use of ISM, including typical regulatory concerns, problems, or incentives that may apply. This section also presents the state of regulatory acceptance for ISM by comparing its use since 2009 as well as practical guidance for working with or within a regulatory agency to gain consensus for using ISM in investigations, risk assessments, and confirmation sampling.
- Section 8 – ISM for Risk Assessment provides guidance for use of ISM data in risk assessment and risk-based decision-making. Key concepts in this go-to resource include the importance of understanding the nature and extent of contamination when designing EUs, the necessity of ISM replicates for calculating EPCs (and how to calculate EPCs), background comparisons using ISM data, and communicating ISM-based risk assessment results.
- Section 9 – Stakeholder Input discusses applicable stakeholder concerns, points of view, and interaction with the

remediation process or other issues discussed.

- Appendix A – Case Study Summaries provides information regarding ISM design, implementation, and assessment methodologies. The case studies presented were selected based on their relevance to the use and application of ISM.
- Appendix B – Statistical Simulations
- Appendix C – Team Contacts
- Appendix D – Glossary
- Appendix E – Acronyms
- Appendix F – References

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.

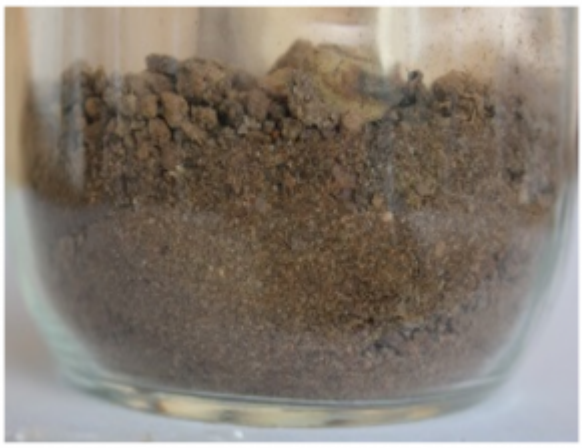
The proper implementation of ISM practices will improve the reliability of soil data, allowing users to avoid many of the problems stemming from decision errors. ISM practitioners have found that ISM improves the reliability of soil data and avoids many problems stemming from the decision errors caused 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.

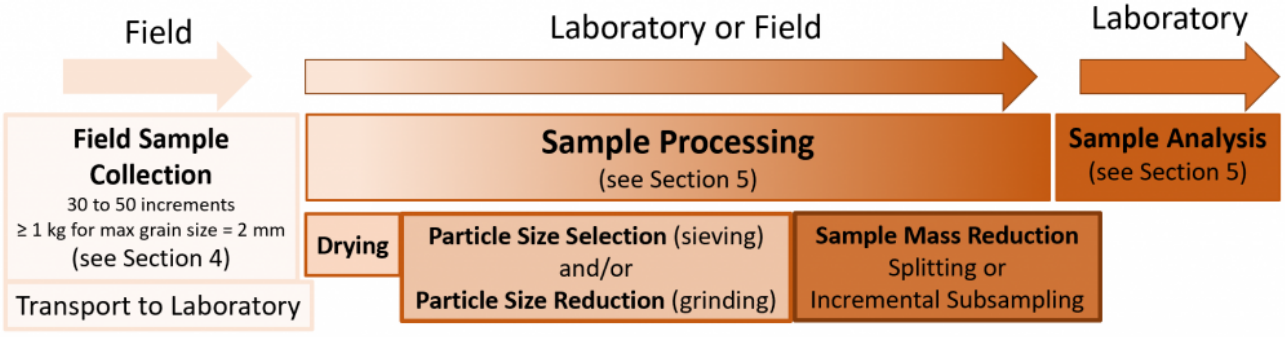


Figure 2-1. Three-stage ISM sampling process. Some processing steps may be optional, depending on soil characteristics, analytes of interest, and project objectives.
Source: ITRC ISM Update Team, 2020.

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).

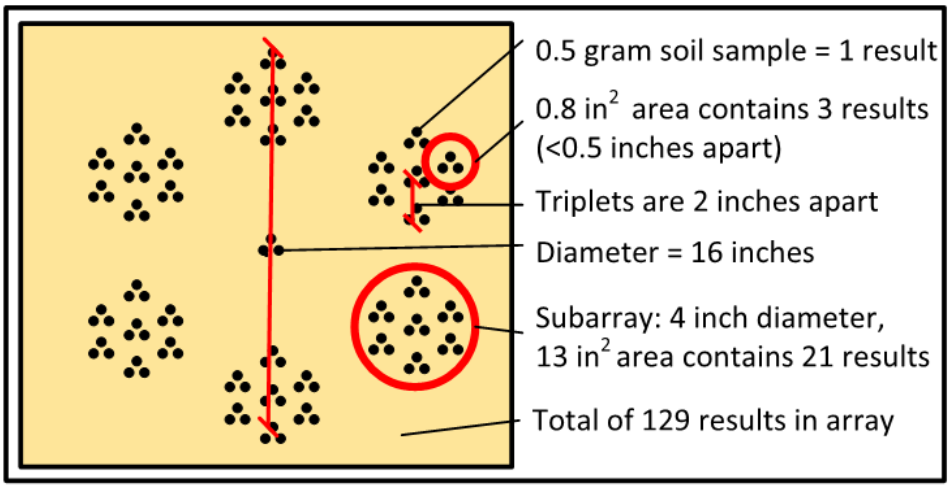


Figure 2-3a. Array pattern for 0.5-g soil samples, which encompass a circular area of 1.4 ft2 and contain 129 individual samples.
Source: Developed by Deana Crumbling, 2014. Used with permission.

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).

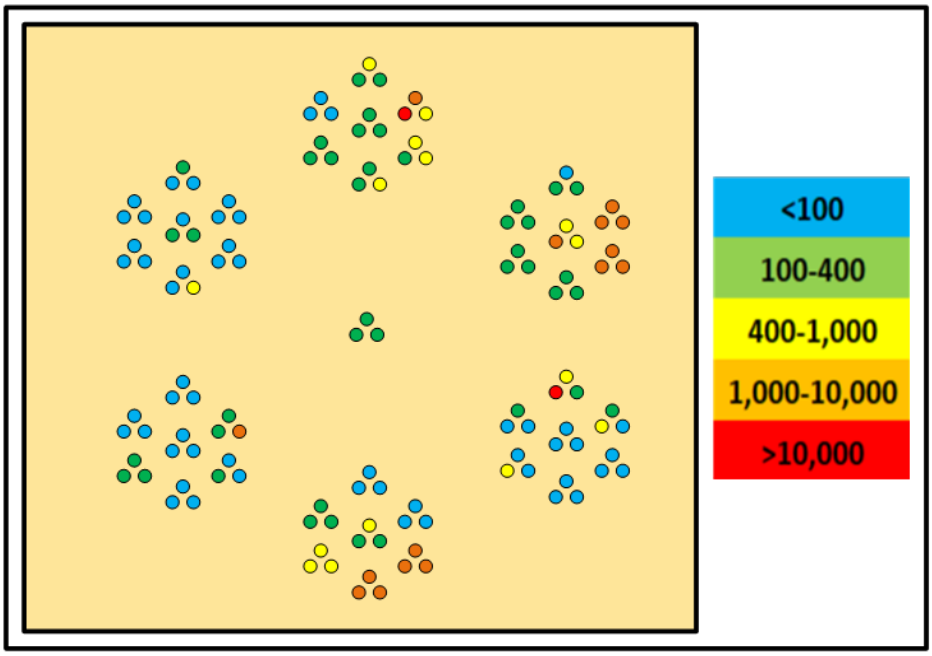


Figure 2-3b. Becker triplets with color-coded sample concentrations. Markedly different concentrations can occur in soil scoops only 1 cm apart. From this dataset, basing a decision for a 400-mg/kg lead threshold on a single random result would underestimate contamination.
Source: Developed by Deana 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 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*.

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).

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).

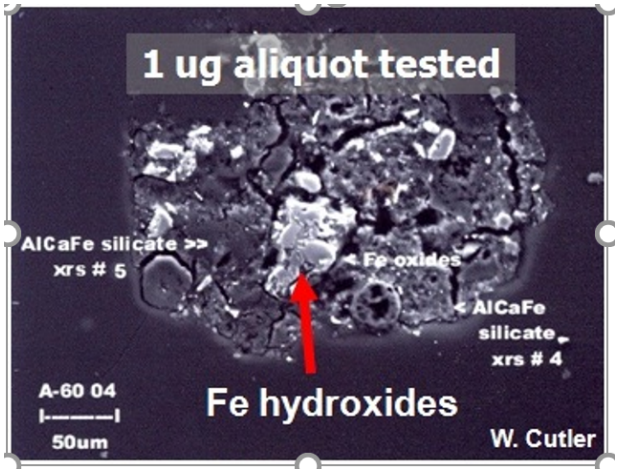


Figure 2-4. Microscopic iron hydroxide grains coated with arsenic, which appears as a light-colored deposit (red arrow). Arsenic-laden iron mineral nuggets are scattered among "clean" particles made of silicate minerals.
Source: Roger Brewer, HDOH, 2012. Used with permission.

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.

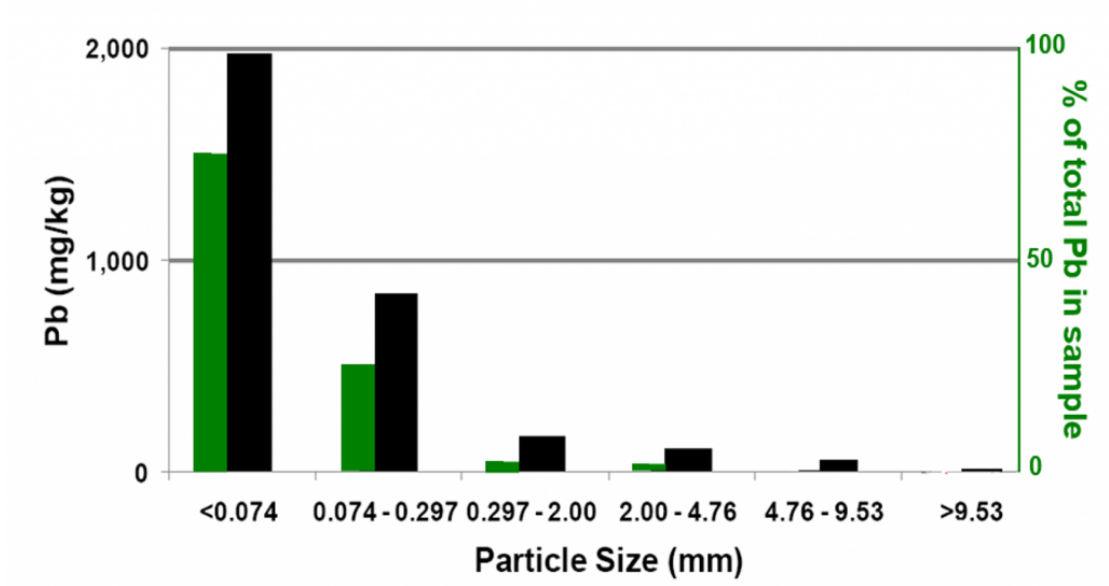


Figure 2-5. The relationship between particle size and Pb concentration for a soil sample collected from a firing range. The black bars show the Pb concentration (in mg/kg) for each of six particle size fractions. The green bars show the percent of the total sample Pb contained by each particle size 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
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.

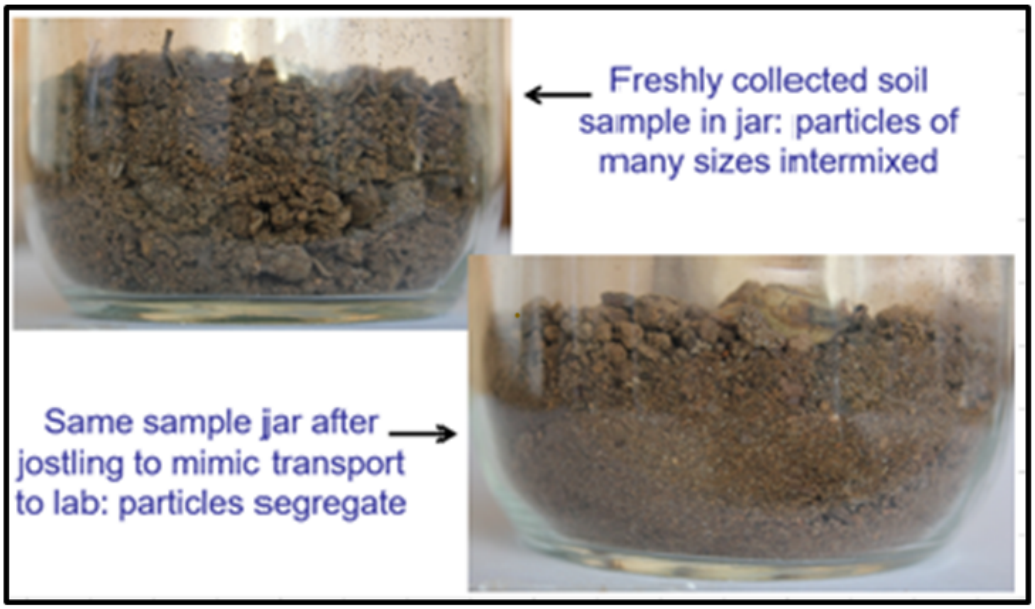


Figure 2-6. Jostling of a sample jar causing smaller particles to settle downward while larger particles remain at the top. Stirring to mix is ineffectual to redistribute particles evenly and often makes segregation worse.
Source: Deana Crumbling, 2013. Used with permission.

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).

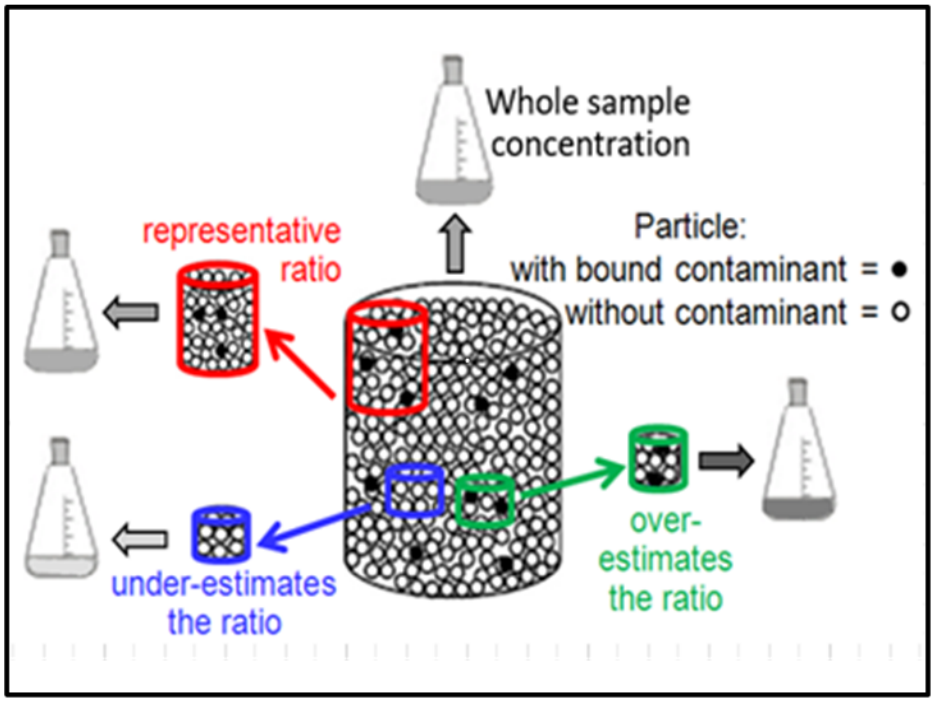


Figure 2-7. A subsample obtaining a representative ratio of nuggets to cleaner particles depends on subsample mass. A sampling error occurs when the subsample does not mirror the ratio of the field sample, producing a significantly higher or lower concentration result.
Source: Developed by Deana Crumbling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

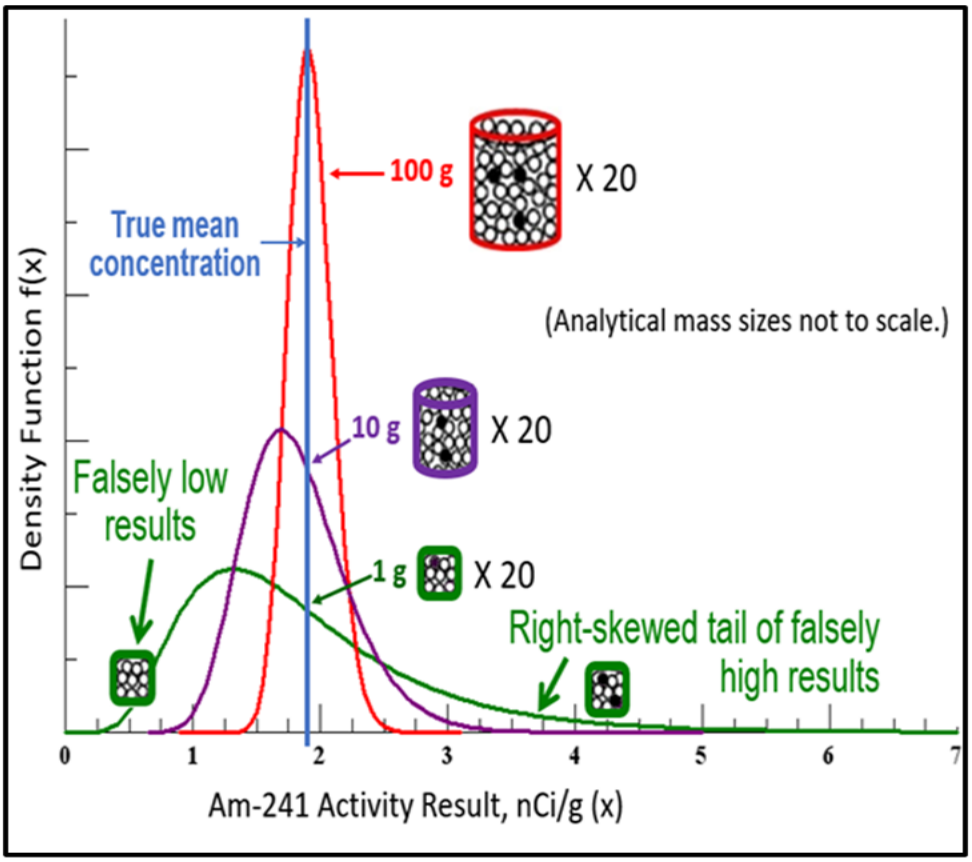


Figure 2-8. Analytical mass influencing the degree of data variability and statistical distribution. Data from 100-g analytical subsample masses (tall, narrow, red peak) were very precise with a normal data distribution. Data from the 10-g analytical mass (purple peak) were less precise, as shown by 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

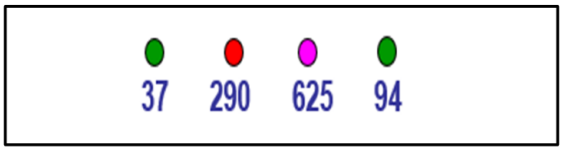


Figure 2-9. Arsenic concentrations for in situ XRF readings spaced about 2 ft apart in a residential yard.
Source: Unpublished data from Deana Crumbling, 2007. Used with permission.

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.

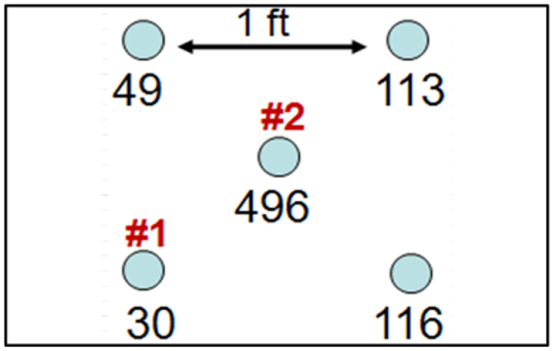


Figure 2-10. Uranium concentrations for five co-located samples located 1 ft or less apart. A decision might be based on location number.
Source: USEPA and DOE, 2008.

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).

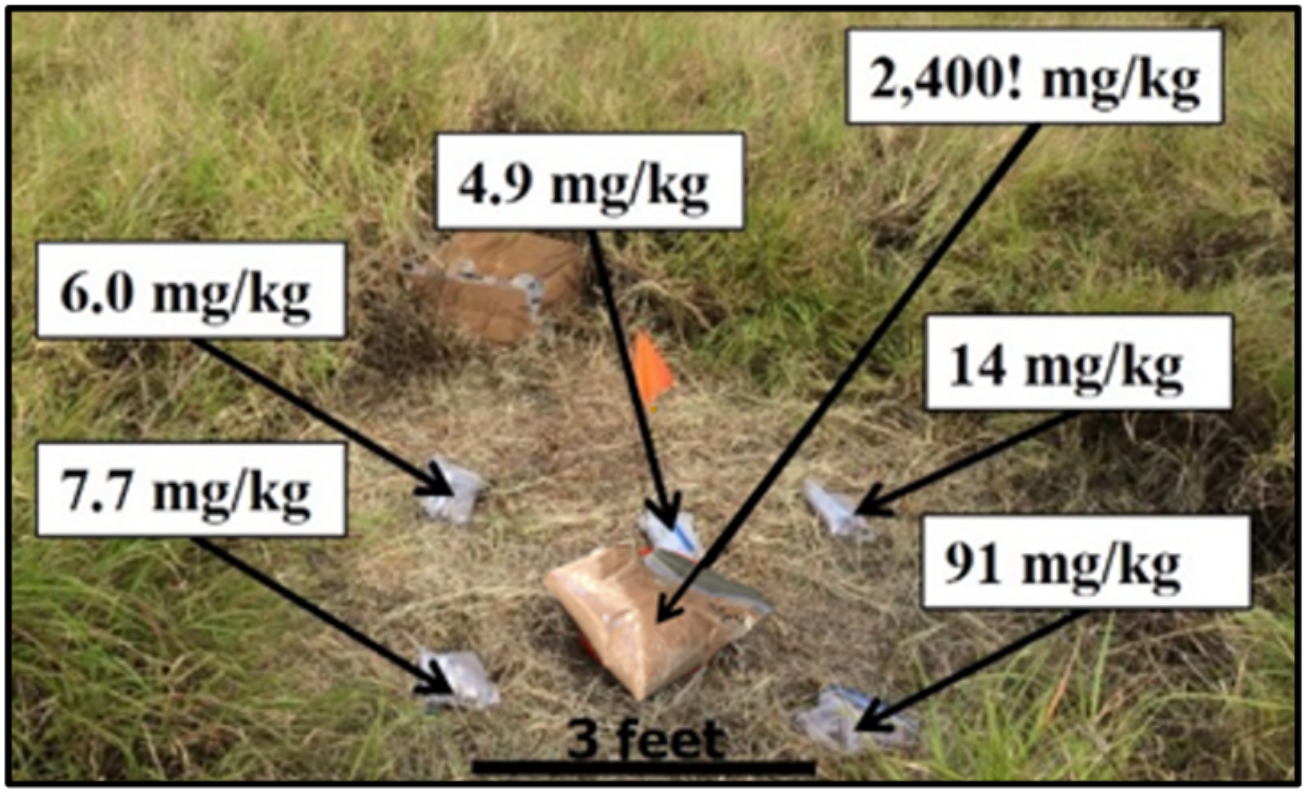


Figure 2-11. PCB concentrations resulting from five small samples and one large sample arranged in an area of

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.

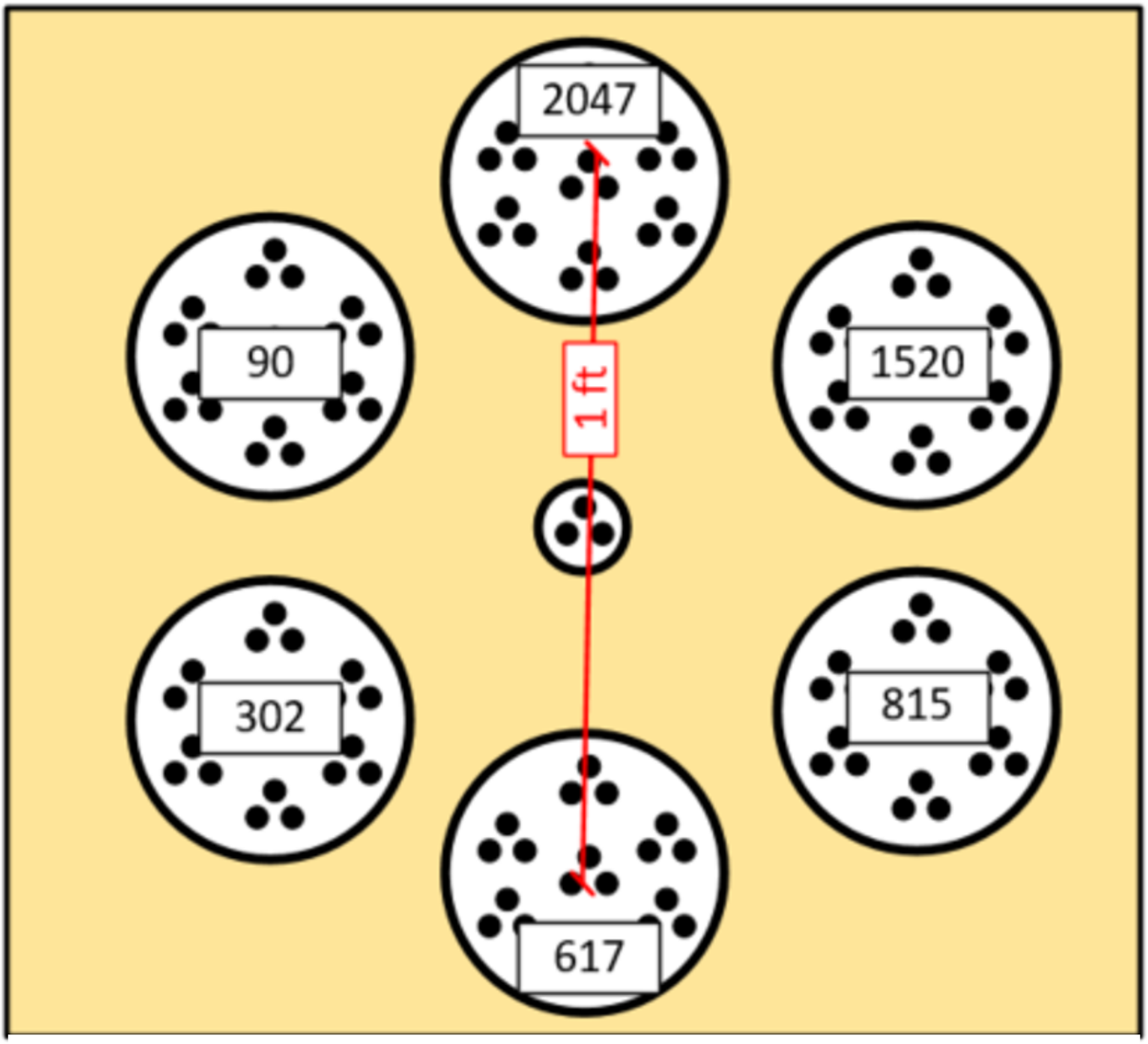


Figure 2-12. Mean Pb concentrations grouped into six 10.5-g subarrays. Results range from 90 to 2,047 mg/kg at a scale of less than 1 ft apart.
Source: Deana Crumbling, 2019, Used with permission.

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.

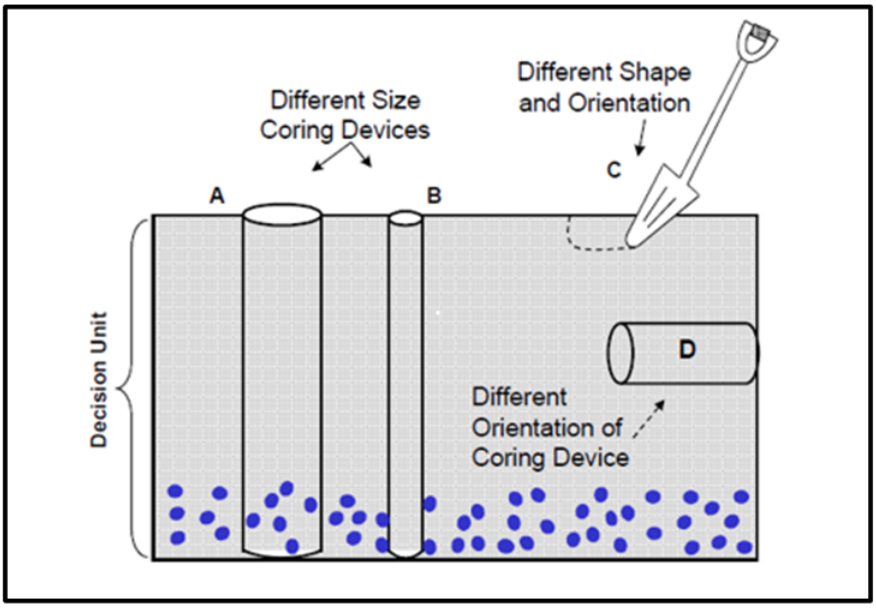


Figure 2-14. Sample support. Only device A has the correct sample support (size, shape, and orientation) to extract a representative sample, which includes the blue dots from the DU.
Source: USEPA, 2002b.

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.

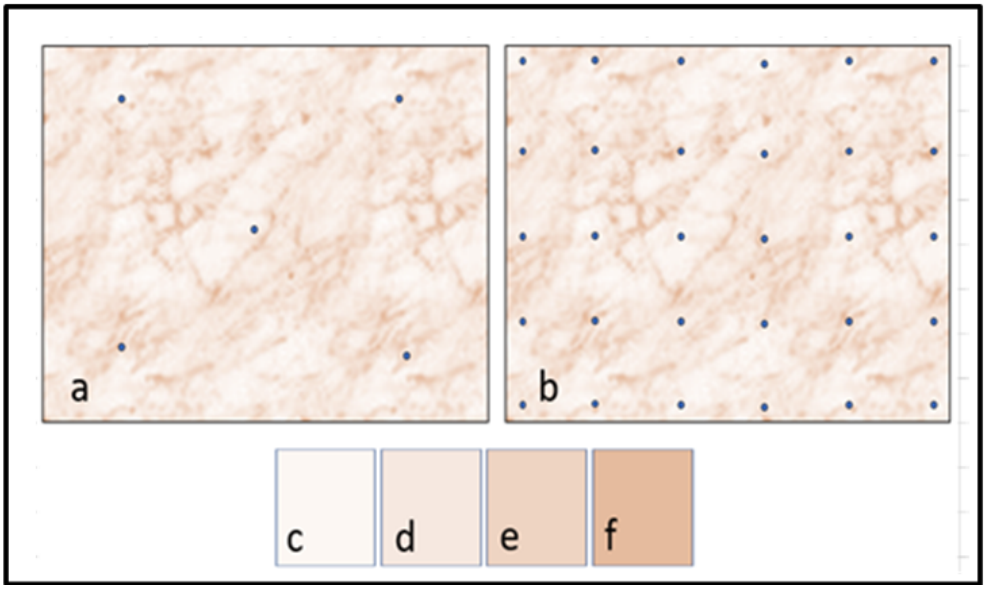


Figure 2-15. Contaminant concentration. Panels "a" and "b" show a heterogeneous DU sampled by five and 30 increments, respectively. Panels "c" through "f" depict concentration levels that range from very low (near white) to high (darkest tan).
Source: Deana Crumbling, 2019. Used with permission.

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.

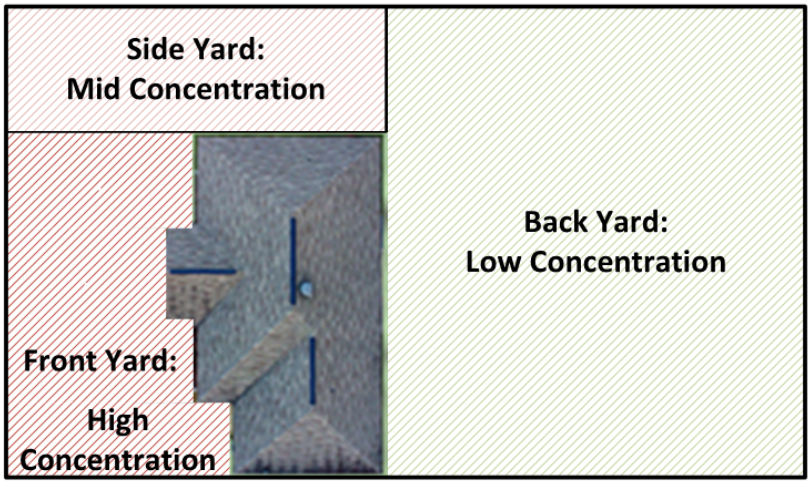


Figure 2-16. Schematic of a DU encompassing a residential property with distinctly different concentrations in the front, side, and back yards. Darker shades represent higher concentrations.
Source: ITRC ISM Update Team, 2020.

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.

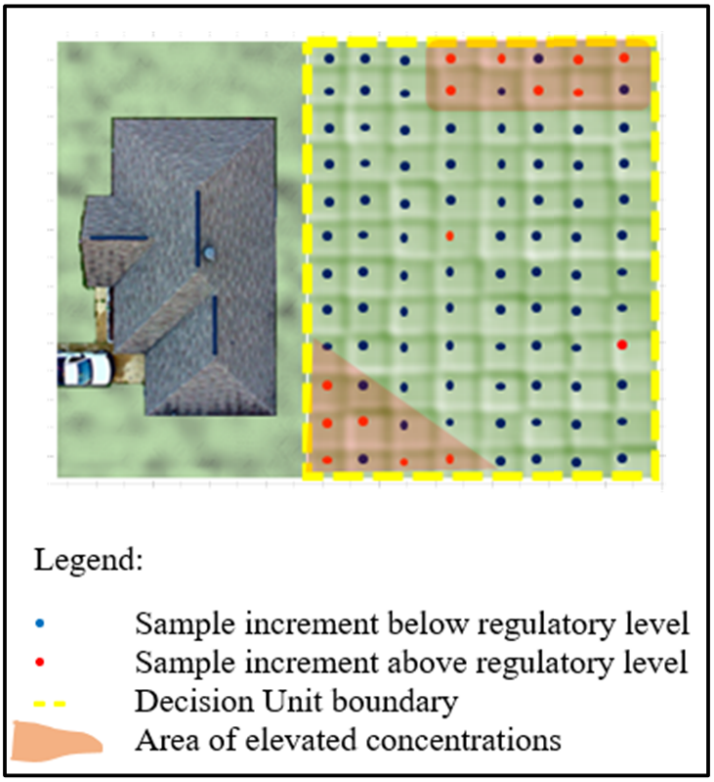


Figure 2-17. Information from real-time in situ analyses. The yellow areas depict zones.
Source: Deana Crumbling, 2020. Used with permission.

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

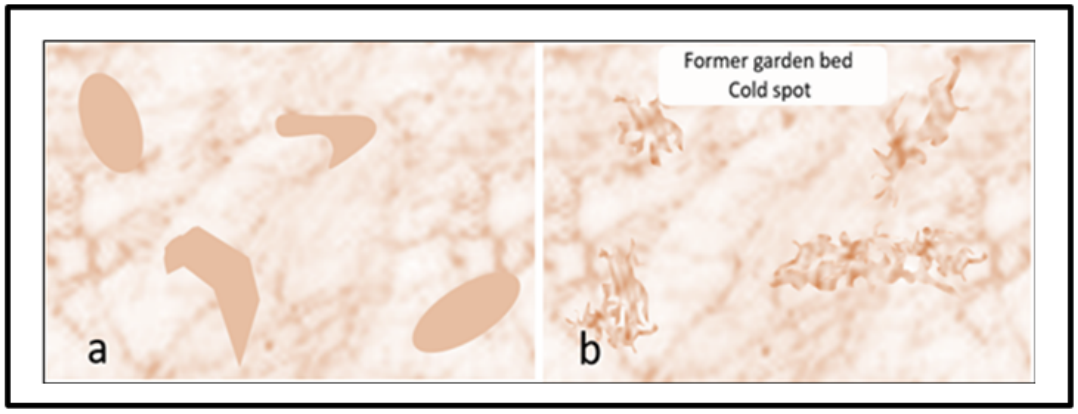


Figure 2-18. Concentration misconceptions. Panel "a" depicts a common misunderstanding about areas of elevated concentration, showing homogenous groupings with defined margins. Panel "b" presents a more realistic representation of such areas along with adjacent areas of lower concentration.
Source: Deana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

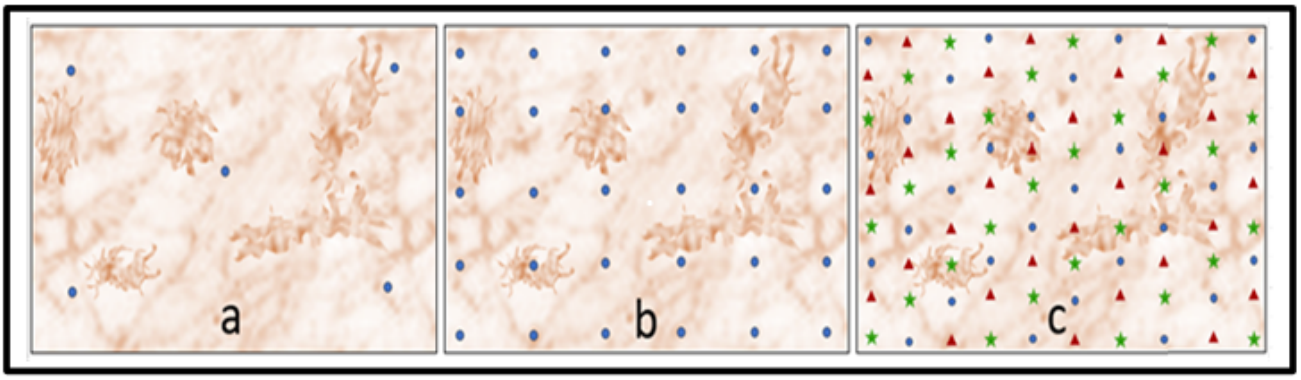


Figure 2-19. Possible design outcomes. Panel "a" illustrates a pattern of five increments, panel "b" shows a pattern of 30 increments, and panel "c" shows 90 increments (three independent replicate samples). The more increments, the better the chance the ISM sample will pick up high concentration areas. 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).

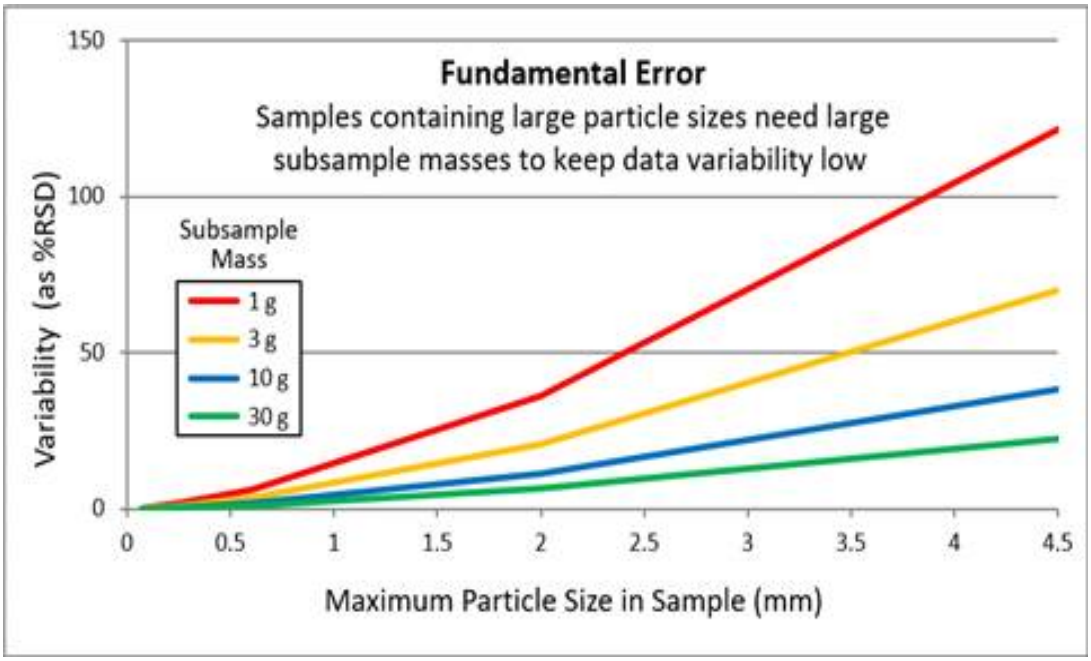


Figure 2-20. Fundamental error as illustrated from equations from ASTM.
Source: Mark Bruce, Eurofins. Based on information provided in ASTM, 2012.

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.

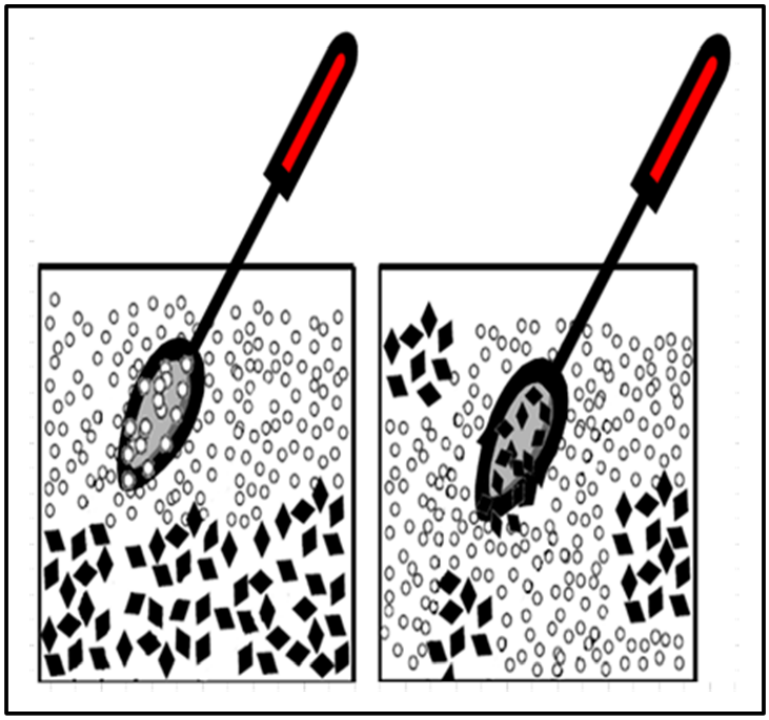


Figure 2-21. Segregation. Non-representative subsampling following the incomplete stirring of a segregated sample creates grouping error.
Source: Developed by Deana Crumbling from data by HDOH, 2015. Used with permission.

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

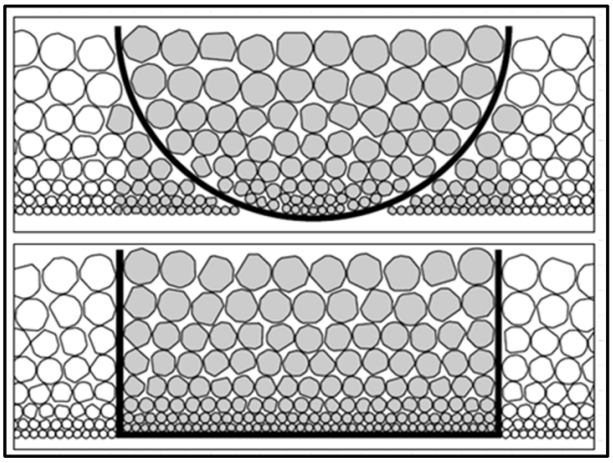


Figure 2-23. Delimitation error. A scoop with a round bottom (top panel) discriminates against the fine

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.

Systematic Planning, Statistical Analyses, and Costs

The following sections describe the process and considerations involved for DU planning, including statistical analysis and cost estimates.

3.1 Systematic Planning and DU Design

Section 3.1.1 through Section 3.1.5 provides a summary of the key aspects of systematic planning and DU design in relation to the collection of soil and sediment samples. Section 3.1.6 provides three examples that illustrate the application of these key aspects of planning for different types of environmental problems:

- an agricultural field, settling pond, and drainage swale being assessed using screening criteria (Example 1)
- a former agricultural field being converted to residential use (Example 2)
- a former industrial facility that is to be redeveloped, with human health and ecological endpoints (Example 3)

3.1.1 Overview

As with any such sampling event, characterization must generate data in three dimensions so that data needs are met for a range of technical users who participate in the site investigation process. This means collecting data to inform each step of an environmental investigation, including source area identification, evaluation of contaminant fate and transport, and assessment of potential exposure and risks.

ISM-related planning guidance is consistent with USEPA's DQO guidance (USEPA 2002c), primarily utilizing the first four steps of the DQO process: problem formulation (step 1), identify study goals (step 2), identify information inputs (step 3), and define study boundaries (step 4). Some material associated with step 5 (develop the analytic approach) and step 7 (develop the plan for obtaining data) is also provided in relation to the examples used to demonstrate systematic planning and DU design with ISM. Use of ISM in conjunction with statistical hypothesis tests, which is the focus of DQO steps 5 and 6, is taken up in Example 2 and addressed in detail in Section 3.2.

Note that implementation of ISM does not require that the DQO process be followed. However, to ensure that data obtained during environmental investigations are adequate for their intended purposes, it is strongly recommended that data collection activities be planned and developed through a systematic planning process (SPP) with end users, including the development and consideration of a CSM. Establishing clear objectives at the beginning of the investigation is crucial to efficient and effective site characterization. As described in this section, the outcome of good systematic planning is well-thought-out DUs and SUs (see Section 2.5.1.2), whose locations and dimensions produce information to support all the investigation questions.

USACE's technical project planning (TPP) process (USACE 1998) provides another example of a systematic planning framework that can readily be used with ISM. More recently, the DQO process has been integrated in the manual for implementation of the Uniform Federal Policy for Quality Assurance Project Plans (USEPA 2005b).A list of guidance documents that can be used with ISM in addition to the DQO, TPP, and uniform federal policy (UFP)-QAPP guidance describing planning processes is provided below:

- "Technical Guidance Manual for the Implementation of the Hawai'i State Contingency Plan" (HDOH 2017b)
- "Improving Environmental Site Remediation Through Performance-Based Environmental Management" (ITRC 2007a)
- "Best Management Practices: Use of Systematic Project Planning Under a Triad Approach for Site Assessment and Cleanup" (USEPA 2010)
- "Triad Implementation Guide" (ITRC 2007b)

3.1.2 DQO step 1: problem formulation (what is the problem, and what decisions need to be made?)

The basic aspects of problem formulation, including establishing the project planning team and developing the CSM, are not unique to investigations employing ISM. When a project team is considering inclusion of ISM as a project tool in the first step of systematic planning, they should consider how ISM might fit into answering related study questions during development

of the CSM by calling upon the expertise of a multi-disciplined team (including, for example, chemistry, data analysis, engineering, field sampling, geology, QA, modeling, regulatory, risk assessment, soil science, statistics, and toxicology experts). Important aspects of a CSM for supporting systematic planning are described below, with particular emphasis on applying a CSM for ISM. Additional information on developing and applying CSMs is provided in Section 3 of ITRC's human health risk assessment guidance (ITRC 2015). USACE Engineer Manual 200-1-12, "Conceptual Site Models," also provides examples of several different types of CSMs and their use(USACE 2012).

CSMs are essential elements of the SPP for complex environmental problems. They serve to conceptualize the relationships among contaminant sources, environmental fate and transport mechanisms, potential exposure media, and the potential routes of exposure to these media for human and ecological receptors. The structured organization of information to form a CSM creates both a summary of the current understanding of site conditions and anticipates future conditions in a manner that can help the project team identify data gaps in the information needed to make project decisions. These gaps are the basis of study goals (sampling objectives) in the next step of the planning process. In this sense, certain study goals can be thought of as hypotheses related to the CSM, and so achieving sampling objectives serves the purpose of increasing confidence in the CSM.

In addition to a narrative description of these component relationships, a CSM commonly includes pictorial and/or graphical representations of the components of the exposure pathway analysis. Figure 3-1 provides an example of a pictorial CSM depicting a contaminated source area and the pathways through which that contamination travels to reach human health and ecological receptors. Another CSM is rendered in graphical format in Figure 3-2. The pictorial representation of a CSM, such as the example in Figure 3-1, can be particularly useful in risk communication with stakeholders. The graphical depiction, shown in Figure 3-2, is particularly useful for framing study goals and the related inputs and boundaries for supporting site-specific risk assessment. See additional examples in Figures 3-14a and 3-14b.

The CSM may also include summaries of available environmental data, information pertaining to source terms such as listings and quantities of process chemicals, and preliminary transport modeling results.

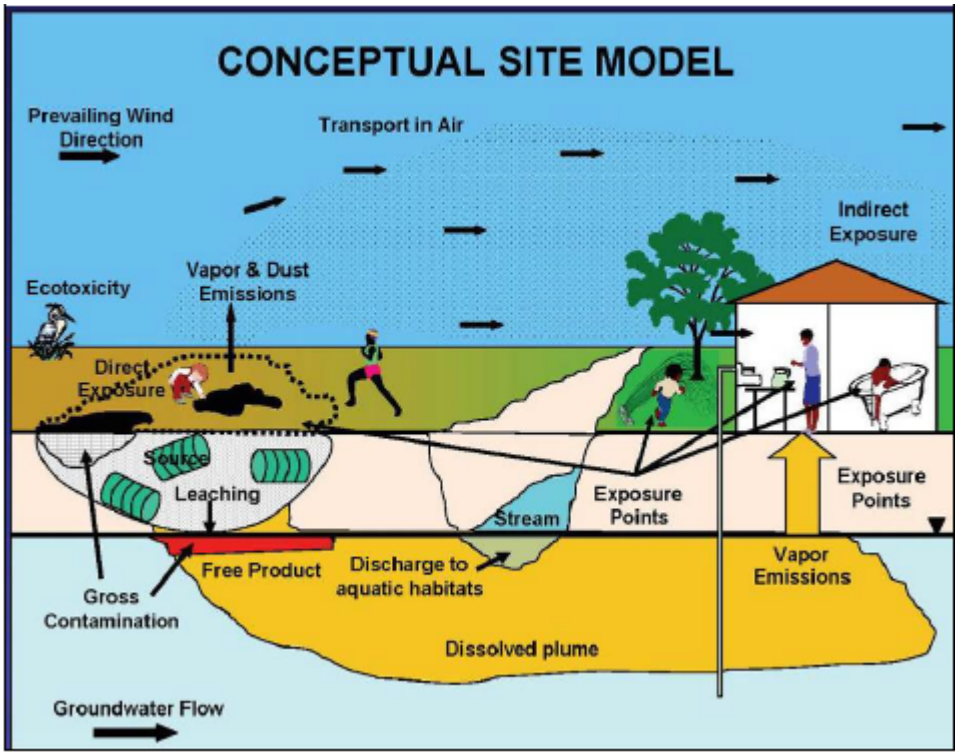


Figure 3-1. Pictorial CSM example.
Source: ITRC ISM-1 Team, 2012.

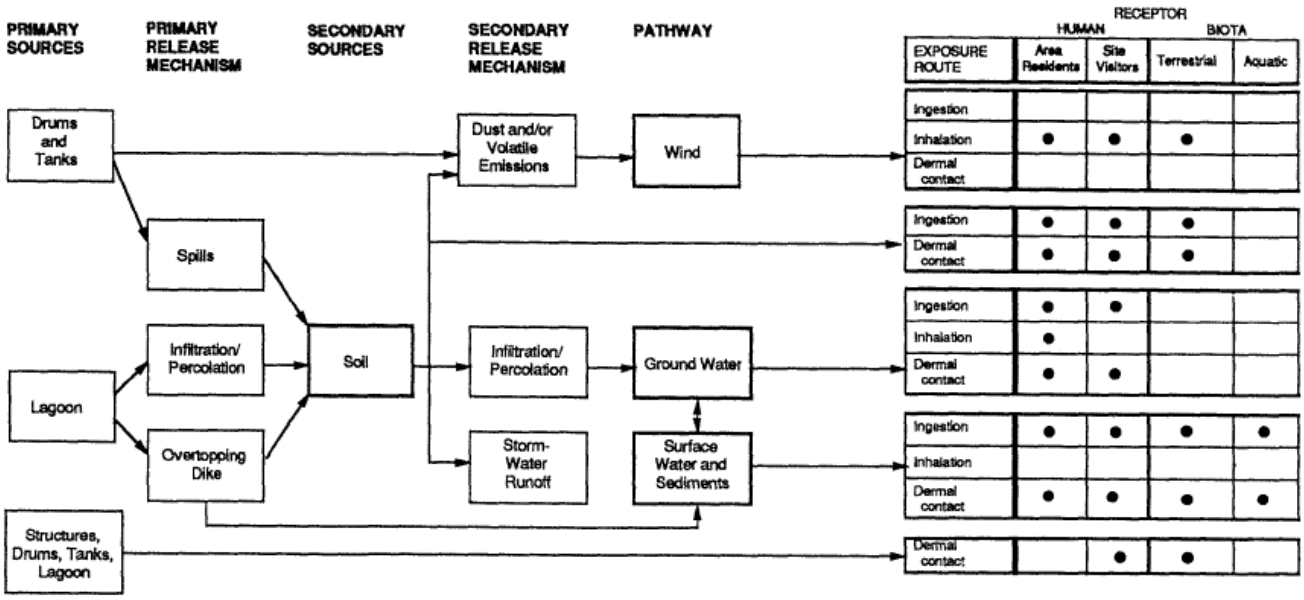


Figure 3-2. Graphical CSM example.
Source: USEPA, 2011.

Decisions about the general sampling approach for a project are crucial in ensuring the data will be adequate to meet project objectives. Project planners may elect to employ ISM or traditional discrete sampling, or even a combination of the two, although these data are not directly comparable and cannot be easily combined (see Section 6.2.5 and Section 6.2.6). The optimum approach depends on the CSM, the sampling objectives, and how the data are to be used. In addition to the technical considerations associated with selecting among different sampling approaches, project planners must also consider relevant regulatory requirements, as well as resource, time, and budget limitations.

Investigation objectives can change as projects progress, which means new information and objectives must continually be reconsidered over the course of the project. Consideration of dynamic or iterative sampling strategies is as essential for ISM as it is for discrete sampling. An example of responding to changing conditions could include establishing additional or alternative DUs to better understand the distribution of contaminant concentrations at a site or to assist in the design and selection of remedial options, based on a review of initial data. Specifically, such a case could involve a relatively large area that was initially thought to be clean and then determined to be heavily contaminated. In this situation, it becomes cost-beneficial to resample subareas in hopes of isolating the contamination and reducing remediation costs. Generally, if DUs are designated in a well-thought-out manner with clear decision statements regarding how the data will be used to answer investigation questions, this will minimize the need for additional unexpected sample collection.

The CSM is essential for DU design. Determining the size, shape, location, depth, and number of DUs and SUs is a critical component of the planning process and is a function of the CSM, the related study objectives, and ultimately the decision mechanisms that relate to the problem formulation.

All contaminant concentrations in soil are heterogeneous on some scale (see Section 2), thus the determination of the sampling scale and the related increment density is very important in all sampling situations. If a finer resolution of contaminant distribution is needed to address the objectives of the investigation, then smaller DUs should be considered. Some basic questions that might be considered include, “How do the definitions of DUs and SUs fit to the study goals of the investigation?” and, “How will the resulting data be used in decision-making to solve the environmental problem?” The designation of DUs and SUs should support and clarify the objectives of the investigation. As the investigation proceeds, if study questions are refined or new questions arise, the DUs, SUs, and decision mechanisms should be reevaluated to ensure they will support the decisions that need to be made.

3.1.3 DQO step 2: identifying study goals (what types of additional information do we need?)

As the goals of the study are defined, the project team should consider the suitability of ISM for meeting those goals. ISM is particularly suited to decision problems related to average soil or sediment concentrations. Through the collection of a large number of increments from multiple locations and a relatively large sample mass, ISM provides better coverage and a more

robust estimate of average concentrations in a volume of soil than is usually achieved with discrete or traditional composite samples. This is particularly important when contaminant concentrations are believed to be near an action level (AL) or decision threshold, or to resolve disagreements among stakeholders.

Following the identification of the study problem and the development of a CSM, the next step in systematic planning is to identify study goals. This is accomplished by developing principal study questions, based upon the CSM, which, when answered, will allow the user to address the study problem identified in step 1. These questions can vary widely and may be different for different phases of the investigation process within a single project, for example:

- Does soil contamination exist (what is the nature of contamination), and if so, has the extent of soil contamination been delineated?
- Does the average concentration of one or more soil contaminants within the investigation area (IA) present unacceptable risk?

These types of questions can be successfully addressed using ISM. Because ISM is applicable to defined volumes of soil and sediment, it is an ideal tool for assessing risks from soils/sediment, comparing site concentrations to regulatory thresholds or other criteria, bulk material characterization for disposal, or other such problems requiring a high degree of confidence in contaminant concentration in a defined volume of soil/sediment. ISM can also be effective for documenting the presence or absence of significant contamination and establishing whether patterns or trends exist within an IA because it allows the user to efficiently obtain information across a large area.

Once the principal study questions have been developed, the user can develop alternative actions, which are logical responses to each potential outcome of the study question phrased as a decision rule. The process of developing alternative actions allows the project team to develop a consensus-based approach at the onset of the investigation, which minimizes the possibility of disagreements further along in the process. Examples of study questions and decision rules relating to hypothetical site investigations and remedial response are provided in Section 3.1.6.

3.1.4 DQO step 3: identifying information inputs (what are the specific inputs for the missing Information we need to evaluate the study goals?)

Having considered ISM during the formulation of the problem to be solved (step 1) and the decisions to be made (step 2), the project team is in a position to state what information is needed and whether/how the ISM methodology can provide some or all of the data needs pertaining to soil and sediment concentrations. It is in this context that the project team should begin to examine and develop ideas pertaining to the attributes of DUs for ISM sampling.

Project teams may need to identify SUs, or the subdivisions of DUs from which separate ISM samples are collected. The boundaries of an SU indicate the coverage of a single ISM sample –SUs define the scale of the ISM sampling and concentration estimation, whereas DUs define the scale of the decision(s) based on that sampling. These definitions allow for the possibility that ISM samples from several SUs composing a DU can be used collectively to make the decision on that DU. It is also possible to employ SUs to address sampling objectives that do not have a clearly associated DU, such as when sampling to evaluate trends in concentrations with distance or depth from a source. Indeed, information from such sampling may itself be used as an input to redefine a DU's boundaries. The final criterion of whether an area sampled using ISM is an SU or a DU is whether or not ISM samples from only that area will be used to support a decision.

SUs define the scale of the ISM sampling and concentration estimation, whereas DUs define the scale of the decision(s) based on that sampling.

One application of SUs is to collect information about average soil contaminant concentrations in subareas of a DU where soil concentrations are suspected to differ based on the CSM. Similarly, SUs might be used to distinguish subareas of a DU where exposure intensity is expected to differ. In either case, the DU is divided into multiple SUs, each of which is separately sampled with one or more ISM samples. Examples related to the use of SUs in such a manner are provided in Section 3.1.6. A general discussion of the concept of stratification in sampling design is provided in Section 2.5.3.1.

SUs may be advantageously used when sampling a very large area where, due to costs or other limitations, sampling 100% of the footprint of a DU is impossible. For example, a 100-acre DU might be sampled by randomly placing fifteen 1-acre SUs within the DU boundary. In this situation, the SU data are treated in an analogous manner as data from traditional

composites or discrete samples to estimate the mean within the DU. As discussed in Section 3.2, ISM data can often be treated as any other data with respect to environmental statistics. An example of how a large DU can be sampled with SUs in this manner is provided in Section 3.1.6.

Caution should be used when applying SUs in ISM study designs. As with other ISM sampling designs, the sizing of DUs should be based on the expected scale of heterogeneity in contaminant concentrations. For example, using a large DU containing non-contiguous SUs may be appropriate to characterize a site where contamination is uniformly distributed based on the CSM, such as aeolian mercury contamination from a power plant or a metals background study. However, such an approach may not be appropriate for a munitions site where range features (target areas, firing lines, and so on) are or were present at the time the contamination was released. For such sites, DUs should be defined for each area representing a unique release profile to aid in site characterization of the nature and extent (N&E). The area and depth of the SU are presumed or already demonstrated through pilot studies to have relatively homogeneous contaminant concentrations that are the results of similar source release mechanisms or dispersion mechanisms.

3.1.5 DQO step 4: define study boundaries (what are the appropriate spatial and temporal boundaries for evaluating the study goals?)

This part of Section 3.1 is the most specific for understanding how to define the number, locations, and dimensions of SUs and DUs to achieve both study goals and support site decisions. The definition of study boundaries for ISM is addressed in the context of informing two interrelated questions that were introduced in Section 3.1.3 as the main objectives of soil and sediment sampling: what is the N&E of contamination, and what is the average contaminant concentration in some defined area?

To address the interdependency of these objectives with ISM, they will be addressed from the premise that understanding patterns of contamination in impacted media as part of an adequate site characterization will assist in designating DU sizes and boundaries. The overarching goal is to determine representative soil contaminant concentrations at a scale that is appropriate for decision-making. For either objective, preliminary data from ISM replicates on the variability of contaminant concentrations can be used to guide delineation of DUs and decisions on the number of increments needed to meet the study goals.

3.1.5.1 Study boundaries related to estimating average soil concentrations in a DU

There are two primary types of DUs that pertain most directly to a study goal of estimating the mean within a defined area: those based on the known locations and dimensions of source areas, called *source area DUs* or *nature and extent DUs* (N&E DUs), and those based on the known locations and dimensions of areas within which human or ecological receptors are randomly exposed, called *exposure area DUs* or simply EUs. In both cases, the primary objective of sampling is to estimate mean contaminant concentrations within a defined volume of soil.

A source area is defined as a discernible volume of soil (or waste or other solid media) containing elevated or potentially elevated concentrations of contaminant in comparison to the surrounding soil such as:

- areas with stained soil, known contamination, or obvious releases
- areas where contaminants were suspected to be stored, handled, or disposed
- areas where sufficient sampling evidence indicates elevated concentrations relative to the surrounding soil over a significant volume of contaminated media

N&E DUs are differentiated from exposure area DUs in that the boundaries of N&E DUs and the scale of sampling are based on a reasonably well-known extent of contamination, while the boundaries of exposure area DUs are determined through the exposure assumptions of the receptors in the risk scenario.

N&E DUs. Source areas are of concern in an environmental investigation because contamination can migrate from source areas to other locations and media (such as leaching to groundwater, volatilizing to soil gas and/or indoor air, overland transport, or running off to surface water), and also because direct exposure to source area contamination may be of concern. The identification and characterization of source areas is an important part of any environmental investigation. N&E DUs can be identified by using various methods, including observation, review of site records, preliminary samples, field analytical samples, wide-area assessments, aerial photographs, interviews, and site surveys. Ideally, source areas are identified based on knowledge of the site before DU designation and subsequent ISM sampling. However, source areas can also be discovered through the interpretation of sampling results.

As discussed in Section 3.1.4, it may be advisable to designate smaller N&E DUs or SUs within larger DUs based on an understanding of potential contaminant distributions. Assessment of a smaller subarea might be motivated by knowledge of site history or topography that could influence fate and transport, leading to an area where concentrations are higher relative to the surrounding soil (that is, a secondary source area). A common example of an N&E DU within a larger DU relates to the investigation of lead soil concentrations in the yards of homes known or suspected to be contaminated with lead-based paint chips. An area around the perimeter of the house might be designated as a separate DU and characterized separately from a larger DU consisting of the entire yard. This is illustrated with an example in Section 3.1.6.2, Example 2B.

Exposure area DUs. Exposure area DUs, or EUs, are a fundamental part of many environmental investigations and are a key tool in risk assessments and risk-based decision-making. An EU in the context of ISM is defined as an area where human or ecological receptors could come into contact with contaminants in soil on a regular basis (refer to exposure area discussion in "Risk Assessment Guidance for Superfund, Vol. I, Human Health Evaluation Manual (Part A)" and "Ecological Risk Assessment Guidance for Superfund; Process for Designing and Conduction Ecological Risk Assessments" (USEPA 1997).

The concentration data collected from an EU can be used to screen risk by using published criteria or to otherwise assess risk to human and ecological receptors. The data are commonly used to develop EPCs, which are generally estimates of the average concentration of a contaminant within the EU. When the remedial decision is to be based on risk assessment results, the EU should represent the area (and depths) where exposure has a high probability of occurring. The size and placement of EUs depend on current use or potential future use of the site, as well as the types of receptors that are expected for each of the land use scenarios. When systematic planning considers soil and sediment data collection to support risk assessment or risk-based decision-making, a primary question is, "Over which area and depth do samples need to be taken to reasonably represent potential exposures of concern?" An EU is commonly a spatially contiguous area within which a human or ecological receptor is generally assumed to be exposed over time in a random manner, and this random pattern of exposure is the basis for using the average to represent the EPC. Practically, we rarely know with a high degree of confidence what the exact size and location of a future exposure area is going to be, although we can make reasonable assumptions or reference default values for certain types of land use. This uncertainty regarding future exposure is why it is important to consider both source areas (based on the known or inferred spatial pattern of contamination) and likely exposure areas in developing DUs.

Lastly, although it is common and practical to discuss EUs based primarily on area, the nature of soil sampling requires that we also consider depth when defining an EU. If, for example, an exposure model states that the activities of humans or burrowing animals might reach a certain depth, then the average soil concentration from the ground surface to that depth is of interest. But here it is especially important to recognize that although, for example, humans *could* excavate soil to a depth corresponding to a basement, we do not necessarily know they will or what the exact location and volume of the excavation will be. If contamination is surficial, it will generally be inappropriate to assume that future excavation will *certainly* result in dilution of the contamination through mixing with clean subsurface soil. These ideas concerning EUs are illustrated with examples in Section 3.1.6.

3.1.5.2 Study boundaries related to evaluating the N&E of contamination

ISM can be used to determine the N&E of contamination in soil and sediment at contaminated sites. This section addresses the use of ISM to evaluate the vertical and lateral extent of contamination, and to identify subareas of elevated soil concentrations. The use of ISM in conjunction with field screening tools is also briefly discussed.

Evaluating the vertical extent of contamination with ISM. Subsurface DUs are an important application of ISM sampling because of the frequency with which subsurface contamination is encountered. In some situations, contamination may be situated entirely below the ground surface. Subsurface DUs are often tabular shaped, like thin books, and the number and thickness of these vertical intervals must be carefully considered based on the CSM, site geology/hydrology, potential receptors, existing data, and applicable state regulation and guidance. Objectives for the investigation related to assessing the N&E of contamination in the subsurface might include one or more of the following:

- determining whether leaching of contamination from soil to groundwater may have occurred
- estimating average soil concentrations by depth interval(s)
- estimating the volume of contaminated soil that may need to be removed or properly managed

Ideally, the nature and quality of ISM subsurface samples should be similar to those collected for more easily accessible surface soils, and in a manner that allows every possible increment in the DU an equal likelihood of being collected. Sampling theory also indicates that the entire cross-section of the DU be sampled in each increment making up the ISM

sample, but in practice, the combined mass of the increments from a large number of borings would likely result in an impractical sample volume. Therefore, field subsampling plans may be needed to achieve sampling objectives.

Sampling approaches for subsurface soils differ from those applied to surface soils because access to the subsurface is more difficult. It is not uncommon to design an ISM sampling approach for subsurface soils that has less increments than are used in the respective surface investigation, but this does not mean that low-quality data are generated for these subsurface samples. Adequate data can be generated with fewer increments in subsurface sampling when geological heterogeneity and the end use of the data are understood, and this should be addressed during the planning process. Moreover, potential limitations of the data should be clearly discussed, and the implications regarding uncertainty in mean soil concentrations should be taken into account in risk management decisions. Section 4 goes into further detail on sampling techniques for subsurface soils. Example of subsurface sampling designs are provided in Section 3.1.6.

Evaluating the lateral extent of sediment contamination with ISM. When existing ISM data indicate high concentrations of contaminants are locally elevated in soil or sediment, such data may be sufficient to establish the boundaries of a source area. However, in other situations, it may be necessary to refine the study goals and redefine the number and boundaries of DUs based on information from additional sampling.

An example of applying ISM to address data needs pertains to the evaluation of trends in contaminant concentrations as a function of lateral distance. Contiguous ISM SUs along a drainage can provide sound information on contaminant concentration trends and also provide information on average concentrations on the scale of one or more SUs. In some situations, designation and testing of anticipated clean *boundary DUs* around anticipated areas of heavy contamination can help to minimize the need for remobilization. Examples of an ISM application to evaluate the lateral extent of contamination is provided in Section 3.1.6.

Evaluating the potential presence of subareas of elevated contaminant concentrations with ISM. Historically, discrete soil sample results with concentrations above an AL have often been assumed to represent a significant volume of surrounding soil containing sufficiently high concentrations of contaminant to warrant concern. The concentrations in these assumed volumes have been considered to represent source areas, which are defined in various ways by different regulatory bodies (ITRC 2008). This range of definitions can lead to a wide range of interpretations and has typically led to additional sampling events to further define the N&E as parties struggle to determine what qualifies as an area of elevated concentration versus a source area. It is highly recommended that project teams include their state regulators early in the planning process and that all stakeholders agree upon the basis for defining and distinguishing elevated concentrations from source areas.

One reason why ISM uses so many increments (a minimum of 30) is to have sufficiently dense *spatial coverage* of the DU. This *spatial density* improves the chance that the field sample will include significant areas of elevated concentration in the same proportions as present across the DU. An important ISM principle is that DUs should not be designed in a way that results in dilution of significant volumes of highly contaminated soil from smaller areas. The location and size of source areas can often be established or hypothesized based upon site history, including waste disposal units, locations of known or suspected spills or releases, and volumes of soil shown by previous sampling to have significant contaminant concentrations relative to the surrounding soil. In other cases, the presence of subareas of soil with relatively high concentrations is suspected, but the locations are uncertain.

A DQO study goal could be to find significant small areas (horizontal and depth) of elevated contaminant concentration(s) above risk-based concentration(s) or an AL within a DU. The DU could be comprised of several SUs designed to meet the "small area" volume requirement. It is in the systematic planning phase that project teams *must define* and designate *what concentration* and *what volume*, *surface area*, *or mass are significant to their decision-making*. To define the size and concentration of a significant small area of elevated contamination, they can use an Excel spreadsheet tool, if the critical condition of a mature CSM is met. For an example and more details on this concept, link to White Paper (Crumbling 2014).

Statistically based sampling designs can be developed to determine whether localized areas of higher soil concentrations exist, even if the locations of such subareas within a larger site are unknown.

The *spacing of increments* (and thus the number of increments needed to fill the DU's area) can be set to have a desired statistical probability of increments being collected from within an area of defined size for incorporation into the field sample. In this case, if the size of a potential subarea of elevated concentrations is specified, sampling can be conducted to determine whether one or more such areas exist within a DU with an objective degree of confidence and scientific

defensibility.

A free software program developed by Pacific Northwest National Laboratory (PNNL) called Visual Sample Plan (VSP) is available to determine the *increment spacing for the DU grid* so as not to miss sampling from a significant small area of elevated concentrations within the DU (VSP 2019). VSP has varied statistical sample size designs built in to support sample collection using ISM. The designs are grouped into two general categories – estimating the mean and detecting elevated regions. Both designs are built with standard statistical sample size design principles – namely, the stakeholders must specify desired Type I and Type II errors and provide estimates for standard deviations associated with the sampling process as well as regulatory thresholds to which the sample values will be compared. VSP does not implement any of Pitard or Gy's equations, although it similarly attacks the goal of accurately estimating concentration levels in soil. A validation study of VSP ISM sampling design for elevated regions at a military training range demonstrated reliable estimates of mean concentrations and corroborated spatial areas with statistically elevated concentrations within the DU for 2,4-dinitrotoluene (2,4-DNT) (USEPA, 2015).

VSP's elevated regions module sampling pattern and design differs from the typical ISM sampling pattern and design described within this document and presented in the examples in both Section 3.1.6 and the case studies in Appendix A. VSP's elevated regions employ a pattern of rows and columns to design increments for an ISM sample in such a way that they can be combined into ISM samples but still used to spatially locate areas of high contamination. Figure 3-3 depicts a VSP 4 x 4 ISM row-column design with 16 cells. VSP can calculate either the number of incremental samples to achieve a desired power of detecting contamination above a specified level or the probability of detecting an elevated concentration, given a specified number of increment samples.

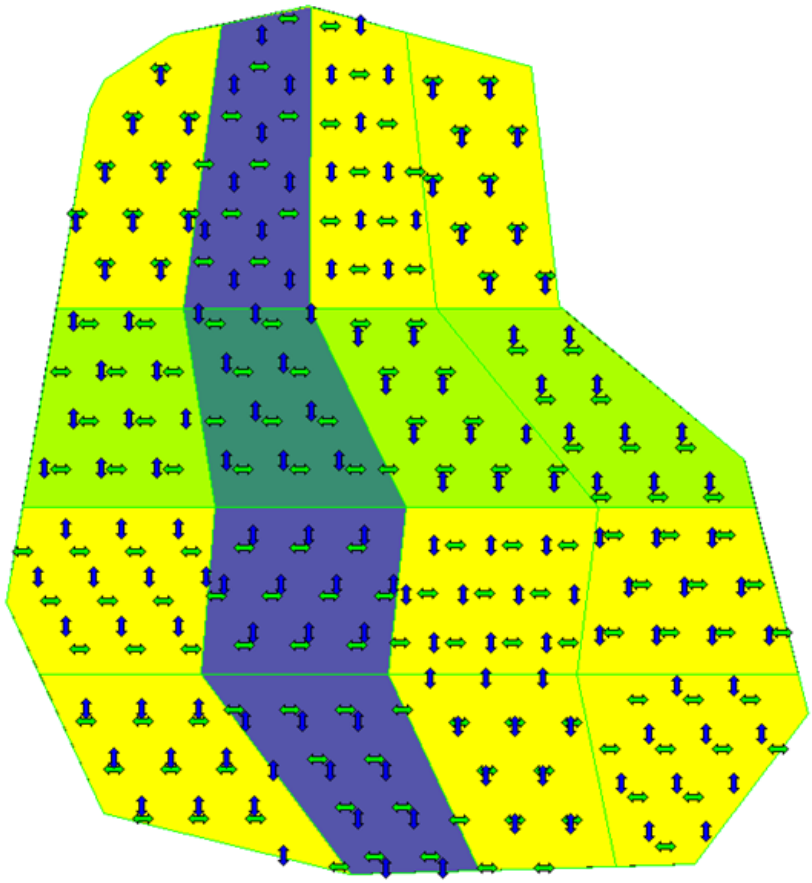


Figure 3-3. VSP 4 x 4 ISM row-column design. The blue vertical arrow increments form the ISM sample for column 2, and the green horizontal arrow increments form the ISM sample for row 2.
Source: VSP help file, https://vsp.pnnl.gov/help/.

As with any statistical tool, there are important assumptions and limitations for the user and project team to consider:

- Users must understand the assumptions of the statistical models used in VSP.
- The closer the analyte's actual data distribution and variability agree with the assumptions of the underlying statistical model, the more accurate VSP's output will be.
- Even when inputs to statistical calculations are reliable, the numerical outputs of statistical calculations are still

imperfect estimates of field concentrations, receptor exposures, and cleanup volumes.

Moreover, there are caveats specific to VSP:

- The user must upload a map of the area (DU) or depict a sampling area (DU) first to enable the ISM Elevated Regions module within the Locate Hot Spots part of a Sampling Goal.
- For very complex shaped sample areas, the site division algorithm does not work well.
- The grids for cells can be square, rectangular, or triangular.
- The user is required to have data on or make a conservative assumption regarding the SD within the small area of elevated concentration and the SD within the remaining IA in the DU. If comparable studies with variance estimates are not available, a pilot study may be needed, which will affect cost. If assumptions on the variance are too conservative, unnecessary costs may be incurred.
- The VSP elevated regions module sampling pattern and design differs from and is more costly than the typical ISM sampling pattern and design, but it provides specific levels of confidence in detecting small areas with significantly elevated concentrations.
- For ISM designs to estimate the mean, VSP does allow the user to input the costs associated with the sample collection and measurement. The costs input are utilized by VSP to propose the most cost-efficient way to aggregate the increments from the DU into the ISM samples with a predicted level of confidence in locating elevated regions in the DU.

During systematic planning, the project team must ensure their study site meets the assumptions and that they have weighed the limitations and caveats for VSP against the study goals. For more details on this concept, see the White Paper (Crumbling 2014). Users are strongly encouraged to fully understand and consult the additional details on VSP designs plus the inherent assumptions and limitations that are available in the VSP help files (https://vsp.pnnl.gov/help/). VSP help for the ISM elevated regions module are under the *Sampling Goal* menu, Locate Hot Spots, *Locate Hot Spots Using MI Samples*.

Another approach, but one that lacks the statistical rigor of a defined statistical probability of increments being collected from within an area of defined size, would be to increase the number of increments and thereby the spatial coverage in the DU, to improve the chance that the sample will include significant areas of elevated concentration in the same proportions as present across the DU. A large relative SD (RSD) among replicates can be used as an indication that a small area of elevated concentration in the DU was sampled in one replicate but not in another. This condition might trigger additional investigation with more replicates from the DU, more increments in the DU, or subdividing the DU into multiple smaller SUs. (See Section 3.2.4.2 text and Table 3-2, which classifies heterogeneity of increments in terms of low, medium, and high coefficient of variation [CV] of replicates.)

Effective detection and delineation of areas of elevated concentrations in heterogeneous soil matrices is a challenge. To avoid the pitfalls of “chasing” areas of elevated concentration, ISM practitioners are encouraged to define an area or volume of concern as part of the SPP. Similarly, the planning team is encouraged to define decision rules related to the assessment of the data acquired. An example of such an ISM application is provided in Section 3.1.6.

Use of field screening methods with ISM. Field screening methods can sometimes be used in conjunction with ISM to expedite evaluation of the N&E of soil or sediment contamination. ITRC provides guidance for the selection and use of field site characterization tools to support development of a CSM, plan for the collection of samples for laboratory analysis, and provide input for considering remedial strategies (ITRC 2019). Field portable XRF and gas chromatography are techniques that can be used to gain an understanding of the N&E of contamination and help define the boundaries of SUs or DUs. “EPA Test Method 6200” (USEPA 2007) provides guidance for the use of field portable XRF spectrometry for determining metals concentrations in soil and sediments. Although the guide was written in 2007 and considers the best available technology at that time, its recommendations are valid and still employed in present-day publications and studies. Field portable gas chromatography can be used to evaluate soil and sediment concentrations of organic chemicals, particularly volatile compounds.

3.1.5.3 Laboratory processing of ISM soil and sediment samples

The manner in which soil and sediment samples are processed can affect measured contaminant concentrations in these samples and whether the concentrations are consistent with the assumptions underlying human and ecological exposure models. During the planning process, the project team should consider the physical and chemical characteristics of suspected contamination and the end use of the data to choose the most appropriate sample processing options. There are

four issues and related questions that the project team should consider during planning:

- moisture management (Is air-drying of the samples acceptable?)
- particle size selection (Should the samples be sieved or otherwise processed to exclude particles larger than a specified diameter?)
- particle size reduction (Should the samples be ground prior to analysis?)
- sample digestion/extraction (Should the mineral matrix of the sample be dissolved, or should digestion/extraction target the contaminants adsorbed in soil particles or otherwise present in soil?)

The specific analytes that are the focus of the investigation can influence sample processing decisions because there can be a wide range of physical and chemical characteristics within analyte groups. Some characteristics that can influence the selection of sample processing options include boiling point, volatility, air reactivity, and sorption characteristics. The presence of high-concentration nuggets of contamination can also influence sample processing decisions. Section 5.2 provides detailed guidance on selecting sample processing options.

3.1.5.4 Considerations for determining the number of increments and sample mass

As covered in Section 2.5 and Section 2.6, the number of increments collected for an ISM sample and the total mass of the sample are the main factors controlling the representativeness of an ISM soil sample, where representativeness is the measure of how well the sample represents the entire mass of soil within an SU or DU.

Section 2.5 and Section 2.6 should be reviewed to understand the basis for selecting the number of increments for a given sample and the target mass of the ISM sample. Collection and analysis of a large sample mass helps to control what is referred to as CH or FE, which refers to the differences in contaminant concentration related to the physical or chemical characteristics of different soil particles. A large number of increments helps to control distributional heterogeneity, which refers to differences in contaminant concentrations due to the large-scale spatial distribution of contamination within the SU or DU.

The selection of the number of increments and sample mass is dictated by the anticipated degree of small- and large-scale heterogeneity, which might be influenced by the distribution of pockets of contamination across a DU, by contaminant chemical characteristics, by soil type and physical characteristics, and by the contaminant release mechanism.

It is generally accepted that between 30 and 100 increments is appropriate for many applications, with a larger number of increments being driven by a larger degree of distributional heterogeneity.

Figure 3-4 presents various factors to consider in deciding on the number of increments to collect from a DU and their influence on heterogeneity. The graphic illustrates the influence of various physical and chemical factors – such as chemical properties, and whether a release is associated with the solid or liquid phase of soil – on potential variability and the related association of each variable to the number of increments to help control heterogeneity.

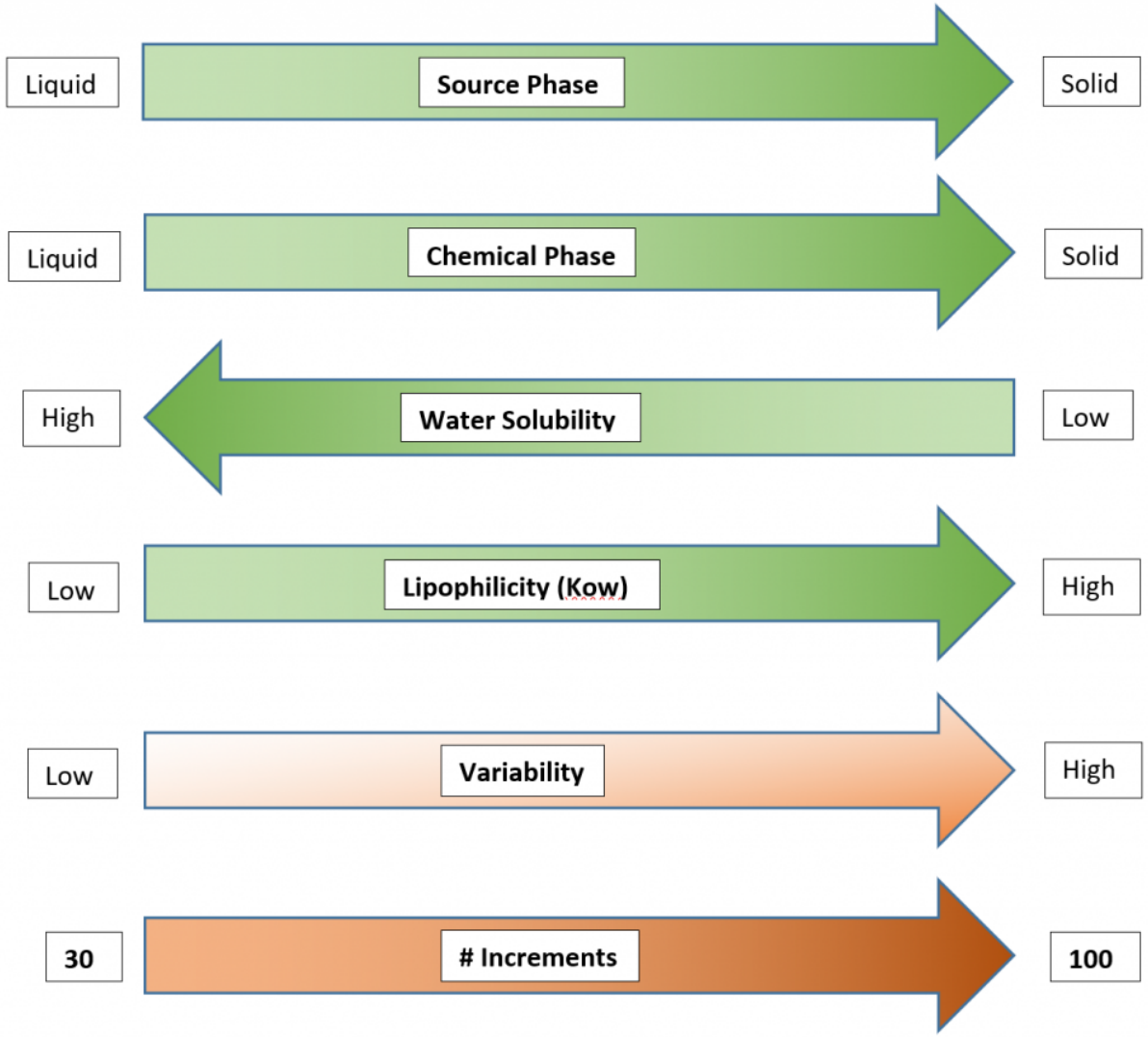


Figure 3-4. Variables to consider in deciding on the number of increments to collect from a DU. The various factors are represented by green arrows, with arrow direction and increase in color gradient corresponding to an increase in a specific physical or chemical property. The consequential effect of each factor on variability is depicted in the parallel light orange arrow. The related association of each variable to the number of increments to help control heterogeneity is depicted with the parallel dark orange arrow.
Source: ITRC ISM Update Team, 2020.

Collection of a field sampling mass greater than 1 kg is recommended. Final ISM field samples typically weigh 500 g to 2,500 g, and as discussed in Section 2.5.3.1, many laboratories will limit soil or sediment sample mass to about 2 g to 3 kg. In general, individual soil increments typically weigh 20 g to 60 g. Based on the target final mass of the ISM field sample and the number of increments specified to control distributional heterogeneity, the minimum mass of the individual increments can be calculated (see equation in Section 4.2.3). The mass of any single increment depends on the depth of interest, soil density, moisture content, and the diameter or size of the sample collection tool. In addition to the function of controlling CH, the mass of the final ISM sample must also be sufficient for the planned analyses, any additional QC requirements, and possible repeat analyses due to unanticipated field, laboratory, and/or QC failures. Note that sieving of soil samples at a specified particle size reduces the amount of soil mass available for preparation and analysis, although as discussed in Section 2.6.2.1, such sieving will also tend to reduce CH.

3.1.5.5 Common sampling designs used with ISM

Planning and design for ISM shares many of the characteristics common to other types of environmental soil sampling. Among the common types of statistically based sampling designs are *simple random* sampling, *stratified random* sampling, and *systematic random* sampling. The element of randomness common to these designs allows statistical inferences to be made about the sampled population, as well as a defensible calculation of average contaminant concentrations within a DU.

Implementation of these types of sampling designs, along with the basis for selecting among them, is discussed in (USEPA 2002e).

Examples of *simple random* sampling, *stratified random* sampling, and *systematic random* sampling are shown in Figure 3-5. In the case of *stratified random* sampling, the strata are shown as regular grids, thus the sampling design is labeled "Random within Grids." For *systematic random* sampling, rather than selecting a random location for each grid cell within a DU, randomization is performed only once, and the randomly selected location within a cell is then applied to all other cells. This *systematic random* sampling design is also shown in Appendix A in Case Study 9, which contains a WP with exceptional articulation of the systematic random placement of increments. For further discussion ITRC 2012.

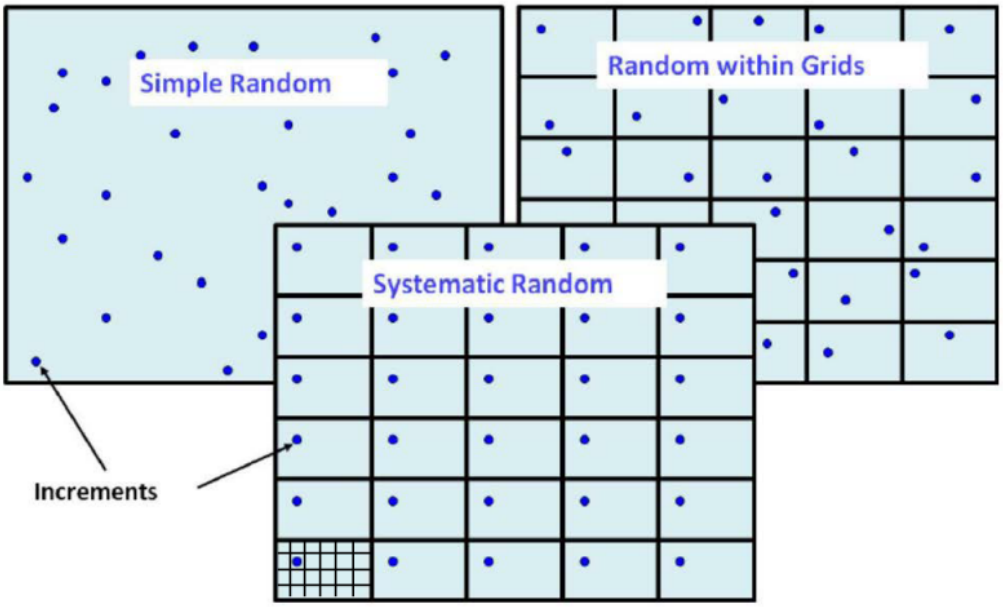


Figure 3-5. Unbiased increment collection designs.
Source: ITRC ISM Update Team, 2020.

Up to this point, this section has provided a summary of the key aspects of systematic planning and DU design in relation to the collection of soil and sediment samples. Section 3.1.6 provides three examples that illustrate these important concepts in different situations.

3.1.6 Examples illustrating planning and design for ISM

The reader will notice that the three examples described here differ in how they were conceptualized and developed. They are presented to illustrate a range of situations and approaches, and to help the reader realize that while thoughtful planning is always necessary, there is no precise formula for how to evaluate a site. Each example illustrates a different application, interpretation, and development of a sampling plan. As discussed in Section 3.1.1, steps 1 through 4 of USEPA's DQO process have been applied to help structure the discussion of systematic planning and to organize these examples. However, some material associated with later steps of the DQO process (particularly step 7, sampling design) is necessarily integrated in these three examples:

- an agricultural field, settling pond, and drainage swale (Example 1)
- former agricultural field and establishing exposure DUs (Example 2)
- former industrial facility that is to be redeveloped (Example 3)

3.1.6.1 Example 1: agricultural field, settling pond, and drainage swale

Four different ISM topics will be addressed through this example set:

- estimating average concentrations in a defined volume of soil or sediment

- evaluating the vertical profile of contamination in soil or sediment
- evaluating the horizontal extent of contamination along a drainage
- estimating average concentrations in stockpiled material for waste management decisions

CSM. A bermed enclosure with a cement floor was used for holding irrigation water runoff for a large agricultural field that had not been actively farmed for decades. Water was supplied as flood irrigation to the field, and on occasions when excess irrigation water was applied, the runoff was captured in a 1-acre holding pond situated at a slightly lower elevation than the field. Organochlorine pesticides (OCPs) were historically used on the field, and soil samples from the field indicate that concentrations of several OCPs are above state risk-based soil screening criteria. The farmers note that there is about 6 ft of sediment that has accumulated in the settling pond, and also that the rates of pesticide application had increased over time when the field was being used, such that the more-recent deposits might have the highest concentrations of OCPs. Furthermore, the farmers point out a notch in one of the berms, on the other side of which is a cement apron that leads to a shallow swale. The swale has a gentle gradient and broadens as it leads toward an ephemeral stream that is about a half-mile away. The excess irrigation water reportedly rarely overtopped the berm, but there is little confidence in that observation (see Figure 3-6).

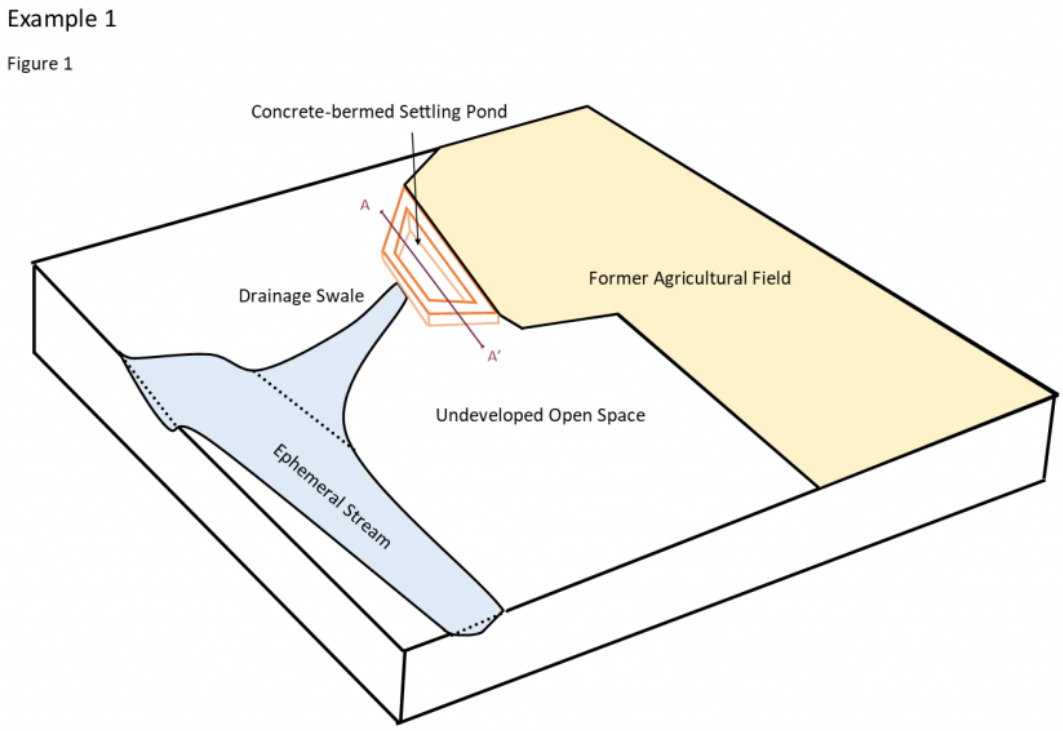


Figure 3-6. Overview schematic for Example 1.
Source: ITRC ISM Update Team, 2020.

Problem formulation. The problem is defined as determining whether sediment concentrations in the settling pond, as well as the swale, could potentially present unacceptable risks to individuals who might currently access the area or to people in the future should the land be repurposed for residential or commercial uses.

Study questions. An initial question (study question 1) is posed as, "Does OCP sediment contamination present unacceptable risk under a residential scenario?"

This question reflects the understanding that residential land use is protective of any other exposure scenario. During the SPP, state soil screening criteria for OCPs are identified as inputs to this question. It is accepted that lateral patterns in OCP

sediment concentrations are unlikely within the settling pond, due to the manner in which contamination was deposited, but the CSM's prediction that the contamination decreases with depth to the cement floor of the settling pond should be confirmed with data. It is further assumed that, because the sediment pond received field runoff directly, OCP concentrations in pond sediments must necessarily be greater than those in the swale.

A second question (study question 2) is therefore posed as, "Are OCP sediment concentrations decreasing with depth in the settling pond?"

Decision rules and sample design for study questions 1 and 2. The first two study questions pertain to OCP sediment concentrations in the settling pond. From these two questions, a decision rule is developed applying the premise that the highest OCP concentrations will be found in the settling pond:

If average OCP sediment concentrations are below residential soil screening criteria in the surface interval, *and* concentrations are decreasing with depth, *then* take no further action, *else* characterize OCP contamination in the swale.

The lateral dimension of the DU area for study questions 1 and 2 is defined as the entire 1-acre surface area of the settling pond within the berms because, as noted in relation to study question 1, systematic patterns in OCP sediment concentrations within a depth stratum are unlikely within the settling pond. For study question 1, a surface sediment interval of 0 to 6 in, where OCP concentrations are expected to be highest based on the CSM, is defined. Because lateral heterogeneity is anticipated to be low, a value of 30 increments is selected from within the recommended range of increments (30 to 100) for the surface soil layer. Three replicates are proposed for the surface interval to support estimation of uncertainty in average OCP sediment concentrations (see Figure 3-7).

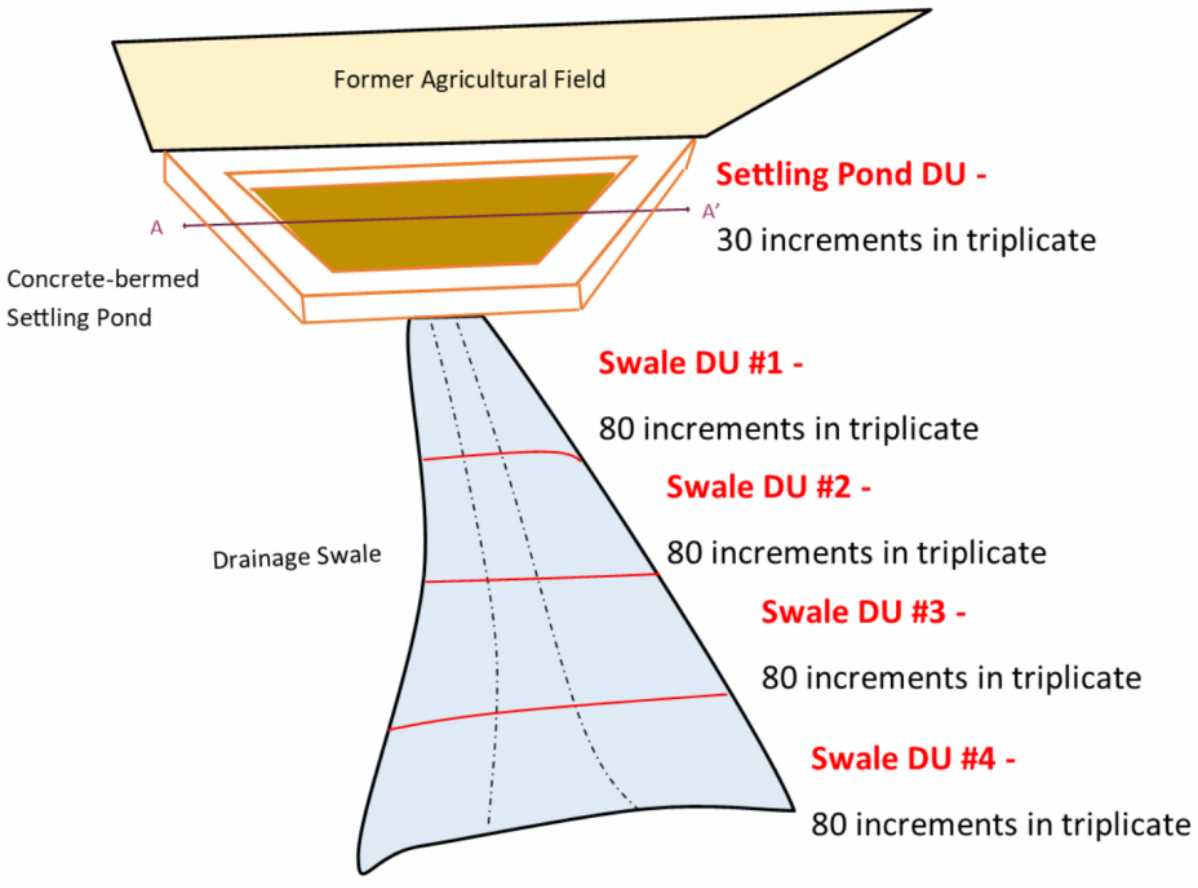


Figure 3-7. ISM design for surface interval sampling in Example 1.
Source: ITRC ISM Update Team, 2020.

To address study question 2, the remaining depth of sediment (approximately 6 ft) is divided into three depth intervals of approximately 1 to 2 ft each. Although ideally 30 increments and three replicate samples would be collected from the deeper intervals, such as were obtained for the surface interval, the project team decides to phase the depth sampling because of the cost of sampling and the expectation based on the CSM that OCP concentrations at depth are likely to be low and relatively homogenous. Ten corings are proposed to obtain 10 core increments from each of three subsurface intervals corresponding to the approximate 6-ft sediment depth in the settling pond (0.5-1.5 ft, 1.5-3 ft, and 3-5 ft), with no replicates.

Figure 3-8 depicts DUs pertaining to subsurface sampling, where DU-1, DU-2, and DU-3 are applicable to Example 1. A 1-kg sample is identified for collection as a plug subsamples from each of the three depth increment (see Section 4.5.1), resulting in a 10-kg sample mass for each subsurface interval. Because laboratories typically limit sample mass to a few kg, field subsampling (per discussion in Section 5.3.5) is proposed for the 10-kg samples to prepare a final 2-kg sample for shipping to the analytical laboratory. A second decision rule is developed specific to study question 2:

If OCP sediment concentrations in a depth interval are clearly below residential soil screening criteria, *then* take no further action, *else* consider either additional sampling to refine the estimate of average OCP concentrations (if concentrations are close to criteria) or remedial action (if concentrations are far above criteria).

When the settling pond analytical data for OCPs are received and evaluated, two key findings emerge. First, it is clear that OCP concentrations in all depth intervals exceed both residential and industrial state soil screening criteria. Also, there is relatively high variability among the three replicate samples of the surface sample interval, meaning the assumption of relatively homogeneous contamination seems to be incorrect. Based on the magnitude of the screening level exceedances, it was determined that proceeding with this relatively large degree of data variability was unlikely to result in decision errors, and that the data were sufficient to proceed to consideration of remedial action in the settling pond without further sampling (see study question 4 below).

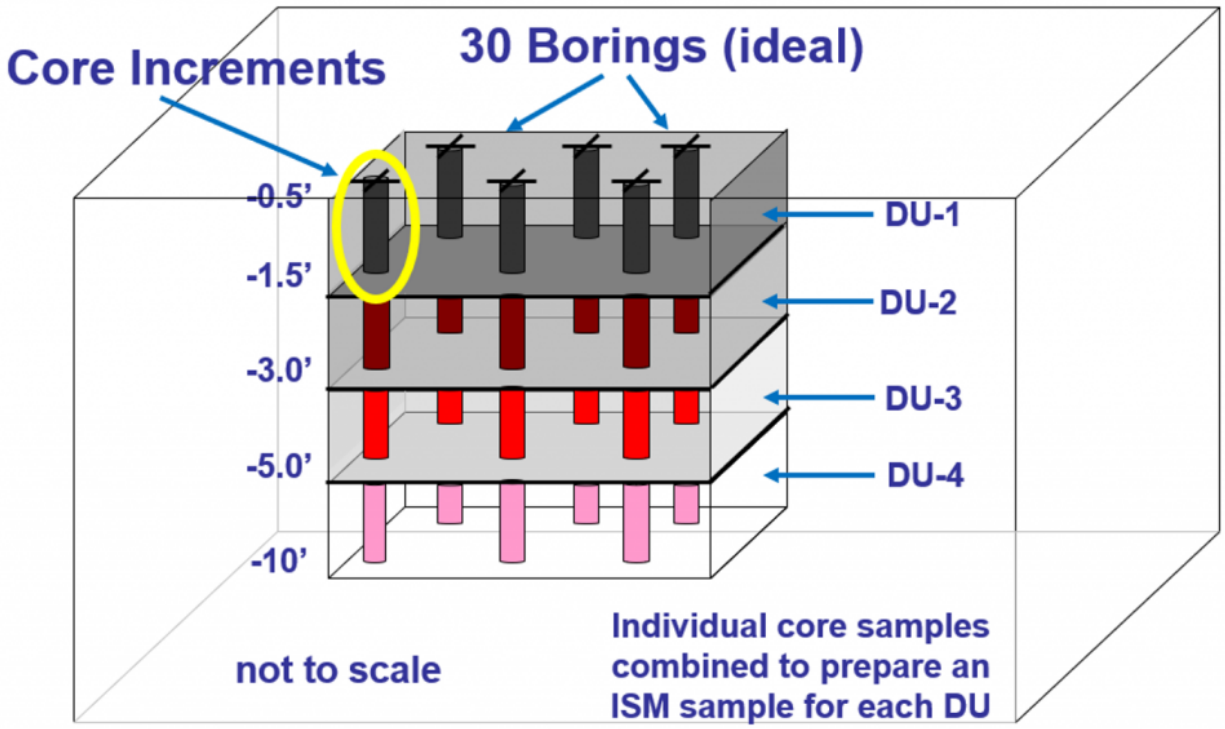


Figure 3-8. ISM design for subsurface interval sampling in Example 1
Source: ITRC ISM-1 Team, 2012.

Decision rule and sample design for study question 3. Consistent with the decision rule for study questions 1 and 2, a design is developed to evaluate OCP contamination in the swale. The swale is divided longitudinally into 500-ft intervals between the settling pond and the ephemeral stream. As the swale broadens with distance from the pond, the areas of these swale segments also increase with distance: 5,500 ft^2, 8,000 ft^2, 15,000 ft^2, 19,000 ft^2, and so on. There is no visual indication of channeling or deposition within the swale. The range of surface areas in the first four swale segments, from about one-eighth to one-half-acre, are sized to fall within the range of areas applicable to both human and ecological exposure scenarios related to state soil screening criteria. Because there is no visual evidence of preferential areas of sediment deposition in the swale, and because the areas of the swale segments are within the range of potential exposure areas, there is minimal concern that there could be subareas of higher concentrations or hot spots within a swale segment. Therefore, contingencies for defining smaller DUs based on data evaluation are not proposed. The residential soil screening criteria applied for the decision rule for study questions 1 and 2 are also applied to the swale segments, since they are determined to be protective of potential ecological impacts.

A third study question (study question 3) is developed: “Do average OCP sediment concentrations in the swale present unacceptable human or ecological risk, and if so, has the lateral extent of contamination relative to such concentrations been established?” From this question, the following decision rule is developed:

If OCP sediment concentrations are decreasing with distance from the settling pond, *and* average OCP concentrations are below residential soil screening criteria, *then* take no further action in the swale, *else* consider additional sampling (to determine extent) and/or site-specific risk assessment or remedial action.

Each of the first four swale segments are defined as DUs. A sediment depth interval of 0 to 12 in is defined for sampling, based on a field survey that showed roughly this thickness of fine-grained material (similar to agricultural field soil) is present within the swale. Because heterogeneity of OCP concentrations in swale sediments is unknown, and given higher than anticipated heterogeneity in settling pond sediments, a value of 80 increments is selected from within the recommended range of increments (30 to 100). Three replicates are proposed for all four segments.

When the swale segment analytical data for OCPs are received and evaluated, OCPs are detected sporadically and only in the first two segments. The average concentrations of OCPs in these segments are below both residential and ecological screening criteria, so consistent with the decision rule for study question 3, no further action is proposed for the swale.

Decision rule and sample design for study question 4. As discussed, average OCP concentrations in all depth intervals of the settling pond exceed screening criteria by a relatively large margin, and evaluation of the three replicate data for the surface interval indicates that there is a high degree of variability in OCP sediment concentrations. Rather than continue in situ sampling, informal cost-benefit consideration suggests that it is advisable to excavate settling pond sediments and dispose of them in an appropriate facility. The OCP concentrations are near levels that differentiate between two disposal facility options with very different disposal costs. An excavation and stockpiling plan is developed to remove sediments by depth and stage them in a long and narrow stockpile that is arranged on the long axis from shallower to deeper sediments, since the analytical data indicate an inverse relationship between OCP concentration and depth.

A fourth study question (study question 4) is developed: “Are average OCP concentrations in segments of the stockpile above the acceptance criteria of the lower-priced landfill?” From this question, the following decision rule is developed:

If average OCP sediment concentrations in a stockpile segment are above the acceptance criteria of the lower-priced landfill, *then* send the material to the higher-priced landfill, *else* ship to the lower-priced one.

The volume of an individual stockpile segment, defined as a stockpile DU, is determined by transportation costs and minimal disposal quantity rules for the hazardous waste landfill. The stockpile is laid out with a depth of 2 ft to allow for cost-effective hand coring. Because heterogeneity is known to be high, and sampling costs are low, a value of 100 increments per segment is selected from within the recommended range of increments (30 to 100). Three replicates are proposed for all segments to support an estimate of a 95% UCL on mean OCP concentrations.

3.1.6.2 Example 2: former agricultural field and establishing exposure DUs

Example 2 focuses on developing and delineating EUs for human health risk-based study questions and will guide you through the development of ISM sampling plans with successively more complex site CSMs. Throughout Example 2, the risk-based study questions focus on current and potential future residential land use with no ecological receptors. The DU size is ¼ acre, the assumed size of a future residential lot. Residential lot sizes vary, thus planning with a regulatory authority and their risk assessor is essential.

Example 2A covers four concepts:

- establishing replicate heterogeneity limits in the DQOs as an MQO in Specific Study Goal data needs
- assessing the assumption of homogeneous contaminant distribution (low heterogeneity) by defining as RSD of 20% in a Decision Rule.
- extrapolating to unsampled DUs within a large study area
- designing background DUs

Example 2B covers three additional concepts:

- Designing source area N&E DUs within EUs
- Designing SUs within DUs (for example, a children’s play area within an adult residential DU)
- Designing for weighted averaging of 95% UCL

The problem formation (DQO step 1) is similar for both Examples 2A and 2B: determine the average concentrations of COPCs in surface soil to assess if potential risks are unacceptable to current and/or future residents. (Note that Example 1 provides guidance on subsurface sampling. Care should be taken to plan for the number of subsurface increments needed to obtain reliable concentration estimates with minimal uncertainty, like surface soil ISM sampling, for use in estimating potential risks.)

Example 2A. The CSM for Example 2A (Figure 3-9a) is a 30-acre agricultural use area that has been farmed since the early 1900s. Legal broadcast application of OCPs and arsenical pesticides, including lead arsenate, is the only suspected potential source of soil contamination and is limited to surface soil contamination with no migration of COPCs to the subsurface. The topography is flat, except for furrows between rows of plants. No localized areas of potentially heavy contamination were identified in a thorough Phase I Environmental Site Assessment (ESA). Moreover, county records indicate that, in recent years, there has been no use of triazine herbicides, carbamates, or organophosphate pesticides. There are no known or suspected pesticide mixing areas, and no existing structures or historical aerial photographs show any evidence of structures dating back to the 1920s. The site is surrounded by agricultural fields, except an area to the west that has never been farmed or had any other known uses based on historical photographs and county records. The site is scheduled to undergo residential development.

Problem Formulation – Identify decisions needed and develop CSM. The goal of the ISM sampling event is to determine the average concentrations expressed as the 95% UCL of arsenic, lead, and OCPs in surface soil to assess potential future residential risks and ascertain if cumulative risks or hazards exceed the regulatory acceptable points of departure of 1×10^{-6} and 1.0, respectively (see Section 1, where 95% UCL is defined, and Section 3.2, which has a discussion on 95% UCL).

For risk-related problems, problem formation will almost always entail the following sequence of steps to generate the preliminary CSM and potentially complete exposure pathways:

1. identify potential primary source areas/release mechanisms
2. identify potential secondary source areas/release mechanisms
3. identify media that could be impacted by such a release/migration (exposure media)
4. identify receptors, both current and future, that could come into contact with these contaminated media and the exposure routes (ingestion, inhalation, or dermal)

First, generate the preliminary CSM and potentially complete exposure pathways to establish EUs.

- Primary source areas/release mechanisms. The only potential source for Example 2A is the agricultural field, with the release mechanism being the legal broadcast application of OCPs and arsenical pesticides, including lead arsenate. There have been no known releases to the adjacent background area that is upwind from the agricultural field.

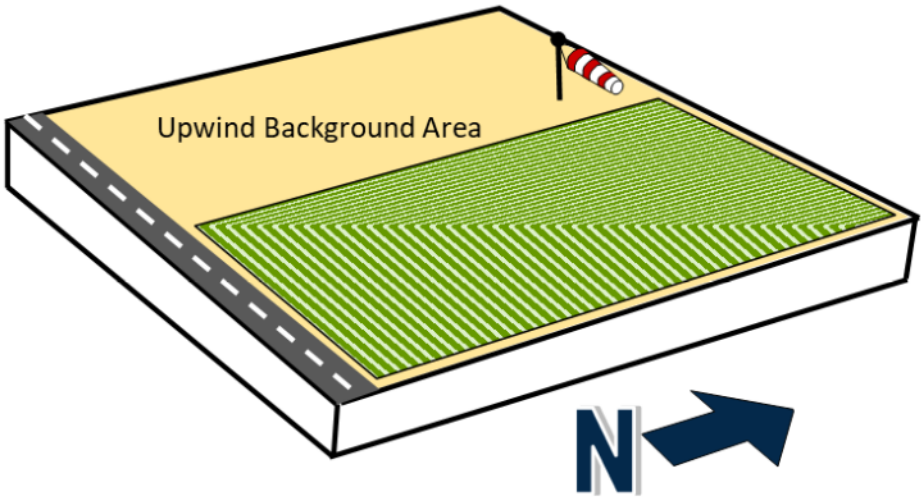


Figure 3-9a. Agricultural field investigation in Example 2A.
Source: ITRC ISM Update Team, 2020.

- *Secondary source areas/release mechanisms*. The broadcast application of pesticides leads to contaminated soils as a secondary source. Secondary releases of COPCs from surface soil can occur from transport of these non-volatile COPCs in surface soil via wind dispersion and plowing of the agricultural field.
- *Exposure media.* The exposure media are limited to surface soil (defined as the top 6 in).
- *Receptors and routes of exposure*. Future residential receptors may be exposed to COPCs in surface soil via incidental ingestion, inhalation of particulates, and dermal contact.

Identify Study Questions – Identify objectives and COPCs. To determine what environmental data are needed to achieve the goals of the ISM investigation, the project team develops the study questions that will guide the sampling and analysis plan in conjunction with the CSM. Example 2A has two study questions; the resulting decision rules are used to develop consensus on ISM results-based actions to help define the data quality needs:

- *Study question 1* – Are the average metals concentrations expressed as the 95% UCL in the agricultural field within ambient background concentrations?
    - Decision rule 1 – *If* the 95% UCL soil metals concentrations are within ambient background concentrations, *then* do not include metals in the risk assessment, *if not*, include metals as COPCs in the quantitative risk assessment.

- *Study question 2* – For each EU in the agricultural field, are the average concentrations expressed as the 95% UCL for each OCP and each metal below risk-based levels of concern?
    - Decision rule 2.1 – *If* OCPs are not detected in surface soil *and* all metals are identified as within ambient background concentrations, *then* no further action, *if not,* proceed to decision rule 2.2.
    - Decision rule 2.2 – *If* replicate RSDs of risk-driving COPCs exceed measurement quality objectives (MQOs), *then* further investigation, *if not,* calculate cumulative risks and hazards *and* proceed to decision rule 2.3.
    - Decision rule 2.3 – *If* cumulative risks and hazards in any EU are above the regulatory acceptable points of departure for risk (1×10^{-6}) and hazard (1.0), *then* for all EUs further action or investigation, *if not*, no further action.

Identify Information Inputs – Specify study goal data needs. Two information inputs are identified for Example 2A.

- Surface soil sampling and analysis are needed for OCPs and metals with detection limits below risk-based screening levels from ¼-acre EUs.
- Define MQOs, particularly the acceptable range of replicate RSDs. For example set 2A, an RSD of less than 20% is established as the MQO.

Define Spatial and Temporal Study Boundaries – Define DUs. The study area's lateral boundary is the 30 acres of agricultural land that is proposed for residential development. The vertical boundary is defined as surface soil based on the release mechanism of broadcast pesticide application and the relatively low mobility of OCP pesticides in soil. Because the study questions are risk-based, the DUs are defined by the anticipated exposure areas, described above as ¼-acre EUs.

Extrapolating to unsampled DUs within a large study area can be achieved in a scientifically defensible manner with ISM.

Extrapolating to unsampled DUs within a large study area can be achieved in a scientifically defensible manner with ISM. Section 3.1.3 and Section 3.1.4 touch on the concept of *sampling a subset of SUs within a large DU*. This concept is applied in this example to extrapolating conclusions from a subset of sampled DUs to a larger group of *CSM-equivalent DUs*, as described in more detail in Section 3.2.8.2. The utility of a pilot study for large areas to assess variability and obtain preliminary COPC concentration ranges is typically very beneficial. In this example, low variance among replicates is anticipated based on the CSM of broadcast pesticide applications, and the 30-acre study area is divided into 120 contiguous equally-sized DUs of a ¼ acre (Figure 3-9b) based on the residential lot size in the area. A subset of DUs can be randomly selected (such as with a random number generator) for sampling. Alternatively, a modified random selection process can be used to ensure that all regions of the 30-acre area are sampled in a proportional manner to reduce the uncertainty from extrapolation if the subset of DUs identified for sampling are grouped too closely together. For modified random selection, the 120 DUs would be allotted into spatial groups and equal numbers of DUs for sampling selected from each group.

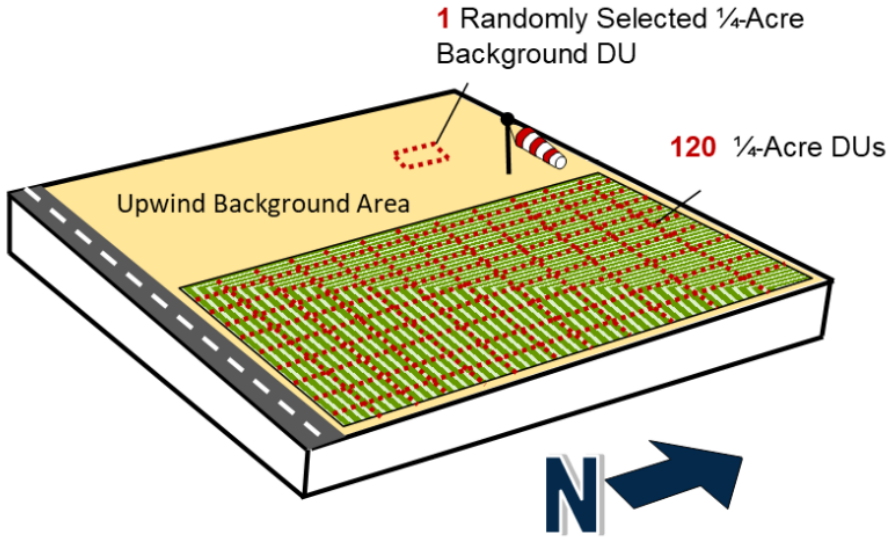


Figure 3-9b. Depiction of DUs in Example 2A.
Source: ITRC ISM Update Team, 2020.

Planning for the number of DUs to sample as the subset of the 120 DUs is a decision that involves all stakeholders and most critically must be sufficient to support the ultimate decisions made based on the extrapolated contaminant average concentration data. Example 2A decisions will be risk-based decisions, and considerations for addressing uncertainty in the risk estimates and risk-based decision errors should err on the side of protecting public health. Generally, practitioners would rather make the mistake of remediating a site that is already clean than make the mistake of not remediating a site that is contaminated.

As presented in Section 3.2.8.2, based on the statistical equations for upper tolerance limits (UTLs) using nonparametric

methods, when there are a large number of DUs (more than 100), a subset of at least 59 DUs must be sampled to conclude that at least 95% of the site area is in compliance with 95% confidence (0.05 = α). From a practical standpoint, confidence in making correct decisions about a large-area site will increase as the proportion of the site area included in ISM sampling increases. Section 3.2.8.2 describes the statistical basis that supports sampling designs that can achieve specified decision error rates, given properties of the data and key assumptions. Based on these numerical simulation studies and statistics commonly applied in environmental investigations, there are conditions when compliance can be achieved by sampling a small portion of the study area (for example, 10% to 30%). The decision regarding the number of small-area DUs to sample should be based on spatial coverage (representativeness) of the site area, the likely degree of variability in soil concentrations across the site area, and the likely proximity of soil concentrations to ALs.

For Example 2A, the project team agrees to determine which DUs to sample using modified random selection and to sample 20 DUs (17% of the total site area, or 5 acres from the 30 acres). This decision to sample 20 DUs by the project team is informed by similar nearby study areas with thorough investigations that had a low CV (<1) and COPC concentrations between 10- and 100-fold lower than risk-based screening levels for all DUs. Therefore, sampling 17% of the site area, or 20 DUs, should be sufficient to avoid a high rate of false compliance decisions while achieving cost savings relative to sampling 59 DUs. Furthermore, the project team agrees that if the CSM assumptions are proven incorrect with either (1) high RSD between replicates or (2) high variability among DUs, then further investigation will ensue with sampling of additional DUs and/or sampling with more increments per DU, rather than extrapolation of the results to the remaining 100 DUs. The anticipated COPC concentrations (that is, 95% UCL to AL ratio of 0.01 to 0.1) plus these two caveats help reduce the uncertainty in drawing conclusions from sampled DUs to other DUs at the site.

Planning by the project team for the number of increments to sample should consider multiple factors. ISM applies soil science and Gy's theory to reduce soil sampling heterogeneity and thereby decrease variability in soil contaminant concentrations among increments and replicates. ISM variability can be reduced by increasing the number of increments collected for an ISM replicate, as described in Section 2. Some of the factors that contribute to heterogeneity in soil contaminant concentrations are taken into consideration when establishing DU boundaries, such as the location of the primary source and secondary or tertiary sources and the soil depths of interest to answering the project team's study questions. Some key factors to consider in deciding on the number of increments in systematic planning are the primary sources and the physical phases of the primary sources (solid or liquid), as well as physical (solid or liquid phase) and chemical properties (water solubility and lipophilicity) that affect COPC fate and transport. The effect of these variables on the heterogeneity in soil contaminant concentrations and the number of increments that should be collected within a DU are illustrated in Figure 3-4. A minimum of 30 increments should be used for each ISM sample or replicate – up to 100 increments may be necessary for some sources/COPCs. The minimum of 30 increments is based on statistical simulations and over a decade of practitioner experience (see Section 2). For certain source types and chemical classes (such as munition residues, metals at small-arms firing ranges, paint chips, ash with dioxins/furans, polynuclear aromatic hydrocarbons [PAHs], and PCBs in transformer oil), more than 30 increments may be necessary due to the highly heterogeneous way these contaminants can be distributed in soil. Case Study 3 in Appendix A demonstrated that 50 increments were insufficient for benzo(a)pyrene (BaP) from a landfill source. Case Study 2 (Clausen et al. 2018a) in Appendix A investigated various numbers of increments (5, 10, 20, 30, 50 100, and 200) to determine how the number of increments affect data quality and concluded ISM samples with 100 increments were appropriate. A field investigation on a diverse set of sources and COPCs from three different sites was undertaken by Brewer et al (Brewer, Peard, and Heskett 2016), which concluded that the magnitude of variability depends in part on the contaminant type and the nature of the release. The sites were (A) a former manufacturer of arsenic-treated ceiling and wall boards, (B) a former municipal incinerator, and (C) a former radio broadcasting station with releases of arsenic, lead, and PCBs in oil, respectively. Variability was well managed for arsenic (site A) and lead (site B), where the use of 54 increments each resulted in RSDs of 6.5% and 20%, respectively. The concentration of PCBs from transformer oil (site C) was so heterogeneous that even in a very small DU, 60 increments were not enough to address the distributional heterogeneity with an RSD of 138%.

For Example 2A, the project team decides to collect three replicates of 50 increments within each of the 20 DUs. Although 30 increments may be sufficient for broadcast application of water-based pesticides that contain arsenic or lead, the use of 50 increments per ISM replicate is decided by the project team to increase the ISM replicate mass because OCPs are hydrophobic (having low water solubility) and have been demonstrated to have relatively large small-scale variability (leading to a higher degree of variability among the increments) at a nearby agricultural field. Table 3-1 presents the variables considered by the project team to determine the number of increments per DU in the agricultural field for Example 2A.

Table 3-1. Variables considered in determining the number of increments per DU for Example 2.

Source: ITRC ISM Update Team, 2020.

Area	Source(s)	COPCs	# Increments	Rationale
Agricultural Field	Pesticides application (lead arsenate)	Arsenic	30	Water-based pesticides
	Pesticides application (OCPs)	OCPs	50	Hydrophobic COPCs
Pesticide Mixing	Spills or ground surface disposal	Full suite of pesticides and petroleum fractions	70	Brewer et al., 2016 (PCBs $n > 60$) $n = 70$ to 100
Residential Area: Current (Example 2B-1)	Paint chips	Metals (lead)	80	Hawaii DOH, 2016 ($n >$ 75)
	Termiticides	OCPs	80	Brewer et al., 2016 (PCBs $n > 60$) $n = 70$ to 100
	Pesticide drift (lead arsenate and OCPs)	Arsenic and OCPs	80	Efficiency of one sampling strategy Unknown heterogeneity ($n = 50$, Hawaii DOH, 2016)
Residential Area: Future (Example 2B-2)	Paint chips	Metals (lead)	80	Hawaii DOH, 2016 ($n >$ 75)
	Termiticides	OCPs	80	Brewer et al., 2016 (PCBs $n > 60$) $n = 70$ to 100
	Pesticide drift (lead arsenate and OCPs)	Arsenic and OCPs	50	Efficiency DUs 1 to 4 for one sampling strategy Unknown heterogeneity ($n = 50$, Hawaii DOH, 2016)
Dump Area Debris	Tires, 55-gallon drums of unknown contents, ash, oil-stained soil, debris	Metals, OCPs, full suite of pesticides, SVOCs, PAHs, dioxins/furans, petroleum fractions	80	Brewer et al., 2016 (PCBs $n > 60$, ash lead $n = 50$ to 60) Sources and COPCs suggest high heterogeneity $n = 70$ to 100

The three replicate locations are established by using systematic random placement as per Section 3.1.5.4 (Figure 3-5), with each DU divided into 50 equally-sized grids. For systematic random selection, after the initial three replicate location is randomly selected, this placement is applied to the remaining 49 grid cells for that DU. If the increments unevenly represent the furrows or crests of the rows, then discuss with the regulatory team the use of a modified systematic random selection process for placement of increments. Similarly, contingencies may be needed if increment locations are inaccessible due to the physical presence of crop vegetation (see Section 4).

Defining background DUs is a concept important for both risk assessment and risk management. Background soil

concentrations for native metals and ubiquitous anthropogenic chemicals such as dioxins and PAHs are often used in risk assessment (see Section 8.4) to establish cleanup goals or verify remedial actions. Background DUs need to be of comparable area and depth (volume) as the site DUs and from a geologically similar area with no known or suspected sources of contamination. ISM background samples need to be of equal sample support – that is, both the field increment volume and number of increments and laboratory subsampling protocols need to match those of the site ISM samples. Ideally, ISM background samples should be comparable to site data and share the attributes of:

- same DU size (volume)
- same sample range of depths
- same soil type (such as sand or loam)
- same volume of soil per increment
- same number of increments and replicates in the DU
- same increment density (such as 30 increments per ½ acre; see Figure 3-10)
- same field methods
- same analytical methods

In situations where nearby background regions are difficult to find, areas equal to the site's DUs from irregularly shaped regions or a combination of discontinuous regions are alternatives (see Figure 3-10 with an example of a background 0.5-acre DU). If different soil types exist within the site, or if soils are derived from different parent material, then multiple ISM background datasets may be needed with one for each soil type.

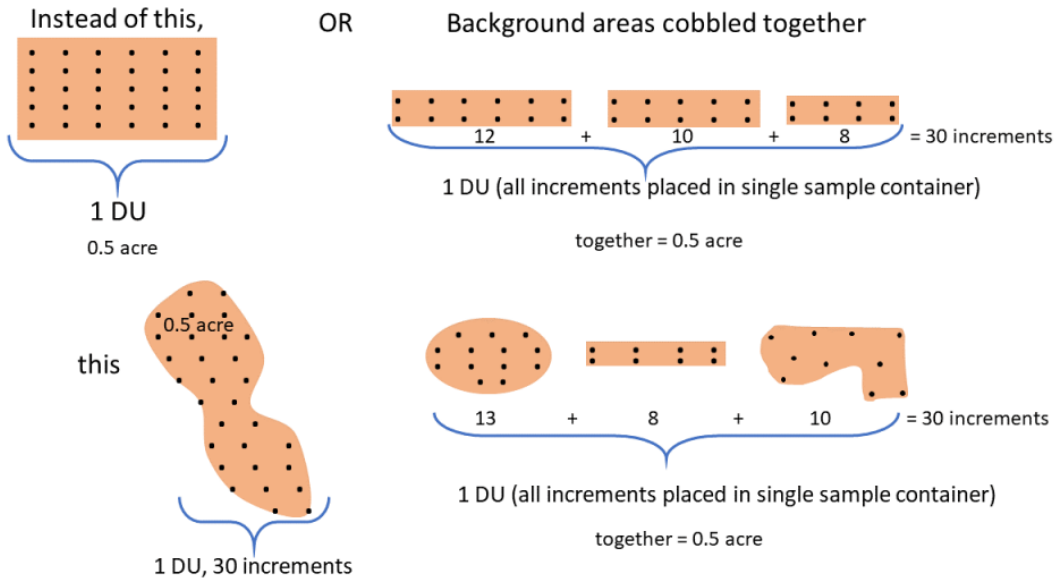


Figure 3-10. Alternative background DUs. When contiguous, regularly shaped, nearby background regions are hard to locate, areas equal to the site's DU from irregularly shaped regions or a combination of discontinuous regions are alternatives.
Source: ITRC ISM Update Team, 2020.

The upwind background area in Figure 3-9a is ideal to avoid any potential cross-contamination from wind deposition during pesticide applications, plowing activities, or windy events. The background sampling must consist of the same number of replicates per DU as the study area, which in Example 2A is three replicates. For Example 2A, one ¼-acre DU is selected from the background area to collect the three replicates of 50 increments each (Figure 3-9b). Similar to the agricultural area DUs described above, the background three replicate locations are established using systematic random placement with the background DU divided into 50 equally-sized areas.

Example 2B. The CSM for Example 2B is a 32-acre parcel proposed for residential redevelopment with a 30-acre agricultural use area, per Example 2A with additional site features and potential source areas (Figure 3-11a). All sources are releases at ground surface. In addition to the legal broadcast application of OCPs and arsenical pesticides, including lead arsenate, the Phase I ESA identified other potential source areas. The parcel had one or more potential pesticide mixing areas, with one currently visible adjacent to an agricultural well that has been present since the 1930s. A rural residence (house) on an acre

of the parcel was built in the 1940s, and there is a children's play area of approximately 0.02 acres (approximately 870 ft^2) that includes a swing set in a bare dirt area. The surrounding agricultural land on the parcel is leased and not farmed by the current occupants. A low-lying area exists in the northeast of the parcel that contains debris from miscellaneous illegal dumping of tires, multiple 55-gallon drums of unknown contents, ash, and oil-stained surface soil. The dump area is approximately 1 acre. There have been no investigations of similar nearby study areas to inform the team on expected COPC concentrations.

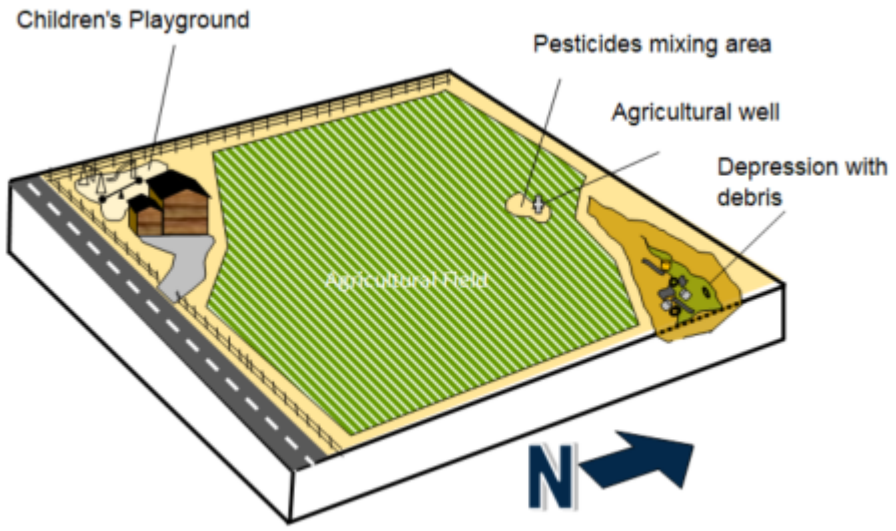


Figure 3-11a. Current residential scenario in Example 2B.
Source: ITRC ISM Update Team, 2020.

Agricultural mixing areas are potential sources of localized heavy contamination of OCPs, arsenic, and lead. Conditions of potential contaminant distribution in pesticide mixing are much more uncertain than in other areas of the agricultural field. In addition to direct exposure hazards, contamination could also pose leaching hazards and subsequent contamination of underlying groundwater resources. Triazine herbicides, arsenic, and other chemicals can pose additional leaching threats to underlying groundwater. While ISM is useful for determining average concentrations of COPCs over a volume of soil and modeling for leaching, it is beyond the scope of this example. Pesticide mixing areas should include the full range of pesticides as COPCs and petroleum fractions.

Potential sources of chemical contamination surrounding the residential structure are lead-based paint and organochlorine termiticide applications around the base of the foundation. The debris/dump area COPCs are SVOCs, including PAHs, metals, dioxins/furans, and petroleum fractions.

Preliminary data have shown site soil arsenic and lead concentrations are above ambient background soil concentrations. Thus, comparison to background is unwarranted as both arsenic and lead will be carried through the risk assessment as COPCs. Although PAHs are ubiquitous in the environment, the ambient concentrations of PAHs are expected to be very low and have no effect on the risk in Example 2B due to the rural location of the site and lack of generalized local source dispersion. Applicable background data for site-related organic chemicals are unavailable, and the project team has decided to review initial risk results to support whether future background studies for other analytes are consequential.

The project team agrees to a phased approach, with the Phase I investigation limited to surface soils based on the sources of release all being to ground surface. The team agrees that if cumulative risks and hazards in any EU are above the regulatory acceptable points of departure for risk (1×10^{-6}) and hazard (1.0), then a Phase II investigation will ensue. The remainder of Example 2B focuses on unique aspects not included in Example 2A and is largely directed to defining the EUs.

Problem Formulation – Identify decisions needed and develop CSM. The goal of the ISM sampling event is the same as in Example 2A, to determine the average concentrations of COPCs expressed as the 95% UCL in surface soil for use to assess potential current and/or future residential risks and ascertain if cumulative risks or hazards exceed the regulatory acceptable points of departure of 1×10^{-6} and 1, respectively. All sources are releases at ground surface, and the decision rules specify additional investigation if surface soil risks are above regulatory thresholds. COPCs in Example 2B include OCPs, the full suite

of pesticides, arsenic, lead, SVOCs, PAHs, dioxins/furans, and petroleum fractions.

The Example 2B preliminary CSM includes the aspects from Example 2A's agricultural field. The exposure media as well as the receptors and routes of exposure are the same as Example 2A.

- *Primary source areas/release mechanisms.* There are four primary source areas/release mechanisms:
  - *Agricultural field*. This potential primary source release mechanism was the legal broadcast application of OCPs and arsenical pesticides, including lead arsenate.
  - *Pesticide mixing area*. There is one known pesticide mixing area where the potential release mechanism included spills or washout from sprayer equipment that was disposed to ground surface.
  - *Rural residence*. Lead-based paint on the exterior of the home was a potential primary source with a potential release mechanism of paint chips dislodged from the exterior's painted surfaces. OCPs for termite treatment is another potential primary source, with the potential release mechanism the application of termiticides to the soil around the base of the residence.
  - *Debris/dump area*: The potential primary sources are the contents deposited to this area, which include debris, tires, multiple 55-gallon drums of unknown contents, ash, and oil-stained surface soil. Potential release mechanisms are degradation of the debris, tires, and 55-gallon drums, which have rusted holes. It is unknown whether the 55-gallon drums were entirely empty or not when they were dumped here.
- *Secondary source areas/release mechanisms*. All the above primary sources can lead to secondary sources of contamination. Secondary source releases of COPCs from surface soil can occur from transport of these non-volatile COPCs in surface soil particulates via wind dispersion, windy events, and plowing of the agricultural field. Runoff from irrigation or rain events is another potential release mechanism, particularly to and within the debris/dump area that is at a lower elevation.
- *Exposure media.* The exposure media are limited to surface soil. The physical and chemical properties of the COPCs indicate low mobility in soil in the absence of co-contamination and co-migration with a solvent such as petroleum fractions or distillates. Surface soil is defined as the top 6 inches in the agricultural field where plowing has mixed the soil and the top few centimeters elsewhere at the site.
- *Receptors and routes of exposure.* Current and future residential receptors may be exposed to COPCs in surface soil via incidental ingestion, inhalation of particulates, and dermal contact.

Identify Study Questions – Identify objectives and COPCs. Example 2B has one study question and two decision rules:

- *Study question 1* – For each EU, are the average concentrations for each COPC expressed as the 95% UCL below risk-based levels of concern?
  - Decision Rule 1.1 – *If* any COPCs are detected in surface soil in an EU, *and* replicate RSDs exceed the DQOs, *then* further investigation, *if not*, calculate cumulative risks and hazards for that EU *and* proceed to decision rule 1.2.
  - Decision rule 1.2 – *If* cumulative risks and hazards in any EU are above the regulatory acceptable points of departure for risk (1×10^{-6}) and hazard (1.0), *then* further action or investigation in that EU, *and if* the EU is in the agricultural field, expand sampling to other EUs, *if not,* no further action.

*Identify Information Inputs – Specify study goal data needs.*Two information inputs are identified for Example 2A.

- Surface soil sampling and analysis are needed from each EU for all COPCs using detection limits below risk-based screening levels.
- Define the DQOs, particularly the acceptable range of replicate RSD. For Example 2B, an RSD of less than 20% is established in the DQOs as the MQO.

Define Spatial and Temporal Study Boundaries – Define DUs,

- The Current scenario (Example 2B-1) study boundaries are the lateral expanses of the 1-acre rural residential property, with the goal to determine the average concentrations of COPCs expressed as the 95% UCL in surface soil for use in assessing potential residential risks, the Example 2B-1. The vertical boundary for this Phase I investigation is limited to surface soil based on the sources of release all being to ground surface. Because the study questions are risk-based, the DUs are defined by the current EUs. Figure 3-11b shows the SUs and EUs for

the current scenario. Adult residents are equally likely to contact any area of the residential property. Because the children's outdoor exposures to soil are expected to be focused within the play area, this area is designated an EU for children (EU_C).. Four SUs are established around the house to sample for potential lead and OCP contamination in soil. (Note that these four SUs could also be used as *source area N&E DUs* for N&E study questions.) *Multiple SUs within a DU* forms the adult residents' EU (EU_A) which consists of SU1 (play area), SUs 2 through 5 (perimeter base of the house), and SU6 (remainder of the residential acre).

Heterogeneity is expected to be large from paint chip nuggets and OCPs that are hydrophobic. From each of SUs 1 through 5, the project team decides 80 increments and three replicates will be collected for metals and OCPs analyses. The project team considers collecting 40 increments for arsenic analysis, ultimately concluding that it is more cost-effective to employ the same sampling strategy for all COPCs. The three replicate locations are established by using systematic random placement, as per Section 3.1.5.4 (Figure 3-5) within each SU that is divided into 80 equally-sized grids. For modified random selection, after the three replicate locations are randomly selected, if they are too close relative to the entire grid area, then additional random selection of a replacement location for one of the replicates is performed to modify the sampling plan.

The ISM data are used to calculate *area-weighted EPCs* for the EU_A (see Section 1 for where 95% UCL is defined, Section 3.2 for discussion on 95% UCL, and Section 3.3.2 and Section 6.2 for weighted means and weighted 95% UCL).

Figure 3-11b. Selection of SUs and EU for risk to human health in Example 2B-1.
Source: ITRC ISM Update Team, 2020.

- The Future scenario (Example 2B-2) study lateral boundaries include the 32-acre parcel that is proposed for future residential development. The vertical boundary for this Phase I investigation is limited to surface soil based on the sources of release all being to ground surface. Because the study questions are risk-based, the DUs are defined by the EUs. Figure 3-11c1 through Figure 3-11c3 illustrate the DUs for Example 2B-2. Table 3-1 presents the variables considered by the project team to determine the number of increments per DU for Example 2B.

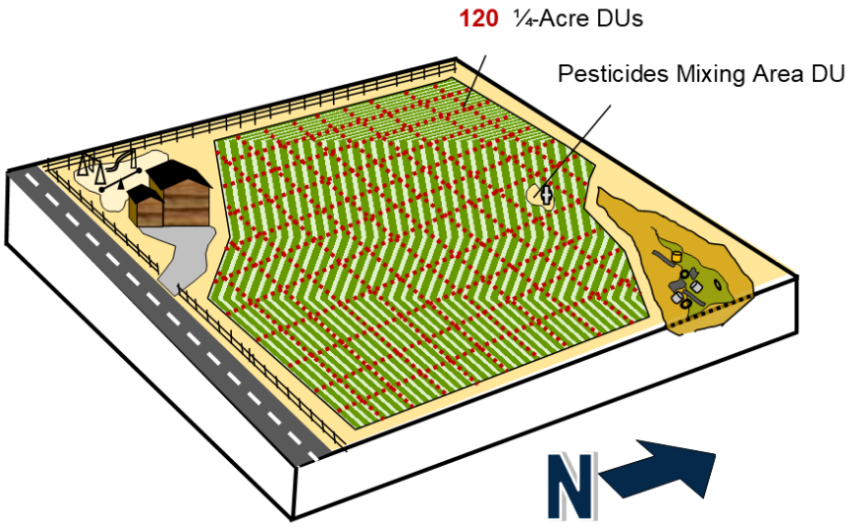


Figure 3-11c1. Future residential scenario in Example 2B-2.
Source: ITRC ISM Update Team, 2020.

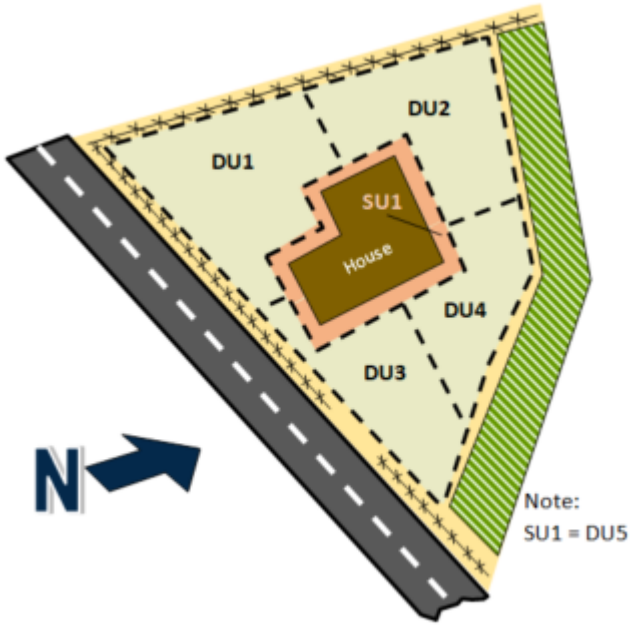


Figure 3-11c2. Selection of DUs and SUs for future residential scenario in Example 2B-2.
Source: ITRC ISM Update Team, 2020.

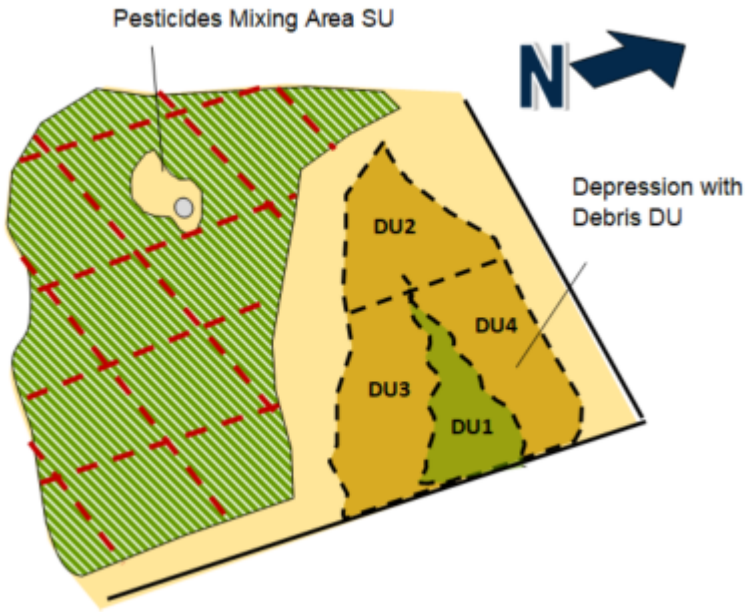


Figure 3-11c3. Selection of DUs for future residential scenario at surface depression in Example 2B-2.
Source: ITRC ISM Update Team, 2020.

- Agricultural field area DUs. As with Example 2A, statistically 59 of the 120 DUs – or 14.75 out of 120 acres – are recommended for the DU subset sampling. An N&E DU is carved out to encompass the pesticide mixing and agricultural well area, thus the total agricultural area DU is 119 acres. The project team agrees to use a modified random selection of 59 DUs for sampling and extrapolating the remaining 60 DUs. Furthermore, the project team agrees that if the CSM assumptions are proven incorrect with high RSDs between replicates and/or high variability between DUs, then further investigation will ensue with more sampling of DUs and/or more increments per DU, rather than extrapolation of the results to the remaining 60 DUs. In addition, the project team agrees that if cumulative risks and hazards in any of the 59 DUs are above the regulatory acceptable points of departure for risk (1×10^{-6}) and hazard (1.0), then further investigation of additional DUs will follow in a subsequent phase of investigation.
  The project team decides each of the 59 DUs will have three replicates of 50 increments by using systematic random sampling locations for the three replicates as described above with analyses for OCPs and metals (see Table 3-1). An alternative that the project team did not choose was to first collect and then analyze the replicates from a portion of the 59 DUs to provide an early indication of whether the ISM sampling design will be successful or if alterations to the sampling design are needed, such as increasing the number of increments. As with Example 2A, the team decides to use a modified systematic random selection process for placement of increments and to plan for contingencies if increment locations are inaccessible.
  Note that a similar approach could be taken if commercial/industrial redevelopment were proposed. Dividing the 30 acres into 30 EUs of 1 acre each and either sampling all 30 EUs or sampling a subset are potential sampling strategies, again using three replicates of 50 increments with analyses for OCPs and metals. EPCs can be calculated because the three replicates are collected from each of the sampled EUs. Section 3.2.8.2 describes the statistical approach for determining the subset number of EUs that should be sampled for 95% power and 95% confidence (0.05 = α), that is, to conclude that at least 95% of all the EUs within the entire area are in compliance with 95% confidence (0.05 = α). If one or more EU results "fail" (exceed the regulatory risk benchmark), we would conclude that the entire EU fails the risk assessment.

- Pesticide mixing area. The size of this area is less than a ¼ acre, so it is assumed that the contamination from the pesticide mixing area is limited to within the ¼-acre area without areas of highly concentrated subareas of contamination into adjacent DUs – that is, there is no migration of contaminants into adjacent DUs. The area is approximately 30 ft by 30 ft, so for simplicity, the project team agrees that the most effective use of resources is to designate a source area SU within this ¼-acre EU (Figure 3-11c3) because of the uncertainty of the yard locations for future residential development. The SU sampling proposed is three replicates of 70 increments

each, with analysis for the full suite of pesticides and petroleum fractions. This decision is based on empirical findings demonstrating the inadequacy of 60 increments for PCB releases (Brewer, Peard, and Heskett 2016). The project team agrees to use the ISM SU concentration data as surrogate EU concentration data, rather than sampling outside the SU within the EU and applying a weighted averaging approach (see Section 6.2). The three replicate locations are established by using systematic random placement as per Section 3.1.5.4 (Figure 3-5) within the SU that is divided into 70 equally-sized areas/grids.

- Rural residence. The one potential source area in the house perimeter with COPCs of lead and OCPs is designated as SU1 and extends from the foundation to 5 ft out from the house. SU1 encompasses the drip-line areas of the house, which is where the potentially elevated levels of lead are expected in the soil. The remainder of the current residential area is subdivided into four DUs of approximately ¼ acre each. With the uncertainty of the locations of the yard locations for future residential development, the project team decides that the most effective use of resources is to consider the source area SU as DU5 for the future residential risk assessment. Note that the source area SU could also be used as a DU for N&E; these DUs are depicted in Figure 3-11c2.
  DUs 1 through 4 are designed to collect three replicates of 50 increments each for metals and OCPs. DU5, with potential paint chips and OCP termiticide application, is designed for collection of three replicates of 80 increments each, with analyses for metals and OCPs. This decision is based on the high heterogeneity in soil lead from lead-based paint chips and the HDOH (HDOH 2016a) recommendation to use 75 increments or more when the source is paint chips, plus the recommendation of 70 to 100 increments due to the empirical findings demonstrating the insufficiency of 60 increments for PCB releases (Brewer, Peard, and Heskett 2016). See Table 3-1 for the variables considered by the project team to determine the number of increments per DU for DUs 1 – 4 at the Rural Residence. Within each SU, the three replicate locations are established by using systematic random placement as per Section 3.1.5.4 (Figure 3-5), with DUs 1 through 4 divided into 50 equally-sized areas/grids and DU 5 into 80 equally-sized areas/grids.

- Debris/dump area. Due to the nature of the area with its slightly sloped sides and a collection of debris in the deepest portion, the 1-acre dump area is divided by the project team into four DUs as depicted in Figure 3-11c3. DU1 is the deepest portion with the anticipated highest potential contaminant concentrations, and DUs 2, 3, and 4 are the sloped sides.
  The dump contents are documented with photographs and physical marking of locations and maintained in study records in the event that further action or investigation is the outcome of decision rule 1.2. The sources and physical/chemical properties of the chemicals suggest potentially high heterogeneity in soil concentrations. Each DU is designed to collect three replicates of 80 increments each, with analysis for OCPs, the full suite of pesticides, arsenic, lead, SVOCs, PAHs, dioxins/furans, and petroleum fractions. Table 3-1 has the rationale used and factors considered by the project team to determine the number of increments per DU for the Debris/Dump Area. The team also used the Hawaii DOH recommendation to use 75 increments or more when the source is paint chips, the recommendation of 70 to 100 increments due to empirical findings demonstrating insufficiency of 60 increments for PCB releases (Brewer, Peard, and Heskett 2016), and the suspected high heterogeneity due to sources and COPCs in this area. Within each DU, the three replicate locations are established by using systematic random placement as per Section 3.1.5.4 (Figure 3-5), with each DU divided into 80 equally-sized areas/grids.

3.1.6.3 Example 3: former industrial site

Example 3 features four major components of systematic planning:

- defining DUs for N&E delineation purposes (source area DUs)
- defining DUs for estimating EPCs for human receptors (Human Health [HH] DUs)
- defining DUs for estimating EPCs for ecological receptors (Eco DUs)
- integrating the sampling needs of all three into one sampling design

It addresses six data collection objectives:

- estimating average concentrations in a defined volume of soil or sediment

- evaluating the horizontal profile of contamination in sediment or soil
- evaluating the vertical profile of contamination in soil or sediment
- evaluating the horizontal extent of contamination along a drainage
- evaluating EPCs for human receptors
- evaluating EPCs for ecological receptors

Current conditions. The former industrial site (Figure 3-12a) currently consists of a 5-acre tract of land within a fenced area. The fencing follows the original property line and has "No Trespassing" signs posted at regular intervals along its length. The entrance to the property is gated and only opened when customers want to access one of the repurposed storage units on the northern portion of the property. The southern portion of the property is currently a large grassy field that is mowed and maintained regularly by the property owner.

To the north of the property, there is a wooded area that runs northward up a sloped area and over the crest of a hill; to the south is other light commercial land; to the west is an interstate highway; and to the east is a natural grassy meadow area that extends to the edge of a wetland. The grassy meadow and wetlands are part of a state wildlife management area, but there are no rare, threatened, or endangered species, species of special concern, or listed species present.

The property is located in an area that is now a rapidly expanding light industrial/commercial area, and the property owner wants a reuse plan for the land that can be quickly implemented. The decision has been made to turn the maintained open grassy area into a surplus/long-term open storage lot for containerized goods and/or vehicles. The owner knows that no buildings will be erected but has not yet decided whether the land will be simply shaped, graded, and paved prior to use. They are expecting to put in a subsurface drainage system in the current maintained grass area to ensure the lot and any containers or vehicles stored there will not be susceptible to flooding.

Figure 3-12a. Current conditions of former industrial site in Example 3.
Source: ITRC ISM Update Team, 2020.

Historical site use. Historically, the industrial site was a redistribution center for a paint manufacturer (Figure 3-12b). Large truck shipments came from manufacturing centers across the country, and their products were off loaded, stored, and then loaded back onto trucks for local and regional redistribution to commercial outlets. No manufacturing or repackaging occurred on the property; products were only off loaded, stored, and then loaded and shipped. However, the product was sometimes damaged, and along with other facility trash, it was dumped in a small, private landfill on the southeast corner of the property (Figure 3-12b).

In addition to the storage units that are still standing and in use today, the site also had an administrative building with an adjacent employee parking lot and a redistribution warehouse with an indoor handling area as well as loading docks. Two roadways connected the areas on the site: the main site road that started at the western (main) entrance gate off the highway and extended to the turnaround area on the eastern part of the property, and a secondary access road by the former employee entrance on the southern property boundary. The main access road has drainage ditches on either side of it as well as around the turnaround area. The sloping topography from north to south facilitates drainage and overland

transport of soil fines, so although these ditches only hold water for brief periods after rain events, the main road ditch system did and continues to require regular maintenance to maintain open drainage. The secondary road leading to the loading docks and the employee parking lot did not need drainage when the facility was operating, but water has been seen to collect in this part of the property in recent years.

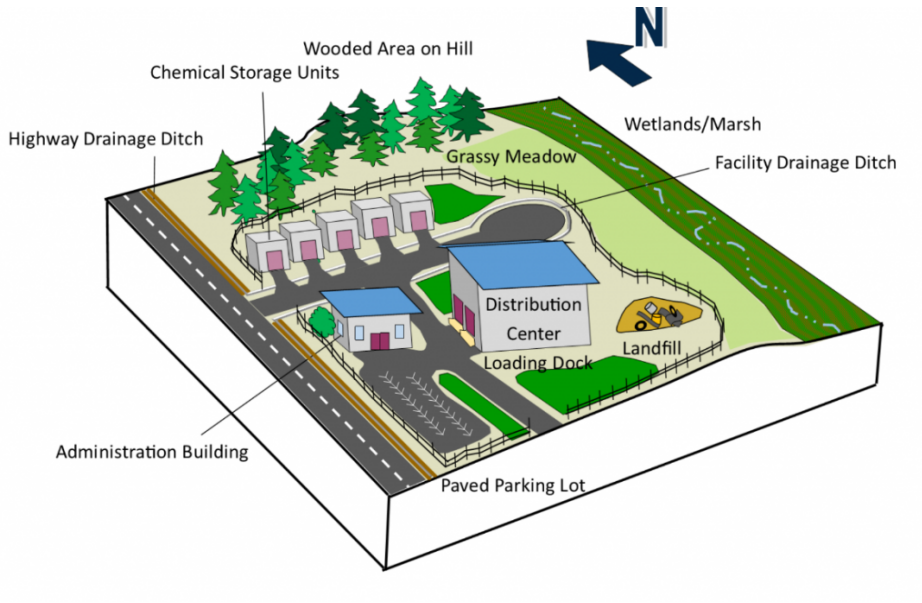


Figure 3-12b. Historical use of the former industrial site in Example 3.
Source: ITRC ISM Update Team, 2020.

DQOs. The following example shows one possible route a site investigation could take. The approach presented here is not intended to portray the only approach to take – rather, it will walk the reader through the logic that the DQO process could take to illustrate how ISM sampling can meet more than one end-use objective.

DQO step 1: problem formulation. Preliminary problem formulation requires the consideration of a sequence of steps:

1. identification of potential primary source areas/release mechanisms
2. identification of potential secondary source areas/release mechanisms
3. identification of media that could be impacted by such a release/migration (exposure media)
4. identification of receptors, both human and ecological, that could come into contact with these contaminated media (potentially completed pathway)

- First, generate the preliminary CSM and potentially complete exposure pathways to establish DUs.
- *Primary source areas/release mechanisms*. Based on past site usage, it has been agreed that there are two potential primary source/release areas: one, the portion of the secondary road in the vicinity of the former loading/unloading docks where material could have been spilled during loading/unloading (Figure 3-13a, DUs 1 through 5), and two, the former landfill where wastes of various kinds were buried and could infiltrate/percolate the soil column (Figure 3-13a, DUs 6 through 13).

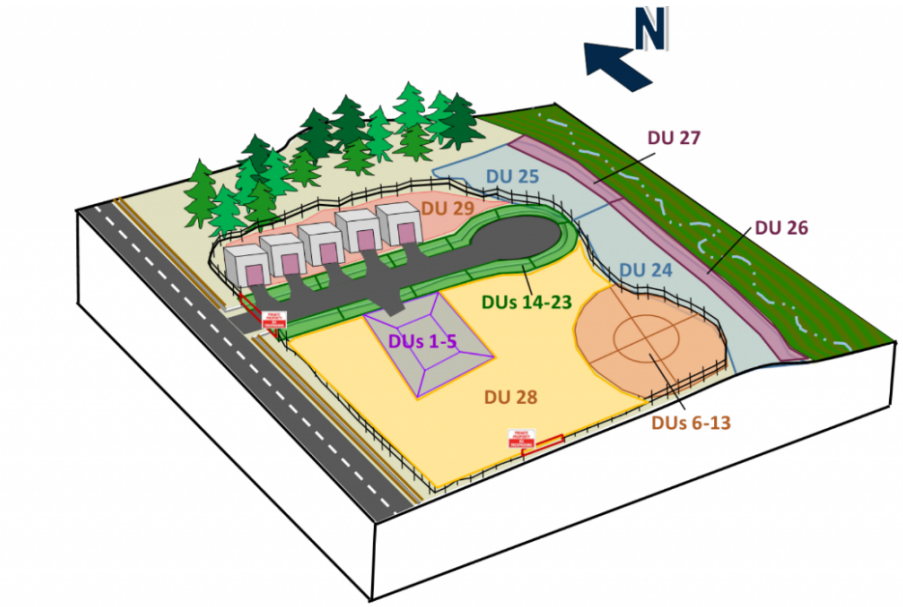


Figure 3-13a. DUs to define N&E in Example 3.
Source: ITRC ISM Update Team, 2020.

- *Secondary source areas/release mechanisms*. The primary release mechanisms in both instances leads to contaminated soil as a secondary source. These contaminated soils can, via a secondary release mechanism, release COPCs through overland transport of fines and particulate (horizontal transport), infiltration/percolation (vertical transport), and dust and/or volatile emissions (horizontal transport). These secondary source areas then become exposure media.
- *Exposure media*. Although some shaping and grading of the site has occurred over the years to direct rainwater to drainage ditches, the topography of the site is essentially level. For this reason, the drainage ditches along the main roadway and the turning area are considered to be potentially impacted. However, due to the hill on the northern border of the property and the wetland on the eastern border, COPCs could also potentially move westward via overland transport and infiltrate/percolate. As a result of this and the proximity of the former landfill to the grassy meadow and wetland, soils/sediments in these areas could impact media differently from the northern shore area. Based on the topography of the site, it is expected that dust and/or volatile emissions could result in widespread and more diffuse impacts. All these successive layers of systematic planning are presented in the final figure for each group of site receptors.
- *Human receptors*. Four human receptors have been identified: industrial/commercial workers who could contact soils anywhere within the property boundaries, as well as the ditches; construction workers who would encounter soils in the maintained portion of the property south of the main road; trespassers who would encounter soils in the grassy meadow east of the property; and trespassers who would encounter sediments along the shoreline the wetlands east of the property (Figure 3-13b).

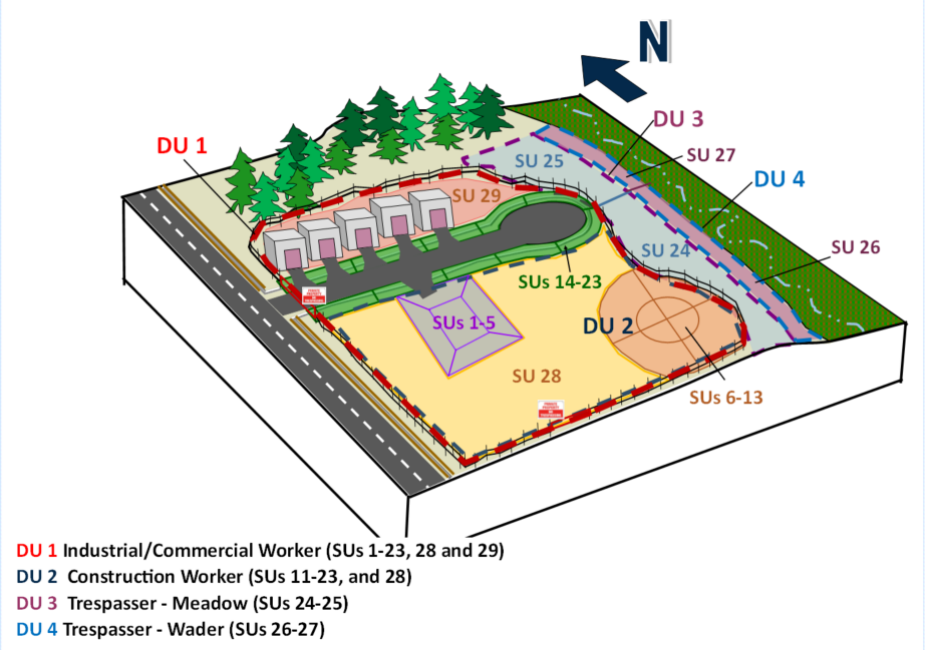


Figure 3-13b. SUs and DUs to define exposure to human receptors in Example 3.
Source: ITRC ISM Update Team, 2020.

- *Ecological receptors*. Two categories of ecological receptors have been identified: terrestrial receptors that would use the grassy meadow, and two aquatic receptors that would use the aquatic habitat. A review of representative receptor categories for the site indicates that all categories of representative receptors would utilize the entire area of the two identified habitats east of the property (Figure 3-11c).

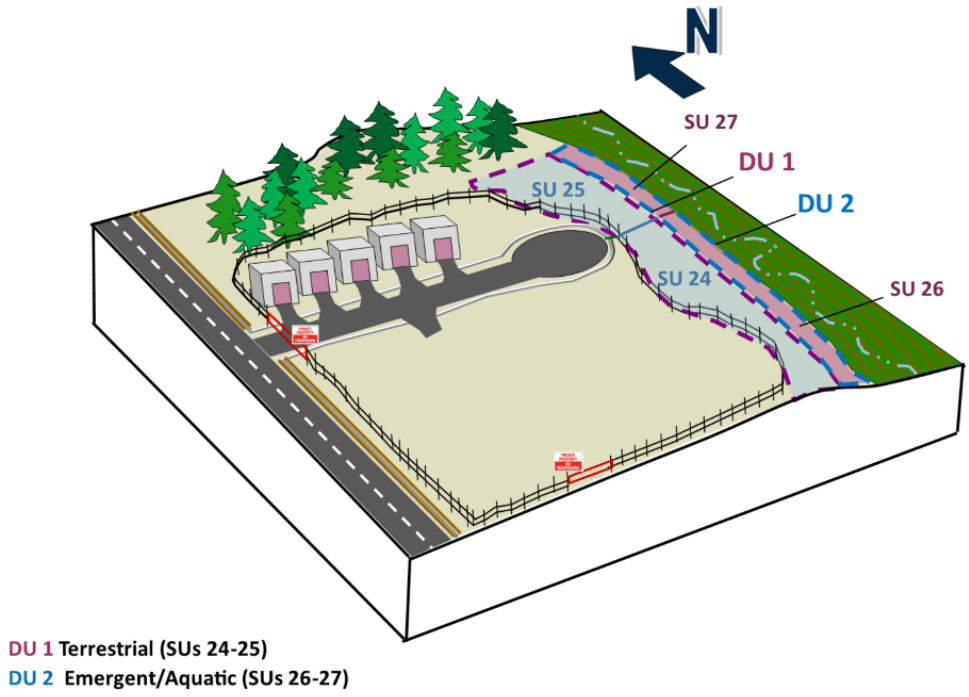


Figure 3-13c. SUs and DUs to define exposure to ecological receptors in Example 3.
Source: ITRC ISM Update Team, 2020.

All these components of problem formulation can be summarized in a CSM (see Figure 3-14a and Figure 3-14b).

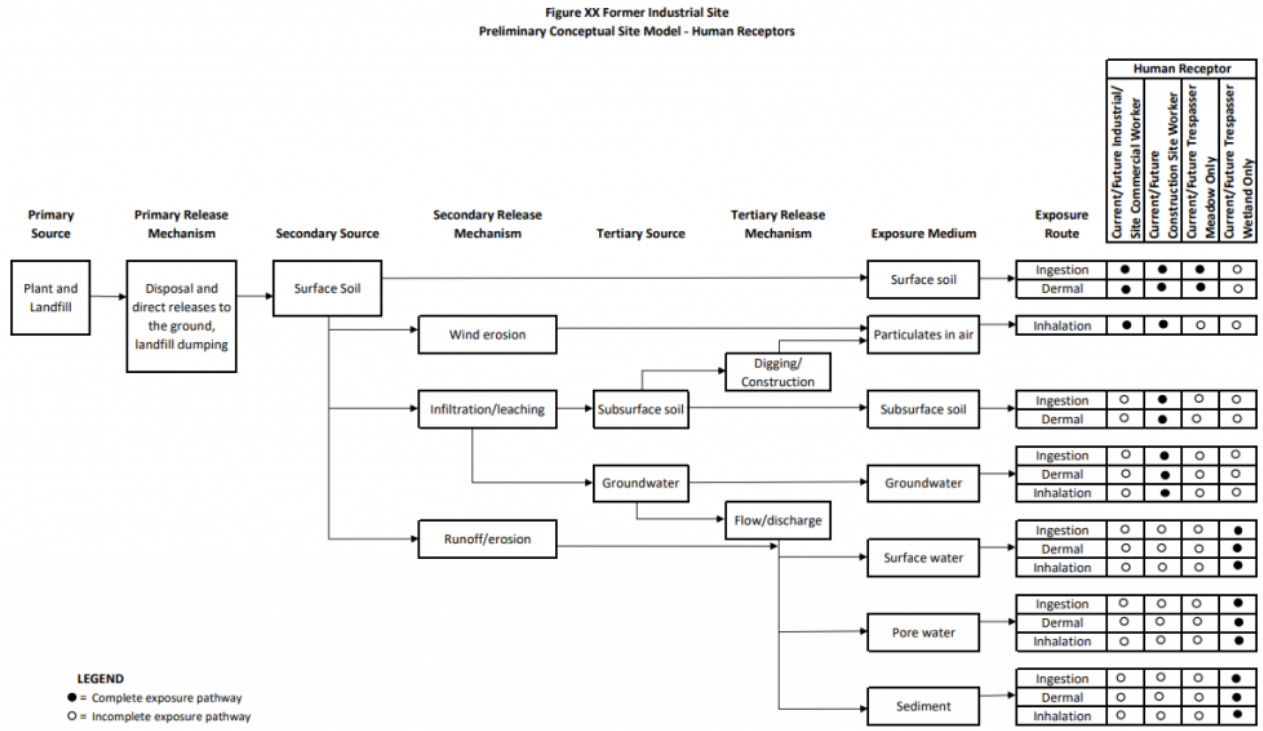


Figure 3-14a. Preliminary CSM of human receptors on a former industrial site in Example 3.
Source: ITRC ISM Update Team, 2020.

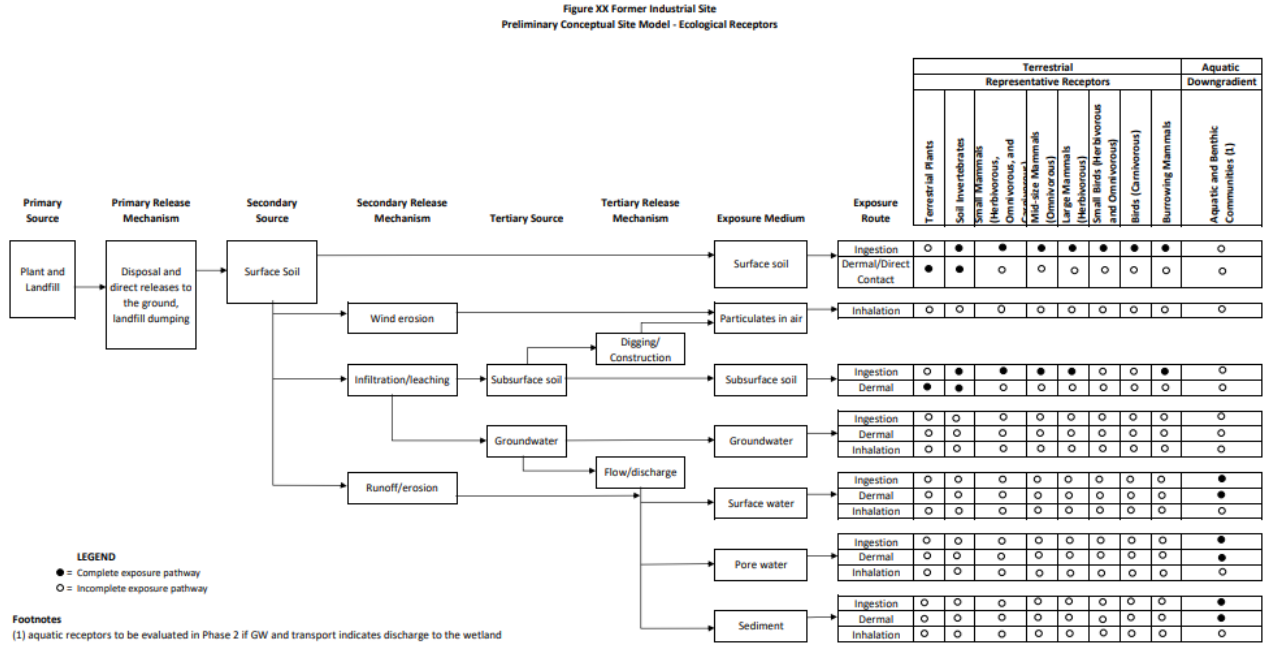


Figure 3-14b. Preliminary CSM of ecological receptors on a former industrial site in Example 3.
Source: ITRC ISM Update Team, 2020.

DQO step 2: identifying study goals.Repurposing this site cannot occur until the site investigation questions are answered. Using the CSM as the starting point for systematic planning and DU design, the project team needs to identify study goals to ensure the data collection will generate data of sufficient quality and quantity to achieve the study goals.

In this case, three study goals have been identified:

- Has the N&E of site-related releases been fully characterized?
- If human or ecological receptors are exposed to site-related COPCs via contaminated soil/sediment, are they exposed at levels that present risk?
- Ultimately, if risk is present, what mitigation measures might be needed?

It is not always possible to plan the full timeline of the investigation/remediation, but consideration should be given to the entire continuum of activities through the involvement of the appropriate technical specialists who will be involved over the duration of the project. Often, minor modifications early in the development of sampling and analysis plans/QAPPs can save time and money at a later stage of the project by removing the need to remobilize to collect similar but not identical data. This planning can contribute data to both the investigation/evaluation phase as well as the remedial measures phase.

DQO step 3: identifying information inputs. As the sampling and analysis plan/QAPP is developed, the project team should evaluate what specific information is needed to evaluate the study goals. In particular, when developing an ISM sampling approach, they should consider the specific purpose of each piece of data collected. ISM samples can be collected from DUs or SUs, but the assignment of an ISM sampling area to one category or the other depends on the use to which it will be put.

Recall from earlier text that both SUs and DUs are separately collected samples, with SUs being subdivisions of DUs. Simply put, if an ISM sample is used as a stand-alone sample to make a decision, it is a DU; if the ISM sample is used in conjunction with another ISM sample(s) to make a decision, it is an SU.

In the example with three study goals (DQO step 2), rather than trying to develop an integrated plan that lays out all SUs and DUs simultaneously, it is strongly advised that the sampling requirements of each study goal be worked out systematically by each technical specialty within the project team, then the team can integrate the various sampling needs into one sampling plan. As will become apparent in this example, systematic planning can define both SUs and DUs and allows for the contingency that a DU for one study goal may become an SU for another study goal.

DQO step 4: define study boundaries. The primary questions to answer when defining study boundaries are (1) "Where is the contamination located, and are there spatial patterns?", and (2) "What is the average concentration that needs to be defined for each study goal?" Answering these questions requires translating the components of the CSM into a sampling plan where the number, location, and dimensions of SUs and DUs can achieve the study goals (DQO step 2) and support site decisions.

- *Study boundaries related to estimating average soil concentrations*. Two types of average soil concentrations are needed to achieve the study goals for the former industrial site: average concentrations to define source areas and average concentrations to define exposure to human and ecological receptors. To collect meaningful data for source areas, the project team can use historical site information, aerial photographs, satellite imagery, and other historical sources to help define the areas that need to be investigated. For exposure estimates, the CSM must be developed (DQO step 1) and integrated with the resource/habitat characteristics on and within the potentially impacted site.
    - *Source area DUs*. Figure 3-13a shows both primary and secondary source DUs, with primary source area N&E DUs 1 through 5 near the former loading dock area and N&E DUs 6 through 13 in the former landfill area. Both sets have a central DU placed where the highest levels of potential releases may be, with the DUs around it determining if the central source area was targeted correctly and indicating how much the original release may have spread.

For this property, secondary source areas are adjacent but more spread out – for example, the drainage ditches (N&E DUs 14 through 23), the grassy meadow (N&E DUs 24 and 25), and the wetlands (N&E DUs 26 and 27) (Figure 3-13a). The drainage ditches were divided into 10 DUs to allow for various transport mechanisms to be evaluated. If the ditches show minimal silting and that concentrations within any particular DU are low, the sampling allows the option of deciding to only clean out silted in areas in a subset of the DUs. Along similar lines, each habitat zone has two DUs to allow the area near the landfill to be compared to the portion of the habitat zone more distant from the landfill, which should provide insights into the fate and transport from the former landfill.

- *Exposure area DUs for human receptors (HH DUs). Figure 3-13b shows DUs associated with areas where humans may reasonably expect to come into contact with site soils and sediment on or adjacent to the site currently or in the future.*

Disregarding the SUs in these figures. Each human receptor has a distinct exposure DU based on (1) an ongoing activity (industrial/commercial worker over the entire property, human health (HH) DU1 in the figure outlined in red dashed lines); (2) a reasonably anticipated future role (construction worker in the portion of the property south of the main access road, HH DU2 outlined in blue dashed lines); or (3) a transient current or future trespasser exposure in the grassy meadow (shown as HH DU3 outlined in dashed purple lines) and the wetland shoreline (HH DU4 outlined in dashed light blue lines).

- *Exposure DUs for ecological receptors (Eco DUs)*. Figure 3-13c shows DUs associated with areas where ecological receptors may reasonably expect to come into contact with soils and sediment on or adjacent to the site currently or in the future.

For this particular site, the ecological receptors are limited because of the limited habitat. The maintained grassy area with the property fence is not considered an ecological resource because it is artificially maintained, however, the areas east of the site are. The grassy meadow that is used to assess potential exposure to terrestrial receptors, Eco DU1 (outlined in dashed purple lines), is the same as HH DU3 for the grassy meadow trespasser. Similarly, Eco DU2 (outlined in dashed light blue lines) is the same as HH DU4.

- Study boundaries related to evaluating the N&E of contamination. N&E must be evaluated both vertically and horizontally.
    - *Evaluation of the vertical extent of contamination*. Three major factors should be considered when considering the vertical extent of contamination: What was the nature of the COPCs? At what depth were they introduced? What depth horizons are required to evaluate exposure for current or reasonably anticipated future receptors? Knowing the nature of the COPCs will allow advance evaluation of the fate and transport of the release. Surface releases may not migrate through the soil column, while containers buried at 10 ft below ground surface that leak starting at depths of 10 ft present a very different scenario for assessing vertical migration. Lastly, along with these two considerations, the depth horizons that need to be sampled to evaluate exposure to human and ecological receptors must also be integrated.

In the secondary source areas (drainage ditch, grassy meadow, and wetland shoreline), spills where potential contamination is expected to be less than in the primary release areas might have you deciding to limit sampling to only surficial soils for N&E and only sampling deeper if the exposure horizons warrant it or surficial soils show elevated levels of site-related COPCs.

- *Evaluation of the lateral extent of contamination.* Lateral extent is handled immediately around the source area DUs with a peripheral ring of DUs that serves two purposes: (1) help to bound the source area, particularly if the placement of the source area had some uncertainty associated with it, and (2) provide insights into possible lateral migration.
    - *Evaluation for the presence of subareas of elevated contamination*. Vertical soil horizons and multiple adjacent DUs commensurate in size with the site history and suspected release allow the team to address the presence of subareas of elevated contamination. Large DUs may help the team assess exposure but run the risk of overlooking subareas where COPC concentrations could be elevated. The needs of N&E DUs and exposure must be balanced. If smaller areas are needed to delineate N&E, they can be reassigned to SUs for the evaluation of exposure; an area-weighted average can then be used to create an EPC for the purposes of evaluating risk.

Integrating the sampling needs for N&E with exposure. Once sampling plans for each study goal are worked out, they should be *integrated into one single project sampling plan*, and the overlap – as well as the differences – between each plan should be understood by the project team. Minor changes in sampling can take data that are useful for only one study goal and make them useful for two or more other goals. Care should always be taken to recognize when a DU for one study goal can serve as an SU for another study goal.

Looking at the more straightforward ecological exposure scenario (Figure 3-13c) first, only two Eco DUs are defined, one for the grassy meadow and one for the wetland shoreline. However, each of these DUs contains two SUs that were designated N&E DUs (Figure 3-13a). The ecological exposure needs to be evaluated for the entire grassy meadow and the entire shoreline, but these zones were divided into two N&E DUs to examine the effect the landfill might have had on environmental media adjacent to the site. If elevated concentrations are noted in N&E DU26 compared to N&E DU27, it would confirm that transport had occurred off the site. However, within the risk context, the area-weighted average of SU26 and SU27 may show no risk is present. At this stage, the project team would need to evaluate this within the context of the project goals to decide how to proceed.

Turning to the human health exposures (Figure 3-13b), identical logic would be applied to trespassers who use the Figure 3-13a N&E DUs as SUs within the exposure scenario to derive an EPC using an area-weighted average. The EPCs for the two receptors within the property boundaries would also be arrived at through the use of an area-weighted mean, using the N&E

DUs as exposure SUs where the industrial/commercial worker includes SUs 1 through 23 as well as 28 and 29 in the derivation of the area-weighted EPC, while the construction worker would include SUs 11 through 23 and 28 in the derivation of the area-weighted EPC. A further refinement could include accounting for the frequency and duration a receptor spends within a particular SU, which could yield an EPC that reflects both spatial and temporal exposure.

Similarly, for the off-site human receptors, the meadow trespasser exposure DU is comprised of SUs 24 and 25 (Figure 3-13b), and the wader trespasser exposure DU is comprised of SUs 26 and 27.

3.2 Statistical Concepts and Applications in ISM Projects

The purpose of this section is to introduce key statistical concepts that are relevant to both the sampling design and analysis of ISM data. Many of these concepts are not unique to ISM and will be familiar to analysts for their use in other contexts. While this section is not intended to serve as a comprehensive guidance on environmental statistics, it does address many of the common questions that practitioners will likely have. Citations and hyperlinks are provided to guidance documents, white papers, peer-reviewed literature, and calculation tools to supplement the information presented here. In addition, several hypothetical examples with ISM data sets are included.

3.2.1 Why use statistics?

Statistical concepts have long been used to guide decisions involving both environmental sampling and inferences based on sample results. Regulatory guidance and quantitative tools facilitate the application of statistics methods that are transparent, objective, and defensible given site conditions (USEPA 2015). Statistical methods can be used to quantify uncertainty and express the level of confidence in estimates of exposure and risk – in turn, this consistency and reproducibility promotes consensus among parties with competing interests. Moreover, including statistical concepts during project planning prior to data collection or as a component of a tiered approach (such as adaptive sampling) can help stakeholders and project teams make scientifically defensible decisions for site investigations. Investigations guided by statistics are also more likely to result in cost-effective outcomes that achieve goals protective of human health and the environment.

This section describes the application of classic inferential statistical analysis methods to ISM data, assuming the sampling design yields a representative sample consistent with a study's objectives. A discussion of statistical concepts that inform the collection of physically representative samples (increment mass given the properties of the medium) and explicit accounting of sources of measurement errors is beyond the scope of this section.

3.2.2 Confidence intervals of the DU mean

Statistics are often used to calculate an upper bound estimate of the AM contaminant concentration of a DU, referred as a *UCL of the DU mean*. While the "true" population mean of a DU cannot be measured exactly, it can be estimated from an ISM sample with some specified tolerance for uncertainty. For environmental investigations, we are typically interested in choosing a method that yields a UCL greater than or equal to the population mean 95% of the time, which is why we call it the 95% UCL. This guidance explains how ISM introduces procedures to cost-effectively reduce sources of measurement errors in the field and the laboratory. This section discusses key statistics concepts and procedures that can be applied to both inform the sampling design and compute a reliable 95% UCL.

Typically, only a small portion of a DU is sampled. A statistician would refer to the set of concentrations reported from the environmental (or soil) samples randomly collected from the same DU as a sample. A variety of summary statistics can be computed from the sample, such as the set of environmental sample concentrations. The sample mean is the numerical mean of the concentrations and is often represented as x bar, or $\underline{x}$. The sample mean estimates the true DU mean, which statisticians refer to as the population mean and often denote with the Greek symbol μ. It is very unlikely that the sample mean ($\underline{x}$) and population mean (μ) will be exactly equal. For any single investigation, may be smaller or larger than μ. In general, we want to choose sampling designs that yield rigorous statistics. One way to achieve this would be to repeat the same investigation of a DU many times, generating a sample mean ($\underline{x}$) each time some of the sample means would be less than the population mean, and some would be greater, but the mean $\underline{x}$ would provide a reliable estimate of μ. Since it is impractical to resample a DU many times, we can rely on statistics instead. Three key statistics concepts tell us important facts about the sample mean generated from any single ISM investigation:

- Random sampling yields unbiased parameter estimates. If a random sampling design is used, the sample mean

will be unbiased, which means that, on average, $\underline{x}$ equals μ. Statisticians often express this equality by stating the average difference ($\underline{x} - \mu$) = 0. This desirable property is true for any sampling design that applies random sampling, including ISM.

- The sample mean from one investigation is more likely to underestimate than overestimate the population mean. The chance that any one sample mean underestimates the population mean (that is, ($\underline{x} - \mu$) < 0) depends on the shape of the probability distribution of the mean concentrations of the concentrations. If the PD is symmetric (or normal), there is an equal probability that $\underline{x}$ will underestimate or overestimate μ. If the distribution is skewed, the probability is unequal and depends on the direction and magnitude of the skewness. Most environmental datasets exhibit positive skew, which means that when they are plotted as a histogram, they have a tail that extends to the right. Under this condition, there is a greater probability that the sample mean underestimates (rather than overestimates) the mean (that is, ($\underline{x} - \mu$) < 0 is more probable than ($\underline{x} - \mu$) > 0).
- A key advantage of ISM over discrete (grab) sampling, which both simplifies the statistics and helps generate reliable results, is that it invokes the central limit theorem (CLT). Each replicate of an ISM sample can be thought of as an independent estimate of $\underline{x}$. If an ISM sample includes three replicates (r = 3), each generated with 30 increments (n = 30), we will have three different values from which we can compute summary statistics, such as the AM and SD. It can be helpful to think of the mean of all DU replicates $\underline{x}$ as the *mean of the means* or the *grand mean*. Importantly, the distribution of $\underline{x}$ values differ from the distribution of concentrations for each increment (n). Specifically, the distribution of $\underline{x}$ exhibits three key properties: (1) the shape of the distribution is more symmetric (less skewed); (2) the SD of $\underline{x}$ is lower; and (3) the grand mean yields an unbiased estimate of μ (as described above). Statisticians refer to the CLT when talking about the distribution of $\underline{x}$. The CLT tells us that as the sample size (n) increases, the shape of the distribution of $\underline{x}$ tends toward a normal distribution, even if the distribution of increments is quite skewed. The fact that the distribution of $\underline{x}$ approximates a normal distribution and that this approximation improves with increasing sample size (n) has important implications for 95% UCL calculations, as discussed further below. ISM basically incorporates superior coverage and sample processing steps to create a physical realization of the CLT.

In risk assessment, an EPC is typically based on a 95% UCL so that risk-based decisions are protective of human health and the environment (USEPA 2002). The purpose of a 95% UCL calculation is to provide an estimate of μ from a single investigation, such that we are unlikely to underestimate long-term average exposure. With statistical methods, we can express the likelihood of under- and overestimation by calculating a confidence interval(CI)for a population parameter. Each CI is defined by a lower confidence limit (LCL) and UCL. There are two relevant properties of a CI that we can specify: (1) the probability that the CI contains the population parameter and (2) whether the CI is one- or two-sided. A one-sided CI is one in which the population parameter is permitted to fall on one side of the CI, either below the LCL or above the 95% UCL but not both. A two-sided CI is said to contain the parameter with a certain probability, but the parameter may be either less than the LCL or greater than the 95% UCL. A 95% UCL is one-sided, such that it has a 95% chance of being greater than or equal to μ. This convention addresses the risk assessment goal of erring on the side of protectiveness of human health and the environment.

Using numerical simulation studies and statistics theory, we can evaluate the performance of different 95% UCL methods under varying site conditions. USEPA, for example, has conducted extensive simulations to evaluate the performance of 95% UCL methods calculated with the software tool ProUCL (USEPA 2015), and ITRC has also conducted extensive simulations to understand the performance of 95% UCL methods applied to ISM datasets. Two key performance metrics are 95% UCL coverage and CI width:

- Statisticians use the term *coverage* to refer to the frequency with which a 95% UCL equals or exceeds μ. A 95% UCL is intended to equal or exceed μ 95% of the time (and fail to exceed μ 5% of the time) if the same sampling design was repeatedly applied to a DU. Therefore, one goal of a 95% UCL method is to achieve a coverage of 95%. Different 95% UCL methods can yield different coverage probabilities, so one decision criteria for 95% UCL method selection is to determine if the method yields reliable (at least 95%) coverage, across a wide range of site conditions. An incorrectly chosen calculation method may provide coverage that is less than 95% (say, 85% or 90%).
- In addition to yielding different coverage probabilities, 95% UCL methods can yield different sizes (or widths) of CIs. The width of the CI is a measure of the uncertainty of the estimate of the DU mean(NIST, 2019). The larger the CI width, the larger the uncertainty. In general, the width of the CI increases as the variability of the data and required level of confidence increases. In addition, different 95% UCL methods can yield different 95% UCLs

when applied to the same summary statistics. If two 95% UCL methods achieve the same coverage, the method that yields a narrower CI (that is, lower than 95% UCL) is preferred (USEPA 2015).

Therefore, to effectively choose between 95% UCL methods, we need to understand the performance of each method under the specific conditions of interest and balance the dual objectives of 95% UCL coverage and CI width. Summaries from simulation studies conducted with ISM samples are presented below to help guide the selection of 95% UCL methods.

3.2.3 Illustration of the CLT using Pb data from the Becker study

The next series of figures is a graphical representation of the CLT. The data used to construct the CLT graphs were taken from the Becker study described in Section 2.2.2. The dataset consists of the 129 Pb results making up one of the four contaminated arrays evaluated for the study (Note that the Becker study evaluated four arrays, one of which is depicted in Figure 2-3a). Figure 3-15 plots a lognormal distribution fit to the dataset consisting of 129 individual Pb sample results. The 129 results range from a low of 48 to a high of 22,000-mg/kg Pb. The distribution is skewed to the right by the presence of a few extremely high results. Approximately 50% of the results are less than 2,400 mg/kg (as indicated by the median, equivalent to the geometric mean for lognormal distributions).

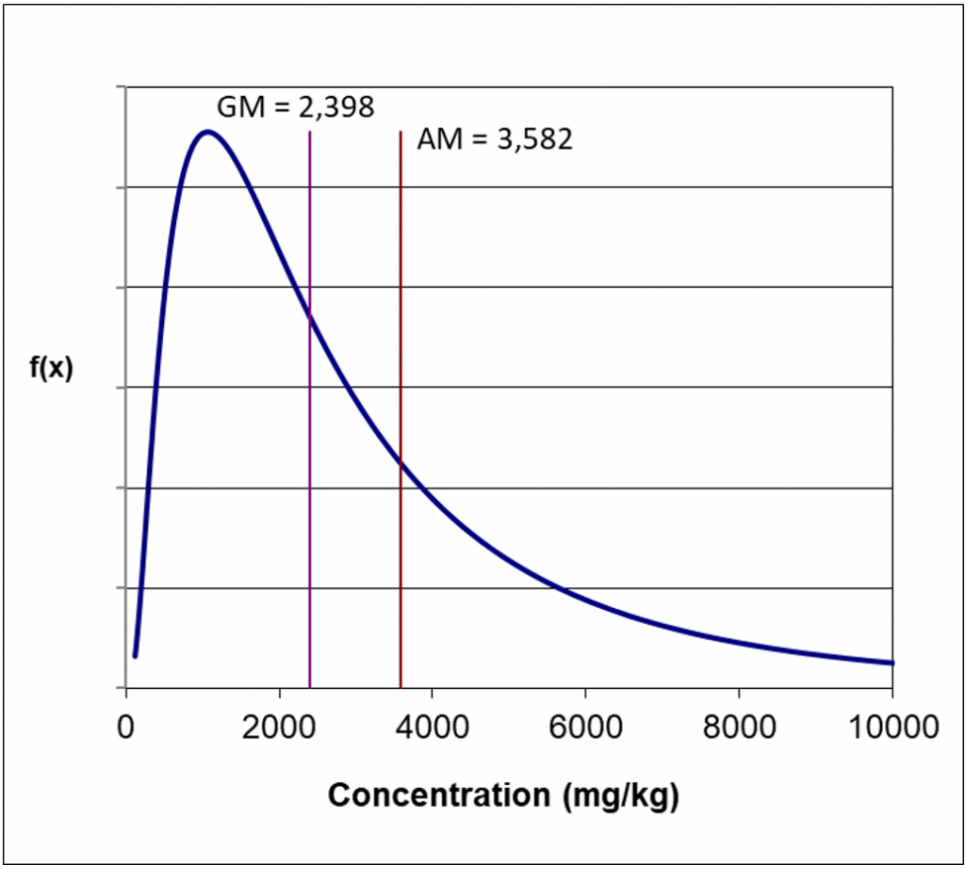


Figure 3-15. Two-parameter lognormal distribution with arithmetic mean (AM) = 3,582 mg/kg, median (equivalent to geometric mean [GM]) of 2,398 mg/kg, and standard deviation of 3,976 mg/kg. Data fit to *n* = 129 measurements of Pb in soil from case study discussed in Section 2.2.2.
Source: ITRC ISM-1 Team, 2012.

The CLT states that, for any random variable X with a (population) mean of μ and finite variance σ^2, the (sample) mean of a set of k independent replicates of X will approach a normal distribution with variance σ^2/k as k increases. In other words, regardless of whether or not X is normally distributed, when the numerical mean is calculated from a sufficiently large number of k replicates, the numerical mean of the replicates will be approximately normally distributed, and the variance of the mean will be equal to the variance of the replicates (σ^2) divided by k. Also, assuming the sampling design involves collecting the samples at random, the sample mean yields an unbiased estimate of the population mean.

For the purposes of illustrating the CLT concept, we can study how the distribution of sample means changes when we repeat many sampling events at a site for which the distribution of Pb concentrations in soil is described by a lognormal with

a population mean (μ) 3,582-mg Pb/kg soil and population SD (σ) of 3,976 mg/kg. Each ISM sampling event generates a replicate r consisting of n increments. With a numerical simulation, we can repeat the exact same sampling program (n, r), drawing random samples from the lognormal (μ, σ) many thousands of times, each time recalculating the sample mean from the r replicate results.

If we repeat the sampling program 150 times (r = 150), we generate 150 sample means. Through simulation, we can examine the following types of questions: (1) How does the distribution shape change as n increases (at what point is the distribution of means approximately normal)? (2) Does the mean of the 150 sample means (the grand mean) change as n increases? (3) Does the SD of 150 replicate means change as n increases?

For this example, the simulation is repeated three times using n = 5, 15, and 30, each with r = 150. The resulting distribution of sample means is illustrated in Figure 3-16 and summarized in Table 3-2 below. We can now provide answers to the questions outlined above:

- It turns out that the greatest effect of changing n is on the distribution shape. The shape of the distribution becomes more symmetric and approaches a normal distribution as n increases. This is a key concept of the CLT.
- According to the CLT, the grand mean should approximate the population mean. In this example, the population mean is 3,582 mg/kg, and the grand means for the three simulations are 3,438; 3,842; and 3,672 mg/kg (see Table 3-2).
- The SD of the set of r = 150 sample means depends on the choice of n. Specifically, the SD of the sample means is approximately proportional to the inverse of the square root of n.

In practice, an ISM investigation will generate very few datasets – for example, r = 1 to 3, and n = 30, from which we calculate one set of summary statistics for the sample mean (such as mean, SD). Therefore, with only 1 to 3 estimates of the population mean, we cannot rigorously explore the shape of the distribution of sample means with the usual goodness-of-fit statistics and data visualization methods. Note that at least n = 30 increments are typically used to prepare ISM samples when it is reasonable to assume there is mild to moderate heterogeneity (see Section 3.2.4.2 and Table 3-3). The extent of heterogeneity is a result of the dispersion of the data and the shape of the distribution.

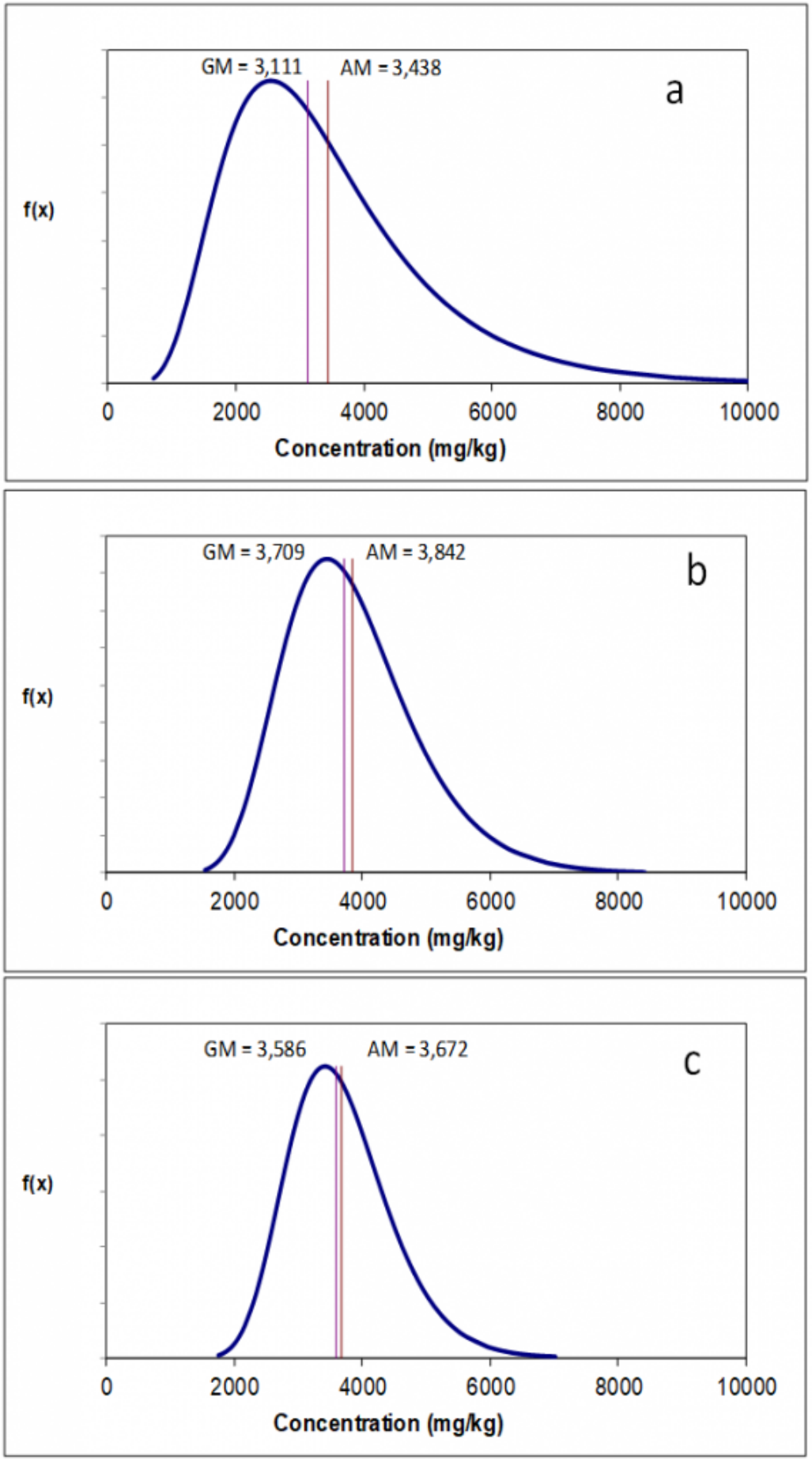


Figure 3-16. Demonstration of shift in distribution shape associated with the CLT. Distributions (blue curve) fit to a simulated set of *r* = 150 ISM replicates generated with three sampling designs: (a) *n* = 5; (b) *n* = 15; and (c) *n* = 30. The underlying distribution of increments is lognormal (AM, SD) where AM is 3,582 mg/kg and SD is 3,976 mg/kg (see Figure 3-15).

Source: ITRC ISM Update Team, 2020.

Table 3-2. Summary statistics for sample means generated by simulating ISM sampling events with *r* = 150 replicates and *n* = 5, 15, and 30 increments.

Source: ITRC ISM Update Team 2020 based on data from Becker, 2005.

Statistic	Population	Scenario A	Scenario B	Scenario C
Sample Size	129	n = 5, r = 150	n = 15, r = 150	n = 30, r = 150
Minimum	48	906	1,896	2,063
Maximum	22,000	9,990	7,216	6,990
Range (Max – Min)	21,952	9,084	5,320	4,927
Mean	3,582	3,438	3,842	3,672
SD	3,976	1,616	1,037	808
CV	1.1	0.47	0.27	0.22
RSD	111%	47%	27%	22%
Distribution Based on Goodness-of-Fit Statistics	Lognormal	Lognormal	Lognormal or Gamma	Normal
95% UCL	Not applicable	Chebyshev = 4,012	App. gamma = 3,986	Student's-t = 3,782

Notes: n = number of observations per event; r = number of repeated sampling events; CV = coefficient of variation = SD/mean; RSD = relative standard deviation = CV x 100%

3.2.4 95% UCLs

ISM samples provide *estimates* of mean concentrations, but many factors can cause an ISM sample's concentration to deviate from the true DU mean concentration. Under some circumstances, those deviations can be large. In one project, BaP concentrations greater than 466 µg/kg triggered DU cleanup. The first ISM sample had concentrations well below that, but additional replicate ISM samples were well above, for example, the first ISM sample = 380, the second = 1,100, and the third = 1,400 µg/kg BaP.

Severe underestimation by a single ISM sample is possible, leading to decision errors unless precautions are taken. Those precautions include replicate ISM samples and the use of 95% UCLs in decision-making.

In theory, all 95% UCL methods that are applied to discrete sampling results can also be applied to ISM. However, in practice, the options for 95% UCL methods with ISM are constrained because the small number of replicates (r = 3) precludes a rigorous evaluation of distribution shape and application of bootstrap resampling methods when distributions are not consistent with normal, gamma, or lognormal distributions.

Goodness-of-fit evaluations that inform the shape of the distribution require at least 8 to 10 observations from a dataset,

possibly more if the data are highly censored (meaning they include non-detects) (USEPA 2016). Typical ISM sampling designs include fewer than 8 to 10 replicates – in fact, three replicates are often used to estimate the SD of sample means. Such small sample sizes limit the options for statistical analysis of ISM data to two methods for 95% UCLs calculations: Student's-*t* and Chebyshev. The formulas for these two methods are presented below. Other 95% UCL calculation methods may be explore for larger sample sizes (e.g., $r \geq 8$ to 10) (USEPA 2016).

The Student's-*t* 95% UCL is restricted to datasets that follow approximately normal distributions. For ISM, physical averaging based on increased sample volume effectively reduces the variance in the underlying distribution of increments, but it cannot guarantee that the distribution is normal – on the contrary, it is expected that some degree of positive skewness will still occur for most sites. It would be incorrect to state that the CLT is always going to sufficiently normalize the distribution of replicate means to support an assumption of normality (USEPA 2016). Therefore, the key assumption for the approximate normality required to select the 95% t-UCL should be considered carefully. In cases where the underlying distribution for increment-sized soil masses is highly skewed (Figure 3-16), 30 increments may not be enough to normalize the distribution of replicate means. In such cases, the Student's-*t* 95% UCL may not provide the desired statistical confidence since it will have a greater than 5% chance of underestimating the population mean. Recalling the discussion of 95% UCL coverage associated with Figure 3-16, another way to say this is that the 95% t-UCL would not actually provide 95% coverage of the true mean. With a highly skewed underlying population, the 95% t-UCL might only cover the true mean 80% or 90% of the time. Of course, since the population mean is unknown, in practice, we cannot calculate or even conduct a simulation study to estimate the coverage for a site. Therefore, we rely on simulation studies of a range of different conditions with known population parameters in order to guide the selection of 95% UCL methods based on properties of the ISM dataset.

Figure 3-17 illustrates the CLT for $n = 30$ ISM samples and for underlying distributions (the individual increments represented by the pink curves) having various level of skew (right-hand tail). Skewness increases as CV increases from 0.5 to 3.0, left to right. Between a CV of 1 and 2, the skewness of the underlying population becomes too great for $n = 30$ to normalize the distribution of replicate means (blue curves).

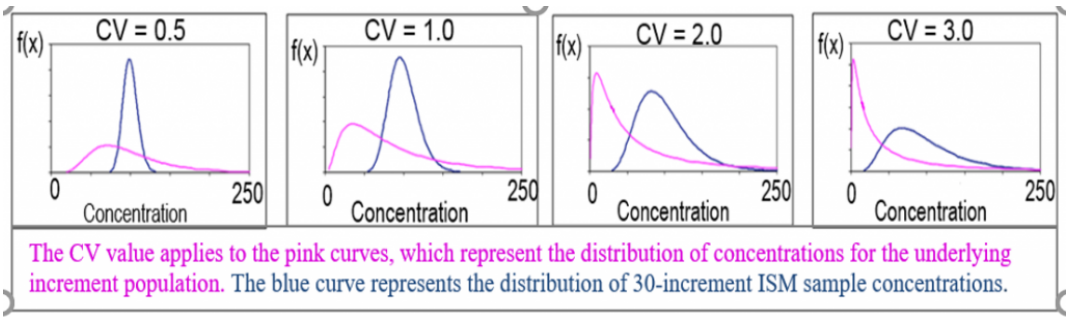


Figure 3-17. Distribution of means. The CLT states that the distribution of mean (blue data distributions) will be more normal (bell-shaped and symmetrical) than the underlying distribution of increments (pink data distributions). But as the underlying (pink) distribution gets more and more skewed (increasing CV), the distribution of means also becomes more skewed (CV = %RSD/100; CV = SD/mean).
Source: ITRC, 2017, ITRC Webinar: Soil Sampling and Decision Making Using Incremental Sampling Methodology – Part 1, Module 4 (Statistics).

Asymmetry in the distribution of means with CV > 1 indicates the assumption of normality is not supported. If the population of replicate means is not near-normal, the 95% t-UCL will likely not provide adequate coverage for the DU mean. As noted in the discussion of the CLT above, one option to improve the symmetry of the distribution of replicate means is to increase the number of increments per ISM sample replicate. As shown in Figure 3-17, the normalizing effect of the CLT is better when *n* is larger ($n = 30$ versus 15 or 5). For this example, the CV of the underlying distribution of increments was 1.1, but no single rule regarding sample size will apply universally because the key is the shape and spread of the underlying distribution of increments.

The Chebyshev method is a viable option to calculate the 95% UCL because it reduces the chance the 95% UCL will underestimate the population mean. The Chebyshev 95% UCL is a nonparametric 95% UCL, which means that it can be used when the data distribution is unknown or is not normal. In many cases, it can achieve the desired 95% coverage even with $r = 3$ replicates and $n = 30$ increments. Reducing the skew in the distribution of means by increasing *n* will improve the performance of all 95% UCL methods, including the Chebyshev. The Chebyshev is considered to be a conservative estimate of the 95% UCL because it generally achieves or exceeds the desired coverage rates, even for non-normal distributions. The Chebyshev is able to achieve the coverage for skewed distributions because the Chebyshev 95% UCL is higher than the 95%

t-UCL. In other words, for a given (1 – α) confidence level, the CI width for a Chebyshev 95% UCL is greater than for a 95% t-UCL, given the same ISM dataset (r, mean, SD). The implication for decision errors is that the Chebyshev 95% UCL is less likely to underestimate the true mean and lead to an erroneous conclusion that a DU is "clean" when in fact it is "dirty."

The initial ITRC ISM document provides recommendations on the selection of 95% UCL methods, given properties of the site and ISM summary statistics. At that time, other than at military ranges, practitioners had little experience with applying ISM to the more common types of sites, so the key observations and recommendations might not have been applicable:

- If the underlying population distribution is only mild to moderately skewed, the default number of increments per DU/ISM field sample required to normalize the ISM data is at least 30 (refer back to Section 3.2.4.2). As a rule of thumb, if the population CV < 1.5, the distribution is likely relatively normal or mildly skewed; if CV = 1.5 to 3, the distribution is moderately skewed; if CV > 3, the distribution is very skewed.
- It is possible that fewer increments per ISM sample will suffice, but data should be collected to demonstrate this statistically. These data are efficiently collected as part of a pilot study, but a pilot study is generally cost-effective only for large projects with many DUs of the same type. For smaller projects, it is more efficient to simply use the default.
- It is possible for DU heterogeneity to be higher than expected, so 30 increments will be too few, and replicate ISM samples will not agree as well as expected.
- As the underlying DU population becomes more skewed, normalization by increasing the number of increments (see discussion of the CLT in Section 3.2.3) becomes less effective. This was illustrated by computer simulations for 30 increments.
- Recall that the same normalization occurred for the Pb mean data in Section 3.2.4.2. The skewed underlying distribution of 129 Pb concentration data (variability of 111% RSD, which is equivalent to a CV of 1.11, refer to Table 3-3) was fully normalized by 30 field samples per sampling event, which is equivalent to 30 increments per ISM sample.

3.2.4.1 Calculating ISM 95% UCLs and a word of caution about ProUCL

The equations for Student's-*t* and Chebyshev 95% UCLs are easily programmed into an Excel spreadsheet file (see ISM 95% UCL Calculator). The methods will yield two different 95% UCL values, prompting a decision as to which of the two to use. Because ISM projects rarely measure the underlying distribution of the increments (that is, analyze at least 10 individual increments and run statistical analysis on the dataset), the CV of the underlying distribution can be estimated from the SD and the number of replicates, and the calculations can be built into a spreadsheet. Such an ISM 95% UCL calculator was built by the first ISM Team and has been updated since then with an improved modeling procedure. The calculator has several benefits:

- The user only has to enter the results of three to six replicate field samples, as well as the number of increments per sample.
- The ISM 95% UCL spreadsheet calculates both the Student's-*t* and Chebyshev 95% UCLs.
- The spreadsheet recommends which 95% UCL should be used.

Many practitioners are familiar with using ProUCL to obtain 95% UCLs for discrete datasets (USEPA 2015). With the release of ProUCL 5.1, ProUCL has been modified to allow calculation of UCLs for datasets with only three sample results. However, results using ProUCL should be interpreted carefully:

- ProUCL may present values from many methods, some of which may lead to an underestimation of the population mean with a greater than 95% frequency. It is important that the selection of a result for a particular dataset is guided by the findings from simulation studies involving small sample sizes and a wide range of types of underlying distributions.
- ProUCL fits the data distribution to several theoretical PDs (normal, lognormal, and gamma distributions). The default assumption is that a dataset fits the theoretical distribution until proven otherwise. However, the statistical tests possess poor power to reject the distribution assumption when the sample sizes are small (r = 3). For example, the underlying distribution of a small dataset may not be normal, but owing to a lack of sensitivity of the statistical test for normality, the assumption of normality will not be rejected (see Section 3.2.3.3). Therefore, ProUCL may recommend the 95% t-UCLs when the sample sizes are small even when that would clearly not be appropriate.

- ProUCL does not perform the calculations to estimate the variability in the underlying increment population and so cannot recommend whether the Student's-*t* or Chebyshev 95% UCL is more appropriate.
- The performance of Student's-*t* and Chebyshev 95% UCL methods applied to censored ISM data (meaning one or more replicates is qualified as a non-detect) has not been explored. At this time, it is unclear what the coverage probabilities can be expected when non-detects (NDs) are represented by an imputed value (such as half the detection limit).
- For larger sample sizes (r = 8 to 10), the ProUCL software can be used to explore a wider range of 95% UCL methods than the 95% UCL calculator. Note that USEPA guidance on the use of ProUCL 5.1 cautions that at least 10 to 15 observations are needed before relying on bootstrap resampling techniques to estimate the 95% UCL (USEPA 2015)."

3.2.4.2 Formulas for calculating ISM 95% UCLs

Calculation of Student's-*t* and Chebyshev 95% UCLs can be readily done using the Excel spreadsheet calculator. Here is the equation for the one-sided (1 – α) Student's-*t* 95% UCL:

$$UCL = \bar{x} + t_{(1-\alpha)(r-1)} \times \frac{SD}{\sqrt{r}} \quad \text{Eq. (3-1)}$$

where

$\bar{x}$ = arithmetic mean of all replicate ISM samples in a DU
SD = standard deviation of all replicate ISM samples in the DU
r = number of replicate ISM samples in the DU
$t_{(1-\alpha)(r-1)}$ = (1 – α)th quantile of the Student's-*t* distribution with (n – 1) degrees of freedom

For a DU with three replicate ISM samples and a 95% UCL, the equation reduces to the following:

$$\text{Student's-}t\ 95\%\ UCL = \bar{x} + 2.92 \times \frac{SD}{\sqrt{3}}$$

Here is the equation to calculate the one-sided (1-α) 95% UCL using the Chebyshev method:

$$UCL = \bar{x} + \left(\sqrt{(1/\alpha) - 1}\right) \times \frac{SD}{\sqrt{r}} \quad \text{Eq. (3-2)}$$

where

$\bar{x}$ = AM of all replicate ISM samples in a DU
SD = standard deviation of all replicate ISM samples in the DU
r = number of replicate ISM samples in the DU

For a DU with three replicate ISM samples and a 95% UCL, the Chebyshev equation reduces to the following:

Chebyshev 95%

$$UCL = \bar{x} + \sqrt{19} \times \frac{SD}{\sqrt{3}} = \bar{x} + 4.36 \times \frac{SD}{\sqrt{3}}$$

For clarity, SD divided by the square root of the number of replicates is equal to the standard error (SE). Therefore, the SD/sqrt(*r*) term is equal to the SE of the distribution of the mean of the replicate means.

The probability that a 95% UCL equals or exceeds the population mean of a DU is referred to as the *coverage*. The desired coverage for a 95% UCL is that, when calculated from an ISM dataset, the value is equal to or greater than the DU mean 95% of the time. Because different 95% UCL methods can yield different coverage probabilities, one criterion for assessing the performance of a method is to examine the coverage probabilities across a wide range of site conditions. For ISM-1 (see Section 1), numerical simulations were conducted to evaluate the coverage probabilities for DUs for which the contaminant distributions exhibited low to high heterogeneity, represented by lognormal distributions with the same AM but different variances. The CV, equal to the ratio of the SD divided by the mean, was selected as the summary statistic to express the dispersion of the distribution. It is important to note that coverage probabilities may vary depending on both the distribution shape and dispersion. Therefore, these simulation results may not apply for all DUs. For cases in which a different positively skewed distribution shape or greater dispersion is suspected, the Chebyshev 95% UCL may be the preferred calculation method because it is more likely to achieve the desired coverage than the Student's-*t* 95% UCL.

For ISM-1, the results from the simulations were presented in a table that recommended either a Student's-*t* or Chebyshev 95% UCL, depending on the expected degree of dispersion (given by the CV and corresponding geometric SD [GSD]) of the contaminant distribution across increments. A practical limitation of that presentation of findings is that the summary statistics (SD, mean, and CV) from most ISM investigations are based on concentrations measured in replicates (*r*) rather than individual increments (*n*). The CV of the increments can be estimated from the CV of replicates by adjusting for skewness of the distribution. For ISM-2, additional numerical simulations were conducted to determine appropriate adjustment factors so that findings from ISM-1 could be applied to statistics based on replicates. Table 3-3 summarizes these findings grouped by the CV of the replicates for $r = 3$. For example, if an investigation with $r = 3$ replicates yields a CV of 0.3, the equivalent dispersion for the distribution across increments is a CV in the range 1.5 to 3.0 (medium dispersion), and the Student's-*t* 95% UCL would not be expected to yield 95% coverage.

Table 3-3. Likelihood that ISM achieves coverage depending on dispersion (*r* = 3 replicates).

Source: ITRC ISM Update Team, 2020.

Degree of Dispersion >>		**Low**	**Medium**	**High**
Dispersion Metric	CV of replicates	< 0.23	0.23 < CV < 0.40	> 0.40
	CV of increments (no adjustment)	< 1.26	1.26 < CV < 2.19	> 2.19
	CV of increments (with adjustment)	< 1.5	1.5 < CV < 3	> 3
95% UCL Method	Student's-*t*	Yes	No	No
	Chebyshev	Yes	Yes	Maybe

Coefficient of variation (CV) = SD/mean. Geometric standard deviation (GSD) = for lognormal distributions.

The difference between Chebyshev and Student's-*t* 95% UCLs can sometimes lead to different decisions for a DU. Project teams must balance larger bias associated with the Chebyshev 95% UCL with the smaller coverage of the DU mean associated with the Student's-*t* 95% UCL when deciding which method to use. If there is no site knowledge available to support an assumption about the degree of dispersion (that is, low, medium, or high) of increments, then the Chebyshev 95% UCL may be the preferred calculation method because it is more likely to achieve the desired coverage than the

Student's-*t* 95% UCL.

Another option some practitioners may want to consider is the bootstrap 95% UCL. For a detailed discussion of bootstrap 95% UCLs, refer to the ProUCL technical guidance (USEPA 2015), but note that ProUCL is able to compute bootstrap 95% UCLs as well. To compute bootstrap 95% UCLs requires at least 10 to 15 field replicates for the DU. The bootstrap method involves treating the sample dataset with *n* observations as the entire environmental population. The population is repeatedly sampled with replacement *n* times to calculate a sample mean. This process is then repeated many times (say, 1,000) to obtain a distribution of sample means (say, 1,000 sample means). The percentile bootstrap 95% UCL takes the th percentile of the bootstrap means. Like the Chebyshev 95% UCL, bootstrap 95% UCLs have the advantage of being nonparametric, so an assumption of normality is not required. In cases of skewed distributions where the 95% t-UCL is not appropriate, bootstrapping methods may produce a more accurate estimate of the mean concentration that is less conservative than the Chebyshev 95% UCL. However, the percentile bootstrap 95% UCL typically falls short of the desired coverage and may not be appropriate for studies with strict coverage requirements. In addition, bootstrapping is likely to be ineffective for small sample sizes and should not be performed with less than $r = 10$ to 15 ISM samples.

The bias-corrected accelerated (BCa) bootstrap method is a modification of the percentile bootstrap 95% UCL that attempts to address the issue of insufficient coverage. The BCa 95% UCL corrects for bias in the bootstrapped means by increasing the percentile to be used – for example, if 95% confidence is desired, the BCa method may recommend instead using the 97th percentile of the bootstrap means as the 95% UCL. The recommendation depends on the degree of bias in the dataset. The coverage for the BCa method is improved over the percentile bootstrap 95% UCL, but coverage for the BCa 95% UCL may still fall slightly short of regulatory requirements compared with Chebyshev.

3.2.4.3 Minimizing the CI width in an estimation problem

A large CI width is not desirable when the goal is to confidently estimate the true DU. A common example is deriving the EPC, which uses a 95% UCL to provide an upper bound estimate of the true mean concentration in the receptor's EU. The 95% UCL is used to avoid underestimating the true mean and thus underestimating risk owing to exposure (USEPA 1992b). The 95% UCL may provide an unreliable estimate of exposure if the dataset is from too few field samples and/or is highly variable (see Section 3.2.2).

Figure 3-17 and the bullets below summarize the factors that affect 95% UCL sizes:

- number of ISM replicates (sample size, *n*) – the more measurements in the dataset, the smaller the 95% UCL (and CI width)
- degree of variability (range of data values) in the dataset – less variability (a lower SD value) gives a narrower CI width
- desired level of confidence – the higher the desired confidence, the wider the CI width must be (CI width will be narrower for 90% confidence than it is for 95% confidence; at 99% confidence, it will be wider than for 95%)
- data distribution of the population – data distributions (the shape of the data's histogram) that need to be modeled by a theoretical PD are referred to as parametric methods

For this last bullet, the normal and lognormal distributions may be the most familiar of the different types of PDs. Nonparametric methods do not require the data to be modeled by a particular PD. When applied to the same data or summary statistics, a parametric method (such as the Student's-*t* 95% UCL) will generally give CI widths that are narrower than a nonparametric method (such as the Chebyshev 95% UCL). Note that when $r = 3$, the only difference between the equations for the Student's-*t* and Chebyshev 95% UCLs is the value of the multiplier in front of the SD. For a three replicate DU and 95% confidence, the multiplier term is 2.92 for the 95% t-UCL, and 4.36 for the Chebyshev, which is why the Chebyshev 95% UCL will always be higher than the 95% t-UCL.

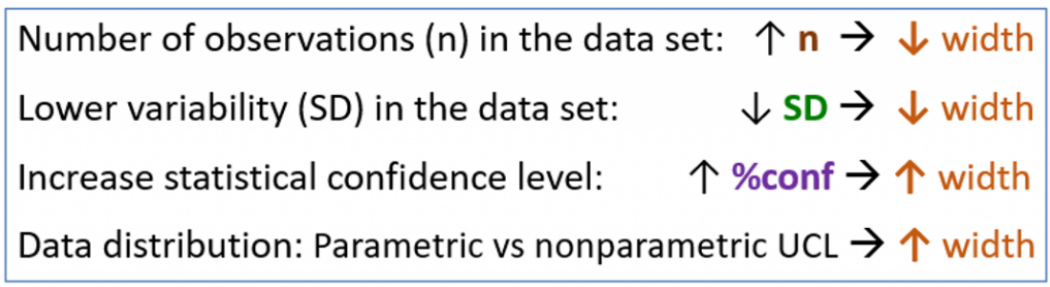


Figure 3-18. Four key factors influencing the distance between the dataset mean and the 95% UCL, and whether this influence widens or narrows that width.
Source: Deana Crumbling, 2020. Used with permission.

Overly large UCLs can be avoided by setting limits on how much uncertainty is tolerable. After coordinating with the risk assessor and stakeholders, the project delivery team might specify the CI width to be no greater than some percentage (%) of the dataset mean ( ). In other words, the 95% UCL value should be no larger than

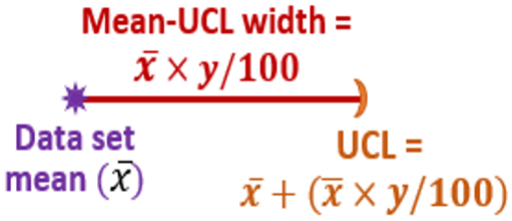


Figure 3-19. CI width expressed as a percentage of the mean.
Source: Deana Crumbling, 2020. Used with permission.

$$\bar{x} + (\bar{x} \times \frac{y}{100})$$

This is illustrated in Figure 3-18. The risk assessor might select the value for *y* by considering the expected concentration range or the point where the risk calculated from the EPC crosses some important benchmark (the concentrations at which the calculated risk increases from 10^{-5} to 10^{-4}). More than one *y* value could be set, depending on what concentrations are found: a wider CI width (and thus a large *y*) may be tolerated when concentrations are low, but a smaller *y* may be triggered if DU concentrations turn out to be larger. Setting a limit on the CI width allows project planning to adapt as the data are reported. The calculated DU mean and replicate variability can be used to calculate the 95% UCL. If the calculated width for a DU is greater than desired, additional replicates may be collected to reduce the CI width for that DU (see Section 3.2.2).

The CI width established in Figure 3-18 should be specified in the WP or QAPP. The project team would monitor the data as the data are generated to ensure the objective is being met and take corrective actions if not.

3.2.4.4 Are ISM 95% UCLs valid for risk assessment?

A common question from decision-makers and/or risk assessors is how an ISM 95% UCL compares with a 95% UCL calculated from discrete samples. Some risk assessors believe 95% UCLs based on composite samples are not valid (Mattuck, Blanchet, and Wait 2005). Both ISM and discrete sampling designs can be used to obtain defensible estimates of DU means. However, owing to the CLT, the variability of ISM results tends to be smaller than the variability of discrete sample results, which tend to yield smaller UCLs. ISM sampling designs also tend to result in superior physical site coverage relative to discrete sampling designs. For example, three ISM sample prepared from 30 increments each would be expected to produce a statistical sample that results in similar physical coverage of the site as 90 discrete samples. Also, as explained in Section 2.4.1, soil data variability is influenced by the mass of the analytical subsample. Increasing the analytical mass of an appropriately prepared and subsampled sample (ISM or discrete/grab) will also result in better representation of the sample and ultimate mean coverage from discrete 95% UCL calculations.

The data distribution in Figure 2-8 is largest for the 1-g subsamples, with the distribution narrowing for the 10-g subsamples; the variability in the distribution is smallest for the 100-g subsample set, which means that all the datasets were nearly the same. ProUCL determined that the distribution for the 1-g set was nonparametric and provided a list of eight potential nonparametric 95% UCLs from which to select. The eight 95% UCLs ranged from 2.49 to 4.63, and the 95% UCL that ProUCL recommended had a value of 2.58 nCi/g. The 10-g subsamples had a gamma data distribution, with a recommended 95% UCL of 2.00 nCi/g.

Risk assessors do not normally enquire about the analytical mass when evaluating a dataset, yet in this example, a 10-fold increase in analytical mass produced a 22.5% reduction in the 95% UCL. The same principle influencing the analytical mass applies to the mass of the field sample. Larger field masses reduce the variability of concentration data, which in turn reduces the 95% UCL. An ISM sample is the ultimate field sample mass since, to the best of our technology's ability, the ISM

field sample represents the concentration of the entire DU, and the ISM analytical subsample is managed so that it represents the concentration of the field sample.

3.2.5 Comparisons of 95% UCLs with project decision thresholds

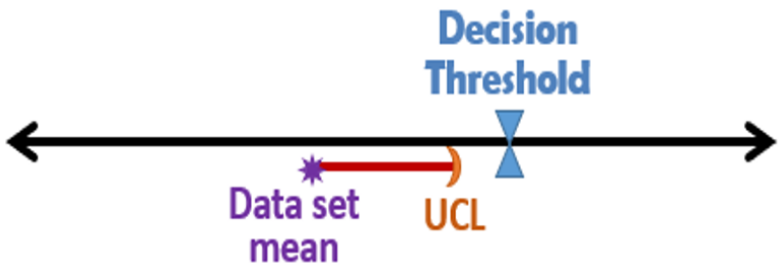


Figure 3-20. UCL and the decision threshold. The UCL is lower, indicating that confidence in the true mean being less than the threshold is lower.
Source: Deana Crumbling, 2020. Used with permission.

A common objective is to determine whether there is sufficient evidence to conclude if the true DU mean concentration is less than some risk-based threshold or other project action limit. As the 95% UCL is the upper end of the CI, the true mean is likely less than the decision threshold if the 95% UCL is below it (Figure 3-19). From the perspective of statistical analysis, the evaluation of compliance with a decision threshold *L* can be thought of as an example of a one-sample, one-sided hypothesis test (see Section 3.2.5.1). If the UCL is below the threshold, it can be stated with 95% confidence that the true DU mean is also below the decision threshold. In this sense, the 95% UCL of the DU mean controls decision errors arising from measurement uncertainty.

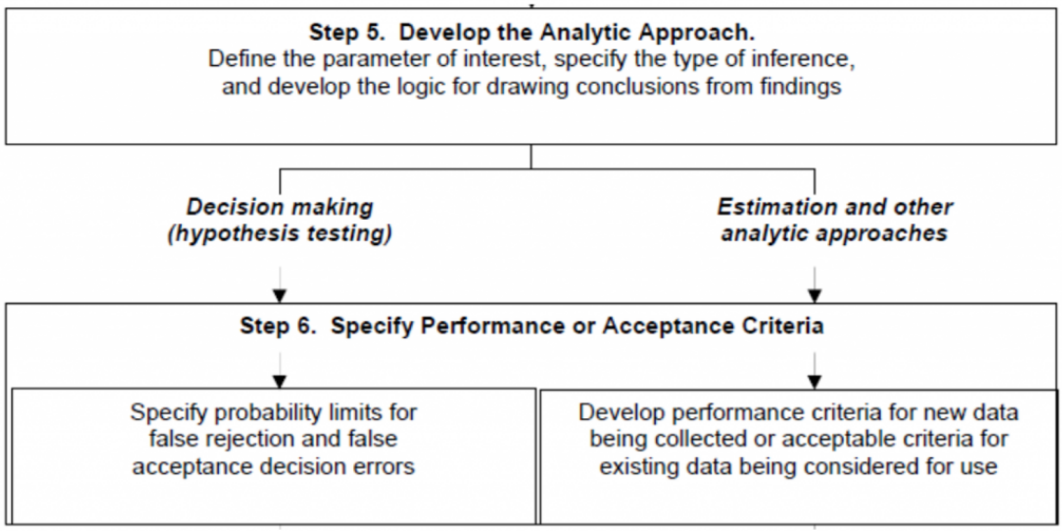


Figure 3-21. Steps 5 and 6 of the seven-step DQO process.
Source: USEPA. 2006b. Guidance on Systematic Planning Using the Data Quality Objectives Process.

Practitioners may be reluctant to tackle steps 5 and 6 of USEPA's DQO process (Figure 3-20). Abstract and unfamiliar terminology can make the statistical component of this process seem more challenging than it actually is, but Chapter 6 of the update is recommended as a resource for more details than are possible in this document. One of the ways that ISM makes steps 5 and 6 simpler is by making clear that decisions apply to individual DUs, not an entire site all at once (as the wording in the DQO guidance implies). Replacing guidance references to "site" with "DU" brings statistical concepts to a manageable level.

Steps 5 and 6 would also typically entail CIs for estimation problems or hypothesis tests for DUs – for example, the use of a CI to estimate an upper bound concentration for a DU mean (previously discussed in Section 3.2.4) is an example of an estimation problem. Comparing a 95% UCL with a project decision limit or project action limit is essentially equivalent to conducting a one-sample hypothesis test. Comparisons of 95% UCLs with decision thresholds can be used to achieve the

same outcome as hypothesis tests and are likely easier to conduct and understand. CIs and statistical hypothesis tests are simply flip sides of the same statistical concepts (USEPA 2006b).

The easiest way to implement the statistical aspects of the DQO process is by using UCLs.

3.2.5.1 95% UCLs and hypothesis tests

Hypothesis tests are commonly used to select from one of two mutually exclusive alternative actions or decisions. They require a null or baseline hypothesis (H_0) and an alternative hypothesis (H_1), with the alternative hypothesis being the condition that needs to be proved. For example, a null hypothesis may be that the DU mean is greater than a compliance level by 10 mg/kg, with an alternative hypothesis that the DU mean is less than or equal to the compliance level plus 10 mg/kg. The null hypothesis is the default condition that data are used to disprove, so a weight of evidence is collected to reject H_0 in favor of H_1. Often, the failure to reject H_0 is an inconclusive result (that is, H_0 may or may not be true).

The probability of rejecting H_0 (in favor of H_1) when H_0 is actually true is referred to as the *false rejection error*, *false positive error*, or *Type I error*. The data user's tolerance for Type I error is usually denoted by the Greek symbol α. It is equivalent to stating the required level of confidence for the hypothesis test (that is, rejecting H_0) is 1 – α. The value of alpha often ranges from 0.1 to 0.01, thus the maximum allowable probability for erroneously rejecting H_0 is commonly 1% to 10%.

The failure to reject H_0 in favor of H_1 when H_0 is false is referred to as the *false acceptance error*, *false negative error*, or *Type II error*. The tolerance for Type II error is usually denoted by the Greek symbol β, whichrefers to the maximum probability that H_0 is false when H_0 is not rejected. The quantity 1 – β is referred to as the required power of the hypothesis test, where the power of a hypothesis often ranges from 0.8 to 0.95 and can be viewed as a measure of the sensitivity of the hypothesis test. The larger the power of the test, the more likely the null hypothesis will be rejected when it is false.

More discussion about these relationships can be found in the 2006 G-4 document beginning on page 63 (USEPA 2006b). The tolerance for Type I and Type II error is summarized in Table 3-4.

Table 3-4. Tolerance for Type I and Type II error.

Source: ITRC ISM Update Team, 2020.

	H_0 Is True	**H_0 Is False**
Reject H_0 (Conclude H_0 False)	Type I error, α (false positive)	Correct decision Power, 1 – β
Do Not Reject H_0 (Conclude H_0 True)	Correct decision Confidence level, 1 – α	Type II error, β (false negative)

Comparing a 95% UCL of the DU mean (μ) with a decision threshold L is equivalent to conducting the following one-sample, one-sided hypothesis test with a Type I error tolerance of 5%:

H_0: $\mu \geq L$, H_1: $\mu < L$

This is one-sample hypothesis test because it entails only one population parameter, the DU mean μ. A two-sample hypothesis test would be conducted to compare the DU mean with a background mean. The null hypothesis that the DU is dirty (that is, $\mu \geq L$) is rejected with 95% confidence when the 95% UCL of the DU mean is less than L. This constitutes what is considered to be an acceptable weight of evidence that the DU is clean. Most cleanup scenarios operate from the assumption that a DU is dirty until proven clean, which is illustrated in Figure 3-21, which is similar to Figure 3-19, but for the addition of the true DU mean (μ) to show the relationship between hypothesis testing and the 95% UCL. A 95% UCL less than the threshold (Figure 3-21b) allows the default dirty assumption to be properly rejected. If the 95% UCL is above the decision threshold, the evidence is not good enough to conclude that the DU is clean (Figure 3-21a). This figure illustrates a false negative – that is, the failure to reject the null hypothesis when it is true. A false positive occurs when the true DU mean is greater than the decision threshold (the null hypothesis is true) but is erroneously rejected (95% UCL < decision threshold).

This is illustrated in Figure 3-22.

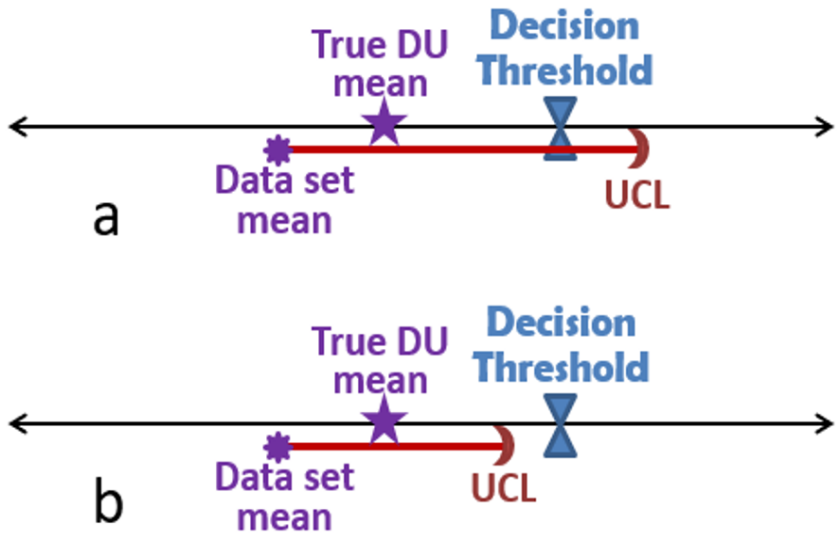


Figure 3-22. True mean and the decision threshold. (a) Although the true DU concentration is less than the threshold, the data are not good enough to bring the UCL below it, so the assumption that the DU is dirty cannot be rejected (false negative). (b) The data provide a UCL that is below the threshold, allowing the dirty assumption to be rejected and the DU declared clean.
Source: Deana Crumbling, 2020. Used with permission.

Note that the terms "false positive" and "false negative" relate to erroneous rejection or acceptance of the null hypothesis, respectively. Consider a second hypothesis test with the following null and alternative hypotheses for the DU mean:

H_0: $\mu \leq L$, H_1: $\mu > L$

For this second hypothesis test, a false positive (incorrectly rejecting H_0) would occur if the null hypothesis that the DU is clean were erroneously rejected for the alternative hypothesis that the DU is dirty. Contrast this with what constitutes a false positive for the first hypothesis test shown – that is, erroneously concluding a dirty DU is clean.

For the first hypothesis test, the Type II error would be the probability of erroneously concluding a clean DU is dirty. Strictly speaking, the failure to reject H_0 is an inconclusive result unless the tolerance for Type II error is met. However, a tolerance for Type II error is not specified for the hypothesis test. For environmental applications, the DU is often conservatively assumed to be dirty when the null hypothesis $\mu \geq L$ is not rejected (e.g., Figure 3-21b).

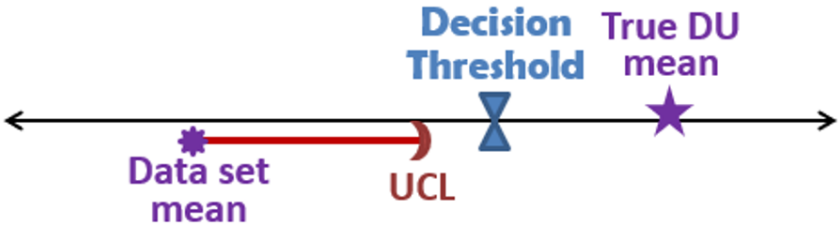


Figure 3-23. False positive for H_0: $\mu \geq L$. The true DU concentration (star) is above the decision threshold. However, the dataset provides a mean and 95% UCL that are below the threshold, resulting in a false positive (erroneous rejection of the null hypothesis).
Source: Deana Crumbling, 2020. Used with permission.

It is noted that this approach is conservative from the perspective of human and environmental risk but does not control false negative errors (that is, erroneously concluding a clean DU is dirty). A false negative can occur when the difference between the true DU mean is small, relative to the magnitude of measurement variability. Data variability could be high because of inadequate laboratory sample processing (meaning high subsampling variability). Alternatively, there could have

been too few increments to handle the degree of field heterogeneity, producing ISM field replicates with poor precision. The variability (or SD) among the replicates could have increased 95% UCL so the CI overlaps with the threshold (Figure 3-21). Recall that the CI width can be narrowed by collecting more data points, so a solution could be to collect additional DU field replicates (n) or to increase the numbers of increments (k) used to prepare each ISM sample, though the former will likely be more effective than the latter to potentially decrease the 95% UCL to a value below the threshold. Note that if there is very large variability, the original 95% UCL can underestimate the DU mean. Under that circumstance, collecting additional ISM samples could increase the 95% UCL.

3.2.5.2 Underestimation of the DU mean

Under certain circumstances, even ISM may misrepresent the true DU mean, although that is much less likely than with discrete sampling. The cause is the presence of small but significant areas of elevated concentration that are missed by insufficient increment density. This increment density may be too low because the default number of increments was used for a large DU without considering potential areas of elevated concentration, or the areas of elevated concentration are too small to be consistently captured. The existence of areas of elevated concentration increases the overall heterogeneity of the DU, increasing data variability and requiring a higher number of increments to manage.

> ***Significant areas of elevated concentration denote small areas of increased concentration that have the potential to change a sample concentration from being below the decision threshold to above it if they are captured in their proper spatial proportions by an ISM sample.***

If areas of elevated concentration exist, but their potential presence and configurations are not anticipated in the CSM, the default number of increments could allow a single ISM sample to miss them and underestimate the DU mean. Collecting replicates and calculating 95% UCLs of the DU mean are usually the best strategy for minimizing false positives and underestimating the true DU mean. When sampling DUs with a poorly understood CSM and unknown spatial distributions of contaminants, it is recommended that at least three replicate field samples be collected to estimate the 95% UCL. If all increments of the replicates are evenly placed across the DU, there is a good chance of at least one field sample incorporating at least one area of elevated concentration. The set of ISM samples will likely represent areas of high and low concentrations in the proper proportions for estimating the DU mean and variance. Even one increment picking up a much higher concentration can provide the warning imparted by imprecise field replicate data. If other causes of data variability can be ruled out by QC data, disagreement among field replicates is an indication that more increments may be needed to manage the heterogeneity caused by small areas of elevated concentration.

The following are two examples of areas of elevated concentration causing imprecision among DU replicates of real projects. One of the lessons from these projects is that when heterogeneity is known or suspected to be high, a DU decision based on simple comparison of a single DU field sample to a decision threshold increases the probability of decision error.

Field Replicate (ppb)	**#1**	**#2**	**#3**
DU-A	380	1100	1400
DU-B	460	490	230

The first example is a property adjacent to a landfill that is contaminated with PAHs. The risk driver is BaP, and as shown in the table below, the variability between field replicates DU-A and DU-B is high. This may be due to buried materials that leach BaP into the soil or that weather to shed particles of nearly pure PAHs (such as chunks of old asphalt). Despite 50 increments per field sample for DUs less than 1/10th acre, and with rigorous sample processing that includes milling, the three replicate BaP results can sometimes resemble the two sets below. Note that the BaP cleanup level = 466 ppb (Crumbling 2019).

The other example comes from the Hawaii PCB study (HDOH, 2015). Figure 3-23 shows a 6,000 ft^2 (~1/7th acre) area that was known to be contaminated with PCBs (from spilled transformer fluid). This DU was sampled with three field replicates and 60 increments per field sample, with the three replicate PCB results coming in at 19, 24, and 270 ppm (the applicable AL

was 50 ppm). The samples were reanalyzed to confirm the accuracy of the results. This level of disagreement is a clear sign of extreme heterogeneity, most likely manifested as small areas of elevated concentration within the DU.

3.2.6 95% UCLs as applied in ISM designs

A question unique to ISM is whether one DU-ISM sample is sufficient (and thus a 95% UCL is not needed) because the ISM sample itself is an estimate of the DU mean. This is a complex topic, so the answer depends on the study question. If the question involves risk-based decision-making, then a 95% UCL may be needed.

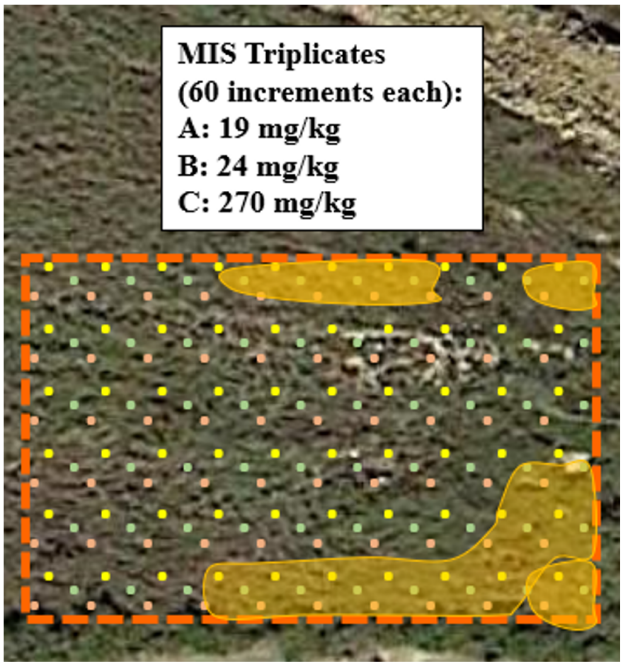


Figure 3-24. Three replicates. Pink, green, and yellow dots represent increments for three replicate field samples of 60 increments each. Yellow blotches depict zones where small areas of elevated concentration likely occur, as suggested by 120 discrete samples.
Source: Developed by Deana Crumbling from data by HDOH, 2015. Used with permission.

Increasing the number of increments, the mass of each increment, or both will increase the likelihood of accurately estimating the true DU mean.

As was previously shown, it is important to remember that any individual ISM field sample can be significantly larger or smaller than the true DU mean, and it is very unlikely that any result will match the true mean. An incorrect conclusion becomes more likely when the DU is more heterogeneous than expected. At the start of site sampling, at least three independent field replicates are needed to assess variability. If the heterogeneity of the site cannot be assessed before ISM sampling, the number of replicates (or increments) needed may be underestimated. If the underlying population is very heterogeneous, the replicates can have very different estimates of the mean, and the 95% UCL may be elevated. To avoid this, heterogeneity should be assessed whenever possible. After heterogeneity is understood, the sampling design can be optimized. The variability in the DU sample mean depends on the sampling design and can be reduced by increasing the sample support or the number of increments (Figure 3-17). Some sites contain hundreds or thousands of DUs, and 95% UCLs for every DU might not be needed to maintain protectiveness and decision confidence. A statistically sound design for such a strategy is more complicated than basic ISM design: it depends heavily on a mature CSM, an experienced ISM practitioner, and continual evaluation of QC measures.

3.2.6.1 Do not default to "maximum sample concentration"

For the estimation of EPCs using discrete samples, it is a common practice to use the sample maximum (the maximum detected concentration) for the EPC for the EU (that is, the DU) when the 95% UCL of the EU mean is greater than the sample maximum. However, this approach is less likely to provide the desired coverage of the EU mean than the 95% UCL of the mean in most ISM sampling designs. Table 3-4 illustrates the relationship between the ratio of the 95% UCL/maximum

for the condition when the underlying distribution of increments is lognormal with CV ranging from 0.1 to 3.

Table 3-5. Probability of the 95% UCL exceeding the maximum concentration.

Source: ITRC ISM Update Team, 2020.

Replicates (*r*)	P (95% UCL>max) for Student's-*t*	P (95% UCL/max) for Chebyshev
3	1.00	1.00
4	0.33	1.00
5	0.04	1.00
6	0.01	0.60
7	0	0.37
8	0	0.18

When $r = 3$, the 95% UCL of the mean will always be larger than the sample maximum. For Student's-*t* 95% UCL, the probability is 33% or less for $r > 3$, whereas for Chebyshev 95% UCL, $r > 5$ is needed for the maximum replicate result to be lower than the 95% UCL for some sampling events. For $r = 3$, the ratio of the Chebyshev 95% UCL to the sample maximum is typically less than 1.5.

3.2.6.2 Extrapolating 95% UCLs among CSM-equivalent DUs

Predicting 95% UCLs for single DUs is a strategy that some ISM practitioners are adopting to obtain the uncertainty management benefits of 95% UCLs while avoiding the time and cost of collecting three replicate ISM samples from every DU. However, this strategy should only be applied with CSM-equivalent DUs and is most useful when one or more factors apply:

- There are many (perhaps hundreds) of CSM-equivalent DUs (DUs for which the mechanism of contamination is expected to be similar).
- Multiple rounds of sampling over months or years will be needed to complete sampling of all site DUs.
- There is one contaminant acting as the primary risk driver, and a numerical cleanup criterion has been established.
- More than 30 increments are needed per DU to manage high short-range heterogeneity.
- Increment collection involves the subsurface and more than one depth interval.
- Increment collection is difficult, and refusal is common.

However, replicates should not be collected solely on a frequency basis (such as when three replicates are collected for 10% of the DUs) or a on a per batch basis similar to the manner in which QC samples such as laboratory control samples and matrix spikes (MS) are processed for laboratory analyses. The success of extrapolation methods usually relies on the ability to statistically model the variance or SD for the DUs for which replicates were not collected – for example, the CSM suggests contamination was released in a similar fashion over a larger number of DUs, so it may be desirable to collected replicates for a subset of the DUs. This may be done either in a separate pilot study before the field work begins or at the start of the field program. A statistical test that compares variances (such as an F-test or Levene's test) may be subsequently used to determine whether the differences in variances of the DUs from which the replicates were collected are statistically significant. If the differences in variances (s) are not significantly significant (95% UCL), the variances can be pooled. The square root of the pooled variance may subsequently be used to calculate s for the DUs for which only one ISM sample was collected. For example, if replicates are collected from m different DUs, Chebyshev UCLs for DUs for which only one replicate was collected may be calculated using the following equation:

$$\text{95\% Chebyshev } UCL = \bar{x} + 4.36 \times s_{pooled}$$

Note that this equation follows from Eq. (3-2) as $n = 1$ for a DU for which only one ISM sample is collected, and s_{pooled} is a pooled SD determined from the variances of the m DUs:

$$s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_M - 1)s_m^2}{n_1 + n_2 + \cdots + n_m - m}}$$

When an equal number of replicates are collected from each DU, the SD formula is simply the square root of the mean variance:

$$s_{pooled} = \sqrt{\frac{s_1^2 + s_2^2 + \cdots + s_m^2)}{m}}$$

3.2.7 Statistical independence in ISM

A common assumption of many statistical methods is that of independent observations – in other words, the measured value of an SU should not be affected by the value of any other SU. Consider the example of height in individuals. Using the height of two identical twins as two discrete samples and treating them as independent would be inadvisable because the two values are likely to be similar and represent redundant information. The heights are correlated in that they are related, and using correlated values undermines the reliability of statistical analyses.

In environmental data, the assumption of independence may be violated if a DU is stratified into multiple SUs, and there is a spatial trend over the scale of the SUs. If there are multiple SUs that follow a large-scale spatial trend, SUs that are near one another would be potentially correlated and would not be statistically independent. Spatial trends should be taken into consideration during the process of DU delineation to ensure that SUs are independent. If a spatial trend exists within a DU, depending on the size of that DU, the nature of the trend, and the scale of the SUs, a biased estimate of the DU mean can be obtained. However, small-scale spatial trends within an SU do not violate assumptions of independence with ISM data because of the composite nature of ISM (see Section 6.2.2).

In addition to the sampling process, care must be taken during handling procedures to avoid violating the assumption of independence. Suppose a volume of soil representing a single ISM result is not homogenous, and a laboratory subsample is prepared by simply sampling small volumes of soil from only the top portion of the ISM sample. Subsamples prepared in this manner may be more similar than subsamples prepared by collecting soil randomly from different portions of the ISM sample, but the subsample is unlikely to represent the concentration of the entire field sample. For that reason, ISM sample processing involves mixing either by disaggregation and sieving or disaggregation and milling. A one- or two-dimensional (1D or 2D) Japanese slabcake technique with incremental subsampling is then used (see Section 5.3.5).

Staff new to ISM sometimes ask whether field replicates can be collected by splitting a single ISM sample three ways, but this is never recommended because the result actually measures the precision of the splitting process rather than providing three independent estimates of the DU mean.

3.2.8 Application of specialized SUs in ISM projects

The following examples illustrate situations where defined volumes of soil (SUs) are sampled for the purpose of gathering information (such as to refine the CSM) but not to make a decision on the SUs per se. For environmental projects that use ISM, note that such activities are typically conducted to ultimately estimate the DU mean or make a decision about it (say, for a future phase of the project). Please refer to Section 2.

3.2.8.1 Statistical SUs to determine the mean of very large DUs

SUs can be used to statistically determine the mean and 95% UCL for a DU so large that it cannot be sampled as a single unit, provided that the CSM supports relatively homogeneous contaminant concentrations across the entire large DU. In brief, (1) the large DU is completely divided into many equally-sized, spatially contiguous SUs; (2) a random sample of at least 10 SUs is selected from the DU for sampling by ISM; and (3) the SU data are used to calculate the mean and 95% UCL for the DU in the same way as discrete data would be used in ProUCL as described below.

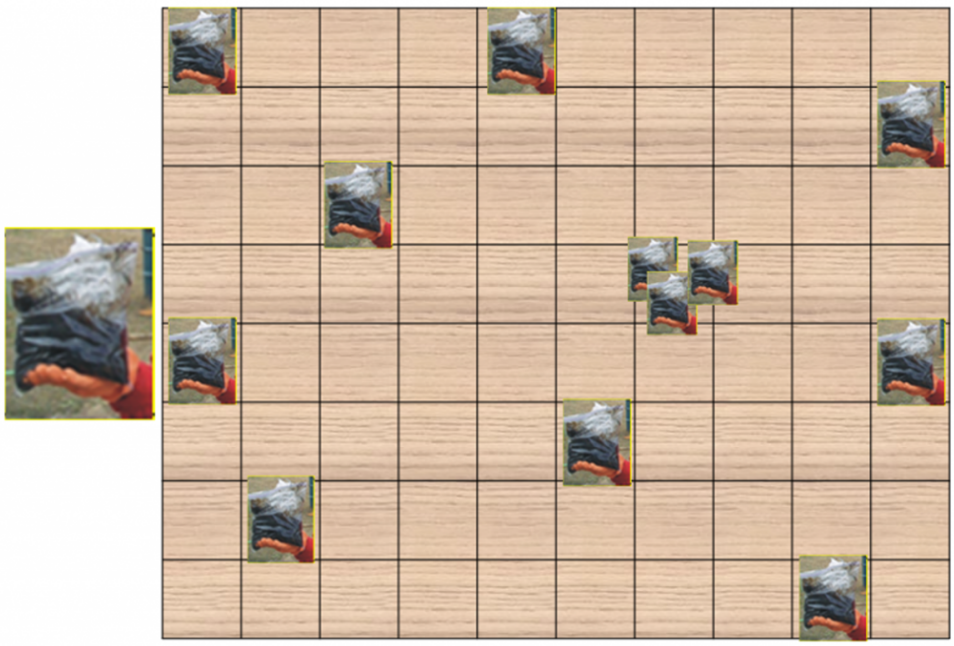


Figure 3-25. An 80-acre DU sampled with 10 statistical SUs, one SU with three replicate field samples.
Source: Deana Crumbling, 2020. Used with permission.

As an example, a risk assessor determines that the EU for a farm worker plowing a potentially contaminated field is the acreage of land that can be worked in a day, say, 80 acres. The 80-acre DU is divided into 80 1-acre SUs, and the DU (population) is defined to consist of a set of 80 SUs. Ten of the 80 SUs are randomly selected for ISM sampling (Figure 3-24). Each ISM sample is prepared by randomly collecting 30 increments from the SU. This random selection of both SUs within the DU and increments within each SU helps ensure that a representative statistical sample will be collected.

It may be desirable to collect three replicate 30-increment ISM sampled from one of the 10 SUs for QC purposes – that is, to ensure 30 laboratory subsamples. Three replicate laboratory subsamples are evaluated to ensure that laboratory sample processing and subsampling procedures can control within-sample heterogeneity. Only one 30-increment ISM sample per SU is collected from the nine remaining SUs.

However, only one result from the SU from which three replicates were collected is included in the dataset used to estimate the DU mean because this requires independent data. The three SU replicates are not necessary statistically independent for that purpose (see Section 3.2.6.7) but are potentially related to each other in a way that the other SU data points are not.

DU summary statistics are calculated from the 10 independent SU data points. Because these data points are from different parts of the DU, those data will not necessarily be normally distributed. However, given a sample size of 10, statistical software such as ProUCL can be used to determine whether the results fit a theoretical PD (that is, a normal, lognormal, or gamma distribution) and calculate the 95% UCL of the DU mean. Because CI width (or 95% UCL magnitude) is partly determined by sample size, a determination of compliance (95% UCL ≤ threshold) may be sensitive to the choice of number of SUs.

3.2.8.2 Statistical SUs to make not-to-exceed determinations for very large DUs

As in the example in Section 3.2.8.1, the DU is too large to be sampled as an entire unit. However, instead of determining whether the DU mean exceeds a decision threshold, the goal is to determine whether a proportion of the DU exceeds a threshold. This decision scenario can occur within the context of certain RCRA situations, such as land disposal restrictions. This statistical strategy can be useful in other applications as well.

The strategy is explained in the RCRA's waste sampling guidance in Section 3.4.2, "Using a Proportion or Percentile to

Determine Whether a Waste or Media Meets an Applicable Standard" (USEPA 2002g). Consider the scenario in which the waste material or media at the site (whose boundaries define the spatial extent of a DU) are comprised of a population of unique SUs, each of a defined size, shape, and orientation. Since it is not possible to know the status of all portions of a waste site, we can collect a representative sample and use statistics to support inferences regarding the characteristics of the population. The relevant statistical methods involve calculations of the CI of a proportion (or percentage) of the waste (or DU) that complies with the standard (USEPA 2002g).

The document describes two statistical strategies that could be used, but only the simple exceedance rule method will be discussed here. It is simple because the outcome is either pass or fail, and statistical tables can be used instead of equations. The method is not constrained to a particular PD of concentrations – nonparametric methods are available to achieve acceptable decision error rates. The method is also reliable even in cases of highly censored data (such as a large proportion of the sample results being qualified as non-detects) (USEPA 2002g).

For this strategy to provide an accurate estimate of the mean concentration, the DU must have a relatively homogenous distribution of contaminants. The strategy consists of the following steps: (1) completely divide a large DU into many (more than 100) SUs of equal size; (2) select a subset of *n* SUs at random; and (3) collect a random sample of increments from each of the *n* SUs. Refer to Table G-3a in the RCRA guidance (USEPA 2002g)and the equation below to determine the number of SUs (*n*) that need to be sampled to demonstrate with (1 – α)100% confidence that at least some desired proportion *p* of the DU is acceptably clean.

For example, based on nonparametric statistics, there is 95% confidence that at least 95% of the DU population is less than the threshold if 59 SUs are sampled and the results reported from all of them are less than the decision threshold. A statistician would describe the maximum reported concentration (from the set of 59 sampled SUs) as a nonparametric 95/95 UTL. By convention, the first of the two values convey the percentile, or coverage (the required proportion of the DU that must be clean), and the second value conveys the magnitude of the upper confidence limit for the percentile. Therefore, a 95/95 UTL is a 95% UCL for a 95th percentile, a 95/90 UTL is a 90% UCL for the 95th percentile, and so on. In general, the number of SUs that must be sampled *n* to demonstrate at least a proportion *p* of the DU is clean with (1 – α)100% confidence when the maximum value is less than the standard can be estimated from this equation:

$$n = \frac{\ln a}{\ln p}$$

If one or more SU results exceed the standard, we would conclude that the entire DU is not in compliance. If exactly one SU result exceeds the standard, one option may be to continue sampling more SUs (selected at random), effectively to increase the proportion of clean SUs (to approach *p*). The total sample size (*n* SUs) required to achieve this result can be calculated explicitly. In general, if one result exceeds the standard, to demonstrate at least a proportion *p* of the DU is clean with (1 – α)100% confidence, the second largest value reported must be less than the standard, where *n* is the smallest positive integer that satisfies the inequality:

$$np^{n-1} - (n-1)p^n \leq a$$

Rather than dividing a large DU into multiple SUs, the same strategy can be used to divide a large study area or property into multiple equal-sized DUs that are randomly sampled using ISM. As an example, a 70-acre former agricultural field is to be developed into ¼-acre residential lots. A review of historical operations suggests the distribution of pesticides is relatively homogeneous across the site, but there is concern that the top 6 in of any single ¼-acre lot could exceed regulatory standards for pesticides. The 70-acre area is divided into 280 ¼-acre residential lot DUs, 59 of which are randomly selected for sampling using ISM samples of 30 increments. Therefore, each SU in this scenario is also a small-area DU. If a 95% UCL is required, either a percentage or all of the ¼-acre DUs would be sampled in three replicates in order to provide an estimate of the variance in the mean concentration.

If none of 59 sampled ¼-acre lots exceed the standard, there is 95% confidence that at least 95% of each ¼-acre lot (sampled and unsampled) in the study area complies with the standard.

The concept of the area sampled within a large DU is very important for risk assessment and risk-based decision-making. When designing a sampling plan to characterize a portion of a large DU and account for potential decision errors from

extrapolation, it is helpful to recognize three key factors that can influence the extrapolation uncertainty and likelihood of making a decision error: (1) the variance of the increments (CV of the underlying distribution); (2) the percentage of the large DU area sampled; and (3) the likely magnitude of the average 95% UCL (across all sampled subarea DUs) relative to a compliance level (that is, the ratio of average 95% UCL divided by compliance level). The situation that results in the highest error rates is when the CV is relatively high and the ratio of the average 95% UCL to the compliance level is between 0.1 and 0.4. Ratios lower than this range are extremely unlikely to yield a false negative in which we conclude from the pilot study that unsampled areas of the site are clean when in fact they are not. Likewise, as the average 95% UCL approaches the compliance level (meaning the ratio approaches 1), it is also very unlikely that all the sampled areas will have 95% UCLs that are less than the compliance level (when in fact one or more mean concentrations truly exceed the compliance level). Results of simulation studies that provide error rates for a range of site conditions (CVs), sampling plans (percentage of areas, number of increments and replicates), and 95% UCL calculation methods (Student's-*t* and Chebyshev) are available in the White Paper by Goodrum and Mendelsohn (Goodrum and Mendelsohn 2018).

3.2.8.3 SUs to collect spatial information to guide cleanup

A key assumption when defining a DU is that concentration differences within DU boundaries are not important to know – rather, the mean concentration (as estimated conservatively by the 95% UCL) is what matters. While this is true for the primary purpose of the DU, follow-on decisions could arise if the DU concentration exceeds the threshold, and cleanup action is required.

Cleanup must target the contaminated soil, which may or may not exhibit a well-defined spatial distribution. If the bulk of contamination is located in only a portion of the DU, there are many advantages to removing only that portion as opposed to the entire DU. Knowledge of contaminant locations can be ascertained by using SUs designed for that purpose. The DU may be split into several SUs based on professional judgment of where localized contamination is likely present. The DQOs and study questions of the project and soil disposal options can also indicate whether the SU data need to be collected using quantitative ISM samples (30+ increments) or whether composites of only several increments each are acceptable. If the DU mean is near the AL, semi-quantitative data from composites of 5 to 10 increments may not be sufficient to establish where cleanup is needed if risk-based cleanup goals are employed. Because concentrations are near risk-based cleanup goals, a higher degree of precision is needed, and a quantitative approach is recommended.

If other information indicates that cleanup of a DU will likely be needed and returning for follow-up sample collection is undesirable, SUs can be defined at the same time as the initial DU sampling. SU composites can be collected at the same time and held for analysis only if required.

3.3 Planning for the Use of ISM Data

Data collection and evaluation is an iterative process, beginning with the CSM's development, project planning, analysis, data quality evaluation, CSM revisions, planning of confirmation sampling, and so on. Systematic planning for sampling and analysis is used to support the collection of data whose quality can be robustly evaluated to be sufficient for the intended use of the data. The DQO process introduced in Section 3.1 is a common method for systematic project planning. This section will discuss steps 5 and 6 of the DQO process as it pertains to ISM project planning. Section 3.2 provides an introduction to the statistical concepts discussed in this section.

3.3.1 Decision Rules and Uncertainty

DQO step 5 involves developing an analytic approach prior to receiving the data that will guide analysis of the study results and then drawing conclusions from those data. In this step, the site team specifies what population parameter is most appropriate for making decisions or estimates. It is important to plan for the analysis of the data before they have even been collected. Considerations such as what parameters will be estimated, how uncertainty will be evaluated, and what statistical analyses will be conducted are important to project planning because they may affect aspects of the sample design, such as the appropriate number of samples or number of increments.

ISM samples are estimates of the mean concentration within the DU, and therefore this type of sampling is useful when the average concentration in a particular area is of interest. If the data to be collected will be compared to a threshold or another comparison will be done for the purpose of making a decision, the site team should also specify what level and decision rule they will use in making their decision – for example, whether a 95% UCL will be compared to a screening level or whether a single ISM sample will be used. If the data are being collected to estimate a site parameter, the estimation

method should be specified. Moreover, defining the parameter of interest and the decision rule and threshold or the degree of acceptable uncertainty at this point in the sample planning process ensures that the data evaluation will be based on the quality of the sampling methods and the intended use of the data. Once the parameter of interest is identified, step 5 is typically stated in an if-then format to explicitly state the decision rule. If a 95% UCL is compared to a screening level, the statement may appear as, "If the 95% UCL exceeds the AL of X, then take remedial action, else leave the area intact." In each case, step 5 should explicitly state what remedial action will be taken.

Whereas DQO step 5 considers what the parameter of interest is for the DU and how the data will be used to estimate that parameter, DQO step 6 considers the impact of error and uncertainty on how the data will be interpreted and what level of uncertainty will be considered acceptable. For decision problems, step 5 typically presents decision criteria in the form of if-then statements, but step 6 is essentially a mathematical expression of step 5 that defines the statistical methodology to be used to make a decision selected from a set of mutually exclusive alternates, conclusions, or outcomes. Decision-making problems represent a considerably different type of intended use of the data compared to estimation and other types of problems. The approach to handling and controlling for error and uncertainty associated with the collected data also differs considerably between these two types of problems. As a result, once step 5 of the DQO process is completed, one of two processes is taken in step 6, based upon the intended use of the data: (1) specify performance metrics and acceptable limits of uncertainty (estimation problem) or (2) specify probability limits for false rejection and false acceptance decisions errors (decision problem). Section 3.2 provides an introduction to these statistical concepts. In the example of comparing a UCL to an AL, step 6 would involve defining the confidence of the UCL to limit the acceptable decision error (such as 95%) and setting performance criteria, which in this case, is the target width for the CI.

In the application of ISM samples to a decision problem, one example of a decision rule is comparing a 95% UCL to a threshold to make a decision, where the threshold could be a screening, action, or cleanup level. Confidence limits, which are estimates of the lowest and highest value the true mean might have based on the data collected at a specified confidence level, are important measures for comparisons to the threshold. In particular, where the 95% UCL is below the threshold, the site team has a high confidence in that determination. Similarly, confidence limits may also be important in defining acceptable limits on uncertainty for estimation problems. For estimation purposes, a maximum width of the confidence limit may be specified, typically defined as the distance between the sample mean and the 95% UCL. As discussed in Section 3.2, calculating the CI requires that each DU have three replicate field samples. An exception to this is the unique case of a site with many DUs where the CSM and a statistical evaluation (such as one that uses pilot data) indicates that the heterogeneity of the DUs should be similar. In this case, if sufficient evidence is available to establish the similarity across DUs, the CV of the similar DU could be applied to those missing replicates.

DQO steps 5 and 6 are critical considerations for developing a sampling plan (DQO step 7) and for evaluating the data collected (see Section 6 for further discussion of data evaluation). Defining the parameter of interest and setting acceptable uncertainty or decision error limits during the sample planning provide a clear understanding of how the data will be evaluated and ensure a study design that is likely to achieve the required data quality. The sections below discuss how these steps are applied when using ISM for risk assessment and site characterization (estimation problems) and for comparisons to screening, action, or cleanup levels (decision problems).

3.3.2 Use in risk assessment and site characterization

Two of the most common applications of ISM are site characterization and risk assessment. During this part of the planning process, considerations for each application are essentially the same, the principle difference being in the delineation of DUs (see Section 3.1). For the purposes of site characterization, delineation of DUs should be based on the CSM. If a goal of sampling is to determine if there are spatial patterns of contamination, the DUs should be delineated such that the relationships are not obscured by the composite nature of ISM (see the example in Section 2.5.3). In risk assessment, DUs are typically delineated according to exposure, with the ISM result being representative of the average concentration across the EU (that is, the EPC, see Section 8). The simplest application of ISM to risk assessment occurs when DUs are defined by size and area assumptions of risk – in other words, the DU and the EU are one and the same. These are called exposure area DUs. Due to the inevitable uncertainty in estimates of the mean, regulatory agencies often require a 95% UCL to represent an EPC or to assess compliance with decision criteria (see Section 8).

A variety of different designs are possible for the layout of SUs with regard to their relationship to DUs or EUs (see Figure 1-1). Most commonly, the EPC is calculated using Student's-*t* or Chebyshev 95% UCL formulas (see Section 3.2.4.2) across field replicates from a single DU. A second possible design is a large DU that is split into multiple SUs (see Section 3.2.8.2). In this case, the ISM results can be used in the same way to compute a 95% UCL if there is a single ISM sample for each SU,

assuming that each SU is equally representative of the DU. This assumption could be violated if, for example, the SUs have different areas or exposure time differs between the SUs.

If there are two or more samples in each SU, a weighted mean could be calculated without having to assume equal representativeness (see the example in Section 6.2.2). However, if multiple replicates are available for one or more SUs, but only a single ISM result is available for other SUs, the method of computing a 95% UCL becomes more complicated because there is no measure of uncertainty available for the SUs with a single result. The results cannot be simply pooled together because the replicates from the same SU will be similar to one another (see Section 6.3.1). In this scenario, three different statistical approaches are possible:

- If there is sufficient evidence to suggest the SUs are similar, the first approach would be to apply the CV from the SUs with multiple replicates to the SUs with a single result (see Section 3.3.1).
- The second approach, which doesn't require the assumption of similar SUs, would be to randomly select a single ISM result from the SUs for which there are multiples and compute a 95% UCL using one result from each SU. If this approach is used, it is important that the selection of which results are used is truly random. The disadvantage of this approach is that there is a loss of information because some samples will be discarded. The selection and subsequent 95% UCL calculation could be repeated many times to obtain information on how the sample selection process ultimately affects the 95% UCL calculation.
- The third approach is to compute the average for each SU and use each average to calculate a 95% UCL for the larger area. This approach has the advantage that all the data will be utilized, and the 95% UCL will be the same each time it is calculated for the same dataset.

3.3.2.1 QA/QC criteria

Particularly if data will be used for risk assessment, the data should be evaluated to ensure adequate quality. The purpose of any QC activity is to determine whether procedures or methods are working as intended. If any data fail QC criteria, the proper action is to determine the root cause, prevent recurrence, and determine what data have been affected by the failure. Most environmental practitioners are familiar with the concept of QC only in the context of the analytical laboratory, but it is critical for field sampling as well (USEPA 2006b). In general, sampling error is much larger than measurement error and consequently needs a larger proportion of resources to control (USEPA 2006a). As early as 1989, studies showed that sample-related variability contributed most of the total data imprecision (USEPA 1992c). "It has been estimated that up to 90% of all environmental measurement variability can be attributed to the sampling process" (Homsher et al. 1991).

Laboratory-related QC criteria address analyte loss and cross-contamination during laboratory processing, as well as the representativeness (or homogeneity) of laboratory subsamples (see Section 5). Sample-related QC checks evaluate how well each aspect of the sampling design is controlling measurement variability. QC problems signal that the data generation process needs to be improved. Failed QC indicates a problem that potentially affects all samples undergoing the same procedure, not just the ones subjected to the QC check. It is necessary to specify performance criteria prior to data collection, as part of the sample plan or WP. Performance criteria will vary from project to project.

The following procedures measure data variability due to soil heterogeneity at different spatial scales:

- Field duplicates are helpful in ISM to quantify precision of samples from the same SU. For discrete data, this is analogous to co-located samples (a set of two separate samples taken a few inches apart so that their results should be nearly the same). This practice is now infrequently used for discrete sampling because the results rarely agree, and no USEPA guidance makes clear what to do when they do not. Yet USEPA's Superfund guidance retains a requirement for co-located samples if soil data are to be considered definitive (USEPA 2006b).
- Laboratory duplicates are two subsamples taken from the same field sample for separate analyses. This is another QC practice that is intended to, but frequently does not, produce similar results, especially if heterogeneity is high. When the degree of difference exceeds some limit, guidance says to flag all data in that same batch as "estimated."
- As part of regulatory oversight, a field sample might be split between the responsible party and regulator field staff for two independent laboratory analyses. Split samples are ostensibly used to ensure the integrity of the implementer's data, with the procedure resting on the assumption that the splits will have similar concentrations. However, this assumption may not be met, and it is not always clear what should be done if not.

Whatever procedures will be used should be specified during the project planning stages, as well as what measures will be

computed, what will indicate an acceptable level of quality, and what action will be taken in the event that quality is not achieved.

Similar measures are used to quantify precision for each of the above procedures. The term *precision* is different from the term *accuracy*. Precision describes the reproducibility of the overall sampling method, whereas the accuracy of the data with respect to the true mean concentration of the contaminant in the DU area and volume of soil can only be known by extracting the chemical from the entire volume of soil and measuring the mass. The true error in the data therefore cannot be determined as part of the sampling design. The potential for significant error in environmental decisions can, however, be assessed based on a review of the collection, processing, and analysis of samples, as well as the precision of replicate sample datasets.

RSD is a measure of the precision in replicate samples (typically three replicates) defined as a percentage. It is calculated as the ratio of the SD to the mean, multiplied by 100. Similarly, CV is simply the ratio of the SD to the mean. The RSD can be calculated from either field or laboratory replicates (at least three) to estimate total measurement precision or the laboratory component of variability, respectively. For example, a low RSD from three field replicates indicates good precision in the sampling method used, which may indicate data for similar DUs collected under the same design have good reproducibility and reliability. However, the determination of what is “good” is highly subjective and may vary from project to project, so this should be specified during project planning. For specific examples of CVs that may indicate 95% UCL estimation reliability, refer to Table 3-3 in Section 3.2.4.2. Note that high RSDs can become unavoidable as contaminant concentrations approach the laboratory method reporting and detection limits.

Similarly, the relative percent difference (RPD) is the percent difference between two sample results that are expected to have the same true concentration. The RPD is calculated for field or laboratory duplicate results to evaluate the precision of the sampling design and measurement system. In concept, the RPD and RSD are similar metrics, but RPD is used for duplicates whereas RSD can be used for upward of three samples. Although a 30% RPD or RSD has emerged as a typical standard, in reality, the RPD/RSD depends on the analytical methodology as well as the analyte. The target RPD/RSD should depend on the project objectives and be within the capabilities of the measurement system.

Completeness is a sampling design term defined as the minimum number of sample results (data points) needed for each analyte and DU to meet each data use objective, relative to the number of usable results that are available or planned. Completeness criteria should be specified in the WP or QAPP. Recovery is a measure of the agreement of an experimental determination and the true value of the parameter, and is used to identify bias in a given measurement system. Standards for recovery criteria depend on the analytical methodology and are available in the U.S. Department of Defense’s Quality Systems Manual (QSM) Appendix C (DOD/DOE 2018) (see also USEPA SW-846 (USEPA 2007)). These standards do not account for ISM sample preparation procedures (see Section 5), but they could be used as a loose guide for determining recovery criteria.

3.3.3 Comparison to endpoints

Another common application of ISM is in a decision problem, as described in USEPA’s DQO guidance, which involves comparing sample results (typically, a 95% UCL) to a numerical criterion of some sort. Cleanup levels, also called remedial goals, are an example of such a criterion. Cleanup levels may be defined in a variety of ways, including regulatory standards, or based on risk assessment calculations under applicable guidance. The specific set of criteria will vary from project to project, but considerations for applying ISM data for these comparisons are consistent across all types of numerical criteria.

Many cleanup levels are developed based on a set of assumptions about the exposure to humans and ecosystems, including the size and location of EUs. When sampling to compare site concentrations to a cleanup level, the area of the EU or DU should be compatible with the exposure assumptions used to develop the cleanup level. If the EU is much larger than the assumed exposure area used to develop the cleanup level, receptors may be exposed unequally to subareas of the EU, and unacceptable levels of contamination may be left unaddressed. For example, comparison of ISM results for a 2-acre DU to a risk-based cleanup level might be appropriate for cleanup levels pertaining to an ecological or recreational endpoint, but might be considered inappropriate for a residential endpoint. These considerations should be addressed in the planning process. Comparisons of ISM results to cleanup levels should also appropriately account for the possibility of a decision error (for more detailed discussion, refer to Section 3.2). Practitioners must consider what probability of each type of error is acceptable. Generally, practitioners would rather make the mistake of remediating a site that is already clean than make the mistake of not remediating a site that is contaminated.

While a common decision rule is comparing the cleanup level to a 95% UCL, the choice of decision rule may vary from project to project (see Section 6.2.1 and USEPA 1989). In the example of calculating a 95% UCL and comparing this value to a cleanup level, the probability of concluding that the mean is less than the cleanup level when it is actually greater is 5%. Therefore, this decision rule can be applied if the acceptable error rate for this decision is deemed to be 5%. Specifics for computing 95% UCLs for risk assessment under various sampling designs (see Section 3.3.2 and Section 8.3.3.1) are also applicable for comparisons to any cleanup level.

3.3.3.1 Considering decision errors and sample sizes

Identification of acceptable rates for each type of error is important information that can be used to determine how much data need to be collected. As acceptable error rates decrease, it is necessary to be more confident that the decision made is the correct one, and more information is needed to make a statistically valid conclusion. This could come in the form of an increased number of increments or additional ISM samples. A power analysis could also be performed to calculate the precise number of samples that would be needed to achieve acceptable error rates.

Increasing the number of increments collected increases confidence that any one ISM result is a good estimate of the mean concentration in the area of interest. Similarly, increasing the number of ISM samples collected increases the confidence that the true mean has been captured in the collection of sample results. Both result in a narrower CI width (see Section 3.2.4.3), therefore, the mass of soil collected ultimately affects the probability of making a decision error in subsequent analyses.

In any decision problem, two types of errors are possible. When a 95% UCL is calculated, the probability of concluding that the site is clean when it is truly contaminated is equal to 0.05 (false compliance, α). The opposite error is concluding that the site is contaminated when it is truly clean (false exceedance). While confidence is one minus the false compliance rate, one minus the false exceedance rate is known as power. Statistically, power is the probability of obtaining a significant result when the result is truly significant. In this case, power refers to the probability of concluding that the site is clean (the 95% UCL is below the AL), given it is truly clean (the true mean is below the AL). For a specified number of ISM samples, lowering the tolerable false compliance error rate will result in decreased power.

In order to inform the number of samples needed for desired error rates, a power analysis could be conducted if information is available on how much variation is expected in the data and how far site concentrations are from the AL. Such analyses must be done a priori. Given known variation in the data, the relationship between power, α, *n*, and the detectable difference is quantifiable by many publicly available online calculators, as well as the pwr package in R. The detectable difference is the minimum difference that can be detected by statistical analysis with a probability equal to the power. In the case of comparison to a numerical criterion, the detectable difference is the difference between the true site mean and the threshold or cleanup level. Because certain information about the site is needed for a power analysis, it is most useful for sites where a pilot study or characterization of a similar area has already been conducted. The variance of those data could be used as a surrogate in the analysis. The more variation in the data, the greater number of samples required to achieve the desired power at the required significance level. In addition, it is necessary to have a general idea of the concentrations that may exist in the area of interest in order to define the difference the analysis should detect. Typically, this is the difference between the threshold or cleanup level and the true site mean. Since the true site mean is unknown, power could be calculated using multiple values for the detectable difference as well as and power, and a somewhat subjective decision could be made taking into account cost, power, the detectable difference, and the ability to collect additional samples.

An example of a power analysis on a real dataset is given in Table 3-6 for Pb. In this case, data from a similar area were available, so the sample SD of the observations (that is, measurements of environmental samples) from that area was used in the calculations, which assumed that the data were independent and approximately normally distributed (see Section 6.3), and a 95% UCL would be compared to the screening level. The sample mean Pb concentration for the surrogate area was 25% lower than the screening level. The estimated detectable difference between the population mean and the screening level is assumed to be 25%, but sample sizes are also computed for 15% and 35% differences to inform decisions. In this example, a more conservative approach would be to assume the true difference is lower than 25% (say, 15%). For a multi-phased sampling design, it may be more cost-effective to use 25% or even 35% and resample the area in subsequent phases if necessary. Note that power calculations could give sample sizes less than three, but three replicates are still required to compute a 95% UCL.

Table 3-6. Example power analysis results for sample sizes required to detect true differences of 15%, 25%, and 35%.

Source: Hayley Brittingham, Neptune and Company, Inc. Used with permission.

15% Difference			
	1% False Compliance (α= 0.01)	5% False Compliance (α= 0.05)	10% False Compliance (α= 0.1)
30% False Exceedance (power = 0.7)	11	6	5
20% False Exceedance (power = 0.8)	13	8	6
10% False Exceedance (power = 0.9)	15	10	8
25% Difference			
	1% False Compliance (α = 0.01)	5% False Compliance (α = 0.05)	10% False Compliance (α = 0.1)
30% False Exceedance (power = 0.7)	6	4	3
20% False Exceedance (power = 0.8)	7	4	3
10% False Exceedance (power = 0.9)	8	5	4
35% Difference			
	1% False Compliance (α = 0.01)	5% False Compliance (α = 0.05)	10% False Compliance (α = 0.1)
30% False Exceedance (power = 0.7)	5	3	< 3
20% False Exceedance (power = 0.8)	5	3	3
10% False Exceedance (power = 0.9)	6	4	3

Another important consideration in estimating sample size is the selected values of false exceedance and false compliance rates. Most commonly, practitioners are satisfied with a 5% probability of false compliance error (α = 0.05) and a 20% probability of false exceedance (power = 0.8). However, it is worth considering deviating from this convention depending on the needs of a particular site. As an example, α = 0.01 might be appropriate for sites with a larger volume of people living close by. If remediation is especially costly, practitioners may decide it is worth collecting additional samples to avoid the risk of remediating a site that is already clean (false exceedance). In the example above, increasing the number of samples from 4 to 5 would decrease the chances of false exceedance from 20% to 10% (assuming a 25% detectable difference and α = 0.05).

The power analysis is only one line of evidence to inform appropriate sample size. For example, in the case of large DUs made up of smaller SUs, a certain level of spatial coverage may be desired to adequately characterize the entire DU (such as 10% of the potential SUs). Qualitative judgments such as the level of skew in the data or a desire for increased confidence or power may also be used as a justification for additional samples or increments. See Section 3.2.8.1 and Section 3.2.8.2 for statistical discussion of large DUs made up of smaller SUs.

3.3.4 Background comparisons

Another common study objective is to compare ISM site data to background concentrations. This comparison requires different statistical methods than those used to compare ISM site data to numerical criterion or cleanup levels. While cleanup levels are known constants, background concentrations are variable, and the true mean and maximum background concentrations are unknown. Comparing site data to background concentrations involves relating two different distributions (site and background) and requires a hypothesis test (such as a two-sample *t*-test) to determine whether the concentrations are similar (see Section 3.2.5.1). Comparing ISM samples to background concentrations is most straightforward when both site samples and background samples are collected using ISM and with similar sampling characteristics such as the number of increments and the total volume of soil per sample (see Section 3.1.6.2). While statistics exist to compare ISM samples to other types of samples, these methods are complex and require the assistance of a qualified statistician. With any statistical method, predefining acceptable error rates is an important part of the planning process.

It is important to check the assumptions of the chosen test after data collection to ensure that the test is appropriate (see Section 6.3). Some practical tests for comparison of site ISM samples to background ISM samples are discussed in the following sections. For more information on these comparison methods, see the White Paper (Georgian 2020).

3.3.4.1 Comparison of means

For risk-based environmental applications, it is preferable to use hypothesis tests to compare mean site and background concentrations. Two-sample *t*-tests are commonly used for this comparison because they only require that both distributions are roughly normally distributed, which is generally achieved for ISM samples through physical averaging and the CLT. A two-sample *t*-test compares the means of the site and background distributions with a prescribed confidence level, such that the null hypothesis is only overturned if there is strong evidence that the site is actually contaminated. Statistical terminology and decision points include error rates, statements of null and alternative hypotheses, and one- and two-sided tests.

Hypothesis testing involves two key concepts regarding the definitions of null and alternative hypotheses. The first concept involves the definition of the null and alternative hypotheses as inequalities. When conducting a background screening analysis, we are most interested in evaluating whether the site mean is greater than the background mean, and whether the difference in means is greater than some prespecified substantial amount *S*. If so, the null hypothesis is expressed as an inequality, and the term one-sided (or one-tailed) test is used. With one-sided hypothesis tests, the false compliance error rate is often $\alpha = 0.05$. The second concept involves the direction of the inequality. When doing a one-sided test, we can elect to use either of the following conditions for the null hypothesis H_0, and alternative hypothesis H_A:

$$H_0: \mu_{Site} \leq \mu_{Background} \text{ and } H_A: \mu_{Site} > \mu_{Background}$$

$$H_0: \mu_{Site} \geq \mu_{Background} + S \text{ and } H_A: \mu_{Site} < \mu_{Background} + S$$

USEPA describes these options as Background Test Form 1 and Background Test Form 2, respectively (USEPA 2002b, 2015). The two approaches can yield different conclusions when the difference between the site mean and background mean approach *S*. Some regulators prefer Background Test Form 2 because it puts the burden of proof on the data to demonstrate that the site mean is not elevated (USEPA 2002b).

Because it may require more samples than are typically collected via ISM to identify small differences between the site and background means, it may be desirable to consider looking for at least a minimum meaningful difference between the site and the background means, instead of looking for just any discernable level of difference. The hypothesis test can be set up to consider any difference or only to discern at least a particular (ideally meaningful) difference.

If the assumption of normality is not well met for either the site or background data, then a nonparametric test for comparison of medians might be appropriate. Nonparametric tests such as the Wilcoxon-Mann-Whitney test, Sign test, or Permutation test may be used to compare the midpoints (means or medians, depending on the test) of the site and background datasets and achieve error rates even under conditions of non-normality. Note that while nonparametric tests relax the assumption of normality, they do not relax the assumption of equal variance (meaning the test can yield higher than expected error rates if the variances of the site and background distributions are different).

An analysis of variance (ANOVA) test may be used to compare the means of two or more SUs. ANOVA is a parametric test

that compares the means of two or more sites to determine whether any of the site means are statistically different from one another. The null hypothesis for the ANOVA test is that all the site means are equal. The alternative hypothesis is that not all the site means are equal (that is, at least two site means are different). If comparing just one site to background, the replicate background samples will be compared to the replicate site samples to determine if they have statistically similar means. A strength of the ANOVA approach is that it is just as possible to compare several sites at once, so one background dataset could be compared to the data from several SUs. Note that an important caveat when evaluating multiple groups is that because the null and alternative hypotheses cannot be defined as inequalities, ANOVA is inherently a two-sided test of differences in means (or medians for nonparametric options). Also, if the null hypothesis is rejected, that would only mean that at least two of the sites differ, and it would not be clear which sites those are without further evaluation. Inclusion of multiple sites simultaneously in an ANOVA may require follow-up by individual comparisons of each site to background to determine which one (or more) site is statistically greater than background.

Power analysis. A power analysis may be conducted using the acceptable error rates in concert with information on the variance and the minimum difference the analysis should detect. Most online calculators and commercial software for statistical analysis can calculate the power. Such an analysis may also be conducted in the context of a background test, but the interpretation of the false exceedance error and corresponding power may differ from that of a one-sample hypothesis test (such as comparison to a threshold) depending on how the null hypothesis for comparison to background is defined. Also, the decision to apply a Background Form 1 or Background Form 2 approach can affect the statistical power, given the prescribed sample sizes and S. In the case of Form 2, the null hypothesis is that the site mean is greater than the background mean by at least S. Therefore, the following definitions of error and power apply:

- False compliance error (α) is the false positive condition – the probability of (falsely) concluding that the site mean is *not* contaminated above background by S (reject H_0), when it truly is elevated by more than S.
- False exceedance (β) is the false negative condition – the probability of (falsely) concluding that the site mean is greater than or equal to the background mean by S (do not reject H_0), when it is *not* truly elevated.
- The corresponding power (1 – β) is the probability of *not falsely* concluding that the site mean is greater than or equal to the background mean by S.

3.3.4.2 Comparison of the upper tails

In addition to comparison of the mean concentrations, it could be useful to compare site concentrations to an estimate of the upper tail of the distribution of background concentrations. The Quantile test compares the upper percentile (such as the 95th percentile) of both distributions (Gilbert 1987). If the upper percentiles of the site concentrations substantially exceed the corresponding upper percentile of background, it could suggest that small pockets of elevated concentrations exist on site that may not be consistent with background. Relating a single ISM site sample to a distribution of background concentrations is also possible (such as a 95% upper prediction limit [UPL] or 95/95 UTL). This comparison is useful if the objective is to merely demonstrate that a release above background has occurred. If normality can be assumed, UPLs may be an option for evaluating future ISM site means (USEPA 2009). Some simulation studies with UPLs demonstrated that desired power can be achieved even for small sample sizes and some non-detects in the background ISM dataset (Georgian 2020). However, simulation studies of the statistical performance of 95/95% UTLs clearly illustrate that the UTL lacks statistical power, given the usual sample sizes for ISM investigations, and can lead to much greater decision error rates than expected (Pooler et al. 2018). If achieving a high degree of confidence in conclusions related to background is critical, point-by-point comparisons to UTLs may not meet project objectives. In general, hypothesis testing is preferred over point-by-point approaches. As a supplemental exploratory data analysis step, it is also always useful to compare the sample results qualitatively, noting whether they appear similar, whether the greatest value falls in the background or the site dataset (this can also be quantitatively assessed using the Slippage test (Gilbert 1987)), and noting whether there are a similar percentage of non-detected results.

3.4 Cost-Benefit Analysis

A sampling protocol is often limited by how much funding resources are available to the property owner and/or responsible party. Thus, a sampling plan is often designed based on limited traditional sampling methodologies (that is, discrete and/or composite sampling approaches) to fit into a preexisting budget instead of an unbiased identification of the data required to meet project objectives. The result is often limited data with a high risk for decision errors. As stated in Section 2.2, heterogeneity causes sample analytical results to fluctuate depending on the precise location, particle segregation,

cohesiveness, and sample mass. ISM generally yields more precise and unbiased estimates of the mean, and the cost-benefit ratio often favors the ISM investigation because it results in fewer decision errors. Less data variability will likely allow for less uncertainty in decision-making, especially when estimates of the mean are close to an AL.

This section will assist in determining potential costs associated with implementing ISM and reference case studies to help determine the cost-benefit of utilizing ISM for a specific site. Proper planning is a major factor for designing the most cost-effective implementation of ISM (described in Section 3).

3.4.1 Costs and benefits of ISM

A simple and direct cost comparison of ISM to traditional sampling approaches is difficult due to several factors, including analytical costs, availability and quality of screening technologies, sample collection methodologies, and establishment of a clear endpoint to a project.

Keep in mind for the ISM versus discrete sample examples presented that the precision and representativeness of a single set of discrete sample data for estimation of the mean is never known, since replicate sets of discrete samples are not collected for comparison. This can impose significant uncertainty and future liability concerns for projects and is the root cause of many failed investigations and remedial actions. Unfortunately, this type of uncertainty is not easily quantifiable. As a common saying in the environmental industry goes, “How much is a lawsuit going to cost you?”

The following section describes comparative costs and considerations for project-specific evaluations. In general, ISM may have higher planning and sampling preparation costs compared to traditional sampling methods. However, providing a reliable and adequate dataset using ISM offers the potential to limit substantial costs that could otherwise be incurred due to additional sampling needs and/or unnecessary remediation.

3.4.1.1 Financial

Variations in costs between ISM and more traditional sampling methods typically occur during system planning, sampling plan review, field sampling, and laboratory services of a project. These four areas are discussed further below.

Systematic planning. Systematic planning that includes the designation of DUs, sample protocol, and associated decision statements to guide data evaluation, while being a key component of ISM investigations, is not unique to ISM, which requires the upfront consideration of DUs and associated decision statements. Although this should be done prior to any project involving soil sample collection, a more detailed planning process is required for ISM.

Omission of a more detailed planning process using discrete and/or composite sample methodology often results in multiple sample collection events to complete site characterization, followed by the designation of impacted areas using limited site characterization data on which final decision-making is based. This process often leads to the need for additional site investigations to fill data gaps or and/or final decisions based on limited data not necessarily reflective of the actual risks posed by contamination at the site.

Systematic planning costs should be roughly equivalent regardless of the sampling design. Developing DUs, discussing hot spots, including additional staff to participate in planning meetings, and stakeholder agreement may increase front-end costs but can significantly reduce cost and the need for lengthy discussions following completion of the field investigation. The intent of systematic planning is to collect usable data and minimize the need for remobilization to collect additional data or situations where parties disagree on the size of a hot spot. Eliminating both would result in lower overall cost in the project lifecycle.

UFP-QAPP worksheets are an ideal format to guide teams through systematic planning and ISM design.

Sampling plan review. A common concern of both regulators and the regulated community is ISM sampling plan review. For regulators not trained in ISM investigation approaches, the sampling plan review can be labor-intensive. For consultants, the time required for regulatory approval of ISM projects from agencies that lack adequate training and guidance documents also increases costs to their clients and at least perceived risk of rejection. In addition, many consultants find it much easier to submit standard sampling plans and assessment/remediation reports to regulators with the goal of a quicker turnaround time for their clients even if they know that this approach will ultimately result in a more drawn out and costly investigation over the life of the project. However, with time, education, and familiarity with the ISM method, the comparative costs for

sample plan review will decrease to become comparable with more traditional sampling technologies.

Field sampling. Many factors can affect the cost of ISM field sampling. All costs are highly dependent on the DQOs, and the size and number of DUs has a direct effect on the financial cost of the field effort as well. A demonstration completed by (USACE 2013a) specifically to evaluate ISM costs and performance at three shooting range sites indicated that "ISM can result in a cost savings of 30% to 60% relative to conventional grab sampling methods." Areas of cost savings in the field include the following:

- A reduced number of samples need to be prepared for laboratory analysis.
- Fewer sample supplies are consumed.
- Decontamination of field equipment is only required between DUs.
- Less time is required to survey as only the corners of the DU need to be identified.

The collection of MIs results in a longer collection time per sample than that for a single discrete sample, but that said, a relatively large number of discrete samples are needed to provide the same data quality as ISM. For the purposes of this chapter, surface sampling was used to compare costs between ISM and traditional methods only. Subsurface ISM sampling may result in higher field costs depending on the sampling method and should be considered on a site-by-site basis.

Laboratory services. As stated above, ISM decreases the number of samples analyzed to characterize a site compared with discrete sampling. For example, a DU will include analysis of three replicate ISM samples, but typically a minimum of 8 to 10 discrete samples are necessary, and as many as 20 discrete samples are commonly recommended by ProUCL guidance for risk assessment purposes, depending on analyte variability. In addition to laboratory analysis, CLP-like data packages and data validation will cost less for ISM projects as there are fewer samples requiring analysis and validation.

> ***If the overall project costs for ISM are higher than that using discrete methods for a specific project; it is worthwhile to double check if you are adequately meeting your DQOs with traditional methods.***

However, the cost savings for laboratory analysis is not solely based on the total number of samples. For ISM, it is recommended that sample processing (such as drying, sieving, and milling) is handled by the laboratory and not completed in the field. The ISM laboratory sample processing and subsampling options are described in Section 5. The price of ISM processing depends on the specific options selected, the amount of soil, the analytes of concern, and other general concerns (such as the number of samples and turnaround time). Understandably, laboratories charge for ISM sample preparation, and rates are in the range of $65 to $200 per sample, depending on the sample preparation methods required to meet the DQOs. Per sample laboratory costs, including processing, QA/QC, and analysis, can double for a single discrete sample; the savings for ISM is achieved because fewer samples are required to adequately characterize an area.

Laboratory costs will increase proportionally with the number of DUs, but savings are realized by using ISM because more discrete samples are typically required for the same spatial coverage. Further, field labor costs using ISM can be absorbed by the decrease in laboratory costs if remobilization for additional discrete sampling is required, to complete needless remediation, or to address the potential damages and legal expenses of not appropriately assessing the magnitude and limits of contamination.

3.4.1.2 Cost variations

The actual cost differences between the sample methodologies will vary from project to project based on several factors. Some of the more critical factors are the number of discrete samples required to generate a statistically defendable dataset and meet the DQOs and the actual time required to collect such samples. Other lesser factors include the following:

- the size of the property and nature of the release
- individual analyte costs
- the use of field analytical methods for specific contaminants
- the ISM sample processing method (drying, particle size reduction, sieving, subsampling)
- the presence of clay, roots, and very wet soil (likely increases overall field processing time and potentially increases costs)
- increased ISM laboratory charges for difficulties sieving the soil (clay, roots, very wet)
- the potential cost of extra ISM QC samples (that is, batch samples)

- the potential cost of large volume shipping and disposal for ISM soil/sediment projects

The link below provides a simplified comparison of field and laboratory costs for ISM and discrete sampling methods but should not be considered comprehensive. It is an interactive table such that the number of DUs and other applicable parameters can be entered, and a series of general assumptions for a generalized estimate will be provided for both methods. This table is meant to provide an example of costs for different sampling scenarios and should not be used for site-specific project estimating purposes.

ISM versus Discrete Cost Comparison Work Sheet for estimates of ISM verses discrete sampling field and laboratory costs

3.4.1.3 Schedule

As described in Section 3.3, completion of ISM requires additional systematic planning to determine the DQOs, determine the number and location of DUs, and sampling requirements. The additional planning will require additional time in the project schedule, potentially more than other sampling methods – the field schedule is more or less equal. Experience has shown that 30-sample increments in three replicates can be collected in approximately 45 minutes with two people provided the DU is less than a ¼ acre. Collection of 8 to 20 discrete surface samples in the same area would be comparable considering that decontamination would be required in between each sample and each sample would be logged and labeled for submittal to the laboratory. Additional time will be required by the laboratory for ISM sample processing, which should be considered if an expedited turnaround on sample results is required. A delay in project schedule may also occur for the review of planning documents, especially as the industry is understanding proper ISM implementation methods and procedures, which is one of the main goals of this document. These delays can be partially mitigated by increased knowledge, experience, and additional staff. Figure 3-26 provides a graphical representation of generalized project schedule with and without ISM.

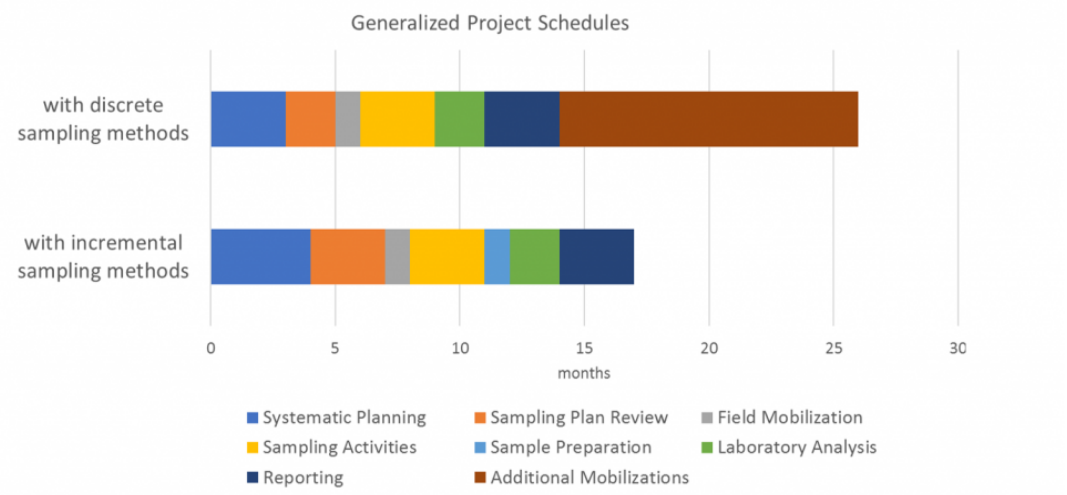


Figure 3-26. Generalized project schedules.
Source: ITRC ISM Update Team, 2020.

3.4.1.4 Decision error

Section 3.2 describes the statistical concepts behind ISM and the importance of producing statistically defendable data. Due to the nature of sampling, it is entirely possible to collect discrete samples from the same area twice and come up with contradicting conclusions regarding characterization. Without proper data, it is possible to conclude a site is dirty when it is clean or that a site is clean when it is dirty. This can lead to multiple investigations, project delays, implementation of an inappropriate remedial action, and an increased liability to the property owner, regulatory agency(ies), and project stakeholders. All of these outcomes will increase costs.

ISM provides much higher confidence in cleanup decisions, which helps ensure effective use of funds toward addressing actual environmental impacts.

3.4.1.5 Environmental impact

While less tangible than previous factors, there is a potential for inadequate characterization as it pertains to undetermined risks to human health and the environment. As environmental professionals, we understand that there are risks of incomplete assessment with all methods and techniques, and limiting that risk is our primary goal and responsibility. Therefore, employing a technically and statistically defendable method for assessment of soil is the most prudent course of action.

3.4.2 Examples

This next section provides examples that compare the costs and benefits of using ISM to more conventional sampling methods.

3.4.2.1 Site screening

ISM provides significant advantages for site screening, especially if the mode of contaminant dispersion is generally uniform across large areas. Specific examples where ISM has been used include munitions sites or ranges (including small-arms and skeet and trap ranges), former mines, landfills, and sites that were contaminated by surface application of pesticides and/or herbicides, radiation, and/or PCBs. Sites where an overall assessment of potential site risk to known receptors is required here. While this may seem counterintuitive, it is important to recognize that the statistical basis for ISM matches well with the contaminant distribution at these types of sites.

Site screening with ISM is very well suited when little information is known about the site and/or access is limited, when data collection will take place in a single mobilization (brownfield site planned or already developed for residential and/or commercial development), and a full analyte list is required. ISM provides a more complete and reliable dataset, minimizing the potential for remobilization, “step out” sampling, and data gap investigations. The case studies in Appendix A offer various examples of ISM applications on range sites, brownfields redevelopment sites, and sites with historical herbicide and pesticide application. The following examples illustrate the potential cost benefits of using ISM for site screening.

Example 1: small-arms practice range. In this example, a former small-arms practice range requires screening to assess if impacts to soil have occurred that are a risk to human health and the environment. No historical sampling has been performed, but impacts to surface soil are likely due to the accumulation and degradation of bullet fragments during range operation. In this case, the site configuration and prior use can be used to select the SUs for investigation. In Figure 3-27, a total of five SUs were selected based on the CSM identifying the likely areas of impacted soil. SUs are used in this example because the data will be used for site screening and not for decision-making purposes. See Section 3.1 for further discussion on the application of SUs and DUs.

Here, three replicate ISM samples from each SU for a total of 15 samples will be submitted for laboratory analysis. Guidance for small-arms range sites recommend that at least 50 increments are collected for each SU due to the typical size of the SUs, nature of contaminant distribution, and likelihood for contaminant heterogeneity. In this example, 50 increments are selected for the range floor and 30 increments for the firing line, target berm, and two side berms as these areas are significantly smaller and therefore less likely to be impacted by contaminant heterogeneity. (See Section 2.5.2 for a discussion on the selection of the number of increments for a specific SU.) This sampling program will collect soil from 510 locations, including increments for each replicate sample.

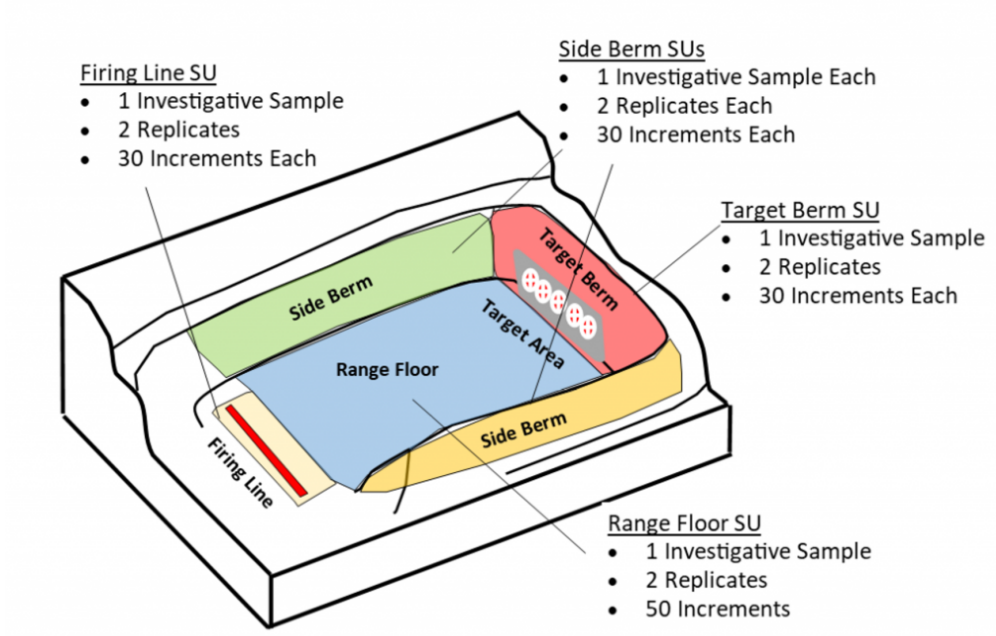


Figure 3-27. ISM sample selection at a small-arms practice range.
Source: ITRC ISM Update Team, 2020.

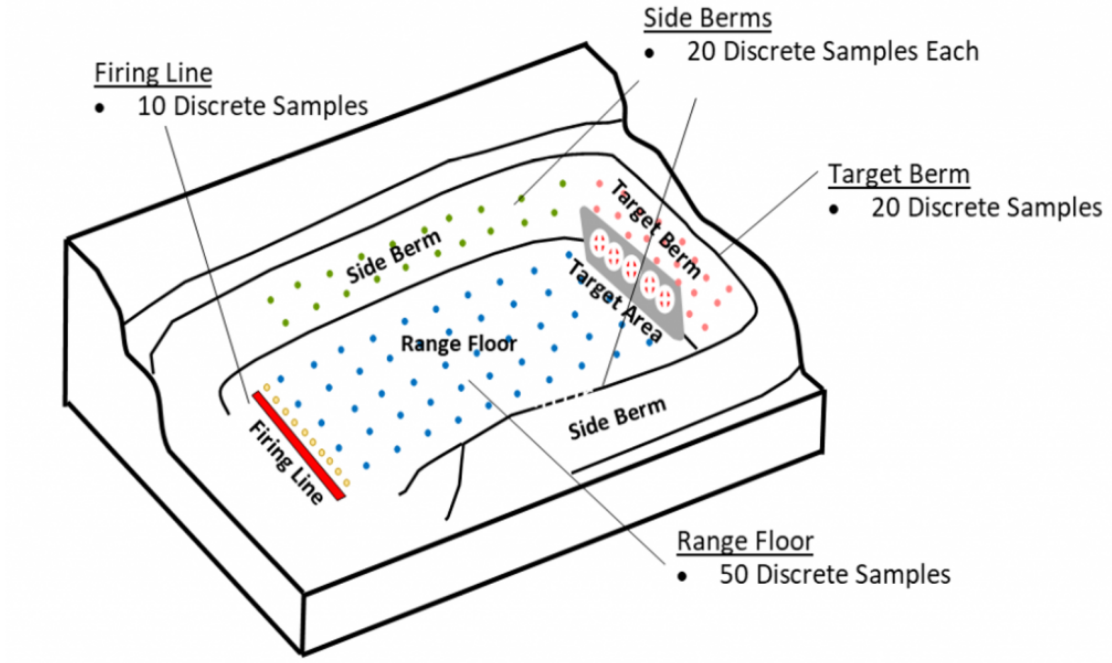


Figure 3-28. Discrete sample selection at a small arms practice range.
Source: ITRC ISM Update Team, 2020.

Table 3-7. Number of ISM verses discrete samples at a small-arms practice range.

Source: ITRC ISM Update Team, 2020.

DU	# of ISM Samples			# of Discrete Samples	
	Investigation	Replicates	Increments	Investigation	Duplicates

Firing Line	1	2	90	10	1
Range Floor	1	2	150	50	5
Side Berm 1	1	2	90	20	2
Side Berm 2	1	2	90	20	2
Target Berm	1	2	90	20	2
Total	**5**	**10**	**510**	**120**	**12**

For comparison, a typical and statistically comparable set of discrete samples were also considered, and a total of 120 discrete soil sample locations were selected: 10 locations near the firing line, 20 locations in each side berm, 20 locations in the target berm, and 50 locations for the range floor.

Consistent with ISM, all samples will be analyzed for metals, and the locations at the firing line will be analyzed for metals and explosive residues. All ISM samples will be prepared by drying, sieving, milling, and subsampling by the laboratory. Using the comparison cost calculator presented in Section 3.4.1.1, the field and laboratory costs are estimated as in Table 3-8.

Table 3-8. Cost estimate for field and laboratory costs for ISM verses discrete sampling at small-arms practice range.

Source: ITRC ISM Update Team, 2020.

Cost Calculator for Comparison Between ISM and Discrete Sampling for Range Site with Five SUs

Parameter	...Using ISM	...Using Discrete Sampling Methodology
Step 1: Determine number of samples for laboratory analysis		
# of SUs	5	n/a
Number of Investigative Samples	5	132
Number of Replicate Samples per SU	2	20%
Number of Increments per SU	30	n/a
Number of Increments for Project	450	n/a
Total Number of Samples for Laboratory Analysis	15	158
Parameter	**...Using ISM**	**...Using Discrete Sampling Methodology**
Step 2: Determine field labor, equipment, and sample shipping costs		
Est. Labor Hours per Sample (for two people)	1	1
Total Labor Hours (for two people, in whole hours)	15	158
Hourly Rate	$100	$100
Subtotal Labor Costs	**$1,500**	**$15,800**
Sampling Equipment	$500	$500
Days of Sampling	2	20
Subtotal Sampling Equipment Costs	**$1,000**	**$10,000**
Sample Shipping	1	5
	15	32
	$50	$50
Subtotal Shipping Costs	**$750**	**$1,600**
Total Estimated Field Labor, Equipment, and Sample Shipping Costs	**$3,250**	**$27,400**
Parameter	**...Using ISM**	**...Using Discrete Sampling Methodology**
Step 3: Determine laboratory and data validation costs		
Sample Processing	$100	$0
Number of Samples To Be Processed	15	158
Total Processing Cost	$1,500	$0
Per Sample Analytical Costs	$250	$250
Total Analytical Costs	$3,750	$39,500
Subtotal Sample Preparation and Analytical Costs	**$5,250**	**$39,500**
Laboratory Data Package (Stage 4)	10%	10%
	2	16
	100%	100%
Subtotal Additional Laboratory Costs for Laboratory Data Packages	**$500**	**$4,000**
Data Validation	$10	$10
	$20	$20
Subtotal Validation Costs	**$170**	**$1,740**
Total Estimated Laboratory and Data Validation Costs	**$5,920**	**$45,240**
Estimated Total Costs	**$12,420**	**$100,040**

Notes:
All costs are in U.S. dollars
Assumptions:
The time taken to collect sample will depend on a number of factors. The values used here are based on the experience. The rate of 0.5 hours for a team of two people to collect 30 increments in triplicate for 1/4 acre site is based on case studies reviewed to date. Calculation provided in this table is adjusted for when more than 30 increments are required.

The financial costs for this field mobilization using ISM are a quarter of the costs for a discrete sampling program. The likelihood of remobilization is unlikely for this site since the area of impact is well-defined to the area of the range. Other

benefits of ISM include greater representation of contaminants, more accurate decision-making that will ultimately result in an appropriate determination of risk, and selection of the most appropriate remedial alternative.

3.4.2.2 Remedial investigation and risk assessment

The remedial investigation process defines the nature and extent of environmental impacts and provides data for risk assessment. Because risk assessment is typically interested in characterizing average exposure to human or ecological receptors, ISM is more directly related to risk assessment than discrete sampling. ISM includes an added component of replicate sample data that can be used to assess the precision and reproducibility of the overall sampling method, increasing confidence in decision-making. Confidence in the resulting data is also increased because ISM typically provides greater coverage of a targeted exposure area and incorporates sampling theory requirements for minimum sample mass, as well as strict processing and analysis requirements. Such considerations are not typically incorporated into discrete sample investigations. Also, because ISM provides and higher sampling density and better spatial coverage within the DU, the confidence in the average analyte concentrations is typically higher than for discrete sampling. Case studies have demonstrated that ISM is applicable to evaluate nature and extent as well as human health and ecological risk assessment processes. Additionally, short-term delay to prepare for and execute ISM has reduced the need to remobilize to the field and resulted in a more complete dataset for evaluation of risk and remedial alternatives.

Example 2: soil sampling of a liquid spill. In this example, a chemical in solution or liquid waste is accidentally released to soil from two abandoned drums on a flat terrain (Figures 3-29 and 3-30). The liquid containing the chemical of interest (COI), in this case an SVOC, flows out from the source. It is assumed the concentration of the COI decreases with distance from the point of the spill, as shown in the diagram below. Three concentric circles of concentration, representing high, medium, and low, have been constructed and are assumed to be the SUs, but these SUs are likely to be of different sizes. Ideally, the boundaries of these SUs would represent a transition boundary at the COI soil screening level, which we assume for this example has been previously established. Area 1 represents 0.4 acres, Area 2 is 1.2 acres, and Area 3 is 2.4 acres. SUs are used in this example because the goal of the assessment is to define the lateral extent and not for risk characterization. See Section 3.1 for further discussion on the application of SUs and DUs.

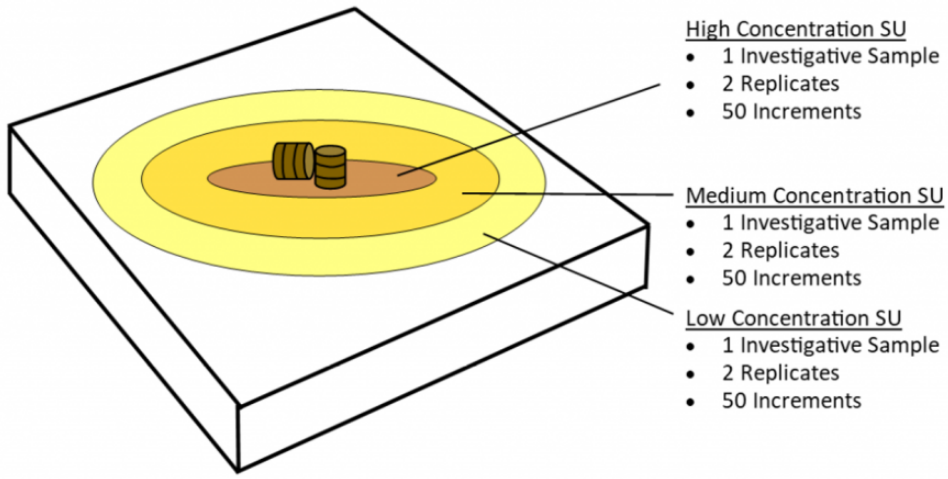


Figure 3-29. ISM sample selection for soil sampling of a liquid spill.
Source: ITRC ISM Update Team, 2020.

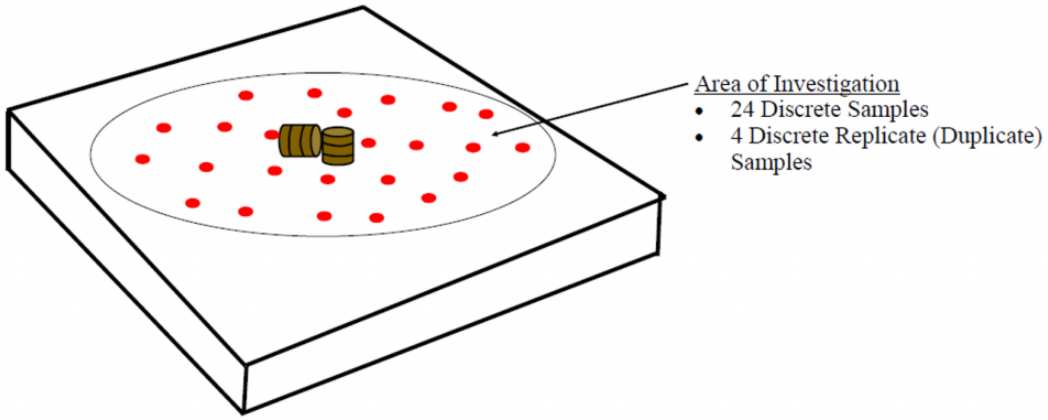


Figure 3-30. Discrete sample selection for soil sampling of a liquid spill.
Source: ITRC ISM Update Team, 2020.

For ISM samples (Figure 3-29), 50 increments are considered better because of potential SVOC soil adherence. Assuming three replicate ISM samples, the calculator indicated a cost of $9,600 see Table 3-9). For comparison, a total of eight discrete surface samples were randomly selected for the same area represented as an SU (Figure 3-30). Using the cost comparison work sheet presented in Section 3.4.1.1, the field and laboratory costs are estimated as in Table 3-9.

Table 3-9. Cost estimate for field and laboratory costs for ISM verses discrete sampling for soil sampling of a liquid spill.
Source: ITRC ISM Update Team, 2020.

Cost Calculator for Comparison Between ISM and Discrete Sampling for Point Source / Spill Site with Three SUs

Parameter	...Using ISM	...Using Discrete Sampling Methodology
Step 1: Determine number of samples for laboratory analysis		
# of SUs	3	n/a
Number of Investigative Samples	3	24
Number of Replicate Samples per SU	2	20%
Number of Increments per SU	50	n/a
Number of Increments for Project	450	n/a
Total Number of Samples for Laboratory Analysis	9	28

Parameter	...Using ISM	...Using Discrete Sampling Methodology
Step 2: Determine field labor, equipment, and sample shipping costs		
Est. Labor Hours per Sample (for two people)	1.7	0.5
Total Labor Hours (for two people, in whole hours)	16	14
Hourly Rate	$100	$100
Subtotal Labor Costs	**$1,600**	**$1,400**
Sampling Equipment	$500	$500
Days of Sampling	2	2
Subtotal Sampling Equipment Costs	**$1,000**	**$1,000**
Sample Shipping	1	5
	9	6
	$50	$50
Subtotal Shipping Costs	**$450**	**$300**
Total Estimated Field Labor, Equipment, and Sample Shipping Costs	**$3,050**	**$2,700**

Parameter	...Using ISM	...Using Discrete Sampling Methodology
Step 3: Determine laboratory and data validation costs		
Sample Processing	$100	$0
Number of Samples to be Processed	9	28
Total Processing Cost	$900	$0
Per Sample Analytical Costs	$250	$250
Total Analytical Costs	$2,250	$7,000
Subtotal Sample Preparation and Analytical Costs	**$3,150**	**$7,000**
Laboratory Data Package (Stage 4)	10%	10%
	1	3
	100%	100%
Subtotal Additional Laboratory Costs for Laboratory Data Packages	**$250**	**$750**
Data Validation	$10	$10
	$20	$20
Subtotal Validation Costs	**$100**	**$310**
Total Estimated Laboratory and Data Validation Costs	**$3,500**	**$8,060**
Estimated Total Costs	**$9,600**	**$13,460**

Notes:
All costs are in U.S. dollars
Assumptions:
The time taken to collect sample will depend on a number of factors. The values used here are based on the experience. The rate of 0.5 hours for a team of two people to collect 30 increments in triplicate for 1/4 acre site is based on case studies reviewed to date. Calculation provided in this table is adjusted for when more than 30 increments are required.

The costs for ISM in this example are lower than that of discrete sampling. The potential advantage of discrete sampling is

that each sample may provide information on the possible boundary for soil above or below the soil screening level, but the method is relying on eight samples to provide an accurate concentration, especially in Area 3 (eight samples for 2.4 acres). ISM sampling will provide a better measurement of the mean for the SU.

3.4.2.3 Remedial action

Following remedial action, ISM can provide greater confidence that target cleanup and risk goals have been met. For example, soil excavation typically requires confirmation samples at the base and sidewalls to the removal of impacted soil. Accordingly, ISM sampling that sets each sidewall and the floor as a DU would reduce the decision error associated with discrete sampling results that mis-characterize the remaining contamination and lower the potential for needless continued costly excavation and soil treatment.

Example 3: remedial excavation confirmation sampling. In this example (Figure 3-31 and 3-32), an excavation was completed to remove soil impacted with petroleum hydrocarbons resulting from historical release at a former underground tank farm. The resulting excavation was initially completed to a depth below the tank invert where no visual or olfactory signs of impacted soil were observed (approximately 15 ft below the ground surface). The lateral extents of the excavation extended over an area approximately 50 ft by 50 ft. To confirm that the contaminated soil was removed to levels that would not pose a risk to human health or the environment, confirmation sample are required. One of two options for sampling could be used.

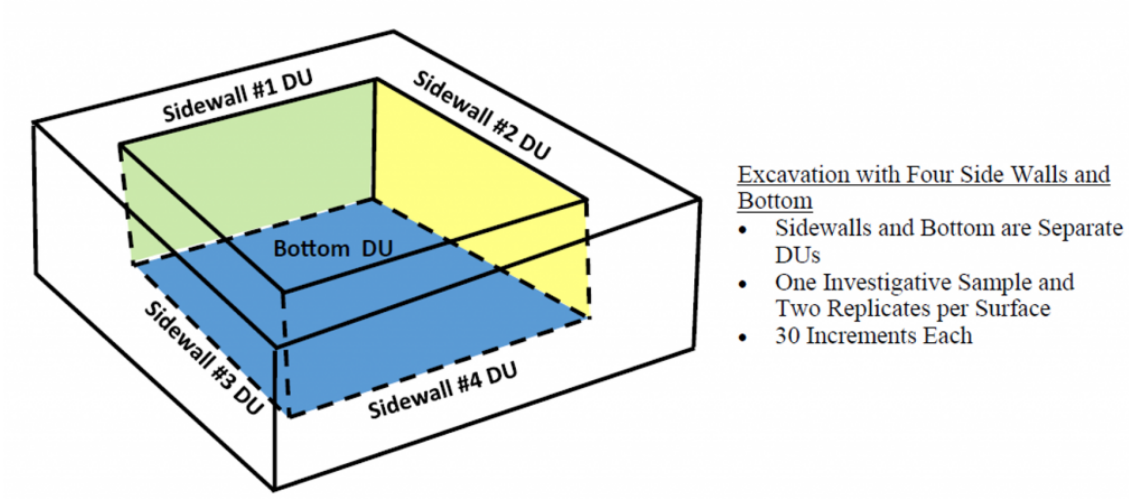


Figure 3-31. ISM sampling for remedial excavation confirmation sampling.
Source: ITRC ISM Update Team, 2020.

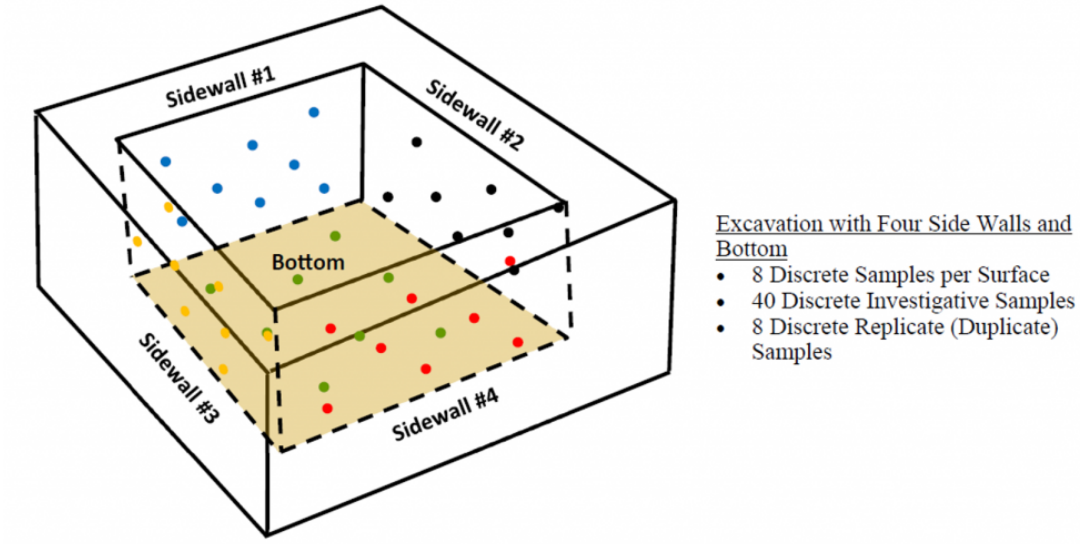


Figure 3-32. Discrete sampling for remedial excavation confirmation sampling.
Source: ITRC ISM Update Team, 2020.

A total of five DUs were selected, each with three replicate ISM samples (Figure 3-31), with a total of 30 sample increments collected for each sample. Alternative, discrete sample methods include a total of eight samples from each sidewall and from the bottom of the excavation for a total of 40 investigative samples and eight additional samples collected as duplicates (Figure 3-32). A total of 48 discrete samples are required to be statistically comparative with ISM and account for soil and contaminant heterogeneity.

Using the cost comparison calculator presented in Section 3.4.1.1 the field and laboratory costs are estimated as in Table 3-10.

Table 3-10. Cost estimate for field and laboratory costs for ISM verses discrete sampling for a remediation excavation.
Source: ITRC ISM Update Team, 2020.

Parameter	...Using ISM	...Using Discrete Sampling Methodology
Step 1: Determine number of samples for laboratory analysis		
# of DUs	5	n/a
Number of Investigative Samples	5	40
Number of Replicate Samples per DU	2	20%
Number of Increments per DU	30	n/a
Number of Increments for Project	450	n/a
Total Number of Samples for Laboratory Analysis	15	48
Parameter	...Using ISM	...Using Discrete Sampling Methodology
Step 2: Determine field labor, equipment, and sample shipping costs		
Est. Labor Hours per Sample (in whole hours)	1	0.5
Total Labor Hours (for two people, in whole hours)	15	24
Hourly Rate	$100	$100
Subtotal Labor Costs	$1,500	$2,400
Sampling Equipment	$500	$500
Days of Sampling	2	3
Subtotal Sampling Equipment Costs	$1,000	$1,500
Sample Shipping	1	5
	15	10
	$50	$50
Subtotal Shipping Costs	$750	$500
Total Estimated Field Labor, Equipment, and Sample Shipping Costs	$3,250	$4,400
Parameter	...Using ISM	...Using Discrete Sampling Methodology
Step 3: Determine laboratory and data validation costs		
Sample Processing	$100	$0
Number of Samples to be Processed	15	48
Total Processing Cost	$1,500	$0
Per Sample Analytical Costs	$250	$250
Total Analytical Costs	$3,750	$12,000
Subtotal Sample Preparation and Analytical Costs	$5,250	$12,000
Laboratory Data Package (Stage 4)	10%	10%
	2	5
	100%	100%
Subtotal Additional Laboratory Costs for Laboratory Data Packages	$500	$1,250
Data Validation	$10	$10
	$20	$20
Subtotal Validation Costs	$170	$530
Total Estimated Laboratory and Data Validation Costs	$5,920	$13,780
Estimated Total Costs	**$12,420**	**$22,580**

Notes:

All costs are in U.S. dollars

Assumptions:

The time taken to collect sample will depend on a number of factors. The values used here are based on the experience. The rate of 0.5 hours for a team of two people to collect 30 increments in triplicate for 1/4 acre site is based on case studies reviewed to date. Calculation provided in this table is adjusted for when more than 30 increments are required.

For this example, costs for ISM are much lower than the costs of discrete sampling, and the data used to make decisions are less likely to mis-characterize remaining contamination and lower the potential for remobilization and continued excavation and soil treatment.

3.4.2.4 Classification for waste deposition

ISM can provide a more defensible method for waste characterization by reducing the sampling errors associated with stockpile sampling based on discrete or traditional composite sampling methods. Composite sampling in the sense of combining multiple samples (or, more specifically, increments) from a single DU volume is an improvement over discrete sampling methods and is an approved method for waste characterization. However, ISM provides a more structured approach and ensures both that final bulk samples are prepared by combining waste from a minimum number of points and that the final mass of the sample meets minimum requirements under sampling theory. ISM will reduce the risk that the soil is mis-classified and either denied at the door of a facility or sent as hazardous material when it may have been classified as non-hazardous.

Example 4: stockpile sampling from agricultural field soil. In this example (Figure 3-33 and Figure 3-34), stockpile soil originating from a former agricultural field will be used as fill for the development of a commercial property. It is assumed that arsenic-based pesticides were applied to the agricultural field uniformly with limited mobility and degradation. A stockpile volume of approximately 20,000 ft^3 (800 yds^3) is assumed and equates to the approximate volume of soil to cover a 1-acre site at a 6-in depth. It is also assumed that the existing stockpile can be flattened or spread out sufficiently so that the interior of the pile can be accessed with a hand sampling device.

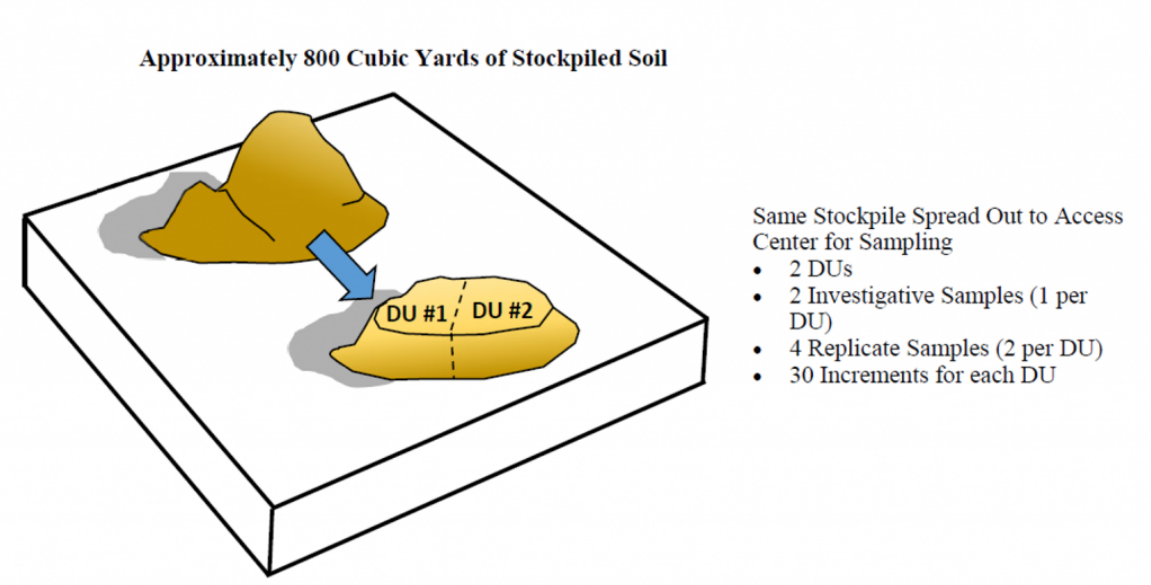


Figure 3-33. ISM stockpile sampling of agricultural field soil.
Source: ITRC ISM Update Team, 2020.

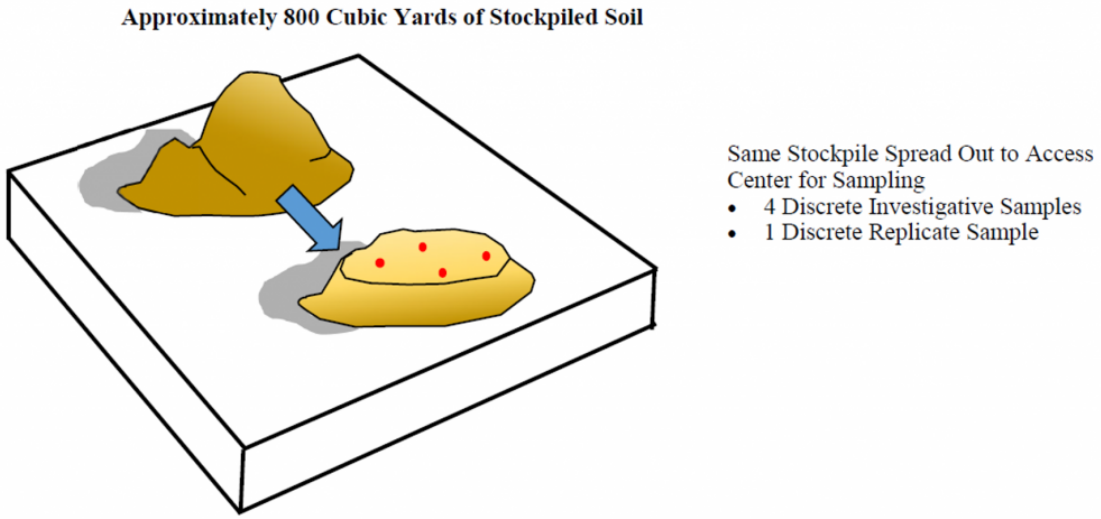


Figure 3-34. Discrete stockpile sampling of agricultural field soil.
Source: ITRC ISM Update Team, 2020.

Using ISM, one DU per 400 yds^3 is recommended as the default exposure area based on the "Guidance for Soil Stockpile Characterization and Evaluation of Imported and Exported Fill Material" (HDOH 2017a). For this example (Figure 3-33), two DUs are required for 800 yds^3 of soil. Arsenic-based pesticides are generally water soluble, allowing homogeneous application, so 30 increments were selected to demonstrate representativeness.

The California Department of Toxic Substances guidance, "Information Advisory, Clean Imported Fill Material" (DTSC 2001), provides a table recommending the number of soil samples for a range of soil volumes. This table was used to determine the number of discrete soil samples to be collected from the stockpile (four soil samples), along with the required laboratory analysis (pesticides and metals; Figure 3-34). The collection of background samples for metals are not included in the cost analysis. Using the comparison cost calculator presented in Section 3.4.1.1, the field and laboratory costs are estimated as in Table 3-11.

Table 3-11. Cost estimate for field and laboratory costs for ISM verses discrete sampling of stockpile from agricultural field.
Source: ITRC ISM Update Team, 2020.

Parameter	...Using ISM	...Using Discrete Sampling Methodology
Step 1: Determine number of samples for laboratory analysis		
# of DUs	2	n/a
Number of Investigative Samples	2	4
Number of Replicate Samples per DU	2	20%
Number of Increments per DU	30	n/a
Number of Increments for Project	180	n/a
Total Number of Samples for Laboratory Analysis	6	5
Parameter	...Using ISM	...Using Discrete Sampling Methodology
Step 2: Determine field labor, equipment, and sample shipping costs		
Est. Labor Hours per Sample (for two people)	1	0.5
Total Labor Hours (for two people, in whole hours)	6	2.4
Hourly Rate	$100	$100
Subtotal Labor Costs	$600	$240
Sampling Equipment	$500	$500
Days of Sampling	1	1
Subtotal Sampling Equipment Costs	$500	$500
Sample Shipping	1	5
	6	1
	$50	$50
Subtotal Shipping Costs	$300	$50
Total Estimated Field Labor, Equipment, and Sample Shipping Costs	$1,400	$790
Parameter	...Using ISM	...Using Discrete Sampling Methodology
Step 3: Determine laboratory and data validation costs		
Sample Processing	$100	$0
Number of samples to be processed	6	5
Total Processing Cost	$600	$0
Per Sample Analytical Costs	$250	$250
Total Analytical Costs	$1,500	$1,200
Subtotal Sample Preparation and Analytical Costs	$2,100	$1,200
Laboratory Data Package (Stage 4)	10%	10%
	1	1
	100%	100%
Subtotal Additional Laboratory Costs for Laboratory Data Packages	$250	$250
Data Validation	$10	$10
	$20	$20
Subtotal Validation Costs	$70	$58
Total Estimated Laboratory and Data Validation Costs	$2,420	$1,508
Estimated Total Costs	$5,220	$3,088

Notes:

All costs are in U.S. dollars

Assumptions:

The time taken to collect sample will depend on a number of factors. The values used here are based on the experience. The rate of 0.5 hours for a team of two people to collect 30 increments in triplicate for 1/4 acre site is based on case studies reviewed to date. Calculations provided in this table are adjusted for when more than 30 increments are required.

In this example, the costs for ISM are approximately twice the costs for discrete sampling. However, the four ISM samples combine soil from 200 points within the stockpile (assuming 30-increment samples), versus four for the discrete samples, and represent a far greater total mass (4 kg to 8 kg versus <1 kg for discrete samples) and were strictly processed at the laboratory to ensure that the data generated were representative of the samples provided (not required for discrete samples). The resulting data are well worth the additional costs and effort, especially if the soil is to be reused in areas where regular exposure could occur.

Regulatory Context Field Implementation, Sample Collection, and Preparation

Before proceeding with field implementation, the following checklist should be consulted to ensure that adequate planning unique to ISM has occurred. Have you:

- DQOs
    - identified the problem/decision to be made
    - identified objectives/goals
    - ensured that the inputs I have designed meet these objectives
    - made sure that I understand the CSM (see Section 3.1.2 through 3.1.4)
- DUs/SUs (see Section 3.1.5.1 and Section 3.1.5.2)
    - evaluated size and depth relative to decisions to be made
    - evaluated if the DU design can be adjusted to serve all data needs (statistics, nature and extent, background, risk assessment, and so on)
    - mapped out shape and considered site's physical constraints
- increments and samples
    - identified the appropriate number of increments per DU/SU (see Section 2.5.2 and Section 3.1.5.5)
    - identified number of replicates based on site characteristics and heterogeneity (see Section 2.5.3.3 and Section 3.1.5.5)
    - considered if pilot or early replicates for unassessed areas are warranted prior to full-scale implementation
    - designed an increment sampling path that ensures unbiased locating of increments (see Section 3.1.5.4)
    - considered resulting ISM sample size relative to scale of decision-making
- laboratory
    - confirmed the processing steps to be conducted by the laboratory (see Section 3.1.5.3 and Section 5)
    - confirmed the sample analysis procedures to be used by the laboratory (see Section 3.1.5.3 and Section 5)
    - considered QC samples and frequency needed for data validation and statistical analysis (see Section 3.2)

After Section 4, the following checklist should also be able to be completed to ensure proper field preparation and implantation. Have you:

- allocated appropriately trained staff to execute the task
- identified site-specific means of marking out a DU
- calculated the increment size based on the sample design total volume
- identified the appropriate tool(s) to obtain each increment from the required depth
- assessed if a mass reduction technique will be needed during increment collection
- requested/obtained sample containers specific to ISM (large volume, special considerations for VOCs for ISM)
- considered the added resources for sample storage (added ice and coolers)
- communicated to the laboratory the required processing and QC requirements for the chain of custody

4.1 Introduction

The objective of ISM field implementation is to collect increments and produce samples that result in a reproducible estimate of the mean concentrations that control sampling error using practical methods.

The objective of this section is to describe the practical methods for collecting consistently sized increments for surface soil, subsurface soil, and sediment from various environments. The sampling method includes guidance on field planning, locating samples, sampling tools, collection and field processing procedures, decontamination, sample handling, and sample

shipping.

This section assumes the systemic planning, DU layout, and statistical sample design have already been completed. Moreover, the DU/SU sizing, location, and thickness have been planned; the number of replicates and increments and the increment layout (systemic or random) have been decided; the laboratory processing steps have been selected; and, from this, a target sample mass has been calculated according to Section 3.1.5. This section will describe appropriate increment sizing to achieve the desired sample mass, the selection of appropriate tools to acquire the increments, procedures for collection and sampling handling, and planning matters to be considered prior to mobilization to allow for efficient and successful field implementation.

4.2 Field Planning

While implementation of ISM sampling has a lot of similar planning steps to conventional discrete soil sampling, there are special considerations to be made prior to field mobilization specifically for ISM. The following sections discuss these considerations, including field personnel training, site conditions and features that impact tool selection, increment sizing to allow for appropriate tooling or bottleware, and special considerations for subsurface investigation.

4.2.1 Field personnel training and staffing

General soil sampling and handling techniques are assumed to be part of the knowledge base for ISM field collection personnel. However, as with all sampling approaches, a basic understanding of the sampling objective specific to ISM can assist in field decisions and allow for collection of improved data. Note that site-specific tooling may vary from those demonstrated in the training; all users should follow manufacturer recommendations to safely and correctly deploy tooling on project sites. A list of commonly used tooling is organized by application in Section 4.3.

With respect to staffing, although ISM sample collection may be performed by a single individual, a two-person team is often the most efficient method. Ideally, one person collects the increments, and the other holds the sample container (such as a clean polyethylene bag) and keeps track of the number of increments. However, site conditions may dictate that three or more individuals are required for the collection of a single ISM sample.

4.2.2 Site conditions, surface features, and soil characteristics

> ***Soil density across the DU should be similar. Contaminant distribution within different soil types should also be considered when determining DUs.***

Whenever possible, data collection within each DU should be completed during stable weather conditions. Sampling across a DU may require more time than other grab sample techniques – for example, variability in extreme heat/cold, wind, rain, or snow may necessitate a delay in sampling activities to ensure moisture consistency between replicates or SUs. Variable moisture can result in longer sample dry times and other sample handling problems. As such, planners should account for the weather forecast prior to mobilization.

Soil density across the DU should be reasonably uniform (meaning the same general soil classification can be expected throughout the DU). Surface features such as asphalt, vegetation, gravel, and so on should be considered as well. When the surface of the DU contains both vegetated and non-vegetated areas, it is likely that less soil (less increment mass) will be obtained from the vegetated regions within the DU. If a site has obvious areas with different soil lithologies and/or densities (such as areas of sand with areas of fat clay, areas of peat, and so on), those different soil type areas should be factored into DU determinations (location, shape, size of DUs). Assumed differences in contaminant concentrations in the different soil types should also be considered. In these cases, it may be necessary to redefine the DU to account for the possible heterogeneity of contaminant concentration. Moderately different soil types within a single DU may warrant different sampling tools (such as for areas of a DU with increased gravel), which requires the sampler to ensure that increment sizing remains consistent between the different tools.

Extensive discussion of DU design is provided in Section 3, however, in some cases, these DUs may have been designed or established in a desktop environment based on historical information. Current site features encountered in the field may warrant reevaluation of DUs, and the project team should be notified. Field personnel should be conscious of the potential

impact of DU heterogeneity on data collection and results.

4.2.3 Increment sizing

Based on the required final mass of the ISM sample as dictated by the sampling design (refer to Section 3.1), the minimum mass and/or volume of individual increments must be calculated prior to sampling to allow for the selection of appropriate collection tools.

The mass of any single increment depends on the depth of interest, soil density, moisture content, and the diameter or size of the sample collection tool. Individual increment mass should be similar, provided the soil density and DU thickness are fairly uniform. Typically, however, individual increments are not weighed in the field during collection. Similar mass per increment is assumed with similar volume collected. Due to practical limitations, increments of similar volume rather than of similar mass are collected, provided that the thickness of the DU is fairly uniform.

The number of increments to be collected per DU, the sampling depth, and the targeted mass of each sample should all be specified in the sampling design. The following formula can be used for estimating sampling equipment requirements (such as the core diameter) based on a predetermined sample mass and number of increments by adjusting the variables as needed (Walsh 2009):

$$M_s = p \cdot n \cdot D_s \cdot \pi \cdot (\varnothing/2)^2$$

where

M_s = targeted mass of sample (g)

D_s = increment length (cm)

n = number of increments

p = soil or sediment density (g/cm3)

$\varnothing$ = diameter of sample core (cm)

Figure 4-1 is an example for estimating how much mass will be collected for the total sample for a given soil density and number of increments.

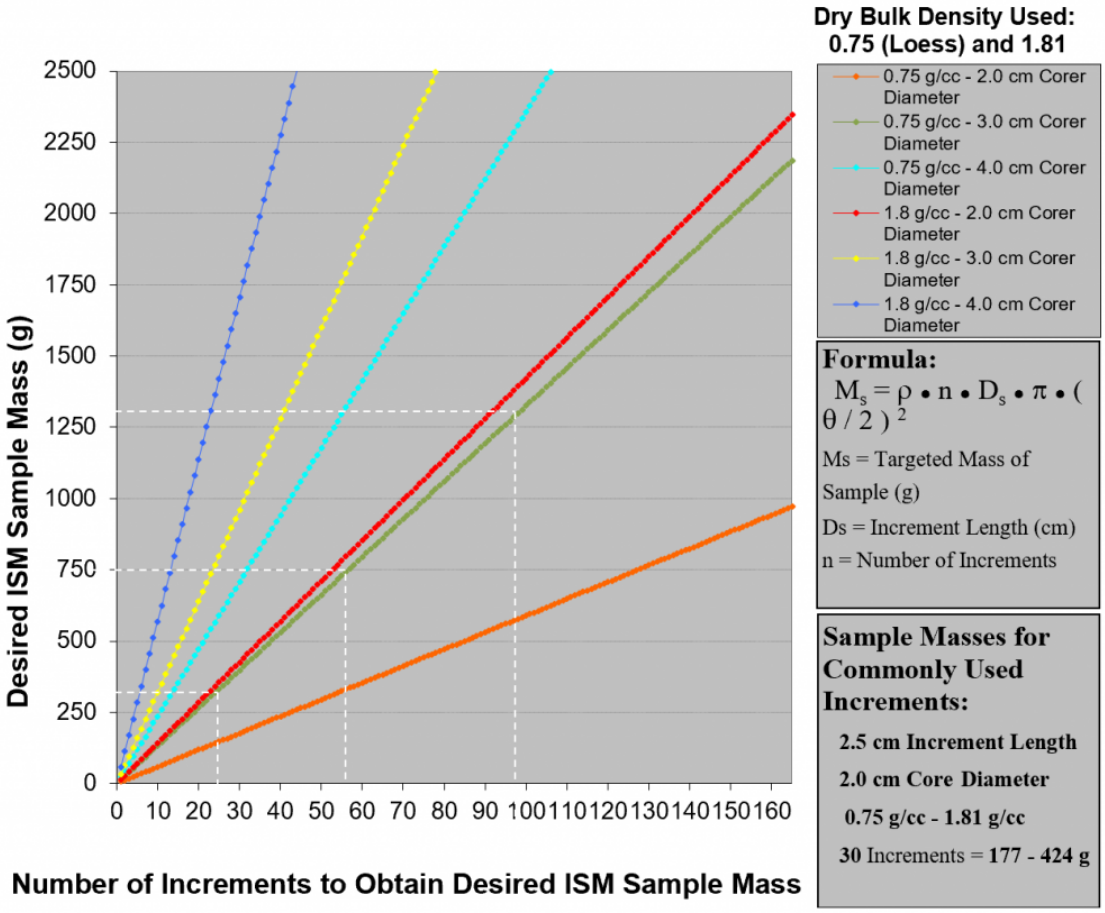


Figure 4-1. Estimated sample mass based on number of increments and soil density.
Source: ITRC ISM-1 Team, 2012.

4.2.4 Special planning for subsurface assessment

As discussed in Sections 2 and 3, DUs are, by definition, three-dimensional in nature and intended to focus the investigation on a specified volume or mass of soil. Obtaining good spatial coverage and data quality for subsurface soils can be more challenging but is nonetheless achievable. The objectives for subsurface investigations may be similar in nature to surface investigations – for example, the goal may be to estimate the representative concentration of targeted contaminants for targeted depth intervals (within the defined vertical limits), to determine or confirm the lateral boundaries of the source area, or for remedial purposes, to estimate contaminant mass within the DU (say, the mass of tetrachloroethene for the design of a soil vapor extraction system or the mass of dioxins to design an in situ thermal desorption system).

The practical application of ISM sampling to subsurface assessment must be carefully considered during project planning and factor in the labor- and equipment-intensive operations associated with obtaining subsurface sample volume. Sampling of a surface DU in the same spatial area may only require a hand tool corer and sample containers. However, sampling the same area in the subsurface may require the sampler to use drilling equipment, perform numerous borings (often 30 to 90 shallow borings), or use mass reduction techniques, all of which significantly contribute to the resources required for sampling. Often, alternative sampling techniques (such as discrete sampling, field screening, or field analytical methods) may be more practical, applicable, and/or cost-efficient (see Section 3.4) compared to subsurface ISM sampling, but under limiting conditions related to budget, time, or site access, subsurface ISM can be adapted to meet project objectives.

4.2.5 Sample containers

The analytical laboratory should be consulted prior to sample collection to discuss sample containers, sample handling, solvent type and volume, shipping of samples in methanol, anticipated analytical detection limits, etc.

Sample containers should be selected based on the predetermined sample volume. Sample containers capable of holding a minimum of 1 kg of soil are recommended. Additional container considerations include analytes, soil type, moisture, and plan for field collection – for example, do you plan to collect increments directly into the sample container or into a larger bucket first for later transfer to the sample container? In most non-volatile analyses, a 1-gallon resealable plastic bag is appropriate, however, plastic is not compatible with certain analytes and regulatory programs, so multiple 1-liter glass containers or high density polyethylene (HDPE) may be needed.

ISM samples can be collected for VOC contaminant analyses, with ISM increments placed directly into the appropriate volume of methanol in the field.

To avoid the loss of volatiles associated with the heavy processing that occurs with typical ISM samples, VOC ISM samples require special bottleware consideration:

- Bottles that have a narrow neck or other means of restricting volatilization losses and containing the volume of appropriate solvent[1] should be prepared prior to the sampling activity. Typically, the bottle and solvent are prepared and pre-weighed at the laboratory prior to shipment to the field. This method allows for laboratory calculation of the final ISM soil mass. The volume of solvent should at least equal the mass of soil that will be introduced (1:1), thus, the sample mass must be predetermined. The headspace to preserved sample ratio (methanol + sample) should be less than or equal to that commonly achieved with discrete methanol VOC preserved samples (that is, ~32 mL headspace to 8 mL preserved sample). Soil increments should remain completely submerged in methanol at all times. If increments are combined in the field, it is important to use a volume of methanol large enough to accommodate all the increments.
- The larger volume of methanol could be subsampled in the field prior to shipment to the laboratory. With this option, the complete ISM methanol-preserved sample is disaggregated and extracted in the field by shaking periodically for at least 24 hours, allowing the solids to settle, decanting or pipetting 20 to 30 mL of methanol into a vial, and shipping this aliquot to the laboratory for analysis. The total mass of the ISM soil sample, as well as the total volume of methanol, must be recorded and provided for the laboratory.
- Increments for VOC analysis could be collected and preserved with methanol individually (for example, 5 g soil in 5 mL methanol in volatile organic analysis vials per USEPA SW-846 Method 5035A (USEPA 2002f)and submitted to the laboratory for a combination of methanol aliquots before analysis. The laboratory would remove equal aliquots of methanol from all individual increment vials and combine them in a single vial to represent the complete ISM VOC sample, using the methanol handling techniques described in USEPA SW-846 Method 5035A (see Figure 5-11). This option also allows for analysis of individual increments or alternate combinations of increment groups, if required. Additionally, this option allows flexibility for varying the number of increments without having a large variety of large volume ISM sample bottles. Disadvantages include increased supplies, labor costs, and sample tracking logistics.
- Similarly, but with some field combination, volatile subsets could be collected in groups, preserved with methanol in the field in larger volumes of methanol, and submitted to the laboratory for combining before analysis. For example, six increments of 5 g each would be collected in an appropriate container containing 30 mL of methanol, then five of these volatile subsets would be collected for a 30-increment ISM sample and submitted to the laboratory. The laboratory would then combine equal methanol aliquots from the five subsets for analysis.
- Individual increments could be collected in separate sampling devices that have vapor-tight seals and are designed for zero headspace (Core N' One™, EnCore, or equivalent type samplers), and then submitted to the laboratory at the appropriate temperature and within appropriate time frames (typically 24 to 48 hours) for combined placement in methanol before analysis.
- Unpreserved soil volume for percent moisture should be collected in the same manner as the ISM VOC samples, with a second increment collected at each ISM increment location and placed in an unpreserved container (4 oz or larger) and submitted to the laboratory.

A potential drawback of ISM for VOCs is that the methanol preservation approach does result in lower sensitivity. The methanol dilution step causes elevated analytical detection limits compared to the direct soil purge-and-trap and low-concentration method techniques. Analytical detection limits could be elevated above relevant screening levels for certain targeted contaminants. Selective ion monitoring (SIM) methods can also be considered to improve the detection limits. The

industry is continually researching and updating approaches for this methodology, and current guidance can be referenced for newer and/or better approaches.

4.3 Field Locating

Field locating refers to the process of defining a DU, increment sampling points, and planned travel paths. The importance of field locating is a process used with other sample collection methodologies and further discussed in the following sections.

4.3.1 Locating increments

For a square, rectangular, circular, or other naturally or structurally defined DU (such as a 5-m perimeter around the exterior of a building) on an open plane, DU boundaries can be located with a field tape and subdivided or gridded into uniform cells or subareas based on the desired number of increments to be obtained. However, for oddly shaped, obstructed, vegetated, or uneven terrain DUs (mountainous, hummocky, steep), a global positioning system (GPS) or global navigation satellite system (GNSS) device could be used to delineate the DU and help space the increments. This approach requires layout of the grid over the DU in computer-aided design (CAD) or geographic information system (GIS) and loading of the grid/increments into the GPS or GNSS prior to mobilization.

Depending on the size of the DU and terrain features, placement of markers (such as pin flags and posts) at the corners or edges can assist with a visual delineation of the cells or subareas where increments are to be collected. These markers can define, for example, lanes, grids, and collection points.

When DUs are square or rectangular, the conversions for the spacing (steps) between increment collection points (cells) are fairly straightforward to calculate – for example, a square-shaped DU can be divided into five rows, with six increments collected from each row. Row lengths and increments per row can be modified as needed for oddly shaped DUs, as well. However, with other shapes, it is recommended that the perimeter be marked and flags prepositioned across the DU in one or more perpendicular lines. Then a trial run with no sample collection can be performed to quickly establish the distance between increment collection points to achieve the desired number of increments, using the flags as guides. In these instances, the grid layout may result in plus or minus the target number of increments.

It is not typically necessary to document the exact location of each increment after collection as is typical with discrete samples. The increments contributing to the DU mean are not assessed individually and will not require precise recollection or reproducibility.

4.3.2 Sample approaches and travel paths

Sample increments can be collected using one of three different approaches: simple random sampling, random sampling within a grid, and systematic random sampling. Examples of these travel paths are depicted in Section 3.1 and summarized in Table 4-1. The travel path approach will have been selected during the planning stages to meet the DQOs, but field planning for the selected travel paths prior to collection of the increments in each DU is recommended, especially for the larger DUs based on the number of replicate samples. For example, a DU with a lot of terrain and vegetation may require more preparation of the travel path than a mown level lawn where a grid is easily laid out in the field.

As discussed in Section 3.3, the goal of increment spacing is to represent each part of the DU equally and without bias. Even spacing of replicates serves two purposes: the classical stats assumption of independent field replicates (Section 3.2.6.7) and spatial consideration of hot spot size that is able to be incorporated (Section 2.5.2.3).

Table 4-1. Unbiased increment collection design (see Figure 3-5 for a visual comparison).

Source: ITRC ISM Update Team, 2020.

Simple Random Sampling	Random Sampling in Grid	Systemic Random Sampling
The selection of all increment's locations in a DU by random selection without gridding (formal approach to random sampling is required).	The DU is divided into a grid and random (computer-based) selection of the location in each cell of the grid is performed.	The DU is divided into a grid, a random sampling location is identified within the first grid cell, and the remaining increments are obtained from adjacent grid cells at the same relative location. Replicates repeat this pattern in another location within each grid cell.

4.4 Sampling Tools and Methods

The selection of the appropriate sampling tool for an ISM sample depends on the cohesiveness and composition of the soil substrate. To minimize the increment extraction and delimitation errors described in Section 2.5.5, the sampling tools should meet the following criteria:

The sampling tool should obtain cylindrical or core-shaped increments of a constant depth from the presented surface, when possible.

The sampling tool should obtain cylindrical or core-shaped increments of a constant depth from the presented surface.

- The sampling tool should be capable of collecting an appropriate increment size considering the target sample volume required to be submitted to the laboratory and the number of increments (see Section 4.1.4).
- The diameter of the sampling tool should be a minimum of three times the diameter d of the largest particle present in a coarse matrix ($d \geq 3$ mm), so $3d + 10$ mm for a fine material (Pitard 2019). In general, sampling tools should have a diameter of at least 16 mm.
- The tools shall equally retain all particles over the entire depth of interest. For less cohesive soils, attempts should be made to retain the entire, complete core increment.
- The sampling tool should consider the nature of the contaminant being analyzed. Sample tools for non-volatile compounds or metals would be different from those used for VOCs.
- The sampling tool should consider operator safety and success. Power tools may be required for dense, hard, dry soil or clays.
- A variety of tools to address different soil types or site conditions should be taken into the field for any given project.

See Table 4-2 for examples of sampling tools for both volatile and non-volatile ISM sample collection and Figure 4-2 for examples of sampling tools for ISM collection of VOCs. These are provided as examples only to depict the general equipment – specific brands are not endorsed, and many different brands of equipment are available on the market. Various other hand augers, core sampling tools, step probes, and so on are available from environmental or agricultural suppliers and are applicable to ISM if the specifications meet project DQOs.

Table 4-2 ISM sample equipment.

Source: ITRC ISM Update Team, 2020.

Equipment[1]	Analysis		Surface Soil		Sediment	Subsurface Soil		Soil Excavation Sidewall/Base	Stockpile Soil
	VOC	Non-VOC[2]	Loose/Fine	Compact/ Course		Shallow	Deep		

Flat-bottomed scoops, trowels, or sampling spoons (to be used in conjunction with appropriate increment measuring device)		X	X		X			X	X
Thin-walled sampling tube or coring device	X	X	X	X	X			X	
Rigid probe (open-slot push probe, alligator probe, tile samplers); be aware of tip cross-contamination issues	X	X	X						X
Interchangeable tip tool (limited in depth; more options for courser soil)	X	X		X	X				X
Electric hammer and spade bit	X	X		X					
Narrow spade, pri bar, and/or mattock		X		X					
Drill with open flight auger, foot plate, and bucket collector (requires overdrilling to obtain bottom of increment)		X		X					
Drill with sample core bit		X		X				X	X
Hand auger		X		X		X		X	X
Drive sampler (manual, hydraulic, or electric)	X	X		X		X			
Plastic syringe/plug tool	X		X	X (?)	X	X	X	X	X
Sediment sampling tube		X			X				
Vibracore device with sampling tube	X	X			X				
Direct push rig with drill rods	X	X				X	X		
Solid stem/hollow-stem augers with split-spoon/barrel samplers	X	X				X	X		
Sonic drill rig with core sleeves/bag	X	X	X	X		X	X		
Backhoe bucket/loader bucket		X						X	X
Rotary Drilling (air and mud)		X				X[3]	X[3]		

Notes:

[1] This list is in addition to the typical equipment required for all sample acquisition.

[2] Sample collection activities for energetics must occur only in the presence of military explosive ordnance disposal (EOD) personnel or qualified unexploded ordnance (UXO) technicians (US ACE, "Protocols for Collection of Surface Soil Samples at Military Training and Testing Ranges for the Characterization of Energetic Munitions Constituents." July 2007.

[3] To be used with caution and for specific application.

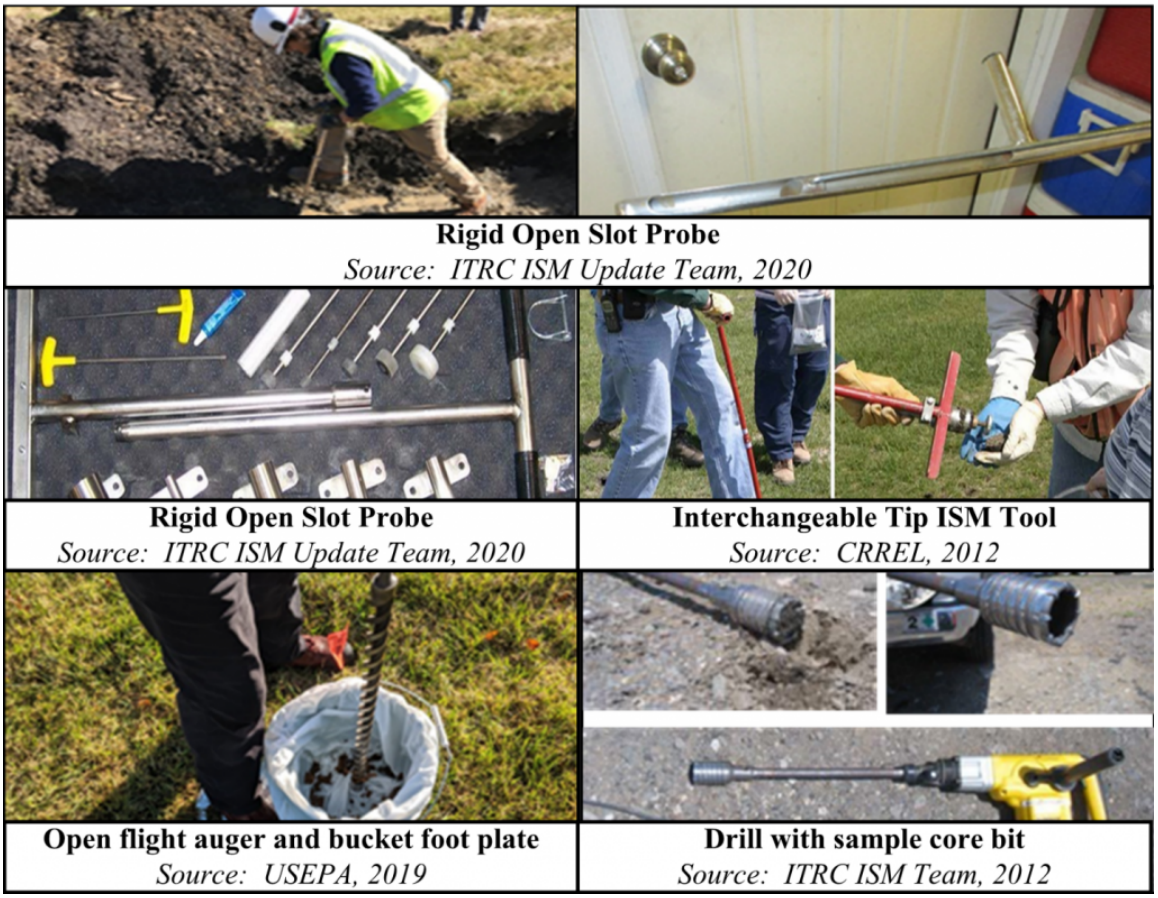


Figure 4-2. Examples of ISM sampling tools.
Source: ITRC ISM Update Team, 2020.

4.4.1 Surface soil and sediment ISM sampling

Surface soil ISM sampling is the collection of increments from an exposed and readily sampled (save vegetation) layer of soil or sediment at the surface. The grid and sampling pattern should be laid out according to Section 4.2, and in some instances, practicing the sampling path prior to the actual collection of increments can facilitate the process.

The ISM sampler starts in one corner or end of the DU and collects an increment using an appropriate sampling tool. Determination of sampling paths is discussed in Section 4.2.2. Increments are placed in an appropriate container (such as a bucket, bowl, bag, or jar), and this process is repeated at all remaining increment positions with all the increments placed together in one container. The total increments now represent the ISM sample for that DU, replicate, or SU, and should be provided to the laboratory and not further homogenized or subsampled (other than the ISM processing procedures described in Section 5). This process can then be repeated for a second and third (or more) replicate or SU.

If surface DUs are thick (more than 1 ft below ground surface), some elements of the following subsurface ISM sampling (see Section 4.3.2) and mass reduction sections (see Sections 4.5.1 and 4.5.2, respectively) may become applicable.

4.4.2 Subsurface soil ISM sampling

ISM subsurface sampling is typically completed by using techniques that produce a core that meets the size and shape characteristics discussed in Section 4.3 and increment size needs discussed in Section 4.1.3. Examples of subsurface

sampling techniques include direct push techniques to extract cores in liners, thin-walled hollow metal tubes, split-spoon samplers, or sonic drilling core bags. The primary difference between sampling accessible surface soil and subsurface soil is the added resources needed to obtain the increments. This often requires numerous (30 to 90) borings for a single DU to obtain the increments.

Collection of the entire core interval depth as the increment is the recommended subsurface ISM procedure.

Ideally, to be representative, the entire core depth interval should be considered as an increment, and then collected, combined with additional increments for an ISM sample, and submitted to the laboratory. Collection of the complete core interval as an increment is the recommended subsurface ISM procedure. However, this method can result in large ISM samples (approximately 5 to 10 kg), making logistics, such as field storage and shipping, problematic. Additionally, the selected laboratory must have facilities available to store, dry (if required), and process these large amounts of soil mass. Consequently, depending on the core diameter and interval depth, inclusion of the entire core increment across a targeted depth interval in an ISM sample may be impractical. In such cases, individual cores may be subsampled to reduce the final mass of the ISM sample. There are some subsurface core sampling options:

- collect an entire core interval (most preferred but results in large ISM sample mass)
- collect a core wedge subsample that reduces the mass of the total ISM sample (most practical)
- collect core slice subsample (least preferred)

When sampling the entire core, the equipment is advanced to the targeted DU depth interval, and the entire core is collected into a container. Decontamination of equipment between increment borings is not required unless two depth intervals are targeted from each core. Similarly, in the case of direct push, core liners can be reused between increments if the sampling design and soil type are such that the liners do not need to be cut to retrieve the increment.

Increments are placed in an appropriate large container capable of containing the final ISM sample volume, and this process is repeated at all remaining increment positions and borings with all the increments placed together in one container. The total increments now represent the ISM sample for that DU replicate or SU, and this sample volume should be provided to the laboratory and not further homogenized or subsampled (other than the ISM processing procedures described in Section 5). This process can then be repeated for a second and third (or more) replicate or SU.

Mass reduction techniques to decrease the final sample volume of the above procedure are described in Sections 4.5.1 through 4.5.2.

4.4.3 Sidewall and confirmatory ISM sampling

An incremental sample result is specifically designed to estimate the mean concentration in a volume of soil designated as a DU. The use of ISM samples to confirm excavation of a source area DU can be highly advantageous over the traditional approach of a small number of discrete samples. The excavation floor and sidewalls can be treated as individual DUs (Figure 3-31) with the investigation objective of assessing whether the estimated mean concentration of COPCs for these areas has been effectively reduced to below targeted screening levels or cleanup goals. This approach may require dividing a large sidewall DU into smaller SUs to meet additional DQOs (such as a certain grid spacing). Additionally, an entire excavation can be considered a DU – individual sidewalls can be sampled as SUs to provide spatial information and then collectively used to assess the effectiveness of the entire excavation (the DU) at meeting cleanup goals. Whatever the sidewall/confirmatory sampling design and objective, collecting ISM samples within these areas rather than single discrete samples ensures good DU spatial coverage and a more representative estimate of mean COPC concentrations at the excavation limit.

There may be regulatory limitations to this approach, however. For example, if regulations require cleanup of releases to a not-to-exceed regulatory level (the maximum concentration determined by discrete samples), then an ISM mean concentration may not be applicable and/or accepted by the regulating authority.

Sidewall/confirmatory sampling is implemented similarly to surface soil since the soil is generally exposed, but sidewalls of certain excavations (deeper than 3 to 4 ft) present their own unique challenges for collection due to limitation on safe access to certain increment locations.

For shallow excavations less than 3 to 4 ft below ground surface, the sampler can enter the excavation (after ensuring that

this procedure is compliant with any site-specific health and safety rules), and depending on the sample design, the DU/SU can be marked on the sidewall using a downscaled approach similar to that described in Section 4.2. Pin flags can be used to mark the grid or increment locations, and the DU replicates/SU ISM samples can be collected using any of the simple random, systemic random, or random sampling in a grid approaches. For deeper excavations beyond 3 to 4 ft that are not shored or sloped for egress (a basic test pit hot spot excavation), safe access may limit the application of systemic random and random grid sampling approaches, and simple random may be the only option.

Samples are collected similarly to surface soil ISM samples. The ISM sampler starts in one corner or end of the DU/SU and collects an increment at the predetermined increment positions using an appropriate sampling tool. Increments are placed in an appropriate container, and this process is repeated at all the remaining increment positions with all the increments placed together in one container. The total increments now represent the ISM sample for that DU replicate or SU, and this sample volume should be provided to the laboratory and not further homogenized or subsampled (other than the ISM processing procedures described in Section 5). This process can then be repeated for a second and third (or more) replicate or SU.

4.4.4 Stockpile ISM sampling

There are special considerations for selecting DUs during the SPP for sampling soil stockpiles:

- the source of the soil in the stockpile
- how the stockpile was created (over time, if applicable)
- how best to access the pile for sampling (for example, is it large or unstable)
- contaminants targeted for laboratory analyses

One of the best options is to coordinate sampling with the formation of any stockpiles on a site. When the stockpile is being formed, there is generally good access to sampling each portion of the pile over time, ensuring access to the entire stockpile DU is provided for good sample representativeness. If an existing stockpile is relatively small, good options may include moving the pile and collecting the increments while it is being moved (such as from the equipment buckets at appropriate intervals) or flattening or spreading out the stockpile so that it is safely accessible to sample with a hand core or other device. If the stockpile is very large or unstable, available sampling tools or methods that safely provide access should be considered with the goal of coming as close as possible to collecting a minimum of 30 increments throughout the stockpile (both vertical and horizontal locations). A resource for additional information on ISM approaches for soil stockpile sampling is the HDOH “Technical Guidance Manual”(HDOH 2008b). Refer to Section 3.6.4.2 of this document for additional information on ISM sampling design for stockpiles.

Once the approach to a specific stockpile is determined, the sampling process is comparable to surface soil ISM samples because the soil is accessible, the primary difference being that, in the case of sampling while a stockpile is being created, increments may be collected slowly over a

longer period of time (hours to days) while the soil is being consolidated. Therefore, it is important to document the number of increments after each collection to keep an accurate count.

Stockpile samples are often collected for waste characterization purposes, which are required to be collected at a certain rate per volume/tonnage. Therefore, it is important to ensure that the increments are collected at a rate so as to evenly distribute increments throughout the targeted volume/tonnage. RCRA land disposal characterization does not allow for ISM/composite sampling and requires only discrete sample collection.

During the sample collection, collected increments are placed in an appropriate container, and this process is repeated at all the remaining increment positions with all the increments placed together in one container. The total increments now represent the ISM sample for that DU replicate or SU, and this sample volume should be provided to the laboratory and not further homogenized or subsampled (other than the ISM processing procedures described in Section 5).

4.5 Decontamination

Sampling devices can be used within a DU without decontamination but should be decontaminated or disposed of between DUs (and also between replicates). If sampling tools will be used for two or more DUs, they should be cleaned of soil particles, decontaminated with the appropriate solutions or solvents, and dried between DUs. Typically, rinse (decontamination) blanks can be used to evaluate the potential effects of cross-contamination, if needed.

4.6 Sample Handling and Mass Reduction

It is recommended that all ISM sample processing be performed in a controlled laboratory setting to minimize sampling errors

ISM sample processing techniques, such as milling and representative subsampling, are designed to ensure the (typically small) mass of sample analyzed by the laboratory is representative of the DU or SU from which it was collected. These techniques reduce data variability as compared with conventional sample handling and processing approaches, but they also introduce some amount of sampling error. It is recommended that all ISM sample processing be performed in a controlled laboratory setting to minimize these sampling errors; discussion of detailed ISM sample processing is reserved for Section 5. However, there are certain projects that may warrant field sampling processing step(s).

When dealing with contaminants that have been deposited as solid particulates (such as energetics, metals at firing ranges, and so on), field subsampling is not recommended. Studies on energetics have shown that representative subsampling prior to grinding is problematic and likely not possible (Hewitt et al. 2009). However, depending on site logistics, type of soil, total number and/or mass of ISM samples, and so on, sample processing can be initiated in the field for some contaminants (SVOCs, pesticides, PCBs, and metals) with appropriate cautions as noted below. Limitations of the field processing of ISM samples include the following:

- Field processing is not recommended for contaminants deposited as solid particulates (energetics, metals at firing ranges, and so on) because subsampling solid particulates without properly grinding the sample mass may lead to samples that are not representative of site conditions.
- There is a lack of commercially available and correct subsampling tools (16-mm flat-bottom scoop with sides) for proper field processing procedures.
- Sample processing requires a controlled environment to air-dry, sieve, and subsample, if necessary, to minimize the potential loss or introduction of contaminants of concern (COCs) during sample processing.
- Additional subsampling replicates are needed to be collected and analyzed in a laboratory setting to evaluate analytical precision. This requires more knowledgeable and trained personnel.
- Moist samples may need to be air-dried to facilitate sieving in an appropriate dust-free location where temperatures and ultraviolet (UV) light are controlled. This requires an indoor location or job trailer.

Samples with little vegetation and composed mostly of sands and silts that naturally have a very low moisture content and soils that have been air-dried can be sieved in the field to remove pebbles and vegetative debris. Prior to air-drying, sieving, or both, the field-moist sample weight should be recorded. Additionally, field sieving is an option that allows the user to calculate the mass of a bulk ISM sample that will go to a laboratory to ensure that the adequate volume is submitted. It also allows the sampler to control the particle size of the sample to meet DQO requirements. Unless field subsampling will be performed, the entire sieved ISM sample fraction should be submitted to the laboratory for appropriate processing and subsampling.

If ISM processing and/or subsampling is performed in the field, subsampling replicates are recommended to evaluate precision.

Finally, if ISM sample processing and subsampling is performed in the field, it is recommended that at a minimum three replicate subsamples be collected and submitted to the laboratory for analysis. The subsampling process (as described above) is repeated on oneISM sample to form replicates, and the replicate results are then used to evaluate the precision of the field processing and subsampling. Note that the subsampling replicates should be collected in addition tothe ISM field replicates described in Section 5.3.5.

Mass reduction is another form of sampling handling that is important to ensure that manageable sample volumes are collected and submitted to a laboratory. Simply dividing an ISM sample (sieved or not) into separate volumes and placing each volume into separate sample containers or selecting a targeted sample volume from a larger ISM sample for analysis is not an acceptable method of mass reduction. Likewise, manually mixing samples to “homogenize” them in the field or laboratory may just serve to further segregate different particle sizes as the particles may settle in layers by weight or size

during mixing. It is imperative that an entire ISM sample volume (the total increment) be submitted for ISM processing/laboratory analysis. The ability to submit appropriate sample sizes is controlled through proper sample design and the sample mass reduction techniques further described in the following subsections.

4.6.1 Sample handling and mass reduction for non-volatile analysis

Ideally an entire soil core would be submitted to the laboratory for processing, however due to storage limitations of the laboratories this is not practical. To resolve this issue the following techniques can be implemented during sample collection in the field so representative samples can be provided to the laboratory for analysis.

4.6.1.1 Wedge sampling

Subsurface ISM increment collection techniques in recommended order are as follows:
Collect entire core interval
Core wedge subsample
Core Plug subsample
Core slice subsample

One option for collecting a representative subsample from a subsurface core increment for non-volatile contaminants is to collect a core wedge sample. The simplest approach is to split the core in half vertically along the axis, reducing the increment mass by half. Alternatively, a smaller wedge of soil (say, one-third of the core) can be taken from the entire length of the targeted depth interval. Removing a wedge of soil across the length of a larger core to encompass the entire depth interval rather than collecting the entire core depth interval as a whole reduces the mass of an individual increment of an ISM sample (see Figure 4-3) while still representing the entire depth. Individual wedges from 30 or more separate DU cores are then combined to form the complete subsurface ISM sample. ISM field replicates require completely separate incremental (that is, core) locations and collection of replicates using multiple wedges from the same core cannot be used as a measure of DU or overall sampling and analysis variability. ISM field replicates are discussed in Section 3.1.5.5. Core wedge replicates (i.e., wedge replicates from the same core) may also be collected when COPCs require separate laboratory processing procedures (see Section 3.1.5.3 and Section 5.2).

Figure 4-3. Examples of wedge sampling.
Source: ITRC ISM Update Team, 2020.

Replicates can be collected from the same core, combined with other wedge increments, and submitted as separate ISM samples to assess the precision of this subsampling strategy. This process reduces the total number of borings required to collect replicates (30 borings compared to 90 for replicates). However, core wedge replicates are not the same as ISM field replicates because ISM field replicates require completely separate incremental (that is, core) locations. Thus, core wedges should not be used as a measure of DU or overall sampling and analysis variability. Core wedge replicates evaluate only the variability in the subsampling process as opposed to collecting the entire core interval as the increment. ISM field replicates, on the other hand, provide information on spatial variability and the variance in the estimate of the mean without specifically separating out the contribution of field and/or laboratory sample processing/subsampling from other sources of variance. ISM field replicates are discussed in Section 3.1.5.5. Core wedge replicates may also be collected when COPCs require separate laboratory processing procedures (see Section 3.1.5.3 and Section 5.2).

4.6.1.2 Plug subsampling

Figure 4-4. Examples of plug sampling.
Source: ITRC ISM-1 Team, 2012.

One option that is less preferred than wedge sampling but more practical in certain soil types is a plug sampler (Figure 4-4). Using this method, a designated number of plugs is collected from the desired increment length, from 30 or more separate cores, and then combined to form the complete subsurface ISM sample. This method is considered inferior to wedge sampling because unlike wedge sampling, where the entire length of the core is represented in the increment, only the randomly selected plug locations represent the increment – other portions of the interval are excluded. However, as some soil types prohibit cutting the soil wedge, the plug approach is increasingly practical.

Replicates can be collected from the same core, combined with other plugs, and submitted as separate ISM samples to assess the precision of this subsampling strategy. Similarly to wedge sampling, this process reduces the number of borings (30 borings compared to 90 for separate replicates), but it does not assess the overall DU variability.

4.6.1.3 Core slice

The least preferred option for subsampling individual subsurface cores for non-volatile contaminants is to collect a core slice from the targeted DU layer (Figure 4-5). In this approach, a randomly selected perpendicular slice from within the larger targeted depth interval is collected as the ISM increment. For example, if the targeted depth interval was 2 ft in length (8 to 10 ft bgs), a 4-in perpendicular slice would be randomly selected from within the targeted depth interval of each individual core and collected as the ISM increment. Individual, randomly selected core slices from 30 or more separate cores would then be combined to form the complete subsurface ISM sample. This option introduces more bias than the whole core increment or core wedge approaches, but by reducing the increment mass, it addresses some of the logistical issues associated with handling the full core or the wedge increments. That said, this is the least recommended approach for subsurface ISM core sampling because it is the least likely to accurately represent the complete vertical length of the targeted DU layer.

Replicates can again be collected from the same core by selecting another slice of the same thickness, thereby reducing the total overall number of borings needed to collect replicates. It is important to prepare a sorting scheme for these increments

so as not to bias one replicate as shallower/deeper than another. Similar to core wedge replicates, core slice replicates are not the same as ISM field replicates because ISM field replicates require completely separate incremental (that is, core) locations. Thus, core slice replicates should not be used as a measure of DU or overall sampling and analysis variability.

Figure 4-5. Examples of core slice sample.
Source: Illinois EPA LUST FAQ and BIOTREE websites.

4.6.2 Sample handling and mass reduction for volatile analysis

The core wedge and slice approaches are not appropriate when VOCs are of concern since they can be quickly lost from an exposed surface (Hewitt, Jenkins, and Grant 1995). For VOCs, multiple plugs representative of the desired core depth are collected and immediately preserved in methanol (see Section 4.3.5).

When sampling for VOCs, the core may be subsampled by collecting numerous, small (say, 5-g) plugs at regularly spaced intervals along the targeted DU depth interval of the subsurface core. As with VOC sampling of any exposed soil, the plugs are immediately placed in a sampling bottle containing a predetermined volume of methanol. Nominal 5-g plugs of soil can be collected across the core using a VOC coring device (see Figure 5-10). The spacing interval of the VOC plugs along the core interval should be determined during the SPP but may require adjustment in the field based on core recovery in the case of subsurface ISM sampling. The syringe (or other coring device) used to collect the increments should be filled completely so that each increment has the same volume of soil. Additionally, the ISM sampler should be aware of potential volatile loss once the core is opened. As with any VOC soil sample, ISM VOC increments should be collected and preserved as quickly as possible to minimize potential loss. Potential loss of COPCs due to volatilization during collection of ISM increments is expected to be similar to discrete sample collection by USEPA SW-846 Method 5035A for the same sample density across a subsurface core (USEPA 2007).

Note that an unpublished study from Hawaii using a large bottle with methanol-preserved VOCs was stored in the sun and repeatedly opened over the course of the day to simulate increment additions, and VOC recovery was better than 80% for all analytes except dichlorodifluoromethane.

4.6.3 Sample storage

The primary concern for sample storage is acknowledging larger sample volumes, which will require more storage space and additional coolers.

Processed, labeled samples ready for packaging and shipment to the laboratory are typically stored in a cooler on ice, which for the larger ISM volumes means greater quantities of wet ice and more coolers. For bulk samples stored prior to mass reduction, larger ice receptacles may be constructed by using storage bins for regularly shaped containers (buckets) or form-boards and plastic sheeting for long or irregularly shaped bulk samples.

4.6.4 Sample shipping

The primary consideration for shipping ISM samples is the weight of large volumes (larger than traditional discrete samples) and the required storage and shipment temperatures. The increased weight should be considered when estimating shipping

costs (see Section 3.4).

When sampling for volatiles by ISM, the shipment of large volumes of solvent to and from the sampling activity can be problematic. When possible, methanol should be transported to the field via a surface transport to avoid or mitigate the volume limitations common in air transport. Guidelines for the transportation of a solvent such as methanol can be found in 49 CFR §172, “Hazardous Materials Table, Special Provisions, Hazardous Materials Communications, Emergency Response Information, Training Requirements, and Security Plans” (DOT 2011). Shipments via air transport may also be required to adhere to International Air Transport Association Dangerous Goods Regulations (IATA 2011).

4.6.5 Sample chain of custody

Typical laboratory chains of custody are not set up to indicate ISM processing and the specific steps requested for a specific project. It is thus important to communicate project-specific requirements to the laboratory and devise a means for requesting ISM processing (and the appropriate steps) in the chain of custody to ensure that the requirements are clearly communicated to the laboratory. Additionally, there are instances where multiple laboratories may be performing different analyses from the same ISM sample, but the ISM sample should not be split between two or more laboratories for separate ISM processing. Because the ISM sample can only be processed by one of multiple laboratories, one laboratory must be selected to perform the processing, and the means of transferring aliquots for analytical analysis to any other laboratory must be coordinated. It is important to ensure that this additional step in the chain of custody is documented. Sample holding times should also be considered in the logistics during field planning. Certain hold times can run short in considering the date of collection, the time to deliver samples to the lab, processing time (especially for wet samples that require extended dry time), transfer to a second laboratory, and laboratory extraction.

---

[1] Methanol preservation is discussed herein as the most common VOC preservative for soil, but there are other volatile preservatives that can be similarly utilized to preserve specific compounds. The methodology and considerations presented would remain the same.

Sample Processing and Analysis

Sample processing options encompass the areas of moisture management, particle size selection, particle size reduction, and sample digestion or extraction. The project team should choose the most appropriate processing options for a specific site characterization. This section includes implementation guidance for the various processing options recommendations for QC.

5.1 Introduction

Section 2 discusses some of the common sources of sampling uncertainty, but to obtain representative samples, sampling uncertainty must be limited or managed (Ramsey and Hewitt 2005). In the absence of uncertainty, a sample result by definition would be accurate. However, it is impossible to completely eliminate uncertainty and produce an accurate result unless all the soil in the DU is included in the analytical determination, which is obviously impractical. Thus, limiting sampling uncertainty is a critical function of any sampling design, implementation, processing, and analysis.

Incremental sampling has been successfully implemented at numerous sites for a variety of contaminants, and multiple processing options are available depending on the contaminants (such as energetics, metals, perchlorate, white phosphorus, SVOCs, VOCs, PCBs, and so on). Prior to implementation of sample processing, the entire sample collected in the field should be shipped to the laboratory. There may be some situations when it is desired to perform some of the sample processing steps in the field, such as air-drying and sieving, but the preferred approach is to conduct this activity in a controlled environmental such as an analytical laboratory.

ISM sample processing techniques, such as milling and representative subsampling, are designed to ensure the (typically small) mass of sample analyzed by the laboratory is representative of the field sample and hence the DU from which it was collected. These techniques reduce data variability as compared with conventional sample handling and processing approaches, but they also introduce some amount of sampling uncertainty. It is recommended that all ISM sample processing be performed in a controlled laboratory setting to minimize this sampling uncertainty. However, depending on site logistics, type of soil, total number and/or mass of ISM samples, and so on, sample processing can be initiated in the field for some analytes (such as SVOCs, pesticides, PCBs, and metals) with appropriate cautions as noted below. The techniques available include air-drying, disaggregation, sieving, milling, subsampling, and digestion or extraction.

5.2 Choosing Appropriate ISM Processing Options

The project team should choose the most appropriate sample processing options for a specific site characterization since there is no single combination of processing techniques applicable to all sites. Site characterization objectives should drive the selection of sample processing options, but there are four general areas to consider: moisture management, particle size selection, particle size reduction, and sample digestion/extraction. Objectives and site characteristics such as COPCs (analytes), surface or subsurface soil, contaminant release mechanisms, and data usage scenarios can also influence the selection of sample processing options.

There are also three important questions that must be answered to select the most appropriate sample processing options: Is air-drying acceptable? Is particle size selection needed? Is particle size reduction appropriate?

The specific analytes must be decided upon before finalizing sample processing decisions. There can be a wide range of physical and chemical characteristics within analyte groups, and these differences can influence the selection of sample processing options. Consider characteristics such as boiling point, volatility, air reactivity, sorption of analyte to particles, and the presence of high-concentration nuggets. Other considerations include particle size, distribution of analytes on or in the particles, particle weathering, contaminant release mechanism(s), and the end use of the data.

5.2.1 Air-drying

Moist samples may need to be air-dried until the soil aggregates are crushable to facilitate disaggregation and sieving, but drying to constant weight is not necessary for this purpose. The samples should be dried in a dust-free location where temperatures and UV light are not expected to cause degradation of COPCs. Air-drying is appropriate if the analytes are chemically stable when exposed to air and have sufficiently high boiling points or are strongly sorbed to the particles such that they are unlikely to volatilize during extended air exposure at the selected drying temperature. Drying at ambient

temperature (15 to 25°C) is most common and may take up to several days, thus impacting turnaround time and remaining holding time. Elevated temperature drying (30 to 105°C) accelerates the drying process but also requires greater analyte stability. The binding (distribution coefficient, soil organic carbon-water partitioning coefficient) between the contaminant and the soil particle should also be considered. Air-drying can be acceptable for strongly absorbed, low boiling point analytes, which would be expected for weathered surface soils.

Table 5-1 lists several example explosives and SVOCs, their boiling points, and their estimated loss potential during the room temperature air-drying step when these analytes are weakly sorbed to the soil matrix (Bruce 2003). Air-drying produces crushable soil particles, but it risks loss of low boiling point target analytes. Table 5-1 is not all-inclusive and is intended only as an example for evaluating COPCs and the possible effects of air-drying. Physical property data for additional COPCs are available in "Technical Guidance Manual Notes: Decision Unit and Multi-Increment Sample Investigations,"Tables 2a and 2b (HDOH 2011).{HDOH, 2003 #254;HDOH, 2011 #254} Applying air-drying to analytes that are not significantly weathered with moderate and large loss risks should be avoided unless there is sufficient site knowledge or experimental data to demonstrate the loss risk is acceptable.

Table 5-1. Potential for loss during the air-drying step for weakly sorbed analytes.

Source: ITRC ISM-1 Team, 2012.

Contaminant	Vapor Pressure (mm Hg)	Boiling Point (°C)	Loss Potential
Acenaphthene	2.15E-03	279	Moderate
Acenaphthylene	6.68E-03	280	Moderate
2-Amino-4,6-dinitrotoluene	3.33E-06	352	Small
4-Amino-2,6-dinitrotoluene	3.65E-06	352	Small
bis(2-Chloroethoxy)methane	1.32E-01	218	Small
bis(2-Chloroethyl)ether	1.55E+00	179	Moderate
bis(2-Chloro-1-methylethyl)ether	5.60E-01	187	Moderate
4-Chloro-3-methylphenol	5.00E-02	235	Moderate
2-Chloronaphthalene	1.22E-02	256	Moderate
2-Chlorophenol	2.53E+00	175	Moderate
Dibenzofuran	2.48E-03	287	Small
1,2-Dichlorobenzene	1.47E+00	180	Large
1,3-Dichlorobenzene	2.15E+00	173	Large
1,4-Dichlorobenzene	1.74E+00	174	Large

2,4-Dichlorophenol	9.00E-02	210	Small
Dimethylphthalate	3.08E-03	284	Small
1,2-Dinitrobenzene	4.55E-05	318	Small
1,3-Dinitrobenzene	9.00E-04	291	Small
2,4-Dinitrotoluene	1.47E-04	300	Small
2,6-Dinitrotoluene	5.67E-04	300	Small
Hexachlorobutadiene	2.20E-01	215	Large
Hexachloroethane	2.10E-01	154	Large
Octahydro-1,3,5,7-tetranitro-1,3,5,7- tetrazocine (HMX)	3.30E-14	436	Small
Isophorone	4.38E-01	215	Large
2-Methylnaphthalene	5.50E-02	241	Moderate
4-Methylphenol	1.10E-01	202	Moderate
Naphthalene	8.50E-02	218	Large
Nitrobenzene	2.45E-01	211	Large
Nitroglycerin	4.00E-04	250	Small
N-Nitrosodimethylamine	2.7E+00	154	Moderate
N-Nitroso-di-n-propylamine	3.89E-01	206	Small
2-Nitrotoluene	1.88E-01	222	Moderate
3-Nitrotoluene	2.05E-01	232	Moderate
4-Nitrotoluene	1.57E-02	238	Moderate
Pentaerythritol tetranitrate (PETN)	5.45E-09	364	Small
Phenol	3.50E-01	182	Small

Hexahydro-1,3,5-trinitro-1,3,5- triazine (RDX)	4.10E-09	353	Small
Methyl-2,4,6-trinitrophenylnitramine (Tetryl)	1.17E-07	432	Moderate
1,2,4-Trichlorobenzene	4.60E-01	214	Large
2,4,6-Trichlorophenol	8.00E-03	246	Small
1,3,5-Trinitrobenzene	6.44E-06	315	Moderate
2,4,6-Trinitrotoluene	8.02E-06	365	Small

Considering analyte loss risks during the air-drying step, it may be necessary on occasion to skip air-drying and proceed with other processing steps on the as-received sample, typically when lower boiling point SVOCs are primary contaminants. HDOH has also expressed concerns about the potential loss of elemental mercury. Wet, sticky samples cause mechanical problems when processing, but 2D slabcake subsampling on the as-received sample is possible when air-drying must be avoided. Note that decreased reproducibility occurs as other processing steps are skipped.

In most cases, air-drying is performed to facilitate disaggregation, sieving, milling, and extraction. When low boiling analytes are weakly retained on the soil particles, they can be lost during the air-drying step and produce a low bias in the final results. This is generally not a concern for surface soil samples unless the COC release is very recent. At most sites, the surface soil has been exposed to air for months, years, or decades since the release. Thus, air exposure at the laboratory for a few days during air-drying is not likely to produce analyte loss for surface soils. Figure 5-1 shows that low boiling PAHs are not lost during air-drying from an aged surface soil. Three replicate subsamples were collected immediately when the soil was spread out for air-drying and after processing (drying, disaggregation, and sieving). The average results demonstrate no analyte loss from surface soil samples during air-drying and that precision is generally improved.

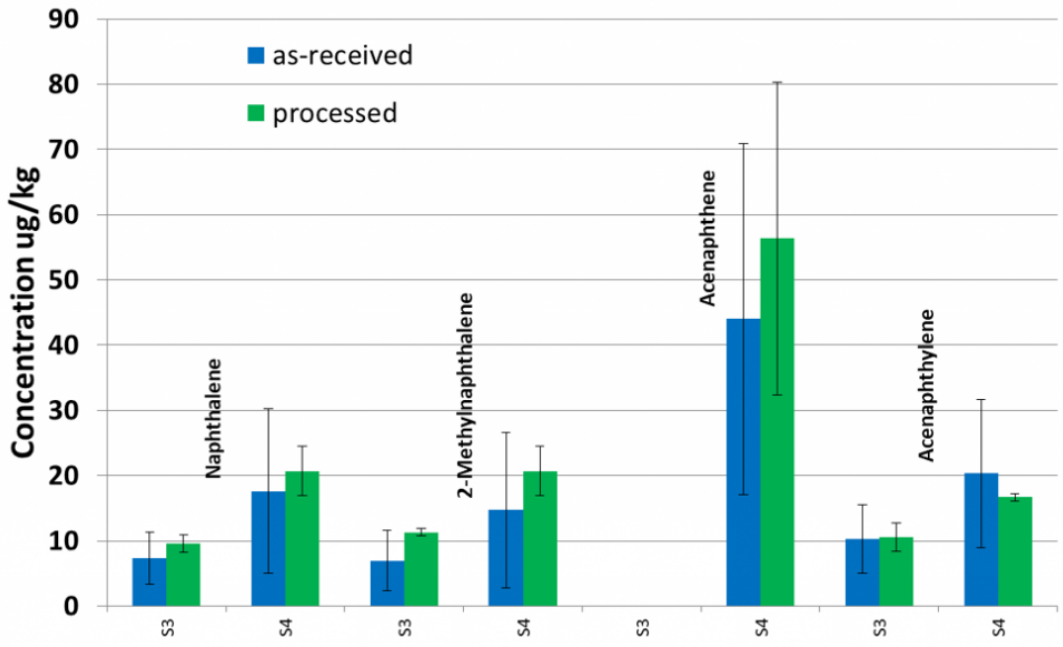


Figure 5-1. As-received subsampling versus processed (dry, disaggregate, sieve, subsample). *Source: Mark Bruce, Eurofins, 2019. Used with permission.*

Conversely, recent releases and subsurface soil samples might produce circumstances where low boiling analytes are

weakly sorbed on the particles and are at risk for loss during air-drying. Subsampling of the as-received sample using the 2D slabcake process is the most common accommodation, but wet sieving and wet milling are viable options if the sample is sufficiently flowable (high water content sediments). As a general recommendation, drying should be utilized only to the extent necessary to avoid potential analyte loss. Moisture content below 5 to 10% is usually acceptable to produce crushable soil aggregates.

5.2.2 Disaggregation prior to other processing

Sample disaggregation is a technique used on dry, crushable soil to break up the aggregates formed during air-drying, but it does not mill small pebbles and other hard particles into smaller particulates, like the particle size reduction techniques (such as milling) listed below. In some risk assessment scenarios, disaggregation is preferable to milling because some metallic COPCs remain “locked” inside the hard particles and are not included in subsequent analyses.

5.2.3 Particle size selection: sieving

Particle size selection and particle size reduction decisions are determined by the DQOs. If the characterization uses an operational soil definition of <2-mm particles, then disaggregation followed by particle size selection with a #10 sieve is a common choice. For the skin-to-mouth and hand-to-mouth exposure pathway, USEPA recommends that soil Pb be measured on the <0.15-mm soil fraction to estimate EPC since this is the particle size most likely to stick to a child’s hands and enter homes as dust (USEPA 2016). Only the particle fraction passing through a 100-mesh sieve is to be analyzed. Other data usage scenarios will push toward other sieve sizes or determine that no sieving is appropriate. Example guidance on sieve selection is available in multiple references (DOD/DOE 2018) (USEPA 2013) (Frederick, Frame, and Vallero 2017, USEPA 2011) (USEPA 2012b) (HDOH 2016b, 2015).

Occasionally, the sample container includes objects not considered part of the sample, so the DQOs should direct whether vegetation and oversized material are included or excluded from the sample. If the Toxicity Characteristic Leaching Procedure(TCLP)is going to be performed on the sample, *all* material must be retained. Vegetation and oversized material can be manually removed with tweezers or spatulas but can be removed more reproducibly with a sieve if the sample is dried. The excluded materials can be documented via photographs and weight removed when appropriate. Note that in the case of energetic materials and other situations where the COPC is deposited on the ground surface as a particulate, the vegetation should not be removed in the field prior to laboratory sieving.

Although sieving to the <2-mm particle size is typical, there may be contaminant investigations or analyses where alternative particle sizes may be of interest. In these cases, the rationale for sieving to other specific particle sizes and associated changes to laboratory processing/analysis should be clearly discussed during the project planning step and documented in the QAPPs.

5.2.4 Particle size reduction: milling

If a COPC is present as a solid particulate, particle size reduction through sample milling can facilitate more representative subsampling by reducing the range of particle sizes and the maximum size present. The selection of the particle size reduction technique and maximum target particle size should be determined during project planning and is part of DQO development. It should be noted that the maximum particle size has a *significant* effect on FE. See the discussion in Section 2.6.2.1 on Gy’s TOS about the relationships among particle size, uncertainty, subsample size, and FE. Selection of the appropriate milling process and equipment to achieve the maximum particle size is determined during project planning (see Section 3.1), with many common options described in USEPA guidance (Gerlach and Nocerino 2003). Examples of when particle size reduction may be appropriate after particle size selection include, but are not limited to, metals at small-arms ranges, clay target fragments at skeet ranges, lead-based paint chips, munitions constituents, and so on.

Milling is recommended for ISM metals analyses, especially for those situations where the particulate form is present (Clausen et al. 2018a, Clausen, Georgian, and Bednar 2013, Clausen et al. 2013) (Clausen, Georgian, and Bednar 2013, Clausen et al. 2016, Clausen et al. 2018c, b). Subsampling that uses a larger subsample mass in lieu of milling should reduce FE, but in some situations, a larger subsample mass increases the measured variability (relative to small subsamples). Smaller subsamples (1 g) can be biased against including larger particles, thus producing artificially low relative SDs (RSDs).

Milling is not universally recommended for organiccontaminants other than energetics (SW-846 Method 8330B) (USEPA 2006c) although it has been demonstrated to improve precision for PAHs on skeet ranges and PCBs. Volatilization loss of organic COPCs may occur due to increased temperatures during milling, and excessive milling can lead to destruction of

organic contaminants, as demonstrated in the mechanochemical dehalogenation (or mechanochemical destruction) soil remediation process. See “Reference Guide to Non-Combustion Technologies for Remediation of Persistent Organic Pollutants in Stockpiles and Soil” (USEPA 2005a) for additional information.

The usefulness of particle size reduction by milling for organic COPCs is usually small because the larger mass (10 to 30 g or more) normally extracted and analyzed and the particulate nugget effect are often minimal. However, nuggets can and do occur for specific organic COPCs – for example, soil analyzed for PAHs at skeet ranges can exhibit a nugget effect due to the deposition of clay pigeon fragments. In such cases, the advantages and limitations of milling for organic COPCs should be evaluated during project-specific systematic planning.

Particle size reduction has been called milling and grinding, but grinding is also used to refer to disaggregation. When using the term *grinding*, specify the equipment to be used to help ensure accurate communication.

Particle size reduction is generally performed to improve the reproducibility of subsampling (Figure 5-2). More aggressive or longer milling typically produces more improvement to precision.

Milling can also be used to intentionally expose the interior of particles to the subsequent digestion or extraction steps. In some instances, this will increase recovery of COCs previously sequestered in the interior of the particles. Figure 5-3 shows this effect for lead particles reduced in size by puck mill and ball mill. In some instances, the increases of the non-anthropogenic metals are negligible and offset by the improved precision and accuracy of the measured metal content (Clausen et al. 2016). The potential improvement in precision and increase in measured metals concentrations should be considered during project-specific systematic planning when determining if milling is appropriate.

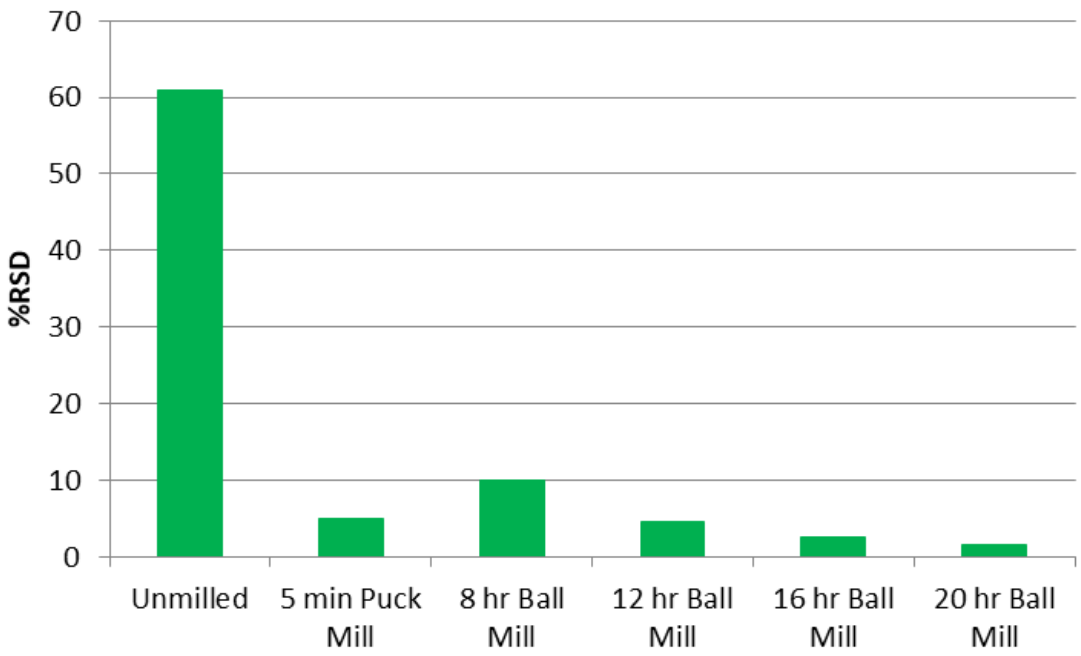


Figure 5-2. Improved precision with milling.
Source: Mark Bruce, Eurofins, 2019. Used with permission.

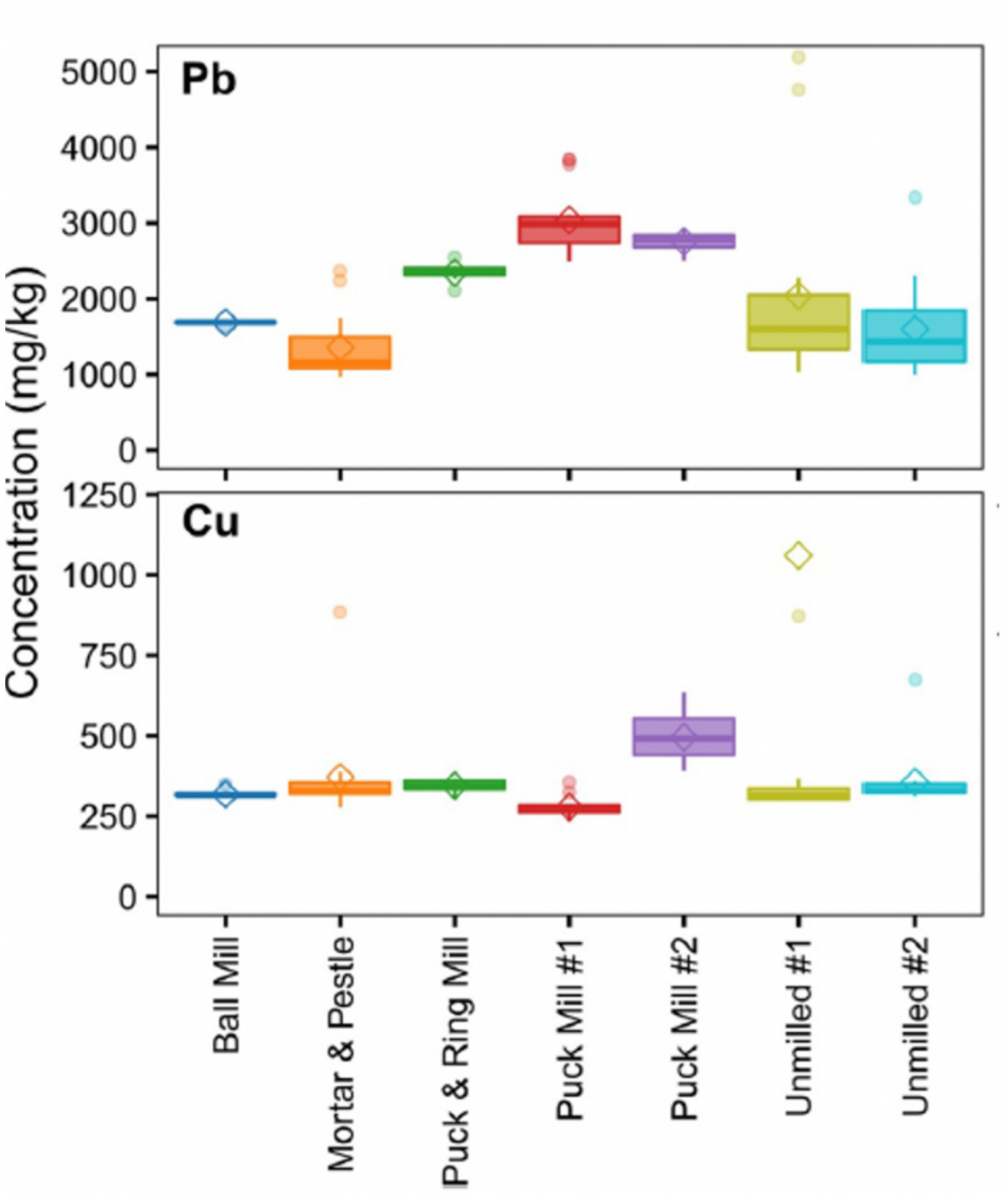


Figure 5-3. Milling effects on measured concentration.
Source: Jay L. Clausen, et al., Microchemical Journal 154 (2020). Used with permission.

Milling can also transfer material from the mill to the sample, which means that mill surfaces with COPCs may interfere with the analysis and should be avoided. If metals are COPCs, then the composition of any metal-containing mill surfaces should be compared with the target analyte list. Ceramic, agate, tungsten carbide, or low chromium steel milling components would be more appropriate. Figure 5-4 shows chromium concentration increasing after the use of a steel puck mill and less contamination from a ball mill with steel can and ceramic grinding media. The use of a milling blank can be useful when a soil-like material with acceptably low background is available. See Section 5.4 for more details.

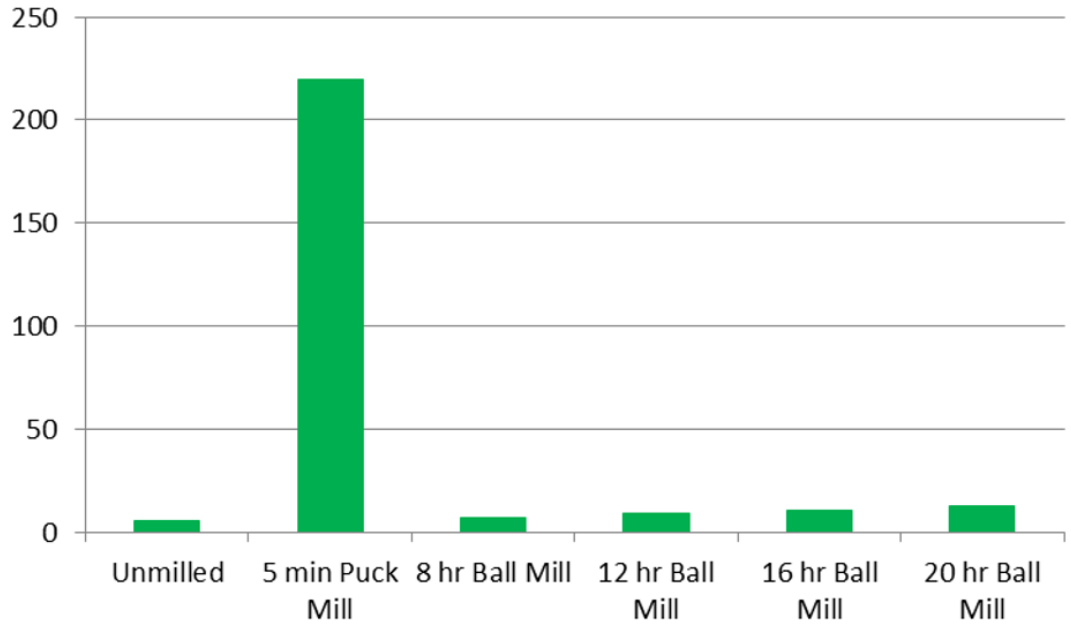


Figure 5-4. Chromium contamination from a steel puck mill.
Source: Mark Bruce, Eurofins, 2019. Used with permission.

Puck mills and ball mills might not be effective at efficiently reducing particle size when the starting size is >2 mm. The use of a jaw crusher can produce particles <2 mm, and for some DQOs, this is sufficient. In other instances, these smaller particles are further particle size reduced in a mill.

5.2.5 Subsampling (2D slabcake on dry or wet samples, sectorial splitter)

Subsampling to produce the analytical samples for extraction, digestion, or leaching is most frequently produced by using the 2D slabcake process. This miniaturized version of what is done in the field is usually a good balance of representativeness and cost. The 1D slabcake can also be used by itself to collect large subsamples or in combination with 2D slabcake to produce small subsamples. The rotary sectorial splitter typically produces a more representative subsample but is much more expensive and labor-intensive.

Of particular concern are methods that require small masses, such as the 1 g typically used for metals digestion. Increasing the initial mass to a minimum of 10 g at a <2-mm sample particle size improves reproducibility (Clausen et al. 2018a). There are generally only two options to reduce FE: increase the sample size to be analyzed or reduce the particle size. For typical soil and analyte concentrations of 1 ppm, to reduce FE to ≤15%, either the sample mass must be increased to 32 g (2-mm particle size), or the particle size must be reduced to less than 325 mesh (0.044 mm) for a 1-g sample.

An important element to consider when using a subsampling process is that the final subsample mass must be used completely in the analytical sample preparation step to maintain subsample representativeness. For this reason, the final target mass for each of the following approaches and the mass needed for analytical sample preparation must be considered when choosing the process.

5.2.6 TCLP

TCLP (USEPA Method 1311) is a method-defined parameter. Some ISM sample processing options have limited application to samples intended for TCLP analysis – for example, sample drying is not mentioned in Method 1311, so additional laboratory-based drying should not be applied to samples intended for TCLP since it might affect analyte stability or leachability. In addition, limited particle size reduction might be needed to meet the <9.5-mm particle size criterion. Two-dimensional slabcake subsampling on the sample can be used to produce 5- and 100-g subsamples for the pretest and leaching steps.

Sometimes, optimal processing options differ for different analytes. In these instances, there are three options: (1) collect separate ISM field samples for each option, (2) split the field sample with different portions going to the different processing options, or (3) collect analytical subsamples at different points in the sequence of processing options.

Collecting separate field samples is the statistically preferred option but also the most expensive. Splitting the field sample prior to processing is inexpensive but can introduce significant reproducibility and uncertainty issues because the splitting process occurs before any ISM processing.

The more cost-effective and reasonably reliable option is to collect analytical subsamples at appropriate points in the processing sequence. The most common intermediate subsampling points and the reasons for choosing them are listed in Table 5-2.

Table 5-2. Subsampling points in the ISM processing sequence.

Source: ITRC ISM Update Team, 2020.

Subsampling Point	Reason to Choose	Limitation
As-received sample when initially spread out for 2D slabcake prior to air-drying	Low boiling point, weakly sorbed analytes might be lost during air-drying	Small-scale heterogeneity is likely to be high
After disaggregation	No particles are to be excluded from the subsamples	Small-scale heterogeneity is likely to be moderate
After sieving	Need to avoid milling because of potential analyte loss	Small-scale heterogeneity is likely to be small to moderate

Often times, the collection of an ISM sample yields a kilogram or more of sample material, but splitting the bulk sample in the field or laboratory is generally not recommended to reduce the sample mass because the entire sample should be processed. The uncertainty introduced by splitting prior to the completion of sample processing can be large when the COPC nugget effect is large, such as in highly heterogeneous samples. Note that bulk sample splitting (or subsampling) without particle size reduction merely increases FE.

Paired ISM sample collection is generally recommended over bulk ISM sample splitting when different sample processing treatments will be needed. Paired ISM samples allow the application of such treatments without the uncertainty introduced by bulk splitting.

5.3 ISM Sample Preparation and Analysis

Sample preparation options include moisture management, disaggregation, sieving, milling and subsampling. Analysis includes sample digestion or extraction and instrumental analysis methods suitable for the matrix and COPCs.

5.3.1 Moisture management

To facilitate air-drying, place the soil sample on a tray made of or lined with a material compatible with the COPC (see air-drying tower in Figure 5-5). The selection of the tray or liner material should ensure that the analytes of interest are neither lost nor gained from the sample to the tray or liner by sorption or reaction, and potential interferences are also avoided. Aluminum trays and liners should be avoided if aluminum is a COPC or if it may interfere or interact with an analyte of interest (such as chromium or elemental mercury). Plastic trays and liners should be avoided if phthalates and plastic components are COPCs, and a paper liner should be avoided if organic carbon or organics that can sorb to paper (such as petroleum) are COPCs. Spread the sample evenly in the drying tray, and if needed, use 2D slabcake subsampling to collect a subsample for moisture determination of the original sample. Place the sample in a ventilated area such as a hood or oven with sufficient airflow to carry away evaporated moisture. Drying time varies from a few hours to several days depending on moisture content, soil characteristics, airflow, and temperature. Intermittent (for example, daily) turning of the soil may be necessary to facilitate air-drying in an acceptable time frame. The soil should be dry enough to allow the agglomerates to be crushed producing a flowable matrix; moisture content below 5 to 10% is usually acceptable. Wet clay samples should be crushed with a pestle partway through the drying process to avoid the formation of large bricks that are difficult to handle with subsequent processing. Drying to a constant weight is not necessary; the sample only needs to be dry enough to facilitate proper mechanical function of subsequent processing equipment. The ventilated air-drying area uses a large

amount of laboratory space during the drying step, so the use of racks to hold the drying trays can facilitate an efficient use of space. Air-drying towers with filters on the front and fans on the back have been effective at consistently and quickly air-drying samples. Drying via ovens at elevated temperatures or high-speed forced air is generally not recommended.

Figure 5-5. Air-drying tower.
Source: Mark Bruce, Eurofins, 2019. Used with permission.

Disaggregation. To disaggregate, take the dry sample and gently rub it into a sieve to promote break-up of the soil agglomerates. A variety of sieve sizes can be used depending on the project DQOs, but a #10 sieve (2 mm) is the most common size. Alternatively, the soil can be disaggregated using a bladed coffee-type grinder or blender, just keep the time as short as possible to minimize wear on the blade, contamination of the sample with the blade materials, and any sample temperature elevation. A mortar and pestle can also be used to gently break up the soil agglomerates, though there is a greater risk of causing particle size reduction of the hard particles than with softer disaggregation techniques such as pestle/sieve and blender. Disaggregation is generally sufficient when SVOC COPCs are the primary concern and subsample sizes are 10 g or larger. Disaggregation and sieving are also commonly used prior to complete particle size reduction using the milling techniques listed in Section 5.3.3.

Other disaggregation tools include spinning blade coffee mills and food processors. A rotary hammer disaggregator breaks up soil clods by hitting them with a cylindrical rod while the soil falls through the path of the hammers (Figure 5-6).

5.3.2 Sieving

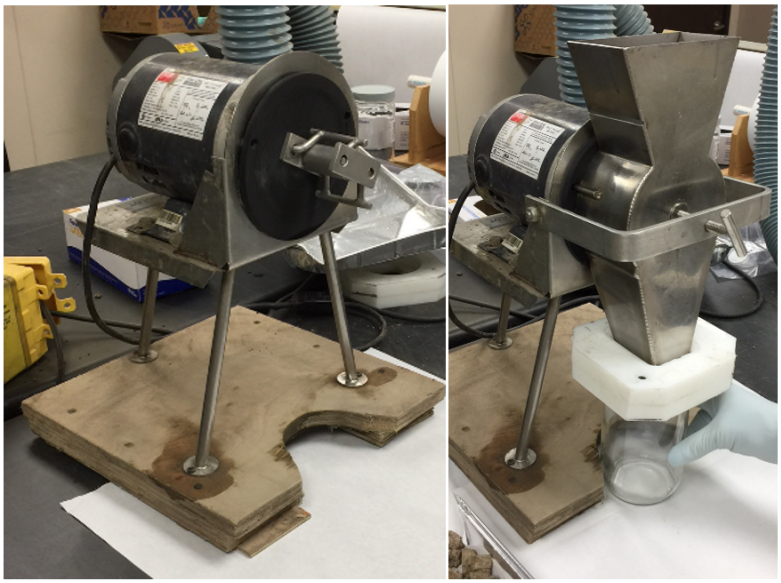


Figure 5-6. Rotary hammer, disaggregator.
Source: Mark Bruce, Eurofins, 2019. Used with permission.

Particle size selection can occur at several different points in the ISM process. However, it generally follows air-drying and disaggregation with the 2-mm (#10) sieve being the most common (Figure 5-7).

Samples with little vegetation and composed mostly of sand and silt (which naturally have a very low moisture content) and air-dried soils can be sieved (typically using a #10 sieve, <2-mm particle size) in the field to remove pebbles and vegetative debris. Prior to air-drying or sieving or both, the field-moist sample weight should be recorded if specified in the SAP. The <2-mm particles are generally considered soil, while larger particles are considered gravel, rocks, or other materials (sticks and roots). Additionally, field sieving is an option that allows the user to calculate the mass of a bulk ISM sample needed to meet DQO requirements based on soil particle size. Unless field subsampling will be performed, the entire sieved ISM sample fraction should be submitted to the laboratory for appropriate processing and subsampling.

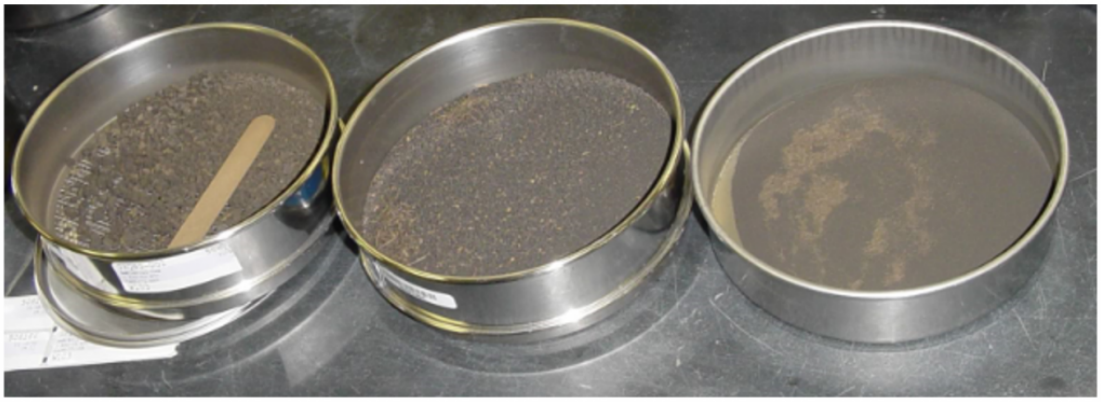


Figure 5-7. Sieves.
Source: Mark Bruce, Eurofins, 2019. Used with permission.

5.3.3 Milling

Extended high-speed milling can elevate sample temperature due to friction, so COPC thermal stability and volatility should be considered when choosing equipment and a milling scheme. USEPA SW-846 Method 8330B for nitrocellulose-based propellant residues specifies a 2-min (or longer) cool-down period between five 60-sec grinding intervals with a puck mill to maintain acceptable temperatures and minimize the loss of volatile energetic COPCs. The hard nature of explosive and metal particulates necessitates a 1-min milling interval.

When milling has been selected as part of the ISM DQOs, the entire conditioned ISM sample is milled. Splitting an un-milled ISM sample with high heterogeneity due to the nugget effect can lead to non-representative subsampling (Clausen et al. 2018a). If the milling equipment is not large enough to process the entire sample, mill smaller portions of the sample and then re-combine and mix them after the milling step. The milling equipment listed below is not an exclusive list of equipment capable of meeting ISM DQOs, but it is an example of equipment that has been used successfully in the past.

Mortar and pestle grinding can be accomplished with either manual or automated systems, with large automated systems recommended because of their increased capacity, better reproducibility, and reduced likelihood of repetitive-stress injuries. The sample contact materials can be steel, ceramic, or others depending on the COPCs. The sample is loaded into a heavy walled bowl, then crushed between the bowl wall and the pestle by manually pushing the pestle or spinning the bowl with a fixed pestle in an automated system.

Rotary pulverizers can reduce particle size from approximately 6 mm to <100 μm, the distance between the grinding plates determines the final particle size. The dry sample is fed into the chute, and the ground sample is collected from a hopper beneath the grinding plates. Note that adequate cleaning of rotary pulverizers between samples to remove any potential cross-contaminants is difficult.

Ball mills consist of both high- and low-speed systems. Typically, the sample is placed in a container along with a grinding medium and shaken rapidly or tumbled slowly. The milling medium (typically steel or agate balls, or ceramic cylinders) crushes the sample particles. High-speed systems consist of high-strength containers and high-speed shakers, which means they can provide more reproducible reduction to <100-μm particle sizes. Typical milling time for high-speed systems is a few minutes.

Low-speed systems typically consist of single-use cans, a milling media, and a low-speed tumbler or roller. Roller mills or paint can shakers are common examples. Typical milling times are several hours, but excessive milling should be avoided due to possible analyte loss. A study of anthropogenic metals in particulate form demonstrated a minimum milling interval of 18 hrs (Clausen et al. 2012).

Dish and puck (shatter box) milling are described in USEPA SW-846 Method 8330B (USEPA 2006c). The sample is loaded into the dish with the puck inserted. If the dish is not large enough to process the entire sample at once, mill smaller portions of the sample and then re-combine and mix after the milling step. The milling cycle time and cooling period (if necessary) depend on the analytes of interest. An example cycle consists of 1 min of milling and at least 2 min without milling to allow the dish and sample to cool. This process may be repeated two to four more times, depending on the materials to be milled. The cooling part of the cycle reduces internal temperatures and hence thermal degradation of the analytes. USEPA SW-846 Method 8330B (energetics) recommends a final particle size of <75 μm, but the optimal milling conditions and final particle size for other COPCs might be different from that guideline. Performance for other COPCs should be demonstrated with reference materials or other known samples.

Sieving can also be used to determine whether the milling step is complete. Particles below the DQO-specified size are removed from the milling process, and those above the predetermined cutoff size are returned to the milling equipment for additional particle size reduction. Common maximum particle size cutoffs are 250 μm (#60), 150 μm (#100), and 75 μm (#200). Alternatively, final particle size suitability can be estimated by touch or visual inspection when less accuracy is acceptable.

5.3.4 Sample mixing

Dry mixing can reduce heterogeneity and facilitate representative subsampling if the sample consists of particles of similar size and density. However, mixing samples with large differences in particle size or density can increase stratification and hinder representative subsampling. Unless, particle size reduction is performed, dry mixing is generally not recommended.

5.3.5 Subsampling (2D slabcake on dry or wet samples, rotary sectorial splitter)

The preferred subsampling methodology is the 2D Japanese slabcake or incremental sampling approach that emulates the field incremental subsampling process in a controlled laboratory setting. The entire sample is spread evenly onto a 2D surface at a depth easily penetrated by a square scoop. A scoop is then taken by removing an increment that equally represents the entire vertical column of the slabcake, and the material is placed in a receiving container. This process is repeated at least 30 times at systematic random locations around the entire sample

The laboratory default should be to use at least 30 increments to build the analytical subsample. If project-specific planning

has determined that other increment numbers are needed to meet the DQOs, then use them, but replicate subsamples are recommended to determine whether the subsampling meets the DQOs.

A process should be established to document that the increments are collected from random or systematic random locations over the entire exposed surface to ensure adequate representation of the sample. To do so, increments for replicate samples should be collected from independent locations, or alternatively, the entire sample may be stirred, re-spread, and replicate increments collected in the same manner as the primary sample, with the process repeating for as many replicate samples as applicable.

A good example setup is a 20- × 30-in aluminum baker's tray lined appropriately. The tray can easily take a 2-kg sample spread across it at a depth of no more than 1 to 2.5 cm. A scoopula pushes the sample around and spreads it to an even depth, ideally, as thinly as practical. As the sample is spread, the fine particles tend to migrate downward toward the tray while the larger, less-dense ones rest on top. A scoop is then used to minimize the discrimination of taking more of the large particles on the top. A square-walled, blunt-end scoop with a minimum 16-mm width tends to perform the best because it facilitates equal collection from both the top and bottom of the slab (Figure 5-9); the sides reduce the tendency of particles to fall off the scoop during increment collection. Before taking increments, the target mass should be considered because each scoop (increment) will ideally represent <1/30th of the desired target mass – for example, for a 30-g subsample, each increment should weigh about 1 g. Before starting the scooping process, a few trial scoops should be taken and weighed to calibrate the amount needed for each scoop. This technique works best when used after disaggregation or milling in conjunction with particle size selection via sieving to reduce the range of particle sizes (Figure 5-8).

Figure 5-8. Two-dimensional slabcake with square scoop.
Source: Mark Bruce, Eurofins, 2019. Used with permission.

The 2D slabcake subsampling process may be applied to moist "sticky" samples as well. The best results are achieved with sieved soils, but this process can also be applied to as-received samples. Spread the moist soil into an even-depth slabcake as described above, then use a square-walled, blunt-end scoop with a minimum 16-mm width (Figure 5-9) to collect 30 or more increments of 2-mm particle size for the final analytical subsample. Coring tools may also be used for subsampling if a moist sample is sufficiently cohesive. See the tool width discussion in Section 4.4.

A 1D slabcake is produced by pouring the sample into a line and using at least 20 passes back and forth to distribute the sample particles over that line. A square scoop is cut across the line to remove a subsample aliquot, and the sampler can combine as many of these aliquots as needed to produce the target subsample size (Gerlach and Nocerino 2003).

Sectorial sample splitting is the statistically preferred process because it results in the least sample heterogeneity of the methods discussed, but it does require investment in a rotating sample splitter and dust-abatement measures. The device consists of a rotating head with several chutes sitting on top of a motor. The chutes are spaced equally apart from each other and are of the same dimensions. A hopper is mounted above the rotating head with a vibrating tray that delivers the soil sample to the splitter at a variable rate, depending on the intensity of the vibrations. The rotation speed should be adjustable. The sample falls from the hopper into the chutes as they spin, and collection devices such as sample bottles are

mounted on the bottom of each chute to receive equal portions of sample material. In general, slower feed rates from the hopper and faster rotational speed make for better subsamples.

The entire sample must be poured into the hopper initially, with the resulting subsamples equal in mass to the initial sample mass divided by the number of subsamples. If the desired target mass is not achieved on the first split, re-combinations of individual splits may be required to achieve a larger final target mass or re-splits of one of the previous spilt samples (serial splitting) if a smaller mass is needed. Small amounts of fine particles may adhere to the device and should be pushed through by tapping or by a small burst of compressed air. Limitations to this technique include equipment cost and availability, trained staff availability for correct operation, equipment cleaning issues, and equipment maintenance.

Riffle splitting generally divides the sample into two equal portions by directing the sample portions into opposite pans with alternating chutes. It can be used sequentially to further subdivide a sample into smaller aliquots (Gerlach and Nocerino 2003). Riffle splitting has been demonstrated to introduce a high degree of error in samples, however, and is not recommended for the collection of subsamples for analytical testing (Pitard 2019).

Figure 5-9. Examples of rectangular and flat-bottom sampling tools.
Source: Mark Bruce, Eurofins, 2019. Used with permission.

If field subsamplingis to be performed, the entire ISM sample should be air-dried (only if necessary) and sieved to the predetermined particle size (typically using a #10 sieve, <2-mm particle size). The remainder of the processing steps are the same as those described in the 2D slabcake section earlier.

The mass of sample required for the analytical test or tests helps determine the mass of each of the 30 or more increments – for example, if a mass of 30 g is required for analytical extraction and analysis, 30 separate 1-g increments would be collected from systematic random locations. Depending on project DQOs, replicates of the field processed soil should be collected and submitted for analysis to evaluate the precision of the ISM field processing procedure. The entire submitted subsample mass must be prepared for analysis due to possible particle size discrimination during sample transit (such as fines settling to the bottom of the sample container). If the entire contents of the submitted container are not to be analyzed, the laboratory must use proper techniques to ensure a representative particle size subsample is used for analysis. Laboratory replicates should be analyzed to evaluate the precision of the laboratory's subsampling procedure.

Simply dividing an ISM sample (sieved or not) into separate volumes and placing each volume into separate sample containers for analysis is not an acceptable method of mass reduction. Likewise, manually mixing samples to homogenize them in the field or laboratory may just serve to further segregate different particle sizes because particles may settle in layers by weight or size during mixing. The process of spreading the entire sample out in a thin layer and collecting many representative increments in a systematic random fashion with a tool that can scoop to the bottom of the sample is the best way to collect a representative subsample of all the different sizes and types of soil particles present.

Finally, if ISM sample processing and subsampling is performed in the field, it is recommended that at a minimum three replicate subsamples be collected and submitted to the laboratory for each different analysis. The subsampling (as described above) process should be repeated on oneISM sample to form replicates, and these replicate results can then be used to evaluate the precision of the field processing and subsampling. Note that, the subsampling replicates should be collected in addition toISM field replicates.

Limitations to the field processing of ISM samples include the following:

- It is not recommended for COPCs deposited as solid particulates (energetics, metals at firing ranges, and so on).
- It requires a controlled environment to air-dry, sieve, and subsample, if necessary, to minimize the potential loss or introduction of COPCs during processing.
- Additional subsampling replicates are needed to evaluate precision.
- More knowledgeable/trained field personnel required.

5.3.6 Sample processing of VOC ISM samples

ISM samples can also be collected for VOCs for contaminant analyses. ISM VOC sampling procedures should minimize soil disturbance and possible VOC loss due to volatilization by using methanol field preservation. Refer to Section 4.2.5 for additional details on the field collection of VOC samples and precautions (hazardous material handling and shipping) for methanol field preservation. Collection is based on the high-concentration method as described in Sections 2.2 and 8.2.2 of USEPA SW-846 Method 5035A (USEPA 2002f), with a minor modification to the sample container (bottle) to accommodate the increased number of increments (soil mass) and methanol volume per ISM sample. Typically, a coring device (Figure 5-10) and larger narrow-mouthed amber bottles (500 to 1,000 mL) with Teflon-lined caps (Figure 5-11) are required for ISM VOC sampling.

ISM samples can be collected for VOC contaminant analyses, with ISM increments placed directly into the appropriate volume of methanol in the field.

Typically, the bottle and solvent are prepared and pre-weighed at the laboratory prior to shipment to the field to allow for laboratory calculation of the final ISM soil mass. The volume of solvent should at least equal the mass of soil that will be introduced, and the headspace to preserved sample ratio (methanol + sample) should be less than or equal to that commonly achieved with discrete methanol VOC preserved samples (say, 32 mL headspace to 8 mL preserved sample). Appropriate surrogate compounds should be added to the sample bottle containing the methanol prior to sample collection, if appropriate to meet project-specific needs. Details should be specified in the SAP, and any alterations due to unforeseen field conditions should be recorded in field logs. When target analytes require immersion in a solvent, trip and field blanks (no sample added) should be included, depending on the DQOs. For example, when sampling for VOCs, if samples are immersed in methanol in the field, then trip blanks and field handling blanks – that is, bottles containing methanol – should travel to and from the field, and the field blank bottle(s) should be opened in the field under the same conditions and for the same amount of time as the sample bottles. Close coordination with the analytical laboratory regarding ISM VOC bottle/preservation requirements, sample kit preparation, sample receipt requirements, and so on is essential.

U.S. Department of Transportation regulations limit air shipment of containers with more than 30 mL of methanol (DOT 2019). As an alternative, a 60-mL bottle with 25 mL of methanol can be used to collect five 5-g increments, then the methanol extracted from the appropriate number of small bottles can be combined to represent the entire DU. A five-increment small bottle facilitates working with 30-, 50-, and 100-increment DUs. Alternate containers with flexible sides to minimize headspace during sampling are in development and can be considered when commercially available.

Figure 5-10. Examples of coring devices for VOC soil increment collection. Core N' One™ tool (left), Terra Core Sampler (center), and Easy Draw Syringe® and PowerStop Handle® (right). 2020.
Source: Courtesy www.ennovativetech.com.

Figure 5-11. Bottles containing methanol and 44 5-g plugs of soil.
Source: ITRC ISM-1 Team, 2012.

A minimum of a 1:1 ratio of solvent volume to sample soil mass (that is, 1 mL of methanol to 1 g of soil) is recommended. This is a conservative recommendation, since a 5-g plug of soil typically has a volume of around 3 mL. Soil increments should remain completely submerged at all times, so additional solvent may be required to ensure that the sample mass meets this requirement (this requirement should also be discussed with the laboratory). The sample container should be selected based on the total mass of soil to be collected and the solvent required – for example, for 30 increments of 5 g each will require approximately 3 mL volume of solid material per increment, so a minimum of 150 mL solvent is recommended (Figure 5-11). The container should be large enough to accommodate additional solvent (if needed) and to prevent loss of solvent through splashing as soil increments are dropped into the container. The headspace to preserved sample ratio (methanol + sample) should be less than or equal to that commonly achieved with discrete methanol-preserved VOC samples. Potential headspace loss in ISM VOC samples is expected to be comparable to conventional discrete methanol-preserved VOC soil samples (see USEPA SW-846 Method 5035A) (USEPA 2002f).

Additionally, a separate unpreserved soil sample for percent moisture determination should be collected, if necessary, to report the ISM VOC results on a dry-weight basis. Typically, the unpreserved soil sample is collected in the same manner as the ISM VOC samples, meaning a second increment is collected at each ISM increment location, placed in an unpreserved wide-mouth container (4 oz or larger), and submitted to the laboratory.

The following equipment and information are necessary for laboratory processing and analysis of ISM VOC samples:

- The bottle tare weight (including sample label) and volume of methanol must be documented to back calculate the soil mass in the submitted ISM VOC sample. The density of methanol (0.7918 g/cm^3) should be used for the calculation.
- The laboratory must have an analytical balance capable of weighing the ISM VOC sample as received.
- A separate unpreserved soil sample, collected in the same manner as the preserved ISM VOC sample, should be submitted for percent moisture determination.
- If required, total volume and moisture correction should be performed by the laboratory for final contaminant concentration reporting, per Section 11.10.5 of USEPA SW-846 Method 8000C.

Typically, a 24-hour period is sufficient to extract VOCs from most soils. Tight clays are an exception and may take several days (Fehsenfeld et al. 1992). Therefore, caution should be taken if the plugs of soil do not readily disperse when submersed in methanol. Soils should be completely disaggregated or dispersed in the solvent to ensure efficient extraction.

A potential drawback of ISM for VOCs is that the methanol preservation (high-concentration method) approach results in lower sensitivity. Specifically, the methanol dilution step causes elevated analytical detection limits, method detection limits (MDLs), reporting limits (RLs), practical quantitation limits, and so on compared to the direct soil purge-and-trap, low-concentration method's techniques. Analytical detection limits could be elevated above relevant screening levels for certain targeted contaminants. If the analytical detection limits (or other issues) present difficulties in using ISM for VOCs, this issue should be discussed with the laboratory and the overseeing regulatory agency prior to sample collection. If the projected analytical detection limits are too high to be of use or some other issue restrains the use of these methods at a specific site,

alternative approaches may need to be used. Options may include alternate analytical methods/techniques (such as SIM) to achieve lower detection limits or select discrete sampling via USEPA SW-846 Method 5035A (low-level VOC sampling).

5.3.7 Sample digestion or extraction

Analysis of energetics uses the solvent acetonitrile for extraction from the soil sample, with a small portion of the acetonitrile extract analyzed by chromatography, usually using SW-846 Method 8330 (USEPA 1994). Typically, a 10-g subsample built from 30 increments of the milled material is extracted with 20 mL of acetonitrile. Walsh and Lambert (Walsh and Lambert 2006) found acetonitrile extraction on a shaker table was equivalent to using acetonitrile in an ultrasonic bath. Other organic analysis subsamples are usually in the 10- to 30-g range.

Metals analysis with Method 3050B involves a standard digestion mass of 1 to 2 g using nitric acid to recover the environmentally available metals and hydrogen peroxide to remove organics. Improved precision is evident with larger digestion aliquots (Clausen et al. 2018c). Consequently, for metals analysis, it is recommended that 10 g of material be obtained during subsampling and subsequently digested if the sample has not been milled.

Subsamples smaller than 10 g tend to have more variability but still might meet project objectives. Collect an initial subsample of 10 to 30 g and split the subsample into the target subsample mass (<10 g) using miniature versions of 1D or 2D slabcake.

5.3.8 Analysis

Method 8330 specifies using High Performance Liquid Chromatography with an ultraviolet detector, which has been the most widely used analytical approach for detecting energetic compounds in soil samples from military sites. Another method (USEPA 1999b) employs the same sample processing steps as Method 8330 but uses a gas chromatography with an electron capture detector for determination. There is no reason that this method of determination could not be used with the sample processing steps specified in Method 8330B.

5.4 Quality Assurance/Quality Control

The overall quality assurance program provides the structure and contains the specific quality control elements used to meet project DQOs.

5.4.1 Introduction

Regardless of the COPCs, it is imperative that sample processing and analysis include appropriate QA and QC measures. QA/QC requires careful upfront planning during DQO development (see Section 3.1) and is dependent upon the intended use of the data. In addition, practitioners must also consider any corrective actions and/or decision rules that will be followed in the event that any QA/QC milestones are not met. Most of the QA/QC criteria applied to laboratory ISM procedures evaluate representativeness of the processed laboratory subsample, analyte loss, or cross-contamination during processing. Therefore, it is emphasized that failure to adhere to processing procedures of samples will affect the ability to meet project quality goals. Users should first have a firm understanding of the differences between QA and QC.

5.4.2 DQOs and laboratory coordination

As outlined in USEPA DQO guidance (USEPA 2006b), the DQO process is used to establish the performance and acceptance criteria that serve as the basis for designing a plan for collecting data of sufficient quality and quantity to support the goals of the study. The project delivery team must decide during the initial project planning phase which of the sample processing and analytical options currently available and applicable to ISM are most appropriate to achieve the ISM project DQOs. This includes sample conditioning steps, such as drying; disaggregation; particle size reduction (if warranted) via milling, crushing, or other means of particle size selection using sieving to focus on a particle size fraction of interest; and finally what analytical subsampling techniques and/or determinative analytical methods will be performed. QA/QC can also be affected if the processing is not in a controlled environment – for example, if any sample processing such as sieving, drying, or subsampling occur in the field.

The project delivery team will determine the project needs for sampling and sample processing as documented in the QAPPs. Before award, the laboratory must review the QAPPs (including all QA/QC criteria) and appropriately determine if

they can meet project needs, so close interaction of the team or the project chemist with the laboratory is critical here. Specifically, although the processing, preparatory, analytical methods, and procedures needed to meet the DQOs are predetermined and documented in the QAPPs, the laboratory will compare in-house capabilities to project QA/QC and determine if they can meet those project needs. It is also critical that the QAPP document any corrective actions that will be necessary in the event that any DQOs are not met – for example, will sample reanalysis be required if any surrogate or grinding blanks fail performance criteria?

The following sections discuss various options regarding sample processing and analysis for ISM. It is imperative that close communication and coordination with the analytical laboratory take place from the initial project planning phase and DQO formulation through ISM sample collection and subsequent sample processing and analysis to ensure that defined data of known quality and usability are obtained for the project. The DQOs, including QA/QC criteria, are always planned for and documented in the project QAPPs.

5.4.3 Quality assurance

Choosing the specific laboratory processes to handle and analyze incremental samples is influenced by the specific COPCs, the nature of the release, and the objectives of sampling event. For example, COPC volatility and thermal stability affect the options for sample conditioning, while exposure scenarios and the objectives for risk assessment may influence choices related to particle size reduction. In addition, sample drying and particle size reduction can create a negative bias due to analyte loss. Biologically degradable analytes with very high boiling points may remain stable when the sample is air-dried, but some of the processing procedures may need to be modified or avoided entirely if a COPC is sufficiently volatile and/or biodegradable. Negative bias created by target analyte loss may also significantly impact attainment of project DQOs, and as these considerations apply to certain volatile metal species, they require careful control to maintain acceptable milling temperature. Negative bias due to analyte loss can affect MDLs and precision as well, so in all cases, the laboratory selected to perform the sample processing and analysis should have empirical data showing that each ISM standard operating procedure (SOP) has been appropriately validated (including the potential affects from all sample processing steps) for each analyte being evaluated and, where possible, in representative matrices. A project chemist should compare laboratory SOPs to the project's QA/QC requirements as documented in the QAPPs.

5.4.4 Quality Control

Monitoring for bias due contamination or analyte loss is accomplished with blanks and the use of known concentration spikes or samples. Precision is monitored with replicates.

5.4.4.1 Monitoring cleanliness and carryover

ISM sample processing can result in a positive bias due to sample cross-contamination of equipment – for example, milling equipment can contribute metal concentrations to the sample through the milling, crushing, or pulverizing apparatus. Common metals include chromium, cobalt, iron, manganese, nickel, and tungsten, with metallic composition analysis and guidance usually available from the manufacturer (such as avoid high chrome steel when low ppm concentrations of chromium are of interest). Other malleable metals, such as lead or copper, may smear in milling machinery. If a significant amount of larger particle size, malleable metals are expected in ISM samples, additional sieving and fractional analysis should be considered, or alternative sample preparation techniques may need to be investigated.

With each of these subsampling techniques, consideration should be paid to the potential for contamination. Decontamination processes must be developed and checked using a matrix such as blank Ottawa sand at an established frequency between samples. The composition of the subsampling equipment should also be considered as a potential contamination source. Plastic parts containing phthalates, for example, should be avoided if SVOC phthalates are COPCs.

Sample preparation blank. Sometimes referred to as a method blank, a sample preparation blank is a clean sample matrix that, when available, is used to establish whether equipment used to process samples (dry, pulverize, split, mix, and so on) has been adequately cleaned between field samples. The sample processing blank is used to assess whether there is carryover or cross-contamination during the processing steps. The blank should be evaluated by milling and analyzing a clean matrix immediately after preparing a sample of known or suspected high malleable metal content, but note that clean soil matrices are more likely to be available for organic analytes. For metallic analytes, no known soil-like matrices are available that will produce ultra-trace (non-detect results) at environmental levels of concern for all likely metals. Again, the project team should work directly with the laboratory to first determine underlying project needs, then establish the

tolerable levels of potential carryover/cleanliness. Method blanks should go through all the phases of sample preparation, subsampling, extraction, and analysis that are experienced by field samples.

5.4.4.2 Monitoring bias

Laboratory control sample (LCS). The LCS is a known matrix spiked with compound(s) representative of all target analytes. It represents a "best case" control of overall method performance as it is performed on a clean matrix spiked with the COPCs. It is also used to document possible analyte loss and/or overall laboratory method performance. LCS control limits must be established by the laboratory for each ISM procedure and analysis performed, and then provided in the final laboratory report. Usually, there is one LCS sample processed and analyzed per batch. As the LCS is intended to evaluate overall method performance, it should experience the same processing steps as that of field samples.

However, many laboratories have introduced the LCS after ISM sample processing because it is costly to add all target analyte spikes of sufficient concentration to a 1-kg or greater ISM sample. This might be acceptable on a project-specific basis, but the potential bias caused by sample preparation procedures prior to LCS processing must be evaluated by other means (such as surrogates). LCS acceptance criteria are typically generated by each laboratory using appropriate statistical evaluation of multiple LCS samples, but monitoring the air-drying and sieving steps may be problematic for many organic compounds. The deposition of the spiking solution onto the LCS might not result in a sample with spiking compounds bound in the same manner as the sample contaminants themselves. Moreover, the association between low boiling point SVOCs and the clean soil or sand matrix might be significantly weaker than in weathered field samples. Thus, potential losses from an air-dried laboratory prepared control sample can be significantly higher than from a field sample.

Theoretically, it is also possible that COPCs are bound significantly to the matrix and will not be dissociated completely during the extraction/digestion step, but these same compounds will be easily extracted/digested in the LCS. However, if, for example, air-drying, sieving, and subsampling are the only ISM sample processing steps being performed, a "standard" (that is, approximately 30-g) clean matrix spiked LCS carried through all these steps would present the potential "worst case" analyte loss to be evaluated for some analytes and the "best case" analyte recovery to be evaluated for other analytes. The best case/worst case scenario for the LCS exists for discrete samples, too. The typical sample size of an ISM sample is 1 kg, but an LCS may not need to be that same size – it may only require the same preparation and analysis process.

Synthetically fortified soils may not produce the same strength of interactions between the COPCs and the soil particles. In particular, QA/QC materials spiked at the laboratory or other commercial providers may overestimate contaminant losses during ISM sample processing steps. Reference materials from weathered "native" contaminated soils are more likely to match the loss rates for field samples; standard reference materials (SRMs) are currently available for explosives (USEPA 2006c), as is a 500-g solid commercially available QA/QC standard for energetics. Such material is often analyzed as an LCS on a per batch basis, but other analyte groups might become available over time.

Project-specific DQOs should be assessed during systematic planning to determine the appropriate analysis frequency. Nitrobenzene, 2-nitrotoluene, 3-nitrotoluene, and 4-nitrotoluene have low recoveries when the QA/QC standard is air-dried at room temperatures, so the DQO process needs to address whether the QA/QC standard will be air-dried or only milled. There can be significant costs associated with a commercial QA/QC standard, but a separate QA/QC standard is available for tetryl (an energetics constituent) and should be considered if it is a target analyte. The frequency at which a QA/QC standard needs to be processed and analyzed should be defined during the SPP, depending on project-specific DQOs. With respect to energetics, additional guidance for laboratory QA/QC can be found as part of USEPA SW-846 (USEPA 1999b) and "DOD Quality Systems Manual for Environmental Laboratories (QSM)" (DOD/DOE 2018).

ISM samples collected for non-volatile metals may also include drying, sieving, and milling preparation steps. It is assumed that this process does not cause the loss of metal analytes, so it may not be necessary to require a large-scale LCS throughout the entire process. The necessity for a metals LCS (large-scale or otherwise) should be defined during the SPP, depending on project-specific DQOs.

Monitoring the effectiveness of the milling step for metals, explosives, or other particulate-based analytes is best demonstrated by adding these analytes in solid particulate form (such as metal salts), rather than the traditional liquid spike solutions used by laboratories. Demonstrating the ability to produce representative subsamples from heterogeneous samples would require the original QA/QC sample to be intentionally heterogeneous and not the highly homogenized reference materials commonly available from providers.

Much of the focus of this QA/QC section is targeted on the milling portion of the ISM process, largely due to the paradigm shift from conventional sample preparation and analyses. Simply put, particle size reduction is an invasive sample handling

technique and therefore requires an additional level of QC. Note that for some organic analytical methods using ISM, particle size reduction may not be necessary. The primary purpose for milling is reduction of FE by reducing the particle size and eliminating larger nuggets that can be the cause for extreme in-sample heterogeneity.

***Matrix spikes*.** MS are intended to evaluate any potential intrinsic matrix properties of the field samples that might affect sample extraction and analysis. An MS is an aliquot of an actual field sample spiked with a known concentration of target analytes, and as the intended objective of the MS is to evaluate matrix issues rather than ISM processing, the spiking occurs just prior to extraction and analysis. Due to the bulk mass spiking limitation, modifications may be necessary for MS analysis. For sites with a large degree of heterogeneity, it may be necessary to collect a duplicate ISM sample to use with this type of MS approach so as not to remove a portion of the primary ISM sample because it could ultimately bias the representativeness of the analyzed sample result to that of the original sampled area. MS/MS duplicates (MSD) might not be needed for precision assessment if subsample replicates are used.

Surrogates. A surrogate is used in the preparation and analysis of samples for organic analytes in order to evaluate the potential bias introduced during sample processing. In this case, surrogates are added to the entire sample as received from the field and go through all the same sample preparation and analysis steps. The surrogates used should be similar to the target analytes and are typically isotopically-labeled analogs of the target compounds of concern. Unlike LCS and MS, which are performed once per batch and in a clean matrix, surrogates are added to each sample, blank, and LCS prior to any processing steps, and the resulting recoveries allow the user to evaluate any potential bias experienced by that specific sample. As previously discussed, an ISM field sample received for analysis must be processed as received in its entirety. Because ISM samples are typically 1 to 2 kg in mass, there are logistical and financial challenges that come with using surrogates with each ISM sample:

- Laboratory costs significantly increase as additional waste is generated. In addition, the concentration of the surrogate fortification stock must be appropriate in order to achieve a final concentration within the linear range of the calibration after sample processing, subsampling, extraction, and analysis.
- Depending on the corrective actions chosen in the event that QC recoveries fail performance criteria, multiple field replicates may need to be collected if the corrective actions require samples to be reprocessed. In addition, multi-method surrogate additions to the sample field replicate may not be feasible as surrogates from one method may interfere with the analysis of another method – for example, surrogates used for Method 8270 interfere with some PCB, organochlorine pesticide, and petroleum hydrocarbon analyses. This may also add to the costs incurred for the collection, transport, and storage of the additional samples.
- Standard recommended surrogate recoveries are provided in each respective organic method in USEPA SW-846, but these method-specific criteria do not include ISM-specific sample preparation procedures (such as drying and milling). Therefore, method-specific surrogate recoveries are not recommended, and each laboratory should define their own performance criteria based on empirical data during method validation.
- As previously mentioned for LCS fortification, because of the difference in the strength of interactions between weathered soil particles and contaminants, surrogates spiked at the laboratory or other commercial providers may overestimate contaminant losses during ISM sample processing steps. Therefore, stable surrogates pre-fortified onto solid spike materials are recommended, rather than simply spiking the surrogate as a solution onto the field sample.

5.4.4.3 Monitoring precision

***Laboratory replicates*.** Laboratory replicates (also known as subsampling replicates) are recommended to assess the precision of ISM subsampling processes and ensure that the subsample selected for analysis is representative of the entire field sample collected for processing. Three subsample replicates are recommended on samples selected by the project team with a targeted RSD as determined during the project-specific SPP. Generally, replicate subsamples of 30 to 50 increments each should be collected after all ISM processing is complete. These replicates should then be carried through the rest of the analytical process. A 15 to 30% RSD for laboratory replicate precision is generally considered confirmation that the processed sample is sufficiently homogenized. The frequency of these laboratory replicates can vary from one replicate set per batch to one set per project, depending on the project DQOs. Therefore, including appropriate subsampling procedures and monitoring subsampling replicate precision is especially critical where sample drying and particle size reduction have not been performed. Historically, analytical processing precision has been evaluated via MS and MSD recoveries, but evaluation of this precision through the evaluation of laboratory subsampling replicates is a superior metric when the analytes are present at sufficient concentrations.

5.4.4.4 Other QC considerations

It is often desirable to perform sample splitting (obtaining identical splits of a field sample post sample collection) in order to send the splits to separate laboratories. Paired ISM sample collection (or field replicates) is generally recommended over bulk ISM sample splitting when different sample processing treatments will be needed. Paired ISM samples allow separate sample processing procedures to be conducted without the uncertainty introduced through bulk splitting. Moreover, the error introduced by splitting prior to the completion of sample processing can be large when the COPC nugget effect is large, such as in highly heterogeneous samples. Note that bulk sample splitting (or subsampling) without particle size reduction merely increases FE.

To help ensure data quality, all field sampling, field processing, and laboratory sample processing activities should be supervised by personnel trained in ISM. Samples should be shipped to an accredited laboratory following recommended protocols for the class of target analytes, and laboratory SOPs for ISM sample processing and analysis should be requested and reviewed by the project chemist. Chain-of-custody, field notes, and data validation reports should also be retained and utilized in the final data usability assessment (DUA).

QC measures should be implemented both in the field and laboratory. When sample processing and subsampling are initiated in the field to reduce the amount of sample shipped off site, replicate samples of the processed soil should be taken to establish the uncertainty introduced by this step (see Section 2.6.2.1). It should be noted that reducing the mass of the sample shipped to the laboratory will tend only to increase FE. Depending on the contaminant, field blanks and/or equipment blanks also may be required. Field blanks are often necessary for VOCs and some SVOCs, particularly when a solvent is involved.

Laboratory accreditation/certification. Project teams must be aware of the accreditation requirements that apply to their projects because such requirements may vary based on the program and state under which the sampling is being performed. They may also vary based on whether the procedure follows a formal published method, is based on a formal published method, or is an internal laboratory procedure. In most systems, accreditation is given at the fields of testing (FOT) level, where each combination of matrix (such as non-potable water, drinking water, and solid and chemical materials), method/technology, and analyte is considered an FOT.

Three primary types of accreditation requirements exist:

- The National Environmental Laboratory Accreditation Program (NELAP) is a national program implemented by member states, with state governmental agencies usually serving as accreditation bodies for state-selected programs and FOTs. A NELAP accreditation body will accept by recognition the accreditation status of a laboratory issued by another NELAP accreditation body (called secondary accreditation). For more information, see www.nelac-institute.org.
- Each state has its own procedures to address accreditation of method modification and internal laboratory procedure. Some states have elected not to participate in NELAP, and some have retained separate accreditation structures for certain programs. Each of these states has its own procedures to address accreditation of method modification and internal laboratory procedure.
- Some federal agencies have their own accreditation programs. The U.S. Departments of Defense and Energy (DOD and DOE) have centralized a Environmental Laboratory Accreditation Program (ELAP). The DOD ELAP program has specific ISM quality requirements listed in the “DOD Quality Systems Manual” (DOD/DOE 2018) that cover laboratory QC requirements for samples collected via incremental sampling. In addition, the DOD QSM provides ISM-specific quality requirements for explosives by 8330B (QSM table B-3); ISM samples being analyzed for parameters other than explosives should utilize QSM table B-23.

Data Quality Evaluation

This section will demonstrate how to determine if ISM data are sufficient and will also provide guidance on using appropriate statistical methods for data evaluation and confident decision-making.

6.1 Data Verification, Validation, and Usability

As mentioned in previous sections, careful planning is needed prior to using an ISM sampling design, and part of that planning should include how the ISM data will be evaluated relative to the intended use as defined in the DQOs. This evaluation should include verification, validation, and an assessment of the final usability relative to the defined study questions. Users are referred to UFP-QAPP worksheets for details on planning and documenting these data evaluation steps (USEPA 2005).

As part of the data evaluation process, the project team should consider what inputs will be used during data verification, data validation, and DUAs. Inputs might include but should not be limited to field records, interim reports and final reports of analyses (including QC results from milling blanks, surrogates, and spike recoveries), replicate results, and further data validation reports. Inputs for data evaluation should document the procedures that will be used to verify and validate project data, including the specific requirements/specification, who is responsible for performing each activity, and how they will be documented.

6.1.1 Data verification

Data verification is a completeness check that all specified activities involved in data collection and processing have been completed and documented and that the necessary records (objective evidence) are available to proceed to data validation. This includes the need to ensure that the sampling plan was properly designed and that the samples were properly collected in the field and processed for testing at the laboratory. For example, if the ISM sampling design called for 30 increments to be collected in each DU but only 25 were taken, this would be documented during the data verification evaluation.

6.1.2 Data validation

The quality of the sample data generated must be reviewed to determine if the data are reliable to answer the risk and/or remediation-based questions prepared at the beginning of the project. This requires a review of the sampling plan design and the methods used to collect the samples.

Data validation is an analyte- and sample-specific process for assessing conformance to stated requirements, methods/SOPs, and specific performance criteria. The scope of data validation needs to be defined during project planning because it affects the type and level of documentation required for both field and laboratory activities. If data validation procedures are contained in an SOP or other document, the procedures should be referenced or included as an attachment to the QAPPs.

Data validation includes confirming the qualifiers assigned to results, with any data qualifiers applied by the data validator clearly defined. Common data qualifiers include estimated values (such as J flags), which is when results are greater than an MDL and less than a method RL (MRL), and non-detects (such as U flags), which is when results are less than an MDL. For ISM datasets that include one or more non-detects (such as censored data), there is greater uncertainty in parameter estimates, including 95% UCL calculations. At this time, numerical simulations used to support the recommendations in this report regarding statistical analysis methods for ISM data (see Section 3.2) are based on datasets that are not censored, so there is uncertainty in applying the same methods to censored data.

Many methods exist for parameter estimation for censored data, and although their computations are no different for ISM data than for any other environmental dataset, caution must be used when sample sizes are small. EPA provides guidance statistical analysis with censored data based on minimum sample sizes of 8 to 10 observations, of which at least five are detects. EPA notes that Kaplan-Meier (KM) estimation has been shown to yield more reliable estimates of the AM and SD than imputation (that is, substitution) methods such as MDL, MDL/2, and under these conditions. However, for ISM datasets with small sample sizes (<5 detects) even KM methods are likely to yield very high uncertainty in parameter estimates. Therefore, it may be advisable to explore both imputation and KM methods as a form of uncertainty analysis, particularly if the MDL is close to a relevant AL (within one SD).

Data validation should also note that when performance criteria are not met, such as failing to meet QC recoveries for surrogates or spikes. Note that final rejection of any data and their use is a decision reserved specifically for the project team and will be made during the DUA.

6.1.3 Data usability

The DUA is performed by key members of the project team, as defined during the SPP at the conclusion of data collection activities. The DUA should be integrated into the definable features of work where decision-making occurs. For phased investigations, the DUA and decision-making will occur at each phase. Table 6-1 summarizes the recommended steps of a DUA.

- Summarize lessons learned and make recommendations for changes to the DQOs or the sampling design for the next phase of investigation or future investigations if needed.
- Prepare the data usability summary report.

Table 6-1. Recommended steps for a DUA.

Source: ITRC ISM Update Team, 2020.

Step 1	**Review project objectives and sampling design.** • Review the DQOs. Are underlying assumptions valid? • Review the sampling design as implemented for consistency with stated objectives. Were sources of uncertainty accounted for and appropriately managed? • Summarize any deviations from the planned sampling design and describe their impacts on the DQOs.
Step 2	**Review the data verification/validation outputs and evaluate conformance to performance criteria.** • Review the data verification/validation reports and supporting data, if necessary (daily/weekly QC reports, assessment reports, and corrective action reports). *Were the corrective actions effective?* • Evaluate the implications of unacceptable QC results. • Evaluate conformance to performance criteria. • Evaluate data completeness. *Were all data inputs satisfied?* • Identify data gaps.
Step 3	**Document data usability, update the CSM, apply decision rules, and draw conclusions.** • Assess the performance of the sampling design and identify any limitations on data use. Considering the implications of any deviations and data gaps, can the data be used as intended? Are the data sufficient to answer the study questions? • Update the CSM, apply decision rules, and document conclusions.
Step 4	**Document lessons learned and make recommendations.** • Summarize lessons learned and make recommendations for changes to the DQOs or the sampling design for the next phase of investigation or future investigations if needed. • Prepare the data usability summary report.

The DUA involves a qualitative and quantitative evaluation of environmental data to determine if the project data are of the right type, quality, and quantity to support the measurement performance criteria (MPC) and DQOs specific to the investigation. It involves a retrospective review of the SPP and the CSM in order to evaluate whether underlying assumptions are supported, sources of uncertainty have been managed appropriately, data are representative of the population of interest, and the results can be used as intended with an acceptable level of confidence. The DUA is a retrospective evaluation of all aspects of the data and project investigation to determine whether the data support their intended uses. Therefore, the DUA represents a critical step in determining if final decisions can be made as stipulated in the DQOs.

A summary matrix of each topic is provided in Table 6-2. The table is not intended to be comprehensive for all aspects of the investigation and should be modified as appropriate on a site-specific basis. Refer to the noted sections of this guidance document and related appendices for detailed information on each topic. Deviations from the recommended methods should

be discussed in the investigation report and resulting limitations of the data collected, described, and considered in the report's recommendations. Methods to help minimize data error when the sample collection and analysis conditions noted in Table 6-2 cannot be met are discussed in the associated appendices.

Table 6-2. Sample data quality and usability matrix.

Source: ITRC ISM Update Team, 2020..

Acceptable?	Site Investigation Stage
CSM and DU Designation (see Section 3.1)	
	Site history and potential sources and type of contamination well understood?
	Site investigation questions used to designate DUs for testing clearly stated and based on risk and/or optimization of anticipated remediation requirements?
	Questions and decision statements developed for individual DUs presented?
	Area and total volume of soil associated with each DU noted and acceptable for intended purposes?
	To-scale map depicting location and size of DUs provided?
Field Sample Collection (see Section 4)	
	Summary of sample collection methods provided, including approximate final mass of each sample?
	ISM samples prepared by collecting and combining a minimum of increments appropriate for the chemical present and nature of contamination?
	Increments appropriately spaced and collected?
	Complete, unobstructed access to all portions of the DU soil available for sample collection?
	Increments collected over a uniform depth and without biasing soil size?
	Samples to be tested for volatile chemicals preserved in methanol in the field or met requirements for alternative preservation and testing methods?
	Minimum sample mass of 1 kg met (minimum 300-g for samples to be tested for volatile contaminants)?
	Three replicate ISM samples collected appropriately to test total data precision or calculate CIs?
Laboratory Processing and Testing (see Section 5)	
	Samples to be tested for non-volatile chemicals air-dried and sieved to target particle size for each specific DU?

	Analytical subsample collected using a sectorial splitter or manually collected from at least 30 points?
	Minimum 30-g analytical subsample mass extracted for <2-mm particle size soil?
	Minimum 10-g analytical subsample mass extracted for <250-μm particle size soil?
	Three replicate analytical subsamples collected from at least 10% of samples submitted (minimum one set)?
	Holding times met?
	Analytical QC and QA criteria met (spikes, blanks, and so on, refer also to USEPA 2002)?
Replicate Sample Collection and Data Precision Evaluation	
	Replicate field sample and laboratory subsample data meet data precision requirements?
	Source of error determined for replicate data that exceed the RSD identified as problematic during planning?
	Laboratory subsampling error identified and subsamples recollected after grinding of primary sample or larger subsample mass collected?
	Data adjusted or new samples collected for DUs with replicate data that exceed the RSD specified as unacceptable during planning?

6.2 Evaluation of DU Results

As in discrete sampling, there is no one decision mechanism dictated by the use of ISM sampling – a variety of decision mechanisms are possible. Each decision mechanism has strengths, weaknesses, and assumptions. In some cases, agency requirements will dictate the approach to be used, but in others, a consensus on the decision mechanisms to be employed needs to be reached among members of the planning team prior to finalization of the sampling plan. Because ISM yields estimates of mean concentrations within a DU, it is important to note the spatial and/or temporal scale that was originally intended in the development of the AL.

6.2.1 Evaluation of DU results

This section provides three examples of decision mechanisms and how to evaluate ISM data. The DQOs may specify that DU results are either used in risk assessment or compared with ALs, which may consist of regulatory screening values, threshold values derived from risk assessment, or other regulatory values.

6.2.1.1 Single DU sample result

The simplest decision mechanism is the comparison of a single ISM sample result for a DU to an AL. Sometimes, more than one set of ALs may apply to a site because they reflect different objectives (human health or ecological endpoints, leaching and groundwater protection, short-term risk, and so on). This decision mechanism is simple, straightforward, and the most cost-effective, since only one ISM sample is collected. The result of the decision is immediately apparent, too – a failure is indicated if the ISM sample result exceeds the AL.

A single ISM sample provides one estimate of the mean concentration, so the practitioner must be aware of the inherent limitations when using a single ISM result in regard to potential uncertainty with the comparison. Single ISM results do not

allow for the calculation of a DU-specific CI or quantification of the precision of the estimate. As a result, it is difficult to predict how far from the actual mean a single ISM sample result might be. Use of a single ISM result might be acceptable when the estimated mean concentration obtained is much greater than (or much less than) the AL such that even a great deal of error in the mean estimate could be tolerated without making a decision error. Note that predetermination of the "much greater than" or "much less than" values should be addressed during the site planning process. In this situation, the ISM sample may provide confirmation of what may have already been strongly suspected – that the DU clearly passes or fails.

These decisions are site-specific and cannot be outlined in a guidance document, but there is consensus that the uncertainty about making the right decision increases as the ISM sample result gets closer to the AL.

The uncertainty of a single ISM sample result may be reduced if numerous single ISM sample results are evaluated, assuming all DUs fall within the same CSM and site assumptions. For example, if 10 single ISM results are all considered comparable, it is unlikely that all of the results either under- or overestimated the actual mean concentration.

6.2.1.2 Mean of replicate DU data

In this decision mechanism, replicate ISM samples are collected in the field from the same DU and provide a measure of the variability of the entire sampling, preparation, and analytical process. The mean concentration of the replicates is calculated and compared to the AL and is likely to be closer to the true mean of the DU than the result from a single ISM sample, therefore, it can be considered to provide a more reliable estimate of the mean. Because a CI has not been calculated, though, there is no assurance that the actual mean concentration has not been under- or overestimated, so this decision mechanism cannot be used for projects requiring the calculation of a CI, such as a 95% UCL.

6.2.1.3 95% UCLs, CIs on the mean of replicate DU data

In this decision mechanism, replicate ISM samples are collected in the field from the same DU as in the previous example, except that a CI may be calculated and applied to the mean concentration. The 95% UCL is commonly selected as an EPC, which can be calculated based on the measured RSD from the replicates. Use of a 95% UCL improves the confidence associated with comparing the ISM results to the ALs. It provides protection against underestimation, but it may also overestimate the true mean concentration.

Project objectives may specify that the estimate of the mean concentration provided by ISM sampling must be health protective, meaning that there is a low chance of underestimating the actual mean concentration within the DU, which is achieved by calculating the 95% UCL. It is important to recognize that the likelihood of underestimating the mean from any sampling method (discrete, composite, or ISM) increases as the degree of heterogeneity increases.

For those accustomed to working with 95% UCL values from discrete datasets, there are some important differences with ISM data. Discrete sampling designs generally require more samples than ISM designs because of the greater variability and reduction in site coverage. A 95% UCL for ISM data may be calculated with as few as three ISM samples (see Section 3.2.4). Additional ISM replicates increase the performance of the mean estimate (providing a 95% UCL closer to the actual mean), and although this increases the cost, it may be worthwhile if the site is relatively large or heterogeneous, and the result is anticipated to be close to the AL.

A second difference involves what to do if the 95% UCL is higher than any of the individual values used in its calculation. With discrete datasets, the maximum concentration observed is often used if it is less than the calculated 95% UCL (though the sample maximum will not necessarily be a conservative estimate of the population mean). With ISM data, the calculated 95% UCL value should be used even if it is higher than any of the individual ISM results. This situation is common when three replicate ISM samples are collected because the 95% UCL always exceeds the highest individual ISM result. Two methods for calculating the 95% UCL from ISM data are available for any ISM design: Student's-*t* and Chebyshev. For designs with at least seven ISM results for each DU, bootstrap methods may also be used. Bootstrap 95% UCLs often result in a more accurate estimate of the true mean than Chebyshev, but coverage may be reduced slightly below 95% (see Appendix B). As discussed in Section 6.3.2, the choice of method depends on the known or anticipated shape of the PD of contaminant concentrations in the DU. Note that software programs for calculating 95% UCL values for discrete sample data (such as ProUCL) contain algorithms optimized to perform well for discrete data only, thus, they are generally unsuitable for calculating 95% UCL values for ISM data. An ISM 95% UCL calculator for deriving 95% UCL values for ISM data is presented in Section 3.2.4.1.

6.2.2 Weighted means and 95% UCLs from multiple SU or DU results

Projects may require combining several SUs into a DU or perhaps determining multiple DUs to make a site decision. There are two types of instances where this might occur:

- A site has areas with different conceptual models in terms of expected contamination, as could happen when there is, for example, a stream channel, a meadow, and a rocky out-cropping in an area that we would like to define as an EU. Each of these areas might be investigated as a separate SU or DU for site characterization but then combined to define a single EU.
- A site with multiple ecological exposures would consider a variety of sizes of DUs to accommodate multiple receptor scenarios. If, for example, the area of a pocket mouse habitat is a quarter that of a muskrat, which is an eighth of that of an eagle, then we might need to sample in DUs of a size defined for pocket mice but then combine DUs for the receptors with larger home ranges.

When DUs need to be combined, the variable sizes of each can be taken into account by using a weighted mean. (In this context, statisticians often refer to the smaller DUs or SUs as *strata*.) Depending on project needs, a weighted 95% UCL could also be calculated.

▼*Read more*

When there are multiple samples in each of the smaller DUs, the overall mean of the larger DU can be estimated by using a weighted mean. Let w_i represent the weight, that is, the relative size associated with region *i* (stratum), $xbar_i$ represent the mean of ISM samples from region *i*, n_i represent the number of samples from stratum *i*, *k* represent the number of regions, and s_i represent the SD of the replicate ISM samples from region *i*. Note that if all strata are of the same size (volume), the weights are equal, so these equations simplify to the more common calculation methods for the mean and SD. The relative size is the percentage of the larger DU that is made up of region *i*. The weighted mean is thus as follows:

$$Weighted\,Mean = \sum_{i}^{k} w_i \overline{x}_i$$

and the SE associated with the weighted mean is as follows:

$$Standard\,Error = \sqrt{\sum_{i}^{k} w_i^2 \frac{s_i^2}{n_i}}$$

which has degrees of freedom approximated by the Welch-Satterthwaite approximation (Cochran 1977):

$$df \approx \frac{\left(\sum_i^n \frac{w_i^2}{n_i} s_i^2\right)^2}{\sum_i^n \frac{\left(\frac{w_i^2}{n_i} s_i^2\right)^2}{n_i - 1}}$$

Table 6-3 provides a numerical example of this calculation where data from two DUs are combined to derive a 95% UCL for a larger DU (α = 0.05). In this example, an elementary school is divided into two DUs representing different play areas: DU1 is the kindergarten playground, and DU2 is the playground for older children. A maintenance worker has contact with both DUs, and a separate DU is constructed to reflect exposure of this worker. Assume the concentrations of replicate results in

DU1 and DU2 are as in the table, based on n = 30 increments per replicate.

Table 6-3. Summary statistics used to combine DUs.

Source: ITRC ISM Update Team, 2020.

Playground Area	Area (Acres)	Sample Statistics		95% UCL	
		Replicates	Mean	Chebyshev	Student's-*t*
DU1 (Kindergarten)	0.25	120, 100, 140	120	170	154
DU2 (Older Children)	1.0	22, 25, 30	25.7	35.8	32.5
Equal Weight	1.25	120, 100, 140, 22, 25, 30	72.8	168	117
Proportionately Weighted	1.25	120, 100, 140, 22, 25, 30	44.5	57.5	50.9

If it is assumed that, on average, a maintenance worker spends equal time in DU1 and DU2, then the replicates from each DU can be weighted equally, yielding the results shown in the third row of Table 6-3. Alternatively, it may be assumed that a maintenance worker's exposure is proportional to the respective areas of each DU, and can be used to generate summary statistics for the combined area. Note that by proportionately weighting the DUs by area, the results are considerably different than if the area was not taken into account for the combined area shown in the last row of Table 6-3. The weighting factors applied to each DU should sum to 1.0, which is achieved by dividing each area by the combined acreage of 1.25:

$$w_1 = \frac{0.25}{1.25} = 0.2$$

$$w_2 = \frac{1.0}{1.25} = 0.8$$

▼*Read more*

The calculations for the proportionately weighted scenario in Table 6-3 follow from the formulas given above:

$$Weighted\,Mean = \sum_{i}^{k} w_i \overline{x}_i = (0.2 * 120) + (0.8 * 25.7) = 44.6$$

$$Weighted\,Standard\,Error = \sqrt{\sum_{i}^{k} w_i^2 \frac{s_i^2}{n_i}} = \sqrt{0.2^2 * \frac{20^2}{3} + 0.8^2 * \frac{4.04^2}{3}} = 2.96$$

$$df \approx \frac{\left(\sum_i^n \frac{w_i^2}{n_i} s_i^2\right)^2}{\sum_i^n \frac{\left(\frac{w_i^2}{n_i} s_i^2\right)^2}{n_i - 1}} = \frac{\left(0.2^2 * \frac{20^2}{3} + 0.8^2 * \frac{4.04^2}{3}\right)^2}{\frac{\left(0.2^2 * \frac{20^2}{3}\right)^2}{3-1} + \frac{\left(0.8^2 * \frac{4.04^2}{3}\right)^2}{3-1}} = 3.8 \approx 4$$

$$Chebyshev\,95\%\,UCL = \overline{x} + \left(\sqrt{\frac{1}{a} - 1}\right) * SE = 44.6 + \left(\sqrt{\frac{1}{0.05} - 1}\right) * 2.96 = 57.5$$

$$Student's - t\,95\%\,UCL = \overline{x} + t_{a,df} * SE = 44.6 + 2.132 * 2.96 = 50.9$$

Download the calculator for the weighted 95% UCL for a combined DU from several smaller DUs (see ISM 95% UCL calculator) from Section 3.2.4.1.

This same methodology can be used to combine a surface DU with its corresponding subsurface DU. The only slight difference would be that the weight term would reflect the proportion of the total soil volume within the DU. For designs with at least seven ISM results, weights could be incorporated into a bootstrap methodology by resampling ISM results with unequal probabilities, in proportion with the desired weights. For more details on bootstrap methodology, see, “The statistical sleuth: a course in methods of data analysis” (Ramsey and Schafer 2012).

6.2.3 95% UCL from multiple SUs or DUs provides spatial distinctions

When replicate samples are collected over the entire DU, each single ISM sample (or singlet) result provides a separate estimate of the mean concentration, and these estimates can then be combined to derive a 95% UCL as discussed in Section 6.2.1.3. Another approach is to divide the DU into SUs and collect one ISM sample from each SU. The results from each ISM sample (that is, each SU) can also be combined to calculate a 95% UCL for the DU. With the latter approach, the ISM samples are not true replicates of the mean throughout the DU in the sense that they provide information on different portions of the DU. Collectively, however, they can provide an unbiased estimate of the mean. Combining singlet SUs or DUs does not provide information about the variability within the singlet if replicates are not available.

The principal disadvantage to this approach is that the 95% UCL often exceeds the true mean by a larger degree than if replicates had been collected across the entire DU. Another disadvantage to combining singlet ISM SUs or DUs is there is no measure of variability within the SU or individual DU when data from replicates are not available. The principal advantage of subdividing the DU for this decision mechanism is that it provides some information on the spatial distribution of contamination. If the DU as a whole fails the comparison with the AL, this spatial information could be valuable if a decision is made to break the DU into smaller DUs for reevaluation.

For example, the DU in Figure 6-1 displays a 95% UCL ISM concentration (calculated based on three replicates) that exceeds the AL of 4.0 (Figure 6-1A). Because spatial information is not available to help determine where in the DU the AL is exceeded, the entire DU is considered to be in exceedance. To obtain spatial information on the contaminants, the DU is divided into smaller SUs based on the CSM, and a 95% UCL calculated from three replicates for each SU (Figure 6-1B). This spatial information shows that only a portion of the DU exceeds the AL.

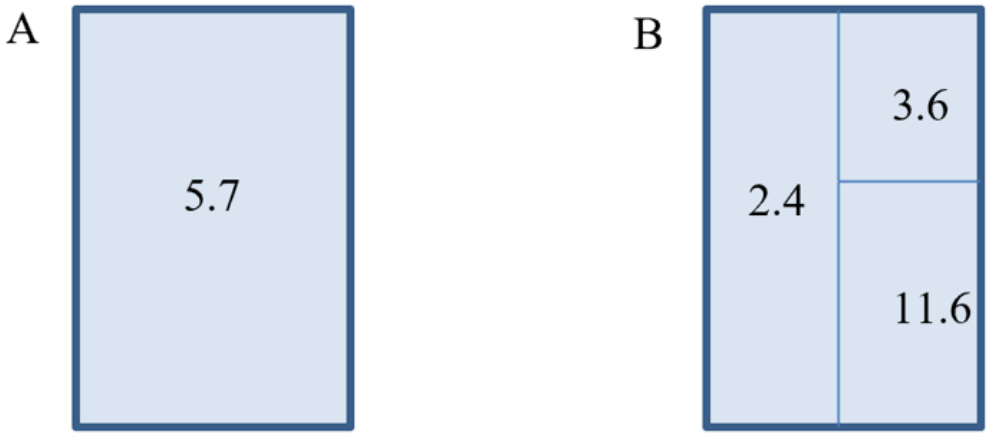


Figure 6-1. DU example.
Source: ITRC ISM Update Team, 2020

As noted above, a disadvantage of using a weighted 95% UCL calculated when a DU is divided into smaller SUs is that it may exceed the 95% UCL produced when increments were collected across the DU. For this example, the area with a 95% UCL of 2.4 is 40% of the DU, the area with a 95% UCL of 3.6 is 20% of the DU, and the area with a 95% UCL of 11.6 is 40% of the DU. Using the weighted means equations from earlier results in a DU 95% UCL of 6.3, which is larger than the 95% UCL of 5.7 that was obtained from the entire DU. Therefore, the decision to divide DUs into smaller SUs will depend on the possible distribution of contaminants at the site and whether concentrations are expected to be relatively homogenous or vary greatly with location. If similar variation is expected across the SUs, an alternative option involves collecting one of the SU or DUs in replicate, applying the measured RSD to each of the SU or DU results, and proceeding with the weighted 95% UCL calculation discussed above. Assuming that the CSM predicts or presumes that the heterogeneity within the measured replicates is applicable to the remaining SU or DUs, this option eliminates the overestimation of the 95% UCL. Section 3.3.2 gives additional details on various possibilities to structure the statistical analysis of DUs with smaller SUs.

6.2.4 Differences between discrete and ISM 95% UCLs

Both ISM and discrete sampling represent a mean concentration for a volume of soil, with discrete samples representing the mean concentration in the volume of soil removed for a single grab sample. This concentration (by itself) does not provide any information about the mean concentration of the DU. Because discrete samples represent only a small volume of soil, multiple discrete samples are needed to estimate an average concentration for the DU. ISM samples are structured composite samples that, when taken correctly, provide an estimate of the mean concentration for the entire DU. Although 95% UCLs can be computed from both discrete and ISM datasets, the two methodologies would produce very different distributions when applied to the same DU (see Section 2.4.1 and Figure 2.8). This is because they represent different properties of the population that is sampled in the DU. While the arithmetic mean of the populations represented by the sampling distributions is the same, the resulting sampling distributions will not be. Efforts to adjust the data and/or summary statistics so that the distributions match are unlikely to fully address the key factors that contribute to their differences (such as differences in sample support and laboratory procedures).

6.2.5 Use of discrete historical data

Discrete sample results (representing individual site measurements) are not directly comparable to ISM sample results, which represent mean estimates, so resampling the discrete data to simulate hypothetical ISM results will not address the disparity between discrete and ISM sampling protocols. Resampling merely provides a means of generating statistics for "composite discrete" samples rather than structured composite ISM samples. Since each ISM sample is a representative mean concentration, the SD of the ISM replicates is similar to the SE. Stated in mathematical terms, the proposed approach is similar in concept to dividing the sample SD (from discrete sampling) by the square root of the sample size n in order to approximate the SD of the ISM replicates (that is, the SE of the population). Because the variability of ISM replicates is an estimate of the variability in mean concentrations, it is similar to a SE, as it includes variability within field and laboratory

measurements. Great care must be applied to understanding hypothetical ISM results as they may be biased high as a result of using the underlying discrete samples and could introduce new decision errors. For example, variability associated with the original discrete samples associated with sample mass may not have been quantified and may not be comparable to the mass of sample collected through an ISM sample.

This is a major limitation in the proposed approaches for mixing simulated ISM results (from historical discrete sampling) with real ISM results and casts doubt on the reliability of the approach for regulatory decision-making.

6.2.6 Mixtures of discrete and ISM data

There are no established methods for combining discrete and ISM data, but ISM and discrete data can both help inform decision-making at the site. Either sampling methodology can be used for site characterization, delineation, confirmation sampling, and comparison to background, although the type of sampling used may vary depending on project goals, size of the site, funding, and other variables. The use of both ISM and discrete data at a site often provides a more complete picture of the nature and extent of contamination than either sampling method alone. For example, historical discrete sampling may be used to inform the delineation of SUs at a site, and because heterogeneity is higher with discrete data, ISM can then be used to obtain a better estimate of the mean concentration across the site. Although these two types of sampling are complementary, they are not equivalent. If combining discrete and ISM samples is deemed absolutely necessary, it should be done carefully and with the help of a trained statistician familiar with ISM site investigation methods.

6.3 Evaluation of Statistical Assumptions

Due to the low sample sizes typical of ISM sampling, many statistical tests designed to evaluate whether data meet assumptions may not be appropriate or may have low power to draw conclusions. A statistician should be consulted to evaluate the appropriateness of various tests.

After identifying the desired method of statistical analysis, it is important to check the assumptions of that method before analyzing data. Common assumptions include independence and normality in the data, and unbiased samples collected using a valid statistical design. Statistical sampling should be addressed in the sampling design in Section 3. The placement of the increments within an ISM sample should generally be selected randomly, but if there are more SUs that need to be sampled, the selection of those SUs should also be done randomly.

6.3.1 Independence

Most analytical methods will require field replicates to be independent – in other words, the value of one ISM result in a particular DU should not have any effect on ISM results in the same DU. Due to the nature of ISM in that increments are collected spatially across the entire volume or surface area, spatial trends within an SU should not cause a violation of the assumption of independence. This is because each composite has more spatial coverage than a discrete sample and has the potential to cover both high and low-concentration areas within a particular SU with a spatial trend.

In the case where a DU consists of more than one SU, the evaluation of spatial scale as it relates to independence of field replicates becomes more complicated. ISM requires the assumption that the SUs are exchangeable in order to meet the assumption of independence. Consider the case of a DU consisting of four SUs. If there is a spatial trend across SUs such that some SUs are more similar to each other than others, then ISM samples collected from each SU would not be exchangeable from one another, and the independence assumption would be violated. The consequence is that the variation will be underestimated. In this case, the groups of SUs that are exchangeable should be analyzed separately and only combined in an additional step to obtain the results for the entire DU (see the example in Section 2.5.3). If, however, a DU is divided into four equally-sized SUs, and each SU is separately sampled, a reasonable estimate of the DU mean should be obtained.

Another common assumption in analytical methods is that laboratory replicates are independent from one another. It is important that, in sample processing, the sample is as homogeneous as possible, such that each subsample in the volume of soil can be assumed to be exchangeable. Refer to Section 5 for a discussion of sample processing.

6.3.2 Normality

Another assumption of many commonly used methods is that the data are normally distributed. The CLT states that the

distribution of the mean is normally distributed (see Section 3.2.5), but in ISM, physical averaging based on increased sample volume is done instead of statistical averaging. In cases of extremely skewed datasets, such as extremely variable PCB or dioxin contamination, the CLT may not apply, but it should apply to the majority of ISM datasets. Because of small sample sizes in ISM data, a formal or statistical evaluation of normality can be difficult, so conceptual knowledge of the site can often be helpful. In general, unless there is a reason to suggest highly skewed data, the assumption of normality is probably appropriate, as shown in Section 3.2, Figure 3-17.

In cases with sufficient sample sizes, the best way to evaluate the assumption of normality is to simply look at the data visually, such as with a histogram. Many researchers recommend formally testing for normality (the Shapiro-Wilk test is an option), but this is problematic with fewer than eight samples. These tests require evidence to reject normality, rather than evidence to confirm normality. ISM data usually will not have sufficient samples to reject normality regardless of the true distributional form. If such tests are used, they should be only part of a larger evaluation of normality that also includes visualizations and conceptual knowledge.

Although assumptions of particular methods should always be checked, it is often difficult to be truly confident that they are met, particularly when it comes to normality of small datasets. Fortunately, most methods (such as the 95% t-UCL and Chebyshev 95% UCL) are robust to small deviations in normality, meaning these methods are still effective and provide the desired level of coverage and accuracy, even if the data are not quite normal.

In studies comparing ISM samples from one site to another, or a site to background, there is often the additional assumption of equal variance in ISM results – in other words, the spread of the values obtained from each site should be similar. Much like the assumption of normality, this can only be evaluated with sufficient sample sizes. For the small sample sizes typical of ISM sampling, reasonable professional judgment and site knowledge may be used to determine whether there is a reason for differences in variance between two sites or datasets. For larger sample sizes (typically at least seven or eight), the assumption of equal variances is best evaluated by visually comparing the data using figures such as histograms or boxplots. In addition, the estimates for SD between site and background datasets may be compared. A formal test for equal variance can be conducted (such as Levene's test), but sample sizes for ISM data are rarely large enough to make a conclusive determination on whether the assumption is met. In reality, site concentrations are expected to be more variable than background concentrations if contamination exists. Common background tests such as the two-sample *t*-test are robust to mild deviations from homogeneity of variances, but for severe differences in variances, other statistical tests may be implemented, such as Permutation tests.

6.4 Decision Endpoints

Each project and corresponding CSM and DQOs will determine the intended data use. Data quality needs for decision-making should be carefully planned by the project team and documented in the planning documents, which should also provide the "if-then" alternatives to be taken in the event that project QC goals are not met. This is particularly critical for complex and/or large-scale ISM projects such as those commonly seen.

Initially, the DQOs should provide the overall structure for data evaluation and use – for example, if the purpose of the DUs is to determine if a specific chemical was over an AL.

In many cases, changes to the CSM or unanticipated larger-scale trends may be detected. In addition to straight comparisons of results to ALs, for instance, larger trends in chemical concentration gradients may become apparent that were not part of the DQO design.

6.4.1 Step 1: initial DU data evaluation

The initial evaluation should begin by evaluating the raw/parent results, that is, ISM sample values. Before conducting detailed reviews, replicate reviews, or statistics, begin with simple, large-scale evaluations that focus on answering the following questions:

- Do the concentrations meet the anticipated CSM? For example, are ISM values in the anticipated range?
- Did the laboratory detection limits meet the sensitivity needs documented in the DQOs? Are the data usable for the intended purpose?
- Does it appear there are anomalies or outliers? For example, are there contiguous DUs with very different concentrations or combinations of concentrations not anticipated relative to the CSM?
- Was information not previously anticipated now present? For example, during field sampling, were there any

new observations about soil types? Were some DUs not sampled because of obstacles or other physical limitations?

This initial evaluation can help determine if the overall objectives were met, or if the project would benefit from splitting existing DUs into smaller DUs in a subsequent sampling event to better understand chemical variability. Several examples of data comparison scenarios are provided in Sections 2.5.3 and 6.2.1.

6.4.2 Step 2: detailed evaluation of ISM results, replicate data and action levels, and CIs

Following the initial evaluation, each project will have its own objectives in terms of detailed evaluation of the results. Examples of the different types of possible detailed evaluation/data use include the following:

- comparison of a summary statistic (single ISM estimate of the mean, the mean of multiple ISM results, the 95% UCL of multiple results, and so on) to an AL
- comparison of results of a quantitative risk assessment that used a summary statistic (commonly a 95% UCL) to an acceptable risk range for carcinogens (such as a 1 × 10-6 excess cancer risk to 1 × 10-5 excess cancer risk) or to an acceptable hazard threshold for noncarcinogens
- comparison of site and background datasets
- combination of data across multiple DUs
- extrapolation of statistics across DUs

Statistical evaluations of field replicate results are used to determine how well the ISM sampling protocol captures the spatial heterogeneity of the contaminant distribution (Table 3-3). Project planning steps should thus include specific metrics to evaluate the acceptable RSD, CIs, or other evaluating factors to be applied to field replicates. The following provides an example of how to evaluate CIs as an example of field replicate evaluation and confidence in decision-making, as demonstrated in Figure 6-2:

- If the parent concentrations are either well below or well above the ALs (Scenarios 1 and 4), then CIs may not impact project decisions. For example, if the parent concentrations at a site are all below 1 mg/kg, and the AL is 100, then likely CIs will not impact the decisions, even if they are above the project RSD or CIs goals. Similarly, parent concentrations at 100 mg/kg compared to an AL of 1 mg/kg render CIs less important in terms of decision-making.
- When parent concentrations are near the AL, and the uncertainty overlaps the decision threshold (Scenarios 2 and 3), CIs become an important factor in determining confidence in decision-making. In this case, evaluation of CIs is critical in terms of overall data usability.

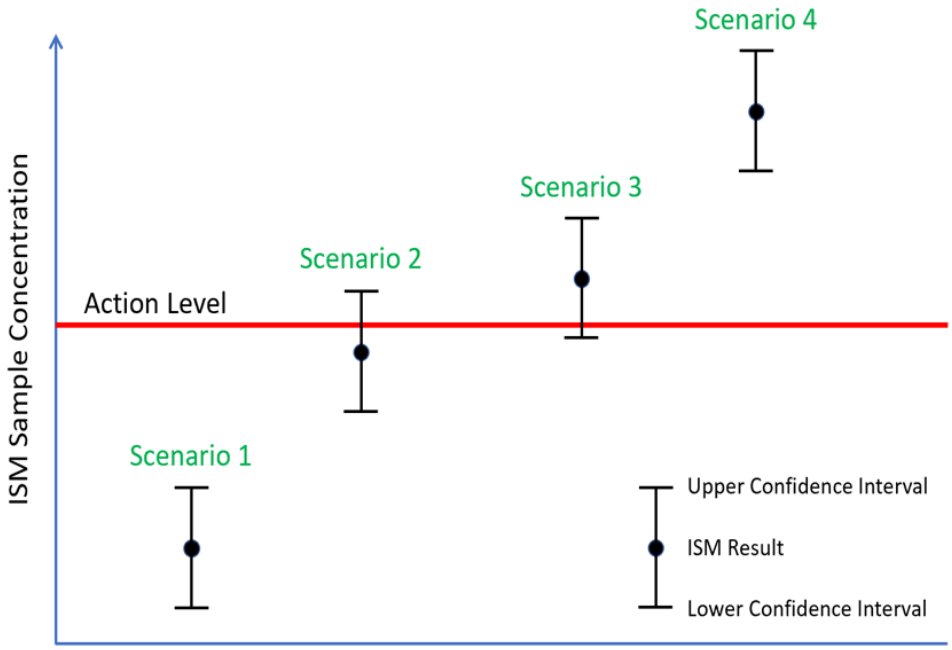


Figure 6-2. CI examples. Source: Jason Brodersen, Tetra Tech, Inc., 2019. Used with permission.
Source: Jason Brodersen, Tetra Tech, Inc., 2019. Used with Permission.

CIs are valuable when use of the parent value alone is not acceptable or desired. While there are several CI types, the 95% UCL of the replicate mean is the most common. CIs can be applied to a single DU with its own replicates, or the CV from those replicates can be applied to nearby or relevant DU concentrations within the project to derive CIs for DUs without replicates (see Section 3.3.1). Parent results with applied CIs can be applied to any of the detailed evaluation types identified above. Use of CIs is a project-by-project evaluation and should be considered thoroughly during planning phases. Some agencies require CIs, some do not – as a result, the evaluation of parent concentrations without CIs may be acceptable, given project DQOs.

6.4.3 Step 3: limitations of replicate results and sampling methodology

If the replicate results do not meet the project objectives, or if replicate results do not provide enough confidence to make site decisions as discussed in Section 6.4.2, then the project team must plan for alternatives that may include changing the initial ISM sampling design utilized – for example, if the DUs were too large, not enough increments were collected, or a larger sample mass is needed.

Regardless of the investigation objectives, it is important to review the replicate results and other factors that might contribute to confidence in the data (such as final number of increments or if the mass was obtained as intended) while making conclusions and findings.

6.4.4 Next steps

Use the results from the three previous steps to determine if the DQOs were met or if they can be optimized by additional ISM investigations. If the DQOs are not met, the methodology may be optimized in several manners, such as collecting more soil mass, collecting more increments, adding analytes to a DU, splitting a DU into multiple DUs, or adding new DUs.

Regulatory Acceptance

ISM is a sampling methodology that, when appropriately applied, will improve decision-making with less time and money, while providing representative, reproducible, and defensible data. ISM is a valuable methodology for consultants, regulators, and environmental professionals working on contaminated sites and can be used with most chemicals/contaminants and for many applications, including risk assessment, investigation, and confirmation sampling. Over the past decade, ISM has been more widely used and has been growing in acceptance, as shown in 2019 survey results. However, there are still regulatory barriers to overcome.

The following subsections present the state of practice for ISM by documenting its use since 2009 as well as practical guidance for working with or within a regulatory agency to gain consensus for using ISM in investigations, risk assessments, and confirmation sampling.

7.1 Comparison of Survey Results from 2009 to 2019

To determine the growth and acceptance of ISM use, ITRC surveyed states, territories, and consultants in 2009 and then again in 2019 to collect data on the current state of practice. A comparison of survey results shows there is increasing demand for ISM guidance and training across the U.S.

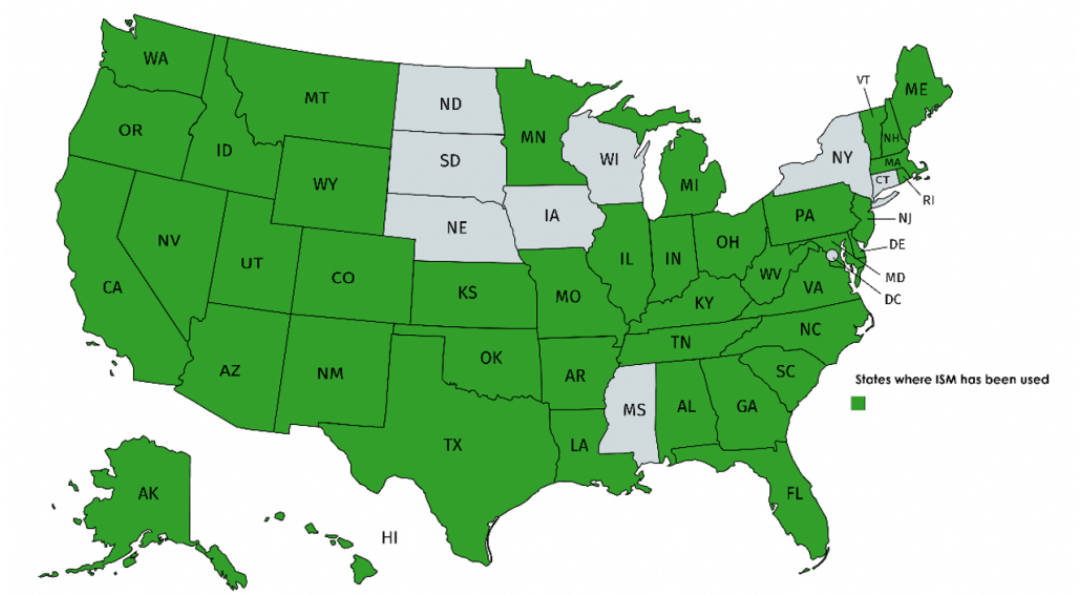


Figure 7-1. States where ISM has been used. Eighty-four percent of state respondents said "yes" when asked if they had used the methodology.
Source: ISM Update Team Survey, March 2019.

7.1.1 Demand for ISM guidance and training

There has been steady interest in incremental sampling use since ITRC published the 2012 ISM guidance document, as shown by the number of technical regulatory guidance document webpage views and the continued demand for internet-based training (IBT) (see Figure 7-2). ISM IBT began in 2012 and based on all subsequent annual IBT offerings, participation numbers and demand across the environmental and regulatory community remain high.

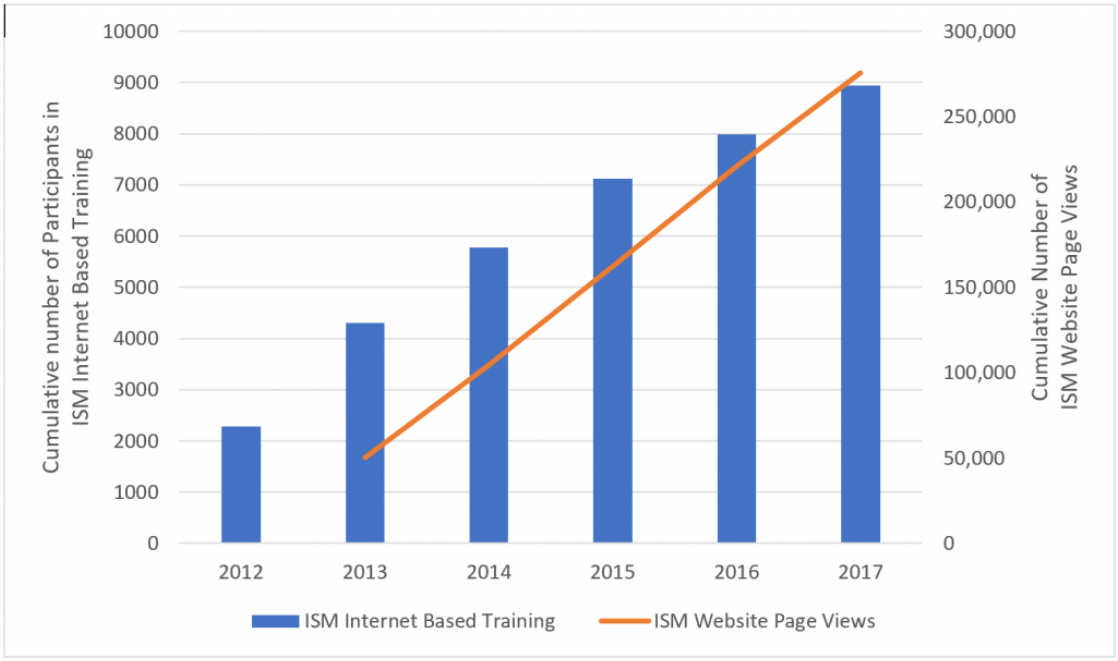


Figure 7-2. Participants in ISM training 2012-17 and ISM website statistics 2013-17.
Source: ITRC, 2019.

As mentioned, the 2012 ISM guidance is one of the most viewed documents on the ITRC webpage and has been downloaded in part or full 3,300 times in its first seven years (Figure 7-2), making it one of the top three ITRC documents downloaded in the last four years and the most downloaded document in 2017.

7.1.2 Regulatory use and acceptance of ISM

The primary respondents of the 2019 survey were state regulators and consultants, but federal regulators, state regulators, and consultants are also common users of the 2012 ISM guidance document and IBT (Figure 7-3).

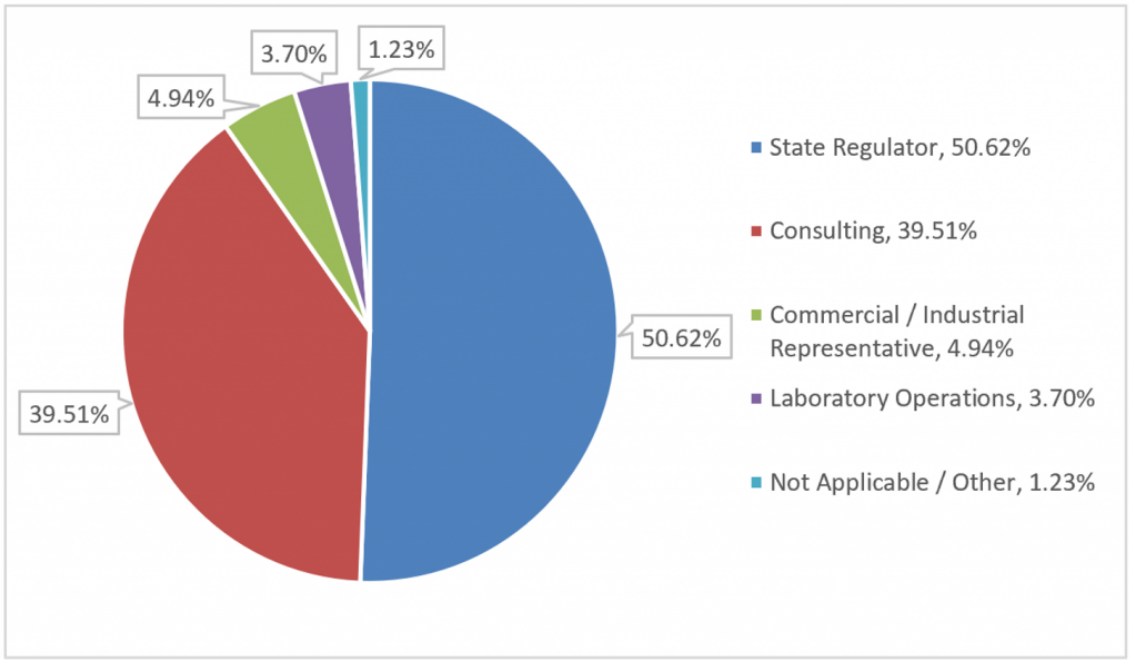


Figure 7-3. Primary job function of 2019 survey respondents.
Source: ISM Update Team Survey, March 2019.

Nearly 87% of survey respondents have used the 2012 ISM guidance to develop or review an incremental sampling WP, and almost 86% of survey respondents indicated that an updated ISM guidance for soils would be valuable to their organization. More people than ever know about ISM, which has helped increase its acceptance by regulators to correct common

misconceptions about the applicability of ISM across the country. While federal regulators were not surveyed, they serve as members of ITRC and regularly use ISM sampling. Many federal agencies have their own guidance and many more state agencies have developed guidance since 2009.

7.2 Current State of Practice

Based on recent survey results, ISM is gaining acceptance across the country in both privately funded cleanups as well as state and federal projects. The legal community is also beginning to recognize the power of representative and reproducible sampling strategies in their cases.

ITRC, USEPA, and several states' guidance documents, coupled with online and classroom trainings, are supporting the expansion of ISM knowledge and practice.

Guidance from several federal agencies has existed for many years. However, some states have supplemental guidance or independent regulations on ISM in lieu of federal regulations as noted below:

- Alaska Department of Environmental Conservation. 2019. "Field Sampling Guidance" (Conservation 2019).
- California Department of Toxic Substances Control. 2019. "Guidance for Screening Level Human Health Risk Assessments." Human Health Risk Assessment Note Number 4 (DTSC 2019).
- (DTSC 2018).
- (HDOH 2017b).

- Idaho Department of Environmental Quality. 2015. "Pile Sampling White Paper" (IDEQ 2015).
- Maine Department of Environmental Protection. 2015. "Standard Operating Procedure." SOP No. RWM-DR-015 (MEDEP 2015).
- Michigan Department of Environmental Quality. 2018. "Incremental Sampling Methodology and Applications." RRD-Resource Materials (MDEQ 2018).
- Ohio Environmental Protection Agency. 2016. "Incremental Sampling for Soils and Sediments." FSOP 2.1.3 (OEPA 2016).
- Washington Department of Ecology. 2019. "Sediment Cleanup User's Manual." Publication No. 12-09-057 (WDOE 2019).

In addition to ITRC IBT, there are also several commercial or private training courses available for interested parties.

7.2.1 Current use of ISM

ISM is most frequently used for surface soil, but it is also commonly used for subsurface soil, stockpiles, and sediment sampling. It is used with increasing frequency for defining the nature and extent of contamination, including brownfield redevelopment projects needed to define large areas (Figure 7-4).

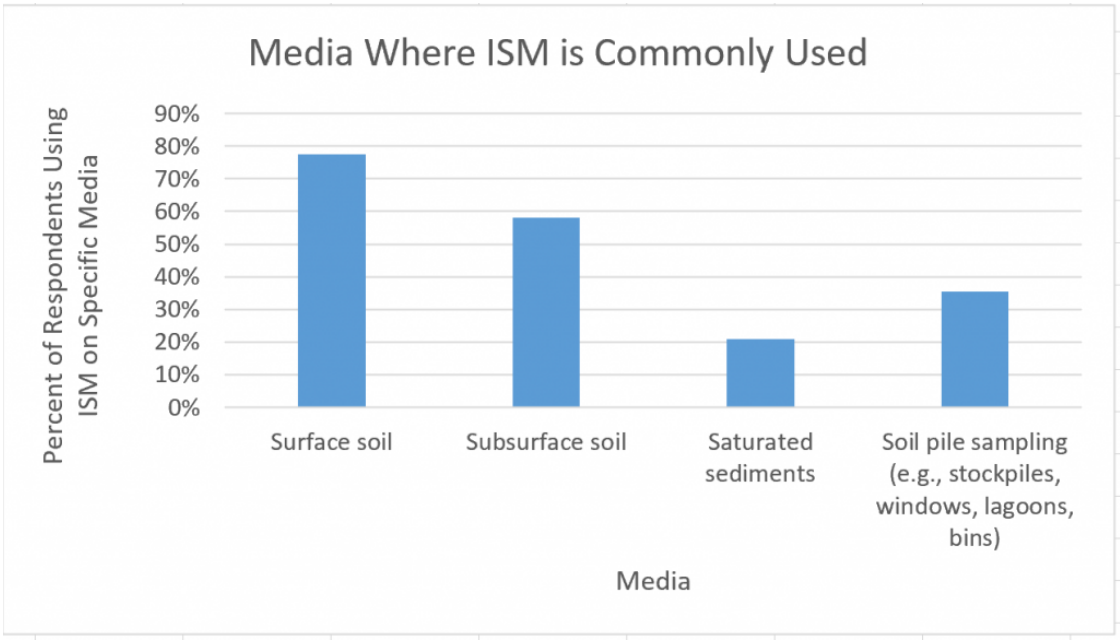


Figure 7-4. Media where ISM is commonly used.
Source: ISM Update Team Survey, March 2019.

ISM is also being used for many COCs, including metals, SVOCs, PCBs, TPHs, VOCs, and PFAS compounds as shown in Figure 7-5. Sections 3, 4, and 5 have further discussion on using ISM for VOCs and PFAS.

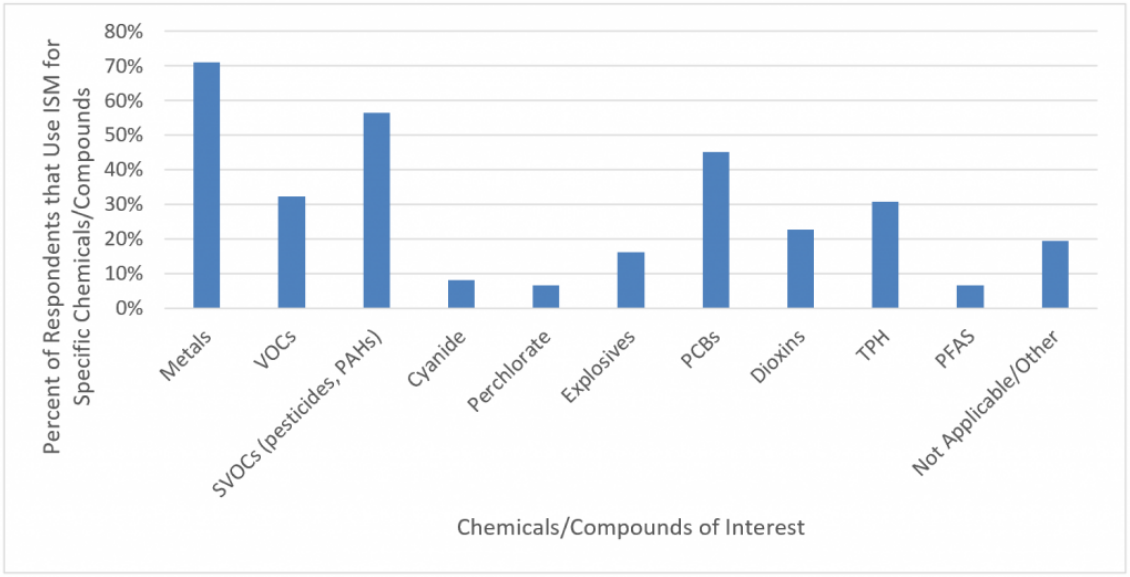


Figure 7-5. ISM's use for chemicals/compounds of interest.
Source: ISM Update Team Survey, March 2019.

7.2.2 Types of sites and uses of ISM

ISM can be used on residential, commercial/industrial, recreational, agricultural, and ecological sites. As such, ISM is most appropriate to make risk-based decisions on any volume of soil or surface area where a reproducible mean concentration is needed. Survey respondents indicated that they most commonly use ISM sampling in order by popularity to:

- determine if a DU met a regulatory cleanup criteria/cleanup level
- compare site data to published screening levels
- compare site data to site-specific human risk-based cleanup criteria
- compare site data to background concentrations
- determine baseline site screening levels
- determine if a remedial action met the site cleanup criteria (Figure 7-6)

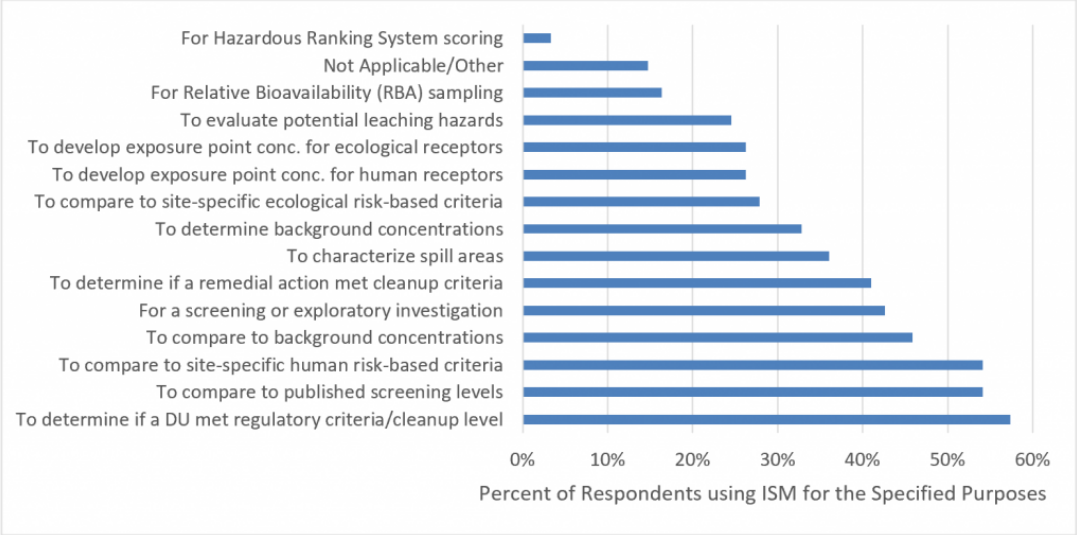


Figure 7-6. Current uses of ISM.
Source: ISM Update Team Survey, March 2019.

7.3 Factors Affecting Regulatory Acceptability

Regulatory acceptance for ISM use is primarily focused around certain concerns for accurate characterization of the nature and extent of contamination. There appears to be four common misconceptions:

- ISM cannot find areas of elevated concentrations.
- ISM will miss the maximum concentration.
- ISM 95% UCLs are unreliable.
- ISM cannot evaluate acute hazards.

Soil heterogeneities lead to increased data variability at spatial scales for which sample collection and analysis are performed (inches and grams), but we make decisions on spatial scales of cubic yards, acres, and tons.

Finally, there are barriers to ISM use related to programmatic or administrative requirements, including provisions for specific types of discrete samples, prohibitions on the use of any compositing techniques, or a lack of laboratory certifications or available facilities. Where there are statutory requirements that are incompatible with ISM, advocates need to work with state agencies over time to influence the development of updated regulations. In cases where the prohibitions are guidance (not regulation) or related to other programs at the agency, ISM practitioners may be able to advocate for ISM application.

7.3.1 ISM can find areas of elevated concentrations

- There is a common misconception that ISM cannot be used to define the nature and extent of contamination or find areas of elevated concentrations. This misconception is primarily based on historic composite sampling done inappropriately. ISM can in fact find areas of elevated concentrations when those areas' size and concentration are defined upfront, and the project teams focus on systematic planning, the DQO process, sample populations, and the delineation of DUs (Crumbling 2014, Hadley and Mueller 2012). To define the size and concentration of a significant small area of elevated contamination, an Excel spreadsheet tool can be used if the critical condition of a mature CSM is met. For an example and more details on this concept, see (Crumbling 2014). As presented in Section 3.1.5.2, statistically based sampling designs can be developed to determine whether localized areas of higher soil concentrations exist, even if the locations of such subareas within a larger site are unknown. The spacing of increments (and thus the number of increments needed to fill the DU's area) can be set to have a desired statistical probability of increments being collected from within an area of defined size for incorporation into the field sample. In this case, if the size of a potential subarea of elevated concentrations is specified, sampling can be conducted to determine whether one or more such areas exist within a DU with an objective degree of confidence and scientific defensibility. A free software program developed by PNNL called VSP is available to determine the increment spacing for the DU grid so as not to miss sampling from a significant small

area of elevated concentrations within the DU. Additional information on VSP, as well as important assumptions and limitations for any statistical tools and those specific to VSP, are included in Section 3.1.5.2. Practitioners of ISM are cautioned, however, that typical SUs are going to be much smaller for these purposes, and there may be a need to assign more SUs, depending on the site. Therefore, costs should be carefully considered.

- The perception that ISM will obscure areas of elevated concentrations is untrue and has not been borne out in case studies. Properly planned and conducted ISM programs have been found to actually produce higher estimates of mean concentrations than traditional grab samples, in addition to producing a more representative estimate than a grab sample of the overall concentration in any collected sample (Brewer, Peard, and Heskett 2016).

7.3.2 ISM will not miss the maximum concentration

There is also a common misconception among the regulatory community that ISM will obscure, miss, or "dilute out" small areas with elevated concentrations. However, most regulatory decisions are based on the upper bound estimate of the mean concentrations of contaminants or 95% UCL, not the maximum soil concentration. In actuality, as discussed in Section 2.5.2, Section 2.5.3.2, Section 3.1, Section 3.2, and Section 8.3.2, because ISM has superior spatial coverage, there is a higher probability that the resulting DU mean will capture the effects of areas with higher concentrations. The dilution misconception is a question of scale and highly depends on sample support (volume of soil collected) both from the field and in the laboratory subsampling process. As discussed in Section 2, at the microscopic level, the maximum soil concentration for any chemical contaminant is always 100%, and the minimum is always 0%. The smaller the sample support, the less representative the sample it. Does that 1 g of soil running through the laboratory instrument truly represent the 3D volume of soil collected and awaiting a decision? Depending on the CSM and nature of the release, there is a high probability that a discrete dataset may not represent the true mean conditions in the field. In the end, both discrete and ISM samples (assuming the same sample preparation, subsampling, and analysis steps are utilized for both types of sampling) represent mean concentrations in each respective sample. Instead of just the 10-g discrete sample location, ISM sample results provide better estimates of the true population mean of the whole DU because of the increased spatial coverage, increased sample support, and rigorous sample processing (see Section 8.3.2). A spreadsheet tool for establishing the size and concentration of a significant small area of elevated contamination can be used if the critical condition of a mature CSM is met. A freeware software program called VSP is available to determine the increment spacing for the DU grid so as not to miss sampling from a significant small area of elevated concentrations within the DU. For more details on this concept and the tools available, see the White Paper (Crumbling 2014) and Section 3.1.5.2.

Concerns about exposure to higher concentration areas in a DU can often occur. Proper utilization of ISM is simply a matter of proper sample planning. For example, in an ecological risk assessment, a high-value receptor (an endangered species, fragile ecosystem, protected waters, and so on) may have a very small home range. This issue can be addressed in the DQO process (see Section 3.3).

7.3.3 ISM 95% UCLs are reliable

A common misconception that affects ISM use acceptance is related to generating statistical estimates of EPCs. Some regulators or stakeholders may believe that a single set of three replicate samples cannot be used to calculate a 95% UCL, but, in fact, ISM is a strongly controlled sampling regime that produces highly reliable 95% UCL estimates (see Section 3.2.4) and can be used for risk assessment or, as appropriate, comparison to regulatory values. This is in contrast to many discrete sampling methods, where statistical evaluations cannot be completed, are ignored, or show poor reliability. The use of a simple RSD analysis rule set ahead of time in the DQO process can reduce disagreement(s) and statistical uncertainties.

7.3.4 ISM can evaluate acute hazards

Although ISM is typically used to evaluate chronic risk, it has also been used to identify acute hazard source soils that the limited spatial coverage of discrete sampling had missed (Walsh et al. 1997). In "Composite Sampling of Sediments Contaminated with White Phosphorus," by M.E. Walsh et al., CRREL Report 97-30. ISM was used to identify portions of a study area that would have likely resulted in acute ecological risk.

7.3.5 Regulatory guidance limitations

As mentioned earlier, there are barriers to ISM use related to programmatic or administrative requirements and/or guidance.

These include provisions for specific types of discrete samples, prohibitions on the use of any compositing techniques, or a lack of laboratory certifications or available facilities.

In several known instances, ISM data are not allowed by rule and/or statute, but there may also be instances where the state/territory regulatory officials do not approve a sampling plan that uses ISM for other reasons. For example, 40 CFR § 268.40 requires that “no portion of the waste may exceed the applicable treatment standard, otherwise, there is evidence that the standard is not met,” where a portion is considered to be a discrete sample. For this reason, ISM cannot be used in RCRA hazardous waste disposal decisions. Moreover, many state petroleum leaking underground storage tank (LUST) programs specify discrete sampling and will not accept ISM samples.

Where there are statutory requirements that are incompatible with ISM, advocates need to work with state agencies over time to influence the development of updated regulations. In cases where the prohibitions are guidance (not regulation) or related to other programs at the agency, ISM practitioners may be able to advocate for ISM application.

7.4 Benefits to Foster Regulatory Acceptance of ISM

Incremental sampling is a powerful tool for making sound decisions about soil contamination. Optimal decisions need to be based on data that are representative, reproducible, and defensible. Soil heterogeneity creates challenges for interpreting analytical data from discrete sampling, but these challenges can be diminished by using ISM, which provides a representative sample and an average concentration over a defined soil mass (or DU), making the ISM data reproducible and defensible.

Regulators want data to answer these common questions:

- What volume of soil in the field was the sample(s) intended to represent?
- Were the samples collected in such a way so that there is a good chance they actually do represent field volume?
- What evidence is there that the representativeness goal was achieved?
- Was the sample analyzed in a way that its representativeness was maintained (meaning, was subsampling variability controlled)?
- Does the QC data support the initial CSM and sampling design?
    - Good replication implies “yes” and increases confidence that data results represent the DU.
    - If replication is poor, the QC process may help identify if more increments are needed to manage heterogeneity, or if sieving or grinding need to be considered in consultation with the laboratory. ISM provides advantages over discrete sampling, for example, when there is not a reason to bias a sample toward a specific known release area. DUs help avoid the problems associated with typical composite sampling. ISM is appropriate any time you need to determine a mean concentration of a volume of soil. The DQO process helps answer two key questions: (1) why sample, and (2) what the result represents. It also provides reproducibility of samples upfront versus after getting the results, which is key to scientifically sound and defensible decisions.

Soil heterogeneities lead to increased data variability at spatial scales for which sample collection and analysis are performed (inches and grams), but we make decisions on spatial scales of cubic yards, acres, and tons. ISM gives representative, reproducible, defensible data at a lower cost than traditional samples with less contention after getting the results, which leads to better decision-making.

ISM for Risk Assessment

Section 8 addresses the key issues risk assessors should consider when planning for the use of ISM data for risk assessment and provides guidance for using ISM data in risk-based decision-making. The focus here is the use of ISM in human or ecological risk assessments, so this section presumes familiarity with the basic concepts and practice of risk assessment and includes references for federal and state guidance documents as well as ITRC guidance (ITRC 2015).

This section provides risk assessors and other users with a go-to resource for using ISM data in risk assessment. Key concepts discussed are as follows:

- systematic planning considerations for risk assessment
    - laboratory analyses
    - EPCs and DUs/EUs
    - nature and extent
- calculation and use of ISM data for EPCs
- benefits, planning, and application of ISM background data
- ISM sampling for post-remediation comparison to risk-based goals
- communicating ISM-based risk assessment results

8.1 Introduction

ISM is a technically sound sampling approach to collecting data for a scientifically defensible risk assessment designed to assist in risk-based decision-making. Specifically, it provides an accurate estimate of the true mean concentration with a limited number of samples. Most agencies accept the use a 95% UCL on mean concentration for comparison of soil or sediment concentrations to screening values or ALs and for use in baseline risk assessments. A minimum of three ISM replicates are necessary to calculate the 95% UCL.

As described throughout this document, the use of ISM samples to characterize the soil within a DU can provide higher-quality data and fewer decision errors than conventional low-density discrete or composite sampling designs. Because ISM yields an estimate of mean concentrations within a DU, it is important to understand the appropriate spatial scale that was specified or implied in the development of the risk-based screening levels.

As discussed in Section 3.1, incorporating a risk assessment strategy into a project should include developing a CSM developed during systematic planning. The risk assessment strategy should also be designed to inform risk management. Communication with those involved in or affected by the risk assessment should start in the planning stage of the project as well. The regulatory jurisdiction overseeing the risk assessment should always be consulted in the study design for ISM sample collection. If the study goal is to complete a quantitative risk assessment, then the strategy for using ISM data should consider the quantity and quality of the data required for the risk assessment, as described in Section 3.3.2 and Section 6.1.3. Important attributes of a CSM related to ISM sampling strategy for risk assessment include the following:

- exposure scenarios
    - receptors (human and/or ecological, current and potential future land use)
    - pathways (direct contact – such as soil ingestion and dermal absorption, transport – such as wind dispersion, surface runoff, infiltration, and uptake by plants and animals)
- exposure media (soil/sediment)
    - area (source characterization, extent of contamination, and the exposure characteristics of potential receptors' activity patterns)
    - depth (activity pattern characteristics of ecological receptors or human soil disturbance activities, and the extent of contamination)

A key concept in risk assessment is the exposure area. Risk assessment generally requires estimating long-term (chronic) exposures, and on occasion, short-term (acute) exposures. A goal of the data collection plan for a risk assessment is to estimate exposure-based chemical concentrations in exposure media based on current or hypothetical future land use.

CSMs, as discussed in Section 3.1.2 of this document and in ITRC's risk guidance (ITRC 2015), are essential to the SPP and for completing well-designed risk assessments. The CSM describes the relationship between potential chemical sources,

media, release mechanisms, fate and transport pathways, exposure pathways, exposure media, exposure routes, and current and future receptor groups. The CSM presents the current understanding of the project area but should be reevaluated and updated as new information is collected throughout the lifecycle of the project. The CSM helps identify data gaps and focus data collection efforts. Various styles of CSM can be useful, including narrative explanations supplemented by one or more figures (pictorial or schematic flow chart) depicting current and potential future site land uses. The ISM sampling strategy should provide inputs for risk assessment study goals.

8.2 Systematic Planning for ISM Data Use in Risk Assessment

Sample design input for risk assessment study goals should provide sufficient data to evaluate all the media each exposure group (such as trespasser, resident, or terrestrial receptor) might encounter. ISM sample results can be collected to provide EPCs for soil or sediment exposure media.

Environmental data must be of the appropriate type, quantity, and quality to manage and evaluate uncertainty so that defensible decisions can be made. To ensure that the data obtained during environmental investigations are adequate for their intended purposes, it is strongly recommended that data collection activities be planned and developed through an SPP, as discussed in Section 3.1.

As described throughout this document, the use of ISM to characterize the soil within a DU can provide higher-quality data and fewer decision errors than conventional low-density discrete or composite sampling designs. In combination with well-considered investigation objectives as well as DU and SU designations, ISM samples can reduce the need for additional sample collection, increase the certainty of decisions, and reduce the time and money required to complete environmental projects. Although a project team may have an ISM strategy in mind during initial planning, a number of sampling and analysis options should be considered, and the sampling strategy selected should be an outcome of the SPP. For risk assessment, the sampling strategy must meet the input needs for the planned risk assessment.

Estimates of mean concentrations in environmental media are generally the appropriate statistic to compare to ALs and to use in risk assessments. ISM provides an estimate of the mean contaminant concentration in a defined volume (area and depth) of soil. An exposure area (also called an EU for ISM risk-based DU) is a geographic area over which a receptor is reasonably assumed to move randomly and is therefore equally likely to contact an environmental medium (for example, soil) at all locations. In addition to assumed activity patterns of receptor behavior, the boundaries of an EU can also be defined to account for knowledge of historical activity and the nature and extent of chemicals in environmental media (USEPA 1989a).

Several crucial decisions must be made when ISM projects are planned. As discussed in Section 3.1, sample design such as the number and size of DUs, the number of replicate ISM samples, and the number of increments for each ISM sample are guided by the problem formulation and study goals. When ISM sample design is chosen as an information input, the DU dimensions should be designed to address a specific study goal or multiple goals. Different goals may require different kinds of inputs. Risk assessments may necessitate multiple sampling objectives and strategies, such as when the CSM shows pathways for multiple receptors or exposure scenarios, which may require sampling EUs of different spatial scales based on different receptors' human and/or ecological activity patterns. The dimensions of each EU must be assessed for appropriateness of the exposure duration (long-term chronic exposures and/or short-term acute exposures), as well as specificity for each receptor under each current and potential future land use. The EU dimensions will depend on the appropriate area and depth pertaining to a receptor's potential exposure. As described in Section 8.2.2.1, the dimensions of EUs for human health and ecological assessments may differ.

8.2.1 Risk assessment considerations for ISM laboratory sample preparation

Laboratory processing of ISM samples prior to analyses for non-volatile compounds usually includes sample sieving and could include grinding of a sample to improve laboratory precision. These topics, as well as consideration of appropriate sample digestion/extraction methods, are discussed in the context of project planning in Section 3.1.5.3 and addressed in more detail in Section 5.2.

The representativeness of the sample after laboratory processing should be considered by the risk assessor with respect to each receptor evaluated in the risk assessment. A thorough discussion of this topic is beyond the scope of this document, but several key points are noted here:

- The surface area of particles per unit volume of soil or sediment is inversely related to particle diameter, and

greater contaminant concentrations can sometimes be observed on fine particles relative to coarse particles.
- The appropriate particle size fraction for evaluating exposures in human and ecological risk assessment can vary as a function of the receptor and exposure route.
- Grinding of soil or sediment samples prior to analysis might produce analytical results that are not representative of environmental exposure conditions.
- Sample extraction methods should be representative of conditions associated with the exposure route (such as ingestion, inhalation, or dermal absorption) evaluated in the risk assessment.

8.2.2 DUs, EUs, and SUs

As described earlier in this guidance, an ISM DU is the smallest (horizontal) area and associated (vertical) depth of soil for which a decision will be made. An SU is either equal in size to – or is a subdivision of – the DU. An SU is comprised of at least 30 increments of soil collected to estimate the mean chemical concentrations for an area and depth of soil or sediment. For risk assessment, EUs represent the area over which a receptor could be exposed for the relevant exposure period corresponding to the particular receptor scenario. The relationship between DUs, EUs, and SUs, discussed in detail in Section 3.1, is described here in the context of risk assessment.

Environmental decisions are often based on the predicted risks from exposure to estimated mean concentrations of contaminants in a volume of soil. In some cases, a decision for additional investigation or remedial action might be made based on a comparison of ISM sample results to benchmark or screening levels, which are often risk-based. In other investigations, the estimate of the mean contaminant concentration provided by ISM samples might be used to calculate risk for human or ecological receptors. ISM results may also be used to estimate background concentrations, to assess the boundaries of source areas, or to evaluate the success of remedial activities. In each case, specifying the dimensions of EUs, SUs, and DUs is a critical component of the sampling design.

8.2.2.1 EUs

EU boundaries are based on exposure assumptions concerning the area and depth of soil where a receptor may be exposed over time. As discussed in Section 3.1, the primary types of DUs are those that are based on the known or expected locations and dimensions of releases (source area DUs or N&E DUs) and those based on the known or expected locations and dimensions of areas within which human or ecological receptors are randomly exposed (EUs). This discussion focuses on applying ISM to EUs.

During systematic planning to support risk assessment for contaminated soil and sediment, a primary question is, “Over what area and depth do samples need to be collected to provide data to represent the potential exposure of a receptor?” EUs based on exposure areas are a key tool in risk assessments and risk-based decision-making. For the purposes of this document, an *exposure area* is an area where human or ecological receptors could come into contact with contaminants in soil on a regular basis (refer to “Risk Assessment Guidance for Superfund, Vol. II” (USEPA 1989b)). Examples include residential yards, schoolyards, playgrounds, gardens, outdoor areas of commercial/industrial properties, or areas designated as exposure areas through other means (such as state laws).

EUs applicable to human receptors may not be readily applicable to ecological receptors, so when sites are evaluated for both human health and ecological receptors, multiple spatial scales may need to be considered for sampling. For example, EUs for human health evaluations may correspond with individual residential properties, while EUs established to address ecological receptors may correspond to the home range of an individual receptor or a population of receptors. This is a good example of why both the CSM and the second step in systematic planning (see Section 3.1.3, “What Types of Additional Information Do We Need?”) are very important – information needs for ecological and human exposure assessment can differ greatly. Other non-risk-based study goals might require different DU dimensions for purposes such as characterizing spatial patterns of contamination or providing data to evaluate remedial alternatives.

The primary use of data representing contaminant concentrations in an exposure area is to assess human or ecological risk. When study objectives involve risk assessment, the EU should ideally be based on the area where exposure is known or anticipated to occur. The size and placement of exposure areas depend on current use or potential future use of the site. However, if there are suspected areas of unique or elevated concentrations within an EU, it may be important to break the EU into two or more parts (see Example 2B in Section 3.1.6.2). Since risk assessment generally assumes long-term or chronic exposure, an underlying assumption of an EU is that the receptor spends an equal amount of time over all portions of the EU for that exposure period (meaning the receptor is randomly exposed across the entire EU).

In some situations, a standard-sized EU might not reflect the known or suspected movement of the receptor – that is, the human or ecological receptor prefers some areas over others. If there are subareas of elevated concentrations within an EU, then potential risks could be underestimated. In this case, the CSM and the resulting sampling plan should consider the suspected or actual movement of the receptor by the use of smaller EUs within which the receptor can be expected to move randomly. Swing sets and sandboxes in residential yards are the classic example for human exposure – in this scenario, a child is expected to spend more time on and around such play areas than in the remainder of the yard. Movement within such smaller areas is expected to result in equal exposure to all parts of the area, and therefore it meets the definition of exposure area. This concept is illustrated in Section 3.1.6.2 Example 2B.

An EU could be based on current land use or potential future land use. Site-specific information should be incorporated into the CSM so that exposure media and exposure dimension assumptions are clear. For example, if future residential exposure is going to be evaluated, then the EU should be designed and sampled at a scale consistent with the local residential property size, and perhaps address the depth of soil that would be excavated for a foundation and used to regrade the area (bringing subsurface soils to the surface). Potential regrading during future development of the property should also be considered in systematic planning of the sampling design to focus proper sample collection in the soil horizon (depths) that may be encountered by future receptors. Some states define the area and depth to consider for a default residential property in the absence of existing residential boundaries, but the uncertainty section of the risk assessment should discuss EU assumptions, such as those associated with a hypothetical future residential EU property size and placement.

The depth and area of EUs should be defined consistently with the exposure scenario under consideration. In many such scenarios, the first few inches or centimeters of the surface is the appropriate sampling interval (recreational receptor). However, the depth and area considered acceptable for evaluation of surface or subsurface soil for exposures varies among agencies (ITRC 2007a) and risk receptor scenarios. Evaluation of risks posed by future excavation and spreading of deeper contamination to the surface could require EU depths many feet below the ground surface. The need for multiple, different vertical EUs depends on the CSM and its expected contaminant distribution with depth as well as on current and potential future land uses.

As discussed in Section 3.2, the number of samples needed from an exposure area (including duplicate and blank samples) depends on variability in chemical concentrations within the exposure area and the level of precision and accuracy required for the project's risk-based decision-making (Hartmann et al. 1993.).

8.2.2.2 SUs

As Section 2 notes, SUs are subdivisions of (or equal to) DUs from which separate ISM samples are collected. The boundaries of an SU indicate the coverage of a single ISM sample, thus, SUs define the scale of ISM sampling, whereas DUs define the scale of the decision(s) based on that sampling. These definitions allow for the possibility that ISM samples from several SUs composing a DU can be used collectively to make a decision on that DU. It is sometimes possible to later redefine SUs as DUs if the resulting scale and number of replicates meets project objectives.

The mean concentration over the entire DU is typically the basis for decision-making. If the DU is properly sized for risk assessment, then it may be comprised solely of replicates of a single SU or several SUs. Estimation of the DU concentration mean from individual and replicate SU samples is discussed in more detail in Section 3.2.8 and Section 6. When ISM samples are collected across the entire DU (where the DU is a single SU), replicates offer information on variability in the mean estimate. Replicates from a single SU or DU do not, however, provide any information on the spatial patterns of concentrations within that single SU or DU. If this information is desirable, an appropriate sampling design is to divide the DU into multiple SUs (or further divide an SU into smaller SUs) and take one or more ISM samples from each newly formed SU. Section 3.1.6.2 Example 2B, Figure 3-11b illustrates this concept of utilizing SUs within an EU for a residential yard, play area, and potential lead-based paint surrounding a home. With this approach, ISM samples from the SUs are not true DU replicates in that they are providing estimates of the mean for different subunits within the DU. Individually, they estimate the mean of a subarea (the SU), and collectively they can be used to estimate the mean of the entire DU. However, collecting three replicates from some SUs within the DU can provide information on spatial variability and help calculate both field variability and DU mean concentration. Dividing DUs into SUs could be used to answer the environmental question, "Is there evidence that remediation of an SU in the DU could lower the risk presented by exposure across the DU?"

Results for individual SUs within a DU should generally not be used to make decisions because, by definition, such SUs are at a smaller scale than appropriate for a decision. This is especially true if sampling involves only one ISM sample per SU (meaning no replicates were collected). Taking only one ISM sample from an SU should generally be done with caution because a single sample will not provide any information about the variability in ISM replicates for a particular SU to assess

uncertainty in chemical concentrations within it. However, such an approach can be advantageous when sampling large areas. In this case, individual SU results are statistically treated in a manner identical to that of traditional discrete or composite samples. This application of SUs is discussed in more detail in Section 3.2.8.

SUs may also be used when there are multiple sampling objectives or sampling scales for a given volume of soil. For instance, when the same area is assessed for multiple receptors with exposure areas of different scales, then the various EUs will define the receptor-specific DUs and multiple SU sizes may be applied in that area to combine in appropriate horizontal and vertical dimensions for the various receptor-specific EUs at the site. For example, Receptor Scenario 1 may consist of a large DU = EU1 that is broken down into multiple SUs, where SU size is equivalent to EU2, which represents the DUs with a smaller spatial scale for Receptor Scenario 2.

8.2.3 Using ISM to evaluate the nature and extent of contamination

As described in Section 3.1.5, delineating the nature and extent of contamination is a common objective of environmental sampling in a remedial investigation. When considering the appropriate area (residential property, ecological receptor home range, and so on) and depths (surface, or 0 to 1 ft; subsurface, or 1 to 10 ft) where receptors may be exposed to soil contamination, the integration of information about the nature and extent of contamination is essential for the risk assessor.

To compile the CSM, understanding the nature and extent, as well as the fate and transport, of contamination is also crucial in determining both where receptors may be exposed to elevated levels of contamination and the complete exposure pathways for all current and potential future receptors. Both the horizontal and vertical extent of contamination must be understood to determine where and what analyses are required from what media for estimating exposures to receptors in the risk assessment process,. For example, for future land use and potential redevelopment where soil mixing may redistribute subsurface soils to shallower depths or the surface, data on the nature and extent of contaminants from deeper soil are needed to evaluate the potential risks from direct contact with soil. Similarly, characterizing the lateral extents are critical for determining potential receptors (for both current and future human and ecological receptors), particularly if the lateral extents go beyond the property boundary on the site to an area where land uses may be different. EU locations and dimensions (breadth and depth) are commonly informed by the nature and extent of contamination, so characterizing the nature and extent is an iterative process that should involve the entire project team so that they obtain the proper data for determining the concentrations of each chemical in each media that could potentially contribute to an unacceptable risk. Evaluating the nature and extent of contamination with ISM is not intrinsically different than doing so via discrete or traditional composite samples. Regardless of sampling technique, however, the sampling design derives from the CSM and the study questions in the SPP.

Sampling to characterize the nature and extent addresses a complementary study goal to the risk assessment study goal (which is to determine EPCs). Typically, soil samples are initially collected at biased locations based on site history or physical features to determine if contamination was released and if chemical concentrations might exceed risk-based or regulatory criteria. After the nature of the contamination is determined (that is, what chemicals have been released), delineation of the extent can be evaluated and used to inform other components of the investigation, such as the potential migration of contaminants (for example, soil contamination at depths merging with shallow groundwater flowing to a surface water body). This data is usually intentionally biased:

- The data will not likely be useful for determining average concentrations over an exposure area.
- The data may over- or underestimate the mean concentration in the area/volume of soil or sediment depending on the proportion of samples collected from the highest area versus the lowest concentration locations (such as in delineating the extent of the impacted area).

One advantage of ISM is that the data from SUs used to define nature and extent are not unduly biased by intense sampling from areas of highest concentrations within the source area or lowest concentrations at the perimeters, thus delineating the extents of contamination. As discussed in Section 8.2.2.1, the dimensions of an EU cannot always be known with certainty. Therefore, understanding the spatial distribution of contamination is important to provide confidence that potential exposures have not been underestimated. ISM results from multiple SUs can be used to estimate EPCs for different possible EUs, providing useful information for risk management. Use of ISM in this manner provides better coverage and a more representative estimate of mean concentrations in a DU, which reduces uncertainties in the risk assessment EPC.

8.2.3.1 Misconception regarding small areas of elevated concentrations.

As discussed in Section 2.5.2, Section 2.5.3.2, Section 3.1, Section 3.2, and Section 7.3, because ISM has superior spatial coverage, there is a higher probability that the resulting DU mean will capture the effects of areas with higher concentrations. Strategies are available for use with ISM to safeguard against missing significant small areas of elevated contaminant concentrations within a DU that can have the potential to change a sample concentration from below to above the decision threshold if they are captured in their proper spatial proportions by an ISM sample.

ISM sample results provide better estimates of the true population mean of the whole DU, and because of the increased spatial coverage, increased sample support, and rigorous sample processing, ISM can in fact find small areas of elevated contaminant concentrations when the sizes and concentrations of concern are defined upfront. See "***Evaluating the potential presence of subareas of elevated contaminant concentrations with ISM***" (within Section 3.1.5.2) to address these types of planning considerations. A DQO study goal could be to not miss significant small areas (horizontal and depth) of elevated contaminant concentrations above risk-based concentrations within an EU because the EU could be comprised of several SUs designed to meet the "small area" volume requirement. It is in the systematic planning phase that project teams must define and designate what concentration and what volume, surface area, or mass are significant to their decision-making. The size and concentration of a *significant* small area of elevated contamination can be established using an Excel spreadsheet tool if the critical condition of a mature CSM is met. For an example and more details on this concept, see the White Paper (Crumbling 2014).

Statistically based sampling designs can be developed to determine whether localized areas of higher soil concentrations exist, even if the locations of such subareas within a larger site are unknown. As discussed in more detail in Section 3.1.5.2, a free software program called VSP is available to determine increment spacing for the DU grid so as not to miss sampling from a significant small area of elevated concentrations within the DU. The spacing of increments (and thus the number of increments needed to fill the DU's area) can be set to have a desired statistical probability of increments being collected from within an area of defined size for incorporation into the field sample. In this case, if the size of a potential subarea of elevated concentration is specified, sampling can be conducted to determine whether one or more such areas exist within a DU with an objective degree of confidence and scientific defensibility (see Section 3.1.5.2). For more details on this concept and the VSP tool, see the see the White Paper (Crumbling 2014). Users are strongly encouraged to fully understand and consult the additional details on VSP designs as well as the inherent assumptions and limitations that are available in the VSP help files some of which are noted in Section 3.1.5.2.

Although somewhat subjective, not statistically rigorous, and less scientifically robust, increment spacing within a DU (or the sampling density) can provide some confidence that small areas of elevated contaminant concentration within a DU are not obscured by either increasing the number of replicates or increasing the number of increments per replicate. Nonetheless, this process may define the size of a small area of elevated concentration within a DU that is *observable* via ISM. As an example, consider a ¼-acre residential EU. If three replicates of 30 increments are collected, they correspond to three increments (one from each replicate) from an area of approximately 18 ft by 20 ft, which results in a sampling density of one increment from each 10 ft by 12 ft area. If 70 increments are collected instead, this translates to three increments from an approximately 12 ft by 13 ft area, or a sampling density of one increment for each 6.5 ft by 8 ft area. In these examples, the small areas of 10 ft by 12 ft or 6.5 ft by 8 ft of elevated concentrations would be represented in ISM samples of 30 or 70 increments, respectively, from the residential lot. If the area sizes and concentrations are significant, sufficient increment sampling density can guard against missing a small area of elevated contaminant concentration and underestimating the true DU concentration.

In addition to (or as an alternative), the SPP DQOs can be used to define the allowable variation in replicates that will be acceptable (the MQO), as discussed in Section 3.3.2.1 and Section 6.4.2. Reproducibility is a hallmark of a scientifically defensible study, and reproducibility among replicates can be used to uncover one or more small areas of elevated contamination within a DU. ISM includes QC procedures designed to measure overall sampling and analysis precision, including the collection of field replicates. High variability between replicates demonstrated by a high RSD is an indication of either a localized small area of elevated concentrations within the DU or unequal distribution of the COPC due to the nature of the source that resulted in the release (such as munitions) or the actual nature of the COPC (such as propensity to form nuggets or hydrophobicity). See Section 3.2.4.2 text and Table 3-3, which classify heterogeneity of increments in terms of low, medium, and high CV of replicates. If results for the replicates do not agree, one reason may be that the number of increments collected was not adequate to representatively include areas of non-random higher concentration scattered throughout the DU. If other causes of data variability can be ruled out by QC data, disagreement among field replicates is an indication that more increments may be needed to manage the heterogeneity caused by small areas of elevated

concentration. Examples in Section 3.2.5.2 demonstrate that large disagreement in replicate concentrations is a clear sign of extreme heterogeneity, most likely manifested as small areas of elevated concentration within the DU. Refinement of the study design to include smaller DUs and/or more increments will shed light on the cause and uncover whether the underlying reason is indeed a small area of elevated contamination within the original DU sampling design.

Examples that integrate the use of ISM sample data collected to investigate nature and extent, as well as support the estimation of EPCs for risk assessment, are included in Section 3.1.6 and Appendix A. Topics include ISM application to different types of releases, such as widespread theoretically homogeneous contamination (for example, agricultural field pesticide applications) and spatially heterogenous contamination from one or more potential source release areas.

8.3 EPCs from ISM Data for Human and Ecological Risk Assessment

The calculation of a receptor's average daily dose is based in part on the average concentration of a contaminant in an exposure medium, which means that the reliability of the chemical concentration data used to develop EPCs is very important. Samples collected by ISM provide reliable data with less uncertainty than those historically collected with discrete or composite sample methodology. Screening level risk assessments using ISM data are as valid as (and have less uncertainty than) those using the maximum concentration from discrete or other traditional sampling approaches.

There are various approaches to defining DUs, but the focus should be on ensuring that the data collected will aid in making the decision associated with the DU area. The approach selected should be consistent with the understanding of the site reflected in the CSM and should support the objectives of the investigation. Human health or ecological EUs – and the spatial distribution of contamination – should provide the basis for designating DUs for risk assessment.

8.3.1 Scenario-specific EUs

The CSM depicting current and potential future scenario-specific receptors is used to define scenario-specific EUs for ISM risk-based study questions. Assumed current and future human and ecological activity patterns associated with exposure media are the primary basis for defining the boundaries of a scenario-specific EU. The concentration of a soil contaminant an individual would be exposed to from long-term random exposure within an EU is, in fact, the average concentration. In other words, it is because a human or ecological receptor is assumed to contact soil across the EU in a random manner over time that we use the average to represent the exposure concentration. This concept can be extended to consideration of exposure to soil as a function of depth as well as area. If, for example, an exposure model states that the activities of humans (or burrowing animals) might reach a certain depth, then the average soil concentration from the ground surface to that depth is of interest. Practically, we rarely know with a high degree of confidence what (future) exposure patterns are going to be. As part of our sample design assumptions, we might state that humans could excavate soil to a depth corresponding to a basement, but we do not necessarily know they will, or what the exact location and volume of the excavation will be. As discussed in Section 8.2.2.1, for future exposure scenarios, the EU should be designed and sampled at a scale consistent for each of the receptors (such as the local residential property size for a future residential scenario). For future receptor scenarios, the depth of soil that would be excavated for a foundation and used to regrade the area (bringing subsurface soils to the surface) should also be considered. This is why it is important to consider both source areas (that is, the known or inferred spatial pattern of contamination) and exposure areas in developing DUs to meet information inputs for study goals in sampling designs (see Section 3.1). As described in Example 2 in Section 3.1.6, if there are specific areas that might receive higher use in an exposure area, such as a play area, that area could be evaluated as a smaller and separate SU or an N&E DU, particularly when localized contamination is suspected. Sections 3.1.6.2 and 3.1.6.3 provide Examples 2B and 3, respectively, for identifying EUs under different conditions and for different study goals. These examples address both human health and ecological risk assessments and integrate ISM sampling for investigations addressing source areas and EUs. The use of ISM to evaluate the spatial extent of contamination is also described in these examples.

8.3.2 Use of ISM replicate data in risk assessment

The number of ISM field replicates required for a scientifically defensible risk assessment is flexible and based on the DQOs established for the study questions in the SPP. A single ISM sample (singlet) does not provide a CI for assessing uncertainty in the mean or provide evidence that the common assumption of relatively low heterogeneity within the DU is met. A minimum of three replicates are needed from a DU to estimate a 95% UCL, and they may be required from some SUs even when a DU is comprised of multiple SUs, depending on study objectives.

In order to calculate a 95% UCL on a mean, at least three replicates are needed from the DU.

The objective of ISM is to provide a reliable estimate of the mean contaminant concentration in a DU, recognizing that any individual ISM sample may over- or underestimate the mean to some degree. This sampling error may be attributed to a variety of factors, as discussed in Sections 2.3 through 2.5. One objective of systematic planning for most sampling designs is to minimize the major sources of error in both the field and the laboratory. ISM standardized protocols minimize sampling errors, and replicates provide a measurement of the cumulative field and laboratory errors. In addition, replicates provide a measurement of the heterogeneity within the SU, with data from replicates used to assess data quality and CSM assumptions, as described in Section 6. In practice, the estimated variance is often viewed as an overall measure that includes the contribution of many sources of error. Section 3.2.4 provides an overview of the attributes of ISM samples that support robust estimates of a 95% UCL, including superior physical site coverage, relatively low variability, and a large sample mass.

Each ISM replicate provides an estimate of the true mean for that SU. As such, the distribution of ISM replicate results is related to but conceptually different from the distribution of discrete samples. The two approaches may share the same grand mean but can be expected to have different estimates of variance (see Section 4.2.1 and Figure 4-3 in (ITRC 2012)). For ISM, the mean of replicates is analogous to repeated trials of discrete sampling (the mean of the means, or grand mean), and the SD is analogous to the SE for the mean in discrete sampling. Even the most comprehensive sampling protocols will introduce some degree of sampling error, so it is possible that a single ISM replicate result can be well above or well below the true mean. The magnitude of the under- or overestimate depends on the overall heterogeneity of the underlying distribution, which increases as heterogeneity increases; this is why, ideally, at least three replicates of each ISM SU should be collected. As described in Section 3.2, both statistical simulation studies and case studies with ISM data support the need for three or more replicates.

8.3.2.1 One ISM result: pitfalls and high uncertainty.

For sites where there is a regulatory requirement to calculate a 95% UCL, at least three replicates should be collected within a DU. For sites where there is no regulatory requirement to calculate a 95% UCL, it is important to understand the potential for decision errors if a decision is to be informed by a single ISM result. Two critical components to a decision error are the likelihood of underestimating the mean and the magnitude of the underestimation, which correlates to the direction and magnitude of the uncertainty in the risk estimate. Section 3.2.5 discusses decision errors, and the 2006 G-4 document beginning on page 63 (USEPA 2006b) further describes estimating decision errors.

Section 6.2.1.1 provides a summary of cautions for drawing conclusions based on comparison of a single ISM result to an AL. Statistical information presented in Section 3.2 provides important concepts about each ISM result. As discussed in Section 6.2.1.1, a single ISM sample gives no information for estimating uncertainty in the mean concentration, which greatly limits the scientific defensibility of this approach. A single ISM result might support a risk-based decision when the estimated mean concentration is much greater than or much less than an AL. In this situation, the ISM sample might provide confirmation of what may have already been strongly suspected – that the DU clearly passes or fails. Obviously, as discussed in Section 6.2.1, uncertainty about making the right decision increases as the ISM sample result gets closer to the AL. The risk assessment uncertainty analysis should clearly discuss the large uncertainty if only one ISM sample is used in risk-based decision-making. Use of one ISM sample rather than three replicates for estimating risks has a very high level of uncertainty that may be inconsistent with many state or agency needs for risk-based decision-making. The uncertainty associated with making decisions with only one ISM sample may make this approach unacceptable to regulators for use in risk assessment.

8.3.3 Calculation of EPCs and 95% UCLs with ISM data

This section discusses the various EPC estimates that may be used in a risk assessment conducted with ISM data. Topics include the applicability of ISM EPCs for risk assessment use, similarities and differences between ISM and discrete EPCs, considerations for ensuring adequate spatial coverage for the EPC, the importance of calculating statistically sound EPCs with ISM, ensuring that the EPC (95% UCL, see Section 3.2.4) calculated from the ISM sample means encompasses the “true mean” of the soil population (see Section 3.2.4 on coverage of the 95% UCL), and circumstances when multiple small DU results may be combined to obtain an EPC for a larger DU. An example of using “weighted means” with multiple SUs is also discussed (see also Section 6.2.2).

The EPC is intended to be a conservative estimate of the mean concentration that the receptor is in contact with over daily exposure (see Section 3.2 and Section 3.3.2). It implies that the receptor's exposure is assumed to be spatially equal in all areas of the EU throughout the exposure period. Project objectives may specify that the estimate of the mean concentration provided by ISM sampling must be health protective, meaning there is a low chance of underestimating the actual mean concentration within the EU. Recall that a 95% UCL of the arithmetic mean is an upper bound estimate of the mean concentration in a given environmental medium. Use of the 95% UCL is health protective because there is only a 5% chance that the mean is underestimated, assuming that appropriate statistical and sampling methods are used such that the coverage is not less than a 95% probability of encompassing the true mean. It is important to recognize that the likelihood of underestimating the mean from any sampling method (discrete, composite, or ISM) increases as the distribution of concentrations becomes more positively skewed (see Section 3.2). Traditionally, with discrete samples, the concern for underestimating the mean has been addressed by specifying an acceptable level of uncertainty (often 5%) and a method for calculating an estimate of the mean with that level of confidence (such as a 95% UCL on the mean). A similar approach is used with ISM data, as discussed below.

For those accustomed to working with 95% UCL values from discrete datasets, there are some important differences with 95% UCLs from ISM data. As discussed in Section 3.2.4, calculation of a 95% UCL for ISM data requires a minimum of three ISM samples, which is generally fewer samples than is required for discrete datasets to yield reliable 95% UCL values. The first ISM Team built an ISM UCL calculator in an Excel spreadsheet file that has been updated since then with an improved modeling procedure. Further information about the updated ISM 95% UCL calculator can be found in Section 3.2.4. Additional ISM replicates above the minimum of three increases the performance of the mean estimate, thus providing a 95% UCL closer to the actual mean; although this also increases the cost, it may be necessary if the site is relatively heterogeneous and worthwhile if the result is anticipated to be close to a level of concern. Alternatively, the CSM may need revision and multiple smaller DUs established if the inherent assumption that there are no significant spatial patterns of contamination within the DU is suspected of being incorrect.

With discrete sample datasets, the maximum concentration observed is sometimes used as the EPC if it is less than the calculated 95% UCL. However, with both discrete and ISM data, the maximum concentration observed may still underestimate the population mean if the sample size is low or the population is highly variable. As discussed in Section 3.2.6.1, the calculated 95% UCL value should always be used as the EPC with ISM samples, even if it is higher than any of the individual ISM results. This situation is not uncommon, particularly when the number of replicates is small. In fact, with three replicates, the 95% UCL will always exceed the highest individual ISM result.

As discussed in Section 3.2.4, two methods for calculating the 95% UCL from ISM data are recommended: Student's-t and Chebyshev. The choice of method depends on the known or anticipated shape of the probability distribution of contaminant concentrations in the DU. Section 3.2 presents the more common 95% UCL equations and decision criteria for selecting the approach that is most likely to provide the desired coverage of the arithmetic mean concentration. An Excel tool (ISM 95% UCL Calculator) is available from the ITRC webpage to facilitate these calculations. USEPA's ProUCL 5.1 software, which implements a wider range of 95% UCL calculation methods, is generally not recommended for use with ISM datasets consisting of fewer than 8 replicates. Furthermore, USEPA guidance on the use of ProUCL 5.1 (USEPA 2015) cautions that at least 10 to 15 observations are needed before relying on bootstrap resampling techniques to estimate 95% UCLs.

The concepts of coverage, CI widths, and accuracy of the 95% UCL are discussed in Section 3.2. In risk assessment, each of these are a measure of uncertainty in the EPC with coverage being the frequency that the 95% UCL is expected to equal or exceed the mean. Typically, a 95% UCL is used for risk assessment, with the resulting uncertainty of a 5% chance that the true mean is above the EPC. As discussed in Section 3.2 and in ISM-1, Section 4 (ITRC 2012), simulation studies have demonstrated that the degree of skewness of the underlying distribution of increments affects whether a 95% UCL method can provide a coverage of 95%. For unknown distributions or highly skewed distributions with CV greater than 1, the Chebyshev 95% UCL method is capable of calculating an EPC with 95% coverage. While we have no way to demonstrate the accuracy of an EPC, the CI width provides a measure of the precision. The more precise the replicates, the lower the RSD and the narrower the CI. A narrower CI reduces uncertainty in the EPC and risk assessment. A small mean-to-95%UCL width is desirable when the goal is to confidently estimate the true DU mean. The degree of variability (that is, the range of data values), expressed as the SD, is a common measure of variability and is used to calculate the 95% UCL – less variability (a lower SD value) gives a narrower mean-to-95% UCL width. Tolerance for the mean-to-95%UCL CI width (see Section 3.2.4.3) can be defined in the DQOs during systematic planning (Table 3-3). The 95% UCL may provide an unreliable estimate of exposure if the dataset is from too few increments or replicates and/or is highly variable (see Section 3.2.2). On occasion, it may be desirable to undertake an additional phase of investigation with redesigned DUs and/or more increments per DU to achieve lower RSD, a narrower CI, and more confidence in the EPC and risk estimates.

8.3.3.1 Combining SUs, EUs, or DUs for 95% UCL calculation of EPC

On occasion, there might be a desire to combine information from multiple SUs, EUs, or DUs into a single larger EU or DU area. Recall that DUs for risk assessment are EUs, although DUs for other study questions may also be investigated at a site. Three types of situations are described here:

- Each SU or EU has three replicate ISM samples with either the same or different spatial coverage.
- Three replicates from one or more random SUs or EUs and a singlet ISM is sampled from all other SUs or EUs.
- A random subset of SUs or EUs is sampled from a very large CSM-equivalent DU or project area.

Cautions on combining SUs or EUs into larger EUs or DUs relate to the increased uncertainty in the EPC and risk estimates. If there is a spatial trend across SUs such that some SUs are more similar to each other than others, then ISM samples collected from each SU would not be independent, meaning sample independence is violated. An example of a spatial trend across SUs is if a release occurred primarily on the top of a hill but migrated down certain parts of that hill, then the SUs comprising the preferential path downhill and the SUs outside the preferential downhill path would each have similar spatial trends but would be different from each other. The consequence of violating the statistical assumption of independence is that the variation will be underestimated, resulting in an underestimation of the EPC and risks. Uncertainty in the EPC and risks is also higher for the situation in bullet 3 above, due to a lack of sampling across the entire DU or large project area.

***ISM three replicates per SU or EU*.** When each ISM SU or EU has three replicates, and there are multiple SUs in an EU or multiple EUs within a larger DU, then a mean can be used to provide an estimate of the 95% UCL EPC across a receptor-specific EU. If the SUs or EUs are all the same size (area and depth) with the same increment density, then a standard 95% ISM UCL can be calculated using the ISM 95% UCL calculator described earlier in Section 8.3.3 and in Section 3.2.4. The variable sizes of each SU, EU, or DU can be taken into account by using a weighted mean. Weighted means take into account the spatial scales of contamination or receptor-specific activity patterns. Section 6.2.2 provides a detailed discussion and equations for calculating weighted means, There are two primary situations when calculation of a weighted mean might occur for risk assessment:

- The CSM of a site anticipates different expected levels of contamination in different areas within a larger area that we would like to define as an EU. Each of those subareas might be investigated as a separate DU for site characterization and then combined to define a single EU. Section 3.1.6, Example Set 3, provides an example CSM with multiple potential source areas within an EU.
- For ecological or human health risk assessment, we might need to consider a variety of sizes of EUs to accommodate multiple receptor scenarios because different receptors' exposure areas are of different spatial scales. For example, if the area of a pocket mouse habitat is a quarter of that of a muskrat, which is an eighth of that of a fox, we might need to sample SUs defined for pocket mice but then combine the SUs into EUs for receptors with larger home ranges.

When these considerations are incorporated in the initial planning stages, they can be addressed by using a stratified sampling design. Within each stratum (smaller SU or EU), it may be appropriate to use ISM, but appropriate systematic planning is needed to ensure that the ISM data from the different strata can be combined for a larger EU. If there are three replicate ISM samples in each SU or EU, a weighted mean could be calculated as described in Section 6.2.2. While the ISM samples in this case are not true replicates of the mean throughout the EU in the sense that they provide information on different portions of the EU, they can provide an unbiased estimate of the mean for the whole EU.

Table 6-3 provides numerical examples of this calculation, where data from two scenario-specific EUs are combined to derive a 95% UCL for a larger scenario-specific EU. In these examples, an elementary school is divided into two EUs representing different play areas: EU1 is the kindergarten playground, and EU2 is the playground for older children. A maintenance worker has contact with both of the smaller EUs, and a separate EU is constructed to reflect the exposure of this worker. If it is assumed that, on average, a maintenance worker spends equal time in EU1 and EU2, then the replicates from each EU can be weighted equally, yielding the results shown in the "Equal weight" row of Table 6-3 for the maintenance worker's EU. Alternatively, it may be assumed that a maintenance worker's exposure is proportional to the respective areas of the two smaller EUs, and the equations from Section 6.2.2 can be used to generate summary statistics for the proportionally weighted combined area, presented in the "Proportionally weighted" row of Table 6-3. The weighting factors applied to each EU should sum to 1.0, which is achieved by dividing each area by the sum of the two areas.

You can click here to download the updated ISM 95% UCL Calculator (ISM 95% UCL Calculator) for a combined EU from

several smaller EUs.

By design, ISM sampling provides an estimate of the average concentration within an EU. As noted above, if areas of elevated concentrations or other spatial patterns of contamination are suspected, these subareas can be evaluated as separate N&E DUs and then combined to define a single EU. This type of ISM design also supports risk management decisions for individual DUs that might contribute to risks within a larger EU.

The same methodology for calculating a weighted average 95% UCL described above for multiple surface soil SUs, EUs, or DUs could be used to combine a surface SU or EU with one or more corresponding subsurface SUs or EUs. The only slight difference would be that the weighting term would reflect the proportion of the total soil volume within the vertically integrated EU.

Three replicates from one or more SUs. If one or more SUs or EUs are represented by a single ISM sample, and one or more other SUs or EUs are represented by at least three replicates, the following methods are options for calculating the 95% UCL for the larger scenario-specific EU:

- One option, discussed in Section 3.2.6.2, is to use pooled variances from SUs or smaller scenario-specific EUs with three replicates that are applied to calculate 95% UCLs for the singlet SUs or EUs. This method is appropriate for CSM-equivalent SUs and EUs where a statistical test that compares variances demonstrates that the differences in variances are not significantly significant (that is, at the 95% level of confidence).
- Another method for computing a 95% UCL could employ the random selection of one replicate result from each SU or EU with multiple replicates (for example, the first replicate) after establishing in the SPP how the replicate used in calculating the 95% UCL would be randomly selected, as described in Section 3.3.2.
- If a similar variation is expected across the SUs or smaller scenario-specific EUs, an alternative option discussed in Section 3.3.1 involves collecting from one of the SUs or EUs at least three replicates, applying the measured RSD to each of the SU or EU results, and proceeding with the weighted 95% UCL calculation discussed above.
- Section 3.3.2 gives additional details on these and various possibilities to structure the statistical analysis of scenario-specific EUs composed of smaller SUs or EUs.

Random sampling subset of SUs or EUs to characterize a very large EU or project area. Occasionally, a risk-based study question may be regarding a very large receptor-specific EU. A brief discussion of using ISM data from multiple SUs or smaller scenario-specific EUs within a larger EU when a singlet ISM sample is obtained for each SU or smaller EU is provided in Section 8.2.2.2, and details pertaining to statistical evaluation of these data and a calculation of a 95% UCL for the very large EU is provided in Section 3.2.8.1. A broader discussion for approaches to evaluate risks or compliance with risk-based concentrations from large EUs follows.

In some cases, a large EU can be divided into equally-sized multiple SUs or smaller scenario-specific EUs, and only a portion of the SUs or EUs need to be sampled to provide either a mean of the entire EU or determine if the means for each SU or EU will likely be less than a risk-based benchmark. While the grids within the EU should be equally spaced and contiguous (see Section 3.2.8.1, Figure 3-25), the SUs sampled do not have to be contiguous. A feature consistently required for sampling a random subset of SUs or smaller EUs is a CSM-equivalent study area supported by a mature CSM.

When the risk assessment is evaluating a large EU, and the area is a CSM-equivalent EU, an approach for sampling a random subset of SUs is described in Section 3.2.8.1. The example is an 80-acre EU for a farmer that is divided into 80 equal-sized 1-acre SUs. A minimum of 10 randomly selected SUs are each sampled with three replicate ISM samples from at least one of the SUs and a singlet ISM sample from the remaining of the randomly selected SUs. Using a minimum of 10 SUs in the sampling design allows for calculating the 95% UCL with ProUCL. Three replicates from one of the SUs provides a measure of the variability between replicates to verify the accuracy of CSM equivalency across the EU. Increasing the number of SUs sampled can reduce the width of a CI (and the magnitude of the 95% UCL) because the number of samples factors into the 95% UCL calculation; likewise, a determination of compliance (95% UCL ≤ threshold) may be sensitive to the choice of number of SUs. Thus, although sampling a subset of 10 SUs is statistically sufficient for calculating a 95% UCL via ProUCL, the variability in ISM results between the SUs should be established in the planning and documented in the DQOs. Furthermore, the spatial coverage of the large EU may be too small and create an unacceptable amount of uncertainty in the risk assessment for broad use.

If the question regarding a very large CSM-equivalent study area is, “Are all EUs in the study area less than a target risk-based concentration?”or“What are the estimated risks (a 95% UCL is needed for the risk assessment EPC)?”, then an alternative approach with less uncertainty is to expand special coverage by increasing the number of SUs (in this case, small

EUs) in the sampled subset. Section 3.2.8.2 details the process for an EU or study area so large that it cannot be sampled as a single unit. The very large CSM-equivalent EU is completely divided into many (more than 100) SUs of equal size, or the same strategy can be used to divide a large study area or property into multiple equal-sized SUs (smaller receptor-specific EUs) that are randomly sampled using ISM. The statistical calculations, which are independent of spatial area, establish a sample size of 59 SUs (smaller receptor-specific EUs), which is sufficient for 95% confidence that at least 95% of the EUs in the very large study area are less than or equal to a threshold (such as a risk-based screening level or cleanup goal) if none of the EUs are greater than the threshold and for calculating a 95% UCL. For a 95% UCL, three replicates are collected from a percentage of or all of the smaller EUs to determine the pooled variance for use in calculating the 95% UCL. Section 3.2.8.2 notes key factors that can influence the extrapolation uncertainty and likelihood of making a decision error:

- the variance of the increments (CV of the underlying distribution)
- the percentage of the large study area or large EU area sampled
- the likely magnitude of the average 95% UCL (across all sampled subset EUs) relative to a compliance level (ratio of average 95% UCL divided by compliance level)

Simulation studies suggest sampling 30% or more of the study area to achieve less than 5% false compliance (false negative) decisions (Goodrum et al. 2018).

In all cases when extrapolations are made from sampled areas and used as surrogates for unsampled areas, it is important to use the uncertainty section of the risk assessment to discuss uncertainty in the EPC and risk estimates. Statements on uncertainty from the extrapolation methods used for unsampled areas can and should be supported with a discussion on the demonstrated low variability as measured by the replicates' RSD and the verification of CMS-equivalent DUs in the risk assessment uncertainty analysis. Greater uncertainty in the representativeness of the subset of SUs sampled to the entire very large EU is tolerated when either the COPC concentrations are far below the risk-based screening levels (or ALs), or the calculated risks and hazards are far below the regulatory thresholds (typically 1 x 10^{-6} risk and hazard index of 1).

8.4 Considerations for Use of Background ISM Data in Risk Assessment

Background concentrations are often used in risk assessment to help refine the COPC list so that chemicals related to site releases can be more easily identified. Risk assessments may compare ambient background concentrations to site concentrations to eliminate COPCs before or after quantitative risk assessment is completed, depending on the stakeholders. ISM can be used for assessing ambient background soil concentrations of both native metals and ubiquitous anthropogenic chemicals such as dioxins and PAHs. While a background threshold value derived from discrete data can sometimes be used to determine if a chemical has been released to the environment and should be included in the quantitative risk assessment, it should not be used to eliminate a chemical based on an ISM site concentration. The comparison of ISM site data to discrete background benchmarks should be done with an understanding of the potential error in the mean based on the ISM result. Details about errors and other important concepts comparing ISM results with a single value are covered in in Sections 6.2, Section 3.2.5, and Section 3.3.3. Also, as stated previously in this guidance document, while statistics exist to compare ISM samples to other types of samples, these methods are complex and require the assistance of a qualified statistician. Therefore, this section will describe the benefits of using ISM background results for comparison with ISM site results, how to properly plan for including a background ISM comparison in a risk assessment, and briefly describe comparison methods. Sections 3.1.6.2, Section 3.3.4, and Section 6.3 cover these topics in more detail.

8.4.1 Benefits of using ISM site and ISM background datasets in risk assessment

If ISM background data are properly collected, they can be used to compare and determine which chemicals detected in ISM site samples are likely present in similar or greater concentrations than background conditions. This information can be used before the quantitative risk assessment to eliminate COPCs or after the risk estimates are completed to aid in determining which chemicals are COCs. If the background comparison is completed after risk estimates are calculated, the comparison of COCs detected in ISM background to ISM site concentrations can provide information about how much of the total site risk for exposure pathways and scenarios is attributable to background conditions. Since both background and site data provide mean concentrations, they are relatively easy to compare and use to make decisions about COPCS or COCs. In addition, planning to collect an adequate number and matching soil types for ISM background and site soils will enable the best comparison results. The statistical comparison methods are also fairly straightforward to use.

8.4.2 Planning for ISM background data collection and use

Background data from an appropriate reference area are often used to evaluate site data for environmental projects. Statistical methods typically applied to compare discrete site sample concentrations to discrete background concentrations are applicable to compare an ISM site dataset to an ISM background dataset. Section 3.1.6.2 presents aspects to consider in planning for selecting and defining ISM background SUs and DUs. Figure 3-10 in Section 3.1.6.2 provides a depiction of the way background DUs could be configured when background areas the same size as site DUs are not contiguous and/or not square/rectangle grids. Ideally, ISM background samples should be comparable to site data in the following ways:

- same sample range of depths (for example, surface soil defined as 0 to 1 ft bgs)
- same soil type (such as sand or loam)
- same increment density (for example, 30 increments per ½ acre)
- same number of increments and replicates in the DUs
- same field methods
- same analytical methods

Assumptions should be clearly stated in the planning stage and revaluated after the data are available. Power is the probability of detecting a difference between background and site concentrations, given that a difference truly exists. A power analysis based on the expected variation and desired CI should also be completed in the planning stage, so the number of increments collected in each SU or EU is more likely to meet the acceptable decision errors (see Sections 3.3.3.1 and Section 3.3.4.1). See Section 3.3.4 for more details on planning for background comparisons.

8.4.3 ISM background comparison methods

Although USEPA guidance on hypothesis testing (USEPA 2002d, b, 2009) was developed with discrete sampling in mind, these methods are also applicable to ISM data comparisons. Two fairly easy statistical methods useful for comparison of ISM background and site data are means versus upper tails (see Section 3.3.4). Evaluating assumptions and completing power analyses before sampling and after data are available are also recommended.

Section 6.3 describes how to evaluate some of the assumptions underlying certain statistical tests, such as independence of the samples and normality of the distribution of the data. ISM sampling designs commonly do not have a sufficient number of samples to reject normality regardless of the true distributional form. However, as noted in the 95% UCL part of Section 3.2.4 and Section 6.3.2, an assumption of normality is generally appropriate for ISM data due to the CLT.

If background and site data are both generated using ISM, comparisons of means can be made using hypothesis testing. Section 3.3.4.1 provides guidance about the use of one- or two-sided hypothesis tests and what to do if the distributions of one or both ISM background or site datasets do not appear normally distributed. ANOVA can also be used to compare the variability of the means between both datasets if the variances are equal. ANOVA is useful for comparing multiple site DU means to background DU means, but it only identifies if there is a statistical difference between the groups, not the groups that differed, necessitating follow-up by individual comparisons for each site DU to the background DU. Furthermore, for 95% confidence ($\alpha = 0.05$), a minimum of five ISM replicates are typically needed from each DU. While ANOVA is a viable option for statistically comparing ISM site data to ISM background data, it is best conducted by a statistician or other professional well versed in environmental statistics.

Section 3.3.4.2 discusses how to compare the upper tails of the two distributions from the ISM background dataset and ISM site dataset. Statistical power to detect differences will be low if there are a limited number of replicates in the ISM dataset. Commonly, at least $n = 8$ observations per group is desired before using hypothesis tests to compare upper tails (such as the Quantile test).

Decision errors and power analyses are discussed in Section 3.3.3.1 and Section 3.3.4.1 Acceptable decision error probabilities should be determined in the planning phase and reevaluated when the data are available. Power analyses should be completed in the planning stage to determine the number of increments to collect in each SU but are also useful to interpret the false exceedance of background error in a one-sample hypothesis test.

A multiple lines of evidence approach with qualitative comparisons of site and background ISM data in conjunction with statistical comparisons of means and/or upper tails can provide more confidence in a decision regarding how to evaluate a native metal or ubiquitous anthropogenic chemical as a COPC or COC in a risk assessment. An example of a qualitative approach is comparison of RSDs among replicates from a site ISM DU to the background ISM DU. Because the ISM results are

means, the variability (RSD) in the ISM replicate results from a contaminated site DU may be higher than that in the ISM background replicates. Discrete data from a contaminated area typically have wider variability in their distribution than discrete background data, but ISM datasets have less variability than discrete because ISM data distributions are from mean concentrations (ITRC 2012). Although ISM field replicates measure sampling precision, comparison of RSDs from ISM site to ISM background data may indicate a contaminant release has occurred on the site but cannot be used to support the hypothesis that site concentrations are similar to background concentrations. Including graphical analyses and figures will aid in showing important differences between ISM site and ISM background distributions. Simple graphical analysis can also provide useful information and serve as a semi-quantitative means of comparison. Background comparisons accompanied by visual presentations of site versus background data such as scatter plots, histograms, or box plots provide more information about how the data compare.

8.5 Use of ISM for Post-Remediation Risk-Based Confirmation Sampling

The application of ISM is most straightforward at the post-remedial confirmation sampling stage of an environmental investigation. The objective of confirmation sampling is commonly to compare average soil concentrations to remedial action criteria within defined DUs that are remediation units. ISM can be used for confirmation sampling conducted to evaluate if a remedial action meets risk-based benchmarks because a properly designed ISM sampling plan can provide a more robust estimate of mean residual contaminant concentrations in a DU than a sampling plan reliant on discrete samples.

In confirmation sampling, the usual objective is to determine whether there is sufficient evidence to conclude that the true DU mean concentration is less than the remedial goal associated with the remediation. Generally, this involves comparison of a 95% UCL to the remediation goal. Section 3.2.5 discusses the topic of statistical hypothesis tests to support this objective and why the use of a 95% UCL is a simple and appropriate way to implement steps 5 and 6 of USEPA's DQO process as it is equivalent to conducting a hypothesis test with a 5% error rate.

The most common example for the application of ISM in post-remediation confirmation sampling design is excavation of contaminated soil from a DU, which allows ready access to the assumed outer margins of the DUs. The sidewalls and floors of excavations can be treated as one or more DUs and can be sampled for confirmation of adequate soil removal. If the spatial boundaries of the excavation DUs are different than the site-specific EU for risk-based cleanup goals, then either combining excavation DUs into EUs or subdividing DUs into SUs with spatial scales equivalent to EU spatial boundaries is appropriate. Use of ISM for the collection of confirmation samples is illustrated in Example Set 1 in Section 3.1.6.

ISM is applicable to confirmation sampling if the criterion for successful excavation is achieving a mean concentration in soil DUs (sized appropriately for the project objectives, for example, risk-based screening levels or remediation goals) below the ALs at the excavation boundaries. Small volumes of contaminated soil within otherwise clean excavation sidewalls or floors do not necessarily pose a significant risk to human health and the environment. The specification of an appropriate DU area in relation to confirmation sampling should be addressed during the planning process.

8.6 Risk Communication Suggestions for Explaining ISM

The goal of risk communication is for all stakeholders to have a common understanding of the results, processes, and assumptions used in risk assessment as well as how the risk assessment is used in risk management. Use of ISM for risk assessment input may necessitate communicating with stakeholders and others the differences between discrete and ISM sampling, as well as how the proper use of ISM enhances the defensibility of risk assessment results in comparison to reliance on discrete sampling. Section 9 of RISK-3 (ITRC 2015) provides a general overview of risk communication, including issues related to risk perception and strategies for communicating technical information to diverse groups of stakeholders. This section focuses on the key differences between ISM and traditional discrete and composite samples and provides suggestions on addressing common misconceptions about ISM. For further information on risk communication, see ITRC's Risk Communication guidance (ITRC 2020).

8.6.1 ISM reduces uncertainty in the mean

Uncertainty in the risk estimates from the use of ISM sampling and analysis protocols derives from the reduced uncertainty in the average concentrations of contaminants in soil or sediment. Scientific studies demonstrate lower variability between ISM replicates than discrete samples. Section 2.2.2 and Section 2.5.2 discuss the soil science and sampling theory that

contributes to the larger variation between discrete samples' mean concentrations and includes a study by (Becker 2005) where lead is the COPC. Studies by Brewer et al. (Brewer, Peard, and Heskett 2016) (Brewer, Peard, and Heskett 2017) examined the effects on variability between discrete samples and ISM replicates within a DU for three different COPCs – arsenic, lead, and PCBs. Arsenic was used as a termiticide and preservative for building products at a former facility that manufactured ceiling and wallboard material treated with arsenic at Site A. Lead in incinerator ash at a former municipal incinerator was investigated at Site B. PCBs at a former radio broadcasting station was the focus for Site C. The ISM design of three replicates from each DU with either 54 increments (arsenic and lead) or 60 increments (PCBs) per DU provided robust evidence of reduced variability between samples as compared to discrete samples. Further sound evidence for reduced variability with ISM was demonstrated upon examination of 20 iterations of random groupings of 10 discrete samples from each of Sites A, B, and C. The largest variation was seen in Site C's PCB random grouping of discrete samples, where the 95% UCL ranged from 9.4 to greater than 1,000,000 mg/kg. The 95% UCL for arsenic discrete datasets of $n = 10$ ranged from 403 to 776 mg/kg from Site A. Lead 95% UCLs from Site B ranged from 201 to 439 mg/kg, straddling the risk-based concentration and preliminary cleanup goal for commercial/industrial receptors. These studies demonstrated that the mean concentrations from discrete samples are more prone to elicit inaccurate risk conclusions resulting in incorrect decision-making for protection of human health and the environment.

From a statistical standpoint, analysis of multiple ISM replicates collected with the same sampling protocol (sampling method and number of increments in an EU) provides a direct measure of the variance in the mean. ISM variances in the mean concentration are typically much lower than that for discrete sample data. A figure may be useful for conveying the population differences in variances between discrete and ISM datasets, and one option is a modification of what was originally Figure 4-3 in ITRC ISM-1 (ITRC 2012). Lower variability, as is characteristic with ISM, yields a 95% UCL with a smaller CI width, and as discussed in Section 8.3.3, a small mean-to-95%-UCL width provides more confidence in the estimate of the true EU mean. In comparison to the results from several discrete sample results, several ISM replicate results typically have a tighter upper CI. The lower variance among the ISM replicates results in less uncertainty in the EPC and in the associated risk estimate.

8.6.2 ISM provides a more representative EPC

ISM protocols typically provide much greater spatial coverage than discrete sampling. Because the use of ISM samples provides better spatial coverage than randomly placed discrete samples, EPCs based on a properly designed ISM sampling plan (one with an adequate number of increments and replicates in a DU) are more representative of a receptor's potential exposure over a long-term period than random discrete samples. The EPC is only one factor in the equations for estimating risks that are upper bound estimates used to meet the goal of ensuring that risks are not underestimated. The higher confidence in the EPC reduces the uncertainty in the risk estimates and strengthens the scientific defensibility of risk-based decision-making. A common concern with the use of ISM is that there could be areas of elevated soil concentrations within a DU, and that ISM will "average away" these concentrations. ISM does provide an estimate of the average concentration within a DU but no information on possible spatial patterns of contamination therein unless the DU is subdivided into mutually exclusive SUs. However, it must also be acknowledged that there is a much higher probability of hitting a small area of elevated soil concentrations with three 50-increment ISM sample replicates than with, for example, 10 or 15 discrete samples collected in the same area. The trade-off is between diluting the concentrations from the elevated area by including the area within the larger DU mean – and potentially missing the area altogether when using discrete samples. Concerns with identifying subareas of much higher soil concentrations are particularly valid when the CSM suggests that such areas could exist, and as noted in Section 8.2.3.1, sampling designs can be developed to address these concerns.

Stakeholder Perspective and Tribal Input

Several key aspects of stakeholder and tribal perspectives on the use of ISM should be considered during planning, including stakeholder and tribal engagement during systematic planning and decision-making; the value of a CSM in risk assessment and communicating risk to stakeholders and tribes; and stakeholder and tribal concerns regarding ISM in general. This section presumes familiarity with systematic planning (Section 3.1) and risk assessment (Sections 8.1 and 8.2) and provides references to those sections to avoid redundancy.

ISM can be used in various stages of site investigation, including source area identification, evaluation of fate and transport, assessment of potential exposure for risk-based decision-making, and confirmation sampling after a site has been cleaned up. The current or future use of these properties determines the level and extent of stakeholder and tribal involvement in the decision-making process. To illustrate the various applications of ISM, this section also summarizes various stakeholder and ecological considerations in ISM investigations as presented in several of the case studies found in Appendix A.

9.1 Introduction

The term *stakeholder* is broadly defined as members of environmental organizations, community advocacy groups, tribal entities, or other citizens' groups who deal with environmental issues, or a concerned citizen such as a homeowner or business owner. Members of public and private lands management organizations (state parks, national parks, and forests) and other ecological and/or protected lands (nature reserves and wildlife refuges) are considered stakeholders in the context of this section, too. The term *tribal* represents the Indian tribes, such as Pueblos, Nations, and so on; Native Hawaiians; and Alaskan Native Americans (Tlingit, Athabascan) and Native Alaskans (Yupik, Inupiat).

A vital difference between stakeholders and tribes is that tribes have government-to-government relationships with regulatory agencies, and stakeholders do not. In fact, many tribes enforce their own USEPA-approved standards. Some tribes are even developing tribal risk assessments that incorporate pathways and scenarios based on traditional and cultural routes of exposure, which in some cases are essentially and profoundly different from traditional risk assessments. Proposals to tribes that include ISM should demonstrate compliance with any tribal regulatory limits and should be part of a process that respects tribes' government-to-government status (ITRC 2012).

A differentiation is made between stakeholders and interested parties (responsible parties, state regulators, and owners and operators of contaminated sites) for discussion purposes in this section only.

9.2 Stakeholder Engagement Through Systematic Planning

Stakeholders and tribes, like interested parties, want to be assured that site investigation activities such as ISM "do no harm" and that the planned activities can effectively confirm the presence or absence of contamination. Stakeholders and tribes generally support planned activities and try to understand the processes used to characterize and/or clean up a site. During investigations, examples of questions that could be asked include, "Does soil contamination exist?" or "Why was this area sampled and why not Sampled over there?" The number one question is, "Does the contamination present an unacceptable risk?" Answering these questions requires open communication during systematic planning (Section 3.1.2 and Section 3.1.3) and throughout the project.

9.3 Communicating with Stakeholders

The pictorial representation of the CSM, such as the example in Figure 3-1, serves to illustrate the relationships among contaminant sources, fate and transport media, and exposure pathways to human and ecological receptors. In other words, the CSM presents a current understanding of the site. Communication with those involved in or affected by the site investigation and assessment of potential risk should start in the planning stage of the project, with the CSM being particularly useful in risk communication with stakeholders and/or tribes (see ITRC's Risk Communication guidance (ITRC 2020). This is a stand-alone ITRC web document by the Soil Background and Risk Team found on the ITRC website.

Establishing clear objectives at the beginning of a project is crucial to efficient and effective site investigation. The means by which these objectives are established is through the SPP, involving end users, interested parties, stakeholders, and/or tribes to help develop the CSM. The outcome of this effort is a well-thought-out DU whose locations and dimensions produce

information to support all investigation questions (see Section 3.1.1).

Two primary concerns of the ISM approach, as expressed by several members of the ITRC stakeholder group, are (1) the idea of averaging away a small area of elevated contaminant concentrations and (2) ensuring that current or future populations (human or ecological) will not be adversely affected as a result of an overlooked small area of elevated contaminant concentration.

9.3.1 ISM versus discrete sampling

Concerns about sampling uncertainty relate not only to ISM but to discrete sampling approaches as well. It is important during the systematic planning stages of an ISM project that stakeholders gain a common understanding of the differences between discrete and ISM sampling, as well as how the proper use of ISM enhances the defensibility of the results in comparison to discrete sampling. Strategies for communicating technical information can be found in Section 9 of "Decision-making at Contaminated Sites: Issues and Options in Human Health Risk Assessment" (ITRC 2015). Section 8.6 discusses the key differences between ISM and traditional discrete sampling and also provides suggestions on addressing common misperceptions of ISM, including the concern stated above – that ISM averages away small areas of elevated contaminant concentration.

9.3.2 Ensuring the safety of current or future populations (human or ecological)

As stated in Section 8.2.2.1, "An *exposure area* is an area where human and ecological receptors can contact contaminants in soil on a regular basis. Examples include residential yards, school yards, playgrounds, gardens, outdoor areas of commercial/industrial properties, or other areas designated as exposure areas by other means." The ISM approach allows for targeted sampling within an EU to address these types of concerns, and the CSM provides the information upon which the DQOs and sampling plans are designed. Examples of this targeted sampling approach are seen in Figures 3-11a, 3-11b, 3-11c1, 3-11c2, and 3-11c3. As stated in Section 3.2.6.2, DQO step 4 (defining DUs), planning for the number of DUs is a decision that involves all stakeholders, but considerations for the uncertainty in risk and risk-based decisions should err on the side of protecting the public health: "Generally, practitioners would rather make the mistake of remediating a site that is already clean, than make the mistake of not remediating a site that is contaminated."

This document may prove helpful in explaining DU design during systematic planning as described in Section 3.1.5. There are times when stakeholders and tribes need a better understanding of how sampling is done and why sample locations are placed in particular locations. Sampling plans should aid these stakeholders and tribes in understanding the challenges associated with soil sampling and how ISM addresses some key uncertainties associated with it. As discussed in Section 8.2.2.1, DUs applicable to human receptors may not be applicable to ecological receptors, so when sites are evaluated for both human and ecological receptors, multiple spatial scales may need to be considered for sampling.

The key takeaway is that ISM provides better coverage with a limited number of samples and a more reliable estimate of the true mean contaminant concentration for a DU (human or ecological), thus reducing the uncertainty of the decisions that need to be made.

9.4 Case Studies

Three case studies in Appendix A provide examples of successful ISM investigations and stakeholder engagement. By acknowledging stakeholder and tribal concerns early and through systematic planning, it is possible to communicate technical investigative approaches such as ISM in an open and transparent forum. This type of communication approach can satisfy stakeholders' expectations of fairness and speaks to their concerns about risk on a level and in terms to which they can relate.

9.4.1 East Kapolei, Oahu, Hawaii

An investigation of 413 acres of former sugarcane fields in East Kapolei, on the island of Oahu, was meant to collect sufficient information to determine if areas of the property are suitable for future residential housing development. The site investigation took place between April and July 2006 and assessed the surface and subsurface soil for contaminants resulting from the application of herbicides and pesticides. The primary COCs were arsenic and dioxin.

The land is currently owned by the State of Hawaii and operated by the State of Hawaii Department of Land and Natural

Resources (DLNR); the agricultural fields are currently leased for commercial fruit and vegetable cultivation. An enclosed area on the westernmost portion of the property is designated as a contingency reserve area (CRA), an environmental preserve that is monitored by the DLNR Division of Forestry and Wildlife due to the presence of endangered plants.

During the presampling site reconnaissance, the investigation team engaged the services of a land agent from the Department of Hawaiian Home Lands (DHHL) Development Division, who took the team on a tour of the site and provided historical information. At the end of the site visit, the team requested that the land agent provide contact information for the tenant farmer. The purpose of this information was to help determine farming schedules and crop rotations to avoid disturbing current operations. The tenant farmer requested that the investigation team identify the areas scheduled for sampling, and following a review of the proposed sampling areas, the tenant farmer agreed that the farming operations would not be affected by the sample collection and confirmed that the farmland in question was not scheduled for pesticide or herbicide application. The tenant farmer did request that sampling personnel attempt to avoid stepping on vines, plants, or vegetables in the fields.

The investigation team also contacted a representative with Forestry and Wildlife to discuss the CRA due to the presence of ecologically sensitive plants. Permission was given to perform sampling even if a DU extended within the boundaries of the CRA, but under certain conditions. It was requested of the investigation team that prior to scheduling sampling events, a representative from Forestry and Wildlife conduct a brief training session to help the sampling team identify the endangered plants and avoid disturbing them. Also, the Forestry and Wildlife representative requested that (1) the sampling team only do surface sampling, (2) vehicles would be prohibited from entering the area, and (3) the team needed to avoid damaging the irrigation system.

Following the pre-sampling reconnaissance with the affected stakeholders, the investigation team concluded that no change in the sampling strategy would be necessary. The sampling proceeded as scheduled.

9.4.2 Anclote Key Lighthouse, Pinellas County, Florida

The Anclote Lighthouse is a cast-iron lighthouse constructed in 1887 and located on Anclote Key, Pinellas County, Florida. The lighthouse was decommissioned as a navigation light in 1985 and is currently part of Anclote Key Preserve State Park. The park is located on a remote island and only accessible by boat; a resident park ranger lives on the site.

The paint used on the lighthouse was lead-based and has eroded and chipped over time. In the 1960s, the lighthouse became battery powered, and over the years, the casings from depleted aid to navigation (ATON) batteries were discarded near the tower or stored in buildings adjacent to the lighthouse. An initial assessment of this site was conducted in 1994, and at that time, 100 batteries were found and removed from the site. The assessment found lead and mercury in soil at concentrations above their respective cleanup target levels, thus making them COCs. The source of the mercury was presumed to be the batteries, and the lead could be from lead acid batteries or from the peeling and chipping lead-based paint.

The purpose of this ISM project was to build on the previous site investigation and support an assessment of human health risk from existing lead and mercury contamination. The primary risk considerations were public visitors to the lighthouse and the park ranger, whose residence is located on the property. A landscaped area surrounds the park ranger’s dwelling.

The interaction between the site investigation team, the State Park Service, and the park ranger was not provided in the case study. Subsequent inquiries to one of the investigators from the Florida Department of Health indicated that all affected entities were affiliated with the State of Florida and were involved in the planning and implementation of the sampling strategy. The sampling strategy and subsequent site investigation were developed in partnership with investigators from the University of Florida.

Key stakeholder considerations for this ISM investigation included the following:

- The park ranger was notified prior to sampling, and there were no other family members or pets on the premises to consider.
- The park ranger cooperated with the team in setting up the investigation schedule.
- The park was closed to visitors until the investigation and subsequent mitigation were completed.
- During the investigation, there was very little disturbance to the property, and boreholes were replaced with surrounding sand/soil.
- An ecological risk assessment was conducted prior to the ISM investigation to determine if there were any risks to nesting birds and found that the investigation posed no risk to the avian population.

9.4.3 Southeast Pennsylvania

A case study about residential properties in Pennsylvania being impacted by a landfill demonstrates how ISM techniques were used to determine if properties contaminated with BaP would automatically qualify for cleanup, if the property was compliant with site-specific cleanup levels, or if additional sampling would be necessary. The determinations were specific to each residential lot, and a separate decision option was applied to every individual residential property or lot within the housing development (Owens 2020). Periodic discrete sampling had been done in the area since 1984, but it wasn't until 1999 when Hurricane Floyd hit, flooding the area under 6 ft of water, that intensive sampling and testing were initiated as part of the emergency response action. In 2001, the landfill was officially listed as a Superfund site, and the remedial investigation commenced. Initial discrete sampling of the residential properties indicated BaP values were below the USEPA cleanup level.

Since this is still an active site, the location and other identifying information is not being revealed. The site history and stakeholder engagement activities were not included in the case study, so an USEPA member of the ISM investigation team was contacted and provided the information.

The residential properties being investigated are in a neighborhood located in southeast Pennsylvania that was developed in the 1960s as part of an urban renewal project where older homes and buildings were razed and the area graded for new homes. The neighborhood consists of approximately 300 homes and lies adjacent to a landfill that existed prior to development and from which excess dirt was used as structural fill during development. It is believed that a creek ran between the landfill and the neighborhood, and that the creek was filled in during the neighborhood development as well. PCBs, PAHs, and other contaminants from the landfill were mixed in with the structural fill material along with runoff from the creek prior to being filled in.

The neighborhood formed a community advisory group (CAG) and contracted with a technical assistance grantee (TAG). Monthly and bimonthly meetings were held between the members of the investigation team, the CAG, and the TAG, where planning documents were shared and discussed among the group. Members of the investigation team gave presentations on the CSM and proposed a sampling plan. The TAG was a great resource in translating the technical information to the CAG, and the CAG agreed to the ISM sampling plan. In most cases, each home was designated a DU, and some larger properties had multiple DUs. Members from the site investigation team worked with the individual residents in developing the sampling schedule for each DU.

The most common concern expressed by the residents during the investigation was, "Why did one home in a row of six not need to be cleaned up when all the homes got flooded?' A member of the site investigation team explained that the flood really had nothing to do with the contaminant distribution in the dirt – rather, the contaminants were already in the dirt before the flood, and the effects of soil heterogeneity and the arbitrary way the dirt was mixed and moved prior to the homes being built determined the levels of contaminant distribution. It was also explained to the residents that although all their homes had some level of contamination, the amount of contamination in some homes was below USEPA actionable levels and the reason why those particular homes did not need to be cleaned up. With the help of the TAG translating the more technical concepts, this explanation was widely accepted by the homeowners. Thirty-three homes were remediated as part of the initial emergency response action, and an additional 170 residential parcels have been remediated with an estimated 10 to 20 remaining.

Appendix A.Case Study Summaries

Summaries of case studies are provided in this guidance document to provide information with regard to ISM design, implementation, and assessment methodologies. The case studies presented were selected based on their relevance to the use and application of ISM, but the methodologies and conclusions of the case studies provided were not independently verified by the ISM Team.

Each summary provides a description of the case study including its key concepts, COCs, media, geographic area, regulatory agency, owner/responsible party, site complexities, field sampling, statistical sampling design, laboratory processing, data quality summary, level of effort, and outcome/lessons learned where available.

The amount of information provided in each published case study varies based in part on the focus, purpose, and/or goals of each. Consequently, the level of detail provided in each summary presented in this guidance generally reflects the amount of information that was available for review. Links to the actual case studies are provided in each summary to enable the reader to review and gather additional details.

Table A-1 lists the 10 case studies summarized in this document followed by the individual case study summaries.

Table A-1. Case study summary table.

Source: ITRC ISM 2020 Team.

Study #	Author	Title
1	J.L. Clausen, T. Georgian, A. Bedna, et al	Demonstration of Incremental Sampling Methodology for Soil Containing Metallic Residues
2	J.L. Clausen, T. Georgian, K.H. Gardner, T.A. Douglas, et al	Applying Incremental Sampling Methodology to Soils Containing Heterogeneously Distributed Metallic Residues to Improve Risk Analysis
3	D. Crumbling	Advanced ISM QC Field Three Replicates Strategy for Managing Hundreds of DUs
4	B. Bachmann, St. Germain	Confirmation Soil Sampling: Remedial Action Report, FairPoint Communications, Utility Pole Storage Area, 11 Mallet Park Road, Brunswick, Maine
5	R. Brewer	Evaluation of Green Island Landfill and Reburial Pit, Former U.S. Coast Guard LORAN Station Kure
6	(1) TetraTech report to HDOH (2) Enviroservices & Training Center, LLC to DHHL	East Kapolei Final Site Assessment and Site Investigation Reports, Kapolei, Oahu, Hawaii
7	K. Hyde, W. Ma, T. Obal, et al	Incremental Sampling Methodology for Petroleum Hydrocarbon Contaminated Soils: Volume Estimates and Remediation Strategies
8	B. Bachmann, St. Germain	Confirmation Soil Sampling: Pole Storage Area Remediation Report, FairPoint Communications Facility, 104-106 Fairbanks Road, Farmington, Maine
9	L. Stuchal	Anclote Key Lighthouse Assessment, Pinellas County, Florida
10	D.B. Stephens & Associates, Inc.	The Mineral Wool Site

Case Study #1: Metallic Residues at Shooting Range, ISM versus Discrete

ISM Concept Demonstrated: Cost savings and ISM sampling performance

Case Study Name: *Demonstration of Incremental Sampling Methodology for Soil Containing MetallicResidues*

Author(s): Jay L. Clausen, Thomas Georgian, Anthony Bednar, Nancy Perron, Andrew Bray, Patricia Tuminello, Gordon Gooch, Nathan Mulherin, Arthur Gelvin, Marc Beede, Stephanie Saari, William Jones, and Shawna Tazik

Date: September 2013

COCs: Metallic residues

Media of Concern: Surface soils at small-arms ranges

Case Study Link: https://www.itrcweb.org/FileCabinet/GetFile?fileID=15091

Background

Metal constituents are introduced into the environment as metal residues from small-arms and pyrotechnic military training areas. The U.S. Army Engineer Research and Development Center created this report for a project that was conducted at two inactive small-arms ranges at Fort Eustis, Virginia, and at the Kimama Training Site in Idaho (both Military Munitions Response Program, or MMRP, sites), as well as at one active small-arms range at Fort Wainwright, Alaska. These locations were selected for the sampling and sample processing of soil samples obtained from the ranges, with the three sites selected to provide a variety of soil types. The project had the objectives of demonstrating improved sampling data quality for metal constituents in surface soils on military training ranges and developing a methodology that would result in the same or lower cost as conventional grab/discrete sampling. This report summarizes the demonstration, which included comparing ISM to conventional grab/discrete sampling through assessing performance and cost.

Site Summary

Regulatory Agency/Program	This report was completed by the Engineer Research and Development Center of the USACE as a partial fulfillment of the obligations for Environmental Science Technology Certification Program (ESTCP) Demonstration Project ER-0918.
Description	Three small-arms ranges located at Fort Eustis, Virginia; Kimama Training Site, Idaho; and Fort Wainwright, Alaska.
Owner/Responsible Party	U.S. Department of Defense
Other Stakeholders	None

Site Complexities

Risks	NA
Characteristics	Three small-arms ranges evaluated, two were inactive MMRP sites and one was active
Other	NA

Planning and Implementation

Field Sampling Conducted	Fort Wainwright: 63 ISM and 50 grab samples Kimama Training Site: 18 ISM and 30 grab samples Fort Stewart: 27 ISM and 33 grab samples

DUs	Fort Wainwright: background, firing point, and berm face Kimama Training Site: background and berm face Fort Stewart: background and berm face
Statistical Sampling Design	Statistical comparisons and summaries were completed for each of the three sites.
Laboratory Processing and Analysis Summary	Laboratory processing included machining or grinding the soil, increasing the digested mass and the digestion interval, improving the digestion efficiency by increasing the acid to sol ratio, and subsampling to build the digestate sample. The following methods were used for analysis: 6010B/3050B for metals, 8330B for explosives, Walkley-Black Method for total organic carbon (TOC), SW-846 9045D for soil pH, ASTM D7503-10 for Cation Exchange Capacity (CEC), ASTM D421/ASTM D422 for grain size, and ASTM D2216 for moisture content.
Data Quality Summary	Acceptable; data were collected to compare ISM to discrete sampling in terms of performance and cost.
Level of Effort (Cost/Benefit, Effectiveness or ROI)	ISM requires additional costs for handling and processing of samples, but this was offset by the need to collect fewer samples using ISM. For each site, the study compared field labor and laboratory analyses costs for ISM and grab sampling. The cost comparisons assumed that three replicate ISM samples were collected for each DU and that grab samples included two scenarios (7 and 15 grab samples). Total costs are compared in Table 37 of the report. For the Fort Wainwright site, total project costs for sampling and analysis using ISM were lower than using grab samples. For the Fort Eustis site, total project costs were virtually identical for both ISM and grab sampling techniques. Total project costs for ISM were slightly higher than total project costs associated with grab sampling for the Kimama site. Overall, the study concluded that total project costs were 5% to 50% lower using ISM: grab sampling would require more samples to be collected.

Decision-making

Outcome and Lessons Learned	This report compared ISM to conventional grab/discrete sampling by assessing performance and cost. The study determined that, at all sites, ISM yielded reproducible and more representative metals soil concentrations than the conventional grab sampling methods. With respect to cost, it was demonstrated that using ISM creates a potential cost savings of 30 to 60% as compared to conventional sampling approaches.
Complicating Factor(s)	The authors of this report are currently working with USEPA to modify Method 3050B and incorporate the recommended changes identified from this project into a Method 3050C. Implementation issues are discussed in section 8 of the report.

Case Study Conclusions

This study was completed by USACE's Engineer Research and Development Center. The report summarizes a project that compares ISM to conventional grab/discrete sampling by assessing performance and cost at three small-arms ranges with metal residues.

Case Study #2: Metallic Residues at Shooting Range

ISM Concept Demonstrated: 100 increments per DU needed for metals at shooting ranges

Case Study Name: *Applying Incremental Sampling Methodology to Soils Containing Heterogeneously Distributed Metallic Residues to Improve Risk Analysis*

Author(s): J.L. Clausen, T. Georgian, K.H Gardner, and T.A. Douglas

Date: January, 20187

COCs: Lead and antimony

Media of Concern: Surface soils

Case Study Link: https://www.itrcweb.org/FileCabinet/GetFile?fileID=15092

Background

This published study was designed to compare grab soil sampling techniques to ISM to characterize the impacts from metals at a small-arms range. Grab and ISM sampling were used to estimate mean metals concentrations in surface soils, particularly for Pb and antimony (Sb), to compare to a background sample previously collected using ISM. The ISM samples were collected within the DU by using a range of increments from 5 to 200 to determine how the number of increments affect data quality. The data were ultimately used to calculate a 95% UCL and then compare it to USEPA residential screening levels.

Site Summary

Regulatory Agency/Program	NA; not discussed in the research paper
Description:	Camp Ethan Allen, Vermont
Owner/Responsible Party:	U.S. Department of Defense, U.S. Army
Other Stakeholders	NA

Site Complexities

Risks	Metals impacts on surface soils
Characteristics	Small-arms shooting range berm (approximately 300 m^2) with metallic and explosive residues
Other	NA

Planning and Implementation

Field Sampling Conducted	For grab samples, 30 grab samples were collected using a steel scoop (or equivalent) and placed in 4-oz glass containers. Grab sample locations were selected using systematic random sampling. For ISM samples, samples were collected to a total depth of 5 cm using a 2-cm diameter corer. Seven replicate samples consisting of 5, 10, 20, 30, 50, and 100 increments each were collected, as was one ISM sample consisting of 200 increments.
DUs	Entire face of the berm
Statistical Sampling Design	Site-specific descriptive statistics (max, min, mean, median) were calculated, as well as SD, percent relative SD, and RPD for the grab and ISM datasets. Additionally, USEPA ProUCL software was used to calculate the UCLs, Student's-*t* tests, and data distributions.

Laboratory Processing and Analysis Summary	Sample processing used a modified method: samples were air-dried, passed through a 10-mesh sieve prior to milling, milled, and then subsampled. Milling involved grinding the sample to <2 mm using a steel ring mill grinder. The milled soil was spread over a sheet of aluminum foil (1- to 2-cm thick layer), and 20 increments were collected using a flat spatula. Increments were combined to yield a 2-g digestion aliquot. Samples were analyzed in accordance with USEPA Method 6010C.
Data Quality Summary	ISM data were acceptable, with higher increment samples providing a better estimate of the mean concentrations.
Level of Effort (Cost/Benefit, Effectiveness, or ROI)	The study concluded that ISM sampling with a minimum of 30 increments and three to five replicates would provide a better estimate of the mean concentration for a DU than collecting a large number (>30) of grab samples, thus reducing the total cost of sample collection and analysis.

Decision-making

Outcome and Lessons Learned	ISM results had much higher precision compared to grab samples (the percent RSD was much smaller for ISM compared to grab), providing a better representation of the DU's mean concentration for Pb and Sb. Grab samples yielded a negative bias when estimating the mean. Ultimately, the ISM datasets indicated that the mean concentration for the DU exceeded the USEPA residential screening level.
Complicating Factor(s)	Elevated chromium concentrations were observed in the ISM samples, the source most likely the milling equipment, which contains chrome-steel grinding surfaces.

Case Study Conclusions

The ISM datasets for Pb and Sb provided much more precise data, with less difference between the mean and median concentrations compared to grab samples and a tighter range between the minimum and maximum concentrations. The study concluded that ISM samples with 100 increments would be appropriate for sampling small-arms ranges for metals, and that ISM samples with a smaller number of increments tended to underestimate the mean concentration.

Case Study #3: Hundreds of DUs, Very Large Sites, Residential, and Landfill

ISM Concepts Demonstrated: Advanced ISM QC field three replicates strategy for managing hundreds of DUs, evaluation of BaP, particle effects caused by data variability

Case Study Name: *Advanced ISM QC Field Three Replicates Strategy for Managing Hundreds of DUs*

Author(s): Deana Crumbling

Date: January 28, 2019

COCs: BaP

Media of Concern: Soil

Case Study Link: https://www.itrcweb.org/FileCabinet/GetFile?fileID=20441

Background

This technical memorandum serves as a description of ISM techniques used to evaluate DU compliance with a site-specific cleanup level of BaP and assess neighborhood residential properties impacted by a landfill. After a single sample was collected from each DU, a decision tree strategy was used to determine if the DU would automatically qualify for cleanup, if the DU was compliant with the site-specific cleanup level, or if additional sampling would be necessary to provide a three replicate DU. This project involved applications of statistical principles to derive an adaptive strategy to speed cleanups and reduce costs while still maintaining full protectiveness and transparency.

Site Summary

Regulatory Agency/Program	EPA Region 3 Superfund Division and Office of Superfund Remediation and Technology Innovation – Technology Integration and Information Branch
Description	A site located in USEPA Region 3; specific location details kept anonymous
Owner/Responsible Party	NA; specific site details are anonymous
Other Stakeholders	NA; specific site details are anonymous

Site Complexities

Risks	Impact to neighborhood residential properties
Characteristics	Residential properties affected by a landfill
Other	Comparisons made to a site-specific cleanup level of BaP to evaluate DU compliance

Planning and Implementation

Field Sampling Conducted	Incremental soil sampling was used to collect DU samples from two depth intervals (0 to 1 ft and 1 to 2 ft bgs). Each incremental sample was composed of 50 1-in diameter increments collected by auger-drill lift into a bucket through a hole in the bottom of the bucket. Because the mass of each incremental sample collected this way usually exceeded 2 kg, the sample was split in the field using incremental procedures to create a representative subsample mass between 1 and 2 kg, which was sent to the laboratory.
DUs	DUs consisted of residential yards
Statistical Sampling Design	Site-specific statistics collected to assess the degree of data variability prior to sampling
Laboratory Processing and Analysis Summary	The samples were processed and analyzed for PAHs, PCBs, and Pb. However, in the technical memorandum, only BaP was considered of the PAH analytes. Sample processing in the laboratory consisted of air-drying, disaggregation, passing the disaggregated material through a 10-mesh sieve to remove non-soil objects, milling of the <10-mesh material, then incremental subsampling of a 2D slabcake using 30 increments to form a 30-g analytical subsample mass.
Data Quality Summary	Acceptable
Level of Effort (Cost/Benefit, Effectiveness, or ROI)	The goal of sampling and analysis was to obtain an estimate of a DU's true BaP concentration for comparison to the site-specific cleanup level.

Decision-making

Outcome and Lessons Learned	The technical memorandum serves as a description of ISM techniques used to evaluate DU compliance with a site-specific cleanup of BaP used to assess neighborhood residential properties impacted by a landfill.

Complicating Factor(s)	Despite laboratory processing activities, considerable variability was observed for BaP results in subsampling replicates. Efforts were made early in the project to troubleshoot the problem, but no procedural modifications were found that consistently controlled subsampling variability. The weight of evidence in the data collected suggested that the problem may be caused by irreducible particle effects. The full discussion of the evidence for particle effects causing data variability was deleted to shorten the case study.

Case Study Conclusions

This unpublished technical memorandum serves as a description of ISM techniques used to evaluate DU compliance with a site-specific cleanup level of BaP and assess neighborhood residential properties impacted by a landfill.

Case Study #4: Confirmation Sampling from Utility Pole Storage

ISM Concept Demonstrated: Decision matrix for using DU mean, DU 95% UCL, and CV

Case Study Name: *Confirmation Soil Sampling: Remedial Action Report, FairPoint Communications, Utility Pole Storage Area, 11 Mallet Park Road, Brunswick, Maine*

Author(s): St. Germain Collins

Date: February 9, 2017

Contaminant(s) of Concern: PCPs, dioxin, metals, and PAHs

Media of Concern: Soil (excavation surface)

Case Study Link: https://www.itrcweb.org/FileCabinet/GetFile?fileID=17102

Background

ISM was used to confirm that remedial actions (such as excavation) removed impacted soils in a utility pole laydown yard. ISM was used to first establish background concentrations for chromium, copper, arsenic (CCA), and PAHs. Excavation of impacted soils was then conducted, removing approximately 3,047 tons of soil in an area approximately 345 ft x 30 ft. The excavation was split into six DUs, and ISM was used to collect confirmation samples from each DU. Those results were then compared to either background and/or Residential Remedial Action Guidelines (RAGs) for COCs.

Site Summary

Regulatory Agency/Program	Maine Department of Environmental Protection (MEDEP)
Description	Utility Pole Storage Area, Brunswick, Maine
Owner/Responsible Party	FairPoint Communications
Other Stakeholders	NA

Site Complexities

Risks	Surface soil exposure (dermal, ingestion, inhalation of particles)
Characteristics	Laydown area for utility poles
Other	The MEDEP approved the response action plan, which proposed the use of ISM for confirmation soil sampling and included a layout of SUs and DUs.

Planning and Implementation

Field Sampling Conducted	ISM was used to collect three background samples (3 DUs) for CCA and PAHs. The excavation boundaries were divided into nine SUs (floor of the excavation had three SUs and sidewalls had 6 SUs). For two specific areas of the excavation, several of the SUs were combined and a total of six DUs to further evaluate the data. Confirmation samples were collected using the USACE multi-increment sampling tool (CMIST) for analysis of dioxins, PCP, and CCA. Each SU sample consisted of 30 increments and collected in three replicates.
DUs	Three DUs for background samples Six DUs for confirmation (two excavation floor samples; four sidewall samples)
Statistical Sampling Design	Comparison of average of three replicates for each SU to remediation goal (either background or RAG). For DUs, 95% UCL of mean concentration for dioxin and CCA for comparison in a pre-approved decision matrix.
Laboratory Processing and Analysis Summary	Nothing reported on laboratory processing. Dioxins analyzed in accordance with USEPA Method 8290. PCP analyzed in accordance with USEPA Method 8270. CCA analyzed in accordance with USEPA Method 6010.
Data Quality Summary	Standard laboratory QA/QC for analyses; data were acceptable for intended use.
Level of Effort (Cost/Benefit, Effectiveness, or ROI)	NA

Decision-making

Outcome and Lessons Learned	Excavation successfully removed impacted soils, and ISM provided more accurate confirmation sampling results. No further remediation was necessary.
Complicating Factor(s)	NA

Case Study Conclusions

The Remedial Action Report documented successful remediation activities to remove surface soils impacted from utility pole laydown operations. Soils were excavated and confirmation soil samples were collected using ISM, with results demonstrating that concentrations were either at below background or RAGs.

Case Study #5: Investigation of a Former Dump Area on an Inaccessible Island

ISM Concept Demonstrated: Incremental sampling conducted successfully at multiple depth intervals. Statistical analyses that determined standard sample collection methodology would have provided unreliable results.

Case Study Name: *Evaluation of Green Island Landfill and Reburial Pit, Former U.S. Coast Guard LORAN Station Kure*

Author(s): Element Environmental, LLC, Haleiwa, HI; prepared for U.S. Coast Guard (USCG)

Date: June 2009

Contaminant(s) of Concern: PCBs and metals (arsenic, cadmium, chromium, lead, and mercury)

Media of Concern: Surface and subsurface soil and sediment

Case Study Link: https://www.itrcweb.org/FileCabinet/GetFile?fileID=19272

Background

Kure Atoll includes one vegetated island (Green Island) that is approximately 1.5 miles long and 0.35 miles wide. The USCG

operated a LORAN station on Green Island from 1960 to 1992, and waste generated during operations were buried at the southwest end of the island in a scrap metal dump (SMD). It is assumed that scrap metal and electrical components containing hazardous materials such as PCBs (capacitors, batteries, and transformers) were disposed at this location from the early 1960s to the late 1970s. Investigations predating this case study were conducted from 1991 through 1994, and a response action ordered the excavation of approximately 800 yds of soil from the SMD in 1993. The excavated soil was placed in a reburial pit near the center of the island with the exception of off-site disposal of 36 yds^3 of soil.

This report summarizes a follow-on investigation conducted in 2008 that included incremental sampling conducted in and surrounding the SMD.

Site Summary

Regulatory Agency/Program	NA
Description	Kure Atoll, Green Island
Owner/Responsible Party	USCG
Other Stakeholders	None

Site Complexities

Risks	Isolated location, no permanent residents or workers
Characteristics	Predominately water deposited coral sands, so Green Island is subject to drastic changes in sand deposits depending on wave action.
Other	A PCB hotspot identified in the SMD from immunoassay field testing of discrete samples resulted in the addition of a DU.

Planning and Implementation

Field Sampling Conducted	Near surface samples collected with trowel. Below near surface to 36 in collected by digging with a shovel to depth and collection of samples with trowel. Samples below 36 in collected with slide hammer-driven soil probe with acetate liner.
DUs	Eighteen total DUs: Ten located outside of SMD DU-1 through DU-8 had a sample depth of 4 inDU-9 and DU-10 were duplicates of DU-8Five located across entire SMDDU-11 sample depth of 4 inDU-12 sample depth of 36 inDU-13 sample depth of 60 inDU-14 and DU-15 were duplicates of DU-13Two located in SMD hotspot with samples from each analyzed for four separate grain sizesDU-17 sample depth of 36 inDU-18 sample depth of 60 in
Statistical Sampling Design	NA
Laboratory Processing and Analysis Summary	Laboratory analytical methods followed USEPA Publication SW-846
Data Quality Summary	Acceptable

Level of Effort (Cost/Benefit, Effectiveness, or ROI)	None provided

Decision-making

Outcome and Lessons Learned	ISM was useful for determining a representative mean level of contamination within a DU. Additional homogenization of samples in laboratory prior to selection and extraction of samples for analysis was critical to obtain representative results due to some isolated hot spots or nuggets of contamination in SMD.
Complicating Factor(s)	None

Case Study Conclusions

A draft unpublished summary report specific to an investigation scale evaluation of the incremental sampling conducted for this investigation is provided as Appendix C of the report. Statistical analyses determined traditional sampling would have resulted in incorrect conclusions versus the incremental sampling that was conducted.

Analysis of four separate grain sizes collected from the two hotspot DUs indicated that the most elevated PCB concentrations were associated with finer grained fractions.

Case Study #6: Pesticide Mixing and Loading Site

ISM Concept Demonstrated: Creation and sampling of 59 DUs in a 400+-acre residential development area with an additional 12 DUs focused on a chemical handling and spill area

Case Study Name: East Kapolei Final Site Assessment and Site Investigation Reports, Kapolei, Oahu, Hawaii

Reports:

1. *East Kapolei Affordable Housing Project, Final Site Assessment Report*, prepared by Tetra Tech EM, Inc., for Hawaii State Department of Health (HDOH), December 12, 2007.
2. *East Kapolei Pesticide and Mixing and Loading Site, Site Investigation Report and Environmental Hazard Assessment,* prepared by EnviroServices & Training Center, LLC, for Hawaii State Department of Hawaiian Home Lands (DHLL), March 2010.

Author(s): Tetra Tech EM Inc.; prepared for Hawaii State Department of Health (HDOH).

EnviroServices & Training Center, LLC; prepared for DHLL.

Contaminant(s) of Concern: Agricultural property – arsenic, dioxin, herbicides, and pesticides; pesticide mixing and loading area (PML) – arsenic, dioxin, and pesticides.

Media of Concern: Surface and subsurface soil

Case Study Links:
https://www.itrcweb.org/FileCabinet/GetFile?fileID=19907
https://www.itrcweb.org/FileCabinet/GetFile?fileID=19121

Background

Investigations of former sugarcane fields and a separate PML area were conducted to identify COCs prior to potential residential and commercial development. The site was used for sugarcane cultivation from approximately 1890 to 1994, with the PML area used for the storage, mixing, and loading of agricultural chemicals for approximately 40 years until sugarcane operations ceased while the remaining field were subject to the general usage of these chemicals. It was suspected that occasional spills of products within the PML area may have resulted in localized contamination of soils. The reports summarize the investigation and assessment of the sugarcane fields and of the PML area, respectively.

Site Summary

Regulatory Agency/Program	Results compared to HDOH health environmental ALs
Description	East Kapolei, agricultural property, 413 acres East Kapolei, PML, 0.63 acres
Owner/Responsible Party	State of Hawaii
Other Stakeholders	NA

Site Complexities

Risks	Area to be redeveloped for residential purposes.
Characteristics	Clay alluvium derived from igneous rocks.
Other	Stand-alone PML area was within overall site boundaries.

Planning and Implementation

Field Sampling Conducted	Surface soil samples (0 to 0.5 ft) collected with a trowel. Subsurface samples in PML spill area DUs were collected using direct push drilling equipment. Subsurface samples in PML non-spill area DUs were gathered by trenching with a backhoe and collecting samples with a trowel.
DUs	Agricultural fields: Fifty-nine total DUs randomly were selected in each of the 7-acre grids, representing the approximate full extent of the agricultural fields. Each DU was a 5,000-ft^2 area in order to represent a typical residential lot. A total of 40 increments were collected from within each DU. PML area: Twelve surface DUs ranging in area from 1,000 to 5,000 ft^2 were selected, with three DUs representing likely spill areas. A total of 50 increments were collected in each DU.Three spill area DUs also included sampling of subsurface soil from the following three layers: 0.5 to 2 ft, 2 to 5 ft, and 5 to 10 ft. A total of 20 increments were collected from each layer.Five of the general surface DUs were randomly selected to also include subsurface soil from the following two layers: 0.5 to 2 ft and 2 to 3 ft. A total of 50 increments were collected from each layer.
Statistical Sampling Design	NA
Laboratory Processing and Analysis Summary	Laboratory analytical methods followed USEPA Publication SW-846
Data Quality Summary	Acceptable
Level of Effort (Cost/Benefit, Effectiveness, or ROI)	None provided

Decision-making

Outcome and Lessons Learned	None noted
Complicating Factor(s)	None noted

Case Study Conclusions

The investigations confirmed that ISM identified small hot spots as well as overall contaminant heterogeneity within the areas investigated. There were no concentrations of COCs in the soil that indicated conditions were not suitable for residential reuse. Additional sampling and evaluation did not appear necessary.

Case Study #7: ISM for Petroleum Hydrocarbons

ISM Concepts Demonstrated: ISM does not underestimate plume extents or magnitude of contamination and is a better predictor of contaminant mass and exposure for risk assessment. For sampling from boring cores, 30 plugs for ISM performed better than wedge sampling.

Case Study Name: Incremental Sampling Methodology for Petroleum Hydrocarbon Contaminated Soils: Volume Estimates and Remediation Strategies

Author(s): Kathlyne Hyde, Wai Ma, Terry Obal, Kris Bradshaw, Trevor Carlson, Steven

Mamet, and Steven D. Siciliano

Date: October 10, 2018

Contaminant(s) of Concern: PHCs

Media of Concern: Surface/subsurface soil

Case Study Link: https://www.itrcweb.org/FileCabinet/GetFile?fileID=19885

Background

This study was conducted to investigate and compare both traditional discrete sampling through a Phase II ESA and ISM for estimating contaminated soil volume and developing management strategies at two sites. Two legacy gasoline and diesel bulk transfer stations located in Saskatoon and Raymore, Saskatchewan, Canada, with known spill and leak histories were selected for sampling. The objectives included estimating the lateral and vertical extent of PHC concentrations, quantifying and determining the causes for the differences in contaminated soil volumes estimated by each method, evaluating and comparing the precision of each method, and determining how to use the information gathered from both methods to manage the two contaminated sites.

Site Summary

Regulatory Agency/Program	NA; analytical data were compared to Saskatchewan Tier I guidance from the Saskatchewan Ministry of the Environment
Description	Two legacy gasoline and diesel bulk transfer stations located in Saskatoon and Raymore, Saskatchewan, Canada, with known spill and leak histories
Owner/Responsible Party	Not provided
Other Stakeholders	Department of Soil Science, University of Saskatchewan, Saskatoon, SK; Department of Scientific Services and Development, Maxxam Analytics, Mississauga, ON; Department of Sustainability, Federated Cooperatives Limited, Saskatoon, SK

Site Complexities

Risks	Potential exposure pathways not yet identified by a risk assessment
Characteristics	Highly heterogenous soils in western Canada formed from glacial till and freeze thaw conditions, which make estimating the average concentration of concern for a site difficult and unreliable.
Other	Supplementary material also available

Planning and Implementation

Field Sampling Conducted	The site areas were conceptually divided into four IAs encompassing the source, plume, plume delineation, and clean areas. Each IA contained three single borehole DUs in unbiased locations from which soil cores up to 7.5 m in depth were taken using direct push core drilling with a Geoprobe® 7822DT. Single borehole DUs had two co-located boreholes within 0.5 m of each other, one for the Phase II ESA based on discrete sampling methods and one for ISM analysis. Traditional discrete samples were collected on site at depth increments of 0.5 m (for up to 6 or 7.5 m) from each initial borehole for the Phase II ESA. A single sample was submitted for analysis for each single borehole DU. The single or additional bias samples were taken based on visual contamination and odor, and whether there were high VOC readings on a photo-ionization detector (PID). Samples for volatiles analysis were collected with a 5-g Terra Core™ sampler and placed into a 40-mL VOC vial pre-charged with HPLC grade methanol. Approximately 200 g of soil was packed into 250-mL jars, from which a 5-g subsample was used for semi-volatiles analysis. From the co-located borehole, the 1.5-m acrylic tube segments were collected and sealed with paraffin wax on site. Cores were stored at -20°C prior to further subsampling in a laboratory setting.
DUs	Each IA contained three single borehole DUs in unbiased locations from which soil cores up to 7.5 m in depth were taken using direct push core drilling. Single borehole DUs had two co-located boreholes within 0.5 m of each other, one for the Phase II ESA based on discrete sampling methods and one for the ISM analysis. Discrete samples were collected at depth increments of 0.5 m (for up to 6 or 7.5 m), and each single borehole DU was divided into three DU layers: (1) surface zone at 0 to 1.5 m, (2) estimated contaminated zone at 1.5 to 4.5 m, and (3) depth delineation zone at 4.5 to 6.0, or 7.5 m depending on the site. From each DU layer, the samples collected included (1) 30 plug increments to combine for one ISM sample, (2) a wedge sample collecting surface soil from the entire length of the core, and (3) a discrete sample from a biased hot spot.
Statistical Sampling Design	The study describes that ISM holds greater statistical power by overcoming the fundamental and distributional error associated with spatial soil heterogeneity and contaminant distribution. Statistical sampling design included false positive and false negative data analysis and 95% UCL calculations using the ITRC UCL calculator.

Laboratory Processing and Analysis Summary	The analytical data for both the Phase II ESA and ISM work were provided by Maxxam Analytics. Stored, frozen soil cores were used for analysis; the cores were thawed for sampling. From each DU, the team collected (1) 30 plug increments to combine for one ISM sample, (2) a wedge sample collecting surface soil from the entire length of the core, and (3) a discrete sample from a biased hot spot. The top layers of the cores were shaved off to expose fresh soil for sampling, and soils were sampled for C6-C10 hydrocarbons, BTEX (benzene, toluene, ethylbenzene, and xylenes), VOC analysis, and SVOC analysis.
Data Quality Summary	ISM holds greater statistical power by overcoming the fundamental and distributional error associated with spatial soil heterogeneity and contaminant distribution. In comparison to ISM, discrete sample data under- and overestimated contaminants.
Level of Effort (Cost/Benefit, Effectiveness, or ROI)	This study was conducted to investigate and compare both traditional discrete sampling through a Phase II ESA and ISM. Incremental samples identify a greater extent of contamination than discrete samples and reduced the occurrence of false positives, which reduced the quantity of mapped contamination. This would in turn reduce remediation costs.

Decision-making

Outcome and Lessons Learned	The results from both methods indicated that the sites were impacted with petroleum hydrocarbons above Saskatchewan Tier I guidance. The study showed that ISM does not underestimate the plume extent or underestimate the magnitude of contamination, and that current Phase II ESA methods are effective in identifying areas of potential concern, but they cannot provide robust estimates of contaminant mass. The study recommended that ISM be used as a remediation planning tool and that following a traditional Phase II ESA with an ISM-directed sampling approach would provide a statistically robust estimate of contaminant mass and exposure for risk assessment.
Complicating Factor(s)	None

Case Study Conclusions

This study was conducted to investigate and compare both traditional discrete sampling through a Phase II ESA and ISM for estimating contaminated soil volume and developing management strategies at two legacy gasoline and diesel bulk transfer stations. The study recommended that ISM be used as a remediation planning tool and that following a traditional Phase II ESA with an ISM-directed sampling approach would provide a statistically robust estimate of contaminant mass and exposure for risk assessment. Supplementary material is also available for review in addition to the published article.

Case Study #8: Remedial Action Report, Utility Pole Storage Area

ISM Concept Demonstrated: Figures with site plan development, grids, and locations of samples within grids.

Case Study Name: Confirmation Soil Sampling: Pole Storage Area Remediation Report, FairPoint Communications Facility, 104-106 Fairbanks Road, Farmington, Maine

Author(s): St. Germain Collins, Keith R. Taylor, and a Certified Geologist (C.G.)

Date: December 10, 2015

Contaminant(s) of Concern: PCPs, dioxin, and PAHs originating from utility pole preservatives

Media of Concern: Surface/subsurface soil

Case Study Link: https://www.itrcweb.org/FileCabinet/GetFile?fileID=17103

Background

The FairPoint Communications (FairPoint) facility in Farmington, Maine, included a utility pole storage area, and preservatives from the utility poles had contaminated surface and subsurface soils with PCPs, dioxin, and PAHs. Since FairPoint planned to end its lease and use of the site, a remedial action was put in place, the goal of which was to remove the soils exceeding MEDEP RAGs. Approximately 312 tons of soil were removed during excavation, and ISM was used during confirmation sampling to ensure that the remedial action had removed the soils exceeding MEDEP RAGs.

Site Summary

Regulatory Agency/Program	MEDEP
Description	Utility Pole Storage Area, Farmington, Maine
Owner/Responsible Party	FairPoint Communications
Other Stakeholders	NA

Site Complexities

Risks	Surface soil exposure (dermal, ingestion, inhalation of particles)
Characteristics	Laydown area for utility poles
Other	MEDEP approved the remedial action plan at the site, and confirmation sampling collection and analysis was completed in accordance with ITRC ISM guidance.

Planning and Implementation

Field Sampling Conducted	Upon completion of soil excavation, 10-ft x 10-ft sampling grids were established over the entire excavated area (sidewalls and floor), and three discrete soil samples were collected from each excavation bottom grid cell. Nine sidewall discrete samples were collected from each sidewall grid cell, as necessary. The discrete samples were combined to create three composite samples, which were then submitted for further processing. The final composite samples represented four DUs, and three replicate results were used to provide the mean concentration for the DU.
DUs	Four DUs: DU-1 (crib excavation bottom), DU-2 (crib excavation sidewalls), DU-3 (debris pile excavation bottom), and DU-4 (debris pile excavation sidewalls)
Statistical Sampling Design	RPD were compared to measure precision, and the three replicate sample results were used to calculate the 95% UCL of the mean concentration of the contaminants. A total of 30 increments were collected per DU and in three replicates.
Laboratory Processing and Analysis Summary	Nothing specifically stated for laboratory processing. Dioxin screening used USEPA Method 4025M.
Data Quality Summary	Standard laboratory QA/QC was utilized for the analyses. Based on the RPD and the 95% UCL calculations, the data were suitable to conclude no further remediation was necessary.
Level of Effort (Cost/Benefit, Effectiveness, or ROI)	NA

Decision-making

Outcome and Lessons Learned	Excavation successfully removed impacted soils below accepted RAGs or below accepted statewide MEDEP background levels, MEDEP urban developed background, or MEDEP urban fill values.
Complicating Factor(s)	NA

Case Study Conclusions

The Remedial Action Report documented successful remediation activities to remove surface and subsurface soils impacted from utility pole laydown operations. Soils were excavated and confirmation soil samples collected using ISM. The results demonstrated that concentrations were below MEDEP RAGs, below detection limits, below MEDEP urban developed background levels, or MEDEP urban fill values. The report concluded that no further remediation was necessary.

Case Study #9: State Park Incorporating Decommissioned Lighthouse and Residential Dwelling

ISM Concept Demonstrated: Example of an ISM WP and method for determining random increments

Case Study Name: *Anclote Key Lighthouse Assessment, Pinellas County, Florida*

Reports:

1. *Workplan for Characterization of Soil Lead and Mercury Levels in Support of Risk Assessment*, Anclote Key Lighthouse, prepared by the Florida Department of Environmental Protection, January 18, 2013.

- *Incremental Sampling Case Study Power Point Presentation*, Anclote Key Lighthouse, prepared by Florida Department of Environmental Protection, March 2013 (Draft).

Contaminant(s) of Concern: Lead and mercury

Media of Concern: Surface soil

Case Study Links:
https://www.itrcweb.org/FileCabinet/GetFile?fileID=16070
https://www.itrcweb.org/FileCabinet/GetFile?fileID=16071

Background

The Anclote Lighthouse was constructed in 1887, decommissioned in 1985, and currently sits within a state park. The paint used on the cast-iron lighthouse was lead-based, and over time, it was subject to erosion and chipping. In the 1960s the lighthouse was converted to battery power and casings from used batteries were discarded or stored in buildings near the lighthouse. More than 100 batteries were removed in 1994, and subsequent discrete sampling identified lead and mercury in surface soil above cleanup target levels. Areas to the east and south of the lighthouse will be subject to remediation based on discrete sampling results with additional assessment to be conducted on the remaining areas of concern, primarily north and west of the lighthouse.

Site Summary

Regulatory Agency/Program	Results compared to Florida Department of Environmental Protection cleanup target levels
Description	Decommissioned lighthouse, surrounding landscaped area, and an adjacent state park ranger residential dwelling
Owner/Responsible Party	State of Florida
Other Stakeholders	None

Site Complexities

Risks	Lighthouse area to be open to public visitors and also single residential dwelling for park ranger
Characteristics	Landscaped
Other	Multiple short-term visitors and a single dwelling for residents

Planning and Implementation

Field Sampling Conducted	ISM surface soil samples (0 to 0.5 ft) were collected following FDEP SOPs using PVC tubes or trowels. In addition to ISM: Discrete surface soil samples were collected for comparison of results to ISM surface soil results. Limited discrete subsurface samples were additionally collected for general characterization purposes only.
DUs	Two DUs based in part on data obtained from previous discrete sampling and site layout characteristics: DU1 – Northwest of lighthouse, 0.18 acres, divided into 32 16-ft x 16-ft cells. Three replicate samples were collected from within each cell, locations determined by systematic random selection. Each of the three replicates combined into three separate samples (32 combined subsamples each).DU2 – Area around park ranger residence, 0.5 acres, divided into 30 28-ft x 28-ft cells. Three replicate samples were collected from within each cell, locations determined by systematic random selection. Each of the three replicates combined into three separate samples (30 combined subsamples each). Discrete surface soil samples were collected from each DU cell for comparison to ISM results, and ISM processing was completed by the analytical laboratory of the submitted replicate samples following internal SOPs.
Statistical Sampling Design	95% UCL for comparison of discrete and ISM mercury results and grand mean for comparison of discrete and ISM lead results in each DU. ISM results were provided for both milled and un-milled samples.
Laboratory Processing and Analysis Summary	Laboratory analytical methods followed USEPA Publication SW-846
Data Quality Summary	Acceptable
Level of Effort (Cost/Benefit, Effectiveness, or ROI)	None provided

Decision-making

Outcome and Lessons Learned	Noted below
Complicating Factor(s)	None noted

Case Study Conclusions

ISM results indicated relative variability, most significantly for lead, in measured concentrations between replicates. This was noted in both milled and un-milled sample replicate sets, so milling of samples did not significantly influence the noted variability. The ISM samples that were milled resulted in generally higher measured concentrations of lead with less

significant variation in measured concentrations of mercury. Discrete grand mean lead concentrations specific to each DU were higher than both milled and un-milled ISM grand mean results. Discrete 95% UCL mercury concentration in DU1 was higher than both milled and un-milled ISM results, while the discrete 95% UCL mercury concentration in DU2 was lower than both milled and un-milled ISM results.

Case Study #10: Metals in Soils, Discrete versus ISM Costs Comparison

ISM Concept Demonstrated: A larger number of increments (30 versus 50) did not improve data quality. ISM sampling costs were less than collecting a large number of grab samples.

Case Study Name: *The Mineral Wool Site*

Author(s): Daniel B. Stephens & Associates, Inc.

Date: May 2015

Contaminant(s) of Concern: Antimony, copper, and lead

Media of Concern: Surface and subsurface soils

Case Study Link: https://www.itrcweb.org/FileCabinet/GetFile?fileID=19887

Background

This study was designed to compare ISM results to conventional investigation sampling results for metals concentrations in soils. Additionally, the study compared ISM results based on the number of increments collected within a DU. The study also compared costs from ISM versus standard sampling techniques. In general, the ISM samples were used to determine representative metal concentrations in each DU, and for the background DU, to develop site-specific background metals concentrations.

Site Summary

Regulatory Agency/Program	Texas Commission on Environmental Quality (TCEQ) – State Superfund
Description	Bell County, Texas
Owner/Responsible Party	NA
Other Stakeholders	NA

Site Complexities

Risks	NA
Characteristics	Former blow wool and batt wool manufacturing facility where aerial deposition, wastewater, and surface water runoff caused impacts to surface and subsurface soils.
Other	An initial investigation of the site had already been completed and identified impacted soils.

Planning and Implementation

Field Sampling Conducted	Both ISM and grab sampling were conducted, with grab samples collected from two of the six DUs. Boundaries were outlined in the field, and the DU was divided into evenly spaced grids with increment samples collected from the center of each grid. For the two DUs where grab samples were collected (DU-3 and DU-6), grab samples were collected in the center of the grids as well. For DU-3, 30 grab samples were collected, and for DU-6, 16 grab samples were collected.
DUs	Six DUs were established: DU-1 and DU-2 were for wastes, DU-3 through DU-5 were for impacted soils, and DU-6 was for background soils. For each DU sample, three replicate samples were collected with each sample having 30 increments. For DU-4, three replicate samples with 50 increments were also collected.
Statistical Sampling Design	For individual DUs, comparison was based on RPD, with an RPD <25% as the goal. For previously collected background data (grab samples), a UCL was calculated to compare to ISM results.
Laboratory Processing and Analysis Summary	Field sieving was conducted.
Data Quality Summary	NA
Level of Effort (Cost/Benefit, Effectiveness, or ROI)	A cost comparison matrix was developed to compare labor and analysis costs for ISM sampling to grab sampling. Grab sampling costs included costs to collect 1, 15, or 30 grab samples per DU compared to three replicate ISM samples with 30 increments per ISM sample. The ISM sampling costs were approximately three to five times less than collecting 15 or 30 grab samples per DU. ISM sampling costs were approximately four times higher than collecting one grab sample per DU.

Decision-making

Outcome and Lessons Learned	ISM did a better job addressing variance than the 16 background grab samples. For the 30 versus 50 increment ISM samples, the RPDs were compared and found to be similar. There does not appear to be a benefit for collecting additional increments for these samples.
Complicating Factor(s)	NA

Case Study Conclusions

For the cost evaluation, labor costs were higher for ISM sampling, but the labor costs for grab sampling begin to converge with ISM labor costs as the number of grab samples per DU increased. ISM sampling costs are lower mainly as a result of reduced laboratory costs.

Appendix B. Statistical Simulation Studies

B.1 INTRODUCTION

This section presents additional details about the simulation studies used to evaluate the performance of alternative ISM sampling strategies applied to DUs with a range of heterogeneities. Monte Carlo methods were used to collect hypothetical incremental samples following various spatial sampling protocols. The following factors were varied:

- number of increments
- range of variability
- number of replicates
- spatial patterns
- sampling methods
- methods of accounting for compositional and distributional heterogeneities
- sampling patterns
- choice of UCL calculation method

The following performance metrics were used to evaluate the influence of these factors on ISM results:

- coverage of UCL (absolute and relative bias in the estimate of the population mean)
- absolute and/or RPD between the UCL and true mean (SD of relative bias in the population mean)
- RSD of replicate means

The main advantage of simulations is that population parameters are known. Therefore, alternative sampling approaches and calculation methods can be explored for a wide range of scenarios. With each simulation, the same sampling method and/or calculations are performed many times, as if a hypothetical field crew repeated the sampling effort over and over. Because each sampling event involves random sampling from the population, no two hypothetical events yield identical results. However, by repeating the exercise many times, we generate a distribution of results from which we can evaluate the various performance metrics noted above.

Note that not every performance metric is captured in every simulation, in part because the simulations use different approaches to represent bulk material heterogeneity in a DU. Summary tables and discussions of each simulation clarify what metrics were evaluated and how this information can be used to guide the selection of ISM sampling protocols. None of the simulations attempt to explicitly define all seven sources of error in estimates of the mean associated with bulk material sampling (see Section 2.6). The simulations focus on representing the compositional and distributional heterogeneities (CH and DH) that can be attributed to FE and GSE.

Simulations were conducted with defined distributions (statistical or spatial) to represent the variability in sample value results that may be expected, given the combined effect of these errors. Simulations allow for the evaluation of different spatial sampling patterns that cannot be evaluated empirically because the true population parameters (such as population mean) are typically unknown. Naming conventions applied to each simulation experiment include a prefix PD for simulations with *probability distributions* and M for simulations with *maps*. The PD simulation approach involved randomly sampling from a two-parameter lognormal PD with a specified mean and variance. The ratio of the population parameters (SD divided by the mean), also known as the CV, provides a measure of variability that facilitates comparisons of results across a wide range of conditions. The M approach involved the use of maps (2D surfaces) to sample from alternative spatial distributions of soil contamination (M). Each set of maps has unique implementations that provide the ability to demonstrate a range of different DU conditions. The method to simulate the soil data for each set of maps follows:

- M-1 is based on a real dataset of more than 200 observations. The sample results were interpolated with inverse distance weighting techniques to yield a completely defined 2D surface of concentrations (see Section B.3).
- M-2 maps are based on real DU data composed of bulk materials. The patterns and concentration values are

established from extensive discrete data (100 increments per DU) gathered as a part of multiple ESTCP projects led by Jenkins and Hewitt (Jenkins et al. 2004, Hewitt et al. 2005) (Jenkins et al. 2004, Hewitt et al. 2005); and Qiao, Pulsipher, and Hathaway (Qiao et al. 2010) (Qiao et al. 2010) document the specific details for how the discrete data were used to establish the completely defined 2D surface of increment values shown Section B.4.

Collectively, the simulation studies presented in this appendix provide a preliminary set of results intended to facilitate the development of ISM sampling designs and corresponding statistical analyses. More detail and underlying assumptions of the different simulation approaches are identified below.

Simulations presented in this appendix refer to different scales of heterogeneity as being "small" and "large," and are not intended to imply a precise dimension for a DU in terms of acres. Instead, the terms are relative to the size of the DU. Small scale refers to the immediate vicinity of the incremental sample, whereas large scale refers to the overall spatial extent of the DU.

B.1.1 Summary of Simulation Findings

Table B-1 summarizes the observations and conclusions from the various statistical simulations that were conducted.

Table B-1. Summary of findings from simulation experiments using PDs (PD) and maps (M).

Source: 2012, ITRC ISM Team.

Effects of the Number of Increments and Replicates on the Estimate of the Mean	
1	Increasing the number of increments and/or replicates reduces variability in the estimate of the mean.
2	Variability in the grand mean (the mean of the replicate incremental sampling estimates of the mean) is a function of the total number of increments collected (increments x replicates).
3	DUs with high heterogeneous contaminant concentrations have greater variability in the estimate of the mean and greater potential for errors in terms of both frequency and magnitude. Underestimates of the mean would be expected to occur more frequently than overestimates for heterogeneous sites with right-skewed contaminant concentration distributions. With equal numbers of samples (that is, individual discrete samples versus ISM replicates), the magnitude of error in estimating the mean would be expected to be lower using ISM.
4	The coverage of the 95% UCL depends on the total sample size (increments × replicates). For the typical number of increments of an ISM sampling design (30 to 100), increasing the number of ISM replicates above three provides marginal return in terms of improving coverage, but increasing the number of replicates decreases (or improves) the RPD, meaning that it will produce estimates of the 95% UCL closer to the DU mean.
5	Simulations produced varying results in terms of improvement in coverage by increasing the number of increments. As with increasing replicates, increasing the number of increments decreases (i.e., improves) the RPD.
6	Coverage provided by the two UCL calculation methods depends on the degree of variability of the contaminant distribution within the DU. For DUs with medium or high heterogeneity, the Student's-*t* method may not provide specified coverage. For DUs with high heterogeneity, the Chebyshev method may not provide specified coverage as well.
7	The Chebyshev method always provides a higher 95% UCL than the Student's-*t* method for a given set of ISM data with $r > 2$. When both methods provide specified coverage, the Chebyshev consistently yielded a higher RPD.
Effects of Sampling Pattern	
8	If the site is relatively homogeneous, all three field sampling patterns yield unbiased mean estimates, but the magnitude of error in the mean may be higher with simple random sampling compared to systematic random sampling. All sampling patterns yield similar coverages.

9	While all three sampling options are statistically defensible, collecting increments within the DU using simple random sampling is most likely to generate an unbiased estimate of the mean and variance according to statistical theory. From a practical standpoint, true random sampling is probably the most difficult to implement in the field and may leave large parts of the DU "uncovered," or without any increment sample locations. It should be noted that random does not mean wherever the sampling team feels like taking a sample, and a formal approach (typically based on a random number generator) to determining the random sample locations must be used.
10	Systematic random sampling can avoid the appearance that areas are not adequately represented in ISM samples and is relatively straightforward to implement in the field. Theoretically, it is inferior to simple random sampling for obtaining unbiased samples and can be more prone to producing errors in estimating the true mean, especially if the contamination is distributed in a systematic way. Random sampling within a grid is, in a sense, a compromise approach, with elements of both simple random and systematic sampling.
Subdividing the DU	
11	Sampling designs with this method yield unbiased estimates of the mean.
12	The principal advantage of subdividing the DU is that some information on heterogeneity in contaminant concentrations across the DU is obtained. If the DU fails the decision criterion (that is, it has a mean or 95% UCL concentration above a soil action limit), information will be available to indicate whether the problem exists across the DU or is confined to guide redesignation of the DU and resampling to further delineate areas of elevated concentrations.
13	Partitioned DU SE estimates are larger than those from replicate data if the site is not homogeneous. Hence, 95% UCL estimates from a subdivided DU will be as high or higher than those obtained from replicate measurements collected across the DU. The higher 95% UCLs improve coverage (generally attain 95% UCL) and increase the RPD. These increases occur if unknown spatial contaminant patterns are correlated with the partitions. In most cases, the Student's-*t* method provides adequate coverage.
RSD	
14	Datasets with a high RSD are more likely to achieve specified coverage for 95% UCL than datasets with low RSD. This tendency is explained by the greater variability among replicates leading to higher 95% UCL values, resulting in better coverage.
15	A low RSD does not ensure specified coverage by the 95% UCL or low bias in a single estimate of the mean. The opposite is, in fact, the case. For situations in which the UCL or one replicate mean is less than the true mean, the underestimate increases as RSD decreases.

The simulation findings presented in this appendix do not represent the totality of simulation exercises conducted as part of this project. It is anticipated that additional research may be needed to further investigate the performance of alternative ISM sample designs.

B.2 Probability Distributions (PDs)

A series of Monte Carlo simulations was run using PDs with different CVs. Table B-2 summarizes distribution variability (based on CV) and results for selected sampling designs and performance metrics (both Student's-*t* and Chebyshev UCLs).

Each scenario can be thought of as a special case of the simulations with maps (M-1 and M-2), presented later in this appendix. With sampling from PDs, each increment is an independent random sample obtained from the same defined distribution (that is, they are identically distributed), which is analogous to using simple random sampling for increment collection for an actual site. The DU is assumed to consist of a single population of lognormally distributed concentrations. It is important to note that, while this approach is useful for conveying important concepts about ISM, sampling from a PD is an oversimplification for the following reasons:

- There is no attempt to quantify the relative contributions of different sources of heterogeneity or errors introduced in both the field and laboratory. The variance is viewed as a lumping term that represents the variability in concentrations in soil if the site were divided into samples of some mass. In practice, the expected error in the estimate of the mean depends in part on the mass of soil collected with each increment (see discussion of Gy's sampling principles in Section 2.6). Therefore, it is convenient to think of the population as having a fixed mean concentration but also a variance contingent on the sample mass. The simulations with defined distributions do not explore the effect of sample mass on performance metrics. Instead, it is assumed that the specified variance simply reflects the collective sources of heterogeneity.
- The defined populations used in the simulations are not described as representing a DU of a specific size. At many sites, it is common for concentrations to exhibit spatial patterns, including subareas of elevated concentrations and overlapping sources (that is, mixtures). This may be true even for very small DUs where concentrations from samples collected within a 1-ft radius differ by more than an order of magnitude. Most of the simulations do not explicitly model these conditions but instead presume that the overall population for the DU can be approximated by a lognormal distribution, regardless of any spatial arrangement of the contaminant mass.
- Only lognormal PDs are defined, and alternative positively skewed PDs were not explored. In general, because lognormal distributions give greater weight to results in the upper tail than alternative choices (for example, gamma or Weibull distribution), the SE for the mean and the corresponding UCLs tends be higher than that of comparable distributions with the same population mean and variance.

Table B-2. Summary of simulation results using lognormal distributions (* %itl = percentile).

Source: 2012, ITRC ISM Team.

95% UCL >= true mean [overestimate of mean]

Statistic*	**Chebyshev 95% UCL**				**Student's-*t* 95% UCL**			
	2 reps	**3 reps**	**5 reps**	**7 reps**	**2 reps**	**3 reps**	**5 reps**	**7 reps**
	m=30, CV=1							
count of simulations	4,571	4,835	4,956	4,981	4,693	4,664	4,689	
95% UCL coverage	91%	97%	99%	100%	94%	93%	94%	94%
mean RPD	27%	22%	18%	16%	37%	16%	10%	8%
5th %ile RPD	3%	4%	5%	5%	4%	2%	1%	1%
50th %ile RPD	22%	21%	17%	15%	31%	14%	9%	7%
95th %ile RPD	65%	48%	34%	28%	91%	34%	20%	15%
	m=30, CV=4							
count of simulations	4,346	4,690	4,852	4,909	4,519	4,430	4,333	4,351

95% UCL coverage	87%	94%	97%	98%	90%	89%	87%	87%
mean RPD	93%	80%	63%	55%	129%	57%	36%	28%
5th %ile RPD	6%	8%	9%	10%	9%	4%	3%	2%
50th %ile RPD	65%	59%	50%	44%	90%	41%	27%	22%
95th %ile RPD	272%	214%	155%	129%	374%	157%	92%	73%
	m=30, CV=7							
count of simulations	4,171	4,532	4,740	4,820	4,414	4,187	4,101	4,137
95% UCL coverage	83%	91%	95%	96%	88%	84%	82%	83%
mean RPD	140%	117%	94%	83%	189%	86%	55%	45%
5th %ile RPD	8%	8%	9%	11%	11%	5%	4%	3%
50th %ile RPD	82%	73%	65%	59%	111%	54%	36%	30%
95th %ile RPD	457%	358%	271%	227%	609%	272%	164%	133%
	m=100, CV=1							
count of simulations	4,604	4,827	4,946	4,979	4,720	4,690	4,687	4,669
95% UCL coverage	92%	97%	99%	100%	94%	94%	94%	93%
mean RPD	27%	23%	18%	16%	38%	16%	10%	8%
5th %ile RPD	3%	5%	5%	5%	4%	3%	2%	1%
50th %ile RPD	22%	21%	18%	15%	32%	14%	9%	7%
95th %ile RPD	66%	49%	34%	28%	93%	35%	20%	15%
	m=100, CV=4							

count of simulations	4,358	4,674	4,858	4,926	4,547	4,435	4,375	4,395
95% UCL coverage	87%	93%	97%	99%	91%	89%	88%	88%
mean RPD	95%	79%	64%	55%	130%	57%	36%	28%
5th %ile RPD	6%	9%	10%	10%	9%	5%	3%	2%
50th %ile RPD	65%	59%	51%	45%	89%	41%	27%	22%
95th %ile RPD	280%	211%	157%	129%	380%	155%	95%	73%
	m=100, CV=7							
count of simulations	4,115	4,509	4,739	4,839	4,362	4,186	4,092	4,119
95% UCL coverage	82%	90%	95%	97%	87%	84%	82%	82%
mean RPD	135%	114%	93%	80%	183%	84%	54%	43%
5th %ile RPD	7%	8%	9%	9%	10%	5%	3%	3%
50th %ile RPD	82%	74%	64%	58%	111%	53%	36%	30%
95th %ile RPD	417%	321%	251%	210%	557%	240%	156%	122%

Chebyshev 95% UCL				**Student's-*t* 95% UCL**			
2 reps	**3 reps**	**5 reps**	**7 reps**	**2 reps**	**3 reps**	**5 reps**	**7 reps**
m=30, CV=1							
4,678429	165	44	19	307	336	311	322
91%	97%	99%	100%	94%	93%	94%	94%
-4%	-3%	-2%	-1%	-4%	-3%	-2%	-2%
-11%	-7%	-7%	-3%	-11%	-8%	-6%	-5%
-4%	-2%	-1%	-1%	-4%	-2%	-2%	-1%

0%	0%	0%	0%	0%	0%	0%	0%
m=30, CV=4							
654	310	148	91	481	570	667	649
87%	94%	97%	98%	90%	89%	87%	87%
-13%	-10%	-7%	-6%	-13%	-10%	-8%	-6%
-30%	-23%	-18%	-15%	-30%	-25%	-19%	-17%
-12%	-8%	-6%	-5%	-11%	-8%	-6%	-5%
-1%	-1%	0%	0%	-1%	-1%	-1%	-1%
m=30, CV=7							
829	468	260	180	586	813	899	863
83%	91%	95%	96%	88%	84%	82%	83%
-18%	-13%	-10%	-8%	-18%	-14%	-11%	-9%
-39%	-31%	-24%	-21%	-41%	-32%	-28%	-23%
-16%	-11%	-8%	-6%	-16%	-12%	-9%	-8%
-2%	-1%	0%	-1%	-1%	-1%	-1%	-1%
m=100, CV=1							
396	173	54	21	280	310	313	331
92%	97%	99%	100%	94%	94%	94%	93%
-4%	-3%	-2%	-1%	-5%	-3%	-2%	-2%
-12%	-8%	-5%	-3%	-12%	-9%	-6%	-5%
-3%	-2%	-2%	-1%	-4%	-3%	-2%	-1%
0%	0%	0%	0%	0%	0%	0%	0%

m=100, CV=4							
642	326	142	74	453	565	625	605
87%	93%	97%	99%	91%	89%	88%	88%
-13%	-10%	-6%	-6%	-13%	-10%	-7%	-6%
-30%	-23%	-18%	-18%	-31%	-25%	-18%	-16%
-11%	-8%	-5%	-5%	-12%	-9%	-6%	-5%
-1%	-1%	0%	0%	-1%	-1%	0%	0%
m=100, CV=7							
885	491	261	161	638	814	908	881
82%	90%	95%	97%	87%	84%	82%	82%
-17%	-13%	-9%	-8%	-17%	-14%	-11%	-9%
-39%	-29%	-22%	-20%	-38%	-31%	-26%	-23%
-15%	-11%	-8%	-6%	-15%	-12%	-9%	-8%
-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%

B.2.1 Methods

Monte Carlo analysis was used to repeatedly apply a specified sampling design (number of increments and ISM samples) to a DU scenario. Typically, between 5,000 and 30,000 trials were used, with the large number of trials expected to yield relatively stable (that is, reproducible) results. Each trial represents a complete sampling event (*n* increments and *r* replicates) and yields an estimate of the population mean, the SE of the mean, and the 95% UCL. Collectively, the results yield a distribution of 95% UCLs that can be used to calculate performance metrics – for example, ideally, the sampling method and UCL calculation yield a PD of 95% UCLs with a 5th percentile equal to (or greater than) the true population mean. This would mean that we can expect the sampling design applied to this type of population to achieve the desired coverage (or percentage of exceedances of the true mean) of 95%. Table B-2 provides examples of simulation experiments with coverages that vary from approximately 80 to 100%.

Multiple ISM samples (or replicates) must be collected to calculate the SE and UCL. The expected small sample sizes (three to seven replicates) for most implementations of ISM preclude the use of bootstrap resampling techniques to calculate a UCL, so simulations were performed using only the Student's-*t* and Cheybshev UCL methods, which are based on sample size, sample mean, and variance. Because the distribution of sample means tends to exhibit less skew than the population due to the CLT, the performance of the Student's-*t* UCL can vary, but Student's-*t* can be expected to yield the most reliable performance metrics for populations with a low (≤1) CV. By contrast, Chebyshev generally yields higher UCLs with higher

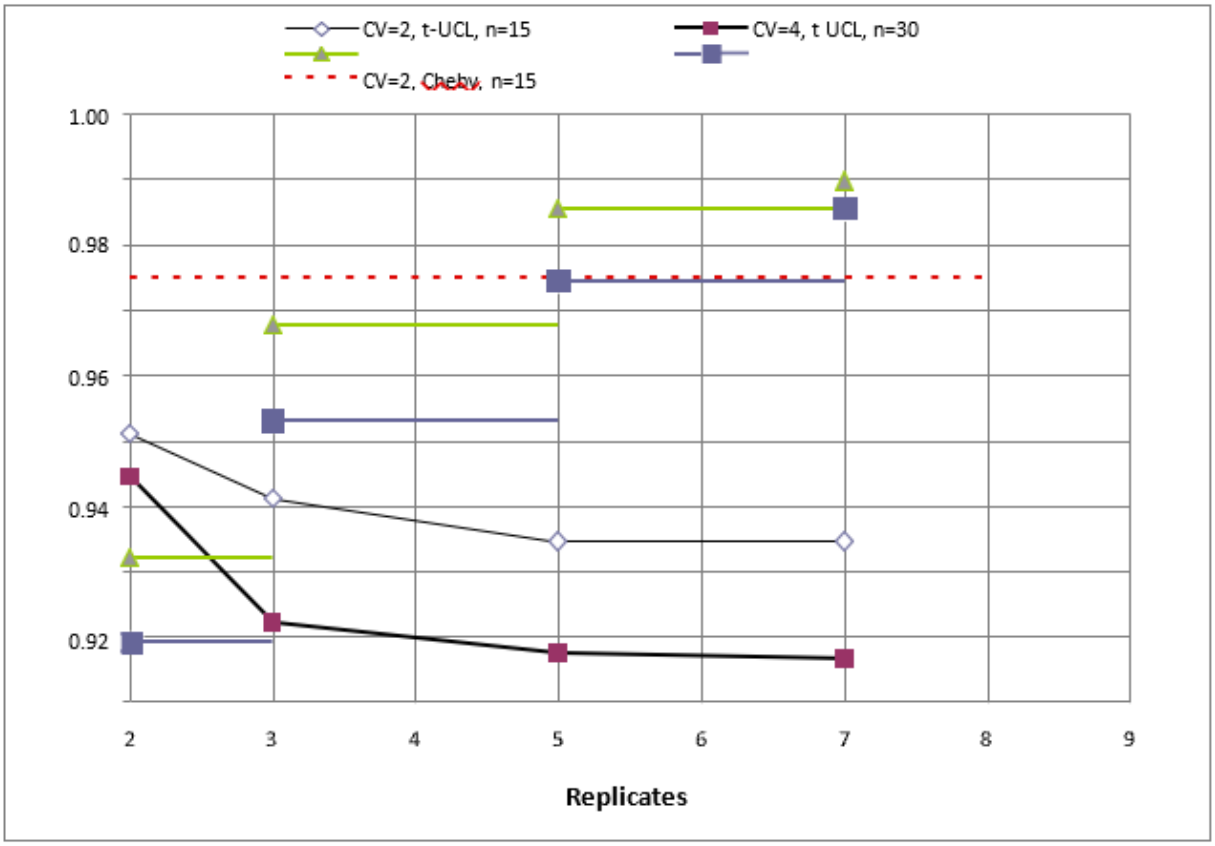
coverage but also higher RPDs. RPD = [(UCL – μ)/100] ´ 100%, where μ denotes the true DU (population) mean.

Generally, sampling designs were varied between 15 and 100 increments and between two and seven replicate ISM samples. The mean of the distribution represents the population mean and is used to calculate the bias and RPD metrics.

The number of replicates is used to represent the degrees of freedom in UCL calculations using ISM.

B.2.2. Results

Figure B-1 illustrates how the coverage of the 95% UCL varies for the Student's-*t* and Chebyshev UCL equations for a range of sampling designs applied to lognormal distributions with a range of variability. The table below the graph gives the coverage statistics as well as the average RPD (based on the full distribution of UCLs calculated).



CV	*n*	*r*	Student's-*t* 95% 95% UCL		Chebyshev 95% 95% UCL	
			Coverage	Mean RPD	Coverage	Mean RPD
2.0	15	2	90.2%	156%	86.4%	108%
2.0	15	3	88.2%	66%	93.5%	99%
2.0	15	5	86.9%	40%	97.1%	82%
2.0	15	7	86.9%	32%	97.9%	72%
2.0	30	2	91.3%	139%	88.5%	96%
2.0	30	3	88.9%	61%	94.3%	91%
2.0	30	5	87.7%	39%	97.1%	79%
2.0	30	7	87.2%	31%	98.2%	70%
4.0	15	2	85.3%	237%	82.8%	163%
4.0	15	3	81.3%	102%	90.8%	152%
4.0	15	5	79.8%	63%	95.7%	129%
4.0	15	7	80.1%	51%	97.2%	115%
4.0	30	2	88.9%	129%	83.8%	187%
4.0	30	3	84.4%	119%	90.6%	80%
4.0	30	5	83.5%	100%	94.9%	49%
4.0	30	7	83.3%	90%	97.1%	40%

Figure B-1. Examples of simulation results using lognormal PDs with CV equal to two and four, increments of 15 and 30, replicates ranging from two to seven, and two 95% UCL calculation methods (Cheby = Chebyshev; t-UCL = Student's-t).
Source: 2012, ITRC ISM Team.

These examples are useful for illustrating the following general patterns that emerge from the simulation experiments with lognormal distributions:

- The Chebyshev UCL generally yields higher coverage than the Student's-*t* UCL, with the exception of scenarios in which two replicates ($r = 2$) are selected. The upper critical value of the Student's-*t* distribution (that is, the t-value) varies with the degrees of freedom (df = $r - 1$), as noted below. For $r = 2$, the t-value is 6.3, which introduces an additional factor of two or more to the calculation of the 95% UCL compared to sampling designs with three or more replicates.

Table B-3. 95th Percentiles of Student's t Distribution

Source: 2020, ITRC ISM Update Team.

Replicates	df = *r* – 1	*t*-value for alpha = 0.05
2	1	6.3
3	2	2.9
4	3	2.4
5	4	2.1
6	5	2.0
7	6	1.9

- The coverage of the Chebyshev UCL generally increases with increasing sample sizes (increments and replicates) but with diminishing returns. The table below lists examples of combination of replicates and increments that can be expected to yield approximately 95% coverage. The coverage of the Student's-*t* UCL generally does not achieve 95% and does not increase with increasing samples sizes (increments and replicates) within a practical range.

Table B-4. Coverage of the Chebyshev UCL

Source: 2020, ITRC ISM Update Team.

CV	Increments	Replicates	Coverage	CV	Increments	Replicates	Coverage
1	15	3	96%	4	30	4	94%
	30	3	97%		50	4	95%
2	15	3	93%		100	3	93%
	15	4	95%		100	4	96%
	30	3	94%	7	30	5	95%
	30	4	96%		100	5	95%

	15	5	95%	
	30	4	95%	
	50	4	96%	
	100	3	95%	

- The RPD between the 95% UCL and the population mean is generally greater for Chebyshev than Student's-*t*, particularly for trials in which the 95% UCL actually exceeds the population mean. Therefore, the trade-off with the Chebyshev UCL is that it achieves more reliable coverage but also higher UCLs.
- The simulations with lognormal distributions yield unbiased estimates of the mean.

B.3 SPATIAL AUTOCORRELATION MAPS (M-1)

For most sites, contaminants in soil exhibit some degree of spatial relationship, meaning that variance in the concentration often reduces as the distance between sample locations decreases. It is well established that strong spatial relationships can reduce the effective sample size of a dataset because each sample provides some redundant information (Cressie 1993). In statistical terms, this redundancy violates the assumption that observations are independent. ISM CIs generated from spatially related data can be too narrow, resulting in a higher frequency of decision errors. Spatial relationships may also introduce bias in estimates of the mean and variance, depending on the sampling protocol. Bias can be reduced by using a truly random sampling strategy (for example, simple random sampling). The issue of spatial relationships applies to discrete as well as ISM sampling.

B.3.1. Methods

Simulations were run to evaluate the effect of spatial autocorrelation on the performance of ISM. Figure B-2 shows a map generated from a real dataset of more than 200 observations. The sample results were interpolated with inverse distance weighting techniques to yield a 2D surface of concentrations. Such spatial smoothing is likely to underestimate the distributional heterogeneity in concentrations that exists at most sites, so the results with ISM may underestimate the variance. Four ISM sampling protocols were applied to this map, assuming the map represents a single DU:

- systematic grid with a random start location (no division of the DU)
- systematic grid with a random start location (division of DU into quadrants)
- simple random sample (no division of the DU)
- simple random sample (division of the DU into quadrants)

For the scenario in which the site is divided into quadrants, each quadrant was sampled with the specified number of replicates, which means simulations with quadrants represent an overall fourfold increase in the sampling effort. Alternative evaluations of the quadrant scenario were evaluated with different maps to illustrate the performance metrics for quadrants in which a single ISM sample is collected from each quadrant, yielding a total sample size of $r = 4$.

B.3.2. Results

Table B-3 summarizes the simulation results with 1,000 Monte Carlo trials using 30 increments and three, five, and seven replicates. The distribution is only mildly skewed (CV = 0.7), and the autocorrelation is high (Moran's I z-score = 3.8). The following observations are noted:

- The spatial autocorrelation does not affect the coverage of either the simple random sampling or systematic grid sampling. With 30 increments and three replicates, Chebyshev yields 96 to 97% coverage, whereas Student's-*t* yields 94% coverage.
- As noted with the simulations using lognormal distributions, increasing the number of replicates results in a

higher coverage for the Chebyshev UCL but generally no improvement in the Student's-*t* UCL.

- The average RPD for the 95% UCL is lower by approximately a factor of two with systematic grid sampling, but introducing spatial autocorrelation tends to result in an improvement in this metric, most likely because autocorrelation affects the correlation between the sample mean and variance. For non-normal distributions, simple random sampling yields a positive correlation between the sample mean and sample variance. When systematic grid sampling is applied to a scenario with high spatial autocorrelation, it is more likely that neighboring samples share similar values, thereby reducing the sample variance.

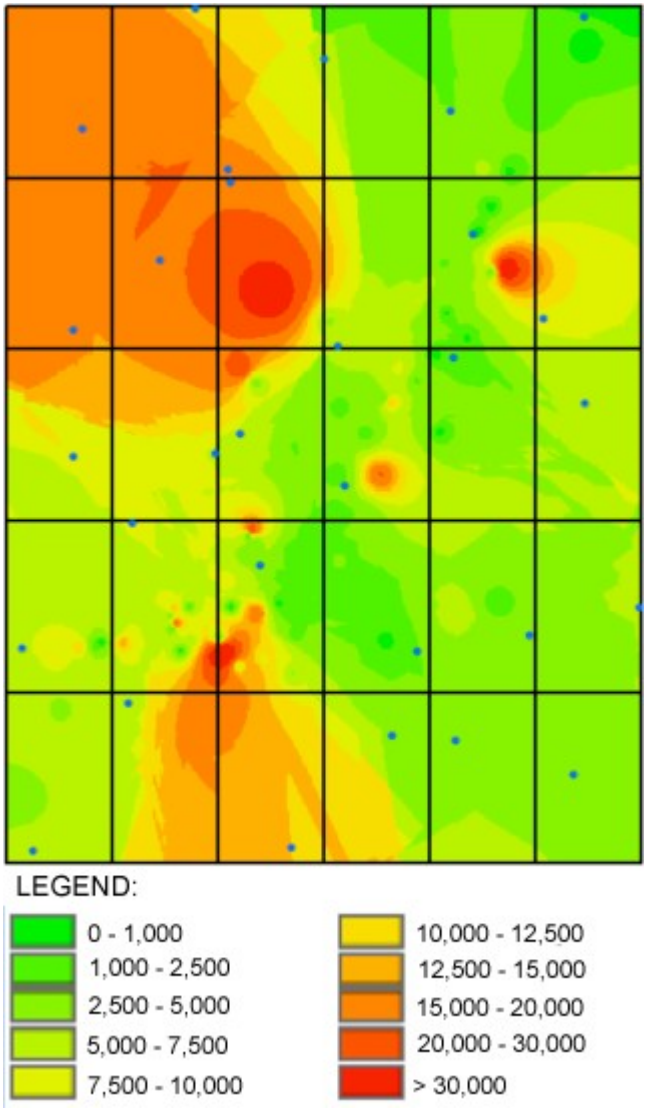


Figure B-2. Example of a map with high spatial autocorrelation (Moran's I z-score = 3.8).
Source: Kelly Black, Neptune and Company, Inc., 2012. Used with permission.

- Throughout the entire DU (all grid cells combined), the population mean is 8,564 and SD is 6,507 (CV = 0.7).

Table B-5. Summary of simulation results for a site with high spatial autocorrelation (see map in Figure B-2).

Source: Kelly Black, Neptune and Company, Inc., 2012. Used with permission.

DU: Map with High Spatial Autocorrelation

Parameters

mean:	8,564	**median:**	6,476
SD:	6,087	**min:**	170
CV:	0.71	**max:**	57,378

Sampling: Simple Random Sampling

(mimics no spatial autocorrelation)

Trials: 1,000

Increments: 30

95% UCL Coverage

	Student's-*t* UCL		Chebyshev UCL	
Replicates	All Site ISM	Quad ISM	All Site ISM	Quad ISM
3	94.4%	NA	97.0%	NA
5	93.6%	NA	99.3%	NA
7	93.8%	NA	99.4%	NA

Bias in Mean

	Grand Mean		Bias	
Replicates	All Site ISM	Quad ISM	All Site ISM	Quad ISM
3	8,561	NA	-0.03%	NA
5	8,563	NA	-0.01%	NA
7	8,564	NA	0.01%	NA

Average RPD between 95% UCL and Population Mean

	Student's-*t* UCL		Chebyshev UCL	
Replicates	All Site ISM	Quad ISM	All Site ISM	Quad ISM
3	19.1%	NA	28.6%	NA
5	11.6%	NA	23.6%	NA
7	9.1%	NA	20.4%	NA

DU: Map with High Spatial Autocorrelation Population

Population Parameters

mean:	8,564	**median:**	6,476
SD:	6,087	**min:**	170
CV:	0.71	**max:**	57,378

Sampling: Systematic Grid and Random Start

Trials: 1,000

Increments: 30

95% UCL Coverage

	Student's-*t* UCL		Chebyshev UCL	
Replicates	All Site ISM	Quad ISM	All Site ISM	Quad ISM
3	93.5%	92.9%	96.2%	96.8%
5	94.5%	95.6%	99.2%	99.0%
7	95.5%	98.2%	99.6%	99.8%

Bias in Mean

	Grand Mean		Bias	
Replicates	All Site ISM	Quad ISM	All Site ISM	Quad ISM
3	8,602	8,612	0.4%	0.6%
5	8,604	8,615	0.5%	0.6%
7	8,605	8,616	0.5%	0.6%

Average RPD between 95% UCL and Population Mean

	Student's-*t* UCL		Chebyshev UCL	
Replicates	All Site ISM	Quad ISM	All Site ISM	Quad ISM
3	10.1%	4.9%	14.8%	7.1%
5	6.3%	3.2%	12.4%	6.0%
7	5.1%	2.7%	10.8%	5.2%

- Both sampling protocols yield relatively unbiased estimates in the mean, which is an expected result for simple random sampling but not necessarily for systematic grid sampling. However, even for a site with high spatial autocorrelation, the bias is negligible when the population has a very low CV.
- Splitting the DU into quadrants results in lower RPDs, mainly reflecting the increase in the total number of replicates.

B.4. MAPS OF RDX AND HMX (M-2A AND M2-B)

Map scenarios M-2A and M-2B represent different spatial structures with both small- and large-scale distributional heterogeneities. These examples are based on a more extensive analysis of ISM conducted for USACE and discussed in a separate report (Qiao et al. 2010). The data are based on results of site investigations involving measurements of concentrations of RDX and HMX in (discrete) bulk surface soil samples. The two histograms in Figure B-3 show each of these sites in 2D histograms with a square-root-transformed count axis to improve the visualization of the tail values. With a standard count axis shown, these distributions would look even more extreme. Their respective means are marked with a dotted green vertical line.

The plots on the left represent a distribution of HMX (mg/kg), and the plots on the right site represent a distribution of RDX (mg/kg) from which increments will be collected. Obstructions such as large rocks and paved roads are excluded to simplify the automation of ISM sampling as well as to simplify the calculation of the population parameter (true mean) from which performance metrics are determined.

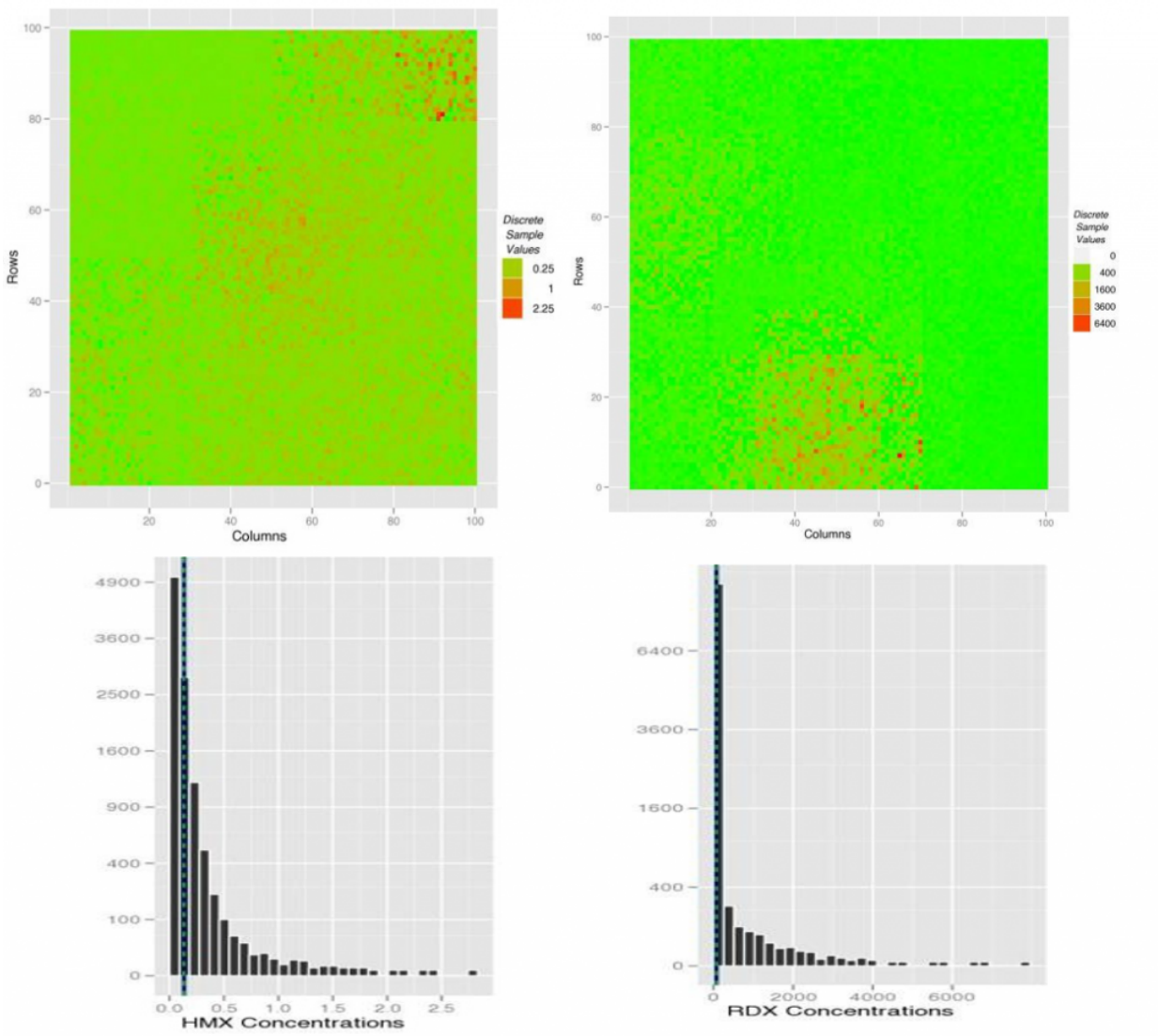


Figure B-3. Spatial distributions and histograms of concentrations for two simulated sites.
Source: J. Hathaway for ACOE, 2012. Used with permission.

B.4.1 Descriptions of DUs

Qiao, Pulsipher, and Hathaway (Qiao et al. 2010) provide details about how the simulated sites were created and values were applied to grid cells representing the DUs. Briefly, each of the 10,000 discrete increment concentration values shown on each site in Figure B-3 are derived from real sites composed of bulk materials. The patterns and concentration values are from extensive discrete data (increments) gathered as a part of multiple ESTCP projects led by Jenkins and Hewitt (Jenkins et al. 2004). Each grid value (increment) in Figure B-3 represents the agglomeration of the bulk material from that area with reported values of constituent levels in units of mg per kg (or parts per million). Thus, as with the simulations with lognormal distributions (PD-1), FE and GSE were not explicitly used in simulating these sites. These errors are implicitly accounted for in the modeled small-scale (local) spatial variability.

B.4.1.1. HMX DU (M2-A)

The HMX concentrations (mg/kg) shown in Figure B-3 (map and histogram on left) depict a 10-m × 10-m DU with moderate heterogeneity. This DU has some spatial patterns, but they are relatively dispersed, and the distribution of values is relatively tight (CV = 1.1). Population parameters include a (true) mean of 0.13, an SD of 0.15, and a maximum of

approximately 2.3 mg/kg.

B.4.1.2. RDX DU (M2-B)

The RDX concentrations (mg/kg) shown in Figure B-3 (map and histogram on the right) depict a 10-m × 10-m DU with more extreme heterogeneity. The map shows one area with extremely high concentrations (bottom middle) and a second area with high concentrations (middle right side) while the rest of the DU has orders of magnitude lower concentrations. This DU represents a site with relatively strong small- and large-scale distributional heterogeneity with a CV of approximately 4.5 (SD = 319 mg/kg; mean = 71.4 mg/kg).

B.4.2 ISM sampling patterns

Figures B-4 to B-7 show the 64 different ISM patterns that are evaluated and summarized in Section B.4.4. For all four figures, each row of plots represents a different number of replicates gathered from the DU (two, three, four, and five), and each column of plots identifies a different number of increments per replicate (16, 30, 49, and 100). Figure B-4 and Figure B-5 show the standard ISM procedure with replicate ISMs over the entire DU for systematic and random grid sampling, respectively. Figure B-6 and Figure B-7 represent the grouped ISM methods for systematic and random grid sampling, respectively. In particular, they show the general structure for each of the evaluated patterns but represent only an example of one random selection for each pattern. Figure B-8 shows the random and systematic discrete sampling types that were evaluated using sample sizes of 9, 16, 30, and 100. Once again, these examples show the general structure for each of the evaluated sampling types and only represent one random selection for each pattern.

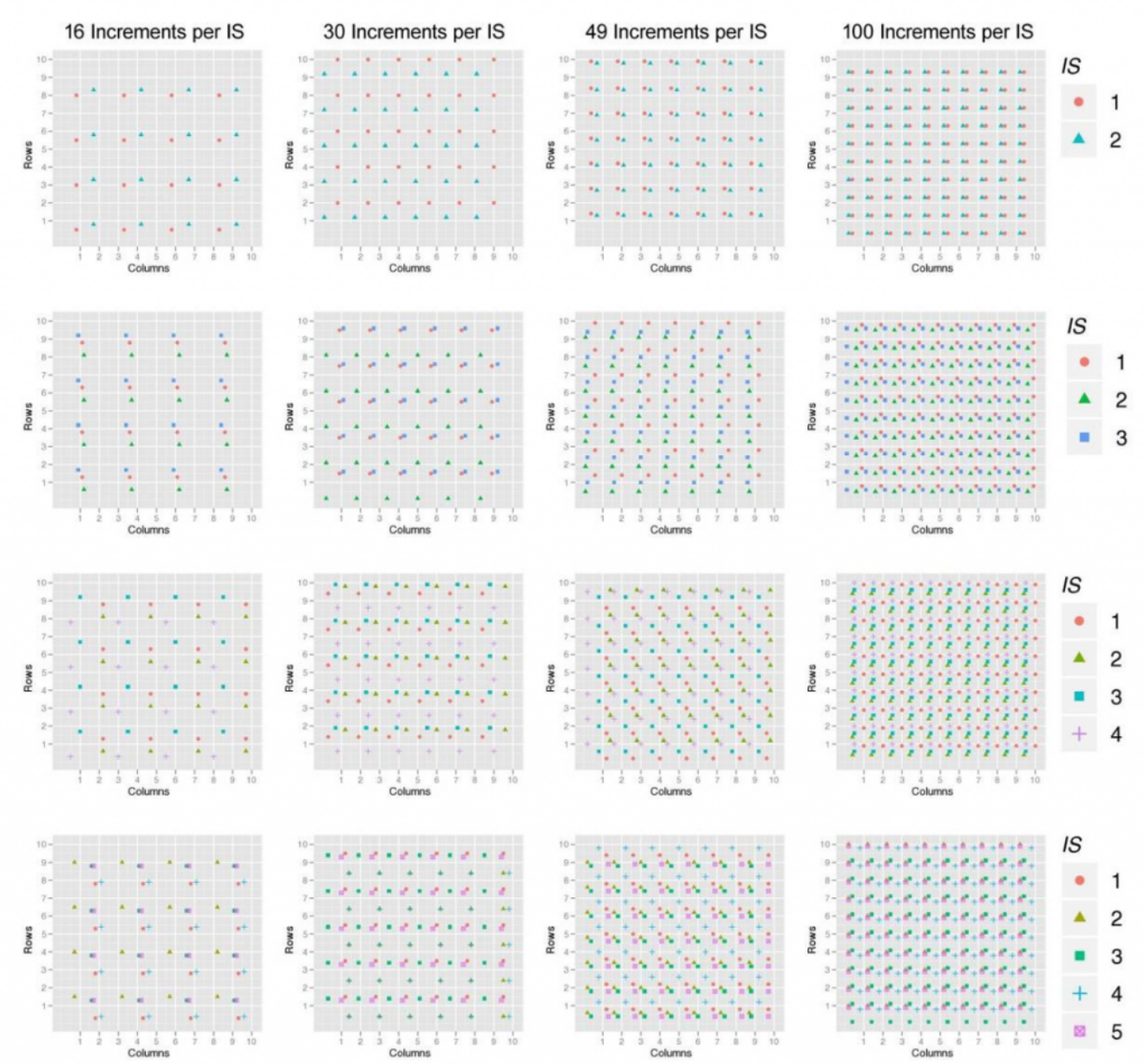


Figure B-4. Standard incremental sampling using a systematic grid sampling approach. Each column represents a differing number of increments per ISM, and each row depicts the differing number of ISMs that were gathered.

Source: J. Hathaway for ACOE, 2012. Used with permission.

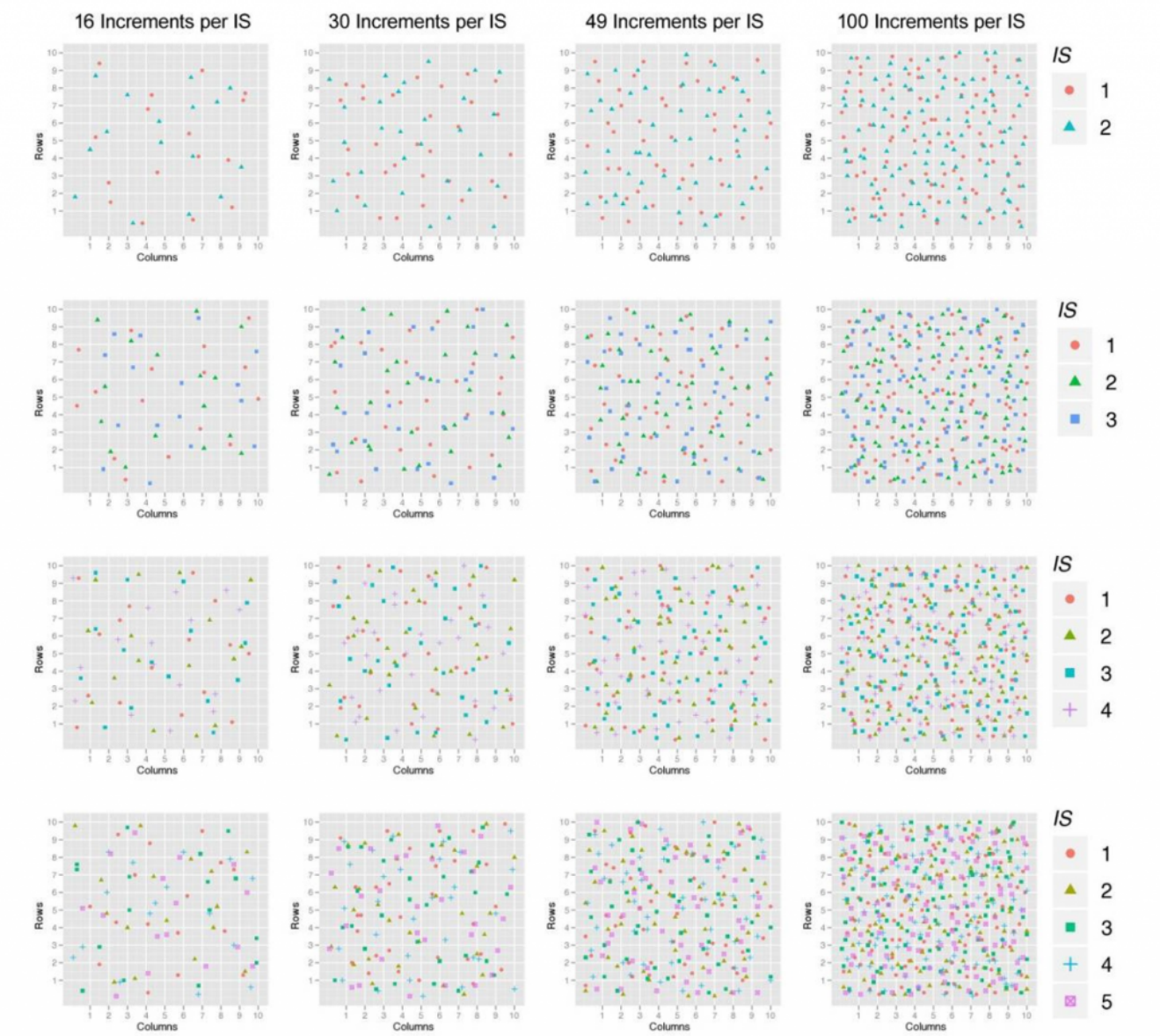


Figure B-5. Standard incremental sampling using a random grid sampling approach. Each column represents a differing number of increments per ISM, and each row depicts the differing number of ISMs that were gathered.

Source: J. Hathaway for ACOE, 2012. Used with permission.

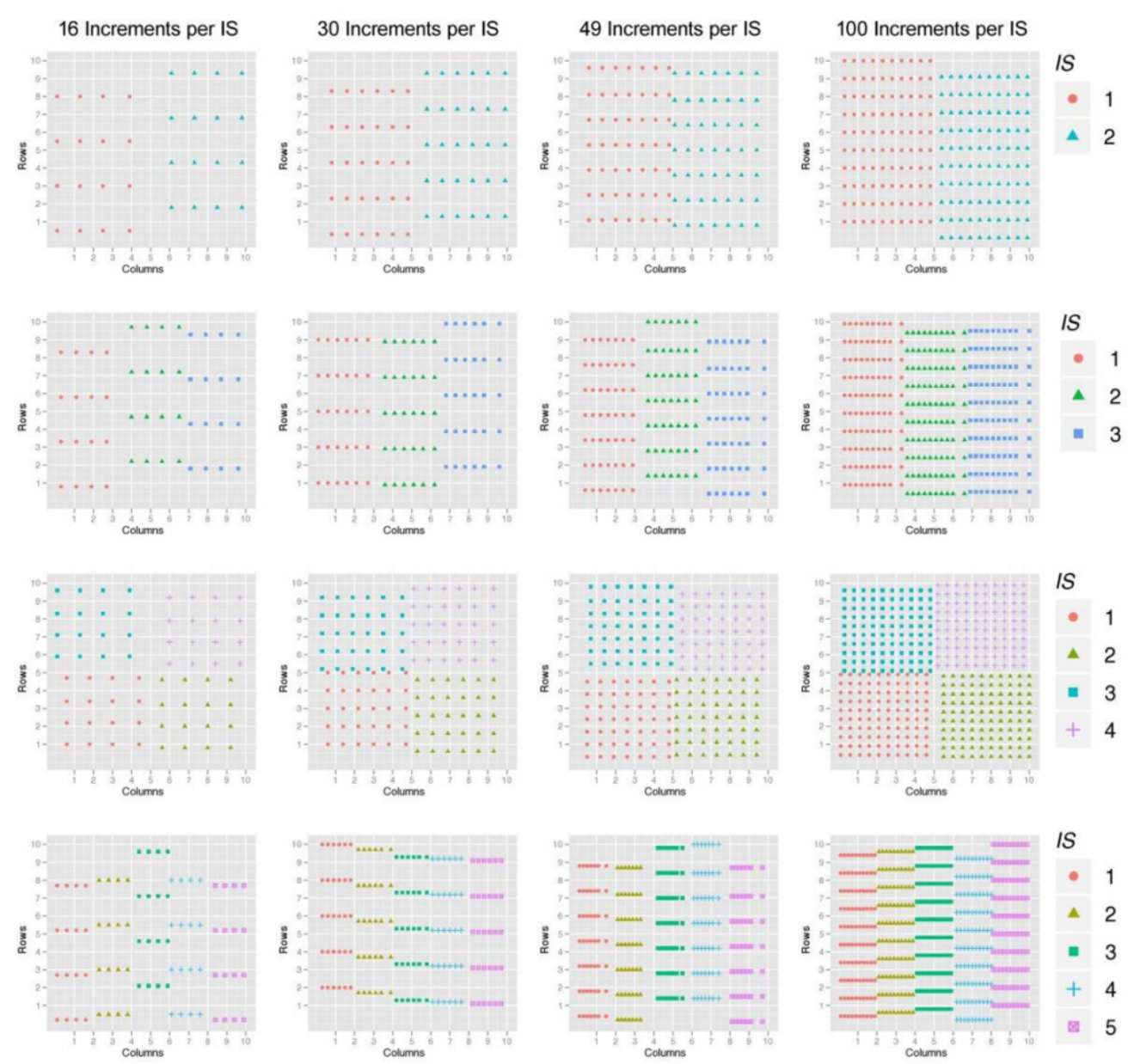


Figure B-6. Grouped incremental sampling using a systematic grid sampling approach. Each column represents a differing number of increments per ISM, and each row depicts the differing number of ISMs that were gathered.

Source: J. Hathaway for ACOE, 2012. Used with permission.

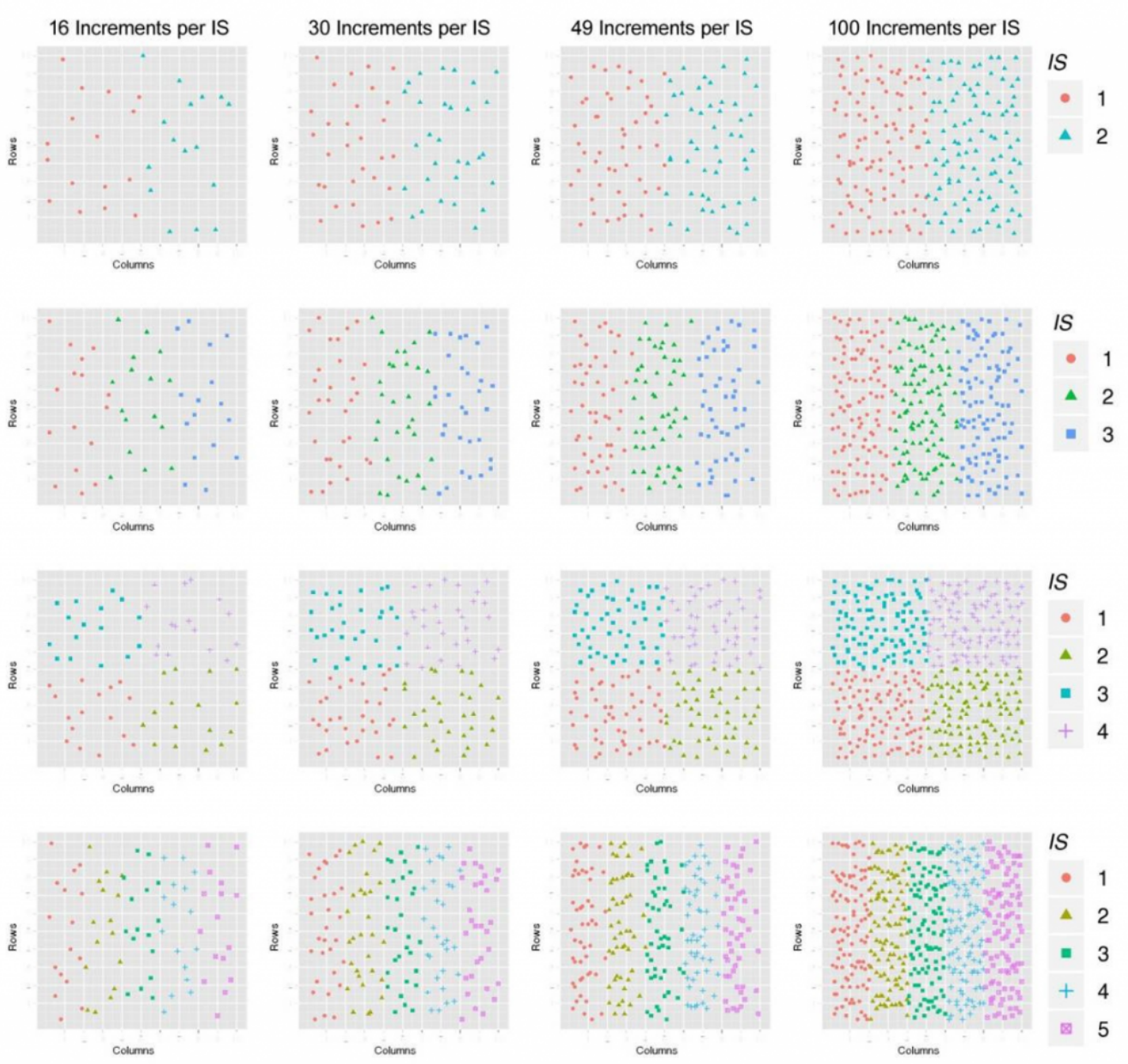


Figure B-7. Grouped incremental sampling using a random grid sampling approach. Each column represents a differing number of increments per ISM, and each row depicts the differing number of ISMs that were gathered.

Source: J. Hathaway for ACOE, 2012. Used with permission.

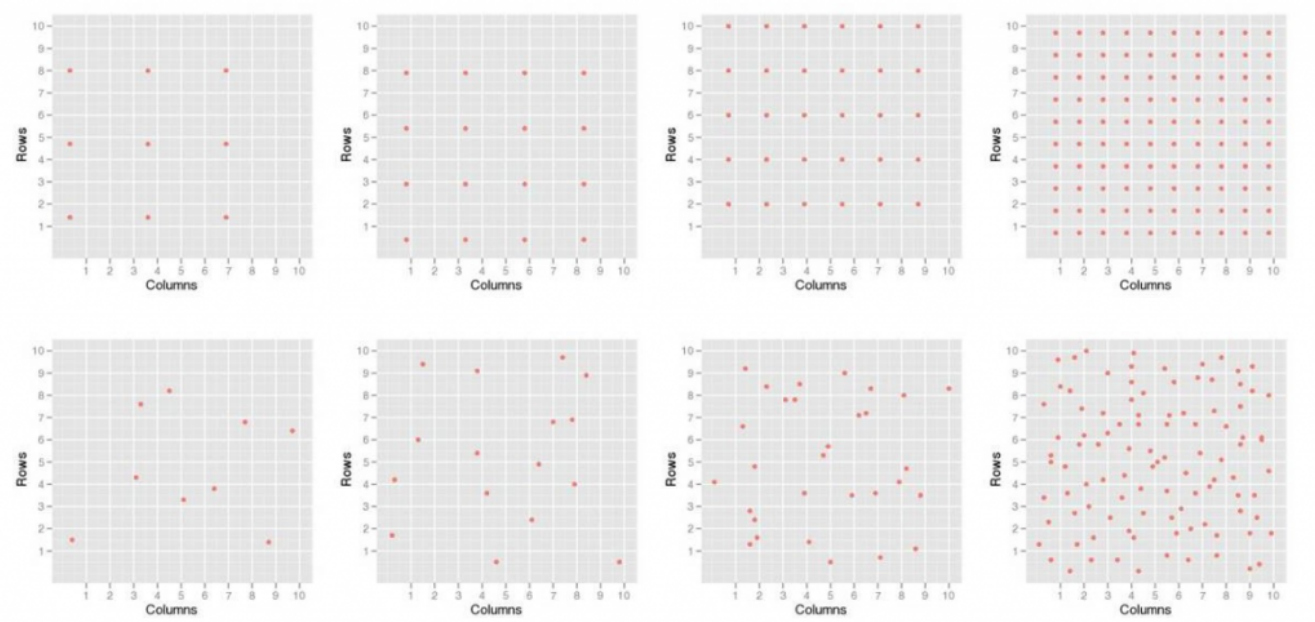


Figure B-8. Discrete sampling using a systematic grid (top row) and random grid (bottom row) sampling approaches. Each column represents a differing number of increments or discrete samples (from left to right 9, 16, 30, and 100 samples per evaluation).
Source: J. Hathaway for ACOE, 2012. Used with permission.

B.4.3 Results using discrete sampling

Table B-4 shows a few of the 2,000 iterations from the UCL calculations based on using the mean and SE calculated from nine systematic grid discrete samples (see upper left plot in Figure A-8) from a DU. These values represent absolute concentrations in mg/kg, and the values from the UCL column are then compared to the true mean. A sampling design achieves the desired statistical coverage if, for example, the UCL values underestimate the true mean in fewer than 100 of the 2,000 iterations (that is, 5%). Figure B-9 shows a histogram of 2,000 UCL values from one simulation scenario where the *y*-axis represents the percentage of 2,000 in each bin (note that the *y*-axis is distorted to show the low bin counts). The red line identifies the location of the true mean. This UCL histogram shows that the coverage was only 76%, which is a significant departure from the theoretical design of 95% UCL. The simulation results provide an example demonstrating how one of the performance metrics (coverage of the 95% UCL) may indicate whether a particular sampling design is unlikely to yield reliable results.

Table B-6. Example of mean and 95% UCL calculations for each iteration of a simulation.

Source: J. Hathaway for ACOE, 2012. Used with permission.

Mean	UCL
0.61	0.76
0.72	0.94
1.01	1.46
0.79	1.18
…	…

0.81	1.02

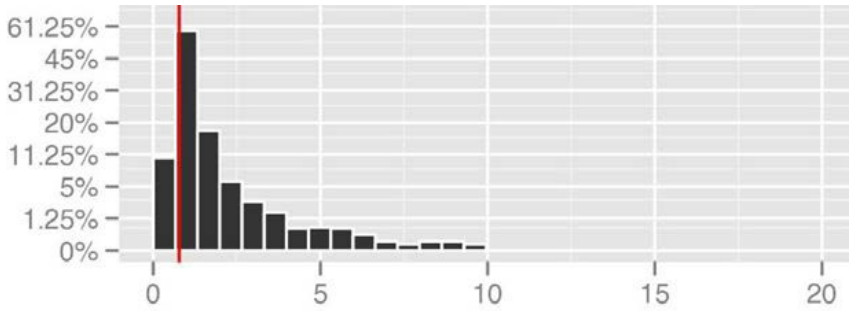


Figure B-9. Histogram of the calculated UCL values using a simulated dataset with 2,000 iterations.
Source: J. Hathaway for ACOE, 2012. Used with permission.

For display purposes, the *y*-axis in Figure B-9 is in terms of percentage of 2,000 and is distorted (not evenly spaced between ticks) to highlight the low count bins. The red line identifies the true mean of 0.776.

The discrete sampling examples were restricted to calculations using Student's-*t* UCL and Chebyshev UCLs. Other methods for UCL calculations are typically considered to attain appropriate coverage by implementing USEPA's ProUCL or comparable software. For sites with heavy right-tailed distributions and distributional heterogeneity, discrete sampling methods with up to 100 samples taken are not sufficient to use a *t*-statistic to calculate a reliable UCL. However, the Chebyshev UCL does provide adequate coverage for many of the DUs at multiple sample sizes. Additional discrete sampling results are discussed in the subsequent sections.

B.4.4. Results using ISM

The following subsections provide results for the RDX and HMX DUs. Within each simulated DU subsection, 40 sets of results are shown using two different UCL calculation methods. Both systematic grid and random grid sampling routines for the grouped and standard ISM patterns were used. Differences in results for these sampling routines were within the range of simulation (stochastic) error. Figure B-10 shows an example of the equal coverage for both M2-A and M2-B using the three different standard ISM sample selection patterns (random grid, simple random, and systematic random) for *t*-based 95% UCLs. For simplicity, only the results associated with the random grid sampling routines are presented in each section.

The tables shown in each section will be separated into the three general sampling patterns: standard ISM, grouped ISM, and discrete sampling. Each table summarizes the results from 2,000 iterations, but the first two columns are different for the ISM and discrete summary tables. For the ISM summary tables, the first column identifies the number of ISMs sampled from within the DU, and the second column shows the number of increments in each ISM. For the discrete summary tables, the first column identifies whether random or systematic sampling was used, and the second column lists the number of increments sampled from the DU that are used to calculate the mean and SD. The third and fourth columns show the UCL coverage for the Chebyshev and *t*-UCL calculations. The last four columns summarize the RPD of the UCL values using the Chebyshev and *t*-distribution UCL multipliers. The RPD above column for each UCL multiplier is the average relative difference from the true mean for those UCL values that were above the true mean. The RPD below columns for each UCL multiplier show the average relative difference from the true mean for those UCL values that were below the true mean.

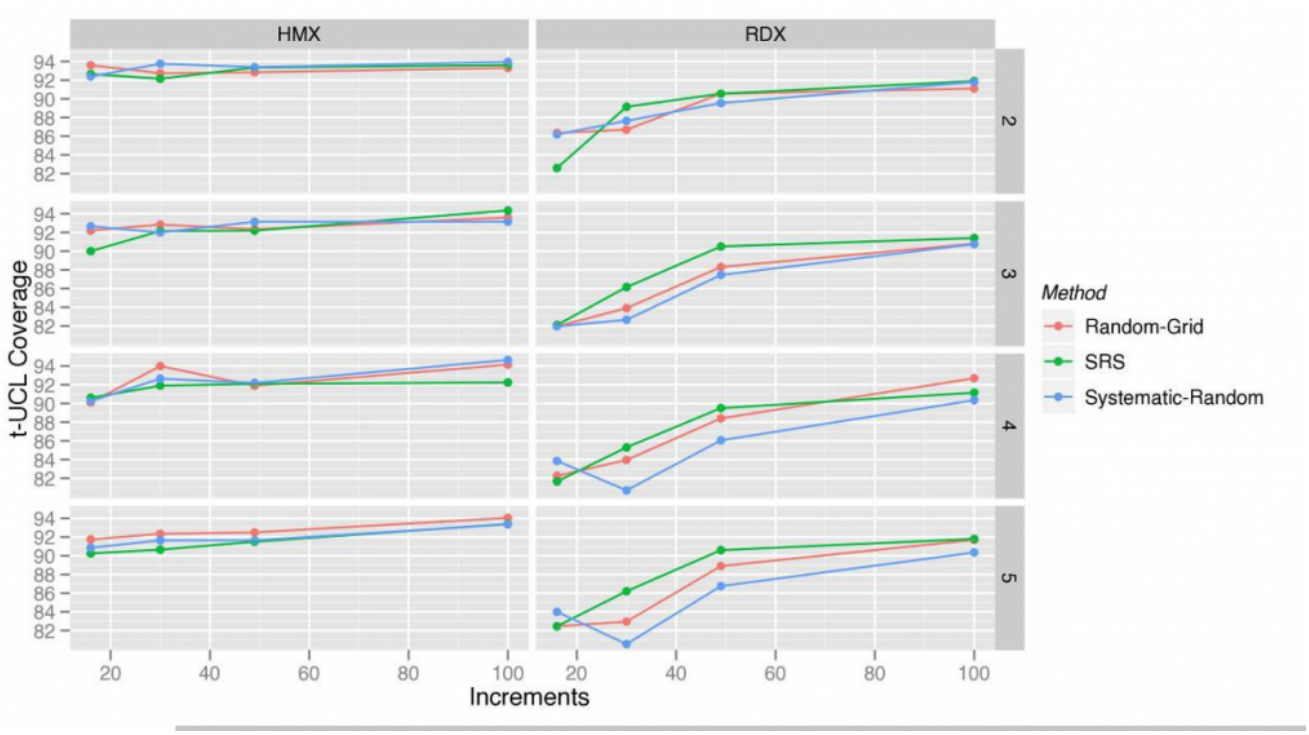


Figure B-10. A coverage plot comparing systematic grid (with random start), random grid, and simple random sampling for the RDX DU (M2-A) and HMX DU (M2-B) when two, three, four, or five ISMs are collected from the DU.
Source: J. Hathaway for ACOE, 2012. Used with permission.

Each subsection contains plots depicting the pertinent information from the coverage tables for an easier visualization of the results from simulation studies. These plots show the designed UCL coverage level (dashed blue line) and the coverage performance of each sampling pattern as a function of the number of increments (in each ISM for the ISM designs and total for discrete designs). Each colored line represents a different sampling pattern with a separate plot for the discrete, grouped ISM, and standard ISM. The dashed line identifies the *t*-UCL calculations, and the solid line identifies the Chebyshev UCL values. Each plotted point represents the results from one line from the tables within the subsection. Coverage results based on 2000 iterations provide estimates accurate to within approximately ±1.5% to ±2.5%.

One figure of 40 UCL histograms with consistent axes is shown in each subsection. These figures are meant to show general distributional and coverage patterns of the calculated UCLs over all sampling patterns and may be difficult to use for evaluating any specific one.

The displayed *t*-distribution UCL calculations are based on a 95% UCL using *t*-distribution with the df equal to 1 minus the number of measures used to calculate the SD for each scenario. For the ISM sampling patterns, df is the number of ISM replicates gathered from the site minus 1, and for the discrete sampling patterns, df is the number of samples gathered minus 1. It is understood that the *t*-distribution is not appropriate for cases where the sample size is small and measured values do not follow a normal distribution. This would generally be the case for the discrete sample designs with 9 and 16 samples as applied to the five simulated sites. In many instances, a different UCL method would be needed for all discrete sample designs (16, 30, 49, and 100), and alternative UCL calculations that do not rely on normal theory should be used in those cases. Such UCL calculations can be found in software such as ProUCL (Singh, Maichle, and Armbya 2007) and VSP (Dowson et al. 2007) for use in environmental studies. There are a variety of choices depending on site-specific needs.

For the proposed ISM sampling methods, the *t*-distribution may not provide adequate coverage, and with the limited number of available data values, it is difficult to use many of the tools in ProUCL for alternative UCL calculations. Thus, a more conservative Chebyshev multiplier is used for attaining an improved coverage percentage; the UCL coverage plots and tables also show the Chebyshev 95% UCL calculations. The SE is multiplied by a prespecified value and added to the mean to identify the UCL. For the *t*-distribution, this value is a function of the number of values used to estimate the mean and SE. The Chebyshev multiplier is 1/sqrt (1 – 0.95) for a 95% UCL regardless of the sample size used. This generally conservative multiplier of 4.472 will shift the coverage statistics up for all sampling patterns except for the two ISM designs. A *t*-distribution with 1 df results in a multiplier of 6.313. The most drastic effects of the Chebyshev multiplier are seen with the discrete designs, as their coverage and bias increase the most.

B.4.4.1 Results for RDX (M2-A)

For the RDX (10-m × 10-m DU) simulations, Tables B-5 through B-7 show the summaries from the evaluated simulations. The coverage, bias, number of increments, and number of ISMs are used to create the coverage plot shown in Figure B-11. Figure B-12 shows the panel of 95% *t*-UCL histograms for all 40 sampling patterns evaluated on the RDX 10-m × 10-m DU.

This site had the strongest small- and large-scale distributional heterogeneity of the two DUs evaluated with a CV of 4.47, with a mean of 71.36 and an SD of 319.1. The coverage results for the standard ISM perform reasonably well for the ISM designs of 100 increments per ISM. The grouped ISM patterns were above the designed criteria of 95% for all but the ISM composed of 16 increments. For this DU, the grouped ISMs are the only patterns that consistently met or exceeded the designed 95% coverage but did have more bias in the mean than the standard ISM or discrete methods.

Table B-7. Discrete summary for RDX DU (M2-A).

Source: J. Hathaway for ACOE, 2012. Used with permission.

Grid Sampling Type	Number of Increments	Chebyshev 95% UCL Coverage	95% *t*-UCL Coverage	Chebyshev RPD above Mean	*t* RPD above Mean	Chebyshev RPD below Mean	*t* RPD below Mean
Random	9	67.20	55.80	596.67	334.23	57.02	61.88
Systematic	9	67.65	54.90	576.75	328.07	56.18	60.07
Random	16	79.25	64.50	431.13	229.60	45.61	49.98
Systematic	16	81.80	65.75	425.83	229.09	47.11	48.37
Random	30	84.60	67.75	292.69	145.30	34.17	40.99
Systematic	30	85.80	67.95	304.20	154.45	39.45	40.97
Random	100	97.50	84.50	182.32	81.15	13.70	20.02
Systematic	100	97.95	86.80	186.52	81.02	12.22	15.26

Table B-8. Standard ISM summary for RDX DU (M2-A).

Source: J. Hathaway for ACOE, 2012. Used with permission.

Number of ISs	Number of Increments	Chebyshev 95% UCL Coverage	95% *t*-UCL Coverage	Chebyshev RPD above Mean	*t* RPD above Mean	Chebyshev RPD below Mean	*t* RPD below Mean
2	16	82.95	86.35	279.99	373.67	37.86	36.14
3	16	88.15	81.95	219.34	157.50	27.98	30.40
4	16	92.35	82.25	199.60	122.60	24.52	26.07

5	16	94.00	82.45	177.73	99.96	20.89	22.80
2	30	82.35	86.70	192.10	257.12	31.80	31.52
3	30	90.50	83.90	150.90	105.86	23.31	24.57
4	30	93.65	83.95	135.61	78.51	20.59	21.45
5	30	95.85	82.95	119.96	64.14	16.60	17.27
2	49	87.85	90.55	147.00	200.34	25.16	23.89
3	49	93.20	88.30	128.19	89.26	16.46	17.75
4	49	96.45	88.40	111.83	64.84	15.40	15.31
5	49	96.85	88.90	101.49	53.30	14.40	15.13
2	100	88.10	91.10	100.46	136.07	16.05	16.26
3	100	94.80	90.80	85.38	59.17	9.62	11.27
4	100	97.60	92.70	76.04	43.07	7.87	10.39
5	100	98.30	91.70	67.41	35.27	8.17	7.79

Table B-9. Grouped ISM summary for RDX DU (M2-A).

Source: J. Hathaway for ACOE, 2012. Used with permission.

Number of ISMs	**Number of Increments**	**Chebyshev 95% UCL Coverage**	**95% *t*-UCL Coverage**	**Chebyshev RPD above Mean**	***t* RPD above Mean**	**Chebyshev RPD below Mean**	***t* RPD below Mean**
2	16	90.55	93.00	408.76	560.08	41.12	42.40
3	16	94.90	90.75	380.21	261.11	31.43	31.17
4	16	95.75	88.45	277.75	159.09	21.51	25.65
5	16	97.95	92.50	297.41	152.63	17.43	23.42
2	30	96.05	97.85	372.63	516.93	29.51	34.85

3	30	98.90	96.15	334.96	223.28	21.55	19.55
4	30	99.35	95.65	239.38	128.93	13.45	18.42
5	30	99.80	96.20	267.77	131.54	13.41	14.40
2	49	99.75	99.95	375.05	528.31	8.90	3.84
3	49	100.00	100.00	342.29	222.02		
4	49	99.75	98.55	240.99	127.20	7.37	17.19
5	49	100.00	97.50	261.90	124.83		12.04
2	100	100.00	100.00	374.57	528.80		
3	100	100.00	100.00	336.40	217.67		
4	100	100.00	100.00	238.50	125.84		
5	100	100.00	100.00	266.15	126.93		

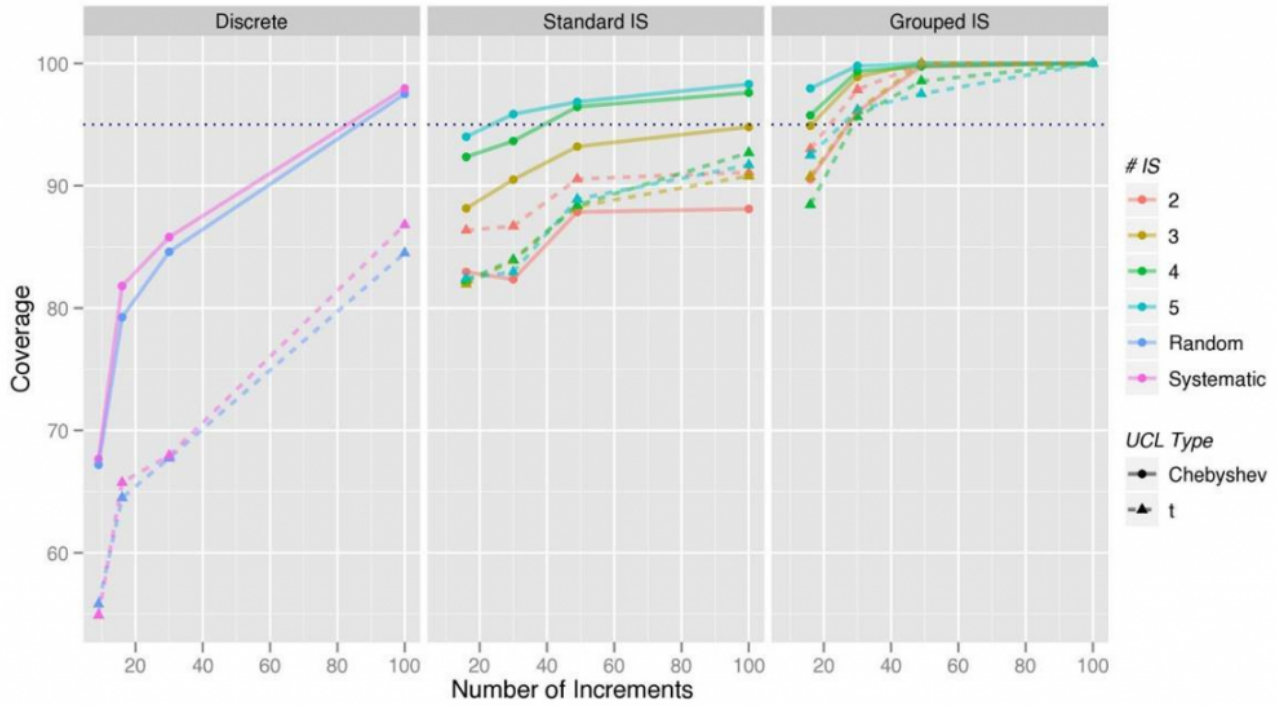


Figure B-11. Plot of the coverage statistics for each of the simulated sampling patterns as applied to the RDX DU. Note that the different sampling patterns are displayed within the plot as well as UCL type.
Source: J. Hathaway for ACOE, 2012. Used with permission.

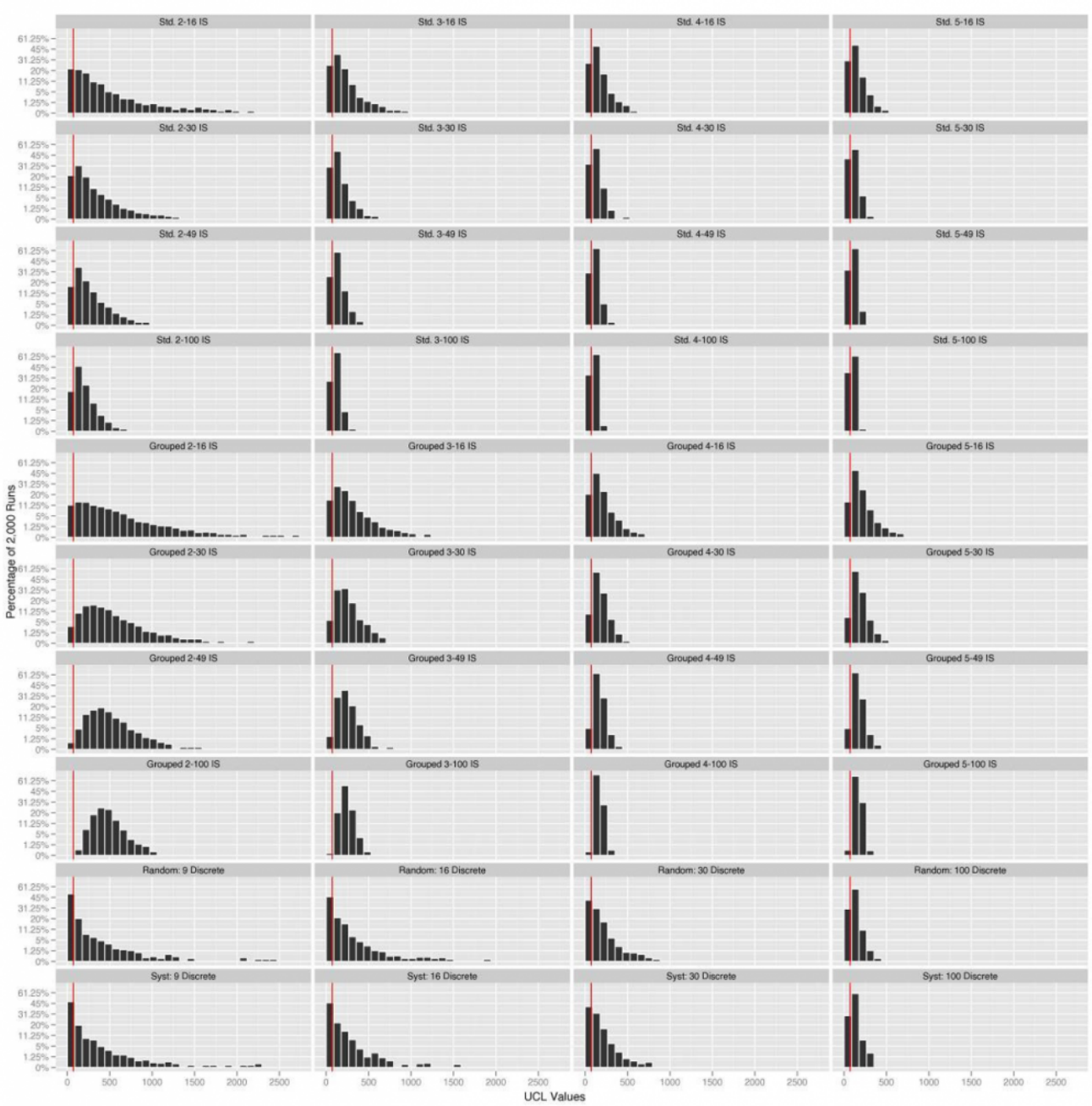


Figure B-12. Panel of histograms of the distribution of *t*-UCL values for the 2,000 simulations. Note that the red line identifies the true mean. The *y*-axis identifies the percent of 2,000 simulations in each bin and is distorted to show the percentage in the low count bins.
Source: J. Hathaway for ACOE, 2012. Used with permission.

Figure B-13 shows the distribution histograms for the 2,000 estimated means from the grouped and standard sampling patterns. This plot is representative of the other simulated sites and shows a few important highlights. As more increments are included in each ISM, the distribution of means becomes more normally distributed. Both the grouped and standard ISM designs provide unbiased estimates of the mean (71.36) and have virtually identical distributions.

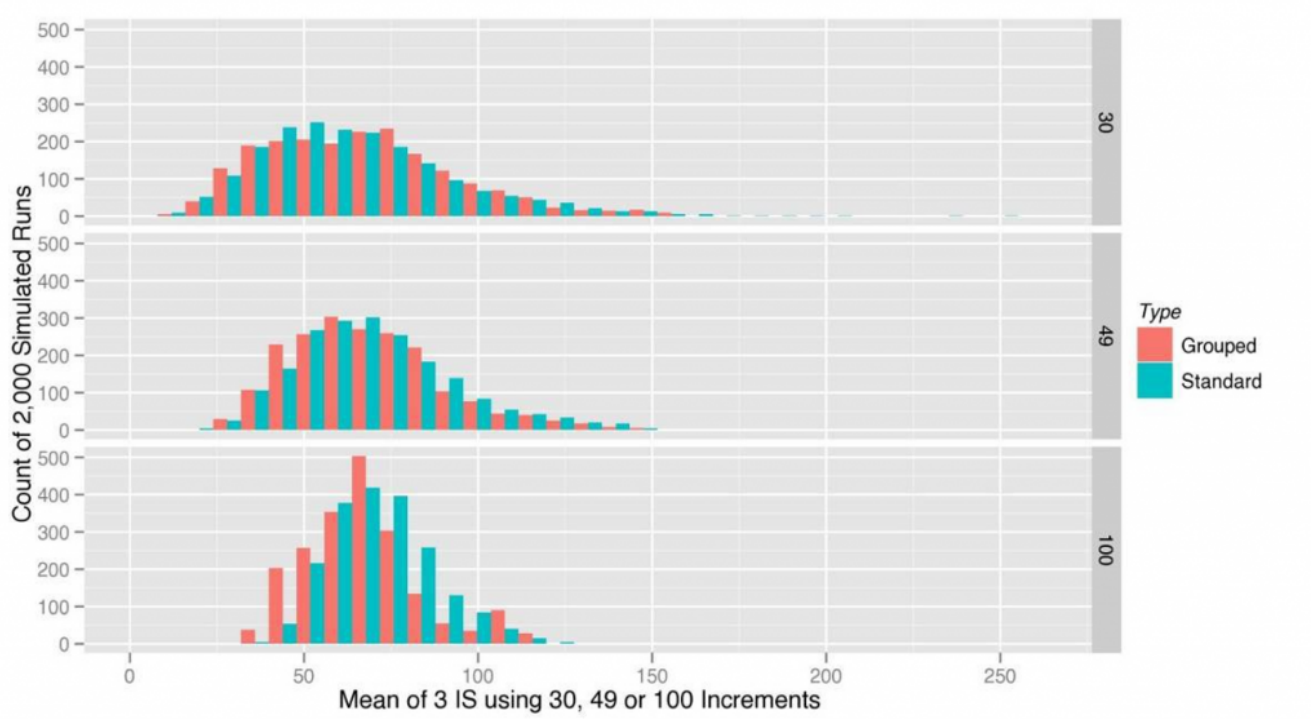


Figure B-13. A comparison of the distribution of means for grouped and standard ISM designs using the RDX DU. Note that the results are similar for all other DUs.
Source: J. Hathaway for ACOE, 2012. Used with permission.

B.4.4.2 Results for HMX (M2-B)

For the HMX (10-m × 10-m DU) simulations, Tables B-8 through B-10 show the summaries from the evaluated simulations. The coverage, bias, number of increments, and number of ISMs are used to create the coverage plot shown in Figure B-14. Figure B-15 shows the panel of t-UCL histograms for all 40 sampling patterns evaluated on the HMX 10-m × 10-m DU.

Table B-10. Discrete summary for HMX DU (M2-B).

Source: J. Hathaway for ACOE, 2012. Used with permission.

Grid Sampling Type	Number of Increments	Chebyshev 95% UCL Coverage	95% *t*-UCL Coverage	Chebyshev RPD above Mean	*t* RPD above Mean	Chebyshev RPD below Mean	*t* RPD below Mean
Random	9	97.65	85.55	140.04	69.15	13.56	15.08
Systematic	9	97.40	83.70	138.07	69.42	11.88	15.48
Random	16	99.05	87.00	110.39	51.50	6.36	11.33
Systematic	16	98.75	86.40	108.63	50.69	5.54	13.01
Random	30	99.55	87.45	83.39	37.49	4.61	8.02
Systematic	30	100.00	90.30	82.82	35.91		6.67
Random	100	100.00	92.20	48.01	19.73		4.15

Systematic	100	100.00	92.80	47.63	19.61		4.64

Table B-11. Standard ISM summary for HMX DU (M2-B).

Source: J. Hathaway for ACOE, 2012. Used with permission.

Number of ISMs	Number of Increments	Chebyshev 95% UCL Coverage	95% *t*-UCL Coverage	Chebyshev RPD above Mean	*t* RPD above Mean	Chebyshev RPD below Mean	*t* RPD below Mean
2	16	90.70	93.60	69.45	94.49	9.84	10.13
3	16	96.50	92.20	59.36	41.10	8.33	7.33
4	16	97.55	90.10	52.77	30.81	5.48	6.15
5	16	98.85	91.75	47.53	24.89	2.97	4.66
2	30	90.20	92.75	50.93	68.96	6.64	6.10
3	30	96.15	92.85	40.70	27.96	5.22	5.69
4	30	98.20	94.00	36.95	20.86	3.59	4.59
5	30	98.75	92.35	33.07	17.44	3.89	3.90
2	49	90.30	92.85	39.87	54.62	6.10	6.01
3	49	96.40	92.35	34.71	23.86	4.60	4.47
4	49	97.65	91.90	29.76	16.71	2.97	3.68
5	49	98.95	92.50	26.96	13.85	2.55	3.56
2	100	91.40	93.30	28.15	38.71	4.67	4.69
3	100	96.95	93.60	22.86	15.56	3.77	3.41
4	100	98.65	94.15	20.29	11.39	1.55	2.27
5	100	99.10	94.05	18.50	9.47	2.10	2.24

Table B-12. Grouped ISM summary for HMX DU (M2-B).

Source: J. Hathaway for ACOE, 2012. Used with permission.

Number of ISMs	Number of Increments	Chebyshev 95% UCL Coverage	95% *t*-UCL Coverage	Chebyshev RPD above Mean	*t* RPD above Mean	Chebyshev RPD below Mean	*t* RPD below Mean
2	16	90.85	92.85	70.55	96.55	9.55	8.54
3	16	96.95	93.05	61.46	42.15	6.42	6.05
4	16	99.80	98.95	89.73	47.36	2.40	5.87
5	16	99.15	93.65	52.03	26.73	5.65	5.19
2	30	91.20	94.00	50.99	69.54	6.84	7.05
3	30	98.55	95.95	47.58	32.33	2.99	3.93
4	30	100.00	99.75	87.32	46.18		3.89
5	30	99.45	95.25	38.94	19.40	4.19	3.44
2	49	92.05	95.35	38.57	52.55	5.02	5.78
3	49	98.70	96.75	38.65	26.05	3.67	3.43
4	49	100.00	100.00	83.60	43.78		
5	49	99.90	97.90	33.76	16.12	2.76	2.63
2	100	93.50	95.00	29.49	40.74	5.66	5.22
3	100	99.20	97.15	27.85	19.02	2.88	2.07
4	100	100.00	100.00	81.47	42.85		
5	100	99.75	98.95	26.67	12.83	1.97	2.38

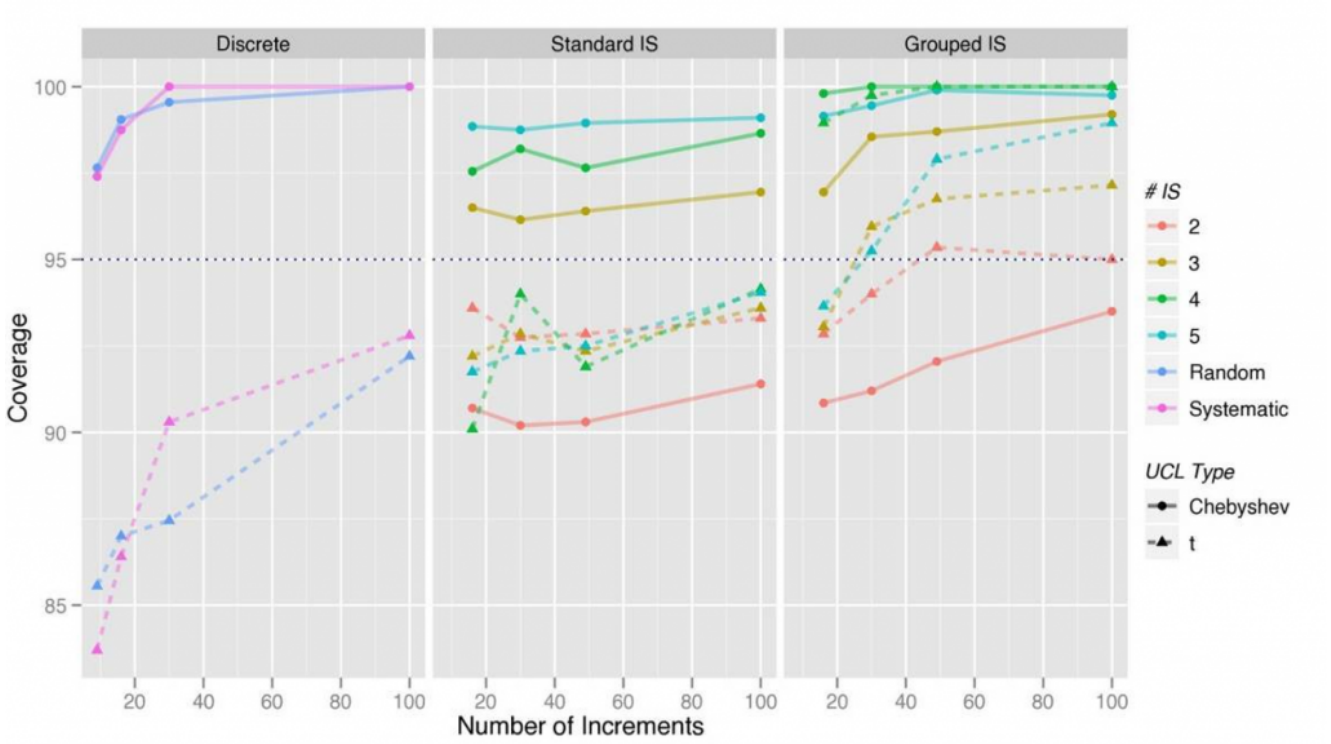


Figure B-14. Plot of the coverage statistics for each of the simulated sampling patterns as applied to the HMX DU. Note that the different sampling patterns are displayed within the plot as well as 95% UCL type.
Source: J. Hathaway for ACOE, 2012. Used with permission.

This DU has some strong distributional heterogeneity, but the distribution of concentration values is not as skewed or heavily right-tailed with a CV of 1.1. The mean is 0.132 with an SD of 0.146. When three or more replicates are used, the coverage results for the grouped ISM patterns were near or above the designed criteria of 95% UCL for all but the ISM composed of 16 increments. The standard ISM performed reasonably well for the 100-increment standard ISM design.

Specific observations from these simulations are noted below and support the consensus points listed in Table B-1:

- The mean concentration estimates for grouped ISM and standard ISM sampling have the same expectation and distribution (see Figure B-13).
- The grouped ISM methods have equivalent or greater coverage than standard ISM when the same number of ISMs and increments are used.
- The RPD of the UCLs for grouped ISM is generally higher than that of standard ISM.
- Grouped ISM, by its definition, provides an improved spatial picture of the concentrations within the site.
- For these maps, the t-UCL may be expected to yield adequate coverage with 100-increment ISM designs.
- As few as 30 increments can be used for DUs with less severe heterogeneity and still maintain coverage with a t-UCL.
- Systematic grid, random grid, or simple random sampling all generally give the same results in terms of coverage, and the use of one or the other can be selected for ease of application (see Figure B-10).
- In general, the Chebyshev method may be necessary to attain adequate coverage depending on the severity of the heterogeneity.
- The improvements in coverage are the more pronounced by increasing the number of increments (for example, 50 to 100) instead of the number of replicates (three to five).

Figure B-15. Panel of histograms of the distribution of 95% t-UCL values for the 2,000 simulations. Note that the red line identifies the true mean. The *y*-axis identifies the percent of 2,000 simulations in each bin and is distorted to show the percentage in the low count bins.

Source: J. Hathaway for ACOE, 2012. Used with permission.

Appendix C. Team Contacts

Caroline Jalanti, ISM-2 Co-Team Leader
New York State Department of Environmental Conservation
518-402-9575
Caroline.Jalanti@dec.ny.gov

Troy Keith, ISM-2 Co-Team Leader
Tennessee Department of Environmental Conservation
423-634-5755
Troy.Keith@tn.gov

Susan Schow, ITRC Program Advisor
Schow Consulting
207-841-0534
schow29@comcast.net

Lisa Alcorn
USDA Forest Service
828-545-2571
lmalcorn@fs.fed.us

Cynthia Auld
U.S. Army Corps of Engineers
978-318-8042
Cynthia.A.Auld@usace.army.mil

Christopher Bartley
Terracon
803-212-0052
cbbartley@terracon.com

Kelly Black
Neptune and Company, Inc.
720-746-1803
kblack@neptuneinc.org

Richard Boone
Ramboll
513-697-2026
rick.boone@obg.com

Hayley Brittingham
Neptune and Company, Inc.
720-639-5460
hbrittingham@neptuneinc.org

Jason Brodersen
Tetra Tech, Inc.
415-497-9060
jason.brodersen@tetratech.com

Mark Bruce
Eurofins
330-966-7267
mark.bruce@testamericainc.com

Douglas Burge
Ramboll
314-590-2963

dburge@environcorp.com

Chris Burrows
RM
317-294-3713
chris.burrows@erm.com

Norlito Cezar
Parsons Corporation
867-688-3352
norlito.cezar@parsons.com

Sandip Chattopadhyay
USEPA/OCSPP
202-564-4806
chattopadhyay.sandip@epa.gov

Chris Christensen
Michigan Department of Environmental Quality
616-446-7582
christensenc@michigan.gov

Jay Clausen
ERDC-CRREL
603-646-4597
Jay.L.Clausen@us.army.mil

Ed Corl
NAVSEA Laboratory Quality and Accreditation Office
757-396-2227
william.corl@navy.mil

Graham Crockford
TRC
734-904-3304
gcrockford@trcsolutions.com

Deana Crumbling
Public
571-263-6356
mccrumb1@msn.com

Karen DiBiasio
CalEPA/DTSC
916-255-6633
karen.dibiasio@dtsc.ca.gov

Annette Dietz
Portland State University
503-725-2328
acdietz@pdx.edu

Richard Doucette
Virginia Department of Environmental Quality
703-583-3813
richard.doucette@deq.virginia.gov

Allison Drouin
Credere Associates, LLC
207-749-1141
adrouin@crederellc.com

Shannon Dunn
Arcadis
206-399-5307
shannon.dunn@arcadis-us.com

Carol Eddy-Dilek
Savannah River National Laboratory
513-529-3218
carol.eddy-dilek@srnl.doe.gov

David Elias
California State Water Board
510-622-2509
delias@waterboards.ca.gov

Calvin Fales
Missouri Department of Natural Resources
573-751-0944
calvin.fales@dnr.mo.gov

Justin Fisher
Tennessee Department of Environmental Conservatism
865-594-5466
Justin.Fisher@tn.gov

Stephen Foster
Geosyntec Consultants
303-790-1340
sfoster@geosyntec.com

Tim Frederick
USEPA Region 4
404-562-8598
frederick.tim@epa.gov

Emily Freitas
Ahtna Engineering Services
315-317-5331
efreitas@ahtna.net

Aaron Friedrich
ERM
317-445-6684
aaron.friedrich@erm.com

Richard Friesner
New England Interstate Water Pollution Control Commission
978-349-2523
rfriesner@neiwpcc.org

Jon Gabry
EPA Region 2
732-321-6650
gabry.jon@epa.gov

Maile Gee
Santa Ana, RWQCB
951-782-4379
Maile.Gee@Waterboards.ca.gov

Thomas Georgian

U.S. Army Corps of Engineers
402-697-2567
thomas.georgian@usace.army.mil

Philip Goodrum
GSI Environmental, Inc.
315-396-6655
pegoodrum@gsi-net.com

Guy Guinot
ERM
504-239-4055
guy.guinot@erm.com

William Harmon
Michigan Department of Environmental Quality
517-284-5110
harmonw@michigan.gov

Allan Harris
EMCBC
513-246-0542
Allan.Harris@emcbc.doe.gov

J. Hathaway
Brigham Young University – Idaho
208-932-8306
hathawayj@byui.edu

Shelley Hicks
Alta Science & Engineering
208-786-1206
shelley.hicks@alta-se.com

Samuel Iwenofu
Washington State Department of Ecology
360-407-6346
siwe461@ecy.wa.gov

West Johnson
Kentucky Department for Environmental Protection
859-319-8170
west.johnson@ky.gov

Celeste Jorge
LNEC
+35-121-844-3348
cjorge@lnec.pt

Konstantinos Kostarelos
University of Houston
832-938-2456
kkostarelos@uh.edu

Matthew Kraus
Tetra Tech, Inc.
412-921-8881
matthew.kraus@tetratech.com

Betty Krupka
CDM Smith

732-590-4646
krupkaea@cdmsmith.com

Matt Lambert
USEPA, OSRTI
703-603-7174
lambert.matthew@epa.gov

Dave Larsen
Utah Department of Environmental Quality
801-536-0226
davelarsen@utah.gov

Mark Leeper
Colorado State University
804-516-3529
mark.leeper@colostate.edu

Julie Lincoln
CA DTSC
gov916-255-3684
julie.lincoln@dtsc.ca

John Lynk
Vatten Geotechnical Services
630-862-9304
jmlynk@vattengeo.com

Thomas Maher
Civil & Environmental Consultants, Inc.
412-429-2324
tmaher@cecinc.com

Alicia McCabe
Ramboll
314-590-2966
amccabe@ramboll.com

Jim McDade
GSI Environmental Inc.
713-522-6300
jmm@gsi-net.com

Ross McFarland
AECOM
+6-141-383-3811
ross.mcfarland@aecom.com

Andrew McKinney
Missouri Department of Health and Senior Services
573-751-6102
andrew.mckinney@health.mo.gov

Lance McMahan
California DTSC
916-255-3674
lance.mcmahan@dtsc.ca.gov

Todd Miller
Kennedy/Jenks Consultants
925-953-1530

toddmiller@kennedyjenks.com

Cheryl R. Montgomery
U.S. Army Corps of Engineers, Engineer Research and Development Center
978-318-8644
Cheryl.R.Montgomery@usace.army.mil

John Montgomery-Brown
EKI Enviroment & Water, Inc.
650-292-9143
jmontgomery-brown@ekiconsult.com

Aditya Moralwar
AECOM
281-675-7674
aditya.moralwar@aecom.com

Kenda Neil
NAVFAC Engineering and Expeditionary Warfare Center
805-982-6060
kenda.neil@navy.mil

Anthony Nelson
U.S. Navy
805-982-2632
anthony.s.nelson@navy.mil

Eric Nuttall
University of New Mexico
505-269-7840
nuttall@unm.edu

Carmen Owens
Apex Companies, LLC
319-360-4128
cowens@apexcos.com

Katherine Owens
Eastern Idaho Community Action Partnership
208-521-3696
kowens@eicap.org

Joey Pace
Arizona Department of Environmental Quality
602-771-4818
jp8@azdeq.gov

Kent Patton
Patton Environmental
503-550-4558
kentpatton@jeffnet.org

Ralph Perona
Neptune and Company, Inc.
360-733-1501
rperona@neptuneinc.org

Stephen Rembish
Parsons Corporation
512-719-6067
steve.j.rembish@parsons.com

Hugh Rieck
U.S. Army Corps of Engineers
402-697-2660
hugh.j.rieck@usace.army.mil

Mark Rigby
Parsons Corporation
801-553-3350
mark.rigby@parsons.com

Christian Romero
Noster Tec
+591-2-243-1080
ductor@nostertec.com

Margaret Roy
Golder Associates, Ltd.
512-913-5142
marymargaret_roy@golder.com

Loren Ruttinger
WWC Engineering
307-672-0761
lruttinger@wwcengineering.com

Craig Sanchez
California State Water Board
916-322-9446
craig.sanchez@waterboards.ca.gov

Shanna Schmitt
Minnesota Pollution Control Agency
651-757-2697
shanna.schmitt@state.mn.us

Ryan Seelbach
Geosyntec Consultants
510-285-2659
rseelbach@geosyntec.com

Richard Sloan
Montana Department of Environmental Quality
406-841-5046
rsloan@mt.gov

Tamara Sorell
Brown and Caldwell
978-983-2046
tsorell@brwncald.com

Jason Speicher
NAVFAC Atlantic
215-897-4914
jason.a.speicher@navy.mil

Judith Strawhecker
U.S. Army Corps of Engineers
402-697-2668
judith.m.strawhecker@usace.army.mil

Michael Stroh

Missouri Department of Natural Resources
573-522-9902
michael.stroh@dnr.mo.gov

Leah Stuchal
University of Florida
352-264-4641
lstuchal@ufl.edu

Theodoros Toskos
Woodard & Curran
609-454-6709
ttoskos@woodardcurran.com

Nicole Wagner
EA Engineering, Science, Technology, Inc., PBC
410-771-4950
nwagner@eaest.com

Tyler West
SC DHEC
864-490-2391
westtl@dhec.sc.gov

Eric White
Texas Commission on Environmental Quality
512-239-2009
eric.white@tceq.texas.gov

Mary KV White
Colorado Dept of Labor and Employment-Div Oil and Public Safety-Remediation
303-318-8506
marykv.white@state.co.us

Jennifer Wilkie
Geosyntec Consultants
312-416-3938
jwilkie@geosyntec.com

Elizabeth Winger
Eurofins
714-895-5494
elizabethwinger@eurofinsus.com

Stephen Witzmann
ERM
907-264-4401
stephen.witzmann@erm.com

Lynn Woodbury
CDM Smith
303-383-2382
woodburyl@cdmsmith.com

Michele Zych
Wood PLC
858-300-4337
michele.zych@woodplc.com

Appendix D: Glossary

accuracy – the degree of agreement between the measurement of the measured value and its actual (true) value. Together, *precision* and *bias* determine **accuracy**.

action level – the generic term applied to any numerical concentration criteria that will be compared with environmental data to arrive at a *decision* or determination about a potential contaminant(s) of concern (from survey through remediation) or for a user-defined volume of media using environmental *sample* data.

aliquot – the term used to refer to a portion of a liquid solution or solid matrix that is taken out of the *sample* or *subsample* for analysis. **Aliquot** is somewhat synonymous with subsample and is often used incorrectly in the site cleanup industry to refer to an *increment*. This is inappropriate and confusing because these terms are not synonymous. Increment is the proper word for field samples that are added together (or pooled) to form a *composite sample* or an *incremental sample*.

analytical quality – the degree to which evidence demonstrates that all steps in an analytical process were performing with acceptable *bias* and *precision*. *Quality* is judged by the ability of the data to be used for their intended purpose. Since there are many types of *decisions* that vary in their need for data rigor, **analytical quality** that is acceptable for making one decision may not be acceptable for making a different decision. If documentation for the *quality control* in critical process steps is lacking, the data are said to be of "unknown" quality.

analytical sample – the portion of a *soil* (or other media) *sample* submitted to the laboratory that actually undergoes extraction or digestion to dissolve target analytes into a liquid that can be injected into an instrument for measurement. The term is interchangeable with the term *analytical subsample*.

analytical subsample – the portion of a *soil* (or other media) *sample* submitted to the laboratory that actually undergoes extraction or digestion to dissolve target analytes into a liquid that can be injected into an instrument for measurement. The term is interchangeable with the term *analytical sample*.

analytical variability – the imprecision in data results that are attributable to the analytical process of extraction/digestion of an *analytical sample*, cleanup of the ensuing extract (if performed), introduction of the extract into the analytical instrument, and the operation (calibration, signal stability, maintenance, etc.) of the analytical instrument itself. The degree and sources of **analytical variability** are measured by analytical *quality control* checks, typically measured in terms of *standard deviation*, *relative standard deviation*, or *relative percent difference*.

area of influence – the area of *soil* surrounding a *sample* that is considered to have the same concentration as the sample. This concept is equivalent to the *sampling unit* concept, which more explicitly considers soil volume, rather than just area.

arithmetic mean – the sum of x measurements divided by x (all measurements are equally weighted).

average – see *arithmetic mean*.

bias-corrected accelerated (BCa) bootstrap method – a modification of the percentile bootstrap *95% upper confidence limit*, which attempts to address the issue of insufficient *coverage.*

bias – the tendency for a measurement to consistently over- or underestimate the actual (true) value. *Precision* and **bias** together determine *accuracy*.

bulk soil – generally, native *soil* that has not been sieved. However, it may contain components that are not considered to be "soil" in its strictest sense, such as twigs and other macro plant fragments, living creatures (insects, worms, etc.), stones larger than the 10-mesh sieve fraction (2 mm), man-made debris, and/or trash.

bulk soil sample – a *soil sample* expected to be representative of native soil. *Soil* is defined by the U.S. Department of Agriculture as mineral and organic material that is less than the 10-mesh sieve fraction (<2 mm in diameter). A **bulk soil sample** contains both the fine fractions, usually considered to be the less than 60-mesh fraction (<0.25 mm in diameter), and the coarse fractions, the material between 60- and 10-mesh (0.25 to 2 mm in diameter). The soil fraction that does not pass through a 10-mesh sieve is often referred to as oversized material and removed from the **bulk soil sample** during processing.

co-located sample – *soil samples* collected a few inches to a few feet apart as a *quality control* check and sometimes performed in traditional discrete or grab sampling. In discrete sampling plans, **co-located samples** provide valuable information about short-scale spatial *heterogeneity* and whether it is causing significant *sampling error* that could lead to *decision* errors. Because there are usually two samples involved, quantitation of the variation/*precision* between **co-located**

samples usually uses the *relative percent difference* as the measure.

composite sample – a *sample* composed of two or more *increments*, which generally undergoes some preparation procedures designed to reduce the errors associated with obtaining a measurement from the combined sample. An ISM sample is a type of **composite sample** whose collection and preparation steps are designed using the general suggestions of Gy's *theory of sampling*. Traditional **composite samples** generally do not consist of a large volume or a large number of increments and do not undergo the same preparation and *subsampling* steps suggested by Gy'stheory. The *mean* for the area covered by a **composite sample** is not expected to be as accurate as a mean produced by an *incremental sample*.

compositional heterogeneity (CH) – the *heterogeneity* arising from the composition of each particle within a *decision* unit. **CH** is inherently dependent on the composition, shape, size, density, etc., of the particles or fragments making up the lot. **CH** is synonymous with *constitution heterogeneity*.

conceptual site model (CSM) – a written or pictorial representation of an environmental system and the biological, physical, and chemical processes that determine the transport of contaminants from sources through environmental media to environmental *receptors* within the system. The **CSM** should include the contamination release mechanisms; fate and transport mechanisms and how long they have been acting on the contamination and creating short-scale *heterogeneity*; what receptors are currently present and potential future receptors; what exposure pathways may exist currently and potential future exposure pathways; what exposure routes of entry may exist currently and potential future exposure routes for each receptor; and what, if anything, will be done about the contamination and/or intact exposure pathways.

confidence limit (CL) – the numbers at the upper (*95% upper confidence limit*) and lower (lower confidence limit) end of a confidence interval. The 95% **CL** is the most frequently used, although other values can be used.

constitution heterogeneity – synonymous with*compositional heterogeneity* (CH).

correct samples – the term used in the Gy-based *theory of sampling* to designate *samples* for which *representativeness* is known and documented. The analytical data from **correct samples** can be relied upon to support correct *decisions*.

coverage – for statisticians, the probability that a confidence interval encloses or captures the true population parameter. For example, a calculated 95% *upper confidence limit* is intended to have a 95% chance of being equal to or exceeding the true (population) *arithmetic mean*. For field investigators, **coverage** is the extent to which the density of sampling locations represents the *sampling unit* (i.e., spatial **coverage**).

data quality – the fitness of data for their intended use. Since *soil* data come from soil *samples*, **data quality** must include *sampling quality* as well as *analytical quality*. The analysis can be perfect, but if the sample was "wrong" (perhaps degraded or mislabeled or the wrong particle sizes analyzed), the **data quality** is "bad" in that the results can give misleading information.

data quality objective (DQO) – a qualitative and quantitative statement derived from the USEPA **DQO** process that clarifies a study's technical and *quality* objectives, defines the appropriate type of data, and specifies tolerable levels of potential *decision* errors that will be used to establish the quality and quantity of data needed to support decisions.

data representativeness – a measure of the degree to which data accurately and precisely represent a characteristic of a population.

data uncertainty – a lack of confidence that data can be used for a particular application. Data are uncertain when there is insufficient documentation to explain how samples were collected, processed, or *subsampled* (i.e., sample *representativeness* is unknown); insufficient documentation of *quality control* (QC) is available to document sources of sampling and *analytical variability* so the *bias* and *precision* of the data are unknown; QC documentation shows that data are too biased; or data precision is too poor to support confident *decisions*. **Data uncertainty** is always present to some degree for *soil* data, however, **data uncertainty** becomes excessive when the risk of incorrect *decisions* exceeds tolerable levels.

data validation – an analyte- and sample-specific process that extends the evaluation of data beyond method, procedure, or contractual compliance (i.e., data verification) to determine the *quality* of a specific dataset relative to the end use. **Data validation** focuses on the project's specifications or needs, is designed to meet the needs of decision-makers/data users, and should note that potentially unacceptable departures from the Quality Assurance Project Plan. The validation process may look at the laboratory *quality control* (QC) checks for *sample* extraction and digestion and for extract cleanup. **Data validation** always includes evaluation of the laboratory QC used to evaluate instrument calibration and performance and may also check mathematical calculations. In this way, **data validation** is important to establish the *precision* and *bias* of the pure analytical process, but it does little to establish *data representativeness* or estimate *data uncertainty*.

decision – a determination made about a potential contaminant(s) of concern or for a volume of media using environmental *sample* data.

decision error – making an incorrect *decision*, such as deciding that cleanup is needed when it is not or missing a cleanup that is warranted based on site-specific information, including measurement from samples collected within the *decision unit*.

decision mechanism – an algorithm or protocol that results in a *decision* about a potential contaminant of concern or for a volume of media. A variety of **decision mechanisms** are possible when using ISM sampling. Each **decision mechanism** has strengths, weaknesses, and assumptions. In some cases, agency requirements will dictate the **decision mechanism** to be used. In other cases, a consensus on the **decision mechanism** to be employed needs to be reached among members of the planning team prior to finalization of the sampling plan.

decision threshold – any type of numerical value used for an exceedance/non-exceedance *decision*, such as a screening level, an *action level*, a cleanup level, a criterion, etc.

decision unit (DU) – the smallest volume (i.e., plan area and depth) of *soil* (or other media) for which a *decision* will be made based upon ISM sampling. A **DU** may consist of one or more *sampling units*. It is an incorrect use of the term **DU** when used to represent all ISM *sample* results, regardless of decision type or intended use.

delimitation error – the error that results from incorrect shape or nonuniform volume of material extracted from the *decision unit* or *sampling unit* to form the *sample*. Often occurs when improperly shaped tools or incorrect equipment are used to collect *increments*.

disaggregation – the act of breaking *soil* peds (clods or clumps) into individual small particles but keeping the small pebbles and hard crystalline (mineral soil) particles intact. **Disaggregation** is often performed by crushing soil peds using fingers, a hand-operated mortar and pestle, a coffee grinder, a rubber mallet, etc. **Disaggregation** does NOT involve particle size reduction, which is the breaking apart or crushing of individual solid particles by *milling*. Mills are able of reducing solid rock particles to the consistency of flour.

discrete soil sample – a *soil sample* obtained from the parent matrix by scooping or coring from a single location at a single point in time. May also be termed a “grab sample,” especially if the sample has been collected without consideration of a statistically valid sampling design or a representative *sample support*. Grab samples are almost always *incorrect samples*.

distributional heterogeneity – the *heterogeneity* describing the nonuniform distribution at all scales of types of fragments or particles within a *sample*, across a *sampling unit* or *decision unit*, or across a site.

energetics – residues that are unreacted explosives and propellant compounds that remain after firing or detonating munitions as defined in USEPA SW-846 Method 8330B.

exposure point concentration (EPC) – an estimate of the concentration of a constituent in an environmental medium to which a *receptor* will be exposed. The **EPC** can be determined for an entire site or for an individual *exposure unit*. The **EPC** is based on a statistical derivation of either measured data or modeled data. In risk assessment, an **EPC** is typically based on a 95% *upper confidence limit* so that that risk-based *decisions* are protective of human health and the environment.

exposure unit (or exposure area decision unit) – a *decision unit* that is used to make *decisions* about risk or a volume of an environmental medium (for example, *soil*) over which a *receptor* is reasonably assumed to move randomly and is therefore equally likely to contact all locations.

extraction error – the error that results from incorrectly extracting the *increment* from the *decision unit* (DU) or *sampling unit*. An example is loose material in the bottom part of the corer falling out and back into the hole when sampling dry, sandy *soils* with an open-bottom corer. 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.

field replicate samples – two or more *incremental samples* independently collected from the same *decision unit* or *sampling unit* using the same number of *increments* but from offset increment collection locations.

fundamental sampling error – the *sample* variability that results from the *constitutional heterogeneity* of *soil*. **Fundamental sampling error** is always present and can be estimated, but its magnitude depends in part on sample mass relative to particle size. It can be reduced by reducing particle size and/or increasing sample mass. For soil with particles up to 2-mm diameter and analyte concentrations of interest in the mg/kg range, a sample mass of 1 kg is typically required to control variability due to fundamental error with 15 to 20% *relative standard deviation*. Particle size must be reduced before

sample mass is reduced (e.g., *subsampling*) to maintain control of fundamental error.

grand mean – the *arithmetic mean* of all ISM replicates from the same *decision unit* or *sampling unit*.

grinding – a generic term for *soil disaggregation* or *milling*. When using the term **grinding**, you must specify the equipment to be used to help ensure an accurate understanding of whether the intent is *disaggregation* or *milling*. *Soil* that is "ground" in a coffee grinder or in a hand-operated mortar and pestle will only be *disaggregated* because those grinders do not have the force needed to fracture non-friable mineral and rock particles (sands, pebbles, etc.). To provide clarity of intent, the term **grinding** should be used when the intent is disaggregation and the term milling should be used when the intent is particle size reduction.

grouping and segregation error – *sample* variability resulting from the short-range *distributional heterogeneity* within and around the immediate area from which a sample is collected (i.e., the sampling location) and developing within the sample container after sample collection. Particles tend to associate into groups of like particles due to gravitational separation, chemical partitioning, differing moisture content, magnetism, or electrostatic charge, which can lead to sampling bias).

Gy-compliant (procedure) – *sample* collection, sample processing, sample splitting, and *subsampling* procedures that comply with the *theory of sampling* by using activities that minimize *sampling errors*, in particular, *fundamental sampling error*, *grouping and segregation error*, and *increment* delimitation and *extraction error*. Using **Gy-compliant procedures** for all steps of the sample collection and analysis process produces *correct samples* for which *representativeness* is known.

heterogeneity – the condition of spatial nonuniformity in the distribution of *soil* constituents. All soil is heterogeneous. In a first analysis, there are two fundamentally different types of **heterogeneity** in soil: **heterogeneity** due to the dissimilar and diverse constituents of the individual particles, and **heterogeneity** due to the nonuniform spatial distribution of different types of particles within the soil. These are identified as *CH* and *distributional heterogeneity*. Compositional **heterogeneity**, also called *micro-scale heterogeneity*, is responsible for *fundamental sampling error* and is reduced by increasing *sample* mass. Distributional **heterogeneity** is present at all scales. Variability due to distributional **heterogeneity** is addressed by site *stratification* into logical *sampling units* or *decision units* and by adjusting the number of *increments*.

hot spot – generally described as an area of elevated contamination (ITRC 2008). A **hot spot** is not typically identified visually (i.e., stained *soil*, free product) but is primarily identified on the basis of chemical concentrations detected in soil *sample* results. For meaningful discussion, the specific size and magnitude of chemical concentrations that constitute a **hot spot** should be agreed on during systematic project planning.

incorrect samples – the term used in Gy-based *theory of sampling* to designate *samples* for which *representativeness* is not known and cannot be determined. No matter how good the analytical *quality* is, data from these samples are not reliable to support decision-making.

increment – a volume of *soil* collected from a single point within a *decision unit* (DU) or *sampling unit* (SU) that is collected with a single operation of a sampling device. Multiple **increments** (typically 30 or more) are collected from a DU or SU and combined to form an *incremental sample*. This term should be used instead of the term*aliquot***,** which actually has the opposite meaning. An **increment** is something added in or added together, an *aliquot* is something taken out, like a portion of extract taken from a flask to inject into an analytical instrument.

incremental sample – a *sample* formed from multiple *increments* collected from a defined volume of *soil*, the *decision unit* (DU) or *sampling unit* (SU), which are combined, processed, and analyzed to estimate the mean concentration in that DU or SU.

independent sample – a stand-alone *sample* whose result is not dependent on any other samples. For example, each *decision unit* is often sampled by collecting three independent *field replicate samples*, not by splitting a single *incremental sample* or by collecting co-located increments.

laboratory replicate – two or more *subsamples* taken from a single field *sample*. Synonymous with *subsample replicate*. Not to be confused with a laboratory instrument replicate, which is repeated measurement of a sample to determine *precision* for the instrument.

laboratory control sample – a known matrix spiked with compound(s) representative of all target analytes.

large-scale distributional heterogeneity – nonuniform distribution of differences in analyte concentration from location to location across an area or differences in how contaminants are spatially distributed throughout the *decision unit* (DU) or

sampling unit (SU). Variability in results due to the nonuniform distribution of analytes across the DU/SU is controlled by increasing the number of *increments* making up the *sample*. This is the spatial scale at which *heterogeneity* becomes important to decision-making. Also synonymous with long-scale heterogeneity.

long-scale heterogeneity – synonymous with large-scale *heterogeneity*.

matrix spike/matrix spike duplicate – environmental *samples* that are spiked in the laboratory or in the field with a known concentration of a target analyte(s) to verify percent recoveries. **Matrix spike/matrix spike duplicate** samples are primarily used to check sample matrix interferences. They can also be used to monitor laboratory performance. A duplicate spike is used to assess *bias* and *precision*.

micro-scale heterogeneity – differences in size and composition between individual *soil* particles. Often due to some soil particles being composed of minerals that more readily adsorb contaminants than other soil particles. See also *compositional heterogeneity*. Not to be confused with *short-scale distributional heterogeneity*.

milling – complete particle size reduction of all *soil* components including hard crystalline materials to a defined maximum particle size (<250 μm or <75 μm). The terms “pulverization” and “comminution” are synonymous with **milling**. The types of mills commonly used with *soil samples* include various types of laboratory-grade ball mills and ring and puck mills. This magnitude of particle reduction reduces *subsampling variability*.

nature and extent decision unit – a *decision unit* (DU) based on the reasonably well-known known location and dimensions of a source area. Synonymous with a source area DU.

nonparametric distribution – when the shape of the *statistical data distribution* curve cannot be plotted by a mathematical formula.

parametric distribution – when the shape of the *statistical data distribution* curve can be plotted by a mathematical formula. Examples of parametric data distributions commonly observed with environmental datasets are normal distributions (another name for bell-shaped curves), lognormal distributions, and gamma distributions.

percent relative standard deviation (%RSD) – a measure of imprecision when two or more replicate procedures are performed. The **RSD** is the *arithmetic standard deviation* divided by the *arithmetic mean* multiplied by 100.

population of soil – a volume of *soil* that shares a common characteristic. The project *decision* to be made defines the soil population of interest, such as the volume of soil exceeding the cleanup criteria or background concentration.

population of potential soil samples – all potential *soil samples* within a *decision unit* (DU), *exposure unit*, or other defining boundary. If a soil sample is considered to be 100 grams of soil in a jar, and the DU is a mass of soil weighing 2000 kg, then 20,000 potential soil samples make up the population defined by the DU.

precision – a measure of reproducibility. Together **precision** and *bias* determine *accuracy*.

quality – the standard of something as measured against other things of a similar kind; the degree of excellence of something.

quality assurance – a management or oversight function that deals with setting policy and running an administrative system of management controls that cover planning, implementation, and review of data collection activities and the use of data in decision-making.

quality control – a technical function that includes all the scientific precautions, such as calibrations and replication, needed to acquire data of known and adequate *quality*.

receptor – a human or ecological individual (e.g., recreational visitor or piping plover) or general ecological population (e.g., benthic invertebrates) that could be exposed to contaminants in environmental media.

relative difference (RD) – a measure of imprecision when only two replicate procedures (i.e., duplicates) were performed. The most common formula for **RD** is to subtract one replicate from the other and divide that difference by the average of the two replicate results. Unlike *relative percent difference*, the fractional result is not then multiplied by 100.

relative percent difference (RPD) – a measure of imprecision that can be used when only two replicate procedures (i.e., duplicates) were performed. The most common formula for **RPD** is to subtract one replicate from the other and divide that difference by the average of the two replicate results. The fractional result is then multiplied by 100. Note the **RPD** is not equal to the *relative standard deviation* of the same two *sample* results.

relative standard deviation (RSD) – a measure of imprecision when two or more replicate procedures were performed. The **RSD** is the *arithmetic standard deviation* divided by the *arithmetic mean*. Also called the coefficient of variation.

replicate samples – two or more independently collected field *samples* or laboratory *subsamples* obtained from the same lot of *soil* by the same sampling or *subsampling* procedure to measure the *precision* of the results. **Replicate samples** are not split but are independently collected *incremental samples*. See also *sampling quality*.

representative soil sample – correctly answers the desired question about a *decision unit* or *sampling unit* with an acceptable level of confidence. A *sample* that is representative to answer one question is not likely to be representative to answer a different question.

representative analytical sample – has the same property of interest as the field *soil sample* from which it is collected.

representativeness – a description of the degree to which an estimate or measurement agrees with the true value of the parameter of interest. The most representative estimate is the one that has the least total error (or greatest *precision* and *accuracy*). "A sample of a universe or whole which can be expected to exhibit the average properties of the universe or whole" 40 CFR § 260.10) (USEPA 2002a).

sample – a small part or quantity intended to show what the whole is like. An *incremental sample* is formed by the reunion of multiple *increments* obtained from a defined volume of *soil* (the *sampling unit* or *decision unit*).

sample support – the size (mass or volume), shape, and orientation of an *increment* or a *sample*.

sampling density – the number of discrete *samples* or *increments* per area or volume of *soil*.

sampling error – anything during *sample* collection and handling that causes the measured properties of a sample to deviate from the actual properties of the population. See also *sampling variability*.

sampling quality – the degree to which evidence demonstrates that all steps related to acquiring representative *samples* from the field and preserving that *representativeness* through laboratory *subsampling* were performed with acceptable *bias* and *precision* in the sample collection, processing, and subsampling.

sampling unit – the volume of *soil* from which *increments* are collected to determine an estimated mean concentration of analytes of interest for that volume of soil.

sampling (or subsampling) variability – imprecision in data results due to various factors in sampling design, field *sample* collection, and laboratory sample processing and *subsampling* procedures. Common causes include insufficient number of *increments*, *incorrect sample support*, and inadequate laboratory sample processing. The term *sampling error* is synonymous with **sampling variability**. The word "error" is commonly used in statistics to refer to variability or imprecision. The degree and sources of **sampling variability** are measured by replicate sampling at various steps in the sampling, processing, and subsampling processes. Variability is typically measured in terms of *standard deviation*, *relative standard deviation*, or *relative percent difference* between *replicate samples*.

short-scale distributional heterogeneity – nonuniformity in the distribution of analytes of interest at spatial scales too small to be relevant at the scale of decision-making. The scale is too small to allow separation of "clean" versus "dirty" *soil* during cleanup and too small to be meaningful to the *receptors* identified during risk assessment. (Note: the meaningful spatial scale can be vastly different depending on the receptor, such as an earthworm vs. a human resident versus a fox.)

slabcake – an entire incremental field sample spread out in a pan (such as a foil-lined cookie sheet) to about ¼- to ½-inch depth.

slabcake subsampling – an entire incremental field *sample* spread out in a pan (such as a foil-lined cookie sheet) to about ¼- to ½-inch depth. At least 30 small *increments* are taken from the full thickness of the slabcake and combined to form the *analytical sample*.

soil – fragmented particulate material consisting of discrete rock and mineral particles less than 2.0 millimeters in size. Small amounts of organic matter (humus). This document uses the term **soil**, understanding that other solid particulate media can also be assessed using this methodology.

specimen – a discrete individual example from a population. The term *specimen* is sometimes used in place of the term *sample* when the data user does not know how a sample was collected and handled to convey the concern that the *representativeness* of the sample or *subsample* relative to the population from which it was take is unknown.

standard deviation – a measure of the dispersion or imprecision of a *sample* or population distribution equal to the

positive square root of the variance.

statistic – a calculated numerical value (such as the *sample* mean) that characterizes some aspect of a sample set of data and that is often meant to estimate the true value of a corresponding parameter (such as the population mean) in an underlying population.

statistical data distribution – assessed by a frequency plot (similar to a histogram) of the data. Frequency plots can take any number of shapes. See also *parametric distribution* and *nonparametric distribution*.

stratification – splitting 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.

stratify – to form, deposit, or arrange into layers.

subsample – a small portion of material selected from a field *sample* by the laboratory for analysis. In ISM, a **subsample** is a representative composite of *increments* collected from an incremental field sample by the laboratory for analysis. See also *aliquot.*

subsampling – in ISM, collecting a small, representative portion of a field *sample* from a processed incremental field sample by spreading the processed field sample in a two-dimensional layer (slabcake) and combining multiple small *increments* taken from random locations through the entire thickness of the layer. Or better, by forming the processed field sample into an elongate pile (one-dimensional line) and collecting increments taken completely through the line from random locations in the *subsample*. See also *slabcake subsampling.*

subsample replicate – two or more *subsamples* taken from a single field sample.

Superfund – the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) established by Congress in 1980. CERCLA is informally called **Superfund** and allows EPA to clean up contaminated sites when responsible parties are either unable or unwilling. It also forces the parties responsible for the contamination to either perform cleanups or reimburse the government for EPA-led cleanup work.

target particle size – the particle size of *soil* that is relevant to the *decision* to be made. Selecting the correct **target particle size** is important because the concentration of soil contaminants generally increases as the particle size fraction analyzed decreases.

theory of sampling – developed by Pierre Gy, a comprehensive approach to representative sampling of bulk particulate materials (including *soil*) that includes a complete analysis of the sources of variability in *sample* results and the *representativeness* of the sampling methods, procedures, and equipment used. 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 and Esbensen 2018).

total sampling error – the cumulative error or variability from all stages of the sampling, processing, and analytical steps.

upper confidence limit (UCL) – a statistical way to derive an upper estimate of the mean. The **UCL** is calculated by adding a "safety factor" to the mean obtained from the *sample* set. The "safety factor" takes into account the number of samples used in the calculation, the variability in the sample results, and the desired level of confidence that the estimate of the mean does not underestimate the true mean. Different mathematical formulas for **UCL** calculation depend on the statistical distribution of the data and the level of confidence desired that the **UCL** is above the true mean.

upper confidence limit, 95% – the calculated statistical value that we are 95% confident above the true value of the mean.**Visual Sample Plan** – a free software program developed by Pacific Northwest National Laboratory that supports the development of defensible sampling plans (including multiple *increment* sampling approaches) based on statistical sampling theory and the statistical analysis of *sample* results to support confident decision-making.

Appendix E: Acronyms

2,4-DNT	2,4-dinitrotoluene
ADEC	Alaska Department of Conservation
AL	action level
AQC	analytical quality control
ASTM	ASTM International, formerly American Society for Testing and Materials
ATON	aid to navigation
BaP	benzo(a)pyrene
BTEX	benzene, toluene, ethylbenzene, and xylenes
BCa	bias-corrected accelerated bootstrap method
CAG	community advisory group
CCA	chromium, copper, and arsenic
CE2	long-range heterogeneity fluctuation error
CE3	periodic heterogeneity fluctuation error
CEC	cation exchange capacity
CFR	Code of Federal Regulations
CH	compositional heterogeneity
CI	confidence interval
CLT	central limit theorem
cm	centimeter
cm^2	centimeters squared
cm^3	centimeters cubed
CMIST	U.S. Army Corps of Engineers multi-increment sampling tool
COC	chemical of concern (or contaminants of concern)
COI	chemical of interest
COPC	chemical of potential concern
CRA	contingency reserve area
CRREL	U.S. Army Cold Regions Research and Engineering Laboratory
CSM	conceptual site model
CV	coefficient of variation
d	diameter
df	degrees of freedom
DGR	Dangerous Goods Regulations
DH	distributional heterogeneity
DHHL	Department of Hawaiian Home Lands
DHS	U.S. Department of Homeland Security

DLNR	State of Hawaii Department of Land and Natural Resources
DNT	dinitrotoluene
DOD	U.S. Department of Defense
DOE	U.S. Department of Energy
DQO	data quality objective
DTSC	California Department of Toxic Substances
DU	decision unit
DUA	data usability assessment
EE	extraction error
ECOS	Environmental Council of the States
EDQW	Environmental Data Quality Working Group
ELAP	Environmental Laboratory Accreditation Program
EPC	exposure point concentration
ERIS	Environmental Research Institute of the States
ESTCP	Environmental Security Technology Certification Program
EU	exposure unit
EU_a	exposure unit for adults
EU_c	exposure unit for children
FDA	U.S. Food and Drug Administration
FDEP	Florida Department of Environmental Protection
FE	fundamental error
FOT	fields of testing
ft^2	square feet
g	gram
GIS	geographic information system
GM	geometric mean
GNSS	global navigation satellite system
GPS	Global Positioning System
GSD	geometric standard deviation
GSE	grouping and segregation error
H_0	null or baseline hypothesis
H_1	alternate hypothesis
HDOH	Hawaii Department of Health
HMX	octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine
IA	investigation area
IATA	International Air Transport Association

IBT	internet-based training
IDQTF	Intergovernmental Data Quality Task Force
IS	incremental sample (or incremental sampling)
ISM	Incremental Sampling Methodology
ISO	International Organization for Standardization
ITRC	Interstate Technology & Regulatory Council
KM	Kaplan-Meier
	lower confidence limit
LCS	laboratory control sample (or laboratory control spike)
LDR	Land Disposal Restrictions
LORAN	Long Range Navigation
LUST	leaking underground storage tank
M	maps
MARLAP	Multi-Agency Radiological Laboratory Analytical Protocols Manual
MDL	method detection limit
MEDEP	Maine Department of Environmental Protection
mg/kg	milligram per kilogram
MI	multiple increment
mm	millimeter
mm Hg	millimeter mercury vapor pressure
MMRP	Military Munitions Response Program
MPC	measurement performance criteria
MQO	measurement quality objective
Ms	mass of the collected sample
MS	matrix spike
MSD	matrix spike duplicate
n	number of increments or sample size
NAEG	Nevada Applied Ecology Group
nCi	nanocurie
ND	non-detect
NELAP	National Environmental Laboratory Accreditation Program
N&E DU	nature and extent decision unit
NIST	U.S. National Institute of Standards and Technology
NRC	Nuclear Regulatory Commission
OCP	organochlorine pesticides
OLEM	Office of Local Environmental Management
ORD	USEPA Office of Research and Development

OSWER	Office of Solid Waste and Emergency Response
oz	ounce
PAH	polyaromatic hydrocarbon
PAL	project action level
Pb	lead
PCB	polychlorinated biphenyl
PCP	pentachlorophenol
PD	probability distribution
PETN	pentaerythritol tetranitrate
PFAS	per- and polyfluoroalkyl substances
PHC	petroleum hydrocarbons
PID	photo-ionization detector
ppm	parts per million
PNNL	Pacific Northwest National Laboratory
ProUCL	USEPA statistical software package for analysis of environmental datasets
QA	quality assurance
QAPP	quality assurance project plan
QC	quality control
QSM	Quality Systems Manual
r	replicates
RAG	Remedial Action Guideline or Risk Assessment Guidance
RCRA	Resource Conservation and Recovery Act
RDX	1,3,5-trinitroperhydro-1,3,5-triazine
RL	reporting limit
RPD	relative percent difference
RSD	relative standard deviation
Sb	antimony
SD	standard deviation
SE	sampling error (or standard error)
SIM	selective ion monitoring
SMD	scrap metal dump
SOP	standard operating procedure
SPP	systematic planning process
sqrt	square root
SRM	standard reference manual
SU	sampling unit
SVOC	semi-volatile organic compound

TAG	technical assistance grantee
TCEQ	Texas Commission on Environmental Quality
TCLP	Toxicity Characteristic Leaching Procedure
TGM	Technical Guidance Manual
TOS	theory of sampling
TPP	technical project planning
t-UCL	Student's-*t*
μ	population mean
UCL	upper confidence limit
UFP-QAPP	uniform federal policy for quality assurance project plans
UPL	upper prediction limits
USACE	U.S. Army Corps of Engineers
USDA	U.S. Department of Agriculture
USEPA	U.S. Environmental Protection Agency
UTL	upper tolerance limit
UV	ultraviolet
VOC	volatile organic compound
VSP	Visual Sample Plan
WP	work plan XRF X-ray fluorescence

References

Anderson, Bryan. 2017. “Taxpayer dollars fund most oversight and cleanup costs at Superfund sites.” *The Washington Post*, September 20, 2017, National. https://www.washingtonpost.com/national/taxpayer-dollars-fund-most-oversight-and-cleanup-costs-at-superfund-sites/2017/09/20/aedcd426-8209-11e7-902a-2a9f2d808496_story.html.

ASTM. 1996. “Standard Guide for Representative Sampling for Management of Waste and Contaminated Media.” D6044-96.

Becker, Joanna M. 2005. “PhD thesis: Centimeter scale analysis of soil heterogeneities within a long-term, heavy metal contaminated site.”, Purdue Univ. .

Brewer, Roger, John Peard, and Marvin Heskett. 2016. “A Critical Review of Discrete Soil Sample Data Reliability: Part 1—Field Study Results.” *Soil and Sediment Contamination: An International Journal* 26:1:1-22. doi: 10.1080/15320383.2017.1244171.

Brewer, Roger, John Peard, and Marvin Heskett. 2017. “A Critical Review of Discrete Soil Sample Data Reliability: Part 2—Implications.” *Soil and Sediment Contamination: An International Journal* 26:1:23-44. doi: 10.1080/15320383.2017.1244172.

Bruce, M. L. 2003. “Turning Dirt into Dust, Evaluation of Kitchen Appliances.” Louisville Chemistry Conference, Kentucky.

Clausen, J. L., T. Georgian, K. H. Gardner, and T. A. Douglas. 2018a. “Applying Incremental Sampling Methodology to Soils Containing Heterogeneously Distributed Metallic Residues to Improve Risk Analysis.” *Bulletin of Environmental Contamination and Toxicology* 100 (1):155-161. doi: 10.1007/s00128-017-2252-x.

Clausen, J. L., T. Georgian, K. H. Gardner, and T. A. Douglas. 2018b. “Inadequacy of Conventional Grab Sampling for Remediation Decision-Making for Metal Contamination at Small-Arms Ranges.” *Bull Environ Contam Toxicol* 100 (1):147-154. doi: 10.1007/s00128-017-2255-7.

Clausen, J. L., T. Georgian, K. H. Gardner, and T. A. Douglas. 2018c. “Optimization of Field and Laboratory Sample Processing for Characterization of Metallic Residues at Military Training Ranges.” *Bull Environ Contam Toxicol* 100 (5):603-608. doi: 10.1007/s00128-018-2311-y.

Clausen, J.L., B. Swope, A. Bednar, S. Beal, D. Rosado, L. Leavitt, T. Cary, M. Catt, K. Armstrong, N. Parker, and A. Staples. 2016. Impact of the Incremental Sampling Methodology on Metal Bioavailability Assessments. edited by Cold Regions Research and Engineering Laboratory.

Clausen, Jay, Thomas Georgian, and Anthony Bednar. 2013. *Cost and Performance Report of Incremental Sampling Methodology for Soil Containing Metallic Residues*. Vol. ERDC.

Clausen, Jay, Thomas Georgian, Anthony Bednar, Nancy Perron, Andrew Bray, Patricia Tuminello, Gordon Gooch, Nathan Mulherin, Arthur Gelvin, Marc Beede, Stephanie Saari, William Jones, and Shawna Tazik. 2013. Demonstration of Incremental Sampling Methodology for Soil Containing Metallic Residues.

Clausen, Jay L. , Thomas Georgian, Julie Richardson, Anthony Bednar, Nancy Perron, and Andrew Bray. 2012. Evaluation of Sampling and Sample Preparation Modifications for Soil Containing Metallic Residues

Cochran, W.G. 1977. *Sampling Techniques – 3rd Edition*, *Wiley series in probability and mathematical statistics*: John Wiley & Sons, Inc.

Conservation, Alaska Department of Environmental. 2019. Field Sampling Guidance. edited by Division of Spill Prevention and Response Contaminated Sites Program.

Crumbling, Deana. 2014. *“Hot Spots: Incremental Sampling Methodology (ISM) FAQs.” Unpublished peer-reviewed white paper, https://www.itrcweb.org/FileCabinet/GetFile?fileID=20542”.*

Crumbling, Deana. 2019. “Personal communication of PAH data analysis for a landfill Superfund site in EPA Region 3.”.

Doctor, P.G, and R.O. Gilbert. 1978. Two Studies in Variability for Soil Concentrations: with Aliquot Size and with Distance. In *Selected Environmental Plutonium Research Reports of the NAEG*. Battelle Memorial Institute, Pacific Northwest Laboratory, Richland, Washington. .

DOD/DOE. 2018. DOD DOE Consolidated Quality Systems Manual (QSM) for Environmental Laboratories. In *DOD/DOE QSM*

DOT. 2011. 49 CFR Part 172 – HAZARDOUS MATERIALS TABLE, SPECIAL PROVISIONS, HAZARDOUS MATERIALS COMMUNICATIONS, EMERGENCY RESPONSE INFORMATION, TRAINING REQUIREMENTS, AND SECURITY PLANS.

DOT. 2019. Transportation, PART 173—SHIPPERS—GENERAL REQUIREMENTS FOR SHIPMENTS AND PACKAGINGS. In *49 CFR 173.4a (c)(2))*. .

Dowson, S. T. , N. L. Hassig, C. J. Murray, B. A. Pulsipher, and S. McKenna. 2007. Visual Sample Plan Version 5.0 User's Guide.

DTSC, CA. 2001. California Department of Toxic Substances Information Advisory: Clean Imported Fill Material

DTSC, CA. 2018. California Department of Toxic Substances: Recommendations for Evaluating Polychlorinated Biphenyls (PCBs) at Contaminated Sites in California. HHRA Note Number 8. edited by DTSC's HUMAN AND ECOLOGICAL RISK OFFICE (HERO).

DTSC, CA. 2019. California Department of Toxic Substances Information Advisory: Guidance for Screening Level Human Health Risk Assessments. HHRA Note Number 4.

Fehsenfeld, Fred, Jack Calvert, Ray Fall, Paul Goldan, Alex Guenther, and C. Nicholas Hewitt. 1992. "Emissions of volatile organic compounds from vegetation and the implications for atmospheric chemistry." *Global Biogeochemical Cycles*. doi: https://doi.org/10.1029/92GB02125.

Frederick, Timothy, Alicia Frame, and Daniel Vallero. 2017. "Comparison of soil sampling and analytical methods for asbestos at the Sumas Mountain Asbestos Site—Working towards a toolbox for better assessment ".

Fye, J. L., H. H. Nelson, R. L. Mowery, A. P. Baronavski, and J. H. Callahan. 2002. "Scanning ultraviolet two-step laser mass spectroscopy of polycyclic aromatic hydrocarbon distributions on creosote-contaminated soil particles." *Anal Chem* 74 (13):3019-29. doi: 10.1021/ac025510l.

Georgian, T. 2020. "Title." Unpublished white paper.

Gerlach, R. W., and J. M. Nocerino. 2003. "Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples " EPA/600/R-03/027.

Gilbert, R.O. 1987. "Statistical Methods for Environmental Pollution Monitoring."

Goodrum, PE, RA Budinsky, E Mendelsohn, and H Summers. 2018. "Use of Thyroid Disease Incidence and Dose-Response Analysis to Reduce Uncertainty in the Dioxin Oral Reference Dose." *Organohalogen Compounds* Vol. 80:429-432.

Goodrum, Philip , and Emma Mendelsohn. 2018. "Title." Organohalogen Compounds.

Gy, Pierre. 1953. "Erreur Commise dans le Prelevement d'un Echantillon sur un Lot de Minerai' (Error made when sampling a mineral lot)." Congres des laveries des mines metalliques français, Paris.

Gy, Pierre. 1988. *Ensemble Coherent de Theories (Heterogeneity, Sampling, Homogenisation. Their Logical Integration)*, *Heterogeneity, Echantillonage, Homogeneisation*. Masson, Paris.

Hadley, P.W., E. Crapps, and A.D. Hewitt. 2011. "Time for a Change of Scene." *Environmental Forensics* 12:4:312-318.

Hadley, Paul W., and Stephen D. Mueller. 2012. "Evaluating "Hot Spots" of Soil Contamination (Redux)." *Soil and Sediment Contamination: An International Journal* 21 (3):335-350. doi: 10.1080/15320383.2012.664431.

Hartmann, H.M., J.P. Butler, A.M. Dorries, and J.R. Beck. 1993. . Use of the Exposure Unit Concept in Risk Assessments: A Case Study for Los Alamos National Laboratory. Los Alamos, NM: Los Alamos National Laboratory.

Hassellov, M. , and F. von der Kammer. 2008. "Iron oxides as geochemical nanovectors for metal transport in soil-river systems. ." doi: DOI: 10.2113/gselements.4.6.401.

HDOH. 2011. Technical Guidance Manual Notes: Decision Unit and Multi-Increment Sample Investigations Roger Brewer & John Peard. edited by Hazard Evaluation and Emergency Response (HEER).

HDOH. 2015. Small-Scale Variability of Discrete Soil Sample Data, Part 2: Causes and Implications for Use in Environmental Investigations.

HDOH. 2016a. Technical Guidance Manual for the Implementation of the Hawai'i State Contingency Plan. edited by Office of

Hazard Evaluation and Emergency Response.

HDOH. 2016b. Technical Guidance Manual for the Implementation of the Hawai'i State Contingency Plan. edited by Office of Hazard Evaluation and Emergency Response.

HDOH. 2017a. Guidance for Soil Stockpile Characterization and Evaluation of Imported and Exported Fill Material edited by Office of Hazard Evaluation and Emergency Response.

HDOH. 2017b. Technical Guidance Manual for the Implementation of the Hawai'i State Contingency Plan. edited by Office of Hazard Evaluation and Emergency Response.

Hewitt, A. D., T. F. Jenkins, and C. L. Grant. 1995 "Collection, Handling, and Storage: Keys to Improved Data Quality for Volatile Organic Compounds in Soil." *American Environmental Laboratory* 7:25–28.

Hewitt, Alan, Thomas Jenkins, Charles Ramsey, Kevin Bjella, Thomas Ranney, and Nancy Perron. 2005. "Estimating Energetic Residue Loading on Military Artillery Ranges: Large Decision Units."63.

Homsher, M.T. , F. Haeberer, P.J. Marsden, R.K. Mitchum, D. Neptune, and J. Warren. 1991. "Performance Based Criteria, A Panel Discussion, Part I and II." *Environmental Lab*

IATA. 2011. "International Air Transport Association Dangerous Goods Regulations (DGR)."

IDEQ. 2015. Idaho Deparment of Environmental Quality – Pile Sampling White Paper. .

ITRC. 2007a. Improving Environmental Site Remediation Through Performance-Based Environmental Management. RPO-7.

ITRC. 2007b. Triad Implementation Guide. Sampling, Characterization, and Monitoring (SCM-3)

ITRC. 2008. Use of Risk Assessment in Management of Contaminated Sites. RISK-2. Washington, D.C.: Interstate Technology & Regulatory Council, Risk Assessment Resources Team.

ITRC. 2012. Incremental Sampling Methodology (ISM) Document.

ITRC. 2015. Decision Making at Contaminated Sites: Issues and Options in Human Health Risk Assessment. RISK-3.

ITRC. 2019. Implementing Advanced Site Characterization Tools. ASCT-1.

ITRC. 2020. Risk Communication Toolkit. RCT-1. Washington, D.C.: Interstate Technology & Regulatory Council.

Jenkins, T. F., A. D. Hewitt, Marianne Walsh, S. Thiboutot, G. Ampleman, T. A. Ranney, and J. C. Pennington. 2004. "Distribution of energetic compounds in soils at training ranges." *Proceedings of the First Conference on Sustainable Range Management – 2004*:1792-1832.

Mattuck, Rosemary, Richard Blanchet, and A. Dallas Wait. 2005. "Data Representativeness for Risk Assessment." *Environmental Forensics* 6 (1):65-70. doi: 10.1080/15275920590913886.

MDEQ. 2018. Michigan Department of Environmental Quality – Incremental Sampling Methodology and Applications. edited by Resource Materials Remediation and Redevelopment Division. Lansing, Michigan.

MEDEP. 2015. Maine Department of Environmental Protection. Standard Operating Procedure. .

Minkkinen, P.O., and K.H. Esbensen. 2018. "Sampling of particulate materials with significant spatial heterogeneity – Theoretical modification of grouping and segregation factors involved with correct sampling errors." *Analytica Chimica Acta*.

Minkkinen, Pentti. 2004. "Practical applications of sampling theory." *Chemometrics and Intelligent Laboratory Systems* 74 (1):85-94. doi: https://doi.org/10.1016/j.chemolab.2004.03.013.

Minnitt, R.C.A., P.M. Rice, and C. Spangenberg. 2007. "Part 1: Understanding the components of the fundamental sampling error: a key to good sampling practice." *The Journal of The Southern African Institute of Mining and Metallurgy*.

OEPA. 2016. Ohio Environmental Protection Agency. Incremental Sampling for Soils and Sediments.

Owens, Kathy. 2020. "Personal communications with RPM for EPA Region 3."

Petersen, L. , P. Minkkinen, and K.H. Esbensen. 2005 "Representative sampling for reliable data analysis: Theory of Sampling. ." *Chemometrics and Intelligent Laboratory Systems* 77:261– 277.

Pitard, F. 2019. *Theory of Sampling and Sampling Practice*. 3rd ed: Chapman and Hall/CRC.

Pooler, P.S, P.E Goodrum, D Crumbling, L.D Stuchal, and S.M Roberts. 2018. “Incremental Sampling Methodology: Applications for Background Screening Assessments.” *Risk Analysis* Volume38 (Issue1):Pages 194-209.

Qiao, Hong Xia, Trenton Pulsipher, John E. Hathaway, Eric E. Richman, and Emil Vergilov Radkov. 2010. “A Statistical Method Analyzing LED Lumen Depreciation and Projecting LED Life.”

Ramsey, C.A., and A.D. Hewitt. 2005. “A Methodology for Assessing Sample Representativeness. .” *Environmental Forensics* 6:71-75. doi: 10.1080/15275920590913877.

Ramsey, F., and D. Schafer. 2012. *The statistical sleuth: a course in methods of data analysis.* : Cengage Learning.

Singh, A., R. Maichle, and N. Armbya. 2007. ProUCL Version 4.1, User Guide. Statistical Software for Environmental Applications for Data Sets with and without Nondetect Observations. edited by USEPA – Office of Research and Development.

USACE. 1998. Environmental Quality Technical Project Planning Process (updated 2016).

USACE. 2007. Protocols for Collection of Surface Soil Samples at Military Training and Testing Ranges for the Characterization of Energetic Munitions Constituents.

USACE. 2009. Implementation of Incremental Sampling (IS) of Soil for the Military Munitions Response Program. In *Environmental and Munitions Center of Expertise Interim Guidance*.

USACE. 2012. Conceptual Site Models. EM 200-1-12.

USACE. 2013a. Cost and performance report of Incremental Sampling Methodology for soil containing metallic residues: Project ER-0918. edited by Engineer Research and Development Center (U.S.) United States. Army. Corps of Engineers, Cold Regions Research and Engineering Laboratory (U.S.).

USACE. 2013b. Demonstration of Incremental Sampling Methodology for soil containing metallic residues: Project ER-0918. edited by Engineer Research and Development Center (U.S.) Army. Corps of Engineers, Cold Regions Research and Engineering Laboratory (U.S.).

USACE. 2013c. Incremental Sampling Methodology (ISM) for metallic residues: Project ER-0918. edited by Engineer Research and Development Center (U.S.) United States. Army. Corps of Engineers, Cold Regions Research and Engineering Laboratory (U.S.).

USEPA. 1989a. Risk Assessment Guidance for Superfund, Vol. I, Human Health Evaluation Manual (Part A)

USEPA. 1989b. Risk Assessment Guidance for Superfund, Vol. II Environmental Evaluation Manual.

USEPA. 1989c. Soil Sampling Quality Assurance User’s Guide, Second Edition. Las Vegas, NV: Environmental Monitoring Systems Laboratory.

USEPA. 1992a. Correct Sampling Using the Theories of Pierre Gy. . edited by Environmental Sciences Division EPA Office of Research and Development, Technology Support Project.

USEPA. 1992b. PREPARATION OF SOIL SAMPLING PROTOCOLS: SAMPLING TECHNIQUES AND STRATEGIES.

USEPA. 1992c. Statistical Methods For Evaluating The Attainment Of Cleanup Standards Volume 3 Reference-Based Standards For Soils And Solid Media. Policy, Planning and Evaluation.

USEPA. 1994. SW-846 Method 8330 Test Methods for Evaluating Solid Waste: Physical/Chemical Methods.

USEPA. 1997. Ecological Risk Assessment Guidance for Superfund: Process for Designing and Conducting Ecological Risk Assessments – Interim Final.

USEPA. 1999a. Correct Sampling Using the Theories of Pierre Gy. OFFICE OF RESEARCH AND DEVELOPMENT, Environmental Sciences Division.

USEPA. 1999b. SW-846 Test Method 8095: Explosives by Gas Chromatography.

USEPA. 2002a. 40 CFR §260.10 – HAZARDOUS WASTE MANAGEMENT SYSTEM: GENERAL. In *40 CFR §260.10*

USEPA. 2002b. Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites. Office of Emergency and Remedial Response.

USEPA. 2002c. Guidance on Choosing a Sampling Design for Environmental Data Collection EPA QA/G-5S. edited by Office of

Environmental Information. Washington, DC. .

USEPA. 2002d. Guidance on Choosing a Sampling Design for Environmental Data Collection for Use in Developing a Quality Assurance Project Plan.

USEPA. 2002e. Guidance on Choosing a Sampling Design for Environmental Data Collection for Use in Developing a Quality Assurance Project Plan.

USEPA. 2002f. “Method 5035A (SW-846): Closed-System Purge-and-Trap and Extraction for Volatile Organics in Soil and Waste Samples.”

USEPA. 2002g. RCRA Waste Sampling Draft Technical Guidance. Planning, Implementation, and Assessment. Solid Waste and Emergency Response (5305W).

USEPA. 2003. Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples. Office of Research and Development.

USEPA. 2004a. Multi-Agency Radiological Laboratory Analytical Protocols Manual (MARLAP).

USEPA. 2004b. Multi-Agency Radiological Laboratory Analytical Protocols Manual (MARLAP).

USEPA. 2005a. Reference Guide to Non-combustion Technologies for Remediation of Persistent Organic Pollutants in Stockpiles and Soil

USEPA. 2005b. Uniform Federal Policy for Quality Assurance Project Plans. edited by Intergovernmental Data Quality Task Force (IDQTF).

USEPA. 2006a. Applicability of Superfund Data Categories to the Removal Program. Office of Solid Waste and Emergency Response. Office of Emergency Management.

USEPA. 2006b. Guidance on Systematic Planning Using the Data Quality Objectives Process. .

USEPA. 2006c. Method 8330B (SW-846): Nitroaromatics, Nitramines, and Nitrate Esters by High Performance Liquid Chromatography (HPLC).

USEPA. 2007. SW-846 Test Method 6200 Field Portable X-Ray Fluorescence Spectrometry for the Determination of Elemental Concentrations in Soil and Sediment.

USEPA. 2009. Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities – Unified Guidance. OFFICE OF RESOURCE CONSERVATION AND RECOVERY, PROGRAM IMPLEMENTATION AND INFORMATION DIVISION.

USEPA. 2010. Best Management Practices: Use of Systematic Project Planning Under a Triad Approach for Site Assessment and Cleanup. edited by Office of Solid Waste and Emergency.

USEPA. 2011. Uniform Federal Policy Quality Assurance Project Plan Template For Soils Assessment of Dioxin Sites – User Guide –

USEPA. 2012a. Protection of the Environment. In *Title 40 CFR. 2012. Section 260.10:16*, edited by USEPA. US.

USEPA. 2012b. Site Characterization For Munitions Constituents. In *EPA Federal Facilities Forum Issue Paper*.

USEPA. 2013. Use of Amendments for In Situ Remediation at Superfund Sediment Sites.

USEPA. 2015. ProUCL Version 5.1, Technical Guide, Statistical Software for Environmental Applications for Data Sets with and without Nondetect Observations. Office of Research and Development.

USEPA. 2016. Recommendations for Sieving Soil and Dust Samples at Lead Sites for Assessment of Incidental Ingestion. Office of Superfund Remediation and Technology Innovation

USEPA. 2019. Incremental Sampling Methodology (ISM) at Polychlorinated Biphenyl (PCB) Cleanup Sites

Visual Sample Plan: A Tool for Design and Analysis of Environmental Sampling Version 7.12a. Pacific Northwest National Laboratory, Richland, WA. .

Walsh, Marianne, Charles Collins, Ronald Bailey, and Clarence Grant. 1997. Composite Sampling of Sediments Contaminated with White Phosphorus (CRREL).

Walsh, Marianne, and Dennis Lambert. 2006. “Extraction Kinetics of Energetic Compounds from Training Range and Army

Ammunition Plant Soils."46.

WDOE. 2019. Washington Department of Ecology. Sediment Cleanup User's Manual. .

Weng, Liping, Erwin J. M. Temminghoff, and Willem H. Van Riemsdijk. 2001. "Contribution of Individual Sorbents to the Control of Heavy Metal Activity in Sandy Soil." *Environmental Science & Technology* 35 (22):4436-4443. doi: 10.1021/es010085j.

Xiong, Juan, Liping Weng, Luuk K. Koopal, Mingxia Wang, Zhihua Shi, Lirong Zheng, and Wenfeng Tan. 2018. "Effect of Soil Fulvic and Humic Acids on Pb Binding to the Goethite/Solution Interface: Ligand Charge Distribution Modeling and Speciation Distribution of Pb." *Environmental Science & Technology* 52 (3):1348-1356. doi: 10.1021/acs.est.7b05412.

Acknowledgments

The members of the Interstate Technology & Regulatory Council (ITRC) Incremental Sampling Methodology (ISM-2) Update Team wish to acknowledge the individuals, organizations, and agencies that contributed to this guidance document.

As part of the broader ITRC effort, the ISM Update Team effort is funded primarily by the U.S. Department of Energy and the U.S. Department of Defense.

The ISM-2 Team greatly appreciates the efforts of the team members who provided valuable written input in the development of this technical and regulatory guidance, as well as those who reviewed and commented on this document. A full team list is provided in Appendix C.

The team recognizes the efforts of the following personnel who contributed to content development of this guidance:

Team Leaders
Caroline Jalanti, New York State Department of Environmental Conservation
Troy Keith, Tennessee Department of Environmental Conservation

State Regulatory Representatives
Chris Christensen, Michigan Department of Environmental Quality
Karen DiBiasio, California Environmental Protection Agency
Richard Doucette, Virginia Department of Environmental Quality
David Elias, California State Water Board
Calvin Fales, Missouri Department of Natural Resources
Justin Fisher, Tennessee Department of Environmental Conservation
Joey Pace, Arizona Department of Environmental Quality
Craig Sanchez, California State Water Board
Shanna Schmitt, Minnesota Pollution Control Agency
Richard Sloan, Montana Department of Environmental Quality
Michael Stroh, Missouri Department of Natural Resources
Eric White, Texas Commission on Environmental Quality

Academic Participants
J. Hathaway, Brigham Young University
Mark Leeper, Colorado State University
Leah Stuchal, University of Florida

Stakeholder Participants
Katherine Owens, Eastern Idaho Community Action Partnership

Emeritus Stakeholder Participants
Deana Crumbling, Public
Eric Nuttall, University of New Mexico

Agency Participants (including federal, city/local government, and international government)
Cynthia Auld, U.S. DOD, U.S. Army Corps of Engineers
Jay Clausen, U.S. DOD, U.S. Army Corps of Engineers, Engineer Research and Development Center
Ed Corl, U.S. DOD, NAVSEA, Laboratory Quality and Accreditation Office
Tim Frederick, U.S. EPA, USEPA Region 4
Thomas Georgian, U.S. DOD, U.S. Army Corps of Engineers
Matt Lambert, U.S. EPA, Office of Superfund Remediation and Technology Innovation
Cheryl R. Montgomery, U.S. DOD, U.S. Army Corps of Engineers, Engineer Research and Development Center
Hugh Rieck, U.S. DOD, U.S. Army Corps of Engineers
Judith Strawhecker, U.S. DOD, U.S. Army Corps of Engineers

Industry Participants
Christopher Bartley, Terracon
Kelly Black, Neptune and Company, Inc.
Hayley Brittingham, Neptune and Company, Inc.
Jason Brodersen, Tetra Tech, Inc.
Mark Bruce, Eurofins
Douglas Burge, Ramboll
Allison Drouin, Credere Associates, LLC
Shannon Dunn, Arcadis

Stephen Foster, Geosyntec Consultants
Philip Goodrum, GSI Environmental Inc.
Betty Krupka, CDM Smith
Thomas Maher, Civil & Environmental Consultants, Inc
Alicia McCabe, Ramboll
Jim McDade, GSI Environmental Inc.
Todd Miller, Kennedy/Jenks Consultants
John Montgomery-Brown, EKI Environment & Water, Inc.
Ralph Perona, Neptune and Company, Inc.
Christian Romero, Noster Tec
Loren Ruttinger, WWC Engineering
Tamara Sorell, Brown and Caldwell
Michele Zych, Wood PLC

Program Advisor

Susan Schow, Schow Consulting