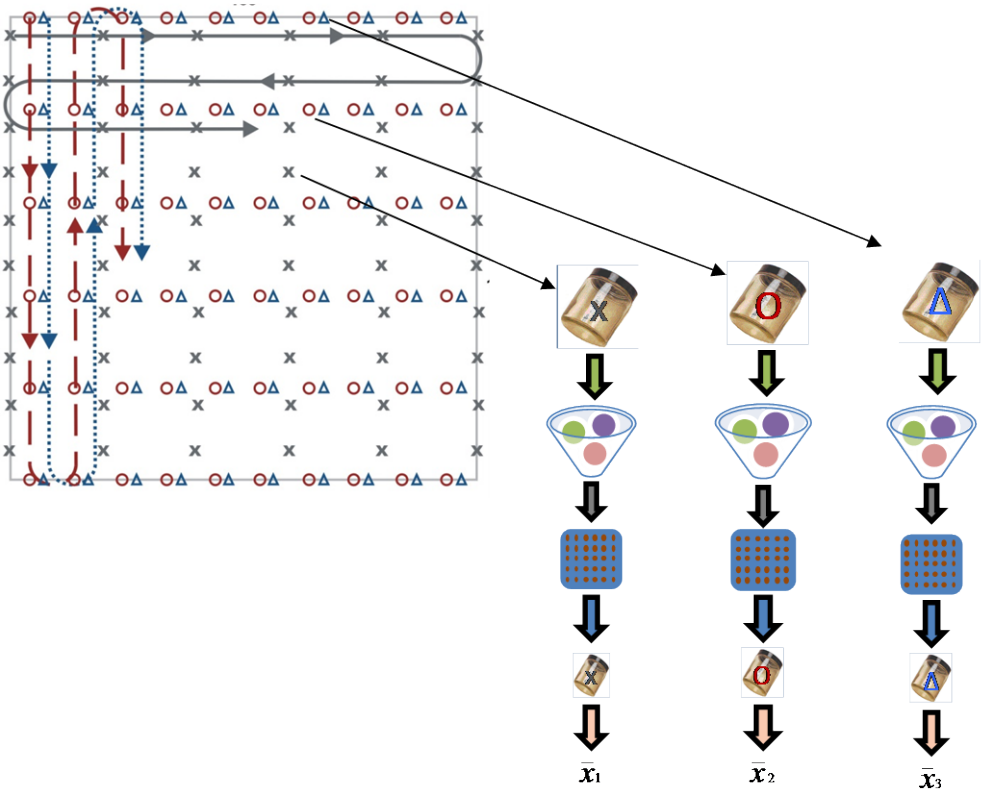
Technical and Regulatory Guidance

Incremental Sampling Methodology



February 2012

Prepared by
The Interstate Technology & Regulatory Council
Incremental Sampling Methodology Team

ABOUT ITRC

The Interstate Technology and Regulatory Council (ITRC) is a public-private coalition working to reduce barriers to the use of innovative environmental technologies and approaches so that compliance costs are reduced and cleanup efficacy is maximized. ITRC produces documents and training that broaden and deepen technical knowledge and expedite quality regulatory decision making while protecting human health and the environment. With private- and public-sector members from all 50 states and the District of Columbia, ITRC truly provides a national perspective. More information on ITRC is available at www.itrcweb.org.

ITRC is a program of the Environmental Research Institute of the States (ERIS), a 501(c)(3) organization incorporated in the District of Columbia and managed by the Environmental Council of the States (ECOS). ECOS is the national, nonprofit, nonpartisan association representing the state and territorial environmental commissioners. Its mission is to serve as a champion for states; to provide a clearinghouse of information for state environmental commissioners; to promote coordination in environmental management; and to articulate state positions on environmental issues to Congress, federal agencies, and the public.

DISCLAIMER

ISM-1

Incremental Sampling Methodology

February 2012

Prepared by
The Interstate Technology & Regulatory Council
Incremental Sampling Methodology Team

ACKNOWLEDGEMENTS

The members of the ITRC Incremental Sampling Methodology (ISM) Team wish to acknowledge the individuals, organizations, and agencies that contributed to this technical and regulatory guidance.

As part of the broader ITRC effort, the ISM Team's effort is funded primarily by the U.S. Department of Energy. Additional funding and support have been provided by the U.S. Department of Defense and the U.S. Environmental Protection Agency.

While all team members listed below contributed to the team's efforts, the co-leaders (Ligia Mora-Applegate and Mark Malinowski) recognize the team members identified with the symbols *****, †, §, or Δ as deserving special recognition as follows:

***** notable writing contributions
† leadership role during document production
§ multiyear member of the team
Δ Internet-based training instructor

The ISM Team co-leaders recognize the efforts of the following individuals:

- Diane Anderson§Δ, Agriculture & Priority Pollutants Laboratory, Inc.
- Matthew Baltusis§, Michigan Department of Environmental Quality-RD-Superfund
- Alfonso Benavides, Texas A&M University
- Nancy Bettinger*, Massachusetts Department of Environmental Protection
- Kelly Black*§Δ, Neptune and Company, Inc.
- Robin Boyd*§Δ, AECOM Environment
- Roger Brewer*§, Hawaii Department of Health
- Mark Bruce*§†Δ, TestAmerica, Inc.
- Matt Butcher, ARCADIS
- Anna Butler§, U.S. Army Corps of Engineers (USACE)
- Frank Camera*§†Δ, New Jersey Department of Environmental Protection
- Janie Carrig§, USACE Omaha District-MM Design Center
- Michelle Caruso§, Tetra Tech, Inc.
- Michael Chacon, Pueblo de San Ildefonso
- Arnab Chakrabarti, Geosyntec Consultants
- Eric Cheng, USACE
- Jay Clausen§Δ, U.S. Army Cold Regions Research and Engineering Laboratory
- Jennifer Corack§, Navy and Marine Corps Public Health Center
- Ed Corl*§, NAVFAC Engineering Command Laboratory Quality and Accreditation Office
- Earl Crapps*§†Δ, Alaska Department of Environmental Conservation

- Deana Crumbling*§†Δ, U.S. Environmental Protection Agency (US EPA), Office of Superfund Remediation and Technology Innovation (OSRTI), Technology Innovation and Field Services Division
- Annette Dietz*§Δ, Oregon Department of Environmental Quality
- Jan Dunker*§ Δ, USACE
- John Esparza, USACE
- Timothy Fitzpatrick§, Florida Department of Environmental Protection (DEP)
- Tim Fredrick§Δ, USEPA Region 4
- Jon Gabry*§, Tetra Tech, Inc.
- Rick Galloway*§†, Delaware Department of Natural Resources and Environmental Control
- Edward Gilbert§, US EPA OSRTI, Environmental Response Team
- John Gillette§, Air Force Center for Engineering and the Environment (AFCEE)
- Phillip Goodrum*§†Δ, Cardno ENTRIX
- Joe Guarnaccia, BASF Corporation
- Paul Hadley*§†Δ, California Department of Toxic Substances Control (DTSC)
- Jeanene Hanley§, Arizona Department of Environmental Quality (ADEQ)
- Bryan Harre§, Naval Facilities Engineering Service Center (NAVFAC ESC)
- John Hathaway*§, Pacific Northwest National Laboratory
- Marvin Heskett, TestAmerica, Inc.
- Keith Hoddinott§, U.S. Army Public Health Command
- Julie Hoskin§, ADEQ
- Paul Jurena§, Past Program Advisor, AFCEE Technical Directorate
- Guy Kaminski, ARCADIS
- Derek Kinder§, USACE Louisville District Environmental Engineering Branch
- Robert Kirgan§, U.S. Army Environmental Center (USAEC)
- Richard Lewis§, HSA Engineers & Scientists
- Peter Lorey§, USACE
- Pornteera Manakun, ERM
- Joseph McElhaney, ExxonMobil
- Kari Meier§, USACE—Environmental Engineering
- Anita Meyer*§, USACE
- Brian Nagy, TestAmerica, Inc.
- Neal Navarro§, USACE
- Osaguona Ogbebor§, CH2M HILL
- Dee O'Neill, Columbia Analytical Services
- Katherine Owens*§, Eastern Idaho Community Action Partnership
- Jeffery Patterson*§†, Texas Commission on Environmental Quality
- Mike Pfister§, California DTSC
- Hugh Rieck§, USACE
- Stephen Roberts*§†, University of Florida
- Randall Ryti§, Neptune and Company, Inc.
- Christopher Saranko§, Environmental Planning Specialists, Inc.
- Angela Sederquist, Program Advisor, Booz Allen Hamilton
- Kathleen Siebenmann, USACE

- George Shaw§, W.L . Gore & Associates, Inc.
- Lizanne Simmons, Kleinfelder, Inc.
- Anita Singh*§, Lockheed Martin
- Patricia L. Smith, Alpha Stat Consulting
- Michael Stroh*§† Δ, Missouri Department of Natural Resources
- Leah Stuchal*§, University of Florida
- Brandon Swope§, SPAWAR Systems Center Pacific
- William Sy§, US EPA Region 2, Superfund Quality Assurance (QA)
- Keith Tolson*§, Geosyntec Consultants
- Deborah Walker§, U.S. Army Engineering and Support Center
- Barry Weand, Past Program Advisor, Team Integrated Engineering
- Gary Zimmerman§, Golder Associates

All parties who contributed to this document whether named or unnamed, be they team member, independent reviewer, or ITRC staff, are thanked by the ISM Team for their effort. Some made major contributions to the project while others made minor ones; all are appreciated for their time and effort.

DEDICATION

This document is dedicated to our friend and colleague

Alan D. Hewitt

whose contributions to environmental science will endure far beyond his time. We admired him for his intellect and enthusiasm and are grateful for the knowledge and energy he brought to this team.

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 (DUs) by collecting numerous increments of soil (typically 30–100 increments) that are combined, processed, and subsampled according to specific protocols.

ISM is increasingly being used in the environmental field for sampling 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 (i.e., discrete) sampling approaches.

In 2009 the ITRC established a technical team to evaluate ISM for sampling soils at hazardous waste sites and potentially contaminated properties. The ISM Team convened national experts in fields such as toxicology, risk assessment, statistics, and soil sampling. Key efforts of the ISM Team included performing a statistical analysis of ISM performance, identifying considerations for unique laboratory processes and procedures, evaluating the suitability of ISM to various contamination scenarios and contaminant categories, and identifying the strengths and weaknesses of ISM.

A key feature of the ISM Team's effort was emphasizing the need to integrate systematic planning for any soil sampling approach. As with any sampling approach, ISM requires the integration of quantitative soil sampling objectives with the conceptual site model. Other topics of interest to the ISM Team included the theoretical underpinnings of ISM, the planning and sampling design process for implementing ISM, and potential regulatory challenges to use of ISM (particularly the requirements for calculating upper confidence limits specified in some regulatory jurisdictions).

The processes and equipment described here are the best available at the time this document was written. As technology advances and new equipment, instrumentation, and processes are developed, they may be included in future ISM implementations provided they meet the data and measurement quality objectives for the site to be characterized.

Overall, members of the ISM Team have found that ISM provides reliable, reproducible sampling results and leads to better, 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 and 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.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

NOTE: "Hyperlinks"

This guidance was developed as Web-based document. The blocks of information presented online as "Hyperlinks" are included in Appendix G.

INCREMENTAL SAMPLING METHODOLOGY

1. INTRODUCTION

Incremental sampling methodology (ISM) is a structured composite sampling and processing protocol having specific elements designed to reduce data variability and increase sample representativeness for a specified volume of soil under investigation. Variability in measured contaminant concentrations between discrete soil samples is due primarily to the particulate nature of soil and heterogeneity in the distribution of contaminants. The elements of ISM that control data variability are incorporated into (a) the field collection of soil samples and (b) laboratory processing and subsampling procedures. ISM is designed to obtain a single aliquot for analysis that has all constituents in the same proportion as an explicitly defined volume of soil. Properly executed, the methodology provides reasonably unbiased, reproducible estimates of the mean concentration of analytes in the specified volume of soil.

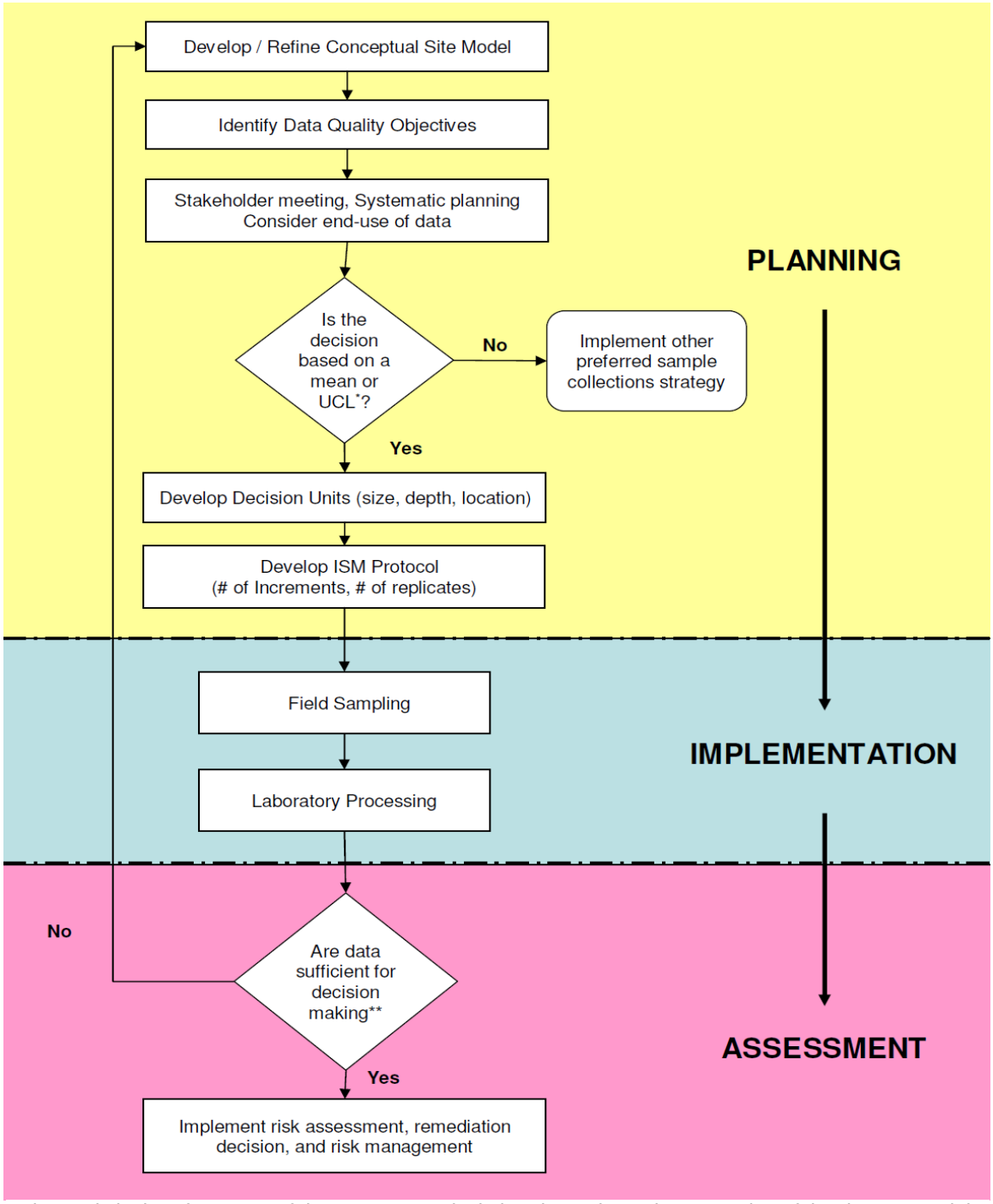
The elements of ISM that control data variability are incorporated into (a) the field collection of soil samples and (b) laboratory processing and subsampling procedures.

In 2009, the ITRC convened the ISM Team to prepare this guidance document, which focuses on soil sampling. ISM addresses all the sources of sampling error in a systematic fashion (Gy 1998, Smith 2006). Other approaches to soil sampling have not emphasized reducing sampling error as much as ISM. Because this methodology requires change from traditional approaches, the ISM Team found it necessary to go into detail about the theory as well as the application of incremental sampling. The team found this to be a valuable exercise and it should be valuable to the reader as well. Because a good deal of new terminology is introduced by ISM, the reader is directed to the glossary in Appendix E. Section 1.8 also illustrates some key terms used in this document.

The ISM Team recommends that, as with any well-conceived sampling approach or plan, all members of the investigation project team (e.g., consultants, regulators, geologist, analytical chemists, risk assessors and toxicologists) be involved in the entire ISM development process.

1.1 Summary of ISM as an Environmental Sampling Approach

Like all sampling approaches, ISM should be applied within a systematic planning framework. Figure 1-1 shows a general ISM flow process. One of the first steps in such a framework is to have the investigation project team establish a working conceptual site model (CSM). Once the CSM has been agreed to, the project team defines the data quality objectives (DQOs) and determines the appropriate decision unit (DU) size(s) and location(s). DUs are based on project-specific needs and site-specific DQOs; both considerations specify and constrain the appropriate end use of the data. The size of a DU is site-specific and represents the smallest volume of soil about which a decision is to be made (USEPA 1999, Ramsey and Hewitt 2005, HDOH 2008a, ADEC 2009). In some cases a DU comprise smaller units known as sampling units (SUs), as discussed in Section 3. The requirement to explicitly and appropriately define the DU that each incremental sample represents is a key component of ISM and is discussed in detail in Section 3.



* The statistical performance of the 95% UCL calculation depends on the properties of the data set and the sampling design. Note that ProUCL or FLUCL does not currently include the statistical algorithms for handling ISM data (see Section 4.0 and Appendix A).
** See Section 7.

Figure 1-1. ISM flowchart.

ISM planning includes the development of an ISM protocol for the number of increments and replicates to be collected for each ISM sample. An incremental sample is created by collecting

many (usually 30–100) equal-volume increments in an unbiased manner from throughout the entire DU. The combined increments (frequently totaling a kilogram or more) are typically processed at the laboratory and subsampled to provide an analytical aliquot of only a few grams that is used for analysis. The final analytical aliquot is the target sample.

ISM addresses major sources of sampling error and increasing sample representativeness.

ISM is designed to provide an unbiased, statistically valid estimate of the mean value of an analyte within the DU. Through adequate spatial coverage of the DU as well as disciplined handling, processing, and subsampling of the single sample formed from the increments collected, ISM works to overcome major sources of error in both sampling and subsampling of soils that have often been apparent with current sampling practices. By design, ISM provides complete spatial coverage within the DU; however, ISM does not provide information on the spatial distribution of contaminants within the DU. Should this spatial variability be important to the decisions being made, a smaller DU should be used. ISM may not be appropriate in certain situations (see Section 8 for further information on the limitations on ISM).

1.2 Traditional Investigation Approach Limitations

Soil sampling is typically done to characterize a site. Historically, the majority of soil samples collected has been discrete samples. Collection of discrete samples is sometimes preferred or mandated by regulatory agencies (see Section 8). Over the years, consultants, environmental scientists, and regulators have become aware of a number of recurring challenges, problems, and deficiencies associated with collecting soil samples as discrete, composite, or any other sampling method, including the following:

- **Lack of clear environmental objectives at the initiation of the investigation**—Often the primary objective is to "find contamination" with little clarity as to how the data will be used to determine whether identified contamination poses unacceptable risks to human health and the environment and often leading to lengthy delays in completion of project and expenditure of funds available for site investigation before adequate characterization is completed.

- **Poor spatial coverage of areas targeted for investigation and inadequate sample density**—Generally, a minimum of 20–30 discrete samples is needed for an adequate characterization of a targeted area and volume of soil; however, only a small number (e.g., <10) of discrete samples are commonly collected to characterize large areas of suspected contamination. The degree of coverage is typically controlled by the amount of funds available for laboratory sample analysis, thus limiting the number of samples needed to provide a representative and statistically valid characterization of a targeted area.

- **Laboratory aliquots prepared for analysis not necessarily representative of the field sample**—Aliquots prepared by random selection of a single, small mass of soil from the field sample container are not representative of the larger volume of soil delivered to the laboratory.

Traditional soil sampling and analysis methods impart a level of uncertainty in the use of data generated to identify potential environmental hazards associated with contaminated soil and to support decisions for or against remediation. In large measure, ISM is evolving to address these limitations.

1.3 How ISM Addresses Traditional Investigation Approach Limitations

The fundamental question with all soil sampling, discrete or incremental, is representativeness. In reviewing sampling results, environmental professionals often find themselves asking, "What does the sample concentration we get back from the lab represent?" With incremental sampling that question is purposefully rephrased as, "What does the (incremental) sample have to represent?" and that question is used to shape the project planning and establishment of DQOs well before any sample is actually collected.

The major problem with discrete soil sampling is the extreme magnitude between the mass of the subsample analyzed by the laboratory and the mass of the target population (area to be investigated or sample volume collected), which can be on the order of 1 in 10 million or more. This increases the chance that the sample misses contamination, which will consequently not be represented in the analytical results at all. ISM builds a sample from increments to provide a good representation of the DU and so is more likely to capture even heterogeneous contamination.

ISM requires that the project team address the spatial dimensions associated with the analyte concentration that is of interest. That is, the project team must define the DU to be represented by each incremental sample. This requirement is inherent in any soil sampling effort but must be addressed head-on and with great deliberation in ISM.

ISM forces the project team to confront the inherent heterogeneity in soil by defining the scale at which heterogeneity will be addressed. ISM does this early in the project life cycle by getting stakeholder agreement on the dimensions of the DUs from which samples will be collected. The scale issue is present for all sampling approaches but has typically not been made an integral part of the sampling strategy as it has with ISM. Furthermore, once the scale of the DU has been decided, the concept of hot-spot delineation within the DU should be moot. If it is not, then the DU may not be appropriately sized and should be reevaluated (see Section 3.5 for further discussion on hot spots).

Early in the project life cycle, ISM forces the project team to confront the inherent heterogeneity in soil by defining the scale at which heterogeneity will be addressed.

ISM addresses common errors associated with sampling soils. As such, ISM embeds the concept of quality assurance (QA)/quality control (QC) in a meaningful way into planning, design, field sampling and sample processing, as well as laboratory work, by explicitly addressing all of the activities necessary to build an ISM sample that will be representative of the DU of interest. Traditional QA/QC approaches have focused primarily on laboratory procedures, particularly those that take place after a subsample of soil has been extracted, and do not address the major sources of error that occur well before an extract solution is introduced into an analytical instrument.

1.4 How ISM Compares to Compositing

Environmental professionals recognize the act of combining increments as being similar to conventional compositing. ISM is an improved type of compositing in comparison to conventional compositing in that great attention is given to establishing the DU. ISM also requires that the total sample mass be sufficient to represent the heterogeneity of soil particles within the DU in proportion to all of the DU soil (i.e., population) and that a sufficient number of equal-volume increments are collected in an unbiased manner from throughout the entire DU so that all particles in the unit have an equal probability of being included in the sample. Thus, the incremental sample has the goal to contain all constituents in exactly the same proportion as they are present in the DU (i.e., the sample is representative of the DU). Proper laboratory processing and subsampling procedures then produce an aliquot for analysis that contains all constituents of the subsample in the same proportion in which they occur in the sample and, therefore, the DU.

ITRC's ISM Team has found that many state regulatory agencies have been reluctant to use composite sampling and that such reluctance spills over to ISM (see survey results in Section 8). One concern expressed with composite sampling is that clean or less-contaminated soil will be mixed in with contaminated soil, therefore diluting areas of high contamination. This problem can be minimized with a clear understanding of sampling objectives that incorporate the concept of a DU.

As shown in Figure 1-2, a hypothetical DU may be sampled with several designs. This figure illustrates a discrete gridded sampling design, various composite designs, and an ISM design. While each has its advantages and limitations, a goal of incremental design is to provide a high degree of spatial coverage of the DU. It is also obvious that the incremental design is similar in appearance to the various composite designs recommended by the U.S. Environmental Protection Agency (USEPA). Compositing is discussed further in Section 2.6.2.

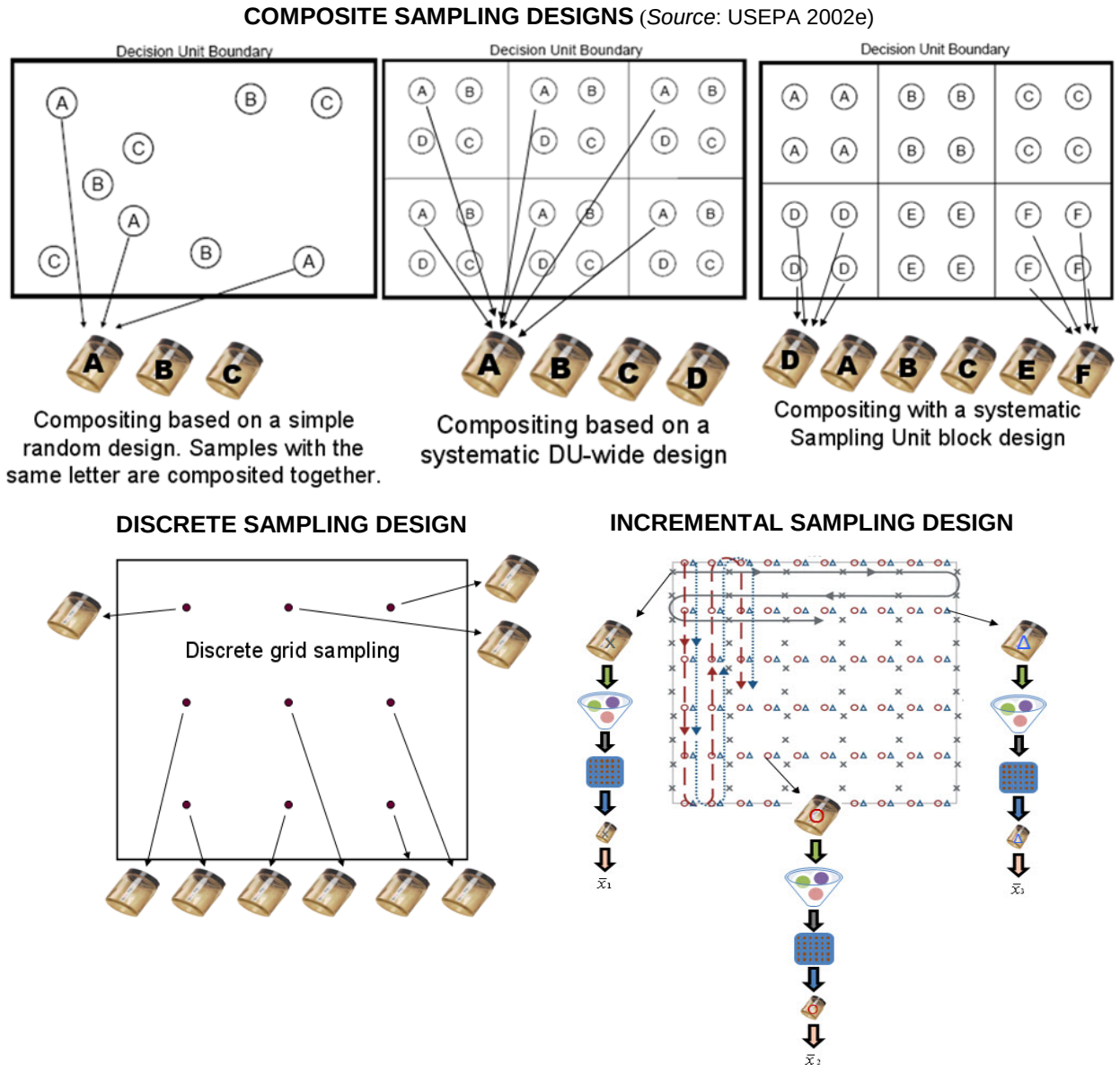


Figure 1-2. Sampling designs.

1.5 Purpose

The purpose of this technical guidance document is to advance the appropriate use of ISM for sampling soils at waste sites and potentially contaminated land or properties. In doing so, this document addresses those challenges that constrain or prohibit use of ISM. Some of these challenges may be directly associated with ISM, but as just described, others may be associated with questions poorly addressed in traditional soil sampling approaches using discrete samples. The challenge for developing this document on ISM is quite broad. See Section 8 for a detailed discussion on regulatory challenges and survey issues regarding ISM and how they can be successfully addressed.

While the focus of this document is on sampling shallow soils, other interests and areas, including sampling of deeper soils, are also discussed. In addition, some of the limitations associated with traditional soil sampling practices are not so much attributable to the reliance on discrete samples as they are due to the lack of clear and quantifiable sampling objectives to achieve project goals. Meeting sampling goals is discussed in Sections 2 and 3 of the document as part of the planning process, as well as part of the sampling design.

ISM usage is increasing in the environmental field. Currently, two states, Alaska and Hawaii, use ISM based on guidance documents that each state has recently developed. In addition, USEPA SW-846 Method 8330B applies incremental sampling procedures for explosive residue field sample collection and laboratory analysis. Thus, it is timely for this document to be issued. Again, it is the intent of the ISM Team that this document advance the appropriate use of ISM, as well as to expand the list of chemical contaminants that can be addressed confidently by ISM.

Currently, two states, Alaska and Hawaii, use ISM based on guidance documents that each state has recently developed. In addition, USEPA SW-846 Method 8330B applies incremental sampling procedures for explosive residue field sample collection and laboratory analysis.

1.6 Frequently Asked Questions

Table 1-1 conveys many challenging points taken on by the ISM Team in preparing this document, including key points and frequently asked questions that are addressed within the referenced sections. Although ISM has advantages over traditional soil sampling practices, it may not be appropriate given the current state of technology for all sampling applications (e.g. low-level VOC analysis, possibly metal speciation, etc.) It is anticipated that technology advances will allow these limitations to be addressed.

Although ISM has advantages over traditional soil sampling practices, it may not be appropriate for all sampling applications.

Table 1-1. Crosswalk for frequently asked questions on ISM

Key Point/Question	Reference
How can a regulator (or anyone) better assess ISM?	• All sections • Contact the ISM Team with questions • Participate in ISM Internet-based training
Section 1. Introduction	
What is ISM and what are the advantages/ disadvantages of using it?	Sections 1.3, 2.6.3, 3.5, and 8.5
Is ISM compositing?	Sections 1.4, 2.6.2, and 2.6.3
Is ISM data more representative than discrete data?	Sections 1.3, 2.3.2, and 2.6
When should ISM not be employed?	Figure 1-2, Sections 3.1, 8.3, and 8.5
Section 2. Nature of Soil Sampling and Increment Sampling Principles	
Is ISM really based on Gy's Theory?	Section 2.5
Why do we care about the mean value in a DU?	Section 2.1

Key Point/Question	Reference
Section 3. Systematic Planning and Decision Unit Designation	
What is a DU, and how is it established?	Sections 3.2 and 3.3
At what types of sites can ISM be used?	Section 3.3
Can ISM be used at any point of an investigation?	Sections 3.1, 3.2, and 3.3
How many increment and replicate samples should be collected?	Sections 3.1, 2.5.6, 4, 5.3, 6.2.2.7, and Appendix A
How can ISM be used in risk assessments?	Sections 3.1 and 3.3
Can ISM be used for ecological investigations?	Sections 3.1, 3.2, 3.3, 4.4.4, 7.1, and 7.2.7
Can ISM be used to delineate "contamination"?	Section 3
Can I use ISM when needed to determine whether contamination is a leaching concern?	Sections 3.1, 3.2, 3.3, 8.3, and 8.5
How do you compare ISM background samples to background generated from discrete samples?	Sections 3.1, 3.2, 3.3, 4.4.3.3, 7.2.4, and 8.5.4.7
Does ISM "mask" areas of high concentration or hot spots?	Sections 3.5, 8.2, and 8.5
What soil sampling depth should be used with ISM?	Sections 3.1, 3.3, 5.3.1, and 5.3.2
Section 4. Statistical Sampling Designs for ISM	
How do you calculate a 95% upper confidence limit with ISM data?	Section 4.2
Can ISM data/results be compared to discrete data/results?	Sections 4.4.3.2 and 8.5.4.4
Section 5. Field Implementation, Sample Collection, and Processing	
How do you sample for volatile organic compounds (VOCs) with ISM?	Section 5.4.2
How do you ship VOC ISM samples?	Section 5.4.2
Section 6. Laboratory Sample Processing and Analysis	
What contaminants are most suitable for ISM?	Section 6.1
Do ISM samples require more laboratory sample preparation?	Section 6.2 and 8.5.5.2
What effects does sample processing (grinding, etc.) have on contaminant concentration?	Section 6.2
How are DQOs addressed in the laboratory?	Section 6.1.1
How do you address low-level reporting requirements of VOCs with ISM?	Sections 6.3.2
Section 7. Making Decisions Using ISM Data	
How do you use ISM data?	Section 7
Section 8. Regulatory Concerns with ISM	
What are the regulatory challenges and what are possible solutions?	Section 8
Are there cost savings when using ISM instead of discrete only sampling?	Section 8.5.3

Key Point/Question	Reference
Is subsurface ISM sampling cost-effective?	Section 8.5.3
Can ISM data/results be compared to regulatory criteria (e.g., "not to exceed")?	Sections 8.5.4.5

1.7 Document Organization

This document is organized into 11 sections which reflect the ISM Team's best effort at presenting a wealth of information concerning ISM in a logical and cohesive manner. Beyond the mechanics of collecting ISM samples, much attention has been given to the planning process, particularly in Sections 2–6.

NOTE: "Hyperlinks"

This guidance was developed as Web-based document. The blocks of information presented online as "Hyperlinks" are contained in Appendix G.

- Section 2 presents the nature of soil sampling and fundamental ISM sampling principles.
- Section 3 focuses on systematic planning and how to determine a DU.
- Section 4 covers the statistical basis of ISM sampling design, the results of statistical simulations and the effects of changing the number of increments, replicates and the effects of sample patterns.
- Section 5 provides information on sampling tools, field sampling collection, and field handling procedures.
- Section 6 presents the current practices and options available for laboratory processing and subsequent analysis.
- Section 7 covers what to do with ISM data.
- Section 8 summarizes the regulatory concerns with ISM and the ISM survey results.
- Section 9 provides selected case studies as examples.
- Section 10 includes input from stakeholders.
- Section 11 provides the list of additional materials referenced throughout this document.
- Appendix A presents additional details regarding the simulation studies used to evaluate the performance of alternative ISM sampling strategies.
- Appendix B presents August 2009 Survey Results.
- Appendix C presents Case Studies.
- Appendix D includes ISM Team Contacts.
- Appendix E includes the Glossary.
- Appendix F provides a list of Acronyms.

1.8 Key Terms

This document includes new terminology introduced by ISM, and Figure 1-3 provides some key terms. See Appendix E for a glossary of additional terms.

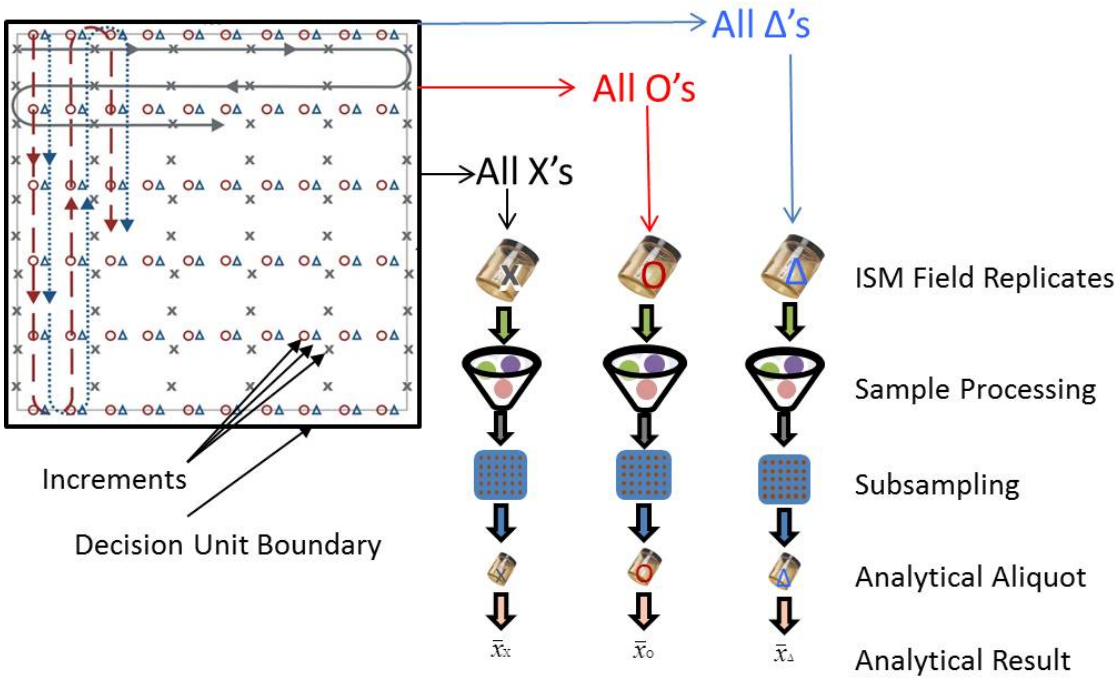


Figure 1-3. Key ISM process terms.

2. NATURE OF SOIL SAMPLING AND INCREMENTAL SAMPLING PRINCIPLES

2.1 Introduction

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

Because it is impractical to collect and analyze the entire volume of soil for which decisions must be made, samples are collected and the results used to represent that entire volume of soil. The industry of environmental investigation, regulation, and laboratory analysis has, to a large extent, developed around the practice of using discrete samples to meet all decision goals, including

estimating mean contaminant concentrations. There are many reasons why a mean concentration may be of interest for decision-making purposes, as discussed in Section 3 and Hyperlink 1.

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

Relying on an estimate of the mean contaminant concentration in a volume of soil using a small number of discrete samples can lead to costly decision errors.

By its very nature, soil is a highly heterogeneous solid with many components. Sampling soil for the purpose of obtaining an estimate of the mean contaminant concentration is highly susceptible to sampling errors from a variety of sources. One goal of a sampling design should be to minimize the errors that can occur in each step of the sampling and analytical process. Historically, the focus has been on controlling errors associated with the analytical part of the process. A great deal of effort is invested in ensuring good data by requiring strict adherence to analytical methodologies and laboratory QA/QC procedures. But all this attention addresses only the tail end of the process. There are many more steps to the data quality chain that require attention for the output to be good data. According to USEPA's soil screening guidance (USEPA 1996b):

> Data users often look at a concentration obtained from a laboratory as being "the concentration" in the soil, without realizing that the number generated by the laboratory is the end point of an entire process, extending from design of the sampling, through collecting, handling, processing, analysis, quality evaluation, and reporting.

Steps usually overlooked when evaluating data quality include sampling design, sample collection techniques, sample processing, and field and laboratory subsampling. However, there is a growing body of evidence that the predominant source of error in the "entire process" to which USEPA refers is sampling error, which occurs because contaminant concentrations in soil are highly heterogeneous. Heterogeneity makes representative sampling difficult. Sampling errors are manifested as variability (i.e., imprecision observed as large differences in results between replicate samples) and/or bias in the data set (i.e., data results significantly over or under the true concentrations). Data variability is easily measured to evaluate the effects of sampling error on data quality. If concentrations are close to a decision threshold and sampling errors are not controlled, data variability can lead to highly uncertain estimates of mean concentrations, which in turn lead to considerable uncertainty about whether the mean is above or below a decision threshold. Hyperlink 2 provides an example illustrating the importance of considering data variability in decision making. Poorly thought out sampling procedures produce misleading

data that can cause decision errors, as illustrated in Figure 2.1, no matter how good the analytical step is.

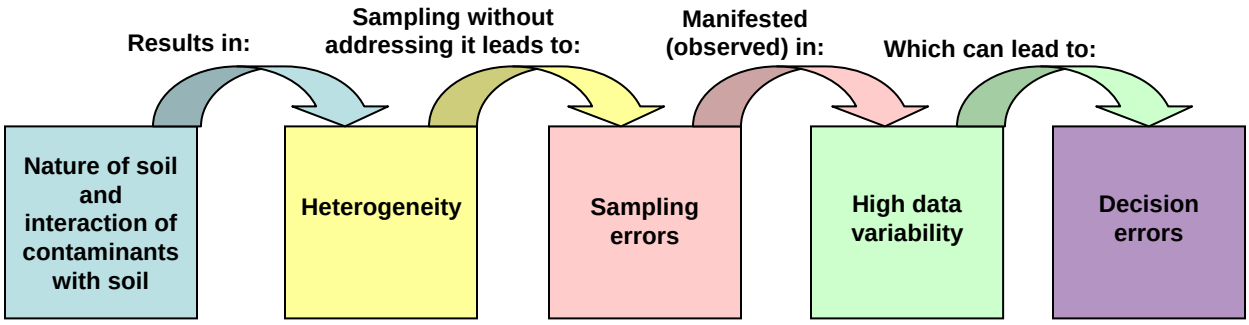


Figure 2-1. Heterogeneous nature of contaminants in soils may lead to decision errors.

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

By controlling sampling error throughout the **entire** sampling and analysis process, ISM can provide a precise and unbiased estimate of the mean using a relatively small number of laboratory analyses.

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

2.2 Soil Heterogeneity and Variation in Contaminant Concentrations

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

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

In contrast to this microscale heterogeneity, which occurs within a single sample, large-scale heterogeneity refers to differences in concentration from location to location across an area, in other words, differences in how contaminants are spatially distributed throughout the DU. For example, contaminants may be released from leaking drums, creating distinct but rather small contaminated areas. Or contaminants may be released by single or multiple large-volume spills, which might create large patterns that are mostly uniform in concentration within the spill area but demarcated by a fairly sharp boundary. Some contaminants, such as pesticides, might have been sprayed only along the edges or in garden pockets of a residential yard. Or pesticide leftovers might be poured out in a single spot. Atmospheric deposition is a common release mechanism with the resulting spatial pattern affected by wind strength and direction and by distance from the source.

Short-scale heterogeneity refers to concentration differences observed at the scale of colocated samples. Colocated samples are taken from the "same" location in the field generally a few inches to a few feet apart and are traditionally considered to be equivalent, meaning that their concentrations are expected to be approximately equal. However, field experience shows that colocated samples often differ in concentration, sometimes quite drastically.

Soil heterogeneity is present at different spatial scales. Each must be considered when collecting representative soil samples.

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

2.2.1 Microscale Heterogeneity

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

Figure 2-2. Grain sizes ranging from 0.016 mm to 2 mm.

2.2.1.1 The smallest particles often have the highest contaminant concentrations

Clay-sized particles are of particular note because of their tiny size and mineral makeup. These particles play a large role in how contaminants interact with soil. Due to their small size, clays have a large surface area per unit mass to which all types of molecules, organic and inorganic, can adhere. The chemical makeup of clay minerals gives them a strong negative charge as well as a weaker positive charge, enabling adsorption of both positive and negative ions. Clay minerals take the form of thin sheets or plates, as depicted in Figure 2-3. This plate structure greatly adds to the surface area of these tiny particles. The layered plates of clay particles also provide spaces for contaminants to absorb into clay particles, as illustrated in Figure 2-4, where the partial negative charges carried by the oxygen atoms on a dibenzo-p-dioxin molecule are attracted to a cation (such as Ca^{+2}) nestled between two clay plates.

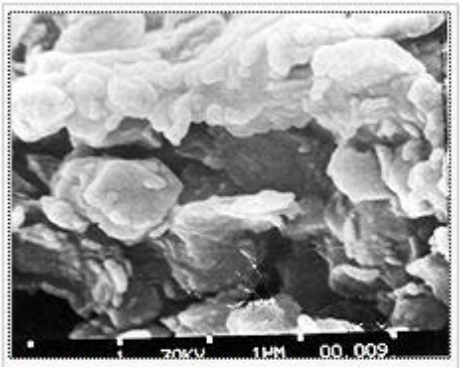


Figure 2-3. Electron microscope photograph of the structure of smectite clay particles.
Source: USGS 2006.

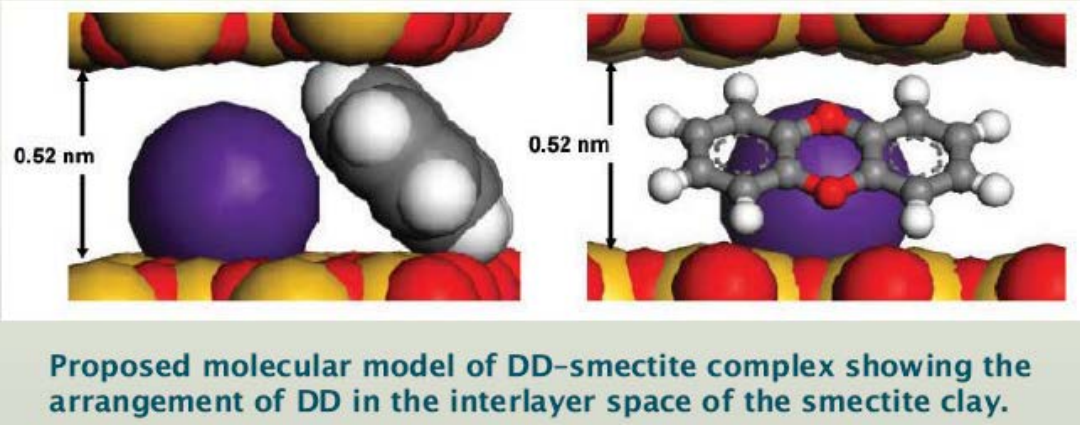


Figure 2-4. Illustration of smectite clay plates and interstitial cation binding with dibenzo-p-dioxin. *Source*: Superfund Research Program 2010**.**

The propensity for smaller particles to attract contaminants was dramatically shown in a study of lead-contaminated soil, as presented in Table 2-1. The pattern is clear: the smaller the particle size (i.e., the larger the mesh size number), the larger the concentration of lead associated with that particle size. Although the smallest particle size, that less than 200-mesh (0.074 mm), made up only one-third of the whole sample mass, it carried nearly three-fourths of the lead mass in the sample. Many experimental studies have documented the finding that for most contaminants, soil fractions composed of smaller particle sizes have higher loadings than fractions composed of larger particles. Important exceptions include metal fragments at firing ranges, explosive/propellant fragments, and ore particles at some mining sites (Walsh et al. 2007; Pavlowsky, Owens, and Martin 2009). Compounds bound to soil particles seldom migrate independently of the particle; they migrate with the particles. The very small particles carrying most of the contaminant load in soil are able to migrate with the wind or be carried by water flow in streams and storm water runoff.

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

Soil grain size (standard sieve mesh size)	Particle size (mm)	Soil fractionization (%)	Lead concentration in fraction by AA (mg/kg)	Lead distribution (% of total lead)
Greater than 3/8 (0.375) inch	>9.53	18.85	10	0.20
Between 3/8 inch and 4-mesh	9.53–4.76	4.53	50	0.24
Between 4- and 10-mesh	4.76–2.00	3.65	108	0.43
Between 10- and 50-mesh	2.00–0.297	11.25	165	2.00
Between 50- and 200-mesh	0.297–0.074	27.8	836	25.06
Less than 200-mesh	<0.074	33.92	1970	72.07
Totals		100	927 (wt. averaged)	100

2.2.1.2 Why laboratory duplicates often fail to match

The ratio between particles of different sizes and densities in a soil sample has a strong influence on the resulting contaminant concentration.

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

2.2.2 Short-Scale Spatial Heterogeneity

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

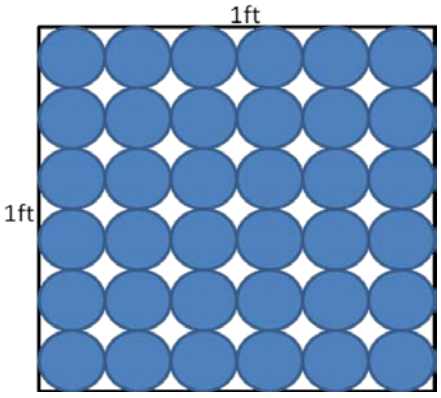


Figure 2-5. A single square foot area of surface soil contains 36 possible 2-inch-diameter core sample locations.

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

High variability in colocated samples is illustrated in Figure 2-6, which presents data from a field investigation for uranium. The original sampling design called for one discrete sample per 270 ft^2. In other words, the result from a single sample would be extrapolated to represent the concentration for an area centered on the sample and encompassing 270 ft^2. Prior to the main investigation, a small pilot study was done to see how much short-scale heterogeneity was present at the 1-foot scale. Figure 2-6 displays the results from the pilot study. It is apparent that the concentration assigned to this 270 ft^2 area could vary by an order of magnitude depending on

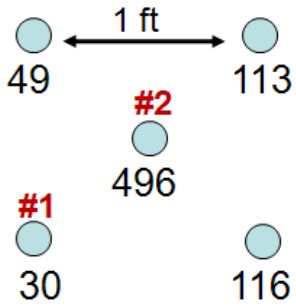


Figure 2-6. Observed short-scale heterogeneity with colocated uranium sample results. *Source*: Unpublished data supplied by Robert Johnson, Argonne National Laboratories.

where the technician happened to place the corer. If the sampler collected from sample placement #1, the concentration would be 30 mg/kg; if from placement #2 about 8 inches away, the result would be 496 mg/kg. In other words, if a single discrete sample were used to represent the sampling area, a conclusion of "contamination is low" vs. "contamination is high" is purely a matter of chance.

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

If unaccounted for, the presence of short-scale heterogeneity can lead one to draw very different conclusions about contaminant concentrations simply depending on where a discrete sample happens to be collected.

2.2.3 Large-Scale Spatial Heterogeneity

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

2.3 Foundational Concepts of Sampling

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

2.3.1 All Concentrations Are Means

At the most basic level, an analytical result represents the overall mean of all the thousands of individual particles in the 1, 10, or 30 g analytical subsample. As explained earlier, different particles carry different amounts of contaminants. By means of the analytical digestion or extraction process, there is a physical averaging of the various concentrations of contaminant particles within

an analytical subsample into a well-mixed liquid extract, as depicted in Figure 2-7. The laboratory result provides an estimate of the mean concentration of those particles making up the analytical subsample. Note that, as shown in Hyperlink 8, the laboratory measures contaminant mass and then derives a concentration. Laboratory results are then extrapolated to represent larger and larger volumes of soil culminating with the volume of the DU.

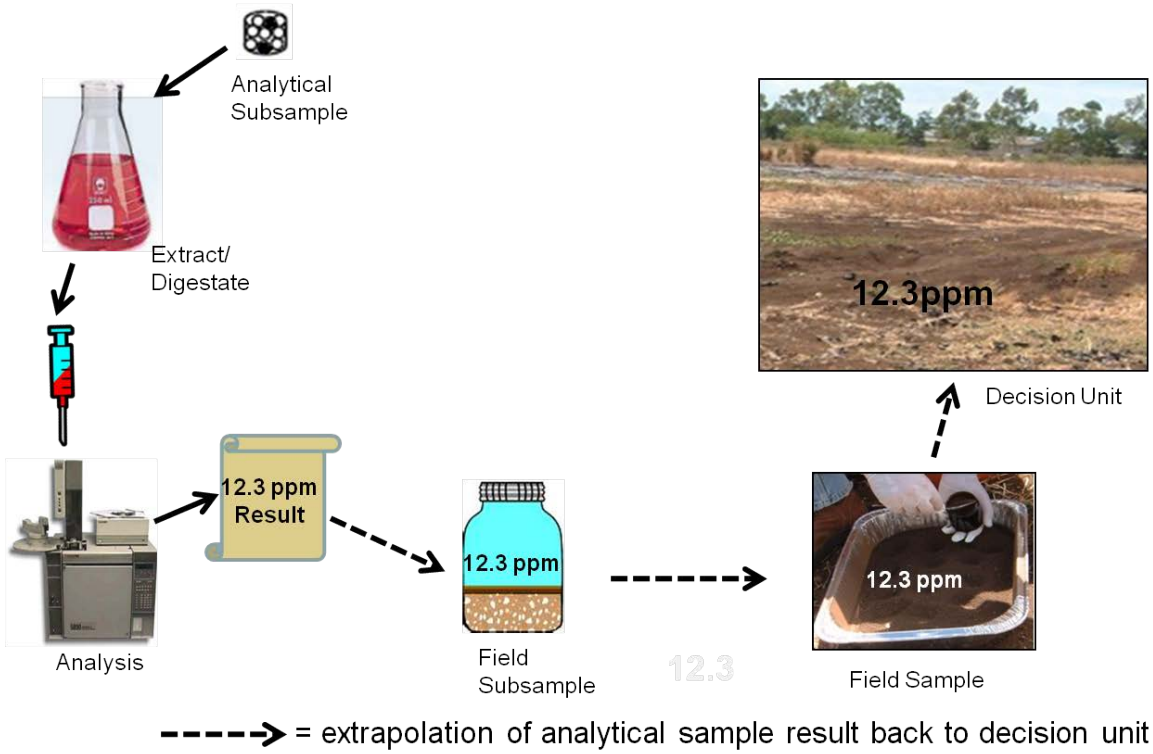


Figure 2-7. Process of extrapolating analytical sample results to soil concentrations.

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

> Scaling up an analytical result obtained from a small soil sample to some larger meaningful volume of soil at a site involves a series of assumptions.

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

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

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

> ISM is intended to provide an unbiased estimate of the mean contaminant concentration within a carefully defined volume of soil.

2.3.2 Representative Soil Samples

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

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

If a sample or set of samples intended to represent the population does not properly do so, a "sampling error" is said to have occurred. This is why systematic planning must be done *before* developing the sampling design. Otherwise, it is impossible to know what a sample is supposed to represent and how to collect it so that it is "representative." Unfortunately, it is common for sampling designs to be developed without a clear picture of how the data will be used. Inadequate sampling designs commonly indicate that "representative samples" will be collected, but often there is no

> A simple test for a good quality sampling plan is whether there are explicit statements declaring that "Samples will be representative of X (enter the population characteristic of interest to the decision)."

indication what the samples are supposed to be representing. On the other hand, a statement such as the following provides an unambiguous statement about the population of interest: "Samples will provide estimates of the true mean concentration of arsenic within the <2 mm soil fraction of the upper 6 inches of soil for each residential lot."

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

2.4 Scale-Specific Sampling Considerations

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

2.4.1 Sampling Considerations—Microscale Heterogeneity

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

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

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

Microscale heterogeneity exerts its effects as soon as soil is placed into a container. The settling of soil that occurs during container movement and sample shipment is governed by particle size and density. Settling stratifies a soil sample such that the larger particles usually end up at the top of the jar as smaller particles work their way to the bottom. If the subsampling procedure involves simply opening the jar and scooping from the top, very few small particles will end up

in the analytical subsample, which may bias the concentrations low. On the other hand, the type of scoop used to take the subsample may discriminate against larger particle sizes if the surface is flat or very small so that larger particles can roll off. The very process of weighing out the analytical subsample can select for small particles if they are preferentially tapped onto the balance to slowly bring the subsample up to the desired weight. Laboratories seldom have standard operating procedures (SOPs) for obtaining a representative analytical subsample. Each laboratory, and each technician in the same laboratory, is likely to handle samples somewhat differently. As a result, the analytical subsample may not be representative of the bulk average in the sample container but may over- or underrepresent certain particle sizes from one subsample to the next (Gerlach et al. 2002).

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

Field duplicates, colocated samples, and laboratory subsampling duplicates data primarily provide information about heterogeneity at different spatial scales and not solely about analytical issues at the lab.

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

2.4.1.2 Measuring the error caused by within-sample heterogeneity

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

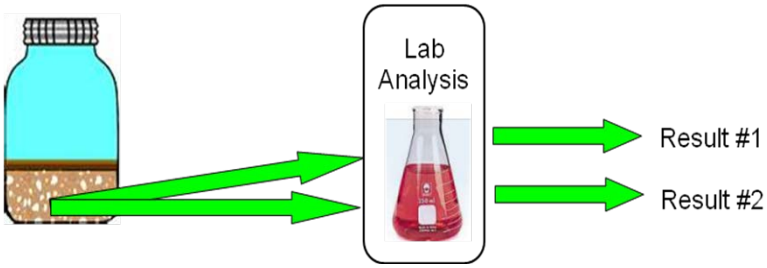


Figure 2-8. Variability among results of laboratory subsample duplicates measures within-sample heterogeneity.

A large difference between results indicates that within-sample heterogeneity is present and is causing sampling error. Field splits and laboratory duplicates for soils are common QC checks that often fail to meet QC acceptance criteria. Unfortunately, nothing is typically done to correct the problem(s) indicated by the failed QC. The data may be qualified as estimated, but in practice they are simply used "as is." Laboratory

duplicate results should not be ignored, for they provide very important information about the quality of sample handling and the magnitude of sampling error.

Variability between duplicate subsample results measures heterogeneity within the sample jar.

Duplicate results may vary so widely that a different decision about "clean" or "dirty" may be indicated, depending on which result is used. The question is often asked, "Which result is right?" The answer is that they are probably both right and both wrong. Both are right in the sense that the analysis of both subsamples was probably correct unless other QC samples indicate otherwise. It is just that the laboratory subsamples are fundamentally different. Both may be wrong in the sense that neither result adequately represents the true concentration for the jar of soil, and by extrapolation, for the concentration in the DU. Highly variable field and/or laboratory duplicates should be an indication to decision makers that the data generation process is excessively imprecise and could lead to decision errors. Hyperlink 13 provides a discussion of approaches for dealing with within-sample heterogeneity.

2.4.1.3 The effect of subsample mass on data variability

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

The data variability caused by heterogeneity affects the statistical distribution of the data, as seen in the three curves in the diagram. Data from smaller subsample masses form more lognormal-like statistical distributions. For example, notice how the right side of the 1 g sample mass curve (blue curve) is pulled out or "skewed" to the right much more than the left side. Because it is easy for small subsamples to miss contaminated particles, many small subsamples have low concentrations. However, sometimes more contaminated particles wind up in a small sample, causing a high concentration data result that is nonrepresentative of the parent material and producing the right-skewed "tail" of a lognormal distribution.

The smaller the analytical sample mass, the more likely that some data results will exceed an action level. Whether or not a volume of soil is considered compliant with an action level can depend on how big the analytical subsample is.

Note that some skewing is still present in the 10 g subsamples (green curve). For the 100 g subsamples the skewing is basically gone and the distribution is normalized (red curve). Figure 2-9 also illuminates how sampling error and the data variability it causes can lead to decision errors. For the sake of illustration, assume that 3 nCi/g is an action level. Because of its skewed distribution, some individual data results from the 1 g small mass data set will sometimes exceed the action level, even though the mean and most of the data results are below the action

level. For the 10 g subsamples, there will be fewer results above the action level. For 100 g subsamples of the same soil, there would be no results above the action level. Recall that these effects are apparent even after the parent sample was milled and sieved to 2 mm. This level of sample preparation goes beyond that which is typical for most environmental analyses.

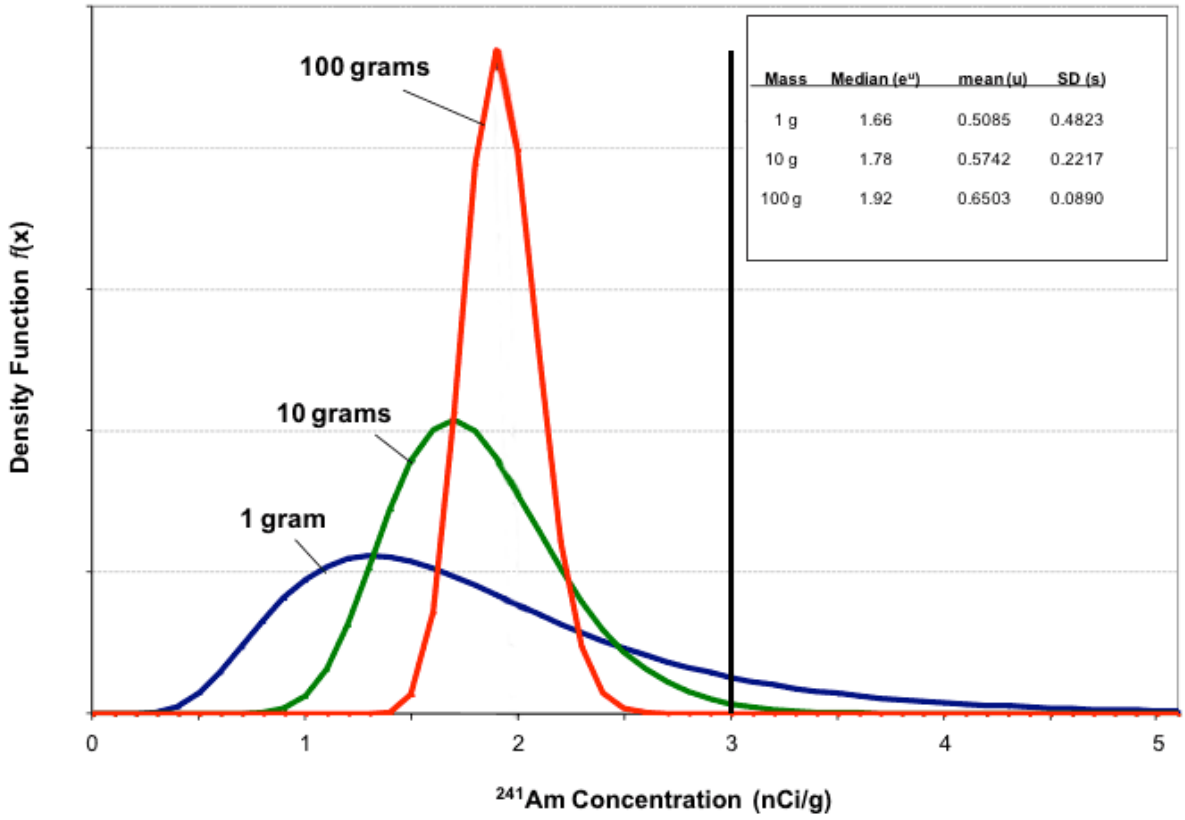


Figure 2-9. Smaller analytical masses contribute to high data variability. *Source*: Data from an experimental study on radioactively contaminated soil (Gilbert and Doctor 1985).

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

2.5 Gy Theory and the Source of Sampling Error

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

2.5.1 Pierre Gy's Sampling Theory and the Seven Basic Sampling Errors

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

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

Gy's Seven Sampling Errors

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

2.5.2 Compositional Heterogeneity

Before sampling errors can be discussed, the Gy theory concepts of constitutional and distributional heterogeneity must be introduced. Constitutional (or compositional) heterogeneity is a measure of the differences in composition between individual fragments or particles of the population being sampled with respect to a given parameter of interest. It refers to the fact that soil is made of many different types of particles that interact with contaminants in different ways. CH is a direct cause of a sampling error termed "fundamental error" (FE). A way to control FE is to have large enough samples (or subsamples) so that the probability is high that the composition of the sample will match the composition of the population. Figure 2-10 presents a population with two samples of different masses. Although both samples were collected from the same population, they are not equally representative of the parent population. The larger of the two samples (Sample B) better represents the composition of the population and reduces FE relative to the smaller sample (Sample A). Also, the larger the particles, the larger the sample mass must be to minimize FE. More illustrations of this concept can be found in Hyperlink 12.

Fundamental error is controlled by collecting samples of sufficient mass.

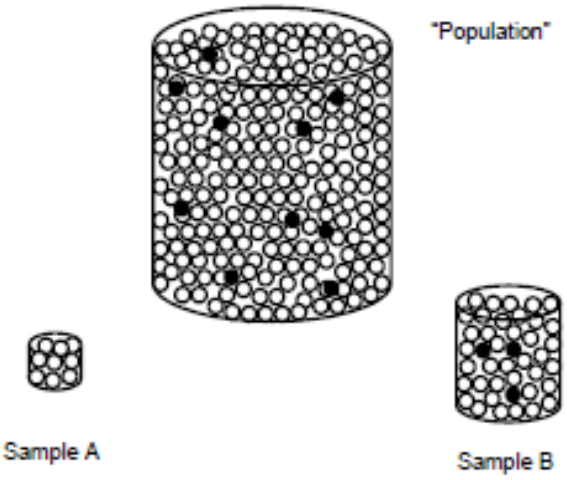


Figure 2-10. Illustration of the effects of sample mass on representativeness of the population. *Source*: USEPA 2002e.

2.5.3 Distributional Heterogeneity

Gy's distributional heterogeneity (DH) is a measure of particle distributions that can take the form of grouping or segregation. Particles may segregate, that is, separate into layers. Segregation

is often a result of gravity. The most common example is jiggling a jar of dry soil, causing finer particles to migrate toward the bottom, while the larger particles end up at the top of the soil mass. Figure 2-11 illustrates the two types of DH that Gy described.

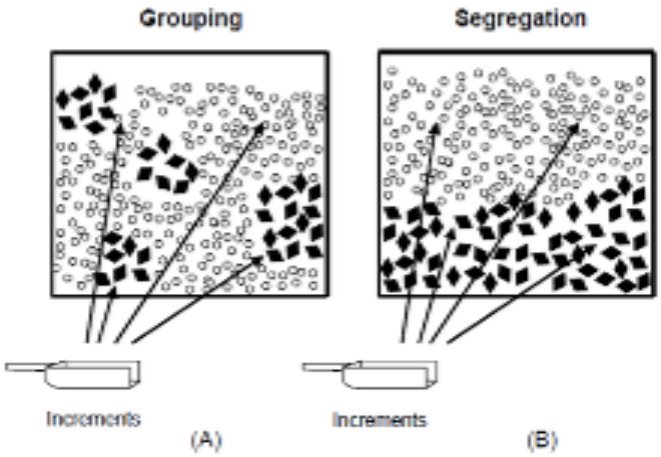


Figure 2-11. Depiction of grouping (A) and segregation (B) of particles. Source: USEPA 2002e.

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

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

GSE is controlled by collecting a sufficient number of increments.

2.5.4 Long-Range and Periodic Heterogeneities

In the Gy paradigm, "long-range heterogeneity" refers to the same contamination pattern as the term "large-scale heterogeneity," as discussed in Section 2.4.3. This heterogeneity involves the nonrandom, nonperiodic distribution of contaminant across the site. Identifying this heterogeneity is often an objective of sampling programs, such as mapping site-specific concentration trends. The question is: what is the volume over which knowledge of this heterogeneity is desired vs. what is the volume over which such heterogeneity is a distraction because the mean is the parameter of interest? In Gy's theory, this heterogeneity is considered the cause of long-range heterogeneity fluctuation error (CE_2). This Gy-defined error may or may not be a relevant error for a sampling design, depending on whether knowledge of contaminant distribution or mean is desired and the spatial dimensions of both have been defined. Gy theory assumes that the parameter of interest is the mean, not contaminant distribution.

In the same way, periodic heterogeneity and its corresponding periodic heterogeneity fluctuation error (CE_3) is the result of cyclical changes in space or time over a site. An example of a cyclical change in time is measuring nitrogen concentration in agricultural fields over several growing seasons. If sampling were always performed at the start of the growing season when nitrogen levels were highest, a misleadingly high value would be obtained if the average over the entire year were desired. Just as for the long-range error above, it is only a true error if it causes an inaccurate estimate of a mean for some defined area, and in this case, for a defined period of time.

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

2.5.5 Device and Preparation Errors

Delimitation error (DE) is a result of using an incorrect shape for the sampling device that removes each increment from the population or the incorrect use of a correct sampling device. For example, an incorrectly shaped sampling tool biases the grain sizes included in that sample. A sampling tool should be of a shape and size so that every fragment of the population of interest has an equal probability of being included in the sample. This error is a common source of bias in environmental samples, both in the field and in the laboratory. Figure 2-12 illustrates that, depending on the sample device, some particles have a greater chance of being included in the sample/subsample than others. The sampling interval depicted in Figure 2-12 has a higher proportion of larger particles at the bottom of the interval. This might be the case, for example, in an in situ soil scenario. On the other hand, this particle distribution pattern might be reversed, for example, in the case of soil jars in the laboratory. Subsampling with the rounded scoop preferentially gathers particles from the top, which tend to be the larger particles when stratification occurs as the sample is arranged in a "slabcake" shape in preparation for subsampling. With its narrower bottom width, a rounded scoop discriminates against the particles at the bottom of the sampling interval, which tends to be the smaller sizes in many if not all subsampling scenarios. By design, the rectangular scoop tool is a more inclusive tool and gathers particles of various grain sizes consistently throughout the sampling interval. In Gy theory, a sampling tool that promotes DE is termed incorrect"; one that reduces DE is called "correct."

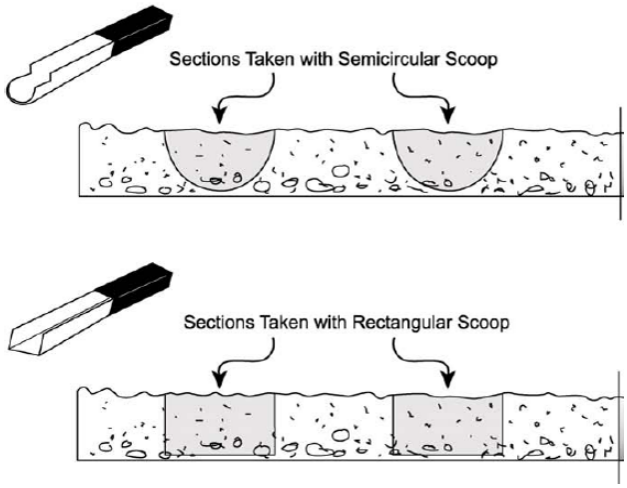


Figure 2-12. Illustration of the effects of sampling device design on particle sizes in a sample. *Source*: Gerlach and Nocerino 2003.

Extraction error (EE) also results from the use of incorrect sampling devices. Unlike DE, which is only a function of the shape of the sampling device, EE is a function of the sizes of both the tool and the soil particles and the correct use of the sampling device. This error occurs because an inappropriate sampling device

Increment DD and EE are controlled by the proper use of correct sampling tools.

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

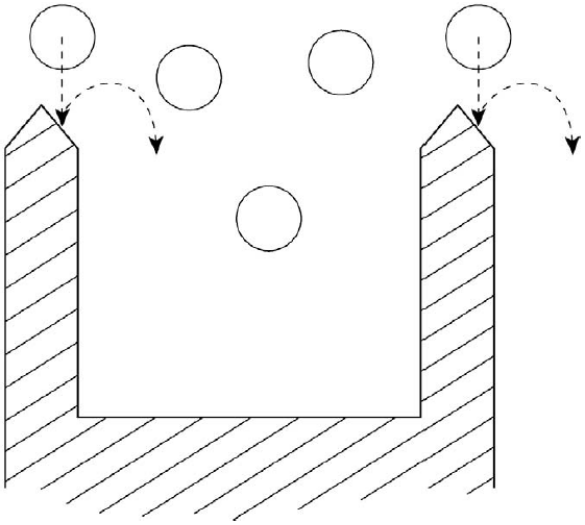


Figure 2-13. Vertical view of a sampling device that minimizes EE.
Source: Gerlach and Nocerino 2003.

Figure 2-14 illustrates both DD and EE. A volume of soil is depicted in a two-dimensional (2-D) plane with larger particles concentrated at the bottom. Coring Device A minimizes the DD and EE because it samples the full thickness of the material and does not discriminate against the larger particle sizes. Coring Device B demonstrates EE and is an incorrect device for this matrix. Its mouth is too small to include larger particles. Coring Device C (the shovel) demonstrates DE because the sample profile it delimits cannot sample the full thickness of the DU. Coring Device D also illustrates DE because the delimited sample profile does not encompass the full shape of the DU in the vertical plane.

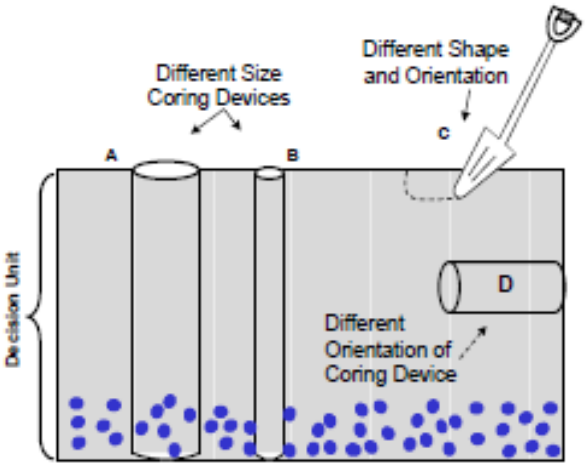


Figure 2-14. Increment DE and EE from sampling device selection.
Source: USEPA 2002e.

Preparation error (PE) is the sum of errors introduced by analyte loss, cross-contamination, or chemical or physical alteration of the sample that biases sample results relative to the true mean. Some of these errors are controlled by traditional QA/QC procedures such as sample preservation, holding times, and blanks.

2.5.6 Controlling Gy Errors

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

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

Factor leading to error	Sampling error	Error results from	How to control
Compositional heterogeneity (CH)	Fundamental error (FE)	Size and compositional distribution of the particles	Increase the sample mass and/or reduce the size of the particles
Distributional heterogeneity (DH)	Grouping and segregation error (GSE)	Heterogeneous distribution of particles within the population	Increase the mass of the sample or increase the number of increments
Large-scale heterogeneity	Long-range heterogeneity fluctuation error (CE_2)	Changes in concentration across space or over time	Reduce the spatial interval between samples
Periodic heterogeneity	Periodic heterogeneity fluctuation error (CE_3)	Periodic changes in concentration over time	Change the spatial and/or temporal interval between samples
Identifying the correct increment geometry	Increment delimitation error (DE)	Incorrect shape (in all three dimensions) of the sample or increment selected for extraction from the population	Use correct sampling plan design and correct sampling equipment that can sample the entire thickness of the population
Shape of the sample extraction device and nature of the soil	Increment extraction error (EE)	Incorrect extraction of the sample or increment because the sampling device is too small	Use correct sampling equipment that does not push larger particles aside, and use correct sampling protocols
Loss or gain of contaminants during sample handling	Preparation error (PE)	Contamination loss or gain due to alteration, evaporation, degradation, cross-contamination, mistake, or fraud	Use appropriate sample handling, preservation, transport, and preparation measures

In practice, the focus is usually on FE and GSE; however, the other errors can be important if correct sampling procedures are not used. As illustrated above, the FE can be minimized by collecting sufficient mass of sample, and the GSE can be minimized by collecting numerous increments.

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

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

When no prior data are available to estimate heterogeneity within a DU, a default range of 30–50 increments is recommended.

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

2.6 Three Sampling Approaches

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

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

This equation emphasizes several important concepts:

- There is a true mean concentration in any volume of soil.
- Any type of sampling and analysis is capable of providing only estimates of the true mean concentration.
- The best estimate is the one with the least total error.

These concepts provide a basis from which to compare different methods for estimating the mean through different sampling approaches: discrete, composite, and ISM sampling. Any sampling design must include consideration of these questions:

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

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

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

2.6.1 Discrete Sampling

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

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

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

> When the true mean is well above or below an action level, even a small number of discrete samples usually results in a correct decision.

2.6.1.1 Heterogeneity and discrete samples

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

result in the same concentration estimate if analyzed. As discussed in Section 2.4.1.1, this is often a poor assumption.

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

2.6.1.2 Discrete sampling plans

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

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

This is not to say that a low-density discrete sampling approach is insufficient for all cases. If, for example, the true mean in a DU is orders of magnitude above or below the action level for a contaminant of interest, it is possible that a correct decision could be made from very few (or even one) discrete samples. The key factors are the degree of heterogeneity present at the various scales, the action level, and the magnitude of the true mean. Since, as is often the case, knowledge about heterogeneity or the magnitude of the true mean is seldom available (which is why sampling is being conducted), relying on data from low-density discrete sampling plans is more likely to result in decision errors.

Relatively High Density. The second type of discrete sampling plan can be called relatively high density discrete sampling plans. In this context the number of discrete samples approaches the number of increments typically collected with ISM (i.e., 30–50). The number of samples may or may not have been statistically derived based on (among other things) an estimate of the heterogeneity of the soil or the anticipated magnitude of the true mean concentration. There is a large body of guidance and reference material that describes how various discrete sampling plans of this sort can be effectively used to investigate soil contamination and make appropriate environmental decisions. However, cost limitations frequently limit the number of discrete samples employed for environmental investigation, and as discussed below, even relatively high-density discrete sampling plans may produce certain characteristics in the data set which are not ideal. The decision quality of relatively high-density discrete sampling plans, especially those derived through statistical methodology, can compare favorably with ISM sampling plans. However, the analytical costs associated with such plans will likely be considerably greater than those of a comparable ISM approach.

2.6.1.3 Interpreting results of discrete sampling

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

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

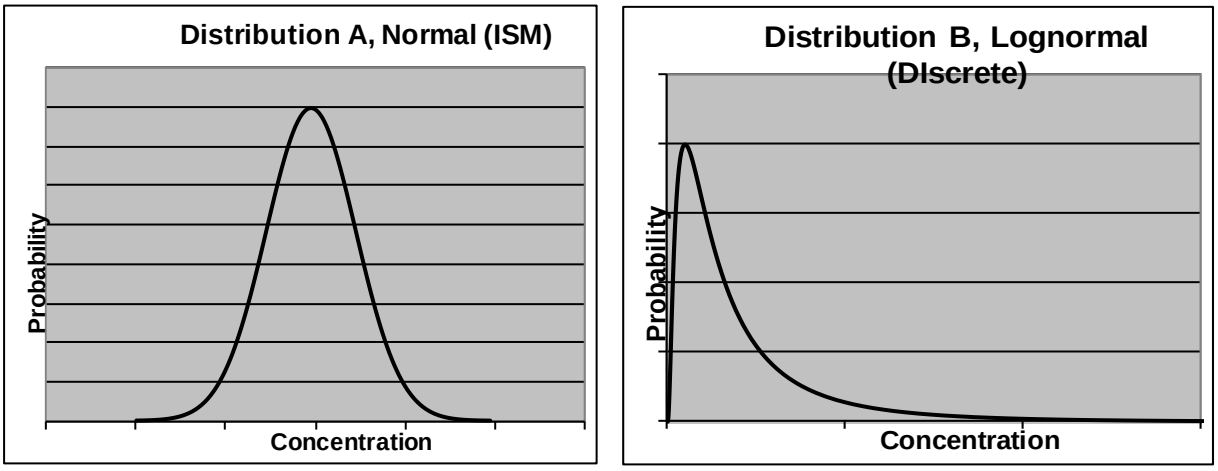


Figure 2-15. Examples of distributions generated by plotting concentration data vs. frequency (i.e., probability) of observation.

Discrete sample data tend to be clustered around the most frequently observed concentration, which is called the "mode." Because Distribution B in Figure 2-15 is skewed to the right, the mode is less than the mean concentration of the distribution. The tail of such distributions can easily contain concentrations one to two orders of magnitude greater than the value at the mode. In contrast, ISM samples can be expected to fit a distribution closer in shape to Distribution A in Figure 2-15, with less tailing and a mode closer to the mean. This fact can have important implications for making decisions based on discrete sampling data.

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

2.6.1.4 When discrete sampling may be successful

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

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

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

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

2.6.2 Composite Sampling

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

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

2.6.2.1 Beneficial uses of compositing

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

With discrete sampling designs there is often the implicit assumption that each discrete sample represents some volume of soil surrounding the area where it was collected. As shown in Figure 2-6, this may be a faulty assumption when short-scale heterogeneity is significant. Recall that short-scale heterogeneity is what occurs at the scale of colocated samples, where kneeling down in one location can give a radically different concentration than kneeling down in a location one foot away. Composite sampling can be used to reduce errors due to short-scale heterogeneity. For example, in Figure 2-6, instead of collecting a single discrete sample from only one of the five placements, an increment of soil could be taken from each of the five placements and composited into a single sample. This method results in a sample more representative of the 1 ft^2 area shown in the figure as compared with any single discrete sample. This process could be repeated in other 1 ft^2 areas, resulting in a number of composite samples. Note that, as with any sampling design, subsequent steps still need to be taken with each sample to address microscale (within sample) heterogeneity to reduce this source of sampling error as discussed in Section 2.4.1. This process is different from an ISM sampling design since it may

not include the prior establishment of a specific DU and the goals may not necessarily be limited to estimating the mean concentration. Also note that the analytical cost of such a composite sample design will typically be much larger than with ISM since many samples will be submitted for separate lab analysis.

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

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

2.6.2.2 Poorly designed composite sampling

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

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

As routinely applied, composite sampling is viewed primarily as a way to reduce analytical costs, without taking more important sampling goals into account. It is not surprising that over the years composite sampling has developed an unfavorable reputation. It is important to understand that ISM differs greatly from the practices common to poorly designed composite sampling applications. It is worth noting that routine applications of composite sampling also differ significantly from the composite sampling designs recommended in USEPA guidance. Yet, ISM transcends even most USEPA compositing guidance because ISM prominently calls out specific error-controlling steps. These were not yet well researched when most USEPA statistical sampling design documents were written.

However, the primary reason that ISM and composite sampling as typically practiced cannot be considered equivalent is that typical composite sampling rarely involves enough aliquots of soil to manage contaminant heterogeneity over an entire DU. Therefore, even when the goal of compositing is to determine an average over some area, it is less likely to estimate the mean concentration with the precision needed by data users. Empirical studies and sampling theory suggest that composite sampling (with inadequate consideration of the steps listed above) simply does not perform as well as ISM sampling. Indications are that low-increment number composite samples combined with insufficient mixing and processing procedures perform about as well as discrete samples. However, it is acceptable to use the composite sampling approach if it meets the user-defined goals for precision and accuracy. Composite sampling approaches should include a methodology for estimating the total precision and take measures to ensure that an unbiased mean is obtained.

2.6.3 Incremental Sampling Methodology

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

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

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

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

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

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

3. SYSTEMATIC PLANNING AND DECISION UNIT DESIGNATION

ISM provides an estimate of the mean contaminant concentration in a defined volume (area and depth) of soil. ISM is particularly useful when practical constraints (e.g., budgets) limit the number of discrete samples that can be collected and therefore limit the precision with which the mean concentration in heterogeneous matrices may be estimated. As discussed in Hyperlink 1, most action levels are derived from risk-based receptor models which assume a specific exposure area (i.e., exposure unit). Therefore, estimates of mean concentrations in volumes of media are generally the appropriate statistic to compare to action levels.

However, it is important to match the project objectives with the type of sampling employed. For some objectives discrete sampling is appropriate (when sufficient numbers of discrete samples are used); for others, ISM sampling is the best option. In certain situations, sampling designs consisting of combinations of discrete and ISM samples may be advantageous. For example, discrete samples might be used to make decisions on obviously contaminated volumes of soil in which contaminant concentrations are very likely to exceed action levels. Even though contaminant concentrations in this situation may be highly variable, this variation would not result in decision errors since any possible sample collected from the volume will likely have contaminant concentrations above the action level. Discrete samples may also be used to estimate the variability within a DU prior to ISM sampling. When field analytical methods (or other cost-saving analytical approaches) are available, sufficient numbers of discrete samples may be used to characterize some contaminants or DUs, while ISM may be appropriate for those contaminants for which these analytical approaches are not available.

3.1 Overview of Systematic Planning

Environmental data must be of the appropriate type, quantity, and quality to manage uncertainty and reach defensible decisions. 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 a systematic planning process, including the development and consideration of a conceptual site model. The planning process should also take into account the decision mechanisms for which the data will be used (Section 7 offers an extensive discussion of decision mechanisms). Establishing clear objectives at the beginning of the investigation is crucial to efficient and effective characterization of the site. As described in this section, good systematic planning is reflected in the field through well-thought-out DUs and SUs that are sampled to answer key investigation questions.

Establishing clear objectives at the beginning of the investigation is the key to obtaining data that will support well-informed management decisions.

As described throughout this document, the use of incremental 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. In combination with well-considered investigation objectives and DU and SU designations, incremental 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 systematic planning process.

The systematic planning process identifies the objectives of the site investigation and establishes the type of information needed to make an environmental decision. The level of detail needed to adequately incorporate a systematic planning approach into a data collection effort varies from project to project; larger or more complex projects typically warrant more detailed planning than smaller, simpler projects. The nature of the ISM process is such that many decisions must be made and detailed plans established and disseminated in advance of sample collection. For these reasons, the principles of the systematic planning approach should be applied on every ISM project.

A clear understanding of the study objectives is important with all sampling strategies, but particularly so with ISM sampling. Different types of objectives may dictate the type, location, and dimensions of DUs. For example, the identification and investigation of small source-area DUs may be especially important for highly mobile chemicals that can pose significant vapor intrusion or leaching risks. In other situations, larger exposure-area DUs or subsurface DUs are appropriate to evaluate risks to specified receptors. ISM should not be used in situations where the resulting data will not meet the project objectives or answer the questions indicated by the systematic planning process or where it cannot be implemented within the constraints of the project. These caveats, however, apply to any potential sampling and analysis strategy.

With ISM it is critical to understand the objective of the investigation since it may determine the type, location, and dimensions of DUs.

Systematic planning involves a series of well-considered steps that result in clear data collection plans and objectives. The USEPA DQO process and the U.S. Army Corps of Engineers (USACE) technical project planning (TPP) process are two examples of systematic planning

frameworks that can readily be used with ISM. Guidance documents which describe the DQO process as well as other systematic planning processes are listed below:

- *A Guideline for Dynamic Workplans and Field Analytics: The Keys to Cost-Effective Site Characterization and Cleanup* (Robbat 1997)
- *Technical Project Planning (TPP) Process* (USACE 1998)
- *Data Quality Objectives Process for Hazardous Waste Site Investigations: Final Guidance* (USEPA 2000a)
- *Using the Triad Approach to Improve the Cost-Effectiveness of Hazardous Waste Site Cleanups* (USEPA 2001)
- *Guidance on Systematic Planning Using the Data Quality Objectives Process EPA QA/G-4* (USEPA 2006b)
- *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 2010a)
- *Technical and Regulatory Guidance for the Triad Approach: A New Paradigm for Environmental Project Management* (ITRC 2003)
- *Triad Implementation Guide* (ITRC 2007b)

3.1.1 Project Planning Team

Systematic planning should be conducted by a project team composed of individuals who are knowledgeable about the site background and investigation goals. Required information likely includes the geographical layout, sampling constraints, laboratory analytical methods, statistical goals, and data interpretation. The ideal size of the planning team varies with the complexity and importance of the problem. Depending on the particular project, the project planning team may include expertise in chemistry, data analysis, engineering, field sampling, geology, quality assurance, modeling, regulatory requirements, risk assessment, soil science, statistics, and toxicology. The ITRC ISM Team recommends all members of the investigation project team be involved in the ISM development process.

All members of the investigation project team (e.g., consultants, regulators, risk assessors, geologists, analytical chemists, and toxicologists) should be involved in the entire ISM development process.

3.1.2 Conceptual Site Models

CSMs are essential elements of the systematic planning process. A CSM serves to conceptualize the relationship between contaminant sources 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 the data collection efforts. The CSM should be maintained and updated as new information is collected throughout the life cycle of the project. Various styles of CSM are useful, from text explanations to a series of figures depicting current and assumed future site conditions in three

DU sizes and locations are key outcomes of the conceptual site model (CSM).

dimensions. Some form of visualization aid (e.g., figures, graphs, charts, tables) that relates site conditions to receptors in a manner that lends itself to the explanation and use of ISM is suggested (Figure 3-1 provides an example). The sampling strategy should reflect the assumptions about the transport phenomena and exposure scenarios reflected in the CSM.

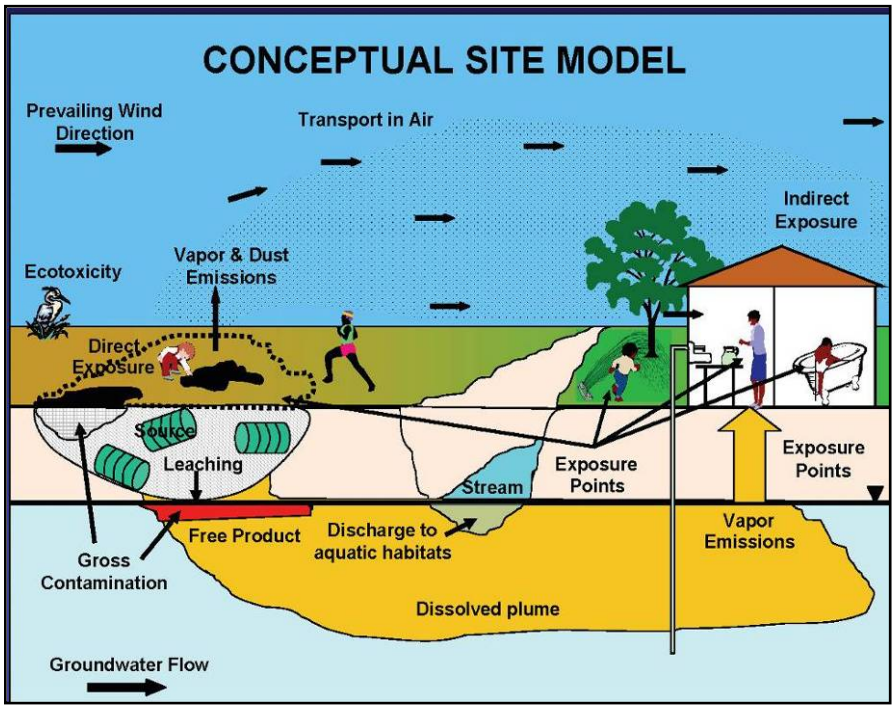


Figure 3-1. Pictorial CSM.

Information on the development of CSMs is readily available in a number of guidance documents, including the following:

- *Guidance for Conducting Remedial Investigations and Feasibility Studies Under CERCLA* (USEPA 1988)
- *Data Quality Objectives Process for Hazardous Waste Site Investigations: Final Guidance*. (USEPA 2000a)
- *Conceptual Site Models for Ordnance and Explosives (OE) and Hazardous, Toxic, and Radioactive Waste (HTRW) Projects* (USACE 2003)
- *Standard Guide for Developing Conceptual Site Models for Contaminated Sites* (ASTM 2008)

The reader is directed to these and other relevant guidance documents for development of CSMs.

3.1.3 Investigation Objectives

Any environmental investigation should include consideration and articulation of project objectives during the planning stages. Ideally, the CSM should lead to the development of these specific project objectives, which vary from project to project and often from phase to phase. The objectives of environmental investigations are not unique to ISM and may include the following:

- identification and characterization of source areas and releases
- delineation of the extent of contamination
- compliance with agency-specific regulatory requirements
- establishment of background conditions and comparison with site conditions
- estimation of exposure point concentrations for risk assessment
- collection of confirmation sampling during and after remediation
- selection, design, and optimization of remedial options

Investigation objectives may change as projects progress; therefore, new information and objectives must continually be considered. Dynamic or iterative sampling strategies may often prove useful with ISM. For example, additional or alternative DUs may need to be established to better understand the site or to assist in the design and selection of remedial options. Data collected for site characterization purposes may need to be supplemented with new information during the remedial phase of the project to identify volumes requiring remediation more precisely.

3.1.4 Data Needs and the Selection of Sampling Approaches

A number of crucial decisions must be made when ISM projects are planned. For example, it is necessary to determine the number and size of DUs, the number of replicate ISM samples, the number of increments making up each ISM sample, and the specific laboratory protocols required. Data needs are generally guided by the overall study objectives as outlined in Section 3.3. However, a fundamental challenge in planning sampling projects is that a sampling event often includes multiple objectives.

3.1.4.1 Spatial scale of DUs

Different study objectives pose very different kinds of assessment questions, and, most importantly for sampling considerations, different objectives require different spatial scales. Some objectives call for characterizing contaminant concentrations over a relatively large scale (up to acres), while others are contingent on distinguishing concentration differences at a much smaller scale (within tens of feet). The scale of a DU can depend upon whether the primary objective is to inform a risk assessment (i.e., DU is an exposure area) or to select a remedy (i.e., DU is a remediation unit).

DUs applicable to human health may not be readily applicable to ecological receptors. Therefore, when sites are evaluated for both human health and ecological receptors, multiple scales and project objectives are likely. For example, DUs for human health evaluations may correspond with individual residential properties, while DUs established to address ecological receptors may consist of much larger or much smaller volumes of soil, depending on the nature of the specified

ecological receptor. Thus, the designation of new DUs and additional data collection efforts after initial sampling are sometimes unavoidable.

The need to characterize concentration variations at different spatial scales poses significant planning challenges. While the ISM approach has gained some acceptance for estimating mean concentrations in large areas, there are lingering concerns about its application where small areas with high contaminant concentrations are hypothesized. These concerns notwithstanding, variations in contaminant concentrations over small spatial scales can be addressed using ISM or a combination of ISM and discrete sampling. However, project planning always requires compromises and decisions, for example:

- Smaller DUs are more likely to indicate the possible presence of areas with higher contaminant levels. However, sampling smaller DUs requires more resources, and project planners must always strike a balance between cost and certainty.
- High-density discrete sampling may offer a reasonable alternative for assessing small-scale variations but may not be practical for large areas.
- While low-density discrete sampling plans may be more feasible, they suffer from the inadequacies discussed in Section 2.6.1.2.

3.1.4.2 Sampling approach

Decisions about the general sampling approach for a project are crucial in ensuring the data will meet the project objectives. Project planners may elect to employ ISM, traditional discrete sampling, or a combination of both. The optimum approach depends on the CSM, the nature and extent of contamination, project objectives, and regulatory requirements. For example, if applicable regulations or policies call for comparison of an action level to the maximum detected concentration, discrete sampling is typically necessary because ISM provides only an estimate of the mean and cannot be used to estimate the maximum. An alternative is to place multiple small SUs over the DU and demonstrate compliance using the maximum value obtained from the SUs. If estimation of the maximum is identified as a sampling objective and a discrete sampling approach is employed, project planners should consider using a high-density discrete sampling plan. A traditional low-density discrete or low-density SU sampling plan is unlikely to produce an accurate estimate of the maximum concentration in the area of concern.

The fact that discrete sample data can be combined and recombined in a variety of data sets after sample analyses is both a strength and a weakness; it is possible to use such data in biased or uninformed combinations. ISM data, on the other hand, are more specifically tied to the individual volumes of soil that were sampled. This too is both a strength and a weakness. Project-specific objectives and constraints as well as the decision mechanism to be used must be considered during the selection of DUs so that multiple issues are addressed.

Project planning decisions are seldom based on technical considerations alone. In addition to the uncertainties associated with different sampling approaches, project planners must consider various regulatory requirements, as well as resource and budget limitations. Nevertheless, to appropriately control decision errors, a clear and complete understanding of the technical

strengths and weaknesses of sampling approaches must be considered before an approach is selected. When weighing these strengths and weaknesses, the available information on soil heterogeneity and the project goals for decision errors should be fully considered. As is always the case in environmental management, sampling plans must balance the necessity of controlling costs with the need for a reasonable degree of certainty. The decision to use ISM, discrete sampling, or some combination to investigate areas that have elevated concentrations should be made with a clear understanding of the limitations of both techniques and clear definitions of what constitutes a potential risk on a scale that considers both concentration and volume relative to specific project objectives.

Table 3-1 lists many items to be considered during the planning, implementation and use of ISM data. Included in this table are example project objectives. The project team must identify the data necessary to address all study questions and meet the project objectives.

Table 3-1. Considerations to address during systematic planning for ISM sampling

Factors	Issues
Conceptual site model	Source(s)
	Contaminants
	Action levels
	Distribution
	Migration mechanisms (fate and transport including preferential pathways)
Parameter to estimate	Parameter(s) the project needs to estimate: • Mean • UCL • Other
Project objectives and decisions	Overall project goal: • Human health risks of exposure areas • Risk to ecological receptors • Effectiveness of cleanup/removal • Extent • Leaching potential • Other
	Source, nature, and numerical value of the action level
	Requirements for precision, total error, and decision quality
	End use of the data
Decision units (see Section 3.3)	Number of DUs
	Rationale for DU selection: • Human health exposure area • Ecological exposure area • Source area • Excavation sidewall or floor • Location of DU

Factors	Issues
	Scale of the decision (spatial and/or temporal scale that was originally intended in the development of the action level)
	Variability within DU (anticipated and/or measured)
	Location of DU
	Size of DU (three-dimensional [3-D]): • Surface dimensions • Subsurface dimensions • Depth interval • Geologic strata or soil horizon
Sampling units (see Section 3.3)	Subdivision into SUs: • Number if SUs composing each DU (from one to many) • Rationale for SUs • Location of SUs • Size (surface dimensions, depth, interval)
	Portion of DU represented by SU: • Targeted areas of suspected high contamination • Random placement • Extrapolation allowed
	Background SU
Increments	Number of increments in each ISM sample
	Increment spacing
	Targeted bulk sample volume/mass
	Approximate increment volume/mass
	Sampling pattern: • Simple random • Systematic random • Stratified random • Other
Replicates (see Section 7.3)	Number of replicates
	Type and purpose of replicates: • DU replicate • SU replicate • Field replicate • Laboratory replicate • Instrument replicate
Targeted soil fraction	Basis for targeted fraction of soil: Targeted fraction of soil (e.g., grain size, geologic unit, etc.) • All fractions • Multiple fractions • <2 mm fraction or a fraction >2mm

Factors	Issues
Field procedures (see Section 5)	Field steps to control sampling errors: • Correct sample device • All potential increments in the SU equally available to sample device • Consistent increment size • Coverage of SU • Cross section of SU collected • Decontamination between ISM samples (not between ISM increments) • Field mixing • Field subsampling
	Sampling procedures including aspects of "correct sampling" (see Section 2.5)
Lab procedures (see Section 6)	Mixing in the laboratory
	Particle size reduction or selection (where appropriate) (see Hyperlink 17): • Grinding or milling • Sieving • Sieve size
	Subsampling
Statistic calculated (see Sections 4 and 7)	Arithmetic mean (of replicates)
	Variance (of replicates)
	95% UCL: • Chebyshev • Student's-*t*
	Metric used to evaluate SUs
Decision mechanism (see Sections 4 and 7)	Source, nature, and numerical value of the action level
	How will the decision be made? Which decision mechanism (DM) will be used (i.e., how will you decide whether further action is necessary or not)? • DM 1: Comparison of one ISM sample from the DU to the action level • DM 2: Comparison of the mean of replicate data from the DU to the action level • DM 3: Comparison of the 95% UCL on the mean of replicate data from the DU to the action level • DM 4: Comparison to background • DM 5: Combining DUs • DM 6: Extrapolating from sampled to unsampled areas • DM 7: Evaluating oversized DUs • Other

3.2 Nature and Intent of Decision Units and Sampling Units

Environmental decisions are often based on the risks resulting from exposure to estimated mean concentrations of contaminants in volumes 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 published screening levels. In other investigations the estimate of the mean contaminant concentration provided by ISM samples might be used to estimate the risk to human or ecological receptors. ISM results may also be used to estimate background concentrations or to assess sources or to evaluate various stages of remedial activities. (See Section 7.2.4 for further discussion.)

ISM is a method for estimating the mean concentration of contaminants in specified volumes of soil called DUs.

- A **DU** is the smallest volume of soil for which a **decision will be made** based on ISM sampling.
- An **SU** is a volume of soil from which increments are collected to **determine an estimate of the mean concentration** for that volume.

While these two concepts are closely related, the subtle differences between them allow for great flexibility in how ISM data may be used to make decisions for volumes of soil. Although it is not necessary to subdivide DUs into component SUs, the option to do so allows additional types of decision mechanisms to be used with ISM data. (Decision mechanisms are discussed in Section 7 and are defined in the glossary as an algorithm or protocol that results in a decision for a volume of media).

3.2.1 Planning DUs and SUs

As discussed in Section 2, all contaminant concentrations in soil are heterogeneous on some scale. Therefore, the determination of the sampling scale and the related increment density is very important in all sampling situations. If a finer resolution of contaminant variability is needed to address the objectives of the investigation, then the scale of the DU is too large. On the other hand, excessively small DUs are impractical at some point. Determining the size, shape, location, depth and number of DUs is a critical component of the planning process. Likewise, the strategy behind the use of SUs as well as their size, shape, and other attributes must be carefully planned.

> The size, shape, location, depth, and number of DUs and SUs must be clearly identified during planning.

Because decision mechanisms and the designation of DUs and SUs are integrally related, the anticipated decision mechanism must also be considered along with the layout of the DUs and SUs during systematic planning. Each of these variables should be considered in relation to the CSM and should support and elucidate the project objectives. Basic questions about the intent of the investigation should be considered; for example, "How do the DUs and SUs fit into the overall objectives of the investigation?" and "How will the resulting data be used in the decision mechanism to address the project objectives?" are crucial. The answers to these questions should flow naturally from the understanding of the site, the CSM, and the project objectives. The designation of DUs and SUs should support and clarify the objectives of the investigation, and as

those objectives are refined further during latter stages in the life cycle of the site, the DUs, SUs, and decision mechanisms should be reconsidered.

3.2.2 Use of Sampling Units

SUs are subdivisions of DUs from which separate ISM samples are collected. The boundaries of an SU indicate the coverage of a single ISM sample; therefore, SUs define the scale of the ISM sampling, while 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 possible to later redefine SUs as DUs (if the resulting scale meets the project objectives) to use the ISM sample results in appropriate decision mechanisms.

Figure 3-2 illustrates the relationship between SUs and DUs; a DU consists of one or more SUs. In the simplest situation shown in Figure 3-2a, the DU consists of a single SU (in which case the term "SU" is not necessary—the DU and the SU are one in the same). In this case, results from one or more ISM samples (e.g., replicates) collected from within the DU are used to make a decision. Figure 3-2b shows the DU divided into four SUs, each of which is separately sampled with one or more ISM samples. When SUs are used, the SU sample results can be used in an appropriate decision mechanism to support a decision for the DU. Valid estimates of the mean or 95% UCL for the DU can be derived from the SU replicates, while they also provide some information on the spatial distribution of contaminants within the DU.

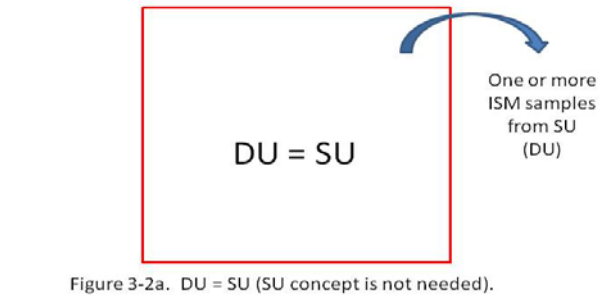


Figure 3-2a. DU = SU (SU concept is not needed).

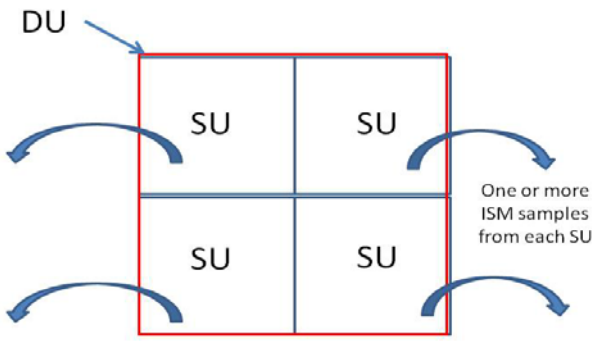


Figure 3-2b. DU is subdivided into 4 SUs

Figure 3-2. Decision units and sampling units.

The mean concentration over the entire DU should be the basis for decision making if the DU is properly sized; that is, the DU passes or fails based on the SU sample results over the entire DU in an appropriate decision mechanism. Results for individual SUs should not be used to make decisions on SUs because by definition SUs are smaller than appropriate for a decision or because they have been insufficiently sampled. This is especially true if sampling involves only one ISM sample per SU (e.g., no replicates were collected) as discussed further in Section 4. Estimation of the mean DU concentration from individual and replicate SU samples is discussed in more detail in Sections 4.3.4.3 and 7.2.3.

In some cases, the planning team may determine that information about spatial variability within the DU is needed in addition to an estimate of the mean concentration of the DU. Sampling at the SU scale can provide such information, which may be valuable if the DU fails the decision and

further investigation or remedial activities are indicated. In this case, additional systematic planning, designation of smaller DUs, and resampling may be necessary (see also Section 7.2.3). However, as discussed in Section 4.3.4.3, there are advantages as well as disadvantages to using SUs in this manner. Therefore, this approach should not be considered the default ISM strategy but merely an option that can be useful depending on the project objectives.

SUs may also be used when there are multiple sampling objectives or sampling scales for a given volume of soil, for instance, when areas must be assessed for multiple receptors with exposure areas of different scales. In this situation the mean of smaller volumes of soil may be estimated through one or more ISM samples on the SU scale, while the sample results of the SUs can also be combined to estimate the mean concentration at the larger DU scale.

SUs may also be advantageous when very large volumes of soil cannot be sampled at the desired scale or sample density due to practical limitations (e.g., costs) or multiple receptors (e.g., human and ecological receptors). In this situation, sampling of SUs and some form of extrapolation to infer the DU mean from the subset of SUs actually sampled may be a feasible, if imperfect, alternative. However, there are a number of assumptions and cautions associated with this approach, as discussed in Sections 4.4.2 and 7.2.6. Decision mechanisms involving extrapolation are not acceptable to regulators in some states.

Finally, SUs may also be useful to characterize smaller source areas encompassed by large exposure areas to determine whether or not they require separate investigation and additional action. More detailed discussions on the use of SU sample data are offered in Sections 4.4.1 and 7.2.5.

3.3 Decision Units

3.3.1 Defining Decision Units

There are various approaches to defining DUs. The approach selected should be consistent with the understanding of the site reflected in the CSM and should support the objectives of the investigation. DUs may be defined in regularly spaced and equal volumes as established by exposure areas, or they may be based on irregular features of the site which define contaminant transport or receptor exposure. Alternatively, DUs may be based on an understanding of the contaminant distributions, for example, in and around source areas. Volumes of soil known or suspected to be contaminated are generally good candidates for designation as DUs because the decision over these volumes is best made separately from less-contaminated surrounding volumes. Human health or ecological exposure areas may provide the basis for the designation of DUs. This approach has the advantage that it is conceptually supported by the exposure assumptions used to derive most action levels. DUs may also be based on the needs of remediation or excavation. For example, landfill construction or other remedial approaches may dictate the dimensions of the DU. Sidewalls and floors of excavations may be designated as DUs to determine whether soil removal was sufficient.

Selection of DUs should also consider the geologic aspects of the CSM. If the boundaries between different geologic formations are important for contaminant transport or exposure, they

may provide a logical demarcation of the DU. In some cases a DU may extend across more than one geologic formation or soil type, but in other situations basing DU boundaries on the geological boundaries may make more sense. Background studies may particularly require consideration of geological conditions.

3.3.2 Types of Decision Units

Two primary types of DUs are suggested: those based on the known or suspected locations and dimensions of source areas, called "source area DUs," and those based on the size assumptions of risk assessment, called "exposure area DUs."

A source area is defined here as a discernible volume of soil (or waste or other media) containing elevated or potentially elevated concentrations of contaminant in comparison to the surrounding soil. Source areas include the following:

- areas with stained soil, known contamination, 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

This definition highlights the difference between types of DUs. Source area DUs are differentiated from exposure area DUs in that the boundaries of source area DUs and the scale of sampling are based on the *known or hypothesized extent of the contamination*, while the boundaries of exposure area DUs are determined through the *exposure assumptions of the risk scenario*. There are a number of purposes for differentiating so exactingly between these two types of DUs. This approach is intended to simultaneously do the following:

- address the concern that source areas will be "diluted out" through the use of ISM sampling by emphasizing the importance of proper DU designation for source areas
- reiterate that action levels derived from risk assessment scenarios are based on exposure assumptions that include a specified areal extent of contamination within which a mean concentration is of interest

3.3.3 Vertical Definition of DUs

DUs consist of volumes of soil. Therefore the depth (how far below ground) and interval (vertical dimension of the DU) of each DU must be carefully considered during planning stages. These attributes should be based on the project objectives and the CSM and should not be left to haphazard decisions in the field. Identifying the correct depth for sampling is not a simple undertaking (true no matter what type of sampling is conducted). Additionally, it is important to remember that a correctly defined DU includes the requirement that all hypothetical increments within the DU have an equal likelihood of being sampled. Therefore, a DU should not be defined to be 5 feet deep when only the first few centimeters are available to the sampling device.

Like all DUs, exposure area DUs necessarily include vertical as well as lateral components. When exposure area DUs are used, the depth and interval of DUs should be defined consistently

with the exposure scenario under consideration. In many such scenarios the first few inches or centimeters at the surface is the appropriate sampling interval. However, the DU depth and interval considered acceptable for evaluation of direct exposures varies among agencies (ITRC 2007a) and risk scenarios. Evaluation of risks posed by future excavation and spreading of deeper contamination to the surface could require DU depths many feet below ground surface.

The vertical dimensions of the DU might also be determined using assumed, current, or future subsurface activities at the site. In this case, evaluation of soil leaching risks to groundwater or surface water might mean that the base of contaminated soil (if known) is designated as the vertical limit of the DU. State requirements might also serve to define the depth and interval of the DU.

When source area DUs are used and the CSM provides an understanding of where the contaminants are most likely to be located, DU depth should be based on this information. Many contaminant releases originate at the surface; therefore, once again, the first few inches or centimeters may be the appropriate sampling interval. If, however, the CSM indicates that contaminants are more likely at greater depths the DUs should be targeted accordingly. Cleanup decisions may also drive the depth of DUs; the depth interval of a remediation "lift" (e.g., a 6-inch scrapping by a bulldozer) is often a rational approach for defining DUs. Geological strata which may affect contaminant transport may also be a method to determine the vertical extent and depths of DUs. Finally, it should be noted that DU depths may also depend on the types of contaminants involved.

3.3.4 Exposure Area Decision Units

DUs based on exposure areas are a fundamental part of many environmental investigations and are a key tool in risk assessments and risk-based decision making. For the purposes of this document, an "exposure area" is defined as 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, areas of commercial/industrial properties, or areas designated as exposure areas through other means (e.g., state laws). Figure 3-3 depicts various types of exposure area DUs.

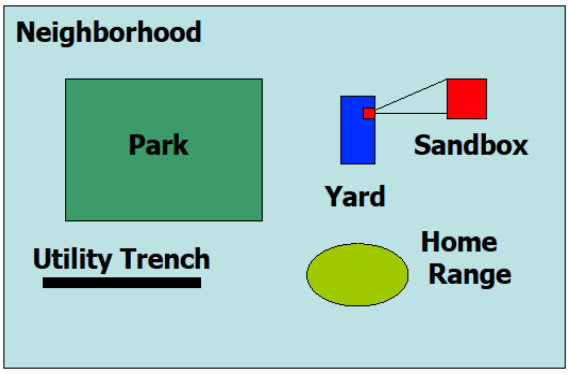


Figure 3-3. Examples of potential exposure area DUs.

The primary use of data from an exposure area is to estimate exposure and, subsequently, risk to human health and the environment. The data may also be used to screen sites for further study using criteria such as risk-based screening levels. (See USEPA's *Soil Screening Guidance* for guidance in development and application of screening levels to assess soil leaching potential [USEPA 1996a, 1996b]). This objective may be accomplished by comparison of the estimated mean concentration in the DU to action levels. If the project is more mature, data may be used to develop exposure point concentrations (EPCs) to quantify risks from exposures to contaminants

by human and/or ecological receptors.[1] When the decision will be based on risk assessment approaches, the DU should ideally be based upon the area where exposure is or potentially could occur. The size and placement of exposure areas depend on current use or proposed future use of the site.

When systematic planning considers data collection to support risk assessment, a primary question is "Over which area and depth do samples need to be taken to represent exposures of concern?" The exposure area could be based on current land use or the likely or possible future use of the area. In some cases, however, data and risk assessment may be required to inform the necessity for deed restrictions for a less likely land use. Site-specific information and the CSM should be used to designate exposure areas as much as possible. In cases where future land use is uncertain, location of future residences, for example, areas suspected of contamination should be sampled at a scale that is consistent with the presumed future land use as much as possible. The size and placement of exposure area DUs may need to be adjusted and the resultant uncertainties documented relative to potential future exposure and risk.

3.3.4.1 Residential exposure area decision units

Exposure areas for residential use can vary in size depending on the location of the site and local regulatory requirements. Consideration of lot-size exposure areas is generally adequate to evaluate long-term, chronic health risks. When *exposure* can be assumed to be relatively consistent across the lot, it is not necessary to investigate concentration trends at a property below the scale of the residential lot. If there are specific areas that receive higher use in an exposure area, such as a play area, that area should be evaluated as a smaller and separate DU (see Section 3.3.4.4).

3.3.4.2 Commercial/industrial exposure area decision units

Exposure areas for commercial or industrial properties are site specific and could be an acre or more in size. Certain maintenance or construction activities at these types of properties may influence the depth of the exposure area. Designation of exposure areas for these sites should be discussed ahead of time with the project risk assessor and should be based on areas of the site where exposure is likely to occur. When practicable, it may be beneficial to designate DUs for current commercial or industrial sites in a manner that assists in evaluation of the property for future unrestricted land use (i.e., residential land use). Consideration of future unrestricted land use during designation of initial DUs (and preparation of risk assessments) at commercial or industrial sites may help *avoid* unnecessary land use restrictions and/or the need to reinvestigate or assess the site should future redevelopment plans call for a more sensitive land use. The added effort and costs for additional characterization should be balanced with the value that the information may bring to the project.

[1] It should be noted that detailed exposure assessments may require evaluations of bioavailability or relative bioavailability of media contaminants, and exposure DU ISM samples may be used for this purpose. However, to be representative of exposure, bioavailability studies must be performed on ISM samples that have not been processed by grinding. If other DQOs being fulfilled by ISM samples in the exposure DU require grinding as part of sample processing, the comparability of ground vs. unground samples should be evaluated as part of the study.

3.3.4.3 Ecological exposure area decision units

Data collected to support ecological risk assessments are based on assumptions similar to those made for human health risk assessments. However, the differences between human health and ecological risk assessment mean that the DUs derived for each of these efforts may be quite different. The selection of the exposure area size during the refinement phases of a baseline ecological risk assessment should be representative of a mean EPC. A further consideration for selecting DUs for ecological risk assessment is that more than one ecological end-point species is typically evaluated and their exposure areas may overlap in some cases or be distinct in others. Obtaining the assistance of a trained ecological risk assessor/toxicologist for exposure area and DU selection and early planning in each phase of an ecological risk assessment is highly recommended.

ISM data can be used to represent estimates of maximum exposure for the purposes of screening level assessments or assessment of threatened and endangered species. This would generally be a sufficiently conservative assumption when other conservative screening level assumptions such as 100% bioavailability and 100% of the diet coming from the most contaminated media are used, but is an assumption that must be agreed on prior to sampling. When ISM is used, the project team must determine a minimum spatial scale for exposure area DUs, often based on the home range or foraging range for representative species. However, since many ecological receptors have exceedingly small or exceedingly large home ranges, it is not always possible for the scale of sampling to exactly correspond with the scale that reflects the wildlife receptors, especially when multiple receptors (multiple species) are considered. In other words, excessively small exposure area DUs may be impractical, while excessively large exposure area DUs may be meaningless. Several options could be considered: use of other sampling methods (see Section 3.1.4.2), extrapolation from sampled to unsampled areas (see Section 7.2.6), and use of statistical approaches to provide conservative estimates of exposure concentrations (see Section 4.4.4). Note that some regulatory agencies do not accept certain options.

Sampling designs for ecological risk assessments must consider not only the size and location of exposure area DUs, but also the sample processing procedures. These procedures must be consistent with the technical requirements and objectives that mandate the exposure assumptions (USEPA 1997).

3.3.4.4 Exposure area decision units based on preferential exposure

As defined in the glossary, an exposure area is the specified area throughout which a potential receptor is exposed. Contact with all parts of the exposure area is equal through random movement within the exposure area over time. But as introduced in Section 3.3.4.1, in some situations standard-sized exposure areas do not reflect the known or suspected movement of the receptor; that is, the human or ecological receptor prefers some areas over others. Therefore, the CSM and the resulting sampling plan should consider the suspected or actual movement of the receptor by the use of smaller exposure area DUs 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 such scenarios 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. In

this example, the sampling plan should reflect two exposure area DUs, where two ISM samples are collected, one confined to the swing set area and the second across the remainder of the yard.

3.3.4.5 Example exposure area decision units

Figure 3-4 depicts the investigation of an older, residential home that is suspected to have been built on top of a former pesticide mixing area. The objective of the investigation is to determine whether pesticides and lead in the soil pose a direct-exposure hazard to the residents. Contaminants include arsenic and dioxins as well as pesticides.

The majority of the yard is included in DU-1 because the exposure scenario consists of the assumption that an equal amount of time will be spent in all parts of the back yard over the assumed exposure duration. Lead-based paint is suspected to have been used on the house; therefore, DU-2 consists of the drip-line of the house, which is suspected to contain elevated levels of lead in the soil. DU-2 can be considered both an exposure-area DU and a source-area DU because it represents the exposure to chips of lead-based paint, which may be of concern for acute or subchronic exposure and is a potential source of lead. Exposure patterns in the front and side yard are different than in the backyard, therefore; DU-3 is designated as the front and side yards. ISM samples consisting of the specified number of increments would be collected from each of the DUs.

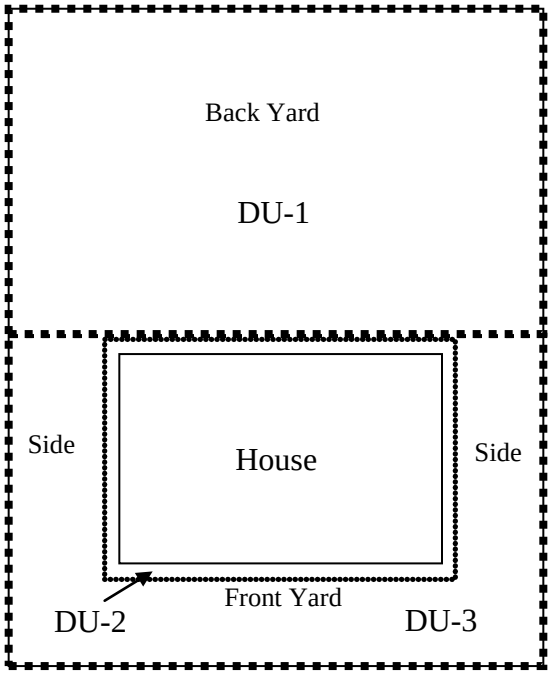


Figure 3-4. Exposure area DUs designated for a residential house lot. The back yard was designated as a separate DU from the front and side yards based on anticipated use patterns due to higher frequency of exposure in backyard.

3.3.5 Source Area Decision Units

Source areas are of concern because contamination can migrate from source areas to other locations and media (e.g., leaching to groundwater, volatilizing to soil gas and/or indoor air, or running off to surface water). Source areas can also result in additional releases, direct exposures, and other issues associated with gross contamination (e.g., risk of explosion, nuisance issues, or inappropriate disposal). The identification and characterization of source areas is an important and generally necessary part of a typical investigation. Source area DUs can be identified 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, whether discrete or ISM. When sufficient sampling evidence indicates high concentrations of contaminants are present in soil, it may be possible simply to

make a decision using those sample results. However, in other situations it may be necessary to refine the sampling plan, redesignate DUs (perhaps on a revised scale) and resample.

Once identified, source areas should generally be designated as discrete DUs and sampled accordingly.

Once identified, source areas should generally be designated as independent DUs (source area DUs) and sampled accordingly. If source areas are successfully demarcated from surrounding soil, ISM is a useful approach to determining the mean concentrations of contaminants within these source areas. DUs derived from exposure areas (exposure area DUs) are not generally recommended for source areas because the environmental hazards represented by the source areas are not directly related to the concept of exposure areas or the scale they represent.

Note that the definition provided here for source area does not presume that any particular type of decision mechanism or action level is required. Sample results derived from source area DUs may be compared in an appropriate decision mechanism to any type of action level. Decision mechanisms are discussed in Section 7, and source area sample results may be compared to any type of action levels (e.g., derived from risk assessment procedures, state or federal regulation, background estimates, disposal requirements, or any other selected approach).

A final point concerning source area DUs: In many situations where the source area is known to be highly contaminated, discrete samples may be the best option. In this situation, even though it is likely that contaminant concentrations are highly heterogeneous, any discrete sample is likely to result in a correct decision because the concentrations are elevated.

3.3.5.1 Smaller source area decision units within exposure area DUs

In some situations, it may be advisable to designate smaller source area DUs or SUs within larger DUs, based on an understanding of contaminant distributions. Assessment of a smaller subarea might be motivated by the suspicion that concentrations are higher in that location relative to the surrounding soil. Depending on site-specific conditions and contaminants, there could be concerns about acute, subchronic, or chronic exposures to soil in these smaller DUs. A common example is the need to investigate soil around a house suspected to be contaminated with lead-based paint chips. The perimeter of the house may be designated as a separate DU and characterized separately from the larger exposure area DU consisting of the entire yard.

In this example, there might also be a need to designate a separate DU because of the need to address the potential presence of large particles (i.e., large chips of potentially lead-contaminated paint) to meet the project objectives for FE. For example, particle size reduction and increased sample mass for extraction might be necessary (see Section 4 for further details).

3.3.5.2 Example source area decision units

Figure 3-5 provides an example of how a source area might be investigated using ISM samples. The CSM suggests that contaminant concentrations should be highest immediately below the source (in this case an aboveground storage tank [AST]) and will gradually diminish with lateral distance from the source. Therefore, the soil immediately below the source area is designated as a

single source area DU, while additional DUs are selected surrounding the source as discussed in Section 3.3.6.

Figure 3-5 depicts a source area DU established at the perimeter of a suspected release. The site is located within a small, industrial complex, and various chemicals are suspected as part of the release. It is also assumed that the greatest potential for contamination is associated with near-surface soils, to a depth of approximately 10 cm. The primary objective of the investigation is to determine the presence or absence of potentially significant polychlorinated biphenyls (PCBs) concentrations in the immediate vicinity of the pad.

Figure 3-5. Source area DU in a suspected release area.

3.3.6 Decision Units Surrounding Source Areas

Contamination often migrates away from source areas over time. In this situation DUs must be selected to identify boundaries of contamination with sufficient resolution to meet the project objectives. However, the boundaries between source areas and clean soil are often not evident prior to sampling. Demarcation of these boundaries is often an important goal of investigation. As discussed throughout this document, discrete sample results can be subject to potentially extreme variability. Therefore, a few discrete samples are not necessarily reliable for demarcating this boundary. As depicted in Figure 3-6, a series of concentric DUs sampled using ISM may be the best alternative. In this situation hypotheses are formed concerning the limits of the contamination associated with the source area and the resolution required for the project. These hypotheses are then tested using ISM samples associated with DUs surrounding or downgradient of the source area. Concentric DUs may be appropriate when there is no readily apparent likely migration direction from the DU. Available information concerning the limits of source areas should always be used when determining DUs for ISM samples. Visual clues and field instrument data in particular should be used if available.

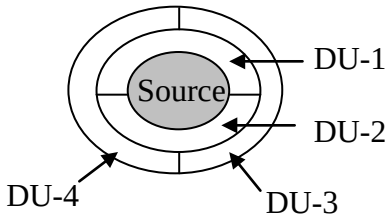


Figure 3-6. Designation of perimeter DUs around a source area DU.

Recall that source areas are defined as volumes of media containing "elevated concentrations," which are generally of concern due to their potential for leaching, vapor intrusion, or acute exposure. They are not simply volumes of media whose concentrations exceed action levels derived using chronic exposure assumptions. Therefore, after moving away from the known

source area, the goal of the investigation changes from delineation of the boundaries of "elevated concentrations" within the source area to determination of DUs whose mean concentrations exceed chronic action levels. In this second situation, exposure area DUs may be more appropriate.

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

Additional DUs are suggested in the situation depicted in Figure 3-6 because the CSM indicates that the contamination gradually diminishes in the lateral directions away from the source. To include increasingly larger volumes of soil in a single DU in this situation (for example combining DUs 1–4 in Figure 3-6 into one DU) eventually results in dilution of the estimated mean concentration of the contaminant in the DU. This is a fact of environmental investigation; however, it is not unique to ISM sampling. Decisions based on discrete and composite sampling must also eventually be based on sample results which represent specified volumes of media, and the dimensions of those volumes of media must be determined with sufficient resolution. This rule always brings up the question "How much soil does a sample represent?" In the case of ISM, the investigator must make reasonable hypotheses about the volume of soil that is contaminated around a source area, sample that volume, and make decisions based on the results. Those DUs surrounding the known source area will be compared to action levels derived from chronic exposure in exposure areas; they are no longer source area DUs. In many situations the volume of soil surrounding a source area that is hypothesized to be contaminated and then tested with ISM samples will be relatively small in comparison to human health exposure areas. This means that decisions based on action levels derived from chronic exposure scenarios will be more conservative than those based on large exposure areas.

In summary, definition of source area DUs should be based on an understanding of the boundaries of the known or suspected source as reflected in the CSM. Rote sizes derived from the exposure areas in risk assessments are not recommended. However, once the investigation has moved away from the discernible boundaries of the source area, DU size and shape must be selected in a manner which reflects reasonable hypotheses about contaminant migration away from the source. Further away from the source, DUs should be based on exposure areas.

3.3.7 Subsurface Decision Units

Because of the frequency with which subsurface contamination is encountered, subsurface DUs are an important application of ISM sampling. Some CSMs may suggest that contamination extends only to depths of a few centimeters. But in other situations the volume of interest may be situated entirely below ground surface (bgs). Therefore, DUs are inherently 3-D and necessarily extend some depth into the subsurface. Vertical depths and intervals must be carefully considered based on the CSM, previous data, screening results, and applicable state laws. Objectives for the investigation of subsurface soils might include assessment of the following:

- leaching of contamination from soil to groundwater
- volume of contaminated soil that may need to be removed or properly managed

- potential for subsurface soil to be excavated during future site development and spread out at the surface, posing direct-exposure hazards
- grossly contaminated soil

Subsurface soils should be subdivided into DUs in layers based on clues from the CSM and the project objectives. One to several layers may be necessary. The DUs should indicate the vertical limit of contamination. Initial subdivisions may later be revised if more exact thickness resolutions are necessary for remediation decisions than for initial investigations.

Subsurface soils should be subdivided into DUs (or SUs) in layers based on clues from the CSM and the project objectives.

Ideally, subsurface DUs should be investigated in a manner that allows every possible increment in the DU an equal likelihood of being collected. Sampling theory also suggests that the entire cross section of the DU be sampled in each increment making up the ISM sample. In practice, however, the combined mass of the increments from a large number of borings would likely result in a sample volume that is impractical. Therefore, field subsampling plans or other compromises may be needed.

In addition to sufficient lateral coverage as with surface DUs, sufficient vertical coverage in subsurface DUs is an important consideration. Sampling approaches for subsurface soils differ from those applied to surface soils because access to the subsurface is more difficult. This does not mean that low-quality data are unavoidable for subsurface soils; sufficient coverage of the DU at depth is still necessary, and improving sampling techniques support higher-density sampling and thus higher-quality concentration estimates are possible for subsurface soils. Section 5 goes into further detail on sampling techniques for subsurface soils.

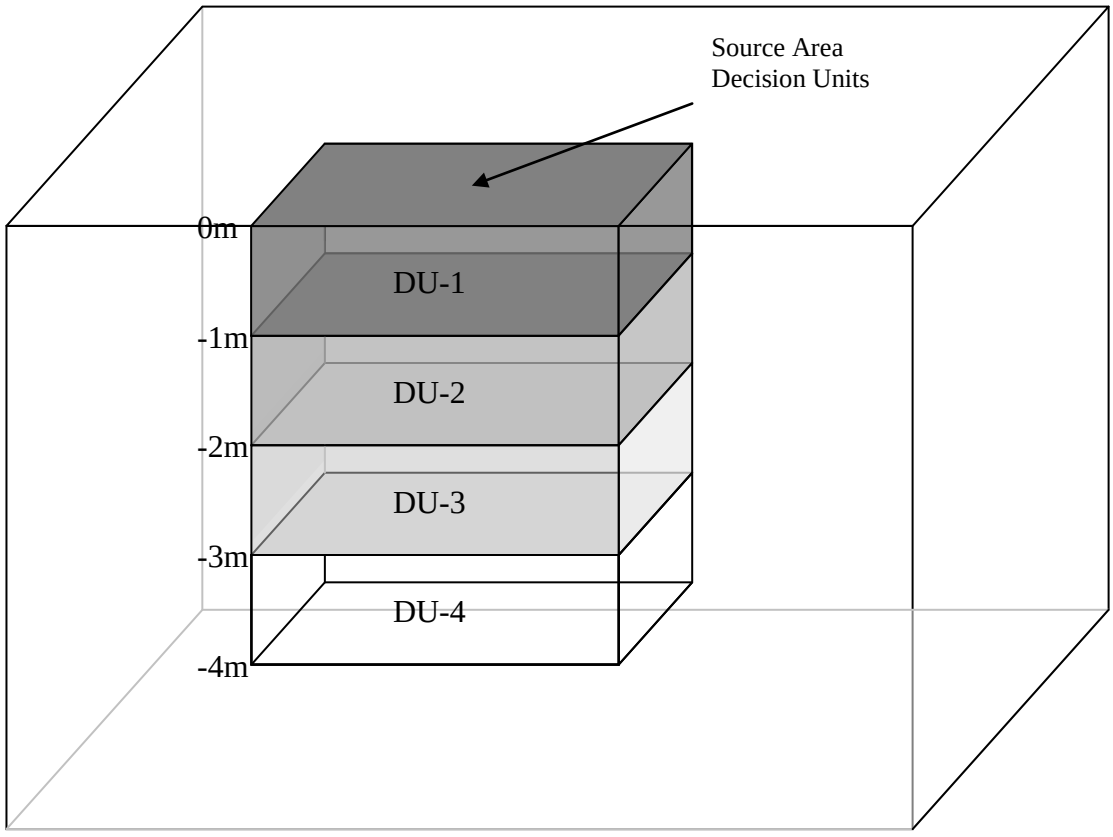
The thickness of each DU is based on balancing factors such as the desired resolution of the investigation and potential disposal costs and the actual time and cost of the investigation. The cost of collecting subsurface ISM samples must be balanced with the cost of analyses as well as the limitations of discrete samples discussed throughout this document. Discrete subsurface soil samples and/or field analytical methods can provide useful screening data prior to, or in conjunction with, subsurface ISM investigations. The results of the screening investigation can be used to determine the number, location, and dimensions of subsurface DUs and ISM samples.

It is important to note that assessing the potential intrusion of vapors into existing or potential buildings may be necessary. Collection of discrete soil gas samples is the recommended sample method for evaluating the vapor intrusion to indoor air route of exposure. During initial site screenings there may be some benefit for the collection of “bulk” soil samples for VOC analysis. ISM might be a useful method at some sites where VOCs may be present and should be used on a case-by-case determination. USEPA Method 5035 should be considered and used where appropriate. Several states have guidance on characterization of the vapor intrusion pathway and individual state guidance should be consulted as appropriate.

Example subsurface source area decision units

Examples of subsurface contamination include soil that has been capped by a layer of clean fill, paving material, or building slabs; leaking underground storage tanks (USTs); buried pipes; buried disposal sites; and surface spills that have spread downward. Figures 3-7, 3-8, and 3-9 depict potential approaches for investigation of shallow, deep, and isolated subsurface DUs.

Figure 3-7 depicts a former pesticide mixing operation with known spills and releases. Contaminants include pentachlorophenol, dioxins, furans, and triazine pesticides. Environmental hazards posed by these contaminants include direct exposure, leaching, and contamination of groundwater. The CSM depicted in Figure 3-7 indicates that contamination extends from the ground surface downward to a relatively shallow depth that can easily be reached with a backhoe. Therefore, excavation provides easy access to the desired sample depth. Because the evaluation of leaching potential is a project objective, focus on source area DUs is appropriate.



(not to scale)

Figure 3-7. One-meter vertical resolution DUs selected to help isolate heavily contaminated soil from less-contaminated soil and assist in evaluation of remedial alternatives.

It is assumed that the lateral boundaries of the source area have already been determined. Most of the contamination is believed to be restricted to the upper 2–3 m of soil and will be excavated and disposed at an off-site facility. Investigation objectives include the following:

- determine the vertical extent of contamination
- identify and separate potential hazardous wastes

In the example shown in Figure 3-8, the CSM suggests that much deeper soils are of interest because a long-term release from the source has migrated to deeper soils. Therefore, a series of vertically stacked source area DUs extending from the ground surface to depth was chosen. Soil borings were necessary to collect the sample increments. Depending on local conditions, direct-push technology might also have been useful.

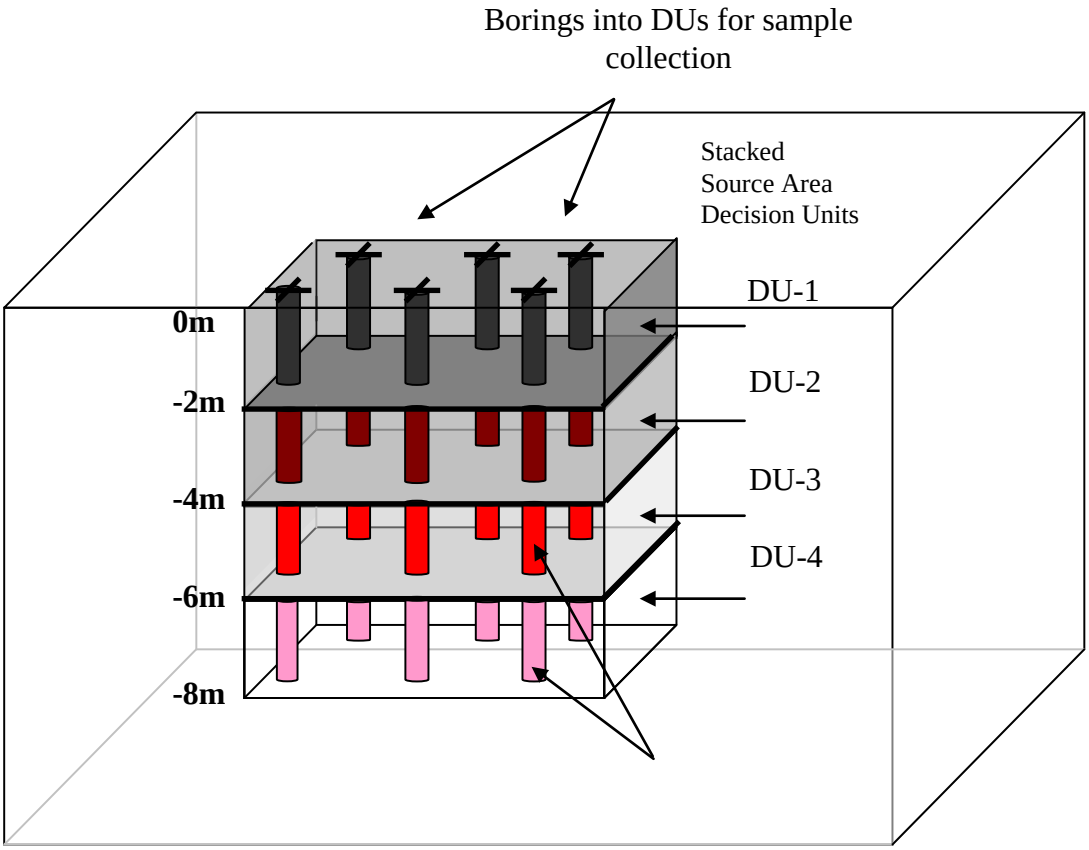


Figure 3-8. A hypothetical investigation of series of stacked source area DUs using borings.

In the final example (Figure 3-9) contaminated soil is believed to have been overlain by clean soil, so once again soil borings are necessary. Figure 3-9 depicts an alternative example where shallow contaminated soil has been excavated but a deeper unit of contaminated soil remained at depth. DU designation and investigation of the situations described in Figure 3-8 and in 3-9 are similar.

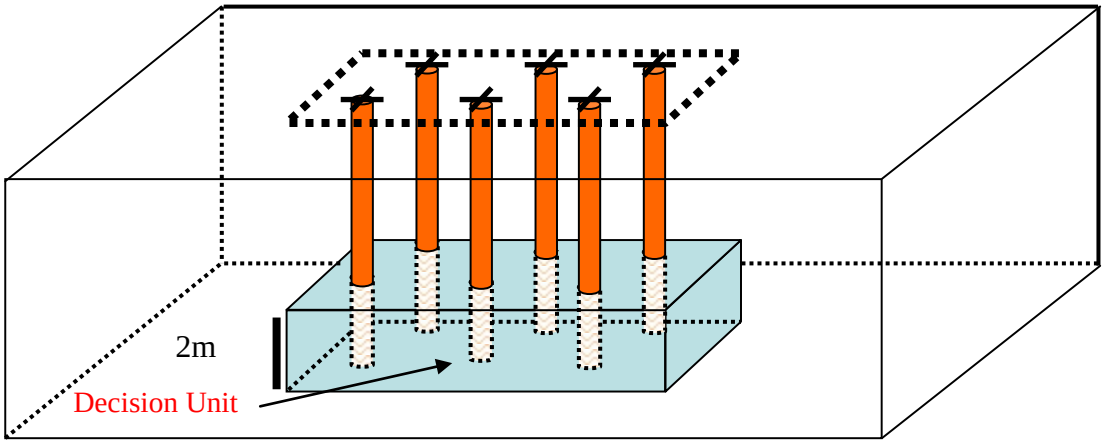


Figure 3-9. An example of a subsurface DU for contaminated soil overlain by clean soil that is not accessible by excavation. Although not shown, 30 borings per DU with incremental samples from borings across the entire depth of the DU is recommended to maximize vertical coverage.

3.3.8 Stockpile Decision Units

Stockpiles should be subdivided into volume-based DUs for sampling.

Stockpiles can offer easy access for sample collection. Data collected from stockpiled soils are often used to make decisions on disposal or reuse of the soil. Soil that is known or suspected to be contaminated should be segregated from soil that is presumed to be clean prior to sampling. Stockpiles should be subdivided into volume-based DUs for sampling based on the target contaminants, potential environmental hazards associated with the soil, and the desired use of the soil.

Example stockpile decision units

Physical sampling is best carried out as the stockpile is being created (e.g., collecting a specified number of increments from individual dump trucks or backhoes). The shape of the stockpile can bias characterization of the DU by limiting access to soil within the pile (ASTM 2006, HDOH 2009). Therefore the stockpile should be flattened to a thickness of approximately 3 feet or less before sampling to help ensure equal access to all soil and ensure that the samples collected are representative of the DU as a whole. Other means to ensure equal access to all soils during sampling may also be used, such as moving the pile from one location to another or digging into the stockpile with a backhoe or hand tool.

Figure 3-10 depicts example DUs for a large stockpile. The stockpile contains approximately 25,000 cubic yards of soil excavated from an agricultural settling pond. The soil is being considered for reuse in a new, residential development project. Each lot will be approximately 1 acre in size; therefore, the exposure areas DUs are designated as 1 acre. Each lot will be covered with approximately 6 inches of fill material from the stockpile. This plan results in an exposure area DU volume for each lot of approximately 800 cubic yards. The stockpile has been flatted to

a thickness of approximately 3 feet in preparation for sampling. The stockpile is subsequently divided into thirty 800-cubic-yard DUs of soil, and each one is tested with one or more ISM samples. This method helps provide equal access to all soil in the stockpile. Samples should be collected from the surface as well as the interior of the stockpile.

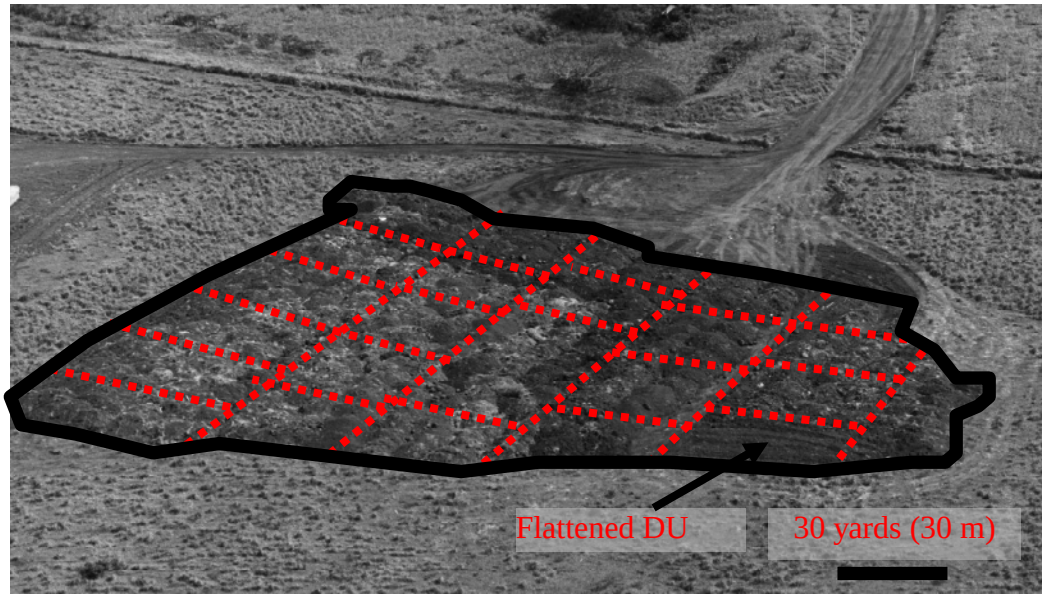


Figure 3-10. Example of stockpile DUs.

3.3.9 Excavation Decision Units

Excavation of contaminated soil from source area or exposure area DUs allows ready access to the assumed outer margins of the DUs, much as excavation of soil exposed at the ground surface allows easy access to the bottom of the DUs. The sidewalls and floors of excavations can be treated as separate DUs and can be sampled for confirmation of adequate soil removal.

This ISM approach is applicable if the criterion for successful excavation is achieving an average concentration in soil below the action levels at the excavation boundaries. Small areas of contaminated soil within otherwise clean excavation sidewalls or floors do not necessarily pose a significant risk to human health and the environment. In these situations use of ISM (or any other sampling methodology) for "confirmation samples" may represent a traditional approach, rather than a sampling plan based on the risk assessment model or targeted removal of source areas. Once an excavation is made, it may logical to base subsequent soil removal in the sidewalls on visual information or field instrument data. Unlike field data, ISM samples collected across multiple depths in the sidewall may not lend themselves to the original definition of DUs based on exposure areas or CSMs.

Example excavation decision units

Figure 3-11 shows excavation sidewall and floor DUs which may be assessed with ISM samples.

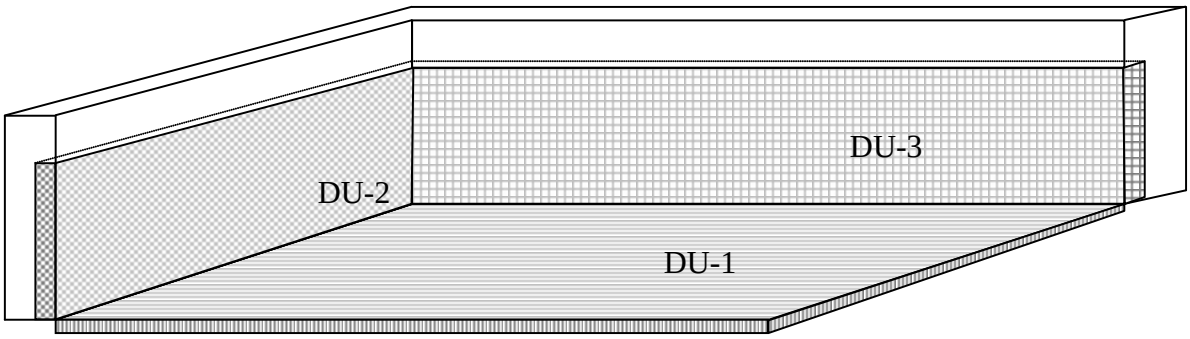


Figure 3-11. Floor and two sidewall DUs for an excavation site.

3.3.10 Decision Units for Very Large Areas

The investigation of hundreds or even thousands of acres of land for potential contamination can be logistically and financially challenging. DU designation should be based on the CSM. Areas of suspected heavy contamination should be investigated separately from areas of different historical use, and soil types and other factors should also be considered when distinguishing separate areas. Nevertheless, some areas for which a decision will be made may be so large that it would require an enormous number of individual DUs for complete coverage at an appropriate and useful scale. In these cases, sampling each DU may not be economically feasible. Sampling only a fraction of the area as SUs and extrapolating those results to make a decision regarding the DU will undoubtedly be attractive from a cost and logistical standpoint. However, this approach introduces additional uncertainty into the evaluation that may or may not be acceptable in achieving site objectives. Tradeoffs between economic feasibility and management of uncertainty must be considered when developing DUs for very large areas (see Sections 4 and 7). However, recall that decision mechanisms involving extrapolation are not acceptable to regulators in some states.

3.3.11 Other Types of Decision Units

ISM may be used to estimate background concentrations for comparisons with potentially contaminated volumes of soil (see Sections 4.4 and 7.2.4). The term SU is more appropriate than DU to describe the area and depth for ISM sampling for background concentrations because these types of areas will not require decisions. As for any background sample, background SUs should be located in a nonimpacted area that is geologically similar to the contaminated area. Ideally, the background sampling area will be of a scale similar to the potentially contaminated areas; however, this may not be feasible if the background data is used for comparing background and contaminant concentrations across multiple DUs.

3.4 Establishing New Decision Units Based on Previous Results

In some situations it may be necessary to develop new DUs based on revised sampling objectives following the assessment of initial results. For example, a DU may have been initially established of such size that remediation of the entire DU is not practical. In this situation, it may be advantageous to resample by dividing the original DU into smaller DUs and sample each smaller

DU to determine which have soils that do not meet cleanup goals and, therefore, should be remediated. Subdividing a source area DU designated for in situ treatment into smaller subunits can also help optimize the design of the remedial system based on a more accurate estimation of contaminant distribution within the DU.

3.5 Hot Spots

Historically, discrete soil sample results with concentrations above an action level have often been assumed to represent a significant volume of soil containing sufficiently high concentrations of contaminant to warrant concern. These assumed volumes have been considered to represent hot spots. The relative nature of this definition results in a wide range of interpretations and typically leads to subsequent remobilization and resampling intended to define the extent of the hot spot, often with insufficiently specified objectives. (Note that ITRC's *Use of Risk Assessment in Management of Contaminated Sites* [ITRC 2008] lists various state criteria and guidance values for hot spots).

To further complicate the issue, the terms "hot spot" and "source area" are sometimes used interchangeably. In this document, the distinction is that information about the location and likely extent of source areas are known or postulated based on a CSM at the outset of a sampling project; source area designation relies on more information than the interpretation of yet-to-be-obtained sampling results. The locations and dimensions of source areas, therefore, may be established or hypothesized a priori. Waste disposal units, spills, releases, and volumes of soil shown by previous sampling to have significant contaminant concentrations relative to the surrounding soil are defined as "source areas" in this document. In contrast, hot spots are considered to be soil volumes with relatively high concentrations that could be present at a site but whose locations and dimensions cannot be anticipated prior to sampling. The designation of a hot spot based on sample data alone has decision-making value only when the chemical criteria (how "hot") and spatial dimensions (what "spot") which will define hot spots are specified. Thus, these criteria must be agreed on by the planning team before sampling. As mentioned above, some states have established criteria for defining hot spots; therefore, it is highly recommended that project teams include their state regulators during early planning and that the project team understand the basis of the criteria used to define hot spots.

Effective detection and delineation of hot spots in heterogeneous soil matrices is a challenge. Results from both traditional low-density discrete sampling approaches and ISM sampling approaches have constraints and uncertainty associated with their interpretation. While, with enough samples, discrete sampling designs provide the user with concentration data on smaller spatial scales from very small discrete areas, there is often lack of sufficient coverage to detect the inflection points in concentration gradients that would establish the boundaries of hot spots. The primary problem with using discrete samples to search for unknown hot spots is that when a few discrete samples are collected in the presence of small- and/or medium-scale heterogeneity, isolated high concentration results can be misinterpreted as high concentrations of contaminants over significant volumes of soil. Similarly, ISM sampling using low-density or noncontiguous coverage of small hot-spot DUs over an area of concern may also be an ineffective means to detect and delineate hot spots. Areas of elevated concentrations could be detected using very high-

density ISM or high-density discrete sampling approaches, but for practical reasons such methods are seldom employed.

When used over relatively large areas (perhaps the size of quarter- or half-acre residential yards or larger), ISM typically captures the broad effects (i.e., proportional representation and thus higher average concentrations) of hot spots due to the improved spatial coverage within the DU, but it does not provide information on the spatial location of smaller volumes of soil containing hot spots of contaminants within the DU, nor does it indicate the magnitude of these areas of elevated concentration if they exist. To detect and delineate potential hot spots using ISM, DUs must be scaled down to be consistent with the area and depth (or volume) of soil of potential concern for hot spots. In other words, to detect a hot spot of a given size, the spatial dimensions of the DUs have to be that size or smaller. Additionally, the hot-spot DUs need to contiguously cover the area suspected of containing hot spots. While smaller DUs may provide better spatial resolution, as discussed above with discrete sampling approaches, there are practical limits on the number of DUs that can be designated, sampled, and analyzed. Therefore, using ISM to detect relatively small hot spots may also be infeasible in many situations.

To avoid the pitfalls of "chasing" areas of elevated concentration with no predefined boundary conditions and for the data to be useful for project decisions, ISM practitioners *must* predefine the area and depth of concern and the chemical criterion that will be used to define hot spots as part of the systematic planning process. It is also encouraged that the planning team understand and agree on the basis of the criteria to facilitate later decisions that will be made with the hot-spot DU data.

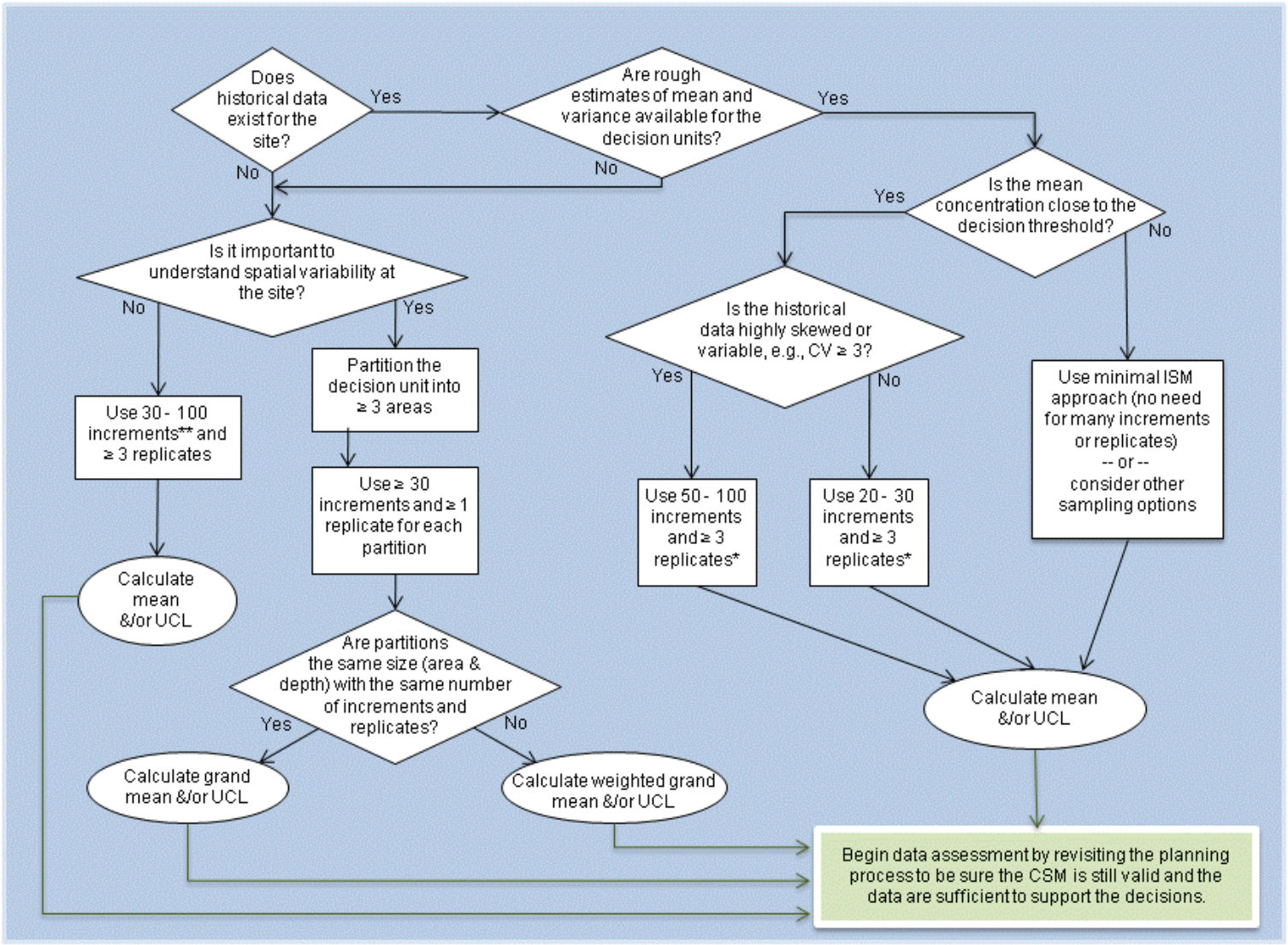
The definitions of source areas and hot spots provided in this guidance are intended to promote meaningful discussion on the purposes, limitations, ability, and *need* to detect and characterize volumes of soil smaller than exposure areas when concentrations are highly variable in heterogeneous particulate materials such as soil. If detection of contaminated volumes such as these is an important objective of an investigation, careful planning is vital.

4. STATISTICAL SAMPLING DESIGNS FOR ISM

This section summarizes results of simulation studies used to evaluate the performance of ISM in estimating the mean under various conditions. Conclusions from these studies are discussed, and recommendations for ISM sampling based on this evaluation are presented in Figure 4-1. The recommendations for ISM sampling design, number of increments, and number of replicates as shown in Figure 4-1 come from the simulation studies discussed in this section and presented in more detail in Appendix A.

As mentioned in Section 3, a variety of sampling designs should be considered during systematic planning. To determine which will fully meet the project objectives most efficiently, it is necessary to have some idea of how many samples will be required as part of the design. Figure 4-1 provides guidance on the number of increments and replicates to collect within a DU if incremental sampling is selected as a sampling strategy. Within each DU, the pattern for spatial collection of the increments is not specified in this figure but is discussed in Section 4.3.

Methods for estimating the mean concentration present at the site based on incremental samples (as depicted in the ovals in this figure) are also not presented in this figure but are discussed in detail in Section 4.2.



*Collecting more than 3 replicates will increase certainty in estimate of mean and UCL and is recommended in these cases. More than 10 have diminishing value.
**The number of increments depends on heterogeneity (highly variable sites require more increments) and on size (a small site may require fewer increments).

Figure 4-1. ISM decision tree.

After data are collected and reviewed, it is important to revisit the outcomes of the systematic planning process to ensure that the data meet the project objectives. The guidance offered in Figure 4-1 is meant as general guidance for the number of replicates and increments necessary to achieve particular objectives. These recommendations are likely to provide sufficient information to meet most basic objectives relating to comparison of an estimated mean or UCL for the mean to a decision threshold. However, the project team must consider whether or not the project objectives have been fully satisfied by the data collected. If the data are not satisfactory for decision making, further consideration and revision of the systematic planning process and outcomes are necessary.

4.1 Factors that Complicate Estimating the Mean Concentration

ISM sampling produces an estimate of the mean contaminant concentration in soil within a specified volume (i.e., a DU). As with any estimate derived from sampling, ISM results are subject to error, the components of which are described in Section 2.5. Understanding error introduced by sampling is squarely in the domain of statistical analysis. Rigorous statistical analysis regarding the extent to which various ISM sampling strategies provide accurate estimates of the mean contaminant concentration have not yet been published. This information is necessary to understand how factors such as number of increments, number of replicates, and contaminant distributions across the site influence the reliability of ISM estimates of mean contaminant concentration. An evaluation of the reliability of ISM based on statistical principles is vital to widespread acceptance of this sampling method for regulatory purposes.

Statistical evaluation of ISM is a new area. Thorough evaluation of ISM is a substantial undertaking, well beyond the scope of this document. Thus, the findings presented here should be viewed as the best available to date but incomplete in terms of addressing all of the points and questions that might be asked. It is also important to note that analyses described in this report have focused on the extent to which ISM field samples represent the true mean of the DU, assuming that the concentration within those samples can be measured with complete accuracy. Statistical evaluation of subsampling methods in the laboratory is also important (see example in Gerlach and Nocerino 2003) but is not addressed due to time and resource constraints.

> The statistical analysis presented in this document evaluates how ISM *field* sampling procedures may influence the error in the estimate of the mean concentration.

Data on chemical concentrations in environmental media present challenges for estimating the mean concentration. This problem applies to both ISM and discrete sampling. If a DU is perfectly homogenous, meaning that the contaminant exists in the same concentration everywhere across the DU, developing a sampling strategy to accurately estimate the concentration is simple. For that case, all sampling approaches, from a single discrete sample to the most expansive set of ISM samples, would yield the same average concentration (within the limits of laboratory error), and thus any can provide a reliable estimate of the mean. Unfortunately, this ideal situation is never encountered in soils. Site concentrations typically exhibit some degree of heterogeneity, and the greater the heterogeneity, the more difficult it is to accurately estimate the mean concentration through sampling. As discussed in the next section, this difficulty gives rise to error in the estimation of the mean, and different sampling approaches yield different values for the mean. This error can be managed so that reliable estimates of the mean can be produced, but

management requires an understanding of how the number of discrete samples, or the number increments and replicates in ISM sampling, affects estimates of the mean. Simulation studies to develop this understanding are described later in this section.

4.1.1 Skewness and Dispersion

The coefficient of variation, geometric standard deviation, and coefficient of skewness are all measures of dispersion of a distribution.

Both the skewness (asymmetry) and dispersion (spread) in the data can affect the confidence in estimates of the mean. Since it is common for environmental data to exhibit positive skewness (i.e, a longer right tail) or a wide range of concentrations, one challenge for sampling designs is to represent the upper and lower tails of the distribution in the proper proportion, thereby yielding a relatively precise estimate of the mean. For data sets generated with discrete sampling, graphical and exploratory data analysis techniques are commonly used to assess the degree of skewness and dispersion. For example, by plotting the data using histograms and probability plots, the distribution shape and the presence of multiple populations may become apparent. This assessment can be further supplemented by a statistical analysis of the goodness-of-fit (GOF) to normal, lognormal, or gamma distributions. Summary statistics can also be informative and readily calculated with both free and commercial statistics software, including (a) coefficient of skewness; (b) the ratio of the standard deviation (SD) divided by the arithmetic mean—referred to as the "coefficient of variation" or "relative standard deviation" (RSD); and (c) geometric standard deviation (GSD), comparable to the coefficient of variation (CV) (see footnotes of Table 4-1) and used specifically with lognormal distributions.

Table 4-1. Data dispersion in terms of CV and GSD

CV[a] (unitless)	GSD[b] (unitless)	Variability/ dispersion
≤1.5	≤3	Low
1.5–≤3	3–≤4.5	Medium
>3	>4.5	High

[a] Coefficient of variation (CV) = standard deviation (SD)/mean.

[b] Geometric standard deviation (GSD) = exp[sqrt(ln(CV^2 + 1))] for lognormal distributions.

For convenience in this document, the degree of dispersion of the concentration distribution in a DU is classified in terms "low," "medium," and "high," as shown in Table 4-1. These categories can be used to guide the selection of methods used to calculate the UCL in the mean, as discussed in Section 4.2.

As discussed below, the distribution of the contaminant distribution in the DU is different from the distribution of DU means that is characterized by ISM sampling. Table 4-1 provides categories of dispersion for the contaminant distribution throughout the DU rather than the distribution of the DU means. For data sets generated with ISM, fewer exploratory data analysis options are recommended due to the relatively small number of samples. For example, one would not generate a histogram or perform a GOF test on a data set consisting of three replicates. Nevertheless, summary statistics of replicates can provide a measure of the precision in the

estimates of the mean, which can be a useful diagnostic for evaluating laboratory DQOs (see Section 4.3.4.4). The mean and variance of the ISM samples can also used to calculate the UCL for the grand mean. However, as discussed in Section 4.3, the RSD statistic does not serve as a reliable performance metric of a UCL calculation because the true DU mean is never known. In addition, simulations demonstrate that, for data sets in which the sample mean is less than the true mean, the likelihood that the UCL also underestimates the mean increases as the sample RSD decreases, due to the positive correlation between the estimated mean and estimated variance.

4.1.2 Spatial Scale, Mixtures, and Autocorrelation

It is important to recognize that the extent of heterogeneity can vary depending on how DUs are defined. In fact, one way to manage the difficulty of estimating the mean when greater heterogeneity is present is to designate DUs based on anticipated concentrations, defining DUs in such a way as to minimize the concentration variability within each. Other approaches for creating DUs, such as designating DUs according to anticipated exposure patterns (i.e., to correspond with exposure units), could result in greater heterogeneity within the DUs but may be appropriate for risk assessment.

Sampling designs, including designations of sampling units and decisions units, may need to accommodate multiple contaminants with different spatial patterns.

Heterogeneity may be different between contaminants being characterized within the same DU. Different sources or release mechanisms, as well as different transport mechanisms, can lead to differing degrees of heterogeneity among chemicals that need to be addressed through a single sampling plan. This fact can complicate decisions regarding the appropriate sampling approach. In general, the sampling strategy must be designed to accommodate the contaminant expected to have the greatest heterogeneity in order for good estimates of the mean to be obtained for all contaminants of interest.

Yet another potential complicating factor is spatial relationships. For most sites, contaminants in soil exhibit some degree of positive spatial autocorrelation, meaning that the variance in the concentration reduces as the distance between sample locations decreases. It is well established that strong autocorrelation can reduce the effective statistical sample size of a data set (i.e., number of samples needed to achieve acceptable decision errors) because each sample provides some redundant information (Cressie 1993). In statistical terms, this redundancy violates the assumption that observations are independent. ISM confidence intervals generated from sampling of a site with high spatial autocorrelation can be too narrow, resulting in a higher frequency of decision errors. Spatial autocorrelation may also introduce bias in estimates of the mean and variance (and corresponding calculations of confidence intervals), depending on the sampling protocol. Random sampling strategies yield unbiased parameter estimates, whereas sampling that is targeted towards areas of suspected high or low concentrations can introduce redundancies that result in inaccurate calculations of confidence intervals and inaccurate estimation of decision errors. For targeted (nonrandom) sampling, the direction of the bias is generally towards overestimation of the mean since suspected source areas may be intentionally oversampled relative to the rest of the site. Nonrandom sampling of sites where contaminants exhibit positive spatial autocorrelation is an issue that applies to discrete as well as ISM sampling. With discrete sampling, spatial weighting methods are sometimes used to reduce the sampling bias. For ISM,

spatial weighting methods do not apply since no information is retained from the individual increments collected throughout the DU. Nevertheless, since most ISM sampling protocols incorporate some variation of random sampling and a relatively large number of increments (i.e., n ≥ 30), spatial autocorrelation is unlikely to impact the statistical performance metrics of ISM (Section 4.3). See Appendix A.3 for an example and additional discussion of this factor.

For both discrete and ISM approaches, random sampling yields unbiased parameter estimates, even when a contaminant exhibits high spatial autocorrelation.

4.2 Uncertainty in Estimates of the Decision Unit Mean

Even the most comprehensive sampling protocols introduce some degree of sampling error. Therefore, one challenge in developing sampling designs is to balance the potential for decision errors against the practical constraints of site investigations, including having incomplete information about potential source locations, as well as time and budget limitations. The objective of ISM is to provide a reliable estimate of the average (i.e., arithmetic 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. A principal objective of systematic planning of most sampling designs is to minimize the major sources of error in both the field and the laboratory. In practice, the estimated variance is often viewed as an overall measure that includes the contribution of many sources of error. Just as with discrete sampling, the estimated variance can be used to quantify a UCL for the mean for ISM samples and the same UCL equations apply. This section describes important concepts relevant to characterizing variance in ISM sampling. Section 4.3 builds from these concepts by presenting the results of simulation studies that examine the performance of alternative ISM sampling strategies applied to a wide range of theoretical site conditions.

4.2.1 One ISM Result

For sites where there is a regulatory requirement to calculate a UCL, at least three replicates should be collected within a DU. For sites where there is no regulatory requirement to calculate a 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.

Each ISM sample provides an estimate of the true mean—the actual average concentration within the DU. As such, the distribution of ISM results is related to but conceptually different from the distribution of discrete samples. The two approaches share the same grand mean but can be expected to have different estimates of variance. For ISM, the mean of replicates is analogous to repeated trials of discrete sampling (i.e., the mean of the means, or the "grand mean"), and the standard deviation is analogous to the standard error for the mean in discrete sampling. Even the most comprehensive sampling protocols will introduce some degree of sampling error, and it is possible that a single ISM sample 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, increasing as the heterogeneity increases. Figure 4-2 illustrates the probability and magnitude of underestimation of a single ISM sample of n=30 increments collected from DUs

with underlying lognormal distributions with CVs ranging 1.0–3.0. The following observations are noted:

- A single ISM sample will underestimate the mean more than 50% of the time for all positively skewed distributions.
- The magnitude of the underestimation depends on the degree of variability, as represented by the CV.

A single ISM result is likely to underestimate the mean more than 50% of the time for most distributions; the likelihood of a decision error increases as the variance in the distribution increases and the difference between the action level and true mean decreases.

- Approximately one-third of the sampling events with a single ISM sample (n = 30 increments) will underestimate the mean by up to 10% for CV = 1 and 20% for CV = 2. For example, if the true mean is 400 ppm, approximately one out of every three ISM samples (n = 30) will yield an estimated mean <360 ppm for CV = 1, and <320 ppm for CV = 2.
- For a distribution with greater dispersion (i.e., CV = 3), approximately one quarter of the sampling events will yield a single ISM result that underestimates the mean by 30%–60%. For example, if the true mean is 400 ppm and CV = 3, approximately one out of every four ISM samples (n = 30) will yield a sample mean 160–280 ppm.

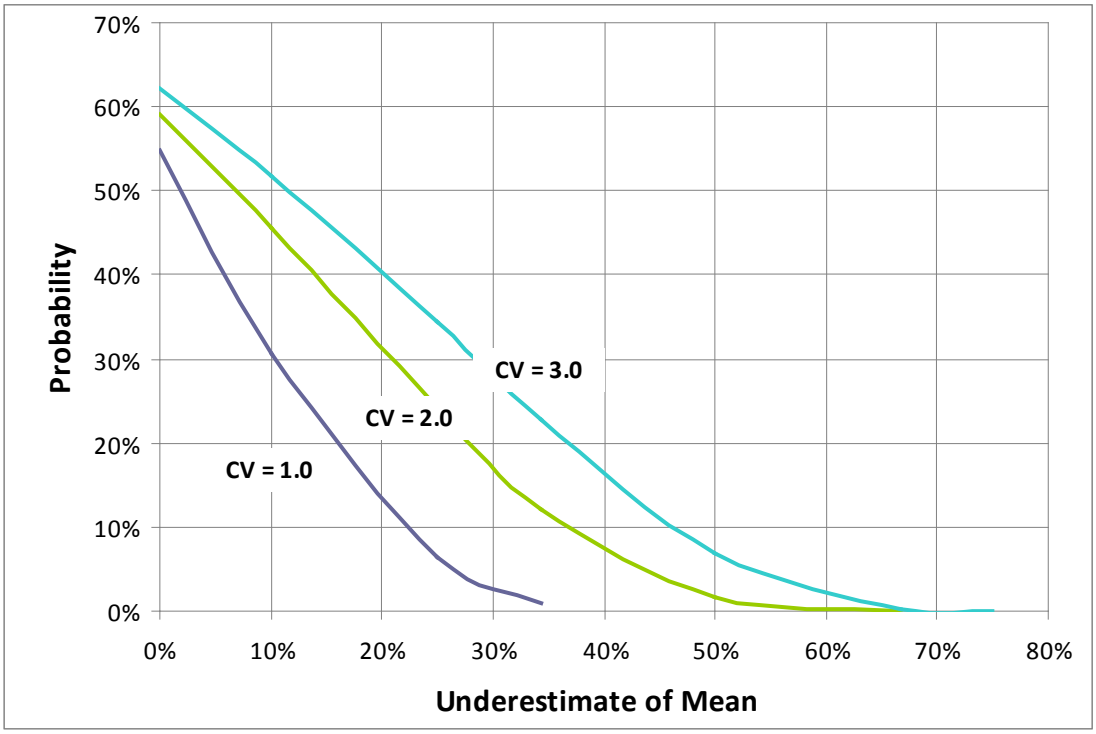


Figure 4-2. Examples of the probability and magnitude of underestimation of the mean from a single ISM sample.

The same issues with underestimation apply when discrete sampling is used to estimate the mean. As heterogeneity of the soil concentrations increases and sample size decreases, the potential magnitude of error in the estimate increases. Consider what would happen if you sent

crews out to the same DU 100 times to collect an ISM sample of 30 increments or a series of discrete samples with which to calculate a mean concentration. If the separated estimates of the mean from these sampling events were plotted as a histogram, they might fit distributions shown in Figure 4-3. The top panel shows estimates of the mean that are normally distributed around the true mean of 100. Given that it is a normal distribution, the estimated mean of approximately half of the sampling efforts is below the true mean and half of the efforts produced an estimated mean above the true mean. The spread of the distribution gives an indication of how far away from the true mean some of the estimates were (i.e., an indication of the potential magnitude of error). As the top panel illustrates, although both distributions are unbiased (centered at the true mean), variability in estimates of the mean are generally less for ISM than for comparable discrete samples due to differences in number of samples collected.[2] The lower panel in Figure 4-3 shows that the potential magnitude of error increases as the estimates of the mean becomes skewed due to heterogeneity.

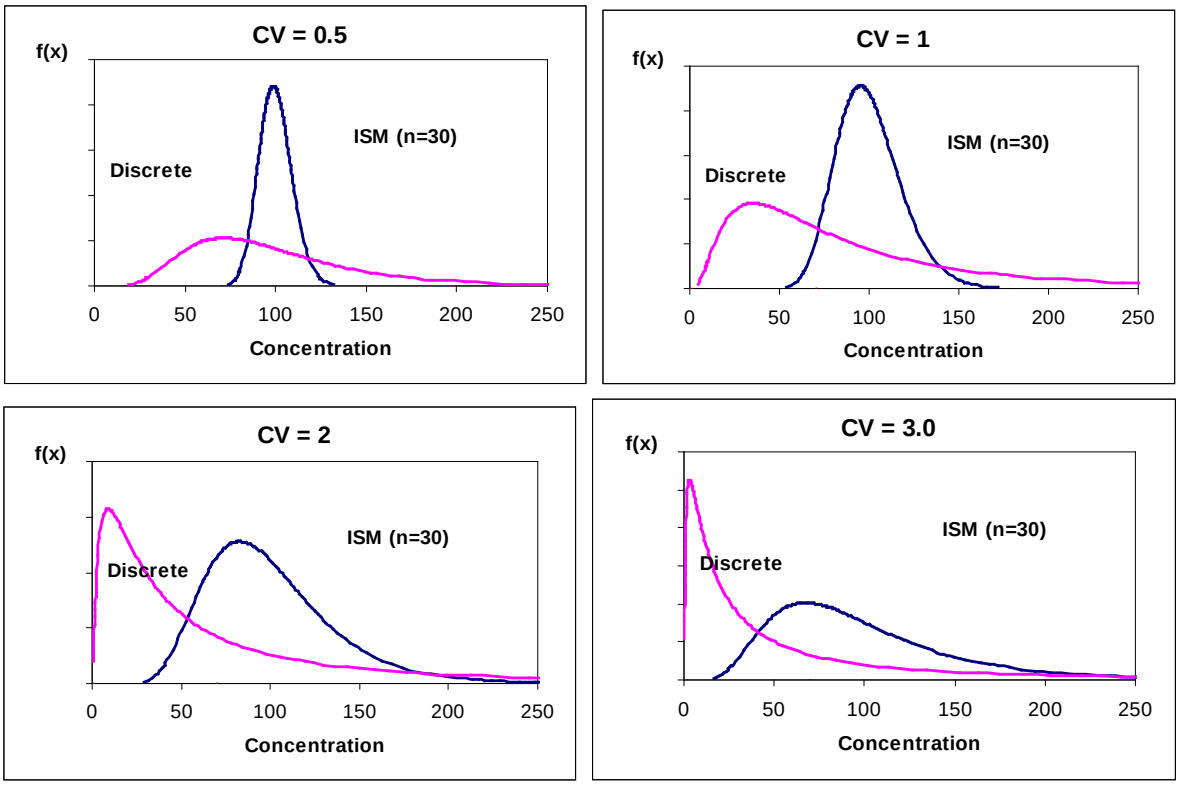


Figure 4-3. Dispersion of means from ISM (based on n = 30 increments) applied to a lognormal distribution (mean = 100) with CVs ranging 0.5–3.0.[3]

[2] Note that the distribution of ISM means (from repeated trials of one ISM sample) and means estimated from discrete samples would be expected to be equivalent if the number of discrete samples was the same as the number of increments.

[3] ISM generates a distribution of means that approaches normality, as predicted by the central limit theorem. However, the ISM distribution can also by asymmetric, and the asymmetry increases with increasing dispersion of the underlying distribution.

From a statistical standpoint, it should be noted that analysis of multiple ISM samples collected with the same sampling protocol (i.e., sampling method and number of increments) provides a direct measure of the variance in the mean. It is important to recognize that the distribution of replicate results is different from, but related to, the distribution of discrete results (X_i) obtained from the same population. As shown in Figure 4-3, both sampling types share the same estimate of the population mean (i.e., 100) but not the same variance. The variance of ISM samples composed of *n* increments is lower than the variance of discrete samples comprised of *n* discrete measurements. While this example is an oversimplification of the differences between ISM and discrete sampling, it highlights an important statistical concept related to sampling from populations.

Discrete and ISM samples yield different distributions for the mean. They share the same (grand) mean but have different shapes and variances.

In practice, you can't send a crew out to sample the same DU 100 times and assess the variability. Instead, you typically have to obtain a reliable estimate of the mean through a single sampling exercise. Through understanding the concept of variability in estimates of the mean and the influence of heterogeneity, the limitation of basing a decision on a single ISM sample becomes apparent. There is no way to know whether any one estimate provided by a single sampling event is above or below the actual mean in the DU as well as the potential magnitude of the deviation from the actual mean without additional sampling data to assess heterogeneity of the concentrations within the DU.

By collecting multiple ISM samples within a DU (i.e., ≥3 replicates), we can obtain a direct measure of variability in sample means and calculate a UCL for the mean with an acceptable decision error rate.

Recognizing that variability and errors in estimates of the mean exist, regulatory agencies often require a 95% UCL to represent an EPC or to assess compliance with decision criteria. Just as with discrete sampling, the variance for replicate ISM samples can be used to estimate the standard error for the mean, which is one factor in the calculation of a UCL as discussed below. Similar to the difference in discrete and ISM variance estimates described above, the UCL calculated from ISM replicates is generally different (and lower) than the UCL calculated from discrete samples with typical sample sizes. In the case of ISM, the UCL can be thought of as a measure of the confidence in the estimate of the "grand mean," or the mean of the means given by replicate samples. In practice, it is expected that a typical ISM sampling protocol will consist of a relatively small number of replicates (e.g., three to five replicates). The small number of samples may have several implications on the performance of the ISM sampling effort, depending on the properties of the contaminant distribution at a site (e.g., heterogeneities, spatial patterns in the distribution, etc.).

With ISM, the UCL can be thought of as a measure of the confidence in the estimate of the "grand mean," or the overall mean of the individual means given by each replicate sample.

4.2.2 UCL Calculation Method

The concept of variability in estimates applies to UCLs as well as to the estimates of the means themselves. Several methods exist for calculating a UCL for estimates of the mean for a set of data. These methods often yield different answers for the same set of data. For example, if a 95%

UCL[4] is estimated for a population 100 times, the 95% UCL will, on average, be greater than or equal to the true mean of the population 95 times. The ability of different methods to produce a value that meets the definition of a 95% UCL depends in part on the number of samples used to estimate the mean, as well as the distribution (e.g., normal, lognormal, gamma) and dispersion of the data. One method might generate 95% UCLs greater than or equal to the true mean for a population 95% of the time, while another 95% UCL method might generate estimates greater than the true mean only 80% of the time. In the latter case, although a 95% UCL method was used, that method did not perform up to the specified level for that population. Had more samples been taken to estimate the mean or if the concentrations were distributed differently, the second method might have performed satisfactorily while the first method was deficient.

In practice, we cannot compare the performance of any UCL calculated at a site because the true mean within the DU is unknown. Similarly, there are no statistical calculations or diagnostics that can be used to compare the individual replicates or UCL to the unknown mean. These are limitations that apply to both discrete and ISM sampling. However, the likely performance of alternative UCL methods can be explored using simulation studies. Such studies have already been conducted by USEPA (2010b) to guide in the calculation of 95% UCLs for discrete sampling protocols. This type of performance evaluation has not been previously conducted for ISM sampling, so initial simulation studies were conducted in the development of this guidance, as summarized in Section 4.4.

In practice, the true mean is unknown, but with simulation we can define the mean. Simulation studies help guide the selection of a UCL method based on simulation-specific information, assumptions, and decision error criteria.

Three or more ISM samples are needed to calculate a 95% UCL. In theory, all of the UCL methods that are applied to discrete sampling results can also be applied to ISM. In practice, however, because fewer than eight replicate ISM samples are likely to be collected for a DU, fewer options are typically available to calculate a UCL compared with discrete sampling data. The small number of replicates precludes GOF evaluations as well as the use of methods that require more samples than typically collected in ISM sampling (USEPA 2010a). Therefore, the options for UCL calculations reduce to the set of methods that require only the parameter estimates themselves: mean and SD. Two candidate UCL equations that can accommodate ISM data sets and which are expected to "bracket" the range of UCLs that may be calculated from a data set are the Student's-*t* (representing the low end of the range) and Chebyshev (representing the high end of the range) UCLs as discussed in Section 4.2.2.1.

Two UCL calculation methods were evaluated for use with ISM samples:

- Student's *t* UCL
- Chebyshev UCL

The online version of this document contains a working calculator for these methods: http://www.itrcweb.org/ISM-1/4_2_2_UCL_Calculation_Method.html.

[4] Note that throughout this document a UCL on a mean estimate is presented as 95% UCL. It is important to note that this is only an example of a UCL. It is possible to use a 90% UCL, 98% UCL, 99% UCL, etc. The specific UCL used should be determined by the project team during systematic planning.

4.2.2.1 UCL of the mean based on Student's-t distribution

The following equation is used to calculate the one-sided (1–α) 100% UCL using the Student's-*t* approach:

$$UCL = \overline{X} + t_{(1-\alpha)(r-1)} \times \frac{S_{\overline{X}}}{\sqrt{r}}$$

where

$\overline{X}$ = arithmetic mean of all ISM samples
$S_{\overline{X}}$ = standard deviation of all ISM samples
r = number of ISM samples
t = $(1-\alpha)^{th}$ quantile of the Student's-*t* distribution with (r–1) degrees of freedom

The Student's-*t* UCL is expected to provide valid 95% UCL values when the distribution of means is approximately normal. The central limit theorem (CLT, Casella and Berger 2001) provides support for the use of a Student's-*t* UCL for composite sampling as well as ISM sampling. The CLT is useful because it defines the distribution of the mean of the samples without having to know the exact underlying distribution of the data. The number of samples, *n*, and the shape of the distribution of the data are the two factors that most influence the accuracy of the approximation of the distribution of the mean. For approximately symmetric or slightly skewed distributions, a relatively small number of samples (e.g., n = 15) may be sufficient for the estimates of the mean to be approximately normally distributed as theorized by the CLT. If the population distribution is moderately skewed, a larger number of samples (e.g., n ≥ 30) is required to reliably invoke the CLT (Casella and Berger 2001). More highly skewed distributions require even larger numbers of samples. When the distribution of replicate samples is right-skewed instead of normal, the consequence of using the Student's-*t* UCL is that it will underestimate the true mean more often than desired.

In ISM sampling, the coverage of the Student's-*t* UCL also depends on the SD of the ISM replicates. The influence of the combination of factors for different sampling regimes can be difficult to anticipate. The simulation results in Section 4.3 demonstrate various performance metrics associated with the use of the Student's-*t* distribution for a wide range of plausible scenarios.

4.2.2.2 UCL of the mean based on Chebyshev inequality

The following equation is used to calculate the one-sided (1–α) 100% UCL using the Chebyshev approach:

$$UCL = \overline{X} + \left(\sqrt{1/\alpha - 1}\right) \times \frac{S_{\overline{X}}}{\sqrt{r}}$$

where

$\overline{X}$ = arithmetic mean of all ISM samples
$S_{\overline{X}}$ = standard deviation of all ISM samples
r = number of ISM samples

The Chebyshev is generally considered to be a conservative estimate of the UCL because it generally achieves or exceeds the desired coverage rates, even for nonnormal distributions. However, with small numbers of samples, the estimates of the mean and SD can be unstable, and the coverage is not guaranteed, especially when the contaminant distribution in the DU is highly heterogeneous. Each simulation discussed in Section 4.3 includes performance metrics for both the Student's-*t* and Chebyshev UCLs to illustrate conditions in which each may be favored.

4.2.3 Nondetects

While nondetects are relatively common for discrete sampling data sets due to spatial heterogeneities in the mechanisms of release of contaminants, it is less likely that an ISM result will be below the reporting limit because some small percentage of the increments are expected to capture source areas. In other words, while individual increments may be nondetect, it is unlikely that the mean of the increments (given by an ISM result) will be nondetect. Exceptions may include constituents that are below reporting limits under ambient (background) conditions and are unrelated to site activities or post-remediation confirmation samples. In both cases, so long as the reporting limits do not approach action levels, nondetect results should not introduce decision errors. If reporting limits approach action levels, users should consider alternative analytical procedures, revisions to the sampling design to characterize the DU, and other lines of evidence that concentrations might be below action levels. If replicate results include a mix of detects and nondetects, then the only option with small sample sizes is to apply substitution methods such as one-half reporting limits to ISM results and to qualify results as potentially biased due to the use of substitution methods. A variety of substitution methods may be applied and the consequences of those options should be explored.

4.3 Evaluating the Performance of Sampling Approaches

This section describes studies used to evaluate the performance of various ISM sampling strategies in providing accurate estimates of the mean and 95% UCL. Metrics used to evaluate performance and the approach used in the simulation studies are described.

4.3.1 Definitions of Performance Metrics

Performance metrics provide a way to systematically evaluate and compare various sampling strategies, including sampling pattern, statistical sample size (both number of increments and replicates), and UCL computation techniques. Collectively, these results can help to establish an optimal decision process for using ISM given a particular set of site conditions and decision

Performance metrics for a 95% UCL that can be evaluated when the true mean is known (or assumed):

- UCL coverage
- relative percent difference between UCL and true mean
- bias
- relative standard deviation of replicate means

criteria. The following four metrics are defined below and evaluated through simulation studies:

- coverage of the UCL
- magnitude of UCL deviation from the mean (i.e., RPD between UCL and true mean)
- bias of the mean of the samples
- RSD of the estimates of the mean from each replicate ISM sample

4.3.1.1 Coverage and magnitude of UCL errors

In repeated trials, an appropriate 95% UCL should exceed or "cover" the true mean 95% of the time. In practice, we never know how well a 95% UCL has performed in terms of coverage[5] because the true mean is unknown. However, in simulation studies we have the opportunity to repeatedly evaluate a theoretical DU for which the mean is known and compute many UCLs. Accordingly, coverage is defined in this context as the percentage of the simulations for which the 95% UCL actually exceeds the true DU mean. As an example, Table 4-2 gives selected results for a simulation with 5000 trials where the mean and 95% UCL were calculated by sampling a lognormal distribution with mean = 100 and SD = 200. The 95% UCL for each trial was based on a Chebyshev equation applied to sample statistics for 3 replicates of 30 increments. The values from the UCL column are then compared to the true mean of 100. If the design were built to theoretically have 95% confidence that the true mean was less than the calculated UCL, then the ideal result from the 5000 iterations would be to find approximately 5% (i.e., 250 of 5000) of the UCL values are below the true mean. Figure 4-4 shows a histogram of the 5000 UCL values from this simulation where the y-axis represents the fraction of total iterations in each bin. In this example, the UCL histogram shows that approximately 5% of the UCL values are below the true mean. This exercise shows that the UCL coverage for this simplified scenario met the design criteria. It is interesting to note that the grand mean of 3 replicates underestimated the true mean nearly 60% of the time (in contrast to the 95% UCL underestimating the mean only about 5% of the time), exemplifying why the UCL is often used to protect against underestimation of the true mean.

For positively skewed distributions (e.g., lognormal), the mean of the ISM samples will underestimate the population mean >50% of the time, whereas the 95% UCL will typically underestimate <5% of the time.

Table 4-2. Example of UCL simulations

Trial	Mean	UCL	RPD
1	64.7	85.0	–15%
2	61.7	102.7	2.7%
3	100.7	105.2	5.2%
4	90.8	107.0	7.0%
⋮	⋮	⋮	⋮
4999	96.1	215.3	115.3%
5000	253.2	855.0	755.0%

RPD = [(UCL – 100)/100] × 100%.

[5] Note that this concept is completely separate and unrelated to that of spatial "coverage" as applied to areal representativeness of samples taken over a DU.

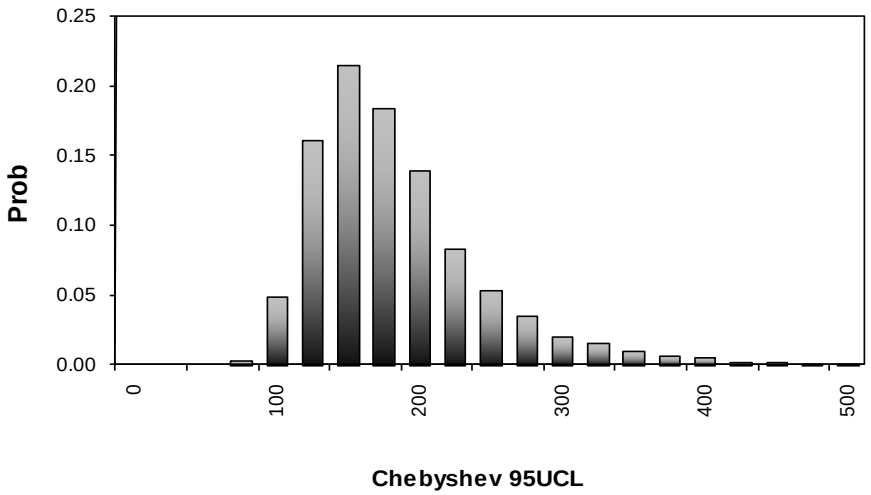


Figure 4-4. Histogram of calculated Chebyshev UCL values using 5000 trials of a lognormal distribution (mean = 100, SD = 200), 30 increments, and 3 replicates.
The "true mean" (100) is exceeded in approximately 95% of the trials.

The optimal methodology for calculating a UCL should provide adequate coverage of the mean and produce a UCL that is not unduly large. The magnitude of difference between the UCL and the true mean can be expressed as the RPD defined as follows:

$$RPD = [(95\%\ UCL - \mu)/\mu] \times 100\%$$

As shown in Table 4-2, RPD may be negative or positive depending on whether or not the UCL exceeds the true mean. RPD may be calculated for all UCL results or can be calculated for those UCLs that fall above (RPD_A) and below (RPD_B) the true site mean, separately. When used for just those UCLs that fall below the site mean, the RPDs reveal the magnitude of the potential underestimation. This calculation is particularly informative in situations where the coverage does not meet the specified 95% criteria.

Figure 4-5 illustrates examples of RPD_A and RPD_B for simulations using lognormal distributions with CV = 1 and CV = 4. Each simulation represents 5000 trials using 30 increments (m) and 2, 3, 5, or 7 replicates (r). Results for both the Chebyshev UCL and Student's-*t* UCL are given side by side. Error bars represent the 5^{th} and 95^{th} percentile RPD values, and the point in the center corresponds to the median. For example, for CV = 1 and r = 3, the Chebyshev UCL generally exceeds the true mean by less than 50% and underestimates by less than 10%. The deviation of the UCL using Student's-*t* is slightly lower for the overestimates and comparable for the underestimates. For CV = 4 and r = 3, the magnitude of the deviations increases for both the Chebyshev UCL (95^{th} percentile RPD_A of 214% and RPD_B of –23%) and Student's-*t* UCL (95^{th} percentile RPD_A of 160% and RPD_B of –25%). Information on coverage and RPD ranges can be combined to yield the following observations:

- Even for distributions with high variance (e.g., CV = 4, r = 3), the 95% UCL using either Chebyshev or Student's-*t* equations can be expected to yield values that exceed the true mean by no more than 150%–200% and underestimate by less than 25%;
- Student's-*t* UCL more frequently underestimates the true mean than does the Chebyshev UCL.
- The magnitude of the underestimate (RPD_B) will be comparable; however, the magnitude of the overestimate (RPD_A) will be greater for the Chebyshev UCL.

The Student's-*t* UCL and Chebyshev UCL provide estimates of the mean that, even for highly variable distributions, generally exceed the true mean by no more than 200% or underestimate the mean by no more than 25%.

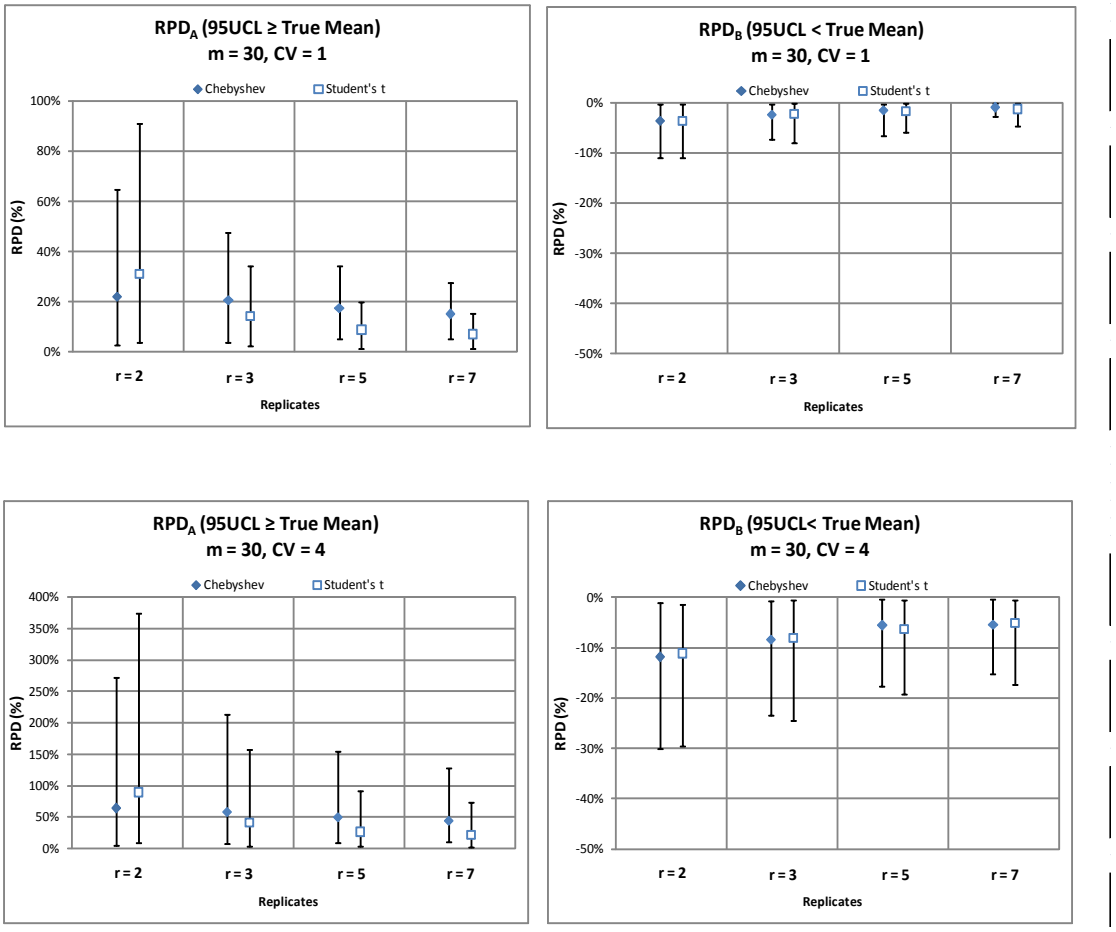


Figure 4-5. Range of overestimation (RPD_A) and underestimation (RPD_B) of 95% UCLs using Chebyshev and Student's-*t* calculation methods for ISM simulations with lognormal distributions (CV = 1 and CV = 4), 30 increments, and 2–7 replicates. Error bars represent 5th and 95th percentiles of 5000 trials.

ISM replicates tend to produce UCLs with smaller RPD_A and RPD_B than a corresponding data set of discrete samples. This desirable quality of ISM is due to the physical averaging of the

It is unlikely that one 95% UCL method excels at all performance metrics. In addition, performance can vary depending on site characteristics. Method selection requires balancing the importance of each metric.

individual increments. Therefore, ISM UCL values may provide reasonably reliable estimates of the site mean even when the desired 95% coverage is not achieved but RPD_B is minimal.

In general, all other conditions being the same, as the number of increments and replicates increases, the error is expected to decrease. This decrease in the standard error will be reflected by an improvement in bias, the coverage and RPD of the UCL. The influence of these components of the sampling design varies depending on characteristics of the population sampled (e.g., magnitude of DH, single or multiple populations) and the sampling method (e.g., systematic random sampling, random sampling with grid, or simple random sampling). The central concept governing the optimization of the sampling design is that while initial increases in the number of replicates and increments improve estimation, there are diminishing returns with increasing numbers of samples. At some point, increasing the number of samples is unlikely to yield an appreciable improvement in either the coverage of the UCL or the magnitude of the over/underestimate of the UCL as indicated by the RPD calculations.

4.3.1.2 Bias in estimated mean

"Bias" is defined here as a systematic over- or underestimation of the true site mean. Bias is generally introduced when the sampling strategy or sample collection method yields observations that are not truly independent or representative of site conditions. For example, use of a systematic sampling pattern that coincides with spatial trends in the data may produce a data set that disproportionately represents a range of concentrations; poor sample collection techniques may underrepresent actual soil characteristics.

> Accuracy reflects a combination of precision (reproducibility) and bias (systematic over/underestimation). The RSD of replicate ISM means is a measure of precision.

4.3.1.3 Relative standard deviation of replicate samples

The reproducibility of ISM replicates collected from a DU can be evaluated in terms of RSD, also known as the coefficient of variation (CV), which is the SD divided by the mean. Because both RSD and CV are commonly used, both are used interchangeably here. Although included as a performance metric in the simulation studies, the RSD does not provide an indication of the accuracy of the estimate of the mean or 95% UCL. Figure 4-6 illustrates the distinction between reproducibility (or precision) vs. bias, which, taken together, represent accuracy. For example, a low RSD indicates the estimates are precise. The values might be reproducible but still yield a biased estimate

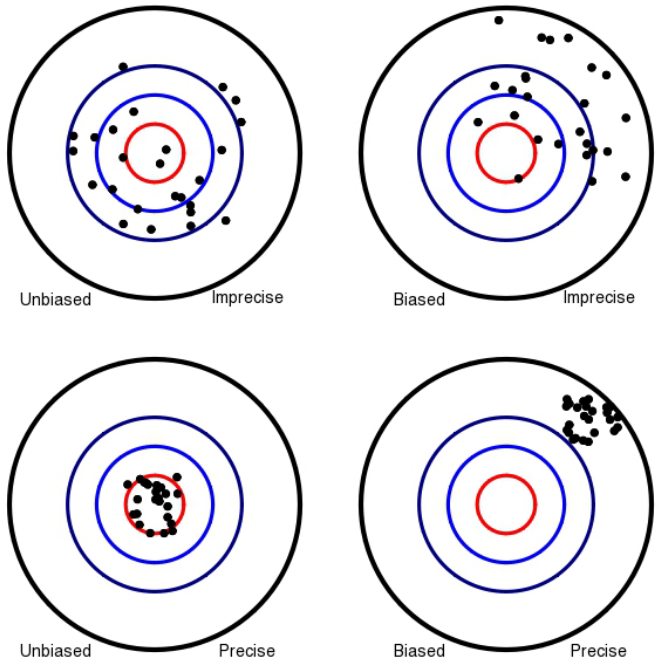


Figure 4-6. Four possible relationships between bias and precision.

of the mean and corresponding UCL.

4.3.2 Simulation Study Approach

A computer-based simulation is a numerical experiment in which a hypothetical site is sampled many times. The key utility of simulations is that the contaminant distribution can be specified so the population parameters are known. This is in contrast to actual sampling with ISM, in which the potential bias in results or coverage of the 95% UCL cannot be quantified since the true mean is not known. Simulation is a convenient tool to evaluate the performance of alternative ISM approaches based on comparisons to the true mean. Furthermore, a variety of different incremental sampling and statistical methods can be simultaneously applied to the same exact scenario to facilitate a comparison of sampling strategies. Each simulation followed this general five-step process:

Simulation studies were conducted to determine the performance of several aspects of ISM:

- number of increments and replicates
- sampling pattern
- heterogeneity and variability of concentrations

1 Define the population. This may be a probability distribution (e.g., lognormal or Gamma), a mixture of probability distributions, or a 2-D map of the concentration surface of a DU. For some scenarios, CH and DH may be explicitly defined, while for others the assumption is that the population variance represents the combination of both elements of heterogeneity and other sources of error.

2 Define an ISM sampling strategy. This step identifies the size and placement of DU, number of increments, sampling strategy (e.g., systematic random, random sampling within grid, simple random sampling; see Section 4.3.4.2 for more description), and number of replicates.

3 Implement a Monte Carlo analysis (MCA). Using MCA (described below), repeat the same ISM sampling strategy many times (e.g., 2000 iterations or more).

4 Calculate statistics. For each iteration of MCA, calculate the DU statistics, including the grand mean (i.e., mean of replicate samples), RSD of the replicate samples, bias in mean (i.e., estimated mean minus population mean), and 95% UCL using Student's-*t* UCL and Chebyshev UCL.

5 Evaluate performance metrics. In this step, the statistics are used to evaluate performance metrics, including coverage of 95% UCL, magnitude of UCL error, bias of the means, and RSD.

Using simulation, we can evaluate a variety of different statistical properties of ISM and determine if factors that can be controlled in the sampling design (e.g., number of increments, number of replicates, DU size, and use of multiple SUs) can be adjusted to achieve the sampling objectives. Furthermore, by running MCA simulations on a variety of different scenarios, we can develop an understanding of the alternative ISM sampling strategies under different conditions. For example, 30 increments and 3 replicates may be sufficient to obtain a reliable 95% UCL for a DU that is described well by a single probability distribution with relatively low DH, whereas

greater numbers of samples may be needed for a DU with multiple overlapping contamination sources and relatively high CH and DH. Pitard (1993) highlights the value of summarizing such relationships with sampling nomographs, which are the "best available tool to quickly estimate sampling errors, effectively design optimum sampling procedures, and find reasonable economical compromises."

Simulations can be used to determine the performance of ISM under very specific conditions and, therefore, the results cannot be expected to apply to all sites. Table 4-3 provides details regarding the range of conditions that have been investigated and summarized in this document.

Table 4-3. Summary of scenarios investigated with simulations

Condition	Levels
Increments	15–100
Replicates	2–5
Sampling method	Simple random sampling, random within grid, and systematic random
Sampling pattern	Entire DU and subdivided DU
Range of symmetry and dispersion	Normal data and multiple skewed data sets (lognormal and Gamma) with CV ranging 0.7–6
DU variability	Homogenous and multiple levels of heterogeneity
DU spatial patterns	Ranged from evenly distributed to localized elevated regions of differing sizes

A comprehensive review of the performance of discrete sampling methods for 95% UCL calculations already exists (USEPA 2010b) and was not evaluated here.

4.3.3 Objectives of the Simulation Studies

The objective of the simulation studies was to address several issues of practical importance in obtaining and using ISM data. As noted above, simulation studies have the unique advantage of evaluating the performance of ISM in estimating the mean under a variety of conditions where the right answer (i.e., the true mean) is known. Thus, they are the best, and in fact the only, way that the accuracy of ISM estimates of the mean concentration can be assessed.

Some of the simulation studies were directed to the basic design of an ISM sampling event (the number of increments) and the pattern in which the samples are taken within a DU. The accuracy of ISM estimates and 95% UCL coverage based on differing numbers of increments, replicates, and sampling patterns were evaluated with attention to bias and magnitude of error (i.e., RPD). Simulation studies evaluated different approaches for computing a 95% UCL using ISM data. Performance of sampling methods was evaluated in terms of coverage provided by a 95% UCL, as well as the extent of overestimation of the mean (RPD_A). Ideally, a calculation method yields a 95% UCL with adequate coverage without excessive overestimation of the mean.

Simulations conducted with hypothesized distributions (e.g., lognormal) did not attempt to distinguish between different sources of error (see Section 2.5) or real-site complexities associated with spatial patterns such as mixtures (i.e., multiple sources of contamination at a site)

or hot spots (i.e., sources with elevated concentrations that occur in relatively small subareas across the DU). By sampling from a single lognormal distribution, the simulations do not explicitly address inherent heterogeneities (e.g., CH and both small- and large-scale DH). However, these simulations are particularly applicable for scenarios in which the contamination is expected to be homogeneous through the DU (meaning that the mean and variance are the same in all subareas) and simple random sampling is applied. These simulations provide a convenient framework to begin to evaluate the performance of different UCL calculation methods with different sampling designs (i.e., numbers of increments and replicates) under a range of skewness and variance in the distribution. The simulations with maps extend the evaluation by exploring the effect of sampling methods (e.g., systematic or simple random) on bias in parameter estimates as well as the effect of DU heterogeneity on the performance of the 95% UCL.

4.3.4 Simulation Study Findings on ISM Performance

The following sections summarize conclusions from the simulation studies. Where possible, results are expressed in terms of the performance metrics outlined in the previous section.

4.3.4.1 Sample size (number of increments and replicates)

One option for reducing errors in sampling designs is to increase the sample size. For ISM, sample size can pertain to the mass per increment (i.e., sample support), number of increments (*n*), and number of replicates (*r*). Assuming a uniform mass per increment, several observations were made regarding the effects of increasing *n* and *r* on estimates of the mean (also see Appendix A, Table A-1):

- Increasing the *n* has a direct effect on the standard deviation of the replicates. Specifically, the central limit theorem suggests the standard deviation of the replicates (which is a measure of the standard error of the mean) reduces by a factor of the square root of *n*. For example, all other things being equal, if the SD of replicates is 4.0 with n = 30, doubling the increments to n = 60 would reduce SD by the square root of 2 (or 1.414) to approximately 2.8.

- Increasing *r* does not reduce the standard deviation of the replicates although it does improve the estimate of the SD by reducing the variability in the estimate. Likewise, increasing *r* reduces the standard error for the grand mean. Specifically, the standard error reduces by the square root of *r*.

- The overall reduction in the standard error for the (grand) mean is a function of the *total* mass collected and spatial area represented (i.e., increments × replicates), and this observation applies to parameter estimation with discrete sampling as well.

- Increasing the number of increments (n) or sample mass reduces the potential for errors in terms of both frequency and magnitude of underestimation of the mean.

Increasing the number of increments and/or replicates reduces the variability in ISM estimates of the mean.

- For nonnormal distributions, increasing *r* above 3 provides marginal return in terms of improving coverage of a UCL when the Chebyshev calculation method is used; however, increasing *r* does not improve coverage of the Student's-*t* UCL.

The difference between Chebyshev and Student's-*t* UCLs can sometimes lead to different decisions for a DU. While the Chebyshev method typically provides greater coverage, it also tends to have higher RPDs. Project teams must balance both properties of UCLs when deciding which method(s) to use.

- Increasing *r* reduces (i.e., improves) the RPD, meaning it will produce estimates of the 95% UCL closer to the DU mean. Therefore, increasing *r* may be an important sampling strategy when errors of either underestimation or overestimation of the mean can have significant consequences.

- Simulations produced varying results in terms of improvement in coverage by increasing the number of increments (n). In some simulations, increasing n produced little or no observable difference. In others, n twofold or more from typical increment numbers used in ISM resulted in marginal improvement. As with increasing replicates, increasing n decreases (i.e., improves) the RPD. The improvement in RPD performance is marginal when the underlying CV is small.

- Simulations showed that coverage provided by the two UCL calculation methods depends upon the degree of variance (or dispersion) of the contaminant distribution within the DU. A variety of statistics provide a measure of dispersion including the CV (i.e., SD normalized by the mean) and the geometric SD (specific to lognormal distributions). Table 4-4 summarizes findings grouped by CV (and GSD). Note that in this case, the CV reflects the SD of the increments divided by the mean and not the SD of the replicates divided by the mean. In practice, individual increments are typically not retained for analysis, so there may be no direct measure of the CV. If there is no site knowledge available to support an assumption about the degree of dispersion (i.e., low, medium, high) of increments, then the Chebyshev UCL may be the preferred calculation method because it is more likely to achieve the desired coverage than the Student's-*t* UCL. The CV (or SD) of the replicates is not a useful metric for determining which UCL method provides sufficient coverage.

Table 4-4. Likelihood that ISM achieves coverage depending on dispersion

UCL Method	Dispersion among individual increments		
	Low (CV <1.5 or GSD <3)	**Medium (1.5 < CV < 3 or 3 < GSD < 4.5)**	**High (CV >3 or GSD >4.5)**
Student's-*t*	Yes	No	No
Chebyshev	Yes	Yes	Maybe

Coefficient of variation (CV) = standard deviation (SD)/mean.
Geometric standard deviation (GSD) = exp[sqrt(ln(CV^2 + 1))] for lognormal distributions.

- The Chebyshev method always produces a higher 95% UCL than the Student's-*t* method for a given set of ISM data with r > 2. When both methods produce specified coverage, the Chebyshev consistently yields a higher RPD.

4.3.4.2 Effects of sampling pattern

Just as with discrete sampling, a variety of sampling methods may be implemented with ISM sampling. One of the more common approaches in ISM is systematic random sampling (a.k.a., systematic grid sampling [Gilbert 1987]), where the DU is divided in a grid pattern, a random sampling location is identified within the first grid cell, and then samples (increments) are obtained from adjacent cells sequentially in a serpentine pattern using the same relative location within each cell (Figure 4-7). Another approach is random sampling within a grid (also called "stratified random sampling" [USEPA 1995b]), wherein samples are obtained sequentially from adjacent grid cells, but the location of the sample within each cell is random (Figure 4-8). A third approach is simple random sampling, where the samples are taken from random locations across the DU (without gridding) (Figure 4-9). Replicate ISM samples are collected with the same sampling method but not the same exact locations. Each sampling method has its strengths and weaknesses that should be considered when selecting the approach for a given site.

Simple random sampling, systematic random sampling, and systematic grid sampling yield unbiased estimates of the mean. The systematic sampling patterns ensure relatively even spatial distribution of samples across the site and are generally easier to implement in the field.

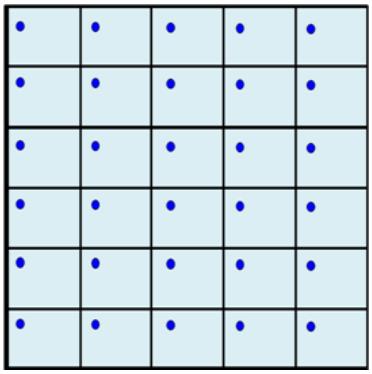


Figure 4-7. Systematic random sampling/ systematic grid sampling with a random start (Serpentine).

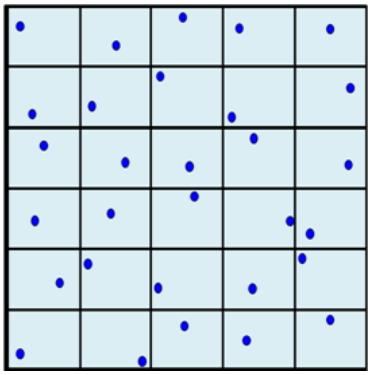


Figure 4-8. Random sampling within grids.

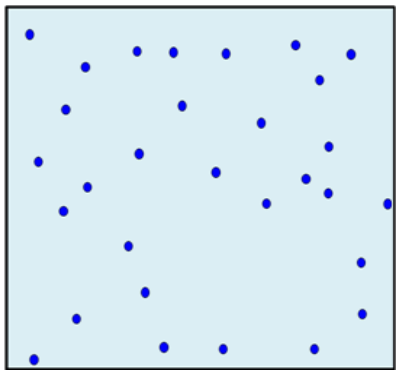


Figure 4-9. Simple random sampling within the entire DU.

- If the site is relatively homogeneous, all three sampling patterns yield unbiased parameter estimates, but the magnitude of error in the mean may be higher with simple random sampling as compared with systematic random sampling. All three sampling patterns yield equivalent coverages.

- 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," meaning without any increment sample locations. It should be noted

that "random" does not mean wherever the sampling team feels like taking a sample: a formal approach to determining the random sample locations must be used.

- Systematic random sampling can avoid the appearance that areas are not adequately represented in the ISM samples. This approach is relatively straightforward to implement in the field. Theoretically, it is inferior to simple random sampling for obtaining unbiased estimates of the mean, especially if the contamination is distributed systematically so that areas of high or low concentrations are oversampled with the systematic design. Random sampling within a grid is in a sense a compromise approach, with elements of both simple random and systematic sampling.

4.3.4.3 Partitioning the DU

Partitioning the DU into multiple SUs is one way to characterize variability on a smaller spatial scale. This can be useful for both exposure assessment (e.g., assessing risks to multiple receptors with different sized exposure units) and remedial design (e.g., delineation of remediation units smaller than a DU).

When taken over the entire DU, replicates offer information on variability in the estimate of the mean provided by the ISM samples. They do not, however, provide any information on spatial variability of concentrations within the DU. Another approach is to divide the DU into multiple SUs and take one or more ISM samples from each. With this approach, ISM samples are not true replicates in that they are providing estimates of the mean for different subunits within the DU. Individually, they estimate the mean of a subarea, and collectively, they can be used to estimate the mean of the entire DU. Sampling designs with this method yield unbiased estimates of the mean.

- 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 (e.g., 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. This information can guide redesignation of the DU and resampling to further delineate areas of elevated concentrations.

- If only one ISM sample is collected per SU, then it is important to understand that each result independently provides an estimate of the mean concentration within the respective SU. Just as a single ISM collected throughout the DU may over- or underestimate the mean by some magnitude (see Section 4.2.1), the information on heterogeneity at the scale of the SU is also subject to uncertainty. If greater certainty is needed at the scale of the SU, then additional increments and/or replicates should be collected at the scale of the SU.

- Collectively, the results from each SU can be used to estimate the mean and 95% UCL at the scale of the DU.

- Error estimates from partitioning a DU into SUs are larger than those from replicate data if the site is not homogeneous. Hence, 95% UCL estimates from a subdivided DU are as high as or higher than those obtained from replicate measurements collected across the DU (using

the same number of total increments). The higher 95% UCLs improve coverage (generally attain 95% UCL) and increase the RPD_A. These increases occur if unknown spatial contaminant patterns are correlated with the partitions.

- It must be clearly understood by all that if the 95% UCL for the DU is below the action level, the entire DU passes, even if the ISM result for one or more of the partitioned areas is above the action level. Even with partitioning, the DU remains the unit over which a decision is made.

Note: "Row-column" is an additional sampling pattern proposed by Patil and Tallie (2001). This sampling pattern has not been widely discussed in the context of ISM and consequently was not explored in the simulation studies. However, this approach is discussed in the composite sampling literature and has the potential advantage of providing spatial information on localized areas of high concentration (see "oversized DUs" in Section 4.4.4).

4.3.4.4 Relative standard deviation

The RSD was calculated from the set of simulated ISM results for each iteration of the Monte Carlo simulation. Collectively, patterns in 95% UCL coverage and other performance metrics can be evaluated for different ranges of RSD. The following were noted:

- Data sets with a high RSD are more likely to achieve specified coverage for 95% UCL of the (population) mean than data sets with low RSD. This effect is explained by the greater variability among replicates leading to higher 95% UCL values, resulting in better coverage.

- A low RSD may intuitively appear to ensure specified coverage by the 95% UCL or low bias in a single estimate of the mean. However, the opposite is in fact the case when the underlying distribution is positively skewed (e.g., lognormal, gamma). For situations in which the UCL or one replicate mean is less than the true mean, the underestimate increases as RSD decreases. This phenomenon reflects the "proportionality effect," whereby the mean and variance are expected to be positively correlated for positively skewed distributions (Goovaerts 1997). Therefore, when the mean is relatively low, so too is the SD. Taken together, there is a greater likelihood that the UCL exhibits insufficient coverage.

4.4 Areas for Further Study

Other potential uses of ISM samples not directly included in the simulation studies are included below. Some considerations are discussed, but additional simulations studies in the future might add further clarification and recommendations for these situations.

ISM is a recent addition to environmental sampling strategies, and there is still much to be learned about it. Areas with recognized information gaps include the following:

- combining information from multiple DUs after sampling
- extrapolating the information from one DU to another
- sampling very large DUs
- comparison of results from multiple sites

4.4.1 Combining DUs

On occasion, there might be a desire to combine information from multiple DUs into a single, larger area. There are two primary explanations for when 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 exposure unit. Each of those areas might be investigated as a separate DU for site characterization but then combined to define a single exposure unit.

- For ecological and human health risk assessment, we might need to consider a variety of sizes of DUs to accommodate multiple receptor scenarios. For example, if 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 these considerations are incorporated in the initial planning stages, they can be addressed by using a stratified sampling design. Within each strata, it may be appropriate to use ISM, but then one encounters the challenge of combining the ISM data from the strata into the larger DU. Conversely, this issue may also arise when ISM data are collected from multiple DUs and combined to estimate the mean in a single, larger DU. Whether preplanned or not, the same treatment of the data is appropriate.

When there are multiple samples in each stratum, the overall mean of the larger DU can be estimated using the following formulae. Let n_i represent the number of samples from region i, $x\text{-}bar_i$ represent the mean of the ISM samples from region i, s_i represent the SD of the replicate ISM samples from region i, and w_i represent the weight, i.e., the relative size associated with region i. Note that if all strata are of the same size, the w_i are equal, and these equations simplify to the more common calculation methods for the mean and standard deviation. The relative size is the percentage of the larger DU that is made up of region i. The weighted mean is thus:

$$\text{Weighted Mean} = \sum_i w_i \bar{x}_i$$

The standard error associated with the weighted mean is:

$$\text{Standard Error} = \sqrt{\sum_i w_i^2 \frac{s_i^2}{n_i}}$$

which has degrees of freedom approximated by the Welch-Satterthwaite approximation (Cochran, 1977):

$$df \approx \frac{\left(\sum_i \frac{w_i^2}{n_i} s_i^2\right)^2}{\sum_i \frac{\left(\frac{w_i^2}{n_i} s_i^2\right)^2}{n_i - 1}}$$

Table 4-5 provides a numerical example of this calculation, where data from two DUs are combined to derive a 95% UCL for a larger DU. 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 shown in Table 4-5, based on n = 30 increments per replicate:

Table 4-5. Summary statistics used to combine DUs

Playground area	Area (acres)	Sample statistics			95% UCL	
		Replicates	Mean	SD[a]	Student's-*t*	Chebychev
DU 1 (kindergarten)	0.25	25, 100, 140	88.3	58.4	187	235
DU 2 (older child)	0.50	5, 25, 305	111.7	167.7	394	534
Equal weight	0.75	25, 100, 140, 5, 25, 305	100	113	193	301

[a] SD = standard deviation.

The 95% UCLs for each DU are given for both the Student's-*t* and Chebyshev methods. Section 4.3.4 provides a discussion of different performance metrics for the UCL that can be used to determine which UCL method may be more likely to achieve the study objectives. Because the true mean for each DU is unknown, the RPD between the UCL and mean cannot be calculated. Figure 4-5 provides examples of the 5th, 50th, and 95th percentile RPDs of UCLs calculated with r = 3 replicates for lognormal distributions with CVs of 1 and 4 when the UCL exceeds the true mean. Recall that the CV in this context refers to the dispersion of the underlying distribution (e.g., distributions given by individual increments), not the distribution of means given by the ISM results. The mean of the ISM replicates can be assumed to approximate the mean of the underlying distribution, and the SD of the replicates can be assumed to approximate the standard error of the mean of the underlying distribution: $SE = SD/\sqrt{n}$. We can rearrange to solve for SD: $SD = SE \times \sqrt{n}$. So for n = 30, we can estimate SD of the underlying distribution by multiplying the SD of the ISM results by $\sqrt{30} = 5.5$. Therefore, the following are estimates of the SD and corresponding CV of the underlying distributions for each DU and the combination of DUs:

- CV of DU1 = SD/mean = (58.4 × 5.5)/88.3 = 3.6
- CV of DU2 = SD/mean = (167.7 × 5.5)/111.7 = 8.3
- CV of DU1 + DU2 (equally weighted) = SD/mean = (113 × 5.5)/100 = 6.2

For r = 3 replicates and CV = 4, Figure 4-5 suggests that the median RPD for both UCL methods is 90% and the 95th percentile is about 200% for Chebyshev and 150% for Student's-*t*. The magnitude of the RPDs is expected to be even more pronounced for CV = 8.

As summarized in Table 4-4, the coverage of the UCLs also depends on the CV of the underlying distribution. Both DUs appear to have high CVs (i.e., >3), and the Student's-*t* UCL is not expected to yield a coverage close to 95%, even if the number of replicates were increased. Therefore, Chebyshev UCL is expected to yield more reliable results (based on coverage).

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 4-5. Alternatively, it may be assumed that a maintenance worker's exposure is proportional to the respective areas of each DU and the equations from Section 4.4.1 can be used to generate summary statistics for the combined area (0.75 acres). The weighting factors applied to each DU should sum to 1.0, which is achieved by dividing each area by the sum of the two areas:

- $w_1 = 0.25/0.75 = 0.33$
- $w_2 = 0.50/0.75 = 0.66$

Weighted mean = $\sum_i w_i \bar{x}_i = (0.33 \times 88.3) + (0.66 \times 111.7) = 103.9$

Standard error for mean (SE) = $\sqrt{\sum_i w_i^2 \frac{s_i^2}{n_i}} = \sqrt{(0.33)^2 \times \frac{(58.4)^2}{3} + (0.66)^2 \times \frac{(167.7)^2}{3}} = 65.5$

Degrees of freedom (df) = $\frac{\left(\sum_i w_i^2 \frac{s_i^2}{n_i}\right)^2}{\sum_i \frac{\left(w_i^2 \frac{s_i^2}{n_i}\right)^2}{n_i - 1}} = \frac{\left((0.33)^2 \times \frac{(58.4)^2}{3} + (0.66)^2 \times \frac{(167.7)^2}{3}\right)^2}{\frac{\left((0.33)^2 \times \frac{(58.4)^2}{3}\right)^2}{3-1} + \frac{\left((0.66)^2 \times \frac{(167.7)^2}{3}\right)^2}{3-1}} = 2.1$

Student's-*t* 95% UCL = $\bar{X} + t_{(1-\alpha)(df)} \times SE = 103.9 + 2.92 \times 65.5 = 295$

Chebyshev 95% UCL = $\bar{X} + \left(\sqrt{\frac{1}{\alpha} - 1}\right) \times SE = 103.9 + 4.36 \times 65.5 = 390$

The online version of this document contains a working calculator for the Weighted 95% UCL for a Combined DU from Several Smaller DUs:
http://www.itrcweb.org/ISM-1/4_4_1_Combining_DUs.html

This same methodology could be used to combine a surface DU with its corresponding subsurface DU. The only slight difference would be that the weight term, w_i, would reflect the proportion of the total soil volume within the DU.

There are other considerations for combining DUs that may benefit from further study:

- a single ISM result in one of the DUs so that a SD cannot be calculated for that region
- the impact of very different numbers of increments in the DUs
- the impact of very different numbers of replicates in the DUs

4.4.2 Extrapolating from DUs

As discussed in Section 4.2, the motivation for collecting replicate ISM samples within a DU is to obtain an estimate of the variance in the mean, from which a UCL can be calculated. When a site includes many DUs, it may be tempting to extrapolate the estimate of the variance (or the CV) from one DU to another. However, we must first consider the extent to which the distributions may be comparable across DUs. Two related questions about the distribution should be considered:

- *Identically distributed*: Does our knowledge of potential sources suggest that similar contaminant distributions can be expected at the spatial scales represented by each DU? In effect, we would like to be able to assume that the distributions are approximately the same.

- *Normally distributed*: Estimates of the means and SDs will vary by random chance across DUs even if the distributions are the same and the same number of increments are used. Is it preferable to extrapolate estimates of the standard SD or CV?

Both questions require that we understand factors that might influence the relationship between the mean and SD of ISM replicate results within a given DU. Statistical theory suggests that we can expect the estimated mean and SD to be independent for normal distributions but positively correlated for positively skewed distributions (Goovaerts 1997). If the ISM mean and variance estimates are independent, this notion presents a challenge because we would have no reason to assume that the ratio of the SD to the mean (as represented by the CV) is the same. DUs with relatively high estimated means may have low SDs and vice versa. Instead of extrapolating the average CV across DUs, we would introduce less uncertainty by extrapolating the average SD. By contrast, if the parameters are correlated because of some asymmetry in the distribution of mean concentrations (despite the CLT, as described in Section 4.2), then it would be preferable to extrapolate the average CV. A priori knowledge about the distribution shape is unlikely, and this source of uncertainty cannot be fully addressed through simulation studies. Therefore, one must be very cautious in how information is extrapolated between DUs and how an extrapolation may ultimately introduce decision errors.

4.4.3 Comparing DUs

Simulation studies and case studies should be conducted to elucidate the advantages and disadvantages and the practical constraints for comparing DUs where some or all have ISM

samples. Without such studies, recommendations on implementation of comparisons are not possible, but there are some general considerations that are clear without the aid of simulations.

4.4.3.1 Site-to-site comparisons

ISM data from one site can be compared to that from another. For example, sampling for two DUs may consist of 30 increments and 5 replicates each. Standard two-sample hypothesis tests (e.g., Wilcoxon Rank Sum or Gehan) can be applied to determine whether the differences are statistically significant under the assumption that the variances are the same. While similar in concept to distribution testing with discrete sampling data, some aspects of ISM data comparisons are unique. First, while it is not necessary for the DUs to have the same number of ISM replicates, it is likely that the number of replicates will be quite small (e.g., 3–5). Therefore, the decision errors may be higher with hypothesis testing using ISM data compared to discrete data. In addition, since the estimated SD of ISM replicate results is a function of the number of increments obtained from the DU, samples can be compared directly only if the same number of increments is collected. To conduct a hypothesis test for ISM data based on unequal increments, a statistical adjustment to the estimated SD may be appropriate to reduce the chance of violating the hypothesis test assumption of equal variance (see Appendix A).

4.4.3.2 Incremental to discrete sample comparisons

Occasionally, it may be desirable to consider comparing or combining discrete data and ISM data. Conceptually, this can only be done when specific conditions are met:

- The design for selecting the discrete samples is known (i.e., simple random sampling, adaptive cluster sampling, etc.), and the discrete sample set is representative of the entire DU (i.e., the sampling design was statistically based and not biased).
- The samples have been collected using the same collection method or methods similar enough to ensure equivalent particle size distributions between types of samples.
- The samples are representative of the same soil conditions (e.g., soil type, depth).
- The samples have been processed in a laboratory using the same sample preparation method or methods similar enough to ensure equivalent digestion and extraction of contaminants from the sample matrix for analysis.
- The samples have been analyzed in a laboratory using the same analytical method or methods similar enough to ensure equivalent analytic results.
- The quality of both data sets is understood (via data validation reports) such that it is known that the data are appropriate for the intended use.

One must be very cautious in how information is compared or combined between DUs since it is likely that one or more of these conditions will be violated to some degree, and in practice, there are no established methods for combining discrete and ISM data.

4.4.3.3 Site-to-background comparisons

A common element of most site investigations is the comparison of the contaminant distribution in

For background screening approaches, summary statistics from discrete sample results (representing individual site measurements) are not directly comparable to summary statistics from ISM sample results, which represent mean estimates.

volumes of soil collected from a DU to the distribution in soil collected from a suitable background or reference area. In some cases, regional background values that represent upper-bound estimates may have been derived or endorsed by regulatory entities. Since these types of background values are derived from discrete samples, their information content is sufficiently different from that of an ISM sample to preclude a direct comparison of summary statistics. For example, a set of discrete sample results provides a measure of the distribution of concentrations in relatively small volumes of soil throughout the DU, whereas a set of ISM samples provides measure of the distribution of mean concentrations, each of which is an estimate of the population mean for the entire DU. Therefore, the SDs estimated from the samples represent very different properties of the contaminant distribution. Regional background levels are typically based on an upper-bound statistic, such as an upper percentile or an upper tolerance limit (UTL, i.e., a UCL for a percentile). The objective is to establish a threshold for point-by-point comparisons to each individual (discrete) site result. If no site result exceeds the threshold, one can be reasonably confident that the distribution is not elevated with respect to background. Similar to the discussion of comparisons with numerical *action level*s, it may not be possible to satisfy decision objectives with ISM when a numerical threshold is intended for comparison to discrete observations (i.e., maximum concentrations in small volumes) rather than estimates of average concentrations. Discrete and ISM data sets have different characteristics, and statistical procedures for comparing DU ISM data with discrete background data, and vice versa, have not been well established.

An alternative background screening approach is to use hypothesis testing to compare the distributions, rather than screening against an upper-bound statistic. This alternative is often used because it is well established that there is a high likelihood with point-by-point screening that one or more site exceedances will be observed by random chance even if the distributions are exactly the same. Furthermore, the error rate increases with increasing numbers of samples for the site. The hypothesis testing approach allows for localized exceedances so long as the difference in the means (or upper tails) is not statistically significant.

For this document, comprehensive simulation studies were not conducted to evaluate the statistical performance of background comparison tests for ISM results (i.e., small number of samples, moderate asymmetry). Since tests are robust to moderate violations of assumptions of normality and equal variance, the fact that formal distribution testing cannot be conducted (see Section 4.1.1) is not expected to be a major limitation for background screening with ISM data. Instead, the two key challenges for ISM are achieving the desired statistical power of the tests (i.e., likelihood of detecting differences in the populations that exist) due to small number of samples and the inability to evaluate upper tails of the underlying distributions. Section 7.2.4 provides a detailed discussion of the assumptions associated with different hypothesis tests, highlighting why results of statistical tests can be misleading when the background and site data sets have fewer than five observations each. In addition, decision errors may be affected if the samples are collected with different sampling designs, including different number of increments/replicates, different sample masses, DU volume, sampling protocols, depth intervals, and sampling patterns. Therefore, the results of hypothesis tests applied to ISM data sets should be interpreted with

For background comparisons, graphical evaluations are preferred over formal statistical tests (e.g., hypothesis tests) because the performance of hypothesis tests has not been evaluated for small sample sizes (number of replicates) expected with most ISM sampling designs.

caution until these limitations can be more thoroughly studied. If formal statistical tests are not used, simple graphical analysis (e.g., dot plots grouping ISM results by study area) may be informative as a semi-quantitative method for comparing background and site distributions.

Comparison of site ISM data to background discrete data using either hypothesis testing or UTLs is not recommended because the variance is represented differently in ISM and discrete sampling. Comparison of an ISM estimate of the mean to a discrete sample collected from soil representing background is likely to lead to decision errors in which one incorrectly concludes that the contaminant distribution on site is consistent with background conditions.

4.4.4 Oversized DUs

Generally, DUs should be no larger than the exposure units used for risk assessment if risk assessment is likely to be needed for the site. However, this limit may be impractical under some circumstances. Examples might include an acute exposure scenario (e.g., single soil ingestion event for a small child) or ecological risk assessment for a species with a very small home range. In these situations, DUs are by necessity oversized, and the average concentration for the DU provided by ISM offers only a crude approximation at best of concentrations that might exist for individual exposure units within the DU. Extrapolations of estimates of dispersion (e.g., SD or CV) across DUs to calculate a 95% UCL or other upper-bound statistic should be performed with caution, as discussed above (see Section 4.4.2).

Another approach is to use estimates of possible upper-end concentrations within a DU to evaluate potential "worst-case" situations, but the information to derive these estimates is limited due to the nature of ISM. This is not a new issue, and an analogous problem exists for composite samples. The literature for composite sampling contains a number of approaches for estimating high-end concentrations within the sampled area. The simplest of these is to multiply the mean value from the composite (or ISM sample) by the number of increments. This method represents the situation in which all of the contaminant is present in one of the increments. Given the number of increments in a standard ISM design, this approach is extraordinarily conservative and can yield quite high values. Other approaches that are less conservative include multiplying the average concentration by the square root of the number of increments or more complicated formulas (Barnett and Bown 2002). It would be advantageous to explore approaches to "decomposite" data in the context of ISM for situations in which the upper end of the concentration range within a DU is an important component of meeting site DQOs.

Note: A computationally equivalent approach is to use the average concentration but divide the soil criterion by the number of increments.

4.4.5 Explicitly Address Additional Sources of Error

Other sources of error, such as blending error and segregation and grouping error could be addressed through more complex simulation studies in the future. The impact of grinding on both the measurable concentration of metals in soil and on bioavailability also merits further study.

5. FIELD IMPLEMENTATION, SAMPLE COLLECTION, AND PROCESSING

5.1 Introduction

Section 2 discussed some of the common sources of sampling error. To obtain representative field samples, sampling error must be limited or managed (Ramsey and Hewitt 2005). In the absence of error, a sample result by definition would be accurate. However, it is impossible to completely eliminate error and produce an accurate result unless the soil in the entire DU is included in the analytical determination, which is obviously impractical. Thus, limiting sampling error is a critical function of any sampling design and implementation. This section addresses those field practices that limit or manage sampling error and provides guidance for obtaining representative samples. It should be noted that for many types of contaminants (e.g., metals, VOCs, semivolatile organic compounds [SVOCs]) specific studies have not been conducted to evaluate the applicability of all of the approaches discussed in this section and in Section 6.

To help ensure data quality, all field sampling and field processing activities should be performed and supervised by personnel trained in ISM. Figure 5-1 is a flowchart for ISM field implementation.

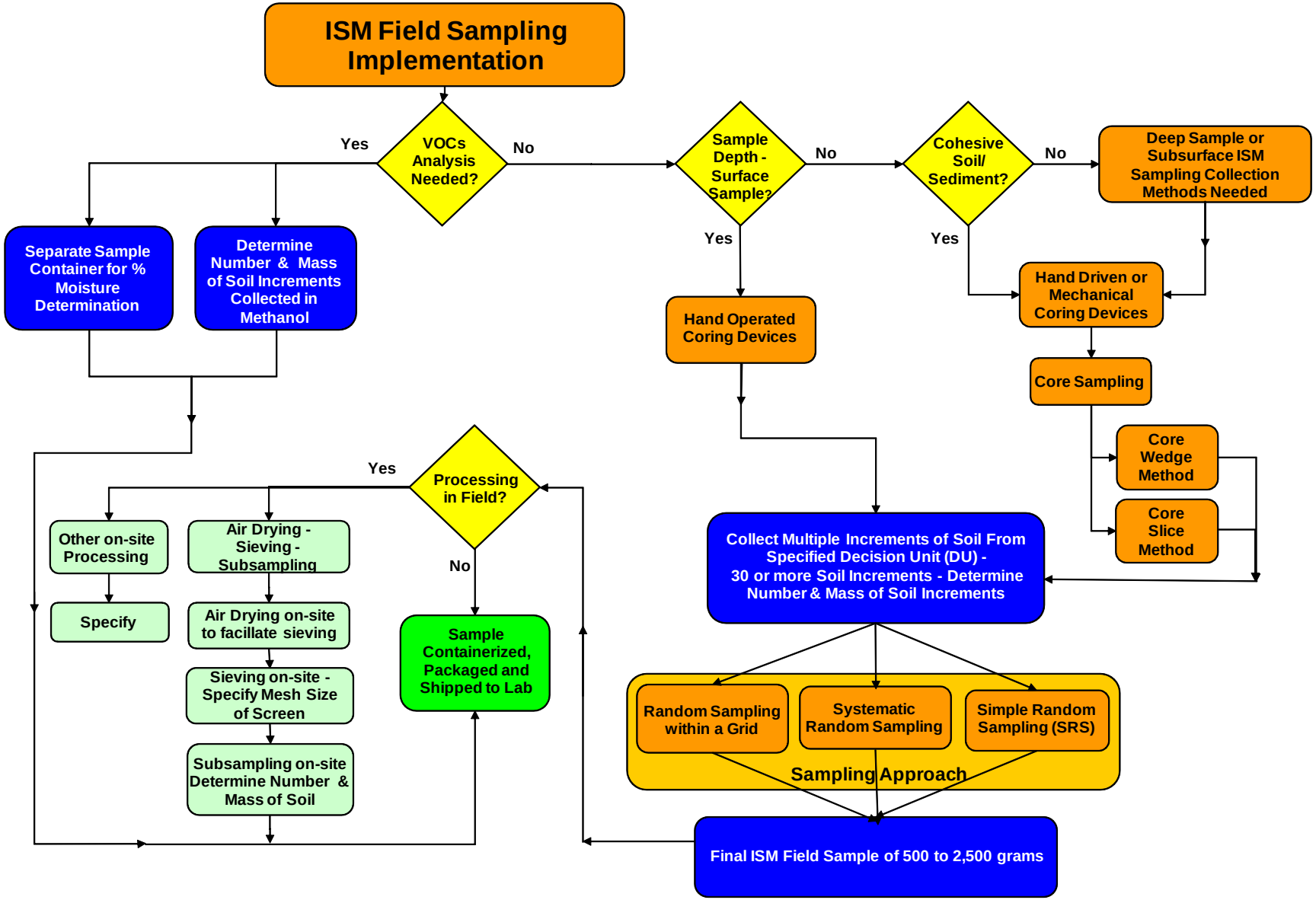


Figure 5-1. Field sampling implementation flowchart.

5.2 Sampling Tools

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 tool should obtain cylindrical or core-shaped increments of a constant depth from the presented surface. 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), and $3d + 10$ mm for a fine material (Pitard 1993). Caution should be taken to select tools that equally retain all of the particles over the entire depth of interest. In general, sampling tools should have a diameter of at least 16 mm. For less cohesive soils, attempts should be made to retain the entire, complete core increment.

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

See Figures 5-2a and 5-2b for examples of sampling tools for nonvolatile ISM sample collection and Figure 5-12 for examples of sampling tools for ISM collection of VOCs. These are provided as examples only. Various other hand augers, core sampling tools, step probes, etc., are available from environmental or agricultural suppliers and are applicable to ISM if the specifications meet project DQOs. Again, the sampling tool(s) selected should minimize increment extraction and delimitation errors.

The sampling tools required to collect core-shaped soil increments of required length in the field are necessarily site specific. Alternate sampling tools that meet the basic ISM principles and project-specific objectives may be available currently or in the future. A variety of tools to address different soil types or site conditions should be taken into the field for any given project.

Cylindrical increments of a controlled depth can be obtained from cohesive soils with a variety of commercially available manually and machine-operated coring tools. For depths of 10 cm (3.9 inches) or less, individual increments often can be rapidly collected and dispensed into a sample container using hand-operated tools. For noncohesive soils and sediments, short- and long-nose scoops (trowels) can be used; however, care should be taken to obtain a "core-shaped" increment over the entire depth of interest. For depths greater than 10 cm, or for hardened and unconsolidated rocky geological materials, coring devices can be advanced with a hammer, slide bar, or some other means of mechanical assistance. Depending on site familiarity, one or several sampling tools should be readily accessible during all sampling activities.

Sampling devices can be used within a DU without decontamination but should be decontaminated or disposed of between DUs. 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.

Sampling devices can be used within a DU without decontamination, but should be decontaminated or disposed of between DUs.

Figure 5-2a. Examples of coring devices for nonvolatile soil increment collection. Top to bottom: Multi-Incremental Sampling Tool (MIST™), EVC Incremental Sampler, JMC Backsaver Handle**,** and Soil Tube.

Figure 5-2b. Example of a drill core bit sampling tool for nonvolatile soil increment collection.

5.3 Field Collection

5.3.1 Surface ISM Samples

ISM samples are composed of increments collected from specific points throughout the DU. The positioning of the collection points can be set using one of three approaches, as described in Section 4.3.4.2: simple random sampling (SRS), random sampling within a grid, and systematic random sampling. SRS involves determining random locations across the entire DU. Note that "random" in this context does not mean wherever the sampling team feels like taking a sample and that a formal approach to determining the random increment locations must be used. With random sampling within a grid, the DU is overlain with a sampling grid and soil increments are collected from random locations determined in each grid cell (see Figure 4-8). Systematic random sampling is similar except that only the initial grid cell sampling location is randomly determined and the same relative location is sampled in each of the other grid cells (see Figure 4-7).

As predicted by statistical sampling theory and demonstrated by the ISM simulations discussed in Section 4.3.4.2 and Appendix A.1, SRS yields the most representative (least biased) estimate of the mean. However, it is also the least practical to implement since field staff have to navigate to predetermined locations nonuniformly positioned within the DU. SRS also may result in a sampling pattern that leaves large portions of a DU unsampled, which may not be acceptable to regulators, risk managers, members of the public, or other stakeholders. In practice, systematic random sampling is most often chosen for ease of implementation and to avoid the appearance of over- or underrepresentation of subareas within a DU, as may occur with SRS. Refer to *Superfund Representative Sampling Guidance,* Vol. 1 (USEPA 1995b) for additional information.

Incremental soil samples are prepared by collecting multiple increments of soil (typically 30 or more) from a specified DU and physically combining these increments into a single sample, referred to as the "incremental sample." When the individual increment mass is adequate, this number of increments (n)

Incremental soil samples are prepared by collecting multiple increments of soil (typically 30 or more) from a specified DU and physically combining these increments into a single sample.

generally results in a soil sample with a contaminant concentration representative of the estimate of the mean contaminant level within a DU (i.e., a representative sample). That is, even when the distribution of individual data points (i.e., discrete sample results) is nonnormal, the distribution of sets of means from the population will approach a normal or Gaussian shape as the number of increments (n) increases (Jenkins et al. 2005). See Section 4 of this document on the statistical basis of ISM for a more detailed discussion of increment number(s), adequate increment mass, and representativeness.

As sampling theory indicates, the number of increments collected depends on the amount of distributional heterogeneity present within the DU for the constituent of interest. A variety of factors may influence the amount of distributional heterogeneity within a DU. These include, but are not limited to, the following:

- contaminant type and physical characteristics
- soil type and physical characteristics
- contaminant release mechanism (e.g., spill, area-wide application, munitions range)
- others

As the DU gets *significantly* larger, the amount of distributional heterogeneity may increase. In these cases, depending on site specifics, CSM, and DQOs, it may be necessary to increase the number of increments per DU to 50 or more. Collection of a greater number of increments in each DU typically reduces the GSE (i.e., minimizes the variation among replicate samples). Alternatively, splitting larger DUs into two or more smaller DUs should be considered. It is not normally necessary to increase the number of increments unless there is reason to believe the DU has more distributional heterogeneity than can be controlled with 30–50 increments. See Section 4.3.4.1 of this document for the statistical information and evaluation of the number of increments for ISM sampling.

In general, a minimum of 30–50 increments is sufficient for most DUs. However, in published reports for solid/particulate-type chemicals of concern (COCs) (e.g., energetics/explosives, particulate metals, etc.) 50–100 increments per DU have been collected. USEPA SW-846 Method 8330B recommends collecting 30 *or more* evenly spaced increments to build a sample with a total mass of >1 kg. It is anticipated that as ISM matures, additional information on the optimal number of increments for other types of contaminants may become more readily available. The number of increments to be collected from each DU of a site investigation should be evaluated during systematic planning as part of the DQO process and documented in the sampling and analysis plan (SAP).

Generally, a minimum of 30 increments should be collected for each DU, with each increment weighing 20–60 g.

In general, individual soil increments typically weigh 20–60 g. Final ISM field samples typically weigh 500–2500 g. To minimize FE to an acceptable level, it may be necessary to calculate the target bulk ISM sample mass for collection prior to field implementation and ISM collection (Pitard 1993, Ingamells and Pitard 1986, see also Section 2 and Hyperlinks 14 and 18) It may be necessary to collect bulk ISM samples >2500 g to reduce FE to an acceptable level. Additionally, note that sieving of soil

samples to the <2 mm particle size reduces the amount of soil mass available for preparation and analysis, so this fact needs to be taken into consideration during systematic planning if minimizing FE is a DQO. Additionally, sieving is not applicable for the collection of VOC samples (see Section 5.4.2). Based on the required final mass of the ISM sample, as dictated by FE considerations and the number of increments determined by distributional heterogeneity, the minimum mass of the individual increments can be calculated. 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. Typically, the mass of the final ISM sample is sufficient for the planned analyses, any additional QC requirements, or repeat analyses due to unanticipated field, laboratory, and/or QC failures. 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 plan as described in the following formula for estimating sampling equipment requirements based on a predetermined ISM mass and number of increments:

> The number of increments to be collected per DU, the sampling depth, and the targeted mass should all be specified in the SAP.

$$M_s = \rho \bullet n \bullet D_s \bullet \pi \bullet (\theta/2)^2$$

where

M_s = targeted mass of sample (g)
D_s = increment length (cm)
n = number of increments
ρ = soil or sediment density (g/cm^3)
θ = diameter of sample core (cm)

These parameters, along with the density of the soil or sediment matrix, assist in the selection of the sampling tool to collect the appropriate individual increment mass for the total ISM sample (Walsh 2009).

Figure 5-3 and Table 5-1 (Walsh 2009) are provided as examples for estimating increment mass that can be collected for a given sampling depth and soil density, once the DU size, number of increments and total ISM sample mass have been established. Generally, a minimum of 30 increments should be collected for each DU, with each increment weighing 20–60 g. 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. For DUs of nonuniform thickness, the available thickness at each increment location is collected to ensure spatial coverage and the increment is not required to have similar volume or mass.

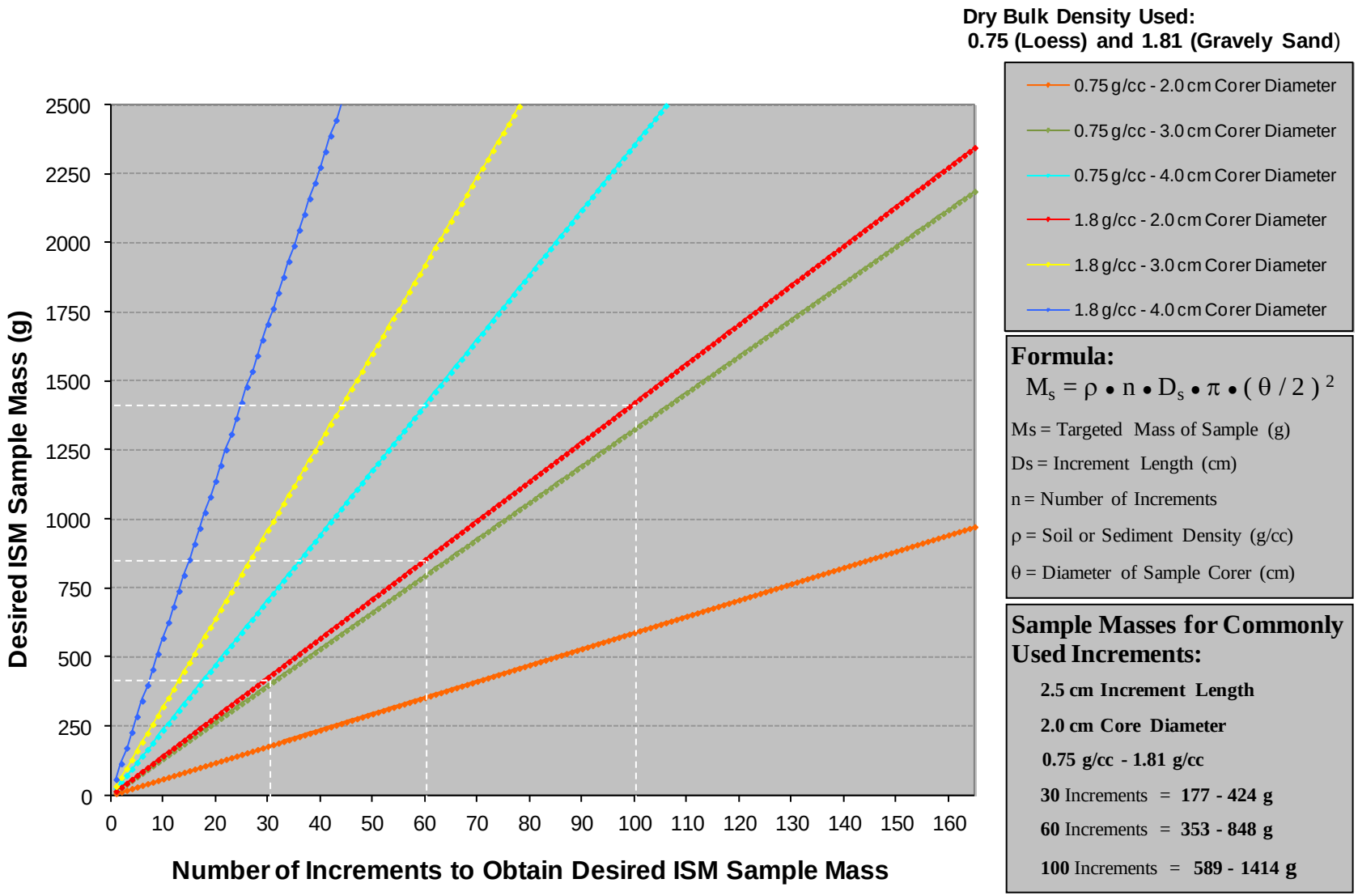


Figure 5-3. Estimated sample mass based on number of increments for set increment and substrate density.

The online version of this document contains a working calculator for incremental soil mass:
http://www.itrcweb.org/ISM-1/5_3_Field_Collection.html

Table 5-1. Estimated sample mass for set increment length and substrate density

Corer diameter (cm)	Number of increments to obtain desired ISM sample mass				
	500 g	**750 g**	**1000 g**	**1500 g**	**2000 g**
Soil density 1.6 g/cm^3, increment length 2.5 cm					
2.0	40	60	80	119	159
3.0	18	27	35	53	71
4.0	10	15	20	30	40
Soil density 1.8 g/cm^3, increment length 2.5 cm					
2.0	35	53	71	106	141
3.0	16	24	31	47	63
4.0	9	13	18	28	35

Substrate density may vary from 0.75 g/cm^3 (for Loess) to 1.81 (for Gravely Sand) with substrate densities typically ranging 1.6–1.8 g/cm^3 (Walton 1988, Domenico and Schwartz 1990).

If replicate ISM results indicate data variability is too high (i.e., interferes with decision making for the DU), additional data evaluation, sample analysis and/or resampling may be required to achieve project-specific DQOs. Note, however, that high variability between ISM replicates may also be a result of laboratory processing and subsampling procedures, which can be evaluated by examining the results of laboratory replicates (if analyzed). High data variability determined to be a result of DU heterogeneity and/or field sampling error may require revision(s) to the ISM design and implementation, including DU modification, additional increments, and/or increased increment mass (see Section 5.3.5).

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

Soil density across the DU should be reasonably uniform (e.g., the same general soil classification can be expected throughout the DU). When the surface of the DU contains both vegetated and nonvegetated areas, it is very 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 (e.g., areas of sand with areas of fat clay, areas of peat, etc.), those different soil type areas should be factored into DU determinations (i.e., 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.

For surface/exposed soil, common sampling depths are 2.5, 5, 10, or 15 cm; however, depths can be greater depending on the DQOs and CSM, including expected vertical distribution of the chemical of potential concern (COPCs) (due to infiltration, buried utilities/facilities, stockpiles, etc.), the exposure scenario, and/or regulatory requirements. Additional depths and/or DUs may be required for vertical delineation. Contaminant dilution should also be considered when determining increment depth. For surface-deposited energetics at active U.S. Department of Defense (DOD) training ranges, soil profile samples have shown 1–2 orders of magnitude decreasing concentrations within the top 10 cm (USEPA 2006c). For these types of sites, the desired sampling depth is approximately 2 cm, based on research conducted at Cold Regions Research and Engineering Laboratory (CRREL); greater increment depths result in dilution of the

contaminant concentration. In general, the location, lateral extent, and depth of the DU should be selected to represent an area of known or expected similarity. For greater depths, use of a smaller-diameter sampling tool may be desirable but often is impractical due to presence of pebbles, rocks, and vegetation. In general, however, the smallest diameter sampling tool applicable to particle size requirements is recommended to minimize delimitation and extraction errors and to attain the necessary soil mass (see Section 5.2). Alternative technologies for site-specific conditions should be considered, as appropriate.

A square, rectangular, circular, or other naturally or structurally defined DU (e.g., 5 m perimeter around the exterior of a building) is first subdivided or gridded-off into uniform cells or subareas based on the desired number of increments to be obtained. That is, the number of cells is equivalent to the number of increments. Using the systematic random design, a random position is established for a given cell, and then the same position is repeated in all of the remaining cells in the DU. For the random sampling within grids design, a random position is designated and sampled in each cell. A random starting point or random position for each cell can be obtained with dice or a random number generator. The process is repeated for replicate samples; i.e., a new random position is established for the single collection point to be repeated in all of the cells, or for each cell, depending on the sampling design. A Global Positioning System (GPS) device should be used to delineate the DU. It may or may not be necessary to determine the exact location of each increment depending on the DQOs specified during the systematic planning process.

Depending on the size of the DU and terrain features, placement of markers (e.g., pin flags and posts) at the corners and or edges can assist with a visual delineation of the cells or subareas where increments are to be collected. That is, the markers can define lanes, grids, or 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 could be divided into five rows, with six increments collected from each row in a systematic random fashion, with an initial random starting point. For more rectangular-shaped DUs, fewer rows might be used with more increments per row collected (Figure 5-4). Row lengths and increments per row may be modified as needed for odd-shaped DUs. However, with other shapes, it is recommended that the perimeter be marked and flags be prepositioned across the DU in one or more perpendicular lines. Then a trial run with no sample collection is performed to quickly establish the distance between increment collection points to achieve the desired number of

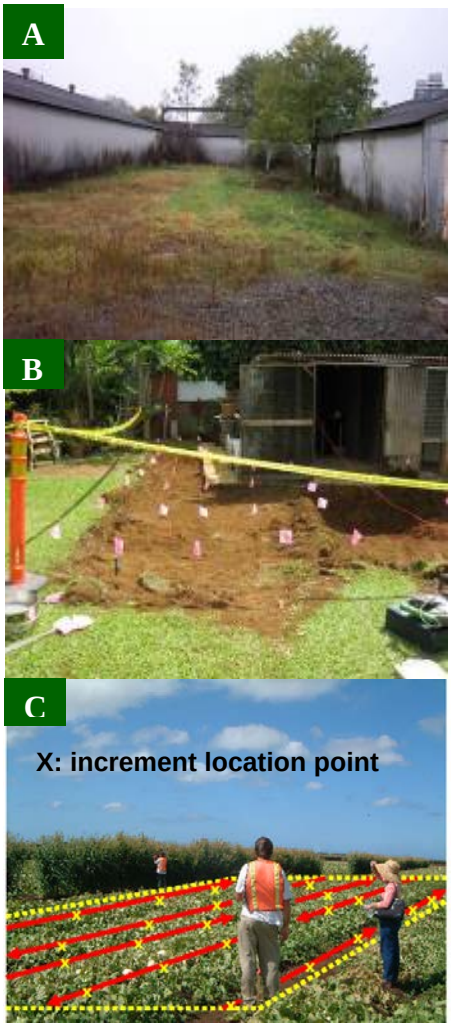
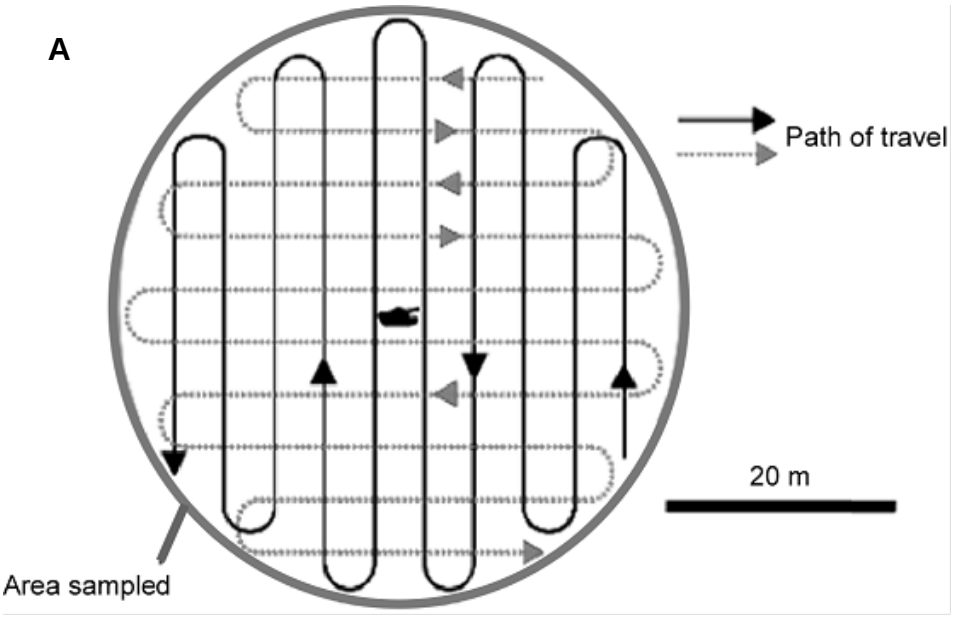


Figure 5-4. Example DUs from industrial (A), residential (B), and agricultural (C) sites.

increments, while using the flags as guides that were positioned within or around the DU.

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 (e.g., 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. The *User's Manual for the CRREL Multi-Increment Sampling Tool* (Walsh 2009) lists common sampling supplies and vendors that would be appropriate for SVOCs and metals. Sampling tools are set for the appropriate depth. Flags may be used to mark DU boundaries and to aid in visualizing the travel paths and/or to mark the actual increment locations. The ISM sampler starts in one corner or end of the DU and collects an increment at the predetermined positions. For the systematic random sampling design, the location of the first increment is determined randomly, and subsequent increments are collected in the same relative location within each grid, resulting in a serpentine collection pattern ending at the opposite corner or end of the DU from where sampling was started (see Figure 5-5). Note that, for simplicity, Figure 5-5A depicts collection of duplicate ISM samples rather than the recommended triplicates. Additional guidance on ISM can be found in the following documents:

- Method 8330B, Appendix A (USEPA 2006c)
- *Protocols for Collection of Surface Soil Samples at Military Training and Testing Ranges for the Characterization of Munitions Constituents* (Hewitt et al. 2007)
- *User's Manual for the CRREL Multi-Increment Sampling Tool* (Walsh 2009)
- *Technical Guidance Manual* (HDOH 2008b)
- *Implementation of Incremental Sampling (IS) of Soil for the Military Munitions Response Program,* Interim Guidance Document (IGD) 09-02 (USACE 2009)



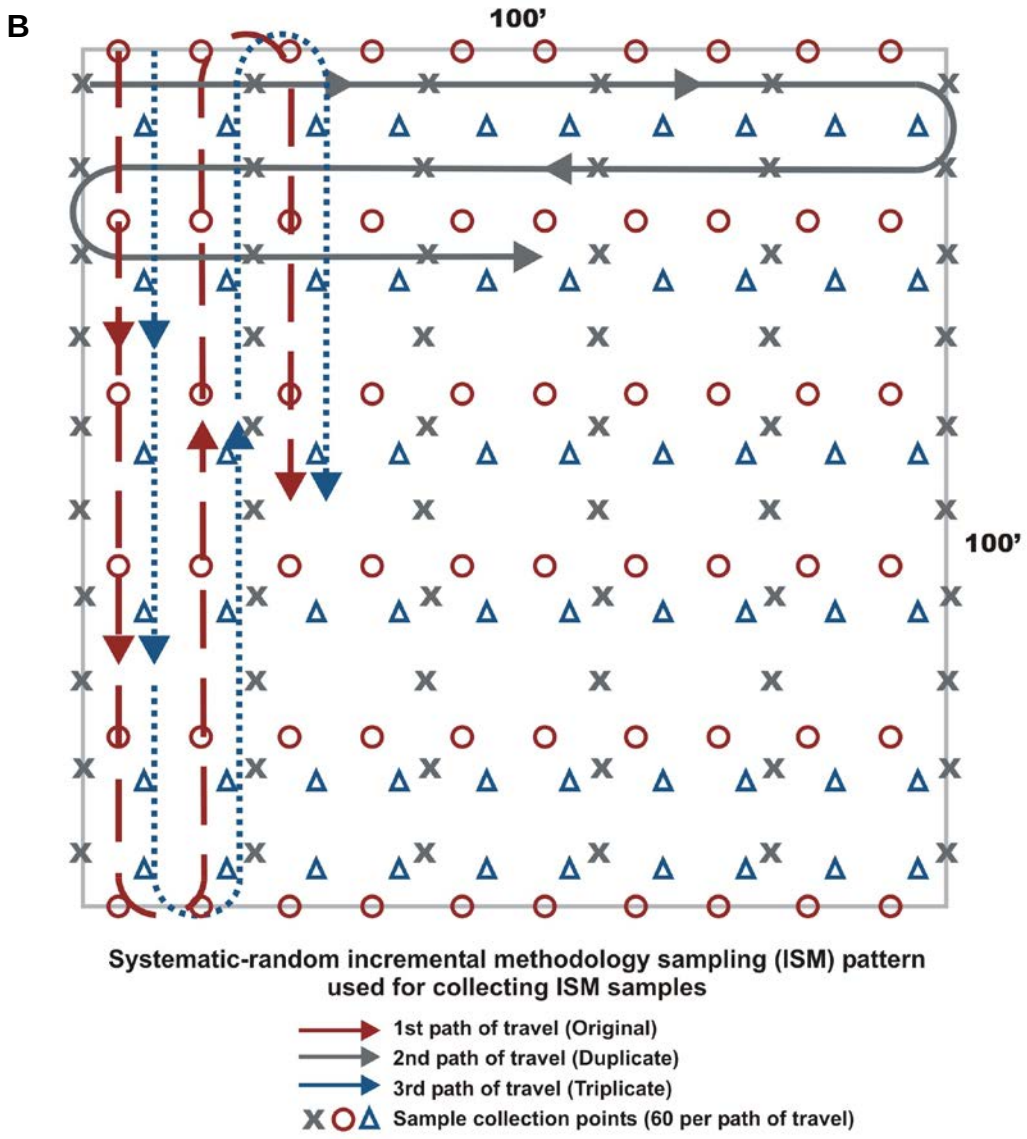


Figure 5-5. Illustrations of systematic random incremental sampling pattern used for collecting samples in circular (A) and square areas (B).

5.3.2 Subsurface ISM Samples

As discussed in Sections 2 and 3, DUs are by definition 3-D in nature and are intended to focus the investigation on a specified volume or mass of soil. Obtaining good spatial coverage and data quality for subsurface soils is more challenging but is still necessary. The objectives for surface vs. subsurface investigations may be similar in nature, for example, to estimate the representative concentration of targeted contaminants for targeted depth intervals (e.g., within the defined vertical limits) or to determine or confirm the lateral boundaries of the source area. For remedial purposes, the estimation of contaminant mass within the DU is also sometimes critical (e.g., mass of tetrachloroethene for design of soil vapor extraction system or mass of dioxins for design of in situ thermal desorption system). The practical application of ISM sampling must be considered during project planning, especially when considering implementing it for nature and

extent investigations of subsurface contamination. Often, alternative sampling techniques (e.g., discrete sampling, field screening, or field analytical methods) may be more applicable and/or cost-efficient.

Soil samples collected as part of a subsurface investigation are intended to be representative of a specific depth interval.

Soil samples collected as part of a subsurface investigation are intended to be representative of a specific depth interval. As discussed in Section 3, this trait can be described as the resolution of the data collected. Discrete soil samples from borings or excavations have traditionally been used to characterize subsurface soils. In most cases, however, discrete samples may provide less spatial coverage of the targeted depth intervals and also increase laboratory analytical costs. As discussed below, alternative sample collection approaches to improve sample data quality and reduce laboratory costs include options for ISM core sampling across targeted depth intervals.

5.3.2.1 Subsurface ISM samples using core sampling

If a coring device is used, samples should be collected from targeted depth intervals in a manner that ensures the best coverage of the interval. For example, the selected subsurface DU investigation strategy may require the collection of soil samples from specified 15 cm (6-inch) intervals over a 1 m depth (see Figure 5-6). In other cases, the mean concentration of a targeted contaminant over the entire 1 m DU (or larger) interval may be desired for risk assessment or remedial purposes (see Figures 3-7, 3-8, and 3-9).

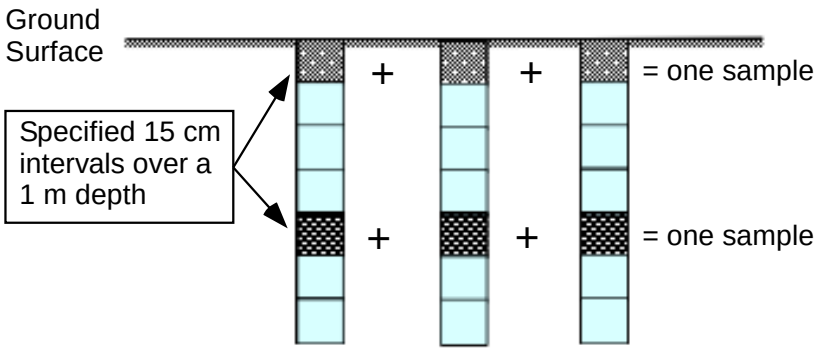


Figure 5-6. Schematic of a procedure to collect an ISM profile sample where two depths have been selected.

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, 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. This method can result in large ISM samples (approximately 5–10 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. Two options are described below.

Another option for collecting a representative subsample from a subsurface core increment for nonvolatile 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 single wedge of soil is 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, constitutes the mass of an individual increment of an ISM sample (see Figure 5-7). Individual wedges from 30 or more separate DU cores are then combined to form the complete subsurface ISM sample. This option results in a more biased and less precise estimate of the DU mean as compared with collecting the entire core. However, since the mass of each increment (and thus the ISM sample mass) is reduced, some of the practical constraints associated with handling full core increments are addressed.

Another option for collecting a representative *subsample* from a subsurface core increment for nonvolatile contaminants is to collect a "core wedge" sample.

Figure 5-7. Example of removing a wedge from the entire length of a soil core.

Replicate(s) can be collected from the same core, combined with other wedge increments, and submitted as separate ISM sample(s) to assess the precision of this subsampling strategy. However, core wedge replicates are not the same as ISM field replicates because ISM field replicates require completely separate incremental 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. The variability of wedge subsamples from alternative areas of the core

is evaluated, e.g., replicate wedge collected 180° opposite the initial wedge subsample. ISM field replicates 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 5.3.5. Core wedge replicates may also be collected when COPCs require separate laboratory processing procedures (see Section 6.2.2.2).

This approach is 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 5.4.2).

The least preferred option for subsampling individual subsurface cores for nonvolatile contaminants is to collect a "core slice" from the targeted DU layer (see Figure 5-8). 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 feet in length (e.g., 8–10 feet bgs), a 4-inch perpendicular slice is 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 DU cores are then combined to form the complete subsurface ISM sample. This option introduces more bias than whole-core increment or core-wedge approaches. However, by reducing the increment mass, some of the logistical issues associated with handling the full core or the wedge increments are addressed. This is the least recommended approach for subsurface ISM core sampling since it is least likely to accurately represent the complete vertical length of the targeted DU layer.

Subsurface ISM increment collection techniques in recommended order are as follows:

1. Collect entire core interval
2. Core wedge subsample
3. Core slice subsample

Figure 5-8. Examples of "core slice" sample. *Source*: Illinois EPA LUST FAQ and BIOTREE websites.

5.3.2.2 Additional subsurface ISM considerations

As with surface ISM samples, it is recommended that a minimum of 30 increments be collected for each DU. In some cases, collecting the recommended minimum of 30 soil increments per subsurface DU may not be feasible or practical. Reducing the number of increments collected per sample may be the only viable option. In this situation, it is important to recognize that collection of a reduced number of sample increments generally increases the GSE and results in a less precise and more biased estimate of the mean contaminant concentration. Depending on the degree of data variability that can be tolerated within the project-specific DQOs, a significant reduction in the number of increments may result in a decision error. A sample containing fewer increments than required to estimate the DU mean concentration within the project-specific uncertainty level may not be considered a defensible ISM sample. Consequently, in these circumstances careful review of DQOs as well as any other sampling options that may be available is warranted. The subsurface sampling strategy chosen, the sampling constraints, and potential impacts on data quality should also be identified in the DQOs in the SAP and or Quality Assurance Project Plan (QAPP).

> Collecting fewer increments generally increases the grouping and segregation error (GSE) and results in a less precise and more biased estimate of the mean contaminant

Increments from the same depth interval throughout the DU can be combined and used to create a single ISM sample for that depth interval. This is a useful approach for the characterization of vertically stacked DUs (see Figures 3-7, 3-8, and 3-9). Data for each ISM sample can be used to create a 3-D map of contaminant levels in the DU. This procedure can be especially useful where a large number of side-by-side DUs are designated for the investigation of large areas (e.g., redevelopment of a former golf course contaminated with pesticides).

5.3.3 Stockpile ISM Samples

Special considerations for selecting DUs during the systematic planning process for sampling soil stockpiles include the following:

- 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, (e.g., large or unstable)
- contaminants targeted for lab analyses

One of the best options is to coordinate sampling with the formation of any stockpiles on the site. When the stockpile is being formed, there is generally good access to sampling each portion of the pile over time, and 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 (e.g., from the front-end loader buckets, at appropriate intervals), or flattening or spreading out the stockpile sufficiently so that it is safely accessible to sample with a hand coring or other device. If the stockpile is very large or unstable, all 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 systematic random or random within grids increments throughout the stockpile (both vertical and

horizontal locations). Replicates are important to evaluate the precision of stockpile sampling and should be collected similarly to the original sample except in separate random locations. Large stockpiles could be divided or segregated into separate DUs (see Figure 3-10), especially if a specific portion or volume of the stockpile will be used in a manner that will become the primary exposure unit of concern in the future (e.g., certain portions or volumes of the stockpile will be hauled to residential lots as surface fill for backyards). A resource for additional information on ISM approaches for soil stockpile sampling is the Hawaii Department of Health (HDOH) *Technical Guidance Manual* (HDOH 2008b). Refer to Section 3.6.4.2 of this document for additional information on ISM sampling of stockpiles.

5.3.4 ISM Confirmation Sampling

Confirmation sampling may be performed during post-removal activities to verify that residual concentrations of target COCs are below the predetermined cleanup goals for the site. Confirmation sampling is often a requirement to achieve final clean closure certification. Confirmation samples are typically collected from the sidewalls and floors of an excavation to confirm that concentrations remaining after excavation are below specified concentration limits. Results from individual grab samples, an average or a 95% UCL from discrete samples, are often compared with the cleanup criteria for the site for this purpose.

An incremental sample result is specifically designed to estimate the mean concentration in a volume of soil designated as a DU. If excavation is performed for a site based on results from ISM sampling, it is usually because one or more DUs "failed" (i.e., had concentrations above the specified cleanup goals). Once the soil in a failed DU has been removed, the motivation for sampling the sidewalls and floor of the excavated DU is presumably to determine whether surrounding potential DUs also require remediation. If adjacent areas have already been designated as DUs, evaluated, and found to have soil concentrations within acceptable limits, confirmatory sampling in the conventional sense may not be necessary. If adjacent areas have not been adequately characterized, collecting ISM samples around the excavation can inform the need for or against further removal. In this situation, the expanded investigation requires new planning, including the designation of additional DUs and the determination of appropriate cleanup goals. One approach is to designate a volume of soil surrounding the excavated area as a new DU and sample from the walls and floor accordingly. This process is somewhat analogous to conventional confirmatory sampling. However, it is important to consider how the areas of the walls and floor relate to the volume of soil in the new DU and take increments in a manner that ensures a sample is representative of the DU. It is also important to recognize that the cleanup goal for a DU consisting of soil immediately surrounding an excavated area might be different from the original cleanup goals used for site evaluation because the objectives (e.g., addressing concern for potential for direct contact, leaching to groundwater, etc.) may be different, given the size and location of the new DU. As always, clear articulation of objectives and proper planning are essential.

In summary, the use of ISM samples to confirm excavation of a source area DU can be highly advantageous over a traditional, small number of discrete samples. The excavation floor and sidewalls should be treated as individual DUs (see Figure 3-11), with the

The use of ISM samples to confirm excavation of a source area DU can be highly advantageous.

investigation objective of determining whether the estimated mean concentration of COPCs for these areas exceeds targeted screening levels. Again, these issues should be evaluated and determined as part of the planning and DQO process. 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. There may be regulatory limitations to this approach, however. For example, if regulations require cleanup of releases to a not-to-exceed regulatory level (e.g., the maximum concentration determined by discrete samples), then an ISM mean concentration may not be applicable and/or accepted by the regulating authority.

5.3.5 Collection of Field Replicate ISM Samples

Replicate ISM samples (triplicates or more) should be taken to quantify uncertainty in the estimate of the mean concentration within the DU.

In the field, replicate incremental samples (three or more) should be taken to ensure reliable estimates of the mean concentration within the DU. The number of replicates and frequency of taking replicate incremental samples should be specified in the SAP and comply with project DQOs.

To statistically evaluate sampling precision for each DU, additional completely separate replicate ISM samples are collected. The increments are collected in simple random, systematic random, or random within grid locations within the DU that are different from those used for the initial ISM sample. ISM field replicates are made of the same number of increments collected in the initial ISM sample and collected using the same sampling pattern from within the same DU. The replicate samples are prepared and analyzed in the same manner as the initial sample. Three replicate samples (i.e., the initial ISM sample plus two additional samples) should be considered the minimum. In some cases, more replicates may be necessary to reduce data variability and/or to calculate a 95% UCL of the mean that is closer to the actual mean of the DU. Section 4.3.4.1 discusses the statistical basis and evaluation of replicate ISM samples.

When sampling in a systematic random sampling pattern, the increments for an ISM replicate sample are generally collected along the same approximate directional lines established through the DU for the initial ISM sample. Increment locations for ISM replicate samples differ from each other by the selection of different random starting locations on the first line/row of the DU and continuing to sample at this different random interval throughout the DU for each replicate (see Figure 5-5). Thus, the increments for ISM replicates should *not* be collected from the same locations or colocated with those used for the initial ISM sample. When using the random sampling within grid pattern, replicates are constructed from increments taken from different, randomly selected locations within each gridded area. With simple random sampling, three sets of random locations across the DU are selected and increments collected for each set are used to create the replicates. Replicate ISM samples should be submitted to the laboratory as "blind" samples, meaning the laboratory does not know they are replicate samples of the initial ISM samples.

If only one DU is being investigated, a minimum of three replicate samples should be collected to provide a measure of variability. For sites with multiple similar DUs, "batch" type replicates may be a consideration; for example, three replicates in one DU could be used to provide an estimate of variability that is extrapolated to a number of similar DUs (similar to how labs use batch replicates for determining lab analysis precision). Each site and/or project is unique in

terms of numbers of DUs and how similar these DUs are, so decisions on numbers of replicates are unique to each site and should be addressed clearly in the SAP. For the batch type of replicates to apply, each DU in the "batch" should have a similar CSM, including the same soil type, site use/history, contaminant deposition, etc. If considered, this batch approach must be discussed, clearly documented, and agreed to by all parties involved during the systematic planning process (see Section 4.4.2). Section 7 discusses how replicate ISM sample data are used to assess sampling error and make decisions.

5.4 Field Handling of ISM Samples

5.4.1 ISM Samples for Non-VOCs

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 that 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. However, these techniques 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. However, depending on site logistics, the type of soil, the total number and/or mass of ISM samples, etc., sample processing can be initiated in the field for some contaminants (e.g., SVOCs, pesticides, PCBs, and metals) with appropriate cautions as noted below.

Moist samples may need to be air-dried to facilitate sieving in an appropriate dust-free location where temperatures and ultraviolet (UV) light are not expected to cause degradation of COPCs. 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 (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 soil particles are generally considered "soil," while larger particles are considered gravel, rocks, or other materials (e.g., sticks and roots). Additionally, field sieving is an option that allows the user to calculate the mass of an bulk ISM sample needed to meet DQO requirements (including FE, see Hyperlinks 14 and 18), based on the soil particle size. 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 lab processing/analysis should be clearly discussed in the SAP. Unless field subsampling will be performed (see paragraphs below), the entire sieved ISM sample fraction should be submitted to the laboratory for appropriate processing and subsampling.

When dealing with contaminants that have been deposited as solid particulates (e.g., energetics, metals at firing ranges, etc.), 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). In cases where sieving is conducted in the field to obtain a targeted particle size (particle size selection), the entire sieved ISM sample should be ground prior to subsampling

(if particle size reduction is part of the SAP). Similar studies evaluating field subsampling for contaminants deposited as liquids (e.g., fuels, solvents, etc.) are not available at this time.

The SAP may specify particle size selection (sieving) and subsampling in the field for the analysis of SVOCs and specific metals. This procedure constitutes, or is similar to, the normal laboratory subsampling step. It should be reiterated that it is recommended that all ISM sample processing be performed in a controlled laboratory setting.

If *field subsampling* is 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 sieved ISM sample should be spread out in a thin layer on a clean surface, e.g., a large, disposable, aluminum baking pan, allowing the entire sample to be accessed. A subsample is then obtained by removing 30 or more equal increments from systematic random locations (see Figure 5-5). The increments collected to form the subsample sample should equally represent the top and bottom of the processed material. This is achieved by using a rectangular, flat-bottom sampling tool with sides and a minimum 16 mm width (see Figure 5-9), as opposed to one that is curved or spoon-shaped (see Figure 5-10). Spoon-shaped sampling tools bias the mass of soil collected.

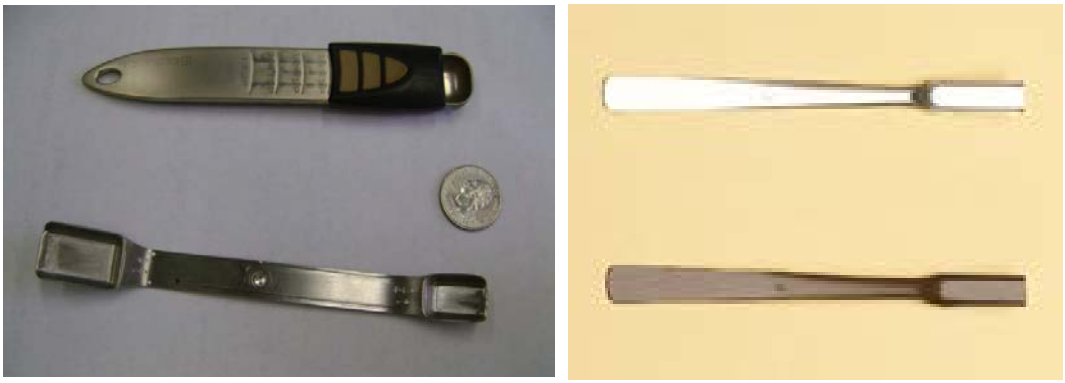


Figure 5-9. Examples of rectangular and flat-bottom sampling tools.

Figure 5-10. Example of subsample being collected in the field.

The mass of sample required for the analytical test or tests is used to determine the mass of each of the 30 or more increments. For example, if a mass of 30 g is required for the analytical extraction and analysis, 30 separate ~1 g increments are collected from systematic random locations. Depending on the 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 (e.g., 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 subsampling procedure. Refer to Section 6.2.2.7 describing analytical subsampling techniques and specifically the description of 2-D Japanese slabcake sampling.

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 (i.e., "homogenizing") in the field or lab 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 to a thin layer and collecting many 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 in the ISM sample.

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 (as described above) process is repeated on *one* ISM sample to form replicates. The replicate results are used to evaluate the precision of the field processing and subsampling. Note that the subsampling replicates should be collected *in addition to* the ISM field replicates described in Section 5.3.5.

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

- not recommended for contaminants deposited as solid particulates (e.g., energetics, metals at firing ranges, etc.)
- lack of commercially available, correct subsampling tools (e.g., 16 mm wide, flat-bottom scoop with sides)
- requires a controlled environment to air-dry, sieve, and subsample, if necessary, to minimize the potential loss or introduction of COCs during processing
- additional *subsampling* replicates need to be collected and analyzed to evaluate precision
- more knowledgeable/trained field personnel required

5.4.2 Volatile Organic Compound ISM Samples

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

ISM samples can also be collected for VOCs contaminant analyses from cores, excavation-pit bottoms and walls, stockpiles, underneath paved areas, etc. USEPA SW-846 Method 5035A Section 8.2.2 (USEPA 2002b) describes the collection of discrete soil samples preserved in the field. The ISM VOC approach is similar to this method and to that described for sampling ISM nonvolatiles in the subsurface, except that numerous soil increments are placed directly into an adjusted volume of extraction solvent in the field (e.g., methanol, shown in Figure 5-11). Individual increment mass should be similar provided the soil density is fairly uniform. Typically, individual increments are not weighed in the field during collection. Similar mass per increment is assumed with similar volume collected.

Figure 5-11. Bottles containing methanol and 44 five-gram plugs of soil.

For exposed soils, such as surface soils or exposed excavation sidewalls/bottom soils, the entire mass of soil collected at a single point represents an increment. These increments are collected using VOC coring devices (see Figure 5-12) and combined in a sample bottle containing a predetermined volume of methanol (see Figure 5-11). Thus, VOC ISM samples of exposed soils are collected and combined in similar fashion as non-VOC ISM samples with the exception that they are field-preserved in methanol.

Figure 5-12. 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).
Source: Courtesy www.ennovativetech.com.

ISM sampling may also be used for VOCs in the subsurface. As previously discussed in Sections 3 and 5.3.2.1, ideally the entire mass of soil collected in a subsurface core across the targeted DU depth represents the increment for that boring. The entire mass, therefore, would be

preserved in methanol and incorporated into the ISM sample for the targeted soil layer. Realistically, this task is impractical, since the volume of methanol required to preserve entire core increments or the combination of increments from multiple cores would be impractically large. Additionally, preserving the entire core would prevent increments for non-VOC contaminants to be collected, if required.

Instead, the core may be subsampled by collecting numerous, small (e.g., 5 g) "plugs" at regularly spaced intervals along the targeted DU depth interval of the subsurface core. As with ISM VOC sampling of exposed soil, the plugs are immediately placed in a sampling bottle containing a predetermined volume of methanol. Figure 5-13 shows an example of this type of ISM VOC sample collection from subsurface cores. Nominal 5 g plugs of soil can be collected across the core using a VOC coring device (see Figure 5-12). The spacing interval of the VOC plugs along the core interval should be determined during the systematic planning process. It is possible to determine the optimal spacing on a site-specific basis, through the collection and analysis of differently spaced plugs along the core interval. However, based on limited field experience to date, plugs should be located no more than 2 inches apart as a starting point. This distance was determined to be adequate to capture the potential heterogeneity of VOC concentrations along the vertical length of the core. It may be necessary to decrease the spacing depending on the site-specific distribution of contaminant concentrations and DQOs. The coring device used to collect the increments should be filled completely so that each increment has the same volume of soil. The complete soil plug must be transferred to the sample container. Additionally, the ISM sampler should be aware of potential volatile loss once the core is opened. 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.

In general, the potential to combine a larger mass of soil from multiple plugs from a subsurface core spaced along the entire length of a targeted DU depth results in a VOC sample that is more representative of the soil core. However, the handling and shipping of large volumes of methanol, as well as tracking and combining preserved increments into a single ISM sample, may present logistical issues, as discussed in the following paragraphs.

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 of the increments. Project planning must determine the number and size of increments (see Section 3). Laboratory personnel should be consulted during systematic planning so that sample size, methanol volume, and bottle size are determined in advance.

Shipment 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 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." Shipments via air transport may also be required to adhere to

International Air Transport Association Dangerous Goods Regulations (IATA DGR, IATA 2011).

Figure 5-13. Example of sample increments being collected and added to a bottle containing methanol for preserving VOC samples.

If the larger volume of methanol presents logistical problems for shipping which cannot be satisfactorily addressed, alternatives can be considered in consultation with the laboratory. With procedures and protocols in place ahead of time, these alternatives may include the following:

- 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/ extracted in the field by shaking periodically for at least 24 hours, allowing the solids to settle, decanting or pipetting 20–30 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 to the laboratory.

- Increments for VOC analysis could be collected and preserved with methanol individually (e.g., 5 g soil in 5 mL methanol in volatile organic analysis vials per USEPA SW-846 Method 5035A) and submitted to the laboratory for 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 5-14). 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.

Figure 5-14. Example of methanol aliquots from individual 5 g field-preserved increments being combined in the laboratory.

- Individual increments could be collected in separate sampling devices that have vapor-tight seals and are designed for zero headspace (e.g., Core N' One™, EnCore, or equivalent type sampler), and submitted to the laboratory at the appropriate temperature and within appropriate time frames (typically 24–48 hours) for combined placement in methanol before analysis.

- To fall under the small-quantity exemption of the shipping regulations, ISM volatile "subsets" could be collected, preserved with methanol in the field, 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. 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.

ISM VOC sampling procedures should minimize soil disturbance and possible VOC loss due to volatilization. For this reason bottles that have a narrow neck or other means of restricting volatilization losses and containing the volume of appropriate solvent 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. The headspace to preserved sample ratio (methanol + sample) should be less than or equal to that commonly achieved with discrete methanol VOC preserved samples (e.g., ~32 mL headspace to 8 mL preserved sample). 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 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 this solvent, 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.

Increments should be collected using tools that minimize the loss of VOCs during sample collection and allow the collection of at least a 5 g mass of soil. Special coring tools should be used for the collection of sample increments to be analyzed for VOCs, and increments should be quickly transferred to bottles containing methanol or another appropriate solvent (Hewitt et al. 2008). Syringe-type devices that can be pushed directly into the soil are preferable (e.g., Core N' One™, Terra Core Sampler, Easy Draw Syringe® and PowerStop Handle®, etc.). Examples of VOC coring tools are depicted in Figure 5-12. These types of devices, which are available in different sizes, can also be used for the collection of samples to be tested for nonvolatile chemicals. The device is pushed into the soil and retracted, and the increment collected is immediately extruded into a container with a premeasured volume of solvent (e.g., methanol). This procedure is repeated with each increment. Sampling devices can be used within a DU without decontamination but should be decontaminated or disposed of between DUs.

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 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 ounces or larger) and submitted to the laboratory.

A minimum of a 1:1 ratio of solvent volume to sample soil mass (i.e., 1 mL of methanol to 1 g of soil) is recommended. This procedure 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. Additional solvent may be required to ensure that the sample mass is completely submerged by the solvent. This requirement should be discussed with the laboratory. Select the sample container based on the total mass of soil to be collected and solvent required (e.g., 30 increments of 5 grams, approximately 3 mL volume of solid material per increment). For 30 increments a minimum of 150 mL solvent is recommended (see Figure 5-11). Use a container that is 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 (refer

to USEPA SW-846 Method 5035A). Note: 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. VOC recovery was better than 80% for all analytes except dichlorodifluoromethane.

Typically, a 24-hour period is a long enough period to extract VOCs from most soils. Tight clays are an exception and may take several days (Hewitt 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.

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.

Guidance on using ISM for the collection and handling of samples for the analysis of VOCs has been published by the State of Alaska (ADEC 2009). The Alaska guidance recommends that consultants provide a sampling and analysis work plan to the overseeing regulatory agency for review and comment prior to collecting any ISM samples. The analytical laboratory should also 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. A potential drawback of ISM for VOCs is that the methanol preservation (high-concentration method) approach does result in lower sensitivity. The methanol dilution step causes elevated analytical detection limits, method detection limits (MDLs), reporting limits (RLs), practical quantitation limits (PQLs), etc., as compared to the direct soil purge-and-trap, low-concentration method techniques. Analytical detection limits could be elevated above relevant screening levels for certain targeted contaminants (see Section 6.3.2). 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, then alternative approaches may need to be used. Options may include alternate analytical methods/techniques, such as selective ion monitoring (SIM), to achieve lower detection limits or select discrete sampling via USEPA SW-846 Method 5035A low-level VOC sampling. Research to improve detection limits from ISM VOC samples is ongoing and expected to improve in the near future. Consult with the laboratory for the latest detection limit capabilities.

6. LABORATORY SAMPLE PROCESSING AND ANALYSIS

6.1 Introduction

This section presents current practices and options available to process and subsequently analyze field samples obtained via ISM. Multiple options are available depending on the contaminants (i.e., explosives, metals, SVOCs, perchlorate, etc.). It is important to note that sample processing for various analytical suites is currently in early developmental stages and/or has experienced limited usage such that in many instances little to no performance information or specific standardized and published procedures are available. Future development of laboratory

equipment and/or sample processing techniques should be evaluated based on their applicability to ISM and their ability to meet project-specific objectives.

Incremental sampling has been successfully implemented at numerous sites for several contaminants other than explosives and metals, such as perchlorate and white phosphorus (Walsh et al. 1997; 75th CEG/CEV 2007; USACE 2009) and other analytical fractions (HDOH 2009, ADEC 2009). However, experience in applying ISM techniques to analytes other than explosives and, to a more limited extent, metals is limited and/or in early developmental stages. Therefore, not all of the possible sample processing and/or analysis approaches discussed in this section may be applicable to all ISM projects and contaminants.

As discussed in Section 2 and Section 3, it is paramount that the project planning team consider the various field-sampling approaches and options during the initial project planning stages when applying ISM (or any other sampling technique) at a site. Equally important during these early planning stages is a dialogue about, and resolution of, the field sample processing and subsequent analytical approaches/options and considerations that must be evaluated and agreed on by stake holders prior to finalizing an ISM field program. For example, will any sample processing such as sieving, drying, subsampling, or other processing occur in the field which is problematic and more likely to increase fundamental error, or will all processing occur at the analytical laboratory? It is recommended that all ISM sample processing occur in a controlled setting to minimize errors. Likewise, what will laboratory processing entail? Within the laboratory, how will bulk sample mass splitting or sample conditioning (if performed) be conducted? What sample conditioning steps, such as drying, disaggregation or hydration, will be performed? Will particle size reduction (if warranted) via grinding, milling, crushing or other means or particle size selection using sieving to focus on a particle size fraction of interest be required? Finally, what analytical subsampling techniques and/or determinative analytical methods will be performed? All have varying degrees of potential ramifications on data quality and usability, and each must be addressed in the project planning stages and DQO decision process along with close coordination with the analytical laboratory to ensure that appropriately defined and agreed on procedures are employed. Specifically, reference Sections 5.2 and 5.3, as well as Sections 2 and 3 of this document, as applicable, for ISM considerations that should be evaluated as part of the DQO decision process.

DQOs and Laboratory Coordination

As outlined in USEPA DQO guidance (USEPA 2006b), the DQO process is used to establish performance and acceptance criteria, which serve as the basis for designing a plan for collecting data of sufficient quality and quantity to support the goals of the study. As discussed above, 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.

Project planning DQOs guide the choices of sample processing and analytical options.

During this decision process, close coordination with the analytical laboratory is particularly important. Additionally, any deviations in the field that may affect laboratory processing and/or analysis must be communicated immediately to the lab and project delivery team. The project

delivery team must know, through close interaction of the team or the project chemist with the laboratory, precisely how sample processing and analysis will proceed within the laboratory. Specifically, what processing/preparatory/analytical methods/procedures are used, what QA/QC is employed and at what frequency, and what acceptance criteria will be applied? All of these greatly impact the specified project DQOs by affecting the project-specific data quality indicators (DQIs) and method quality indicators (MQIs) determined during the DQO process.

The subsections that follow discuss various options regarding sample processing and analysis for ISM. Figure 6-1 presents a flow diagram of these options and where decision points typically occur for an ISM sample. As ISM sample processing and analysis techniques advance, modifications and/or additions to this generalized schematic are expected to occur. Hence, 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 defined data of known quality and usability are obtained for the project.

6.2 Laboratory Processing

6.2.1 Volatile Organic Compounds

ISM for VOCs uses a methanol preservation technique when collecting increments.

ISM samples to be analyzed for any VOCs are collected using methanol field preservation. Refer to Section 5.4.1 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 2002c), 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 (see Figure 5-12) and larger narrow-mouthed amber bottles (500–1000 mL) with Teflon-lined caps (see Figure 5-11) are required for ISM VOC sampling. Close coordination with the analytical laboratory regarding ISM VOC bottle/ preservation requirements, sample kit preparation, sample receipt requirements, etc. is essential.

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 should be collected in the same manner as the ISM VOC samples, e.g., a second increment collected at each ISM increment location placed in an unpreserved wide-mouth container (4 ounces 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.

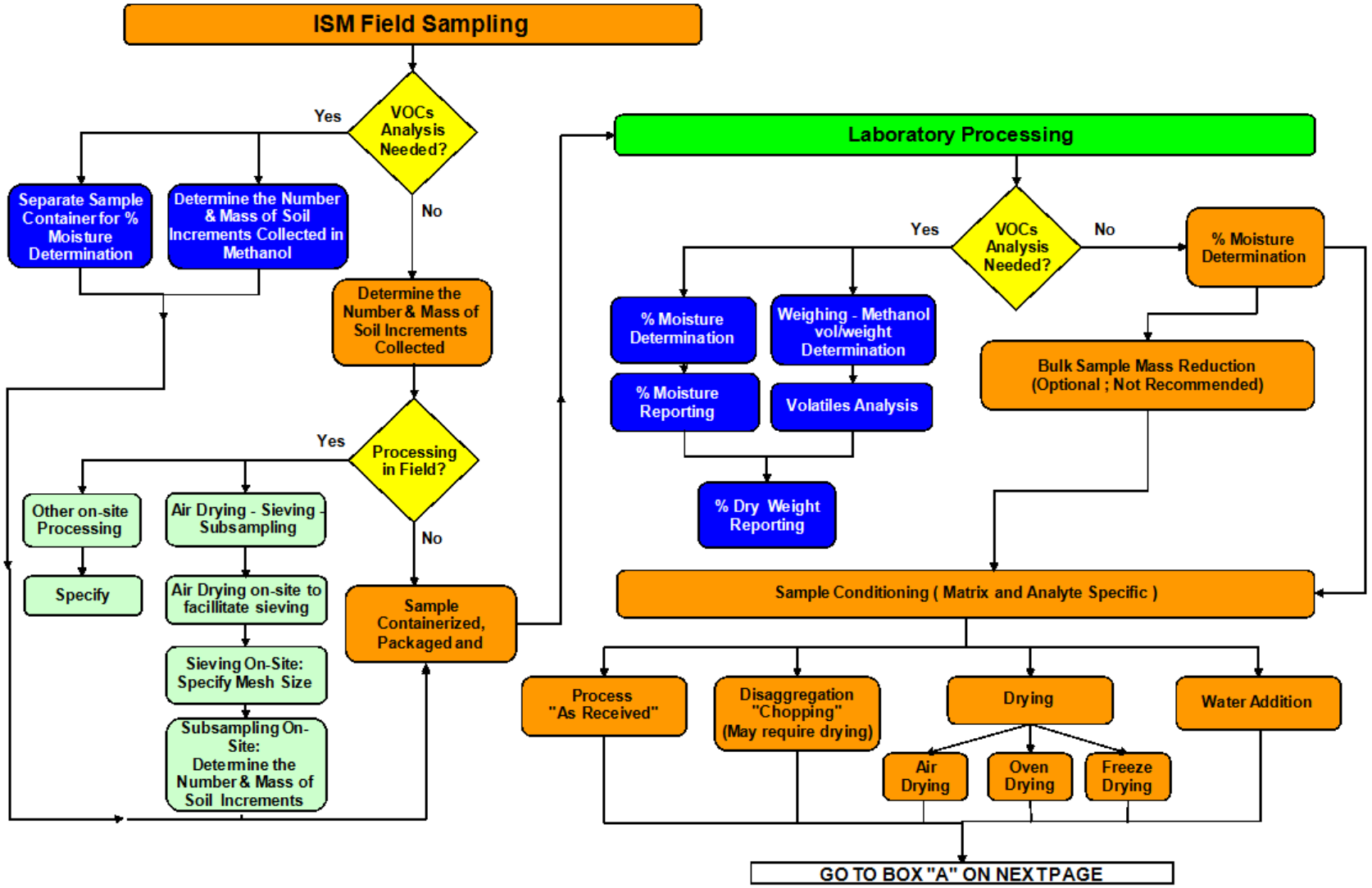


Figure 6-1. Sample processing and analysis flow chart.

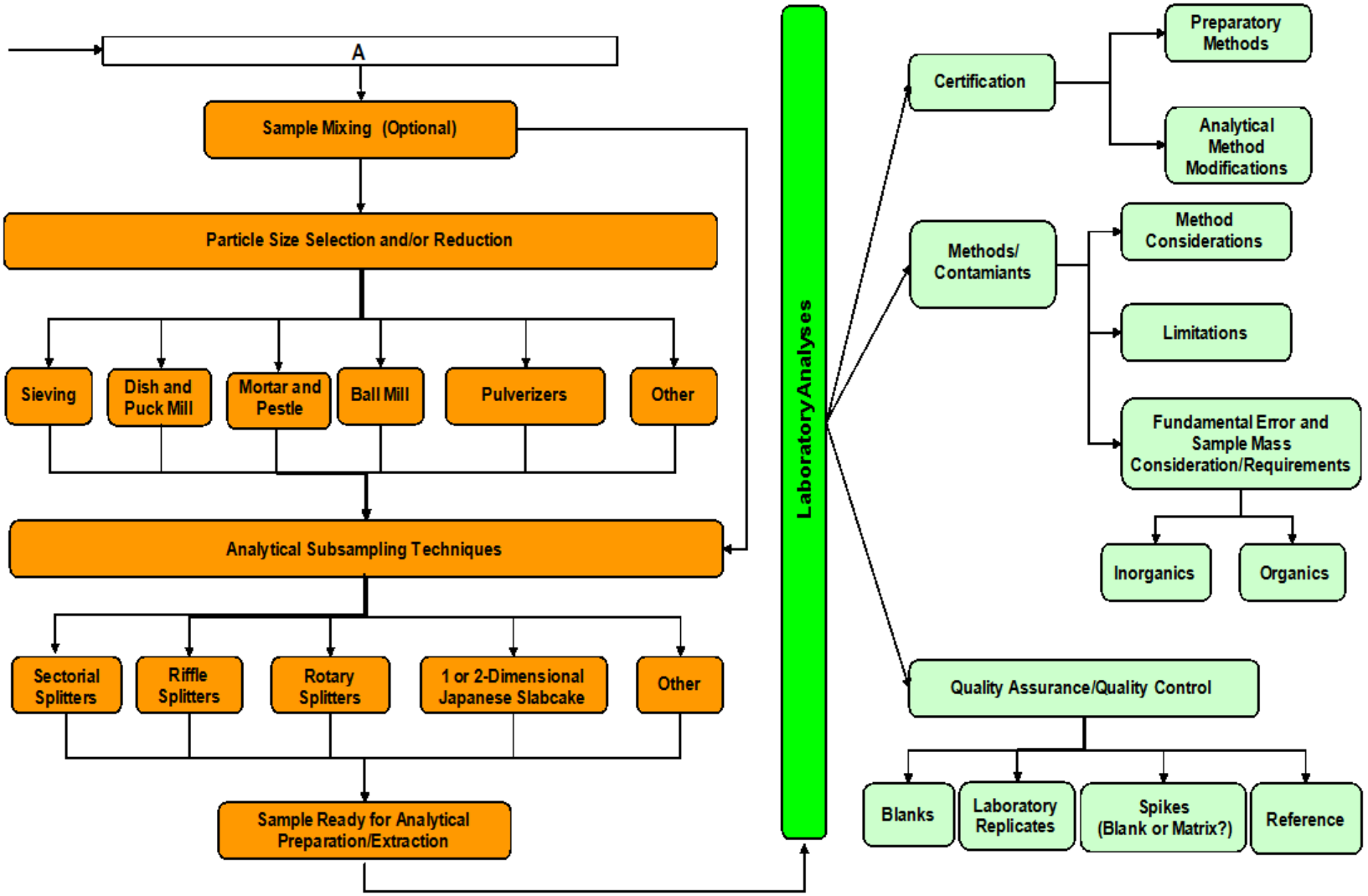


Figure 6-1. Sample processing and analysis flow chart (continued).

Analysis of the methanol extract should proceed according to USEPA SW-846 Method 5035A and the applicable analytical method for VOCs (e.g., USEPA SW-846 Method 8021B, 8260C, etc.).

6.2.2 Semivolatile Organic Compounds and Inorganic Contaminants of Concern

Choosing the specific laboratory processes to handle incremental samples is influenced by the specific COPCs and the objectives of the characterization process. In particular, the volatility and thermal stability of the contaminants affects the options for sample conditioning and the objectives for risk assessment influence the choices related to particle size reduction. USEPA SW-846 Method 8330B for explosives is a good detailed example of the complete laboratory process for the analysis of an ISM soil sample (USEPA 2006c). The remainder of this section provides guidance on adapting the principles displayed in USEPA SW-846 Method 8330B to other COPCs.

6.2.2.1 Identification of sample

During systematic planning determine materials (e.g., vegetation or oversized materials), if any, that should be excluded from the sample.

Occasionally the sample container includes objects that are not to be considered part of the sample. DQOs should direct whether vegetation, oversized material, or decantable water are to be included or excluded from the sample. Decantable water can be poured off the top of the settled sample. 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 most common sieve size for ISM samples is <2 mm (standard #10 sieve), but specific objectives may necessitate a smaller or larger sieve. The excluded materials can be documented via photographs and weight removed when appropriate.

6.2.2.2 Bulk sample mass reduction via sample splitting

Bulk sample splitting is not recommended and its limitations should be addressed during systematic planning if these techniques are to be considered.

Paired ISM sample collection 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. 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 the fundamental error (FE).

Splitting the bulk ISM sample into two or more portions may be necessary if only one ISM sample is collected and two (or more) different sample handling processes are required to optimize for different contaminants. For example, targeting both explosives and SVOCs from subsurface samples requires air-drying prior to the explosives sample preparation, but air-drying may not be appropriate prior to SVOC analysis due to the potential volatilization losses of the lighter, low-boiling-point compounds. It is recommended that these techniques be performed in a

controlled laboratory setting. If appropriate, splitting of an unprocessed bulk ISM sample may be accomplished with alternative shoveling, fractional shoveling, cone and quartering, or other splitting or subsampling techniques that may be appropriate depending on the nature of the soil matrix, contaminant, and DQOs (see Section 6.2.2.7). These techniques may not always be effective (Gerlach and Nocerino 2003; Petersen, Dahl and Esbensen 2004; Gerlach et al. 2002). Splitting of an unprocessed bulk ISM sample is not recommended for solid, particulate type contaminant (e.g., energetic, metals from firing ranges, etc.), as discussed below.

These bulk sample mass reduction techniques might compromise subsample representativeness when the sample is highly heterogeneous (e.g., explosives on training ranges or sample particles with a wide range of density or size). The techniques also increase the FE. Collecting duplicate ISM samples is generally more appropriate than trying to split a single ISM sample. These issues should be evaluated and addressed as part of project planning and DQO process. Specifically, reference Sections 6.2 and 6.3, as well as Sections 2 and 3 of this document as applicable, for ISM considerations that should be evaluated as part of the DQO decision process.

6.2.2.3 Sample conditioning

Sample conditioning prepares the sample for subsequent processing, e.g. sieving, grinding, and subsampling.

Sample conditioning is usually needed before additional mixing or particle size reduction techniques are used because most require a flowable sample. Processing the sample as received is generally the best way to retain the widest range of analytes. There are two primary exceptions. VOCs should be handled as described in Section 6.2.1 since analyte loss even in a bulk, unpreserved sample container is almost certain. Analytes with very high boiling points but biologically degradable may remain stable when the sample is air-dried; however, some of the procedures discussed below should be avoided if any analyte of interest is sufficiently volatile and/or biodegradable to introduce possible biases impacting attainment of project DQOs. A few soil samples are dry enough as received to be handled with the mixing, particle size reduction, and subsampling techniques described below. However, most soil samples require moisture modification before further possessing. Drying the sample to reduce moisture content is the most common approach, but increasing water content can be beneficial in a few selected instances as discussed below. If the moisture content of the original field sample is needed, then use the 2-D Japanese slabcake approach described in Section 6.2.2.7 on the sample prior to any moisture modification.

Air-drying is appropriate when the analytes are chemically stable when exposed to air and have sufficiently high boiling points such that they are unlikely to volatilize during extended air exposure at the selected drying temperature. Drying at ambient temperature (15–25°C) is most common. This process may take up to several days, thus impacting turnaround time. Elevated temperature drying (25–105°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 contaminants. Table 6-1 lists several example explosives and SVOCs, their boiling points, and estimated loss potential during the air-drying step when these contaminants are weakly sorbed to the soil matrix (Bruce 2003). Air-drying produces

crushable soil particles; however, it risks loss of low-boiling-point target analytes. Table 6-1 is not all-inclusive and is intended only as an example for evaluating contaminants and the possible effects of air-drying. Physical property data for additional contaminants is available in *Technical Guidance Manual Notes: Decision Unit and Multi-Increment* Sample Investigations* (HDOH 2011), Tables 2a and 2b. Applying air-drying to contaminants 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 6-1. Potential for loss during the air-drying step

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

Contaminant	Vapor pressure (mm Hg)	Boiling point (°C)	Loss potential
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

After weighing the contaminant 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. This is most likely when lower-boiling-point SVOCs or elemental mercury (Hg) are primary contaminants. Wet sticky samples cause mechanical problems, but coarse sieving and 2-D slabcake subsampling are possible though labor-intensive. See Sections 6.2.2.6 and 6.2.2.7 on 2-D Japanese slabcake for further details.

Place the soil sample on a tray made of, or lined with, a material that is compatible with the contaminant of interest and the drying temperature. The selection of the tray and/or liner material should ensure that the analytes (or interferents) of interest are neither lost nor gained from the sample to the tray and/or liner by sorption or reaction. Aluminum trays and liners should be avoided if aluminum is a contaminant of interest or if it may interfere or interact with an analyte of interest (e.g., chromium, elemental mercury). Plastic trays and liners should be avoided if phthalates and plastic components are contaminants. A paper liner should be avoided if organic carbon or organics that may sorb to paper (e.g., petroleum) are contaminants. Spread the sample evenly in the drying tray. If needed, use 2-D 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 (e.g., 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%–10% is usually acceptable. Wet clay samples should be crushed with a pestle part way through the drying process to avoid formation of large “bricks” that are difficult to handle with subsequent processes. 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. The use of racks to hold the drying trays can facilitate efficient use of space.

Freeze-drying is useful for analytes that volatilize or degrade under extended air exposure or at elevated temperature. Split the soil sample into multiple freeze-drying containers if necessary. Operate the freeze-drying equipment at reduced temperature and pressure (e.g., –80°C, 0.375 torr) for several hours (ERDC 2000). Additional freeze-drying guidance is available from International Organization for Standardization (ISO) (ISO 2005).

Water addition facilitates subsequent processing but can interfere with recovery of some analytes.

Water addition can also be used to produce a mixable sample with less air exposure than during the air-drying processes described above, thus improving retention of low-boiling-point analytes. The added water can interfere with some subsequent sample preparation techniques for high-boiling-point, nonpolar analytes (e.g., solvent extraction of high-molecular-weight polyaromatic hydrocarbons [PAHs]). Place the soil sample in a heavy-duty mixer (e.g., bread dough mixer) constructed of appropriate sample contact materials. Add sufficient reagent water to produce a thoroughly mixed wet paste. Do not add too much water as that will produce a slurry that separates quickly when the mixing process is stopped.

Sample disaggregation is a gentle grinding technique used on dry, crushable soil. It breaks up the soil agglomerates but does not mill small pebbles and other hard particles into smaller particulates, as the particle size reduction techniques (e.g., 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. Disaggregation can facilitate mixing and subsampling. Take the dry sample and crush it on a sieve with a pestle to promote breakup of the soil agglomerates. A variety of sieve sizes can be used depending on the project DQOs. A #10 sieve (2 mm) is the most common size. Alternatively, the soil can be disaggregated using a bladed "coffee" type grinder or blender. 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 the 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 is also commonly used prior to complete particle size reduction using the milling techniques listed in Section 6.2.2.5.

6.2.2.4 Sample mixing

Dry mixing can reduce spatial heterogeneity and facilitate representative subsampling if the sample consists of particles that are similar in size and density. However, mixing samples with large differences in particle size or density can increase stratification and hinder representative subsampling. Tumbling the sample in a container with sufficient headspace is a simple mixing process. The bladed mills and blenders mentioned previously for sample disaggregation can accomplish mixing if they are large enough to contain the whole sample. The same is also true of large-scale mills mentioned in relation to particle size reduction (Section 6.2.2.5). Dry mixing has been used after puck milling and prior to 2-D Japanese slabcake subsampling in USEPA SW-846 Method 8330B.

Wet mixing converts the sample into a thick but mixable paste that does not quickly stratify or separate. See the discussion of water addition in Section 6.2.2.3.

6.2.2.5 Particle size reduction

Particle size reduction can facilitate more representative subsampling.

If the contaminant is present as a solid particulate, particle size reduction through sample grinding can facilitate more representative subsampling by reducing the range of particle sizes and the maximum size present. The determination 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 the FE. See the discussion in Section 2 and Hyperlink 7 on Gy sampling theory regarding the relationship between particle size, uncertainty, subsample size, and FE. Select the appropriate grinding process and equipment to achieve the maximum particle size that was determined in project planning (see Section 3). Many common options are described in USEPA guidance (Gerlach and Nocerino 2003).

Depending on project DQOs, particle size reduction may or may not occur in combination with particle size selection (see Section 6.2.2.6). 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, etc.

Extended high-speed milling can elevate sample temperature due to friction. The thermal stability and volatility of the contaminant(s) should be considered when choosing equipment and a grinding scheme. For example, USEPA SW-846 Method 8330B for nitrocellulose-based propellant residues, specifies a 2-minute (or longer) cool-down period between five 60-second grinding intervals to maintain acceptable temperatures and minimize loss of volatile energetic contaminants.

Milling is not recommended for general purpose application for *organic* contaminants at this time (other than energetics, USEPA SW-846 Method 8330B), as it has not been extensively evaluated. Loss of organic COCs may occur due to increased temperatures during milling, as stated above. Additionally, excessive milling may lead to destruction of organic contaminants, as demonstrated with 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 2005) for additional information. The usefulness of particle size reduction by milling for organic contaminants is usually small because the larger mass (10–30 g or more) normally extracted and analyzed and the particulate "nugget" effect is often minimal. However, "nuggets" can and do occur for specific organic contaminants; e.g., 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 contaminants should be evaluated during project-specific systematic planning.

Milling is recommended for ISM metals analyses. Grinder types and the applicability for the processing of metal particulates are still being evaluated. Care must be taken to avoid the loss of fines during the grinding and transfer process. Milling may also increase some measured metals concentrations when metallic content in the center of the larger particles is subsequently released

through particle size reduction and included the metals analytical process. 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. In some instances it may be possible to meet precision DQOs without milling by increasing the metals subsample size to 10 g. See Section 6.3.3 for further discussion.

Grinder surfaces that contain the contaminants of interest or compounds that may interfere with the analysis of the contaminants should be avoided. If metals are contaminants, then compare the composition of any metal-containing grinding surfaces with the target analyte list. For example, high chromium steel puck mills should not be used when total chromium is a contaminant. Ceramic, agate, tungsten carbide, or low chromium steel grinding components would be more appropriate. The grinder should be checked for contamination by processing a suitable blank material to demonstrate that the grinder does not release contaminants at the concentrations of interest. Confirm that the laboratory has the proper grinding equipment during project setup.

When grinding has been selected as part of the ISM DQOs, the entire conditioned ISM sample is ground. Splitting an unground ISM sample with high heterogeneity due to the "nugget" effect can lead to nonrepresentative subsampling. If the milling equipment is not large enough to process the entire sample at once, then mill smaller portions of the sample and then recombine and mix after the milling step (see Section 6.2.2.4). The milling equipment listed below is not an exclusive list of equipment capable of meeting ISM quality objectives; it is an example list of equipment that has been used successfully in the past.

Mortar and pestle grinding can be accomplished with either manual or automated systems. The large automated systems are recommended because of increased capacity, better reproducibility, and reduced likelihood of repetitive-stress injuries. The sample contact materials can be steel, ceramic, or others depending on the contaminants. The sample is loaded into a heavy walled bowl. The sample is 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. Dry sample is fed into the chute, and ground sample is collected in a hopper beneath the grinding plates.

Ball mills consist of both high-speed and low-speed systems. Typically, the sample is placed in a container along with a grinding medium and shaken rapidly or tumbled slowly. The grinding medium (e.g., steel or agate balls, ceramic cylinders) crushes the sample particles. High-speed systems consist of high-strength containers and high-speed shakers; thus, they can provide more reproducible reduction to <100 μm particle sizes. Typical grinding time for high-speed systems is a few minutes. The low-speed systems typically consist of single-use cans, a grinding medium, and a low-speed tumbler or roller. Roller mills or paint can shakers are common examples. Typical grinding times are several hours; however, excessive overgrinding should be avoided due to possible analyte loss.

Dish and puck mill (shatter box) grinding is described in USEPA SW-846 Method 8330B (USEPA 2006c). The sample is loaded into the dish with puck inserted. If the dish is not large enough to process the entire sample at once, then grind smaller portions of the sample and then recombine and mix after the grinding step. The grinding cycle time and cooling period (if necessary) depend on the analytes of interest. An example grind cycle consists of 1 minute of grinding and at least 2 minutes without grinding to allow the dish and sample to cool. This process may be repeated two to four more times, depending on the materials to be ground. 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. The optimal grinding conditions and final particle size for other contaminants might be different than those described in USEPA SW-846 Method 8330B for energetics. Performance for other contaminants should be demonstrated with reference materials or other "known" samples.

6.2.2.6 Particle size selection

Particle size selection can occur at several different points in the ISM sample processing. It can be used as part of the dry sample conditioning or disaggregation process (see Section 6.2.2.3), with the 2 mm (#10) sieve being the most common. For example, a general ISM sample preparation procedure may be to air-dry, sieve to the selected particle size (e.g., <2mm), subsample by the appropriate method (see Section 6.2.2.7), extract, and analyze. Particle size selection via sieving must be evaluated during the systematic planning and DQO determination process. Careful consideration should be given to the particle size of interest and whether it meets the project DQOs. Additionally, sieving may "dilute" contaminant concentrations by removing larger contaminant particles (e.g., lead or clay target fragments from small arms ranges, lead-based paint fragments, etc.), or "enrich" contaminant concentrations by removing larger, less-contaminated fractions (e.g., rocks, pebbles, organic matter, etc.).

To meet some DQOs it may be necessary to sieve moist "sticky" soil or clay samples when particle size selection to exclude coarse matrix particles is required and air-drying is not acceptable. Mechanical sieve shaking is generally not effective; rather the sample must be gently pushed against the sieve screen and extruded through the sieve (see Figure 6-2). At present this is a very labor-intensive process, and the risk of damaging the sieve is high.

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), 75 μm (#200). Alternatively, final particle size suitability can be estimated by touch or visual inspection when less accuracy is acceptable.

Figure 6-2. Example of wet-sieving soil on an as-received basis.

6.2.2.7 Analytical splitting and subsampling techniques

These sample splitting and subsampling options can provide representative aliquots.

After the entire sample has been dried, sieved, or otherwise prepared, a variety of techniques may be employed to complete the incremental subsampling process for target analyte or moisture determination. Of the multiple processes that exist, several are omitted from this section based on their low performance rankings in regard to grouping and segregation and agreement to calculated sampling error as seen in Table 8 of *Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples* (Gerlach and Nocerino 2003). The following techniques are reviewed in this section according to the preferential order in the aforementioned Table 8: sectorial sample splitters, paper cone sectorial splitters, simple incremental sampling (1-D or 2-D Japanese slabcake), and

riffle splitting. Alternative shoveling, fractional shoveling, and cone and quarter techniques are available but generally not recommended unless sample characteristics prevent the first four techniques from functioning properly or DQOs can be met even with these less rigorous techniques (Gerlach et al. 2002).

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. For example, plastic parts containing phthalates should be avoided if SVOC phthalates are contaminants.

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. 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.

As with all aspects of field sampling, coordination should take place between the laboratory and the end data user to determine which method would be most appropriate. Each of these processes may have different biases. In general, projects requiring a greater amount of reproducibility should be processed with the smallest particle sizes and the largest final target masses acceptable (ASTM 2003).

Of particular concern are methods that use small masses such as the 1 g amount typically used in metals digestions. Increasing the initial mass to a minimum of 10 g at a <2 mm sample particle size improves reproducibility. See the discussion of FE in Hyperlinks 7 and 18 for further details. There are generally only two options to reduce the FE: increase the sample size or reduce the particle size. For a typical soil and analyte concentrations of 1 ppm, to reduce the 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.

The following techniques for splitting and subsampling may or may not be appropriate, depending on project-specific DQOs.

Sectorial sample splitting is the preferred process that results in the least sample heterogeneity of the methods discussed. It requires 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 also be adjustable. The sample falls from the hopper into the chutes as they spin. Collection devices such as sample bottles are mounted on the bottom of each chute to receive equal portions of sample. 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 and the resulting subsamples are equal in mass to the initial sample mass/the number of subsamples. If the desired target mass is not achieved on the first split, recombinations of individual splits may be required to achieve a larger final target mass or resplitting 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 the device 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.

Paper cone sectorial splitting achieves a result similar to that of the rotating sectorial splitter and does not require the purchase of expensive equipment, but is far more labor-intensive and more sensitive to operator technique. A square piece of paper is folded in such a way as to have several equally spaced ridges and valleys in a downward conical shape. A funnel is held in one hand and a container holding the sample in the other. The entire sample is poured from the container into the beaker while rotating the funnel around the top of the cone. Individual containers are placed at the base of each valley to receive the sample as it falls.

One-dimensional Japanese slabcake is produced by pouring the sample into a line using at least 20 passes back and forth to distribute the sample particles over the line. A square scoop is cut across the line to remove a subsample aliquot. Combine as many of these aliquots as needed to accomplish the mass reduction (Gerlach and Nocerino 2003).

Two-dimensional Japanese slabcake or incremental sampling is a method that emulates the field incremental subsampling process in the controlled laboratory setting. The entire sample is spread evenly onto a 2-D surface at a depth that can be 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. This technique may introduce more bias than the previous three techniques, as it is impossible to extract an ideal increment (a cylinder or rectangular solid) from a noncohesive soil, even when using a square scoop with vertical sides (the bottom of the slab is underrepresented in the increment).

The laboratory default should be to use 30 increments to build the analytical aliquot. If project-specific planning has determined that other increment numbers are needed to meet DQOs, use them. 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. 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. Repeat the process for as many replicate samples as applicable.

A good example setup is a 20 × 30 inch 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 1–2.5 cm. A scoopula is used to push the sample around and spread it to an even depth and ideally as thin 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 used that minimizes 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. The sides reduce the tendency of particles to fall off the scoop during increment collection. Before taking increments, the target mass should be considered. Each scoop (increment) will ideally represent $1/30^{th}$ of the desired target mass. 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 (see Figure 6-3).

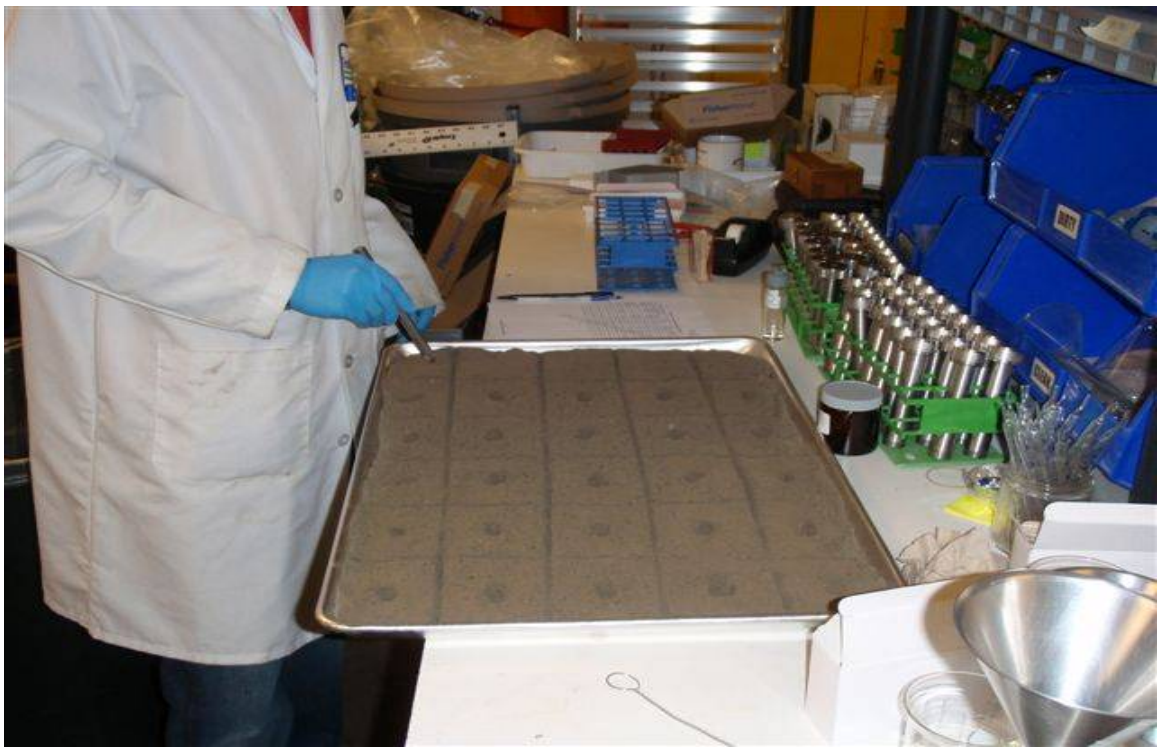


Figure 6-3. Example of 2-D Japanese slabcake incremental subsampling on dried and sieved soil.

The 2-D Japanese slabcake subsampling process may be applied to moist "sticky" samples as well. The best results are achieved with moist sieved soils (see Figure 6-4), but this process can also be applied to as-received samples. Spread the moist soil into an even-depth Japanese slabcake as described above. Use a square-walled, blunt-end scoop with a minimum 16 mm width for 2 mm particle size to collect 30 or more increments to produce the final analytical subsample. Coring tools may also be used for subsampling if the moist sample is sufficiently cohesive. See the tool width discussion in Section 5.2.

Figure 6-4. Example of 2-D Japanese slabcake incremental subsampling on moist sieved, "as-received" soil.

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).

Alternate shoveling divides the sample into two subsamples by placing alternate subsample scoops of the original sample into two separate sample containers (Gerlach and Nocerino 2003).

Fractional shoveling is similar to alternate shoveling except the sample is divided into three or more subsamples (Gerlach and Nocerino 2003).

Cone and quartering splits the sample into two subsamples by pouring the sample into one large cone, flattening the top, and dividing into four sections. Opposite sections of the sample are then combined to form the two subsamples (Gerlach and Nocerino 2003).

6.3 Laboratory Analysis

In general, ISM samples that have been collected and subsampled as described in the previous sections can be prepared and analyzed using standard analytical methodology, such as USEPA SW-846 or other applicable methods for soil matrices. There are, however, contaminant- and/or method-specific considerations that should be evaluated prior to ISM sample collection, preparation, and analysis, as outlined in the following sections.

6.3.1 General Sample Processing Considerations

Other than Appendix A of USEPA SW-846 Method 8330B, there are no reviewed, published SOPs for laboratory processing of ISM samples. Therefore, on a sample- and contaminant-specific basis, the possible effects of ISM sample processing should be considered. Possible

contaminant interference, interaction with, and/or contribution from all equipment used in ISM sample processing should be evaluated during systematic planning and discussed with the laboratory prior to ISM processing and analysis.

Drying, even at ambient room temperature, may contribute to the loss of more volatile and/or photosensitive contaminants. Drying at elevated temperatures is not recommended unless it is documented that the analyte of concern is thermally stable at that temperature. Additionally, ambient drying is not in accordance with accepted methodology of retaining samples cooled or refrigerated (e.g., $4 \pm 2^{o}C$) until extraction. Excessive drying periods for high-moisture soils may contribute to increased biodegradation of contaminants. Finally, increased ISM drying and processing periods should also be considered in regard to method-specific holding times. Holding time violations may result in estimated or rejected data.

Most contaminant concentrations are reported on a dry-weight basis; thus, a subsample for percent moisture determination should be collected using the same techniques as for the contaminant. Subsampling and percent moisture determination should be performed on the as-received ISM bulk sample prior to any processing in the lab and on any moisture-modified samples (intentionally dried or wet, see Section 6.2.2.4), depending on the sample processing decisions made to achieve the ISM project DQOs by the project team during the initial project planning phase. The original sample moisture content can be determined for informational purposes by using the 2-D Japanese slabcake subsampling process to collect a subsample of the as-received bulk ISM sample for percent moisture analysis.

6.3.2 Organics

VOCs

Sample processing considerations for organics include analyte properties, soil interactions, and processing tools.

As previously discussed, ISM VOC sampling involves soil increment collection directly into methanol. In addition to the logistical sampling, shipping and laboratory processing issues, the analytical detection levels of the VOC analytes of concern should be evaluated. Methanol preservation dictates higher analytical detection levels (DLs), MDLs, limits of detection (LODs), limits of quantitation (LOQs), RLs, or PQLs, due to the required methanol dilution into water prior to purge and trap analysis. In general, PQLs and MDLs may be elevated by a factor of 20–50. The use of alternative VOC trap(s) which retain methanol to a lesser degree thus allowing reduced methanol dilution, e.g. “J” or “BTEX” traps, should be considered if appropriate for the volatile contaminant. Reduced methanol dilution will result in lower analytical limits.

If the detection limits from the ISM VOC sample will not meet project DQOs, alternative VOC sampling, such as low-level discrete sampling, should be considered. In some cases, the use of SIM may be necessary to achieve adequate analytical sensitivity in methanol-preserved samples.

SVOCs

In general, the sample processing and drying issues previously discussed should be considered for SVOCs. Analyte loss of the more volatile (e.g., chlorinated benzenes, naphthalene-range PAHs) or less stable SVOCs (e.g., phenols, benzidines, etc.) during the ISM sample processing has not been extensively studied. Large losses of the lower-boiling-point SVOCs are possible during the air-drying step if they are not strongly sorbed to the soil matrix (Bruce 2003). If, on the other hand, the soil is a highly weathered surface soil with many years of air exposure, then weakly sorbed SVOCs have already been lost, and air-drying at temperatures no higher than what the sample has already been exposed to in the field should not result in additional analyte losses. This principle may not apply to subsurface and other samples that have had limited air exposure in situ or to site-specific soils not normally exposed to indoor air temperatures. See Section 6.2.2.3 for a headspace test description to evaluate the potential loss risk due to volatilization.

Sample processing tool materials can contribute contaminants to the sample, particularly in abrasive operations. The composition of plastic tools should be considered, and soft plastics should be avoided when phthalates are of interest. ISM sample processing should evaluate and consider all these process limitations on a chemical-specific basis.

Milling is not recommended for general purpose application with *organic* contaminants other than energetics (see USEPA SW-846 Method 8330B). Excessive milling may lead to destruction of organic contaminants, as demonstrated with 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 2005) for additional information. The usefulness of particle size reduction by milling for organic COPCs is usually small because of the larger mass (10–30 g or more) normally extracted and analyzed and the particulate "nugget" effect is often minimal.

6.3.3 Inorganics

Metals

A number of issues should be considered when evaluating ISM metal samples:

Sample processing considerations for inorganics cover analyte stability and subsample size for digestion.

- While drying temperature and time are generally not an issue for most metals, they should still be considered if certain organometallic compounds, metal species, and/or Hg are analytes of concern, due to compound volatility.

- Volatile metals, such as certain organometallic compounds, metal species, and/or Hg, require careful control to maintain acceptable grinding temperature (see the grind and cool technique described in USEPA SW-846 Method 8330B) or choose a grinding technique that does not elevate the sample temperature during the grinding process. Adjust the subsample size to 10 g if the final particle size is >0.25 mm.

- Milling (grinding) is recommended for ISM metal samples (see Section 6.2.2.5). Dish puck mill–ground samples with particle size <0.25 mm can often meet DQOs with a 2 g subsample. Standard metal digestions are based on 1–2 g soil aliquots.

- Alternatively, if milling is not required to meet project-specific objectives and/or not applicable for the selected COPCs, etc., the recommended minimum mass for unmilled (unground) ISM analysis is 10 g, based on an FE of <15% for a <2 mm sample particle size. If a larger FE meets project-specific DQOs, a smaller subsample mass may be acceptable. Laboratories may need to modify the standard metal digestion procedure or perform multiple digestions (e.g., five individual 2 g digestions) and combine the digestates to account for the increased soil mass. Sample digestion procedures that modify the conventional sample mass to reagent ratios and final sample volumes should be verified for recovery of the metal(s) of interest by successfully digesting a reference material using the modified sample mass to digestate ratio. Alternatively, demonstrate that the results of triplicate 10 g preparations meet the project-specific objectives when compared to thirty 1 g preparations using an ISM site-specific sample.

- Grinding may release metals either naturally occurring or anthropogenic, such as arsenic, mercury, etc., from larger particles resulting in elevated results. These metal concentrations may not be available in the unground sample and, therefore, may not be applicable to the site-specific CSM and/or DQOs.

- The grinding equipment may contribute metal concentrations to the ISM sample. This possible contribution is due to the metal composition of the grinding, crushing, or pulverizing apparatus. Common metals include chromium, cobalt, iron, manganese, nickel, tungsten, etc. Metallic composition analysis is usually available from the manufacturer. For example, avoid high chrome steel when low parts per million concentrations of chromium are of interest.

- Malleable metals, such as lead or copper, may smear in grinding 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.

For additional information and considerations for inorganic ISM samples, see *Implementation of Incremental Sampling (IS) of Soil for the Military Munitions Response Program* (USACE 2009).

6.4 Quality Assurance/Quality Control

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 a certified laboratory following recommended protocols for the class of target analytes (e.g., 4°C for VOCs and SVOCs) to be analyzed. See Section 5.4.1 for ISM field implementation details. Laboratories should have well-trained analyst(s) that follow

Laboratory SOPs for ISM sample processing and analysis should be requested and reviewed as part of the systematic planning process.

documented SOPs while processing, subsampling, and analyzing samples. Laboratory SOPs for ISM sample processing and analysis should be requested and reviewed as part of the systematic planning process. Chain-of-custody, laboratory notes, and completeness reports should accompany all data packages.

QC measures should be implemented both in the field and laboratory. When sample processing is 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 5.4.1). It should be noted that reducing the mass of the sample shipped to the lab will tend only to increase the FE. Depending on the contaminant, field blanks and/or equipment blanks also may be required. Field blanks often are necessary for VOCs and some SVOCs, particularly when a solvent is involved.

General laboratory QA principles apply to incremental sampling methodology samples.

The laboratory must have QA/QC procedures for documenting ISM method performance (i.e., precision, accuracy, method sensitivity), as well as QA/QC procedures for documenting matrix effect(s). At a minimum, these procedures may include the analysis of QA/QC samples such as a method blank, a matrix spike/matrix spike duplicate (MS/MSD), sample replicates, and a laboratory control sample (LCS) in each analytical batch as appropriate. All QA/QC samples should be subjected to the same analytical procedure as those used on actual samples, as applicable to the contaminant and analysis.

Two or more laboratory replicates are recommended to assess the precision of the ISM subsampling.

Laboratory replicates are recommended to assess the precision of the ISM subsampling processes. Generally, two or three replicate subsamples should be collected after all ISM processing is complete. These replicates should then be carried through the rest of the analytical process. The frequency of these replicates can vary from one replicate set per batch to one set per project depending on the project DQOs. Note that there is a difference between replicates collected during sample processing and replicates collected during the field sampling effort. ISM replicates collected from a DU provide information on the variance in the estimate of the mean without specifically separating out the contribution of laboratory sample processing from other sources of variance.

A clean sample matrix, when available, can be used to establish whether equipment used to process samples (e.g., pulverize, split, mix, etc.) has been adequately cleaned between field samples. Clean soil matrices are more likely to be available for organic analytes. For metallic analytes there are no known soil-like matrices that are nondetect at environmental levels of concern for all likely metals. Reagent water or fluoropolymer boiling chips have been used in some instances. A well-characterized soil sample that has low or nondetect concentrations of the critical metals contaminants for a site can also be used to assess equipment and process cleanliness.

The LCS is a known matrix spiked with compound(s) representative of the target analytes. It is used to document possible analyte loss and/or laboratory method performance. LCS control

limits must be established by the laboratory for each ISM procedure and analysis performed and provided in the final laboratory report.

The MS/MSD is an aliquot of sample spiked with a known concentration of target analyte(s). The spiking occurs prior to sample preparation and analysis. An MS/MSD is used to document the bias of a method in a given sample matrix.

Currently, most LCSs and MS/MSDs are introduced after ISM sample processing since it is costly to add surrogate or target analyte spikes of sufficient concentration to a 1 kg or greater ISM sample. Large-scale LCSs consisting of a clean matrix spiked at the laboratory with the analytes of interest and run through most if not all of the incremental processing steps should be considered. The current state-of-the-art materials limit the ability to accurately assess all processing steps. The current materials costs are high when considering 1 kg soil and associated spiking quantities. Additionally, storage and disposal of large volumes of this type of LCS matrix are an issue. Costs are expected to decrease over time if market demand for the material increases. As the technical issues and cost of large spike quantities are addressed, spiking prior to ISM processes may be more common.

Monitoring the air-drying and sieving steps is problematic for SVOCs and/or VOCs. 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. 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 a laboratory prepared control sample that is air-dried can be significantly higher than from a field sample. Theoretically, it is also possible that COCs 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” (i.e., approximately 30 g) clean matrix spiked LCS carried through all of 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. An LCS may not need to be the same size as the ISM sample; it may only require the same preparation and analysis process.

In summary, synthetically fortified soils may not produce the same strength of interactions between the contaminants and the soil particles. Several studies demonstrate that extraction rates for short-term fortified soils can be as much as 10 times higher than weathered “native” contaminants from the same soils (Grant et al. 1995). This phenomenon indicates that 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. However, the number of contaminants covered and the true concentrations are unknown. Neither type of QC material meets all QA goals. The limitations of each should be considered when interpreting the data.

Similarly, due to the bulk mass spiking limitation, modifications may be necessary for MS/MSD analysis. If systematic planning DQOs allow, a small portion (e.g., 100 g) of the as-received ISM sample could be removed using an appropriate splitting or subsampling process (e.g., 2-D slabcake). Before beginning any sample processing, this portion would be spiked with a known concentration of target analytes and then carried through the complete ISM process. This process would increase the uncertainty of the original ISM sample results. 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/MSD approach so as not to remove a portion of the primary ISM sample.

Processing for ISM samples collected for energetic contaminants includes air-drying, sieving, and grinding preparation steps. The associated QA/QC samples for energetics should also be ground. Grinding a sample may generate heat (see Section 6.2.2.5). The size (mass) of a ground LCS must be decided to more closely replicate the heat generated in the matrix samples. A 500 g solid QC standard for energetics is commercially available (e.g., Environmental Resource Associates, Wibby Environmental). This material is often analyzed as an LCS on a per batch basis. The project-specific DQOs should be assessed during systematic planning to determine the appropriate analysis frequency. Additionally, nitrobenzene, 2-nitrotoluene, 3-nitrotoluene, and 4-nitrotoluene have low recoveries when the QC standard is air-dried at room temperatures. The DQO process needs to address whether the QC standard will be air-dried or only ground. There can be significant cost associated with a commercial QC standard. A separate QC standard is available for Tetryl (an energetics constituent) and should be considered if it is a target analyte. The frequency at which the QC standard needs to be processed and analyzed should be defined during the systematic planning process dependent on project-specific DQOs. With respect to energetics, additional guidance for laboratory QA/QC can be found as part of USEPA SW-846 Method 8330B and is available through the Environmental Data Quality Working Group (EDQW 2008) and *DOD Quality Systems Manual for Environmental Laboratories* (QSM) (DOD 2009). Prior to using a grinding step for an ISM project with compounds other than those listed in SW-846 Method 8330B, the associated QA/QC samples must be defined.

ISM samples collected for nonvolatile metals may also include drying, sieving, and grinding preparation steps. It is assumed that this process does not cause the loss of metal analytes; therefore, it may not be necessary to require a large-scale LCS through the entire process. The necessity for the metals LCS (large-scale or otherwise) should be defined during the systematic planning process dependent on project-specific DQOs. If metals contributions from the grinding apparatus are of concern (see Section 6.3.3), a method blank carried through the entire process should be performed.

Monitoring of the effectiveness of the grinding steps for metals, explosives, or other particulate-based analytes would be best demonstrated by adding these analytes in solid particulate form (e.g. 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 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 "grinding" or "milling" portion of the ISM process, largely due to the paradigm shift of grinding from "traditional" sample preparation and analyses. Simply put, grinding or particle reduction is an invasive sample handling technique and, therefore, requires an additional level of QC. Please note that for most organic analytical methods using ISM, particle reduction by grinding or milling is not necessary. The primary purpose for grinding is reduction of the FE by reducing the particle size and eliminating nuggets that can be the cause for extreme heterogeneity. The following methods are known candidates for particle reduction, milling or grinding:

Analyte group	**USEPA SW-846 method**
Energetics (explosives)	8330B
Metals	3050B/6010B, 6020B

Studies are currently being performed by CRREL for evaluating sample processing for metals that may lead to revisions of USEPA SW-846 Method 3050B.

As previously stated, to establish whether the sample processing protocol is achieving the level of precision stated in the SAP, subsample replicates should be taken at a predetermined frequency. Typically, two or more post-processing subsample replicates are collected and analyzed with every batch of 20 samples, with a targeted RPD or RSD as determined during the project-specific systematic planning process. The milling step can be evaluated for analyte losses under USEPA SW-846 Method 8330B with the aforementioned QC standard. Additionally, a separate QC standard is available for Tetryl and should be considered if it is a target analyte. However, caution must be used when using this QC standard since several of the more volatile analytes are susceptible to vaporization losses during prolonged (>1 hour) air-drying exposure at room temperatures (24°C).

Where technically feasible and practical, it is recommended that QA/QC samples that can accurately measure recoveries of target analytes throughout the entire preparatory and analytical process be included with every sample batch of 20 samples as is the current standard practice with discrete samples. As noted, the grinding and air-drying/sieving processes are areas of concern. Because of the highly invasive nature of the grinding procedure involving great force and production of heat, this step leaves a reasonable potential to affect target analyte recoveries, both high and low. Although large analyte losses are not suspected for select energetic (Walsh and Lambert 2006) and inorganic (metal) compounds, to date, limited data are available to verify this hypothesis. Likewise, limited data are available regarding the possible analyte loss due to the air-drying/sieving procedure. Therefore, demonstration of analyte recovery should be performance based and demonstrated through acceptable QA/QC samples.

The long-term goal for technical, method, and material development is to use QA/QC samples that can accurately measure the recoveries of all target analytes throughout the entire preparatory and analytical process with every batch of samples. This is possible for most nonvolatile and high-boiling-point SVOC analytes. The technical issues surrounding lower-boiling-point analytes have not been resolved as of 2010. As QC standards become commercially available for other

analyte groups, they should be incorporated into the laboratory QA program for ISM samples. All information presented in this section should be taken into consideration.

6.4.1 Laboratory Accreditation/Certification

Laboratory accreditation and certification is possible for ISM processes even without reference methods by using laboratory-specific SOPs.

Project teams must be aware of the accreditation requirements that apply to their projects. Accreditation 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. Each combination of matrix (e.g., nonpotable water, drinking water, solid and chemical materials), method/technology, and analyte is considered an FOT.

There are three primary types of accreditation requirements:

- **National Environmental Laboratory Accreditation Program (NELAP).** NELAP is a national program implemented by member states. State governmental agencies usually serve 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.
  Note: Each member state has its own procedures to address accreditation of method modification and internal laboratory procedure.

- **Non-NELAP state accreditation**. 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 (e.g., Alaska).

- **Agency-specific accreditation programs.** Some federal agencies have their own accreditation programs. DOD recently centralized the Environmental Laboratory Approval Program (ELAP). Additional information on DOD ELAP can be found at www.navylabs.navy.mil.

Some accrediting bodies certify laboratories based on the laboratory-specific SOPs, e.g., NELAP, DOD ELAP, and other appropriate accrediting bodies.

Laboratories must demonstrate compliance with the DOD QSM through the DOD ELAP, and the SOPs should be in accordance with the *Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples* (Gerlach and Nocerino 2003). Assessment by DOD ELAP according to the DOD QSM (DOD 2009) covers laboratory procedures for incremental sampling for explosives analysis. Laboratories proposed for analysis of parameters other than explosives should be assessed and approved for ISM sample processing

in accordance with *Incremental Sampling: MIS-Based Laboratory Requirements for the Analysis of Explosives (USEPA SW-846 Method 8330B) and Metals in Solid Matrices* (USACE 2008b). *Standard Guide for Laboratory Subsampling of Media Related to Waste Management Activities* (ASTM 2003) gives guidance on sample splitting, particle size reduction, and the mass of subsample necessary to reduce the FE to <15%.

7. MAKING DECISIONS USING ISM DATA

7.1 Introduction

This section provides guidance on using data generated from ISM samples to make decisions about a DU. Since the data may inform one or more decisions; it is helpful to establish a structured approach to making decisions, referred to here as "decision mechanisms."

"Decision mechanism" refers to the different ways that environmental concentration data can be used to make decisions at a site.

In the context of this discussion, decision mechanisms include the procedures, inputs, and algorithms that are used to aid decision making based on environmental concentration data. The simplest decision mechanism is a comparison of a single measured concentration to an action level. The inputs in this case are the concentration measured in the sample and the action level. The procedure may be, for example, to compare the concentration to the action level to determine whether further sampling, evaluation, or other action is needed. A common example of this type of decision mechanism might be the comparison of individual discrete soil sample results obtained during a CERCLA Preliminary Assessment to Regional Screening Levels (RSLs) for chemical contaminants at Superfund sites. Discrete soil sample results are often compared directly to the RSL benchmarks. Exceedance of a benchmark by one or more discrete soil sample results can be used to identify a contaminant as a COPC.

More complex decision mechanisms may involve procedures that include the use of advanced statistical analysis or numerical models. For example, a surface soil investigation may involve the use of geostatistical modeling or krieging to estimate the distribution and extent of contaminants across a site using high-density discrete soil sampling data. Decision mechanisms may involve a series of procedures that are iterative or progressively more complex.

The specific decision mechanisms that may be needed to make a final decision for a DU should be determined as part of the planning at the start of the investigation as noted below and in Section 3. As discussed below, decision mechanisms that apply to ISM are analogous to decisions with discrete data, and include the following:

- comparison of a summary statistic (e.g., single ISM estimate of the mean, the mean of multiple ISM results, the 95% UCL of multiple results) to an action level
- comparison of results of a quantitative risk assessment which used a summary statistic (typically a 95% UCL) to an acceptable risk range for carcinogens (e.g., 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 data sets
- combination of data across multiple DUs
- extrapolation of statistics across DUs

One of the primary benefits of ISM sampling is that the volume of media to which a decision will be applied must be determined prior to sample collection. It is also essential to have an understanding of the manner in which ISM samples will be used to make decisions during project planning. Decision mechanisms must be consistent with the rationale behind the sampling plan design, as discussed in Section 3, and should be based on the following:

- CSM
- goals of the project and end use of the data
- scale of the decision
- requirements for precision, total error, and decision quality
- assumptions of the statistical method(s)
- anticipated and/or measured degree of variability within the DU

Although the primary component of the decision mechanism is the actual procedure, algorithm, or statistical test employed to evaluate the data and make the decision, such variables as the location of the sample, the number of samples or increments involved, and the rationale behind the action level must be considered. The following are important aspects of decision mechanisms that must be included:

- number of, rationale for, and size of DUs and SUs
- number of SUs composing each DU (from one to many)
- number of increments collected to form each ISM sample
- bulk mass of ISM sample
- mass of analytical subsample
- aspects of "correct sampling" (Pitard 1993)
- number of replicate ISM samples in each SU
- particle size reduction or selection (where appropriate)
- statistic calculated
- source, nature, and numerical value of the action level

ISM samples can be used for a number of different applications. The type of decision mechanism employed must be consistent with the type of decision being made. The specific size and location of DUs are guided by site knowledge regarding the spatial distribution of contaminant(s) and the movement or behavior of receptors that may contact different areas of the DU. Estimates of mean concentration provided by ISM can be useful in evaluating risk from direct contact with soil, where a DU is designated to correspond with a presumed or actual exposure area for human health or ecological risk assessment. Likewise, because most soil-to-groundwater leachate models assume a volume of contaminated soil as the source of contamination to the groundwater, estimates of mean concentrations in targeted volumes of soil are directly applicable to assessment of soil concentrations using soil-to-groundwater leachate models. Another useful application of

ISM is when multiple decisions must be made for very large volumes of soil, for instance, when large former agricultural fields or dredge piles are intended for future residential uses. ISM has been used for the exploration of concentration gradients because, in the presence of small-scale heterogeneity, ISM provides a better understanding of contaminant distribution than a few widely spaced discrete samples. ISM samples may also be used over concentric SUs surrounding a suspected source area. Finally, a variety of different strategies may be used in subsurface investigations with ISM samples.

An ISM-based sampling project should be tailored to the decisions for which the data will be used. Careful planning is the key to ISM data usability.

Regardless of the decision mechanism, the standard steps of data quality assessment as discussed in Section 3 apply. After data are collected, it is important to revisit the CSM and determine whether it is supported by the data or should be modified. Methods for statistical analysis of data should be selected based upon the sampling design and project objectives. Key underlying assumptions associated with the statistical test must be identified and determined to be valid for the data to be analyzed.

7.2 Decision Mechanisms

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. In other cases, 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. This section discusses the benefits and drawbacks as well as the assumptions involved in several decision mechanisms available for ISM sampling. Although decision mechanisms 1–3 are cast in terms of comparison with action levels, the same considerations apply when using ISM data to develop exposure point concentrations for baseline risk assessments.

7.2.1 Decision Mechanism 1: Comparison of One ISM Sample from the DU to the Action Level

The simplest decision mechanism is the comparison of a single ISM sample result for a DU to an action level, which is typically a threshold value derived through risk assessment, regional background estimate, or other regulatory means. Sometimes, more than one set of action levels may apply to a site because they reflect different objectives (e.g., protection of acute and chronic health end points). 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 action level.

This decision mechanism is simple and straightforward. The result of the decision is immediately apparent; a failure is indicated if the ISM sample result exceeds the action level. This approach is frequently used with individual discrete samples under the CERCLA preliminary assessment process; however, when the action level is intended to represent a mean concentration for a risk assessment exposure area, it is more logical to compare an estimate of the mean concentration (e.g., 95% UCL) in the DU from an ISM sample to the action level or, similarly, the mean (or 95% UCL) of multiple discrete sample results.

A single ISM sample provides one estimate of the mean concentration, which may be above or below the actual mean concentration in the DU. Unless replicate ISM samples are taken, there is no indication of the extent to which estimates of the mean vary, and consequently, it is difficult to predict how far from the actual mean a single ISM sample result might be. This uncertainty greatly limits the scientific defensibility of this approach. Use of a single ISM result might be acceptable when the estimated mean concentration obtained is much greater than, or much less than, the action level such that even a great deal of error in the mean estimate could be tolerated without making a decision error. In this situation, the ISM sample provides confirmation of what may have already been strongly suspected—that the DU clearly passes or fails. However, when the ISM sample result is close to the action level, this decision mechanism is unreliable, and decision errors in both directions are possible (i.e., concluding that the DU fails when the average concentration is in fact below the action level, or vice versa). How big a difference from an action level is big enough to conclude confidently that a DU passes or fails? The problem with this approach is that there is no clear answer. Because only one concentration is available, there is no indication of the potential magnitude of error, and a decision that a concentration difference is large enough that a decision error will not be made is arrived at subjectively. Obviously, uncertainty about making the right decision increases as the ISM sample result gets closer to the action level. Comments from many states suggest that the uncertainty associated with making decisions with only one ISM sample would make this approach unacceptable.

Comparison of a **single ISM sample result** to an action level is useful when the ISM result is either well above or well below the action level and error in the estimate of the mean is unlikely to lead to an incorrect decision.

Decision Mechanism 1 example

A single, 5000 ft^2 DU is established across an area suspected to be a former transformer dump site. The objective of the investigation is to determine whether the mean concentration of PCBs in surface soil (0–4 inches bgs) exceeds an action level of 0.22 mg/kg for residential land use. Forty increments of soil are collected using systematic random sampling and combined into a single ISM sample for sample preparation and analysis. The reported concentration of PCBs in the sample is 6.2 mg/kg. The result is substantially higher than the action level, and the planning team comes to consensus that plausible error in estimating the mean would not likely change the ultimate decision that the DU fails.

7.2.2 Decision Mechanism 2: Comparison of the Mean of Replicate Data from the DU to the Action Level

In this decision mechanism, replicate ISM samples are collected in the field from the same DU. These replicates 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 action level. The mean concentration from replicate samples is likely to be closer to the true mean of the DU than the result from a single

Comparison of the **mean of replicate ISM sample results** to an action level is most useful when quantifying the uncertainty in the mean is not an important element to the decision.

ISM sample (see Section 4) and could therefore be considered to provide a more reliable estimate of the mean. There is no assurance, however, that the actual mean concentration has not been underestimated. Consequently, this decision mechanism would not be useful in circumstances where project objectives dictate that estimates of mean concentrations must be conservative (e.g., EPC values in most USEPA human health risk assessments).

Decision Mechanism 2 example

This example is similar to the one provided for Decision Mechanism 1, except three replicate samples are collected from the DU. The reported concentrations of PCBs in the replicates are 0.12, 0.16, and 0.26 mg/kg. The mean concentration of PCBs based on the three replicate samples is 0.18 mg/kg. This does not exceed the action level for residential land use of 0.22 mg/kg, so it appears that no further action may be warranted. Note, however, how close the estimate of the mean is to the action level. This fact may have important implications for decision making. This same example data set is used again in the example for Decision Mechanism 3 to further illustrate these implications.

7.2.3 Decision Mechanism 3: Comparison of the 95% UCL on the Mean of Replicate Data from the DU to the Action Level

Project objectives may specify that the estimate of the mean concentration provided by ISM sampling must be health protective, meaning that there is low chance of underestimating the actual mean concentration within the DU. 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. 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 a conservative estimate of the mean (e.g., a 95% UCL). A similar approach can be used with ISM data as discussed below.

For those accustomed to working with 95% UCL values from discrete data sets, there are some important differences with 95% UCLs from ISM data. As discussed in Section 4, calculation of a 95% UCL for ISM data requires a minimum of three ISM samples. This is fewer than is required for discrete data sets to yield reliable 95% UCL values. Additional ISM replicates increase the performance of the mean estimate (i.e., provides a 95% UCL closer to the actual mean), and although this increases the cost, it may be worthwhile if the site is relatively heterogeneous and the result is anticipated to be close to the action level. 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 data sets, the maximum concentration observed is often used as the EPC if it is less than the calculated 95% UCL. With ISM data, the calculated 95% UCL value should always be used as the EPC 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 UCL *always* exceeds the highest individual ISM result.

Two methods for calculating the 95% UCL from ISM data are available: Student's-*t* and Chebyshev. As discussed in Section 4, the choice of method depends on the known or anticipated shape of the probability distribution of contaminant concentrations in the DU. Note that software

programs for calculation of 95% UCL values for discrete sample data (e.g., ProUCL) contain algorithms optimized to perform well for discrete data only. They are generally unsuitable for calculation of 95% UCL values for ISM data. A calculator for deriving 95% UCL values for ISM data is provided in Section 4.

Comparison of the **95% UCL on the mean of replicate ISM results** is most useful when the chance of underestimating the true mean must be minimized.

When replicate samples are taken over the entire DU, each is a true replicate and provides a separate estimate of the mean concentration. These estimates can be combined to derive a 95% UCL. Another approach is to divide the DU into SUs and take one ISM from each. The results from each ISM sample (i.e., 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. The principal disadvantage to this approach is that the UCL often exceeds the true mean by a larger degree than if replicates had been collected across the entire DU. 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 action level, this spatial information could be valuable if a decision is made to break the DU into smaller DUs for reevaluation. (Note: The single ISM results from each SU would not be adequate to make confident decisions regarding them. Systematic planning would be needed to establish the smaller DUs and resampling would be required.)

Decision Mechanism 3 example

This is similar to the example for Decision Mechanism 2. The same three replicate samples are collected from the DU with reported concentrations of total PCBs of 0.12, 0.16, and 0.26 mg/kg. The 95% UCL of these results is 0.30 mg/kg with the Student's-*t* method and 0.36 mg/kg with the Chebyshev method, both of which exceed the action level for residential land use of 0.22 mg/kg. Therefore, while the sample arithmetic mean is less than the action level (as we saw in the previous example), there is sufficient variability in the results that there is a relatively high likelihood that the true mean exceeds the action level. Uncertainty in the shape of the underlying distribution does not factor into this result, since both 95% UCL methods yield the same conclusion. Options in this situation include deciding that the DU fails or taking more samples to reduce uncertainty and lower the 95% UCL, perhaps to a value below the action level.

7.2.4 Decision Mechanism 4: Comparison to Background

Background data from an appropriate reference area are used to evaluate site data for many environmental projects. With discrete sampling, comparisons between site and background data are generally done in one of two ways: point-by-point comparison of site data to an upper bound of background conditions (e.g., UTL) or distributional comparison using hypothesis tests to determine whether the differences in the central tendency (i.e., mean or median) or upper tails are statistically significant. USEPA guidance on hypothesis testing (e.g., USEPA 2002c, 2007, 2009) was developed with discrete sampling in mind and includes the following elements:

- Set the null hypothesis to state that the central tendency (e.g., mean) for the site distribution is statistically greater than that of the background distribution. This places the "burden of proof" on the data to show that site concentrations are not greater than background and is considered a more conservative (health-protective) approach.

- The use of nonparametric procedures such as the Wilcoxon rank sum (WRS) test relax the assumption of normality but not the assumption of equal variance. Therefore, it is possible that a test outcome is influenced more by differences in variance than by differences in central tendency, for example. For this reason, statistical tests should be accompanied by exploratory graphical analysis (e.g., histograms) to support the overall conclusion regarding background/site comparisons.

- Welch's test (also called Satterthwaite's-*t* or the unequal-variance *t*) is a modified Student's-*t* test that attempts to correct for unequal variances, though it still requires the assumption of normality. Simulations suggest that results are robust to moderate deviations from normality (i.e., moderate asymmetry).

- Both central tendency and upper tail tests should be evaluated to determine whether background and site concentrations are significantly different. A difference in either may suggest significant difference from background. The emphasis on the use of upper tail tests is that it is informative to understand whether subareas of the DU are elevated compared to background.

- Decision errors and determinations of statistical significance are closely tied to sample size and distribution shape, as well as the specified significance level (e.g., $\alpha = 0.01$, 0.05, 0.10, etc). When sample sizes are small for either data set, a formal statistical test may not be appropriate. For example, using WRS with $n = 4$ in both data sets and $\alpha = 0.01$, one can *never* identify a significant difference between two populations. This principle is true no matter what the sample concentrations are, even if all four site measurements are larger than background. WRS requires at least $n = 5$ in a group, or a higher (less-protective) level of significance (e.g., $\alpha = 0.05$ or 0.10).

As discussed in Section 4.4.3.3, ISM results are not suitable for point-by-point comparison with UTLs generated from discrete sample background data because ISM and discrete data sets have fundamentally different characteristics. If background and site data are both generated using ISM, comparisons of central tendencies (e.g., medians) can be made using hypothesis testing, but statistical power to detect differences will be low due to the limited number of replicates in most ISM data sets. Similarly, at least $N = 8$ observations per group is desired before using hypothesis tests to compare upper tails (e.g., quantile test). Nonetheless, hypothesis tests are not the only tool available to determine whether there are important differences between site and background distributions. Simple graphical analysis can provide useful information and serve as a semiquantitative means of comparison.

Decision Mechanism 4 example

Continuing with the example presented in Decision Mechanisms 1–3, five replicate samples are collected from a reference area unimpacted by site contamination for comparison with the site data. The reported concentrations of benzo(a)pyrene from the reference area ISM samples are 0.05, 0.10, 0.12, 0.20, and 0.40 mg/kg. The sample mean and SD of the reference area samples are 0.17 and 0.14 mg/kg, respectively. By comparison, the site sample ISM replicate results are 0.12, 0.16, and 0.26, and the mean and SD are 0.18 and 0.07 mg/kg, respectively. Therefore, the sample means are almost the same, but the SD is greater in the reference area by a factor of 2.

Figure 7-1 provides a graphical comparison of the two ISM data sets using side-by-side dot plots. For context, the action level for benzo(a)pyrene of 0.21 mg/kg is also shown on Figure 7-1. Presenting the information this way, it is clear those concentrations in the reference area exhibit greater variability and that the difference may be partly explained by the difference in sample sizes. If more ISM replicates had been collected at the site, then perhaps more extreme high and low concentrations would have also been observed.

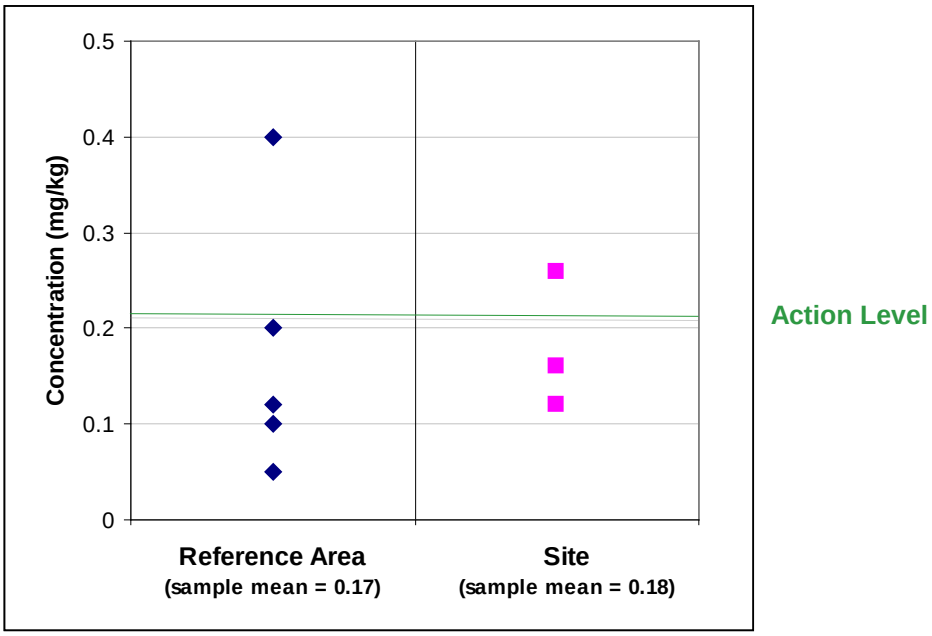


Figure 7-1. Dot plot comparison of background (reference area) and site ISM results.

Since the sample sizes are too small to evaluate a GOF to a normal distribution, a secondary line of evidence may be provided by hypothesis testing (noting the limitations in applying these tests as discussed above). For purposes of this example, a nonparametric WRS test ($\alpha = 0.05$) was applied. Using a one-sided null hypothesis specified as site median less than or equal to the background median, one would not reject the null hypothesis ($p = 0.33$) and, therefore, conclude that the distribution on site is comparable to background. By contrast, using a one-sided null hypothesis specified as site median greater than or equal to the background median, which is consistent with USEPA guidance (e.g., USEPA 2002c, 2007, 2010a), one would again fail to reject the null ($p = 0.77$) and, therefore, conclude that the distribution on site is elevated with

respect to background. The result is contingent on the form of the hypothesis test that is selected. Since the latter hypothesis puts the burden of proof on the data to demonstrate that the distributions are comparable, the small sample sizes from ISM data sets very often yield a conclusion that site > background even when the ranges overlap as shown in this example. Therefore, statistical significance should be interpreted with caution.

7.2.5 Decision Mechanism 5: Combining DUs

There are circumstances when it may be advantageous to combine results from two or more DUs into a larger DU. This situation might occur when there are multiple sampling objectives for a given area. For example, delineation of source areas might necessitate creation of several small DUs, while evaluation of risk from exposure is based upon a DU that encompasses two or more of these DUs. DUs can be constructed in such a way as to meet both objectives efficiently if results from smaller DUs can be combined to produce an estimate of the mean for a larger, "super" DU. In constructing the "super" DU, each of the smaller, component DUs is in a sense like a SU. However, all are true DUs in that a decision must be reached for each, based upon one site objective or another.

Another example is a situation in which sampling objectives require assessment of exposure of different receptors or scenarios such that differently sized, superimposed exposure areas must be evaluated. Here again, the ability to combine results from small DUs to estimate mean concentrations for larger DUs would be advantageous.

> Combining results from two or more small DUs to estimate the overall mean concentration in a larger combined DU is advantageous when the data must support more than one decision (e.g., overlapping exposure units for ecological and human health receptors).

Operationally, the mechanism requires a stratified sampling plan. The overall mean of the larger DU can be calculated using replicate data from the smaller, component DUs using formulas described in Section 4. These formulas take into account the size of the smaller DUs, weighting their contribution to the larger DU accordingly. The ability to combine DUs extends vertically as well as horizontally; that is, results from DUs from different soil depths can be combined if needed to meet sampling objectives.

Decision Mechanism 5 example

An elementary school is divided into three DUs based on anticipated exposure of students and maintenance workers to soil. The kindergarten children have their own playground that is designated as DU1. The older children have another playground that is designated as DU2. School maintenance workers come in contact with soil from both DUs equally, and their area of exposure is DU3, which consists of DU1 + DU2. DU1 and DU2 are each sampled using systematic random sampling with a total of three ISM samples from each. The results from the six ISM samples are combined, with appropriate weighting as described in Section 4, to derive the average concentration for DU3. The weighting factors applied to each DU result should reflect the assumptions in the CSM.

7.2.6 Decision Mechanism 6: Extrapolating from Sampled to Unsampled Areas

This decision mechanism entails using estimates of the mean obtained from areas where ISM samples are taken to make decisions regarding other DUs that are unsampled. The fundamental assumption made with this mechanism is that the distributions of contaminant concentrations in the unsampled areas are essentially the same as in the sampled areas. The most common rationale for this assumption is that the source of contamination, mechanism(s) of transport, etc. are similar for each of the areas and that these conditions should lead to similar levels of contamination and similar variances. This decision mechanism is typically considered when large tracts of land or large volumes of soil must be assessed with a limited budget.

> Extrapolation from a sampled area to an unsampled area requires an assumption that the distributions of contamination in the unsampled areas are sufficiently similar to the sampled areas.

The key to this decision mechanism is confidence that the fundamental assumptions are valid and that there are no significant differences in contaminant distribution among the sampled and unsampled areas. In the absence of data to verify the assumption, that confidence is subjective. There is nothing unique about ISM that enables this extrapolation with reduced uncertainty—the same issue of whether or not to extrapolate exists whether the sampled areas are evaluated with ISM or discrete samples. Based on feedback obtained in development of this report, this decision mechanism is not acceptable for many states.

A distinction should be made between extrapolation between DUs and extrapolation within a DU. It is sometimes suggested that because there is precedence for using results from discrete samples to make inferences about unsampled areas within a DU, the same uncertainty applies to ISM. In this context, there is a difference between how information from discrete and ISM data may be used. With discrete data, spatial interpolation methods (e.g., geostatistics, inverse distance weighting) or discretization methods (e.g., Thiessen polygons) can be used to provide more reliable estimates of the mean and standard deviation throughout the DU. These methods also have the advantage of using information across DUs (i.e., when a site is split into multiple DUs) to derive estimates of the mean and standard deviation within each individual DU. With ISM, this degree of spatial resolution is lost because the increments are composited, so there is no basis for estimating concentrations in subareas of the DU or for developing a mathematical model that uses data from across the DUs. One exception would be for a site that is divided into many DUs—if a sufficient sample size is available, each estimate of the mean may be considered representative of a portion of the site such that spatial patterns and interpolation method may be explored.

A variation on this approach is to collect replicates in subset of the DUs and extrapolate the estimate of the variance (or the CV) to DUs with a single ISM sample. Although this approach appears to be a less uncertain way to extrapolate findings among DUs, the extent to which the distributions may be comparable across DUs must be considered. The chance that the distributions differ among DUs increases as the number of sources and the complexity of the contaminant transport mechanisms increase. In addition, sites with multiple subareas of elevated concentrations can be expected to introduce inherent variability within and between DUs, making

a successful extrapolation of the variance more difficult. In general, the greater the number of DUs where replicate ISM samples are collected, the more likely that the average measure of variance will be representative of DUs with single ISM results (see Section 4.2).

As noted in Section 4, it is unclear whether the appropriate statistic for extrapolation is the SD or CV. The CV is preferred if it can be reasonably assumed or demonstrated that there is a positive correlation between the mean and SD. Based on the proportionality effect, the mean and SD are expected to be positively correlated for positively skewed distributions (Goovaerts 1997). If replicate data are available for multiple DUs, plots of the SD vs. the mean should be developed to explore patterns in the relationship between the sample statistics.

A related situation exists when a DU is subdivided into SUs and only a fraction of the SUs are actually sampled. In this approach, the results from each of the sampled SUs are compared with the action level(s). If all are lower than the action level(s), the entire DU passes. The same assumptions and considerations discussed above apply in this situation as well. If one or more SUs are above the action level, the DU does not pass, and the systematic planning team should be reconvened to plan the next steps, which may include additional sampling.

7.2.7 Decision Mechanism 7: Evaluating Oversized DUs

An oversized DU is one that is larger than can be justified based upon site objectives but is evaluated nevertheless because of practical considerations. While oversizing DUs is strongly discouraged, there are some situations where it is unavoidable. Examples include DUs that are larger than the home range of some of the species of interest in an ecological risk assessment or are larger than the exposure area for some of the receptors/risks of interest in a human health risk assessment, such as when acute exposure to soil is a concern. In this situation, the DU evaluated actually consists of a few to perhaps thousands of smaller, ideally but impractically sized DUs. The problem faced is determining what information the sampled DU can provide concerning concentrations in the smaller sub-DUs.

There are no optimal answers to solve this dilemma, unfortunately, because typical ISM sampling designs are devoid of spatial information on contaminant distribution within the DU. This is not a new problem, as it has been documented in the literature for composite sampling and there are a number of approaches for estimating high-end concentrations within a sampled area. The simplest of these is to multiply the mean value from the composite (or ISM sample) by the number of increments. This approach represents the situation in which all of the contaminant is present in one of the increments. Given the number of increments in a standard ISM design, this approach is *extraordinarily conservative* and can yield quite high values. Given the conservative nature of this method, it is useful only to support "no further action" decisions or decisions to characterize the area further. Other approaches that are less conservative include multiplying the average concentration by the square root of the number of increments or more complicated formulas (for an example, see Barnett and Bown 2002).

Note: A computationally equivalent approach is to use the average concentration but divide the soil *action level* by the number of increments.

7.3 Assessment of Error

It is desirable to seek quantitative information on the potential magnitude of error in ISM data when using those data to make decisions. In all environmental sampling, two basic types of error are produced:

- error associated with the collection of sample(s) in the field
- error associated with the processing and analysis of those sample(s)

The ISM approach, which includes both field and laboratory steps, is intended to minimize the potential error and produce more technically defensible data by specifying the targeted volume and the parameter to be estimated and collecting, subsampling, and processing the sample(s) in accordance with the recommendations of sampling theory. An important component of this process is, to the extent possible, to assess the errors generated in the sampling and analysis scheme from beginning to end (i.e., from collection of soil in the field through the production of an analytical result).

One means of evaluating ISM data is through comparison of the results of replicates, both as taken in the field and in the laboratory. As discussed in Section 3.6, field replicates consist of separate ISM samples taken by the field team from the same area (SU or DU). They are not field splits—they are collected and processed as separate samples. Laboratory replicates are samples taken from a single ISM sample, usually in the laboratory. They can be taken from the bulk ISM sample at a number of points during sample processing, depending on the process step(s) being evaluated. Replicates taken at the beginning of laboratory processing of the bulk ISM sample are used to evaluate potential overall error resulting from laboratory processing and analysis.

Replicate ISM samples collected from each DU in the field provide a measure of total sampling and analysis error.

Use and interpretation of replicate data depends in part on the decision mechanism being applied. For example, field replicate data allow calculation of 95% UCL values needed for Decision Mechanism 3 and allow statistical comparison of site and background results using hypothesis testing in Decision Mechanism 4. Decision Mechanisms 5–7 can also rely on 95% UCL values calculated from field replicates.

Three or more ISM samples are needed to calculate a defensible 95% UCL.

Replicate data can also be used to calculate an RSD, which is used to evaluate the precision of the data. RSD is a measure of reproducibility of estimates of the mean provided by replicates. Just as the sample mean and standard deviation are estimates of the corresponding population parameters, the sample RSD is an estimate of the ratio of the population parameters. It provides a measure of the total error associated with the data, although not necessarily the accuracy of the estimate. To calculate appropriate statistics, at least three field replicate samples are needed. Ideally, the project team then designates one of these replicates for separation into laboratory replicates. Replicate RSD data are intended to quantify the total error

The precision of ISM data can be quantified from replicate ISM sample results.

of the measurement system and attribute that error to either field sampling or laboratory procedures.

The total error is estimated based on the field replicate RSDs. Laboratory error can also be estimated based on the laboratory replicate RSDs. The field sampling component of error can then be estimated by subtracting the laboratory error from the total error. Therefore, the collection of field and laboratory replicates allows the error to be attributed to either the laboratory or the field sampling processes.

High RSD values for the laboratory component indicate potential problems with laboratory subsampling of the bulk ISM sample or other sources of analytical error. In this situation, the source(s) of laboratory error should be investigated and resolved.

High RSD values for the field component can have different implications depending on the decision mechanism being applied. For example, a high RSD (e.g., exceeding 30%–35%) from field replicates, but with acceptable RSDs from laboratory replicates, strongly suggests a substantial degree of heterogeneity in the DU contaminant concentrations. For Decision Mechanism 2, where a simple average of the replicates is used to derive the average concentration, this situation represents a problem. It means that estimates provided by the individual ISM replicates are quite variable and that the estimate of the average for the DU they provide may be unreliable. If the results are close enough to an action level that decision errors are possible, resampling with an increased number of increments may be used to reduce error. For Decision Mechanism 3, potential error created by heterogeneous concentrations is handled through calculation of the 95% UCL. Simulation studies discussed in Section 4 show, that with appropriate choice of 95% UCL method, conservative estimates of the mean to satisfy sampling objectives for this decision mechanism can be obtained despite high RSD values. This principle applies as well to other decision mechanisms where a 95% UCL is calculated.

A low RSD is not an indication that the mean is accurate or that the 95% UCL exceeds the population mean unless the distribution can be reasonably assumed to be relatively homogeneous.

A low RSD indicates that the field replicates are providing reproducible estimates of the average and generally triggers no additional steps to refine the estimate. However, it must be recognized that RSD is a measure of precision, not accuracy (see Section 4 for addition discussion of these concepts). Thus, an estimate of the average from replicates with a low RSD is not necessarily close to the actual mean. The opportunity for significant error is greatest when the DU is relatively heterogeneous and the replicates by chance give similar results. Unless information on heterogeneity of contaminants within the DU is available, it is difficult to judge whether this situation may have occurred and consequently the degree to which a low RSD should be reassuring. This is certainly an issue for the simple average of replicate data in Decision Mechanism 2. It is also an issue for Decision Mechanism 3 and others where a 95% UCL is calculated. Simulation studies discussed in Section 4 have shown that the UCL does not always ensure that a conservative estimate of the mean is obtained when the RSD is low. That is, when the RSD is low, the mean can be underestimated even by a 95% UCL. In short, a low RSD from field replicates offers information on the reliability of the

estimate of average only when the contaminant distribution within the DU is known, or can be confidently assumed, to be relatively homogeneous.

For Decision Mechanism 6, replicates are often collected from a fixed percentage of DUs; however, the selection and number of DUs from which field and laboratory replicates are collected is not a simple matter—there is no one size fits all approach. Therefore, the number of DUs from which replicates are collected must be determined using site-specific considerations. Simply relying on a fixed percentage and arbitrary decisions to select which DUs will have replicates is ill advised.

If budgetary considerations limit the number of samples, field and laboratory replicates should be collected from those DUs that will provide the most useful information. Knowledge of source areas and areas likely to have high or low concentrations should be used to make deliberate choices. If there is a choice between a DU with anticipated high concentrations (i.e., above the action level) vs. one with low concentrations (i.e., close to the action level), the DU with concentrations closest to the action level should be selected for replicate samples. The closer contaminant concentration gets to the action level, the more important replicate statistics are in making a decision. Detection limit may also be a consideration in some situations. DUs with detectable concentrations provide more information than DUs were concentrations cannot be measured.

It is advisable to collect field and laboratory replicates from DUs that are believed to have different characteristics in terms of contaminant distribution, contaminant concentration, sampling design, or sample matrix. When less than 100% of DUs have replicate samples, the RSD (same as CV) from one or more DUs can be applied to similar DUs, subject to the limitations described for Decision Mechanism 6 above. If different sources of contamination or different release mechanisms are identified, field and laboratory replicates should be collected from each different DU. Furthermore, other factors that may influence the number of DUs with replicates are significantly different soil types that could cause different contaminant distributions and/or sample preparation efficiencies and different numbers of increments in ISM samples.

8. REGULATORY CONCERNS WITH ISM

8.1 Introduction

In August and September 2009, ITRC's ISM Team developed and conducted a survey designed to collect data on incremental sampling practices from regulators, consultants, commercial laboratory personnel, and project managers. The purpose of the survey was to gain an understanding of how incremental sampling is being used, how widespread is its use, what problems have been encountered, and the current level of understanding of ISM among the respondents. Nearly three-fourths of the respondents were either state regulators or consultants (see Figure 8-1).

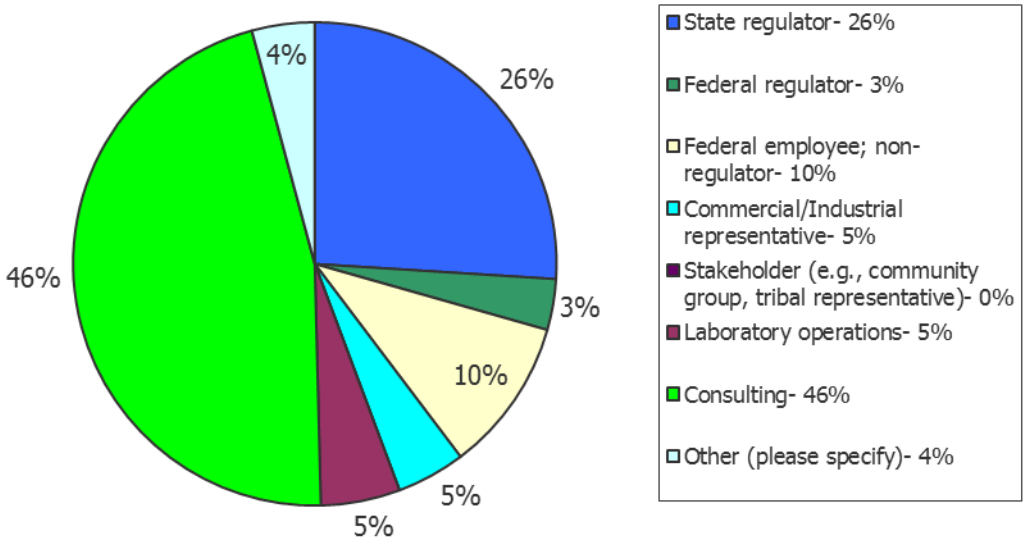


Figure 8-1. Distribution of survey respondents (n = 263).

Specific focus areas covered by the survey included regulatory challenges to using ISM, comparison of ISM to discrete sampling techniques, and the type of projects/ programs for which ISM is used. Appendix B presents the results of the survey. A subsequent survey is planned to learn of advances made during the course of this ITRC effort.

8.2 Perception Issues

Few regulators and consultants had heard of ISM, and even fewer indicated they had appreciable experience with ISM. Based on the survey results, respondents' experience with and knowledge of the limitations of composite sampling appear to have colored their acceptance of ISM.

> Survey respondents indicated that the primary difficulties with ISM are delineation of hot spots, regulatory acceptance, inability to collect ISM VOC samples, delineating the extent of the release, and determining the size and shape of the DU.

To be able to address regulator and consultant perceptions of ISM, the ITRC Team asked a number of questions to rate ISM utility and the difficulty of its application. The results indicated that as a whole, inability of ISM to delineate hot spots was the top difficulty in applying ISM, followed by lack of regulatory acceptance, problems with collecting VOC samples using ISM, inability of ISM to delineate the extent of contamination, and lack of knowledge on how to determine the size and shape of the DU. Regulators saw regulatory issues as the top difficulty in applying ISM. Nonregulators (consultants and laboratories) saw the inability of ISM to delineate hot spots as the top difficulty.

Respondents with lower and higher experience and knowledge perceived ISM differently. Individuals who rated their experience "high" indicated delineation of hot spots as the top problem but at less than half the frequency of less experienced individuals. Those with minimal experience responded that regulatory issues and acceptance as the top difficulty in applying ISM.

Respondents also listed limited training and understanding, failure to properly apply systematic planning, application of ISM data, and dealing with VOCs as other major difficulties in applying ISM. It is interesting to note that the main issues are not technical in nature but are related to the application of ISM.

The survey asked respondents whether they had any personal opinion about ISM. For regulators and nonregulators, about one-third had evaluated ISM but used it only rarely. The personal opinions indicated caution regarding the use of ISM. The opinions ranged from questions about cost-effectiveness (laboratory preparation, smaller sites, etc.), applications such as VOC sampling and analysis, applicability to COCs other than explosive compounds, site-screening, and sediment sampling. The responses suggest that the reluctance to use ISM stems from a lack of experience.

8.3 Regulatory Challenges for ISM

When asked during the 2009 survey, 40% of regulator and 20% of the nonregulator respondents agreed there are specific applications for which ISM would not be endorsed. Both groups agreed that the least likely application to be endorsed for ISM is to identify areas of high concentration (i.e., hot spots).

Several states have regulations and guidance that specifically address hot spots. These states include Massachusetts and Oregon. See Table 3-2 of the *Use of Risk Assessment in Management of Contaminated Sites* (ITRC 2008) for more details.

Seventy-eight regulator respondents representing 25 states felt that ISM is discouraged (56%) or even expressly prohibited (3%). Three percent of respondents indicated ISM is recommended in their states (38% responded “other” or no comment).

Some states have statutory/rule prohibitions on compositing, while five states have policy/guidance restrictions on specific applications of compositing.

Only a few states (Alaska and Hawaii) generally accept the use of incremental sampling. Several states indicated that they are debating the widespread use of ISM, and Washington is updating its regulations to include ISM. Table 8-1 lists states that provided links to documents restricting or prohibiting ISM/compositing.

Table 8-1. List of states with specific restrictions on compositing[a]

State	Reason for Restriction	Link
Iowa	Discrete and maximum concentrations required	www.iowadnr.gov/land/ust/techindex.html
Florida	Some action levels based on acute exposure; compositing not allowed; if using 95% UCL, must use discrete samples; maximum may not exceed three times action level for many sites; leaching not to exceed action levels	www.dep.state.fl.us/waste/categories/wc/pages/ProgramTechnicalSupport.htm
Michigan	Composites of samples are not accepted without prior DEQ approval	www.michigan.gov/documents/deq/deq-erd-stats-s3tm_250015_7.pdf
Wisconsin	Sampling average by permission only	http://dnr.wi.gov/org/aw/rr/technical/results.pdf
New Jersey	Discrete required and composite prohibited	www.state.nj.us/dep/srp/regs/techrule N.J.A.C. 7:26E-3.4(c), N.J.A.C. 7:26E-3.6(a)5

[a] This information was true at the time of the survey 2009. Please contact the appropriate state to see whether this information is still current.

Language in applicable statutes or rules may specify delineation of the horizontal and vertical extent of contamination in a way that requires consideration of point concentrations rather than area averages. This would include, for example, situations in which the boundary of the contaminated area is defined through comparison of concentrations at various locations with not-to-exceed values (e.g., risk-based preliminary remediation goals or background levels).

8.4 State of Knowledge, Experience, and Training

One segment of the survey was designed to assess the current state of knowledge among the regulator and nonregulator communities and to gain a feel for respondents' level of experience and training with ISM sampling.

Responses to series of survey questions were used to gauge respondents' level of knowledge and acceptance of ISM. Responses were tabulated and scored, and statistical comparisons were made between groups of respondents. The analysis of the survey results shows that nonregulators were generally stronger proponents of ISM than the other respondent groups. Analysis of the data geographically also shows a stronger level of understanding and support for ISM within USEPA Regions 6, 9, and 10 as compared with other USEPA Regions.

> The survey results indicated that USEPA Regions 6, 9, and 10 generally understand and support ISM more than other EPA regions.

The 2009 survey data indicate that while there is a basic level of understanding of ISM, the level of actual experience with ISM is fairly low. Sixty percent of respondents rated their understanding of ISM concepts as moderate or very good, while nearly 70% rated their level of experience as modest to none.

8.4.1 How ISM Is Being Conducted

ISM sampling has been used nearly twice as frequently on sites with commercial/industrial land use as compared with residential land use. When queried about the programs in which ISM sampling was most often conducted, survey respondents identified CERCLA and state-lead cleanup sites most frequently, followed by RCRA and petroleum sites (see Figure 8-2).

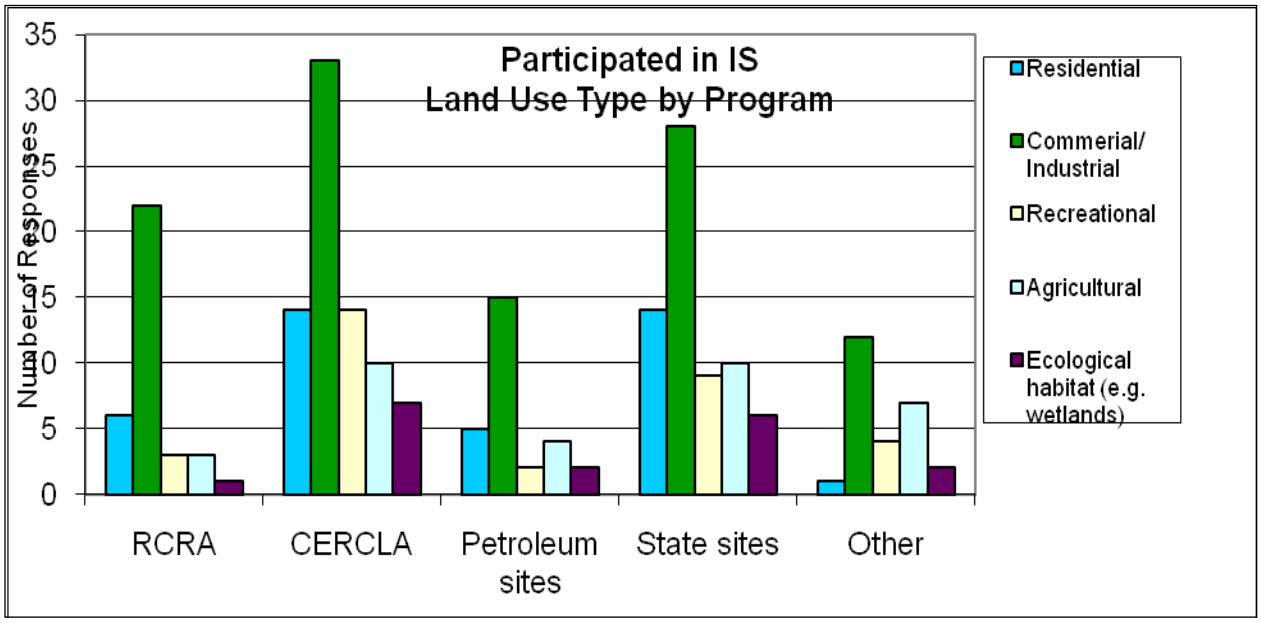


Figure 8-2. Survey response of ISM sampling in land use type by program.

According to the survey responses, ISM is being used primarily during screening investigations or to obtain data for meeting regulatory or cleanup criteria (see Figure 8-3) and primarily on surface soils (see Figure 8-4). Fewer than half of the respondents indicated they had used ISM for subsurface soil sampling. Very few respondents cited use of ISM for other matrices, such as sediment, soil gas, and water.

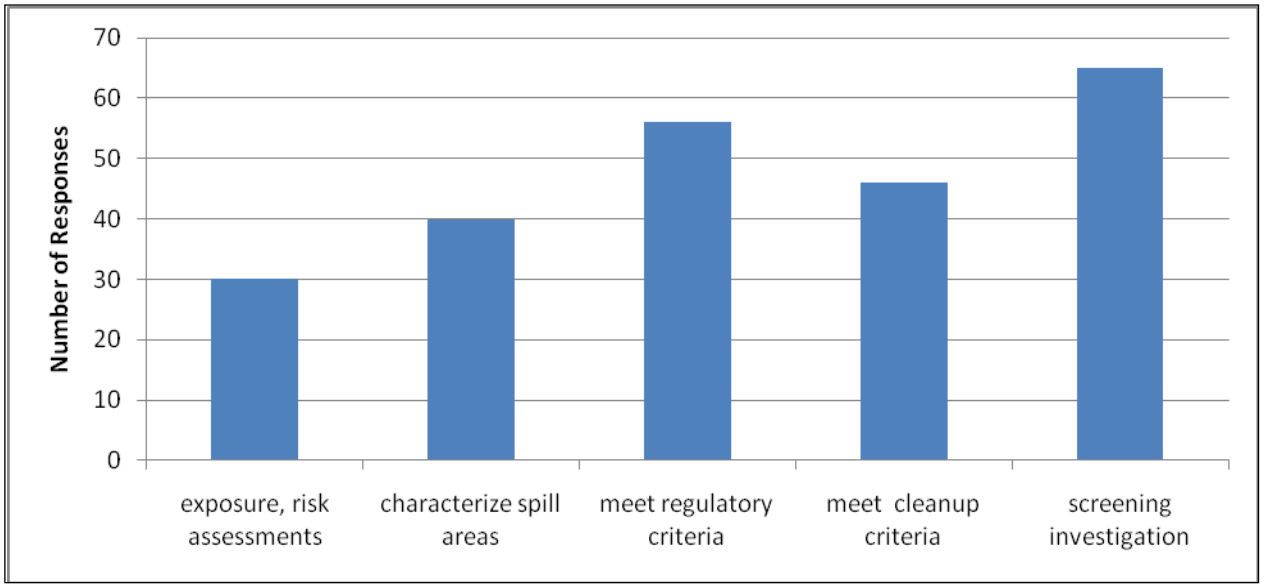


Figure 8-3. Survey responses identifying the objectives of ISM sampling.

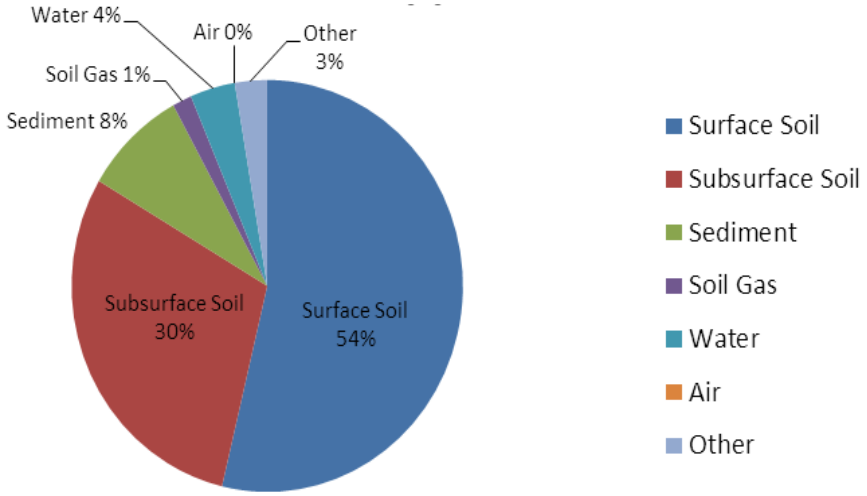


Figure 8-4. Survey responses of the ISM media applications.

ISM has been used primarily to assess heavy metals and explosive residues. Other contaminants cited less frequently include VOC and SVOCs, PCBs, and total petroleum hydrocarbons (TPH).

Hawaii, California, and Alaska together account for more than 40% of all ISM sampling activity.

Although survey data show ISM has been conducted in 36 states, over half of the ISM sampling activity has been conducted in a relatively small number of states, with Hawaii, California, and Alaska together accounting for over 40% of all ISM sampling activity (see Figure 8-5). The survey found that about half of the ISM sampling projects have been conducted on commercial/industrial land use types (see Figure 8-6).

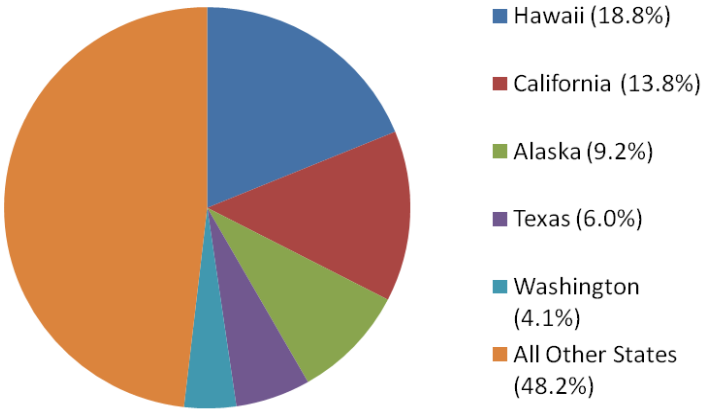


Figure 8-5. Survey responses of states where the organization has participated in ISM.

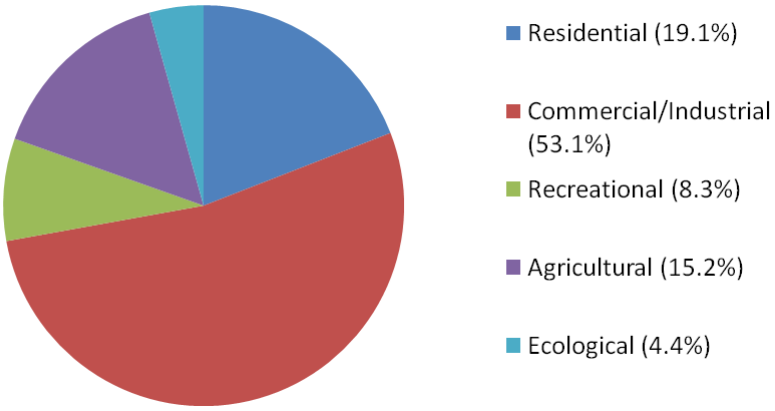


Figure 8-6. Survey responses of ISM sampling participation per land use type.

8.4.2 Written Guidance

There is relatively little written guidance available on the use of ISM for environmental contaminants. Most survey respondents cited one of three primary written sources:

Hawaii Department of Health. 2008b. *Technical Guidance Manual*, in preparation. Office of Hazard Evaluation and Emergency Response. www.hawaiidoh.org.

Alaska Department of Environmental Conservation. 2009. *Draft Guidance on MULTI INCREMENT Soil Sampling*. Division of Spill Preventions and Response, Contaminated Sites Program. www.itrcweb.org/ism-1/references/multi_increment.pdf.

U.S. Environmental Protection Agency. 2006c. "Method 8330B: Nitroaromatics, Nitramines, Nitrate Esters by High-Performance Liquid Chromatography (HPLC)," Appendix A. "Collecting and Processing of Representative Samples for Energetic Residues in Solid Matrices from Military Training Ranges." www.itrcweb.org/ism-1/references/8330b.pdf.

The following are other guidance documents cited by survey respondents which address the use of ISM:

- USACE *Protocols for Collection of Surface Soil Samples at Military Training and Testing Ranges for the Characterization of Energetic Munitions Constituents* (Hewitt et al. 2007)
- USEPA *Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples* (Gerlach and Nocerino 2003)

8.4.3 Misconceptions

With proper identification of sampling goals and use of the systematic planning process, an ISM sampling design can be created to identify specific areas of high contaminant concentration within an area of interest (DU) (see Figure 8-7). However, a significant percentage of survey respondents either felt that ISM cannot accomplish this goal or were undecided. This opinion was particularly pronounced in responses by regulators relative to nonregulator respondents.

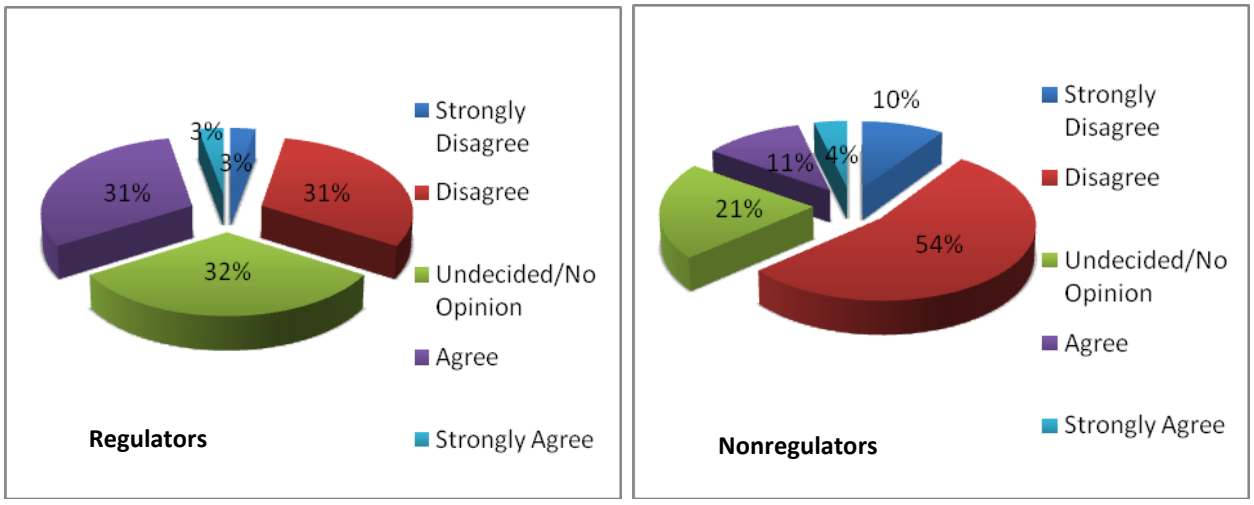


Figure 8-7. Survey responses for the statement "ISM is ineffective because it cannot identify specific areas of high concentration."

As described in Section 4.3, an ISM sampling plan can be designed to collect information about contaminant distribution variability and provide input for conventional risk assessment calculations. About 30% of regulators and 46% of nonregulators are undecided whether this is the case (Figure 8-8).

As stated in Section 2, the contaminant concentration reported by the lab is a ratio of the mass of contaminant measured to the total mass of the analytical subsample. Approximately 47% of respondents disagree that contaminant concentrations are related to the amount of the soil sample (see Figure 8-9).

There is no clear consensus among respondents about whether ISM is more expensive than discrete sampling (see Figure 8-10). Refer to Section 8.5.3 for additional discussion of this topic.

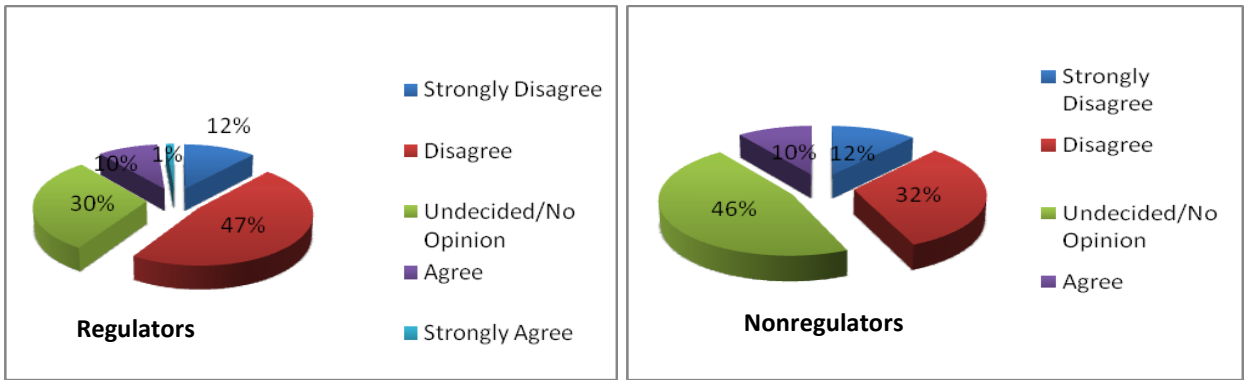


Figure 8-8. Survey responses for the statement "incremental sampling cannot be used for risk assessment because it does not address variability."

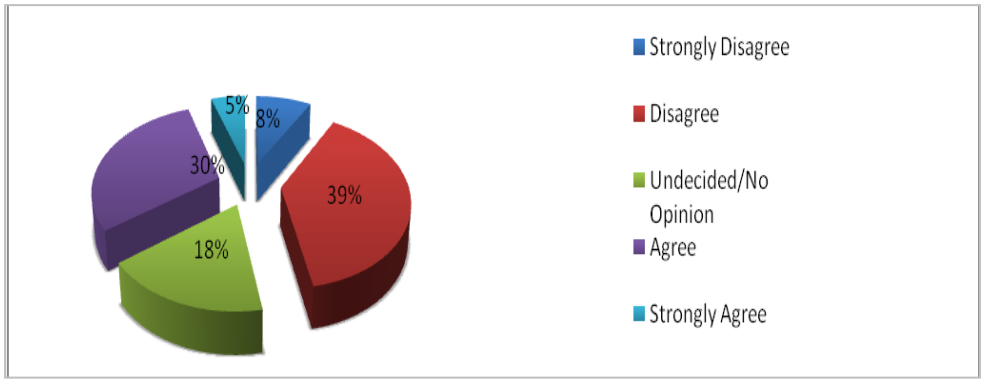


Figure 8-9. Survey responses for the statement "contaminant concentration depends on the amount of soil sample."

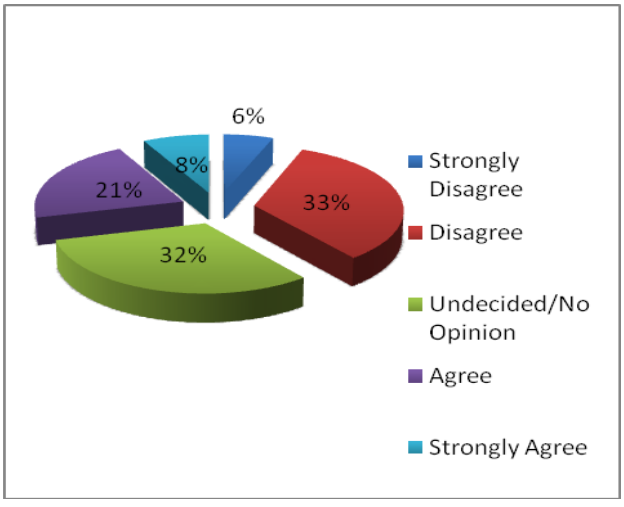


Figure 8-10. Survey responses for the statement "ISM is generally more expensive than conventional discrete sampling."

Although an important element of ISM is sample processing procedures that are typically employed prior to sample analysis, there appears to be a significant level of misunderstanding regarding this issue. Over half of regulator survey respondents were either undecided or disagreed that these additional measures are commonly needed (see Figure 8-11).

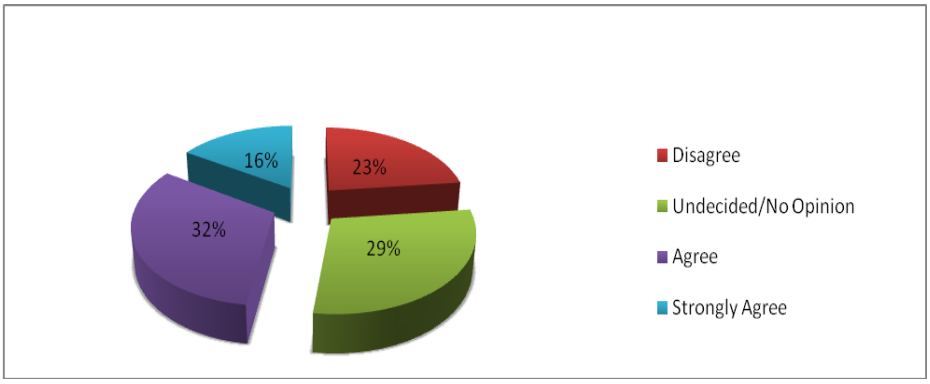


Figure 8-11. Survey responses for the statement "Incremental samples commonly require additional laboratory sample preparation."

8.4.4 Limitations

When asked to identify the most significant challenges to broader implementation of ISM, survey respondents identified delineation of local areas of high concentration as the top difficulty, with the lack of regulatory acceptance at or near the top. In fact, regulators identified regulatory acceptance as the most important challenge, followed by collecting ISM samples for VOC analysis and determining the DU. By contrast, nonregulator respondents identified delineation of hot spots as the top difficulty, followed by regulatory acceptance and delineation of the extent of contamination.

Other limitations identified by respondents include a general lack of training/knowledge of ISM concepts and techniques in both the regulatory and consulting communities, difficulties in using ISM techniques on projects involving collection of subsurface samples, and comparing ISM data sets with historic discrete sampling results.

8.5 Implementation Issues

This section identifies potential obstacles to implementing ISM, particularly those identified in the survey, and provides recommendations to assist project teams in addressing the challenges. For purposes of this section, the project team includes the project manager, geologist, engineer, toxicologist, laboratory personnel, and in some cases, interested public stakeholders, consultants, and one or more regulatory agencies. The project team personnel are determined in the DQO process and by other project considerations.

8.5.1 Systematic Planning

Any type of sampling design (ISM or discrete) should be based on a systematic planning approach to ensure that there are clear objectives and that the data obtained are of sufficient

quality to make an environmental decision (see Section 3). Good systematic planning involves a series of well-thought-out steps, but often projects omit this process entirely or miss one or more key elements. The project team needs to continually review and understand the key elements listed below and apply a systematic approach to site investigation to bring the site to completion.

- Develop the CSM.
- Identify the COPCs.
- Identify data info needs. (What is the reason for the sampling, and what is the function of the data?)
- Determine the need to find hot spots. Agree on concentration and size of hot spot (i.e., volume).
- Define the DU.
- Develop decision statements.
- Develop and implement SAP.
- Ensure data quality.
- Control decision error using defensible decision rules.
- Conduct data assessment and identify environmental hazards.
- Reevaluate the CSM.

Lack of a clear and concise CSM at the start of a project can lead to confusion and disagreements throughout the planning, implementation, and data assessment phases. The primary objective of most site investigations is to determine the presence or absence of potential environmental hazards associated with environmental contamination. Sampling objectives tie directly to development of the DU and the CSM.

The project team needs to discuss all the systematic planning elements. Experience has shown that some elements are more difficult to develop and agree to than others. Some of the common ISM systematic planning challenges are discussed in the following subsections.

8.5.1.1 CSM

Challenge: Developing a CSM acceptable to the entire planning team.

Recommendation: Making decisions based on a poor CSM can lead to incorrect environmental decisions. Whenever practical, conduct face-to-face planning meetings and use site maps and figures to aid in developing the CSM. The planning team should visit the site prior to or as part of planning meeting(s) to aid in developing the CSM. Revise and update the CSM, if appropriate, with new information. Indicate whether the new information supports the existing CSM.

8.5.1.2 Sampling objectives and developing the decision unit

Challenge: Some project stakeholders are concerned about identifying potential hot spots.

Recommendation: Because the size of the DU sets the scale of the resolution of the investigation, it is critical to ask and resolve the following questions before dividing an area under investigation into DUs:

- What is the overall objective(s)?
- What is the amount of soil to be concerned about?
- Do acute hazards need to be addressed?

With discrete sample data, assertions about hot spots are typically made after looking at the results of the investigation. Claiming that a single discrete soil sample result represents a meaningful volume of highly contaminated soil is rarely defensible and more importantly often not practical. A hot spot identified by an individual discrete soil sample within a DU may simply denote the degree of heterogeneity of contaminant concentrations within the DU. It is sometimes assumed that a sample or samples from a data set which contain the highest contaminant concentrations represent the highest concentrations actually present at a site, when in fact, due to short-scale heterogeneity, an area only an inch away may contain higher concentrations. An extremely large number of samples is required to estimate and delineate with any degree of certainty relatively small areas with the highest concentrations that might be present at a site. Estimating the actual true maximum contaminant concentrations in soil is therefore often an impractical endeavor.

ISM cannot identify hot spots smaller than the DU.

ISM replicates provide a measure of variability in estimates of the mean concentration in the DU but do not provide information about the spatial distribution. To characterize spatial variability, either the scale of the DU or the scale of the areas sampled within the DU need to be adjusted. DUs can vary greatly in size from very small (e.g., much less than ¼ acre such as a sandbox) to very large (e.g., hundreds of acres). During systematic planning, options for either combining or splitting DUs may be considered to address multiple objectives. It is possible that prior site knowledge can be used to refine the sampling plan to account for potential source areas as separate DUs, but there must be high confidence that this information is accurate. If a DU is subdivided to evaluate smaller volumes of soil, it is likely that additional ISM samples will need to be collected before decisions can be made at the scale of the smaller DU. For example, a large DU may be subdivided into four areas with ISM samples collected from each subarea. The data would support calculations of an area weighted mean concentration and 95% UCL for the large DU, as well as preliminary estimates of means within each subarea. Additional ISM samples could then be collected to reduce uncertainty and delineate areas of elevated mean concentrations at the smaller scales. More detail on combining and splitting DUs and calculating area weighted means can be found in Sections 4.4.1 and 3.4.

As a general rule, misunderstandings about potential hot spots can be avoided with proper DU scaling during systematic planning and by keeping the investigation and evaluation focused on the DU identified, not to the scale of an individual sample.

Challenge: Some project stakeholders are concerned that potential areas of higher concentration within a DU (i.e., hot spots) will be diluted out when combined through ISM with increments of soil from less-contaminated portions of the DU.

Recommendation: Even the best systematic planning and underlying CSM could result in a sampling design that fails to identify small areas of extreme high and low concentrations within the DU in the proper proportion compared to the total mass of the soil. This reality contributes to a concern about dilution of hot spots. There are at least two issues with the issue of hot-spot dilution:

There are two concerns regarding hot spots: sampling density and defining the DU. ISM effectively addresses compliance when action levels are based on the mean concentration within a DU. Concerns related to spatial resolution can be addressed only by changing the scale of the DU (i.e., DU equals hot spot size).

- Sampling density. If the small area(s) within a DU with extremely high concentrations are not represented by a sample in the proper proportions (compared to the total mass of soil in the DU), the estimate of the sample mean can be highly variable. Unless the ISM sampling process is repeated many times (replicates), there remains a concern about the performance of any one sampling event. Specifically, if the difference between the estimate of the mean and the true population is large, there is a greater chance of reaching the wrong conclusion regarding compliance with an action level (see performance metrics, Section 4). For this reason, it is very important to consider collection of replicates to ensure that the small areas of high concentration are collected in the proper proportion relative to the total mass.
- Defining the DU. Different sampling designs provide different information concerning the location, spatial extent, and magnitude of subareas of high concentrations. Uncertainty regarding the toxicological significance of acute exposures leads to uncertainty in the definition of a hot spot. This uncertainty creates a challenge for any sampling design—without a clear definition of what constitutes a hot spot, it is difficult to delineate a DU and develop a sampling design that provides information to adequately address hot spots. For this reason, the definition of hot spot and size of the DU must be agreed on by the team during the systematic design phase.

The chance that any single sampling event will include subareas of high and low concentrations in the proper proportion is directly related to the number of samples collected within a DU. An advantage of ISM over other sampling designs is that it accommodates large sample sizes (i.e., large number of increments as well as multiple replicates). For this reason, while any individual sample collected in a hot spot is diluted within the larger group of samples, we are more likely to achieve an estimate of the mean that is representative of the true mean within the DU. This advantage of ISM addresses the first concern (compliance with action levels) but not the second concern (spatial resolution). If the DQO includes the identification and delineation of small areas of elevated concentrations, ISM sampling can address this objective only by changing the scale of the DU (i.e., DU must be the same size as the hot spot of concern).

Challenge: Developing the correct DU.

Recommendation: The DU is properly sized when knowledge about spatial variability/spatial patterns of contaminant concentrations within the DU are no longer of interest. Section 3.3 provides information to consider when developing the DU(s). Keep in mind that a site may be subdivided into multiple DUs to accomplish investigation objectives. Consider how DU sample results will be used. If the objective is to assess potential exposure concentrations over a ¼-acre residential lot, then the DU is ¼ acre, and subdivision into smaller DUs is not necessary. If the objective is to investigate and locate areas of potentially higher contaminant concentration within the site as a whole (e.g., source areas or separate exposure areas), using multiple smaller DUs is appropriate.

Although DUs are ideally sized no larger than the volume of soil in which the average concentration is sufficient to make a decision, there may be situations where DUs cannot be sized small enough for practical reasons. For example, the Florida Department of Environmental Protection has criteria for some contaminants in soil for protection from acute exposure. These criteria are based on a scenario in which a small child ingests, on a single occasion, a handful (10 g) of soil. For this scenario, the DU should be an area approximately the size of child's hand. Obviously, it would be impractical to divide a site into DUs of this size, and use of a larger DU encompassing thousands of these exposure areas would raise legitimate questions of whether acute toxicity potential can be evaluated. This is a problem with the use of discrete data as well, but with discrete data, some information on variability of concentrations within an area is obtained with which to estimate what the concentration in the most contaminated exposure area might be. For sites in Florida in which acute exposure and toxicity are a concern, the regulatory acceptance issue is whether to allow use of a method such as ISM that appears incapable of providing reasonable assurance that acute-toxicity based criteria have been met. A similar issue may apply to ecological risk assessments, where the area of exposure for some species of interest is smaller than can be accommodated by an affordable number of DUs. Again, discrete sampling has its own set of problems dealing with this issue, but a sufficient number of samples can provide information on variability of concentrations over space from which predictions regarding worst-case exposure areas can be made.

A refinement of initial DUs flagged for remediation may be useful to better isolate areas of high contamination and optimize resources available for cleanup. See Section 3 for a thorough discussion on this topic.

8.5.2 Lab Availability

Application of ISM to environmental site assessment has primarily occurred in the last few years. Thus, only a few technically advanced commercial laboratories have developed ISM sample processing and handling capabilities, primarily in support of federal agencies such as USACE and USEPA or state agencies such as HDOH and ADEC. As of the survey date in 2009, laboratories providing support work for military sites or sites in Hawaii or Alaska were the most likely to have experience supporting the advanced processes used for incremental samples.

ISM sample processing techniques are so new that many have not been fully documented in laboratory SOPs. Laboratory support for USEPA SW-846 Method 8330B (explosives) is probably the best-documented ISM method. Certification and reference material are available for

this method. There are currently no other USEPA-approved ISM methods for other contaminants; thus, the laboratory processes have been developed on a case-by-case basis. Technical conversations between the laboratory and project chemists and other data users early on during the systematic planning process are strongly recommended to ensure that the appropriate support processes are selected to meet the DQOs. Support processes can vary greatly from one laboratory to the next for parameters other than explosives. Guidance about the strengths and limitations of the various options within the laboratory is discussed in Section 6.

It is expected that as use of ISM increases, more laboratories will gain experience, and finding laboratories familiar with the necessary procedures for handling ISM samples will no longer be an issue.

8.5.3 Costs

A cost comparison of ISM to discrete sampling approaches is difficult. Cost-effective sampling is important, but it is more critical that the sampling approach(es) meet the sampling objectives.

The cost of collecting and processing an individual ISM sample is nearly always more than that for a single discrete sample. In general, the number of analytical samples to be analyzed for ISM is less than for discrete samples, so ISM analytical cost may be lower. However, costs differences are based on various issues, including specific analytical costs (e.g., metal vs. dioxin), availability and quality of screening technologies, and ease of collecting samples. Costs should be evaluated on a case-by-case basis.

It is also important to remember that ISM generally yields more precise and unbiased estimates of the mean (for example, three 30-increment ISM samples as compared to three discrete samples). This difference is important because, from a decision-making standpoint, investigations based on limited discrete sample data are in many cases more likely to result in a decision error. In those instances where an ISM investigation costs more than discrete sample investigation, the cost-benefit ratio might still favor the ISM investigation because it may result in fewer decision errors.

When ISM and discrete sampling costs are similar, variability in the ISM data can be significantly less. This decreased data variability might allow for less uncertainty in decision making, especially when estimates of the mean are close to an action level. Ultimately, making a correct decision at a site reduces overall project costs by eliminating costly and unnecessary remediation of DUs incorrectly identified as dirty or dealing with the consequences of mistakenly walking away from a DU that is actually contaminated.

Although the survey did not query the question of cost vs. benefit, this section discusses general costs for ISM. It should be noted that the Florida field study presented in Appendix C did not have a detailed cost-benefit analysis as a project objective, and for this reason costs analysis for the Florida field study are not presented.

Challenge: How do costs for ISM compare to discrete sampling approaches?

Recommendation: The recommendations below discuss four areas of cost differences between ISM and discrete sampling approaches: systematic planning, sampling plan review, field costs, and laboratory costs.

Systematic Planning. Systematic planning, including the designation of DUs and associated decision statements that guide evaluation of the data collected, while being key components of ISM investigations, are not unique to ISM. ISM requires the up-front consideration of DUs and associated decision statements. Although this should be done prior any project involving soil sample collection, traditionally, systematic planning is often omitted completely or only partially conducted prior to many site investigations. This omission often results in multiple sample collection events purely for site characterization purposes, followed by the designation of what are essentially DUs using already-collected data on which final decision making is based. This process often leads to the need for additional site investigations to fill data gaps or, especially at small sites with limited budgets, final decisions based on low-quality data that may or may not reflect the actual risks posed by contamination at the site.

Very little data exist on how much systematic planning for ISM costs vs. more traditional discrete sample approaches. The reason is likely that the ISM approach is new and that conditions vary so greatly from site to 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 costs and the need for lengthy discussions following completion of the field investigation. The intent of systematic planning is to 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 costs in the project life cycle.

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. Many regulators stated that they currently do not have time to review standard sampling plans and reports, let alone a more labor-intensive ISM plan. 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. Many consultants find it much easier to submit standard sampling plans and assessment/remediation reports to regulators in an attempt to get 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. While this statement may be true currently, the publication of this ISM document and subsequent training should allow practitioners to develop ISM work plans with less risk of rejection and regulators to review sampling plans more quickly.

Field Costs. Many factors can affect the cost for ISM field sampling. Only limited data on ISM sampling costs were available at the time this ITRC document was developed. All of the costs are highly dependent on DQOs. The discussion of costs presented below is only for surface soil sampling. Generally, field costs for ISM and the equivalent number of discrete samples (e.g.,

three ISM or 30 discrete samples) are approximately equivalent. Cost considerations include the number of increments in an ISM sample, replicate collection, and field processing.

Based on experience reported by the State of Hawaii, the average time needed to lay out up to a 1-acre DU in the field and collect a single 30–50 point ISM sample is approximately 45 minutes for a three-person field team (two to collect samples and one to manage samples, decontaminate, manage paperwork, etc.). For the same number of discrete sample or increment points (e.g., 30 points within a targeted area), the collection of a single ISM sample will be faster than the collection of 30 discrete samples due to the need to label, pack, and document a much larger number of the discrete samples. In cases where a relatively low number of discrete samples are required for characterization of a targeted area (e.g., 10 discrete samples), the field time required to collect the discrete samples is likely to be significantly shorter than the time required to collect and process an ISM sample, especially if replicate ISM samples are to be collected.

The real cost saving is in the analysis effort needed to produce equivalent precision, where, for example, instead of analyzing 30 discrete samples from a targeted area, the lab analyzes three ISM samples (one incremental sample and two replicates). An example presented at the 2010 Environment, Energy Security and Sustainability Conference (Penfold 2010), indicated that the total field and lab costs for one ISM sample and two replicates from a single DU was $3,150 vs. $6,975 for 30 discrete samples. The ISM samples contained 100 increments each. The total field and lab cost for 10 DUs was $20,700 for ISM vs. $62,725 for discrete samples. The samples were analyzed by USEPA SW-846 Method 8330B for explosives.

The 2009 Environmental Security Technology Certification Program (ESTCP) *Cost and Performance Report* (Hewitt et al. 2009) prepared for characterization of energetic residues provides an excellent discussion on the cost issues associated with ISM sampling for USEPA SW-846 Method 8330B. According to the report, extra costs could include ISM sample shipment and disposal (due to extra weight) as well as QA/QC costs associated with batch samples. In addition, the report noted that ISM was not projected to be cost-competitive on smaller scale due to the relative increased processing (e.g., sieving and handling) and analysis; however, the report concluded that there is a cost saving of 50%–80% using ISM (Hewitt et al. 2009).

Laboratory. ISM increases the amount of sample handling in the laboratory. There is a wide variety of ISM laboratory sample processing and subsampling options. The price of ISM processing depends on the specific options selected, the amount of soil to be processed, analytes of concern, and other general business concerns (e.g., number of samples, turnaround time). As of mid-2010, the additional cost of ISM sample processing ranged from $50–$250 for a 1 kg soil sample. Normal sample preparation and analysis charges depend on the contaminant(s) of interest and are not included in this price range estimate. Processing equipment blanks, LCSs, and MS/MSDs through the ISM laboratory steps is recommended, but the lack of readily available and suitable reference materials makes it challenging to estimate potential costs. Depending on QA/QC samples necessary to meet DQOs, per batch cost could increase significantly. Despite increased costs, the added value of processing known QA/QC samples may be worthwhile. Discuss batch QA/QC options with the laboratory during project planning to get specific cost estimates.

The ISM approach might not be the most cost-effective option when low-cost field screening tests provide acceptable accuracy and sensitivity (such as XRF for selected metals) and can be used inexpensively on large numbers of discrete samples.

Challenge: Are there cases where ISM is not cost-effective or when is ISM most cost-effective?

Recommendation: Costs need to be evaluated on a site-by-site basis as DQOs and other site-specific factors make it very difficult to predict which sample method will be the most cost-effective. The following issues related to costs should be considered:

Cost considerations for ISM include individual analyte costs, availability of field analytical methods, type of sample processing necessary (drying, particle size reduction, sieving, subsampling), difficulty in sieving, whether the lab has an ISM SOP, shipping and disposal costs associated with larger ISM sample mass, and need for additional laboratory QC samples.

- What are individual analyte costs?
- Are there field analytical methods that can be used for specific contaminants?
- What type of sample processing has to be done (drying, particle size reduction, sieving, subsampling)?
- If sieving in the field, does the soil contain clay, roots, or very wet soil? These will likely increase overall field processing time and increase costs.
- Will the laboratory charge be based on how difficult is it to sieve the soil (clay, roots, very wet)?
- Does the lab have an ISM SOP, and if not, will it charge to develop one?
- What are the costs for extra QC samples (i.e., batch samples) often needed with ISM?
- Are there added costs for shipping and disposing the large volumes of soil collected with ISM?

Note, however, that cost should not be the most important issue. The priority should be whether the sampling design meets the sampling objectives.

8.5.4 Challenges in Developing and Using ISM Data

Regulatory agencies are most accustomed to working with discrete sample data sets, with a variety of concentration measurements from within an area of interest. It is not surprising that some regulatory criteria are written specifically to deal with these data sets, specifying decisions to be made based upon data parameters readily obtained (e.g., mean, maximum concentration observed). ISM data, while providing in most cases a better estimate of the mean, cannot provide all of the parameters that may be called for in some regulations (e.g., maximum concentration observed, upper percentile concentrations observed). This shortcoming can constitute a regulatory barrier to acceptance, as noted in some of the sections below. Other sections address other practical problems associated with the nature of ISM and discrete data that may limit regulatory acceptance of ISM.

8.5.4.1 Validation of statistical analysis within ISM

A rigorous statistical analysis regarding the extent to which various ISM sampling strategies provide accurate estimates of the average has not yet been published. Information from simulation studies can be used to address this issue, and results from efforts conducted during development of this document are described in Section 4 and Appendix A. However, it is reasonable to state that this is not a "mature" area of study, meaning that the strengths and limitations of ISM from a statistical standpoint are only just now being rigorously explored. Given that the statistical foundation of ISM is critical to understanding its reliability in providing estimates of the mean for site evaluation, regulatory agencies may be reluctant to embrace ISM until more thorough statistical evaluation has been conducted and formal guidance that addresses these issues is available.

8.5.4.2 Meeting regulatory requirements for average and maximum concentration estimates

EPCs that present a spatial average are often required to be the upper 95% confidence limit (95% UCL) on the mean. If only a single ISM sample is taken from a DU, a UCL on the mean cannot be calculated to satisfy regulatory requirements. In this situation, a sampling strategy involving replicate measurements is required. (See Section 4 for a discussion of calculation of 95% UCL values from ISM data.) Also, some programs that accept an expression of the mean for risk and cleanup evaluation also specify an upper percentile or a maximum concentration that can remain on site. For example, some states currently require for most sites that if the 95% UCL is below the soil criterion, contamination is still not within acceptable limits unless the maximum concentration is at or below three times the criterion. Because ISM provides no direct indication of the maximum concentration, its ability to demonstrate compliance with this regulatory requirement is questionable. Approaches to estimate maximum concentrations within a DU using ISM data are discussed in Section 4 and might be useful to address this issue but have not been widely used.

8.5.4.3 Decision unit versus exposure unit

Some DUs selected for a project may not match the definition of an exposure unit, for example, when DUs are designed solely for comparisons with cleanup values to reach "remediate/don't remediate" decisions for specific plots. Another situation is when there are different exposure units over the same area for different receptors but only one set of DUs for a site. Depending on how DUs are defined, it is possible to have a DU that is larger or smaller than an exposure unit. There are no statistical procedures in place to estimate an EPC when the DU is larger than the exposure unit, although some possibilities for estimating high end concentrations are discussed in Section 3.5. Similarly, when an exposure unit is composed of more than one DU, unless replicate ISM are available for each, there is no established method to combine results from the DUs to produce a robust, demonstrably conservative EPC. Limitations of methods in situations where the DUs do not match the exposure units may be a significant obstacle to the use of ISM data in risk assessment. Methods for overcoming these limitations are discussed in Section 4.2.

8.5.4.4 Comparison of discrete samples and ISM samples

Many sites have historical discrete sampling data, and some have concurrent discrete sampling data taken in response to specific needs or regulatory requirements. Qualitative comparisons can be very instructive, but quantitative comparison of discrete and ISM data should be done only with caution. Comparison of ISM samples to discrete samples is discussed in Section 4.4.3.2. Key issues that should be considered whenever comparing ISM and discrete samples are sampling design, sample collection method, similar soil conditions, similar sample processing and analysis, and data quality being understood and appropriate for intended use.

8.5.4.5 Comparison of ISM means and "not to exceed" basis regulation

Regulatory requirements in some situations compel evaluation of concentrations within an area on a "not-to-exceed" basis. This may include screening levels, action levels, or leachability values, depending on the state. In this situation, derivation of a mean concentration by ISM alone does not satisfy the requirement. The development of statistical approaches that use ISM data in some form to estimate variability across the sampled area could overcome this challenge and allow ISM to be used in these circumstances, as discussed in Section 4.

Under some state regulations, leachability-based cleanup goals may be considered to be not-to-exceed values for any single, discrete sample collected within the targeted volume of soil. In this situation, ISM samples would not allow a direct comparison to cleanup objectives; however, it is important to consider that the cleanup goals need to consider not only the concentration but the mass of the contamination. ISM data can provide an estimate of the mass of the contaminant within the DU. This allows for better comparison to the cleanup goals. As an update to the survey, some states are moving in the direction of emphasizing mass of contamination over leachability goals by establishing a minimum volume of soil and contaminant concentrations that need to be considered for potential leaching hazard.

8.5.4.6 Decisions based on a single ISM

This limitation applies specifically to an approach where only one ISM sample is taken per DU. Because a 95% UCL cannot be calculated from a single ISM result, this approach is precluded when a 95% UCL is required by regulation. It may also not be accepted because of the inherent uncertainty associated with using a single, unreplicated estimate of the mean and the potential to underestimate the actual mean.

8.5.4.7 Background/geochemical limitations

Comparison of site data with background data may be necessary to establish the extent to which chemicals present are naturally occurring. As discussed in Section 4.4.3.3, the two key challenges for ISM are the likelihood of detecting differences in the populations that exist (ISM background data to ISM site data) and the inability to evaluate upper tails of the background to site underlying distributions. In addition, decision errors may be affected if the background samples are collected with different sampling designs from the site samples, including different number of increments/replicates, different sample masses, sampling protocols, depth intervals, and

sampling patterns. Therefore, the results of hypothesis tests applied to ISM data sets should be interpreted with caution until these limitations can be more thoroughly studied. Even if formal statistical tests are not used, simple graphical analysis (e.g., plots grouping ISM results by study area) may be informative as a semiquantitative method for comparing background and site distributions.

Comparison of site ISM data to background discrete data using either hypothesis testing or upper tolerance limits is not recommended because the variance is represented differently in ISM and discrete sampling. Careless comparison of an ISM estimate of the mean to a discrete sample collected from soil representing background is likely to lead to decision errors in which one incorrectly concludes that the contaminant distribution on site is consistent with background conditions.

8.5.4.8 Extrapolation between and within DUs

In some situations, the area to be evaluated is larger than can be effectively sampled. This is typically the case for large tracts of land where available resources may preclude sampling each properly sized DU. One approach in this situation is to sample a portion of the area to be evaluated using ISM and extrapolate data to other, unsampled areas. This can take a number of forms, including (a) dividing the area into DUs, sampling some fraction of the DUs, and extrapolating the mean and/or variance of ISM data from sampled to unsampled DUs and (b) creating SUs within a DU that cover some but not all of the area. Results from the SUs are used to make a decision on the DU.

Justification for extrapolation from sampled to unsampled areas is usually based on a CSM that predicts a similar distribution of concentrations in both areas. Generally, this assumption is based solely on judgment and can be associated with considerable uncertainty. This issue is not unique to ISM and applies equally if extrapolation is considered using data from other sampling strategies such as composite and discrete. It is also important to note that ISM offers no special advantage in reducing this uncertainty. Regulatory acceptance of uncertainty associated with extrapolation can vary considerably, depending on the agency and sometimes site-specific circumstances such as the intended use of the property (e.g., agricultural vs. residential). Discussions within the ITRC ISM Team and feedback from states indicate that extrapolation will be accepted by some states, under some circumstances, but not by others.

8.5.5 Matrix and Parameter Issues

8.5.5.1 Laboratory experience

Based on the survey results, most laboratories' ISM experience has been with metals projects. Explosives and SVOC projects are next in frequency. PCBs and TPHs make up the third tier. A few laboratories have ISM experience with VOC and dioxins, and fewer still with perchlorate and cyanide. Parameter-specific certification of the ISM laboratory processes is generally not available due to a lack of reference methods. Section 6 provides specific guidance about appropriate laboratory processes for the various parameter groups. Some laboratories have experience with a wider range of contaminants, including metals, pesticides, dioxins, PCBs,

SVOCs, VOCs, and petroleum, so it is important to ask laboratories how much experience they have for contaminant analytes required by the DQOs.

8.5.5.2 Sample processing

Two key issues are the applicability of air-drying and the use of particle size reduction to facilitate representative laboratory subsampling. These techniques generally work well for higher-boiling-point, thermally stable contaminants. The question of whether to grind samples prior to metals analysis should be carefully considered, since this can both improve reproducibility and release metals previously bound inside particles that were less environmentally accessible. The ISM principles have been applied to VOCs in a manner that does not require air-drying or additional air exposure. See Section 5 for field activities and Section 6 for the corresponding laboratory activities.

Meeting traditional holding times is more challenging when air-drying and particle size reduction techniques are used at the laboratory, due to the lengthened sample processing; however, data from USEPA SW-846 Method 8330B studies indicate that contaminant stability can be much longer in dried samples than is common in as-received moist samples. The Method 8830B process has been validated for energetic residues, but it should not be assumed to apply to all other contaminants or contaminant groups. See Section 6.2 for a detailed discussion of the strengths and limitations of the various ISM sample processing options.

Lab certification for ISM sample processing procedures may require certification by laboratory SOP except for USEPA SW-846 Method 8330B. This may be a significant limitation for certain regulatory entities. Please see Section 6.4 for more details.

Below is a list of recommendations for addressing matrix and parameter issues:

- Minimize error by processing ISM samples in a controlled setting.
- Do not use particle size reduction on ISM samples to be analyzed for organic contaminants other than energetics.
- Request that the lab analyze a laboratory control sample at a minimum frequency of one per analytical batch of 20 ISM samples.
- When analyzing ISM samples for SVOCs, confirm that drying is acceptable for the specific target compounds.

8.6 Summary

The survey identified several issues concerning the use of ISM, including how to successfully collect VOC samples with ISM, misconceptions about hot spot identification, how to use ISM data, how to apply ISM cost-effectively, and when ISM may not be the best choice. The ISM Team used the survey information to aid in developing this technical-regulatory guidance document. If the guidance document is successful, the perception of ISM will be improved, and regulatory challenges can be broken down, thus allowing ISM to be used more often and in an appropriate fashion. Table 8-2 provides a summary of the limitations and possible solutions for more widespread implementation of ISM.

Table 8-2. Limitations, solutions, and section references for using ISM

Limitations	Solutions	Section reference
Hot spots	Address during systematic planning with proper scaling, combining, or splitting DUs	Sections 3.5, 8.2, and 8.5
Vertical and horizontal DU delineation	Address during systematic planning with proper scaling or splitting DUs	Section 3.3
Acute exposure	Development of approaches for "decompositing" ISM data to estimate variability in concentrations within a DU	Section 3.1, 3.3, and 3.5
Background	Development of formalized guidance on statistical methods for comparison of site and background ISM data.	Sections 3.1, 3.2, 3.3, 4.4.3.3, 7.2.4, and 8.5.4.7
Leachability	ISM provides probability statement	Sections 3.1, 3.2, 3.3, 8.3, and 8.5
Compare with regulatory standards	Discuss with stakeholders during systematic planning	Sections 3.1, 7, and 8.5.4.5
ISM cost-effectiveness	Cost-effective when large DU, expensive analyte cost, remobilization is expensive	Section 8.5.3
Statistical challenges—compare ISM and discrete	Development of statistically sound methods for comparison of discrete and ISM data	Sections 4.4.3.2 and 8.5.4.4
Statistical challenges—95% UCL	Use Student's-*t* or Chebyshev	Section 4 and Appendix A
Statistical challenges—DUs that do not correspond to exposure units	Development of statistically based methods for combining and subdividing DUs	Sections 3.1, 3.3, 4, and 7
Grinding	Not recommended for organics other than energetics by USEPA SW-846 Method 8330B; recommended for nonvolatile metals; may not be appropriate for project-specific DQOs	Section 3.1 and 6.2
Lab-processing, equipment—sieving, grinding, drying	Close coordination with laboratory is essential throughout ISM; lab business decision to have specific equipment; may need to evaluate different grinding equipment based on method detection limit requirements; laboratory should be familiar with the project-specific ISM requirements and have the facilities (space) and equipment (air-drying racks, grinders, etc.) to meet project-specific DQOs	Section 6.2

Limitations	Solutions	Section reference
Lab—lack of nationally recognized methods	USEPA/DOD methods development	Section 6.1
Field—shipping VOC container	Complete extraction in field and ship subaliquot to the lab; transport via lab pick-up or appropriate method for hazardous goods	Section 5.4.2
Lab—VOC elevated method detection limit	Analyze by USEPA SW-846 Method 8260C SIM; additionally, may be necessary to use low-level VOC discrete sampling and/or a combination of ISM and discrete	Section 6.3.2
Lab—certification	Check with the appropriate regulatory agency; some states have certification process for lab SOPs; continued research is necessary on possible effects of ISM sample preparation procedures on COPCs, especially organics; develop and implement lab certification for ISM, possibly through NELAP	Sections 6.4.1 and 8.5.5.2
DU size and shape	Establish based on site history during systematic planning; may require remobilization if concern over results at the end of sample collection	Sections 3.1, 3.2, and 3.3
Regulators reluctance to use ISM	Review ITRC document and attend ITRC training	All

Although this document attempts to cover all the relevant topics to ISM, there are several issues which were not addressed, including the following:

- consequences of sample grinding on assumptions made during ecological risk assessments
- use of ISM for sampling air, sediment, or other environmental media
- additional statistical simulation to evaluate:
  - combining DUs (Section 4)
  - comparing site to site DUs
  - comparing IS vs. discrete
  - comparing site vs. background
  - comparing oversized DUs
  - other types of sampling errors

9. CASE STUDY SUMMARIES

This section introduces four case studies that examine the application and comparative findings of incremental sampling methods to discrete sampling methods. Appendix C presents the complete case studies and detailed findings.

9.1 Case Study 1. PCB-Contaminated Landfill

Site Name: Green Island Landfill and Reburial Pit, Kure Atoll, Hawaii

Contact Name: Roger Brewer, HDOH

Site Location: Kure Atoll is the northernmost island in the Hawaiian Island chain, located approximately 1400 miles northwest of the island Oahu and 56 miles northwest of Midway atoll. The atoll consists of a lagoon encircled by a reef and a single vegetated island, Green Island. Green Island is just under 1.5 miles long and about 0.35 miles in width and has a maximum elevation of 15 feet.

Background: A U.S. Coast Guard (USCG) station was located on the atoll from the 1960s through the 1990s. A ½-acre area located on the southwest corner of the island was used to dispose of old electrical components and scrap metal. Discreet confirmation soil samples identified concentrations of PCBs as high as 170 mg/kg within the formal landfill footprint. Soil, sediment, and biota samples collected in the surrounding area indicated that PCB contamination was primarily restricted to the landfill site. Debris and approximately 700 yd^3 of PCB-contaminated soil were removed from the site in 1993.

A follow-up study of the former landfill area was carried out in 2008. As part of the site investigation, USCG took the opportunity to evaluate the potential advantage and limitations of incremental soil sampling approaches over traditional discrete sampling approaches. The investigation focused on the use of DU and ISM investigation strategies published by HDOH (2008b).

Statistical Evaluation: A statistical evaluation of discrete vs. incremental sample data was conducted by Anita Singh, a contractor to USEPA with Lockheed Martin in Las Vegas and member of the ITRC ISM Team. One objective of the review was to compare estimates of the mean concentration of PCBs in the DU soil based on a specific number of discrete samples vs. one to three incremental samples drawn from the same data set. Another objective of the Kure atoll data set was to determine the equivalent number of discrete samples to a triplicate set of 30–50 point incremental samples. This effort will help to evaluate the cost-effectiveness of collecting incremental samples over discrete samples.

Lessons Learned—ISM Data Collection:

- Isolation of areas of suspected higher contamination is important at a site-wide scale but not at a DU scale.

- Identify and investigate suspected spill areas separately via historical knowledge and/or preliminary sampling.
- Subdivide remaining area into risk-based DUs based on human or ecological health concerns.
- Incorporate an adequate number of increment points within a DU to capture contaminant distribution and heterogeneity.
- A range of 30–50+ increment points is required to adequately characterize a DU—anything less is probably just sampling the mode.
- Use replicate samples to verify that contaminant heterogeneity has been adequately characterized.
- Tight grids of discrete samples can be useful for an initial screening of sites and DU designation, as well as subdivision of "hot" DUs for smaller areas for isolation and characterization of concentrated contamination.

Lessons Learned—ISM Simulation:

- Include at least 30–50 increments per ISM field replicate sample for initial DU characterization.
- Always collect and use replicate sample data (e.g., triplicates) from one or more DUs at a site to evaluate the representativeness of incremental sample data.

Some combination of both discrete and ISM sampling data was ideal for estimating the PCB mean.

- Determining the appropriate number of ISM increments and replicates is critical to ensuring that the ISM sample is representative of the conditions in the field and to assess precision. Between 60 and 90 increment replicate samples are needed to ensure the incorporation of isolated hot spots.
- ISM helped identify the primary spill area, but ISM for the entire core would have yielded the same answer.
- It is not possible to test data representativeness of field data with a small set of discrete samples (e.g., <30 samples), as lognormal outliers would likely be missed.

9.2 Case Study 2. Petroleum-Contaminated Soil Stockpile

Site Name: Petroleum Contaminated Soil Stockpile, Prince of Wales Island, Alaska

Contact Name: Earl Crapps, Alaska Department of Environmental Conservation (ADEC)

Site Location: The site is located on the Prince of Wales Island near Craig, Alaska. Craig is on a small island off the west coast of Prince of Wales Island and is connected by a short causeway. It is 56 air miles northwest of Ketchikan and 220 miles south of Juneau.

Background: The purpose of this project was to test the protocols in the ADEC draft MULTI INCREMENT® sampling guidance (ADEC 2009). The test site was a petroleum-contaminated soil stockpile located in a rock quarry on Prince of Wales Island, Alaska.

During the 2006 excavation and removal of an underground heating oil tank, discrete samples were collected that documented diesel range organics (DROs) at 300–900 mg/kg. Stockpile

tilling and fertilizing were conducted by the responsible party several times after the soil was moved from its original location in May 2006.

Lessons Learned:

- Although the stockpile was shallow, it was compacted and difficult to excavate by hand.
- Field sampling was labor-intensive, requiring approximately 15 person hours to complete. Data quality may have been affected.
- It is recommended that ISM sample processing occurs in a controlled laboratory setting.

9.3 Case Study 3. Former Golf Course Field Demonstration of ISM

Site Name: Former Golf Course

Contacts Names: Kelly Black, Neptune and Company, Inc.; Deana Crumbling, USEPA; Ligia Mora-Applegate, Florida Department of Environmental Protection; Mark Malinowski, California Department of Toxic Substance Control; Phil Goodrum, ERM; Keith Tolson, Geosyntec Consultants; Ed Corl, NAVFAC Laboratory Quality and Accreditation Office; Hugh Rieck, USACE; Steve Roberts, University of Florida; Leah Stuchal, University of Florida; Richard Lewis, Conestoga-Rovers & Associates, Inc.; Chris Saranko, Environmental Planning Specialists, Inc.

Site Location: Florida

Background: The ITRC ISM Team identified this site for a field demonstration of ISM. The site was a former golf course where both fertilizers and herbicides containing arsenic had been applied. This former golf course will become a residential development. While it was an active golf course, arsenic was applied in two ways. Monosodium methanearsonate (MSMA) was used as a herbicide to stunt the growth of unwanted plant life, mostly on the fairways. Also, arsenic-rich fertilizer was used frequently on the course. Fertilizer was used more heavily on the tee boxes and greens than on the fairways. Arsenic in soils was the media and COC. Preliminary characterization showed that arsenic is the only COC and that it ranges from 0 to nearly 100 mg/kg in some areas, with significant contamination limited to the top 6 inches of soil.

Lessons Learned:

- Only in cases with strongly skewed or variable data was there much value in collecting more than 30 increments per sample.
- Discrete samples spanned a much wider concentration range and were more variable than the ISM results.
- The data collected via discrete samples and the data collected via incremental sampling methods lead to different results and potentially different decisions.
- Partitioning DUs into subareas may provide an opportunity to discern spatial differences that would not be apparent if incremental samples were collected from the entire DU as a whole.

9.4 Case Study 4. Hawaiian Homelands Development

Site Name: Hawaiian Homelands Development, Kapolei, Oahu, Hawaii

Contact Name: Roger Brewer, HDOH

Site Location: The East Kapolei Affordable Housing Project property is located in East Kapolei, Kapolei, Oahu, Hawaii.

Background: This case study summarizes the investigation of the 401-acre former sugarcane field and a ½-acre pesticide-mixing area located within the field which is being developed for residential and commercial use. The primary COCs were arsenic, pentachlorophenol (PCP), dioxins (associated with past use of PCP), and triazine herbicides, each used in the past for weed control. A detailed discussion of the sugarcane field investigation is provided in the report *East Kapolei Affordable Housing Project Kapolei, Oahu, Hawaii, Final Site Assessment Report* (TTEMI 2007). A summary of the pesticide mixing area investigation is provided in the report *Site Investigation Report and Environmental Hazard Evaluation, East Kapolei II Pesticide Mixing and Loading Site* (ESTC 2007, 2010).

Lessons Learned:

- There are no elevated concentrations of COPCs in the soil that suggest conditions are not suitable for residential reuse or that any additional sampling or evaluation is necessary.
- The investigations confirm that ISM samples, essentially very good composite samples with additional lab requirements, are better able to capture small hot spots and overall contaminant heterogeneity within a targeted area.

10. STAKEHOLDER AND TRIBAL INPUT

ISM can be used in various stages of site investigation, including site characterization to evaluate whether a contamination problem exists, to identify and isolate contaminant source areas or high levels of concentrations (e.g., hot spots), or for 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. For the purpose of this ITRC document, the term "stakeholder" represents the citizen stakeholder, community or environmental advocacy members, tribal members, and members of the affected public. "Tribal" represents the Indian tribes, pueblos, nations, et al.; Native Hawaiians; and Alaskan Native Americans (e.g., Tlingit, Athabascan) and Native Alaskans (e.g., Yupik, Inupiat).

Stakeholders want to be assured that site investigation activities and subsequent decisions "do no harm."

Stakeholders and tribal members/environmental staff, like regulators, want to be assured that such activities "do no harm" and that the planned activities find all the contamination so it can be cleaned up. Active public stakeholders and tribes generally support

planned activities and try to understand the processes used to characterize or clean up a site. During investigations, questions often are asked about how to know whether a chemical is there or not: "Was this area sampled, and why not sample over there?" Answering these questions requires open communication starting with initial planning and continuing throughout the project.

However, 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 EPA-approved water quality standards. Some tribes are now developing tribal risk assessments, which incorporate pathways and scenarios based on traditional and cultural routes of exposure, which in some cases are essentially and profoundly different than 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 tribe's government-to-government status.

There are times when stakeholders and tribes need a better understanding of how sampling is done and why sample locations are placed in particular spots. Sampling plans should aid the stakeholders and tribes in understanding those issues. This document may also prove helpful in explaining the challenges associated with soil sampling and how ISM addresses some key uncertainties associated with soil sampling.

The primary concern of the ISM approach, as expressed by several members of the ITRC stakeholder group and by some members of the ISM Team, is the idea of averaging away a hot spot. Even if a DU meets the regulatory threshold of 95% UCL for the COC, it may not alleviate the nagging question of whether or not that hot spot might someday become the location of a child's sandbox or play area.

To reinforce this point, an example was cited at a location in New York City where the site of an old railway yard was being redeveloped for an elementary school. The site name is Mott Haven. The NYC School Construction Authority proposed to build a four-campus school on an old railyard in the Bronx. The sampling of the site isolated a hot spot. The contractor proposed compositing the samples which would meet the 95% UCL and avoid the need to remediate the hot spot. The local citizens group did not support this decision and intervened. The Bronx Committee on Safe Schools retained the New York Lawyers for Public Interest to review their concerns over the cleanup of the site, resulting in a reversal of that decision and the eventual remediation of the hot spot.

This example illustrates the power of stakeholder influence and the degree to which citizens can be instrumental in the decision-making process. It is imperative that, during the systematic planning stage of a proposed ISM project for future public use, all affected stakeholders and tribes be identified, engaged, and included in defining the sampling plan and cleanup objectives of the site. In the case of the tribe(s), this may be because the tribe is a regulatory agency.

> It is imperative that affected stakeholders be identified, engaged, and included in defining the sampling plan and cleanup objectives of the site.

The information provided in this document may aid all parties in understanding how ISM is more likely to find contaminations rather than dilute the results. Questions regarding not all the soil being sampled or areas not being sampled and how to make sure nothing is missed relate not only to ISM but to discrete sampling approaches as well. The key point about ISM is that ISM results provide a more reliable estimate of the average concentration for the area being sampled.

The four case studies presented in this document include a PCB site on an isolated and uninhabited pacific atoll; a petroleum dump site in Prince of Wales Island, Alaska; a former golf course that is being redeveloped for an upscale housing development in Florida; and a former sugarcane plantation slated for an affordable housing development in Hawaii. These case studies illustrate the wide range of contaminants and geographical regions applicable to ISM.

The first two case studies are located on federally owned property and are not intended to be used by the public, but many federal facilities include citizen advisory groups or public comment periods that allow for community outreach and stakeholder involvement. The last two case studies are slated for housing developments and may have a direct impact on future residents. It is likely that stakeholder involvement will be critical for these redevelopment projects, and stakeholder interest may be heightened. Although these case studies do not specifically highlight the level of stakeholder involvement at these sites, it is important to note that stakeholder involvement can be a crucial element to the overall success of an investigation, particularly with using ISM. Also, it is imperative to remember that decision-making protocols with Indian tribes, Native Hawaiians, and Alaska Inuits need to be recognized. As stated earlier, in many cases, tribes have treaties with the federal government granting them regulatory authority over environmental cleanup on native lands. Keep in mind that political boundaries are a creation of the dominant culture; areas of tribal concerns may go beyond modern political boundaries to ancestral homelands. For example, a major DOE facility in New Mexico is sited entirely on the ancestral homeland of a neighboring tribe. In this case, DOE has honored the government-to-government relationship and has partnered with the tribe in monitoring efforts and in communicating early and often with the tribe on proposed actions which might affect the tribal members and resources.

By appreciating stakeholder and tribal concerns early on and through effective communication, it may be possible to better explain the proposed ISM for a particular site in a more open, transparent, and understandable fashion that meets the stakeholders' expectations of fairness, as well as possible tribal regulatory requirements, and speaks to their concerns about risk on a level and in terms to which they can relate. The bottom line is that citizen stakeholders and tribes need to come away from any discussion with the sense that they and their loved ones are safe, and that no threat exists to their continued well-being. In the case of tribes, this confidence may need to extend to seven generations, which for many tribes is the length of time their stewardship of Mother Earth extends.

11. REFERENCES

75th CEG/CEV (75th Civil Engineering Group Environmental Flight). 2007. *Repeated Multi-Incremental Explosive Residue Characterization at the Utah Test and Training Range.*

ADEC (Alaska Department of Environmental Conservation). 2009. *Draft Guidance on MULTI INCREMENT Soil Sampling*. Division of Spill Preventions and Response, Contaminated Sites Program. www.itrcweb.org/ism-1/references/multi_increment.pdf.

ASTM (ASTM International). 2003. *Standard Guide for Laboratory Subsampling of Media Related to Waste Management Activities.* Method D 6323-98. www.astm.org/Standards/D6323.htm.

ASTM. 2006. *Standard Guide for Sampling Strategies for Heterogeneous Wastes*. Method D 5956-96. www.astm.org/Standards/D5956.htm.

ASTM. 2008. *Standard Guide for Developing Conceptual Site Models for Contaminated Sites*. Method E 1689-95. www.astm.org/Standards/E1689.htm.

ASTM. 2009. *Standard Guide for Representative Sampling for Management of Waste and Contaminated Media.* Method D 6044-96. www.astm.org/Standards/D6044.htm.

Barnett, V. and Bown, M. 2002. "Statistically Meaningful Standards for Contaminated Sites Using Composite Sampling," *Environmetrics* **13**: 1–13.

Bruce, M. L. 2003. "Turning Dirt into Dust, Evaluation of Kitchen Appliances," presented at the Louisville Chemistry Conference, Kentucky.

Buchman, M. F. 2008. *NOAA Screening Quick Reference Tables.* NOAA OR&R Report 08-1. Seattle: Office of Response and Restoration Division, National Oceanic and Atmospheric Administration. www.itrcweb.org/ism-1/references/SQuiRTs.pdf.

Casella, G., and R. L. Berger. 2001. *Statistical Inference*, 2nd ed. Pacific Grove, Calif.: Duxbury.

CCME (Canadian Council of Ministers of the Environment). 2001. *Reference Method for the Canada-Wide Standard for Petroleum Hydrocarbons in Soil: Tier 1 Method*. www.itrcweb.org/ism-1/references/final_phc_method_rvsd_e.pdf.

Cochran, W. G. 1977. *Sampling Techniques*. New York: Wiley.

Cressie, N. 1993. *Statistics for Spatial Data*. New York: Wiley.

DOD (U.S. Department of Defense). 2010. *DOD Quality Systems Manual for Environmental Laboratories*, vers. 4.2.

Dixon, W. J. 1953. "Processing Data for Outliers," *Biometrics* **9**: 74–89.

Domenico, P. A., and F. W. Schwartz. 1990. *Physical and Chemical Hydrogeology*. New York: Wiley.

EDQW (Environmental Data Quality Working Group). 2008. *Guide for Implementing EPA SW-846 Method 8330B.* U.S. Department of Defense. www.itrcweb.org/ism-1/references/Final_8330B_Implementation_Guide_070708.pdf.

ERDC (U.S. Army Engineer Research and Development Center). 2000. *Protocols for a Rapid Cleanup/Extraction Procedure and an Improved P450RGS Dioxin Screening Assay for Sediments*. ERDC TN-DOER-C10. www.itrcweb.org/ism-1/references/doerc10.pdf.

ESTC (EnviroServices & Training Center, LLC). 2007. *Site Investigation Report and Environmental Hazard Evaluation, East Kapolei II Pesticide Mixing and Loading Site.*

ESTC. 2010. *Site Investigation Report and Environmental Hazard Evaluation, East Kapolei II Pesticide Mixing and Loading Site.* http://hawaii.gov/dhhl/publications/ekii/SIR-EHE,%20Mar2010.pdf.

Gerlach, R. W., D. E. Dobb, G. A. Raab, and J. M. Nocerino. 2002. "Gy Sampling Theory in Environmental Studies, 1. Assessing Soil Splitting Protocols," *Journal of Chemometrics* **16**: 321–28.

Gerlach, R. W., and J. M. Nocerino. 2003. *Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples.* EPA/600/R-03/027. www.itrcweb.org/ism-1/references/guidancerl.pdf.

Gilbert, R. O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. New York: Van Nostrand Reinhold.

Gilbert, R. O., and P. G. Doctor. 1985. "Determining the Number and Size of Soil Aliquots for Assessing Particulate Contaminant Concentrations," *Journal of Environmental Quality* **14**(2): 286–92.

Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.

Grant, C. L., T. F. Jenkins, K. F. Myers, and E. F. McCormick. 1995. "Holding-Time Estimates for Soil Containing Explosives Residues: Comparison of Fortification vs. Field Contamination," *Environmental Toxicology and Chemistry* **14**: 1865–74.

Gy, P. 1998. *Sampling for Analytical Purposes.* A. G. Royle, trans. West Sussex, U.K.: Wiley.

HDOH (Hawaii Department of Health). 2008a. *Evaluation of Environmental Hazards at Sites with Contaminated Soil and Groundwater*. Office of Hazard Evaluation and Emergency Response. www.itrcweb.org/ism-1/references/HTGM_Section_04-00.pdf.

HDOH. 2008b. *Technical Guidance Manual*, in preparation. Office of Hazard Evaluation and Emergency Response. www.hawaiidoh.org.

HDOH 2009. *Technical Guidance Manual*, interim final. Office of Hazard Evaluation and Emergency Response. www.hawaiidoh.org.

HDOH. 2011. *Technical Guidance Manual Notes: Decision Unit and Multi-Increment* Sample Investigations*. Office of Hazard Evaluation and Emergency Response. www.itrcweb.org/ism-1/references/DU-MIS_Notes_(Final_March_15_2011).pdf.

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, A. D., T. F. Jenkins, M. E. Walsh, M. R. Walsh, S. R. Bigl, and C. A. Ramsey. 2007. *Protocols for Collection of Surface Soil Samples at Military Training and Testing Ranges for the Characterization of Munitions Constituents*. ERDC/CRREL TR-07-10. Hanover, N.H.: U.S. Army Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. www.itrcweb.org/ism-1/references/GetTRDoc.pdf.

Hewitt, A. D., T. F. Jenkins, M. E. Walsh, M. R. Walsh, and S. Brochu. 2009. *ESTCP Cost and Performance Report: Validation of Sampling Protocol and the Promulgation of Method Modification for the Characterization of Energetic Residues on Military Testing and Training Ranges*. ESTCP/CRREL ER-0628. Hanover, N.H.: U.S. Army Engineer Research

and Development Center, Cold Regions Research and Engineering Laboratory in partnership with Defense Research and Development Canada-Valcartier, Quebec, Canada. www.serdp.org/Program-Areas/Environmental-Restoration/Contaminants-on-Ranges/Identifying-and-Evaluating-Sources/ER-200628.

Hewitt, A. D., P. H. Miyares, D. C. Leggett, and T. F. Jenkins. 1992. "Comparison of Analytical Methods for Determination of Volatile Organic Compounds in Soils," *Environmental Science and Technology* **26**: 1932–38.

Hewitt, A. D., C. Ramsey, and S. Bigl. 2008. "MULTI INCREMENT® TCE Vadose-Zone Investigation," *Remediation* **19**(1): 125–40.

IATA (International Air Transport Association). 2011. "IATA Dangerous Goods Guidance." www.iata.org/whatwedo/cargo/dangerous_goods/pages/dgr-guidance.aspx.

Ingamells, C. O., and F. F. Pitard. 1986. *Applied Geochemical Analysis*. New York: Wiley.

ISO (International Standards Organization. 2005. *Soil Quality: Pretreatment of Samples by Freeze-Drying for Subsequent Analysis*. ISO16720:2005. http://www.iso.org/iso/catalogue_detail.htm?csnumber=30427.

ITRC (Interstate Technology & Regulatory Council). 2003. *Technical and Regulatory Guidance for the Triad Approach: A New Paradigm for Environmental Project Management*. SCM-1. Washington, D.C.: Interstate Technology & Regulatory Council; Sampling, Characterization and Monitoring Team. www.itrcweb.org.

ITRC. 2007a. *Improving Environmental Site Remediation through Performance-Based Environmental Management*. RPO-7. Washington, D.C.: Interstate Technology & Regulatory Council, Remediation Process Optimization Team. www.itrcweb.org.

ITRC. 2007b. *Triad Implementation Guide.* SCM-3. Washington, D.C.: Interstate Technology & Regulatory Council; Sampling, Characterization and Monitoring Team. www.itrcweb.org.

ITRC. 2008. *Use of Risk Assessment in Management of Contaminated Sites.* RISK-2. Washington, D.C.: Interstate Technology & Regulatory Council, Risk Assessment Resources Team. www.itrcweb.org.

Jenkins, T. F., A. D. Hewitt, M. E. Walsh, T. A. Ranney, C. A. Ramsey, C. L. Grant, and K. L. Bjella. 2005. "Representative Sampling for Energetic Compounds on Military Training Ranges," *Environmental Forensics* **6**: 45–55.

Patil, G. P., and C. Taillie. 2001. "Use of Best Linear Unbiased Prediction for Hot Spot Identification in Two-Way Compositing," *Environmental and Ecological Statistics* **8**:163–69.

Pavlowsky, R. T., M. R. Owens, and D. J. Martin. 2009. *Distribution, Geochemistry, and Storage of Mining Sediment in Channel and Floodplain Deposits of the Big River System in St. Francois, Washington, and Jefferson Counties, Missouri.* Report to the U.S. Fish and Wildlife Service, Cooperative Ecosystems Studies Unit. www.itrcweb.org/ism-1/references/Big_River_Chan_&_FP_Mining_Sediment_Report_6-18-10_MSU_FINAL.pdf.

Penfold, L. 2010. "Incremental Sampling and 8330B for Explosive Residues at Two Military Ranges—Analytical Considerations," presented at the Environment, Energy Security and Sustainability Symposium and Exhibition, National Defense Industrial Association, June 17–14, Denver. www.itrcweb.org/ism-1/references/10234.pdf.

Petersen, L., C. K. Dahl, and K. H. Esbensen. 2004. "Representative Mass Reduction in Sampling: A Critical Survey of Techniques and Hardware," *Chemometrics and Intelligent Laboratory Systems* **74**(1): 95–114.

Pitard, F. 1993. *Pierre Gy's Sampling Theory and Sampling Practice: Heterogeneity, Sampling Correctness, and Statistical Process Control*, 2nd ed. Boca Raton, Fla.: CRC Press.

Ramsey, C. A., and A. D. Hewitt. 2005. "A Methodology for Assessing Sample Representativeness," *Environmental Forensics* **6**: 71–75.

Robbat, A. 1997. *A Guideline for Dynamic Workplans and Field Analytics: The Keys to Cost-Effective Site Characterization and Cleanup*. www.itrcweb.org/ism-1/references/dynwkpln.pdf.

Smith, P. L. 2006. *A Primer for Sampling Solids, Liquids, and Gases.* Philadelphia: American Statistical Association and Society for Industrial Mathematics.

Superfund Research Program. 2010. "Sequestration of Dioxin by Clays." Research Brief 183. http://tools.niehs.nih.gov/srp/researchbriefs/view.cfm?Brief_ID=183.

Tindall, S. 2006. "Managing Uncertainty with Systematic Planning," presented at the Data Quality Assessment Workshop for State of Hawaii Department of Health.

TTEMI (Tetra Tech EM, Inc.). 2007. *East Kapolei Affordable Housing Project Kapolei, Oahu, Hawaii, Final Site Assessment Report.* Prepared for the Hawaii Department of Health Office of Hazard Evaluation and Emergency Response.

USACE (U.S. Army Corps of Engineers). 1998. *Technical Project Planning (TPP) Process*. EM 200-1-2. www.itrcweb.org/ism-1/references/USACE_TPPEngrMan.pdf.

USACE. 2003. *Conceptual Site Models for Ordnance and Explosives (OE) and Hazardous, Toxic, and Radioactive Waste (HTRW) Projects.* EM 1110-1-1200. http://140.194.76.129/publications/eng-manuals/em1110-1-1200/toc.pdf.

USACE. 2008. *Incremental Sampling: MIS-Based Laboratory Requirements for the Analysis of Explosives (Method 8330B) and Metals in Solid Matrices*.

USACE. 2009. *Implementation of Incremental Sampling (IS) of Soil for the Military Munitions Response Program.* Interim Guidance Document (IGD) 09-02. Environmental and Munitions Center of Expertise. www.itrcweb.org/ism-1/references/IGD_9-02v2.pdf.

USCG (U.S. Coast Guard). 2009. *Evaluation of Green Island Landfill and Reburial Pit, Former U.S. Coast Guard LORAN Station Kure*. Prepared by Element Environmental for Civil Engineering Unit, Honolulu.

USDA (U.S. Department of Agriculture). *2010. National Soil Survey Handbook*, Section 618.43. Natural Resources Conservation Service. http://soils.usda.gov/technical/handbook.

USEPA (U.S. Environmental Protection Agency). 1985. *Verification of PCB Spill Cleanup by Sampling and Analysis*. EPA/560/5-85/026. Office of Toxic Substances.

USEPA. 1986. *Field Manual for Grid Sampling of PCB Spill Sites to Verify Cleanup*. EPA/560/5-86/017. Office of Toxic Substances.

USEPA. 1988. *Guidance for Conducting Remedial Investigations and Feasibility Studies Under CERCLA*. EPA/540/G-89/004. www.itrcweb.org/ism-1/references/guidancecri.pdf.

USEPA. 1989a. *Methods for the Attainment of Cleanup Standards, Vol.1: Soils and Solid Media*. EPA/230/02-89/042. www.itrcweb.org/ism-1/references/vol1soils.pdf.

USEPA. 1989b. *Risk Assessment Guidance for Superfund, Vol. II: Environmental Evaluation Manual*. EPA/540/1-89/001. www.itrcweb.org/ism-1/references/RASUPEV.pdf.

USEPA. 1992. *Supplemental Guidance to RAGS: Calculating the Concentration Term*. OSWER Publication 9285.7-08I.

USEPA. 1995a. *EPA Observational Economy Series, Vol. 1: Composite Sampling*. EPA/230/R-95/005. www.itrcweb.org/ism-1/references/composite.pdf.

USEPA. 1995b. *Superfund Program Representative Sampling Guidance, Vol. 1: Soil Interim Final*. EPA/540/R-95/141. OSWER Directive 9360.4-10. www.itrcweb.org/ism-1/references/sf_rep_samp_guid_soil.pdf.

USEPA. 1996a. *Soil Screening Guidance: Technical Background Document.* EPA/540/R-95/128. www.epa.gov/superfund/health/conmedia/soil/index.htm.

USEPA. 1996b. *Soil Screening Guidance: User's Guide.* EPA/540/R-96/018. Superfund Publication 9355.4-23. www.epa.gov/superfund/health/conmedia/soil/index.htm.

USEPA. 1997. *Ecological Risk Assessment Guidance for Superfund: Process for Designing and Conducting Ecological Risk Assessments* (Interim Final). EPA/540/R-97/006. Office of Solid Waste and Emergency Response. www.epa.gov/oswer/riskassessment/ecorisk/ecorisk.htm.

USEPA. 1999. *Correct Sampling Using the Theories of Pierre Gy*. Office of Research and Development Technology Support Project. www.itrcweb.org/ism-1/references/csutpg.pdf.

USEPA. 2000a. *Data Quality Objectives Process for Hazardous Waste Site Investigations: Final Guidance*. EPA/600/R-00/007. Office of Environmental Information. www.itrcweb.org/ism-1/references/g4hw-final.pdf.

USEPA. 2000b. *Guidance for Data Quality Assessment: Practical Methods for Data Analysis*. EPA QA/G-9. EPA/600/R-96/084. www.itrcweb.org/ism-1/references/g9-final.pdf.

USEPA. 2000c. *Short Sheet: TRW Recommendations for Sampling and Analysis of Soil at Lead (Pb) Sites.* EPA/540/F-00/010. www.itrcweb.org/ism-1/references/sssiev.pdf.

USEPA. 2001. *Using the Triad Approach to Improve the Cost-Effectiveness of Hazardous Waste Site Cleanups*. EPA/542/R-01/016. Office of Solid Waste and Emergency Response. www.itrcweb.org/ism-1/references/triad2.pdf.

USEPA. 2002a. *Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites*. OSWER Publication 9285.6-10.

USEPA. 2002b. "Closed-System Purge-and-Trap and Extraction for Volatile Organics in Soil and Waste Samples." Method 5035A. www.itrcweb.org/ism-1/references/5035a_r1.pdf.

USEPA. 2002c. *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites*. EPA/540/R-01/003, OSWER 9285.7-41. Office of Emergency and Remedial Response.

USEPA. 2002d. *Guidance on Choosing a Sampling Design for Environmental Data Collection for Use in Developing a Quality Assurance Project Plan, EPA QA/G-5S*. EPA/240/R-02/005. www.itrcweb.org/ism-1/references/g5s-final.pdf

USEPA. 2002e. *RCRA Waste Sampling Draft Technical Guidance*. EPA/530/D-02/002. www.epa.gov/epawaste/hazard/testmethods/sw846/samp_guid.htm

USEPA. 2005. *Reference Guide to Non-Combustion Technologies for Remediation of Persistent Organic Pollutants in Stockpiles and Soil*. EPA/542/R-05/006.

USEPA. 2006a. *Applicability of Superfund Data Categories to the Removal Program.* EPA/540/F-05/005, OSWER 9360.4-21FS. www.cecosweb.com/resources/OSWER%209360.4-21FS%20Removal%20Program%20Jul06.pdf.

USEPA. 2006b. *Guidance on Systematic Planning Using the Data Quality Objectives Process EPA QA/G-4.* EPA/240/B-06/001. www.itrcweb.org/ism-1/references/g4-final.pdf.

USEPA. 2006c. "Method 8330B: Nitroaromatics, Nitramines, Nitrate Esters by High-Performance Liquid Chromatography." www.epa.gov/epawaste/hazard/testmethods/sw846/new_meth.htm#8330B.

USEPA. 2007. *Performance of Statistical Tests for Site Versus Background Soil Comparisons When Distributional Assumptions Are Not Met*. EPA/600/R-07/020. Office of Research and Development.

USEPA. 2008. "Regional Screening Levels for Chemical Contaminants at Superfund Sites." http://www.epa.gov/region9/superfund/prg/ or www.epa.gov/reg3hwmd/risk/human/rb-concentration_table/index.htm.

USEPA. 2009. *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities: Unified Guidance*. EPA/530/R-09/007. Office of Resource Conservation and Recovery. www.itrcweb.org/ism-1/references/unified-guid.pdf.

USEPA. 2010a. *Best Management Practices: Use of Systematic Project Planning Under a Triad Approach for Site Assessment and Cleanup.* EPA/542/F-10/010. www.itrcweb.org/ism-1/references/epa-542-f-10-010.pdf.

USEPA. 2010b. *ProUCL 4.00.05 Technical Guide* (draft). EPA/600/R-07/041. Office of Research and Development. www.epa.gov/esd/tsc/TSC_form.htm.

USGS (U.S. Geological Survey). 2006. "USGS Photo Library." http://libraryphoto.cr.usgs.gov/cgi-bin/show_picture.cgi?ID=ID.%20McKee,%20E.D.%20%20316.

Walsh, M. E., C. M. Collins, R. N. Bailey, and C. L. Grant. 1997. *Composite Sampling of Sediments Contaminated with White Phosphorus.* CRREL Special Report 97-30. www.itrcweb.org/ism-1/references/GetTRDoc1.pdf.

Walsh, M. E., C. M. Collins, C. A. Ramsey, T. A. Douglas, R. N. Bailey, M. R. Walsh, A. D. Hewitt, and J. L. Clausen. 2007. *Energetic Residues on Alaskan Training Ranges: Studies for the US Army Garrison Alaska 2005 and 2006*. ERDC/CRREL TR-07-9. www.itrcweb.org/ism-1/references/TR-07-9.pdf.

Walsh, M. E., and D. Lambert. 2006. *Extraction Kinetics of Energetic Compounds from Training Range and Army Ammunition Plant Soils: Platform Shaker vs. Sonic Bath Methods*. ERDC/CRREL TR-6-6. www.itrcweb.org/ism-1/references/TR06-6.pdf.

Walsh, M. R. 2009. *User's Manual for the CRREL Multi-Increment Sampling Tool*. ERDC/CRREL SR-09-1. www.itrcweb.org/ism-1/references/umcrrel.pdf .

Walton, W. C. 1988. *Practical Aspects of Groundwater Modeling.* Worthington, Ohio: National Water Well Association.

Appendix A

Statistical Simulations

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

STATISTICAL SIMULATION STUDIES

A.1 INTRODUCTION

This appendix presents additional details regarding 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. As noted in Section 4 of the main text, a range of DU scenarios was investigated to explore the effect of various factors on statistical performance metrics. 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 relative percent difference between the UCL and true mean—standard deviation of relative bias in the population mean
- relative standard deviation 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.

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 in 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 (refer to the main text and Appendix E for a discussion of Gy's principles). The simulations focus on representing the compositional and distributional heterogeneities (CH and DH) that can be attributed to fundamental error (FE) and grouping and segregation error (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 (e.g., 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 probability distribution with a specified mean and variance (PD). The ratio of the population parameters (i.e., standard deviation divided by the mean), also known as the coefficient of variation (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 (2-D 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—Based on a real data set of more than 200 observations. The sample results were interpolated with inverse distance weighting techniques to yield a completely defined 2-D surface of concentrations (see Section A.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. Hathaway and Pulsipher (2010, NOTE: See references end of this appendix in Section A.8) document the specific details for how the discrete data were used to establish the completely defined 2-D surface of increment values shown Section A.4.
- M-3—"Bulk material" DUs, hypothetical homogeneous and heterogeneous DUs mimicking bulk material (e.g., soils) DUs are generated using the "MIS Module" of the USEPA software Scout 1.1 (USEPA n.d. "Scout 1.1"), as discussed in Section A.5.

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." These 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.

Appendix E is a glossary of terms relevant to ISM. A glossary is also included at the end of this appendix to provide an expanded discussion of the definitions of key terms and concepts.

A.1.1 Consensus Points

The presentation of simulation results is organized by points of consensus among the ISM workgroup. Each consensus point was guided by statistical theory and supported by results of simulation experiments. Table A-1 lists the consensus points, grouped by topics that are relevant to the overall sample design.

Table A-1. Consensus points guided by simulation experiments using probability distributions (PD) and maps (M)

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 (i.e., the mean of the replicate incremental sampling estimates of the mean) is a function of the total number of increments collected (increments × 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 (i.e., individual discrete samples vs. 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 (e.g., 30–100), increasing the number of ISM replicates above 3 provides marginal return in terms of improving coverage; however, increasing the number of replicates decreases (i.e., 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 yields 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," meaning 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 (e.g., based upon 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 the ISM samples. This approach 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 unit fails the decision criterion (e.g., 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 standard error 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.
Relative standard deviation	
14	Data sets with a high RSD are more likely to achieve specified coverage for 95% UCL than data sets 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. Some sets of simulations were subject to different interpretations, yielded inconsistent findings, or were repetitive. That some sets of the simulations were inconsistent or viewed differently within the ISM statistics workgroup is not surprising given that exploration of the statistical implications of ISM is a relatively new field. The reasons for differences were still being considered during development of this ITRC document. It is anticipated that additional research may be needed to further investigate the performance of alternative ISM sample designs. To avoid confusion and limit presentation to

essential material, Table A-1 includes only findings related to consensus points. Two of the simulation approaches below (Sections A.4 and A.5) have documented their work, and additional simulations can be found in technical documents (Hathaway and Pulsipher 2010; Singh, Singh, and Murphy 2009). The simulation approaches in Sections A.2 and A.3 are presented here for the first time.

A.1.2 Points of Nonconsensus

There were subjects on which the statistics workgroup was unable to reach consensus. The most significant of these regards the M-3 simulations presented in Section A.5. Some members of the workgroup were of the strong opinion that these simulations addressed the bulk material sampling nature of ISM and Gy theory in ways that the other simulations did not and make an important contribution to understanding of how sampling patterns affect ISM performance. Other members of the workgroup held the strong opinion that the methods for these simulations and their interpretation of these simulations are flawed, specifically in terms of how the "true" mean is defined. To aid the reader in understanding the basis for this disagreement, the contrasting viewpoints are presented below.

A.1.2.1 Rationale for the M-3 simulations

A major portion of this ISM technical and regulatory guidance discusses Gy's increment sampling methodology, Gy sampling errors, bulk material heterogeneities, sample support, and sampling patterns. It should be noted that, for all statistical data distributions, a simple random sample (discrete or composite) always yields an unbiased (representative) estimate of the population mean. For bulk materials, it is the correct sample support and sampling scheme that matters to obtain an unbiased estimate of a bulk material DU. In bulk material sampling, we are sampling bulk material (e.g., soils) and not values from a data set following some known or unknown statistical distribution. In bulk material sampling concentration distribution does not matter to obtain a representative sample yielding an unbiased estimate of the DU mean. However, concentration distribution plays a role in computing a defensible UCL providing desired coverage to the DU mean.

In Section A.5 an attempt has been made to evaluate ISM incorporating heterogeneities, sample support, and sampling patterns. Examples discussed in Section A.5 confirm Gy's findings, and simulation results described there lead to the conclusion that an ISM sample is a representative sample (yields an unbiased estimate of DU mean) provided increments of appropriate sample support are collected using the simple random sampling scheme.

One of the main objectives of this document is to evaluate the capability of Gy's increment sampling methodology on environmental bulk material DUs in obtaining unbiased estimates of DU means. To address CH, small-scale DH, and GSE, the concept of sample support is introduced in M-3 DU simulations. To demonstrate the importance of sample support and sampling patterns used in obtaining unbiased estimate of the population (DU) mean, ISM increments of specified sample support from M-3 DUs were collected using the three sampling patterns. In M-3 simulations, the

concept of sample support is used to demonstrate how the use of an appropriate sample support can address small-scale DH and GSE resulting in unbiased estimates of the DU mean.

Most of the simulations from bulk material sampling maps represent idealized scenarios in which the DU is relatively homogeneous with respect to bulk material particle mass. For example, maps are used in some simulations to represent a distribution of concentrations throughout a DU but without specifically noting the scale (sample mass or volume) that each coordinate location on the map actually represents (see Section A.3). Other simulations conducted with probability distributions (instead of maps) are equivalent to sampling from a "smoothed" surface with homogeneous concentrations at a small scale (see Section A.2). It is implicitly assumed that the "bulk material" within the DU is homogeneously distributed (however, concentration values within the DU can be highly skewed and may follow spatial patterns) with one and only one point (particle) at each sampling location; and therefore GSE is not present within the DU. It should be noted that Gy proposed the use of incremental sampling to address GSE. The set of simulations for Scenarios M3-A and M3-C (Section A.5.1) attempted to introduce the concept of how differential particle mass and concentration can be addressed through ISM sampling.

To evaluate the performance of ISM in producing unbiased estimates of means of "bulk material" DUs, hypothetical homogeneous and heterogeneous DUs mimicking bulk material (e.g., soils) DUs are generated using the "MIS Module" of the software Scout 1.1 (USEPA n.d. "Scout 1.1"). In addition to bulk material particulates of varying sizes and shapes, a typical DU also consists of uncontaminated items (e.g., trash, twigs, rocks, dead creatures) that are discarded before submission to lab analysis. Moreover, some locations within a DU are inaccessible (e.g., construction, trash, trees, bushes, boulders, ponds etc.). All these factors also contribute to CH and DH within a DU. Due to the presence of CH and DH in a bulk material DU, each location of the DU consists of none (e.g., buildings, bushes and trees representing inaccessible locations) to multiple (e.g., a training range used multiple times) particulates of the bulk material (e.g., soils).

In M-3 bulk material DUs, locations with no points (empty spaces) are considered representing inaccessible locations which cannot be sampled. Keeping these practical scenarios in mind, while generating and sampling M-3 DUs, it is not assumed that each location (e.g., [x,y] location) of the DU consists of one and only one value (particle, point). This phenomena can be best illustrated by using increment samples collected using a pogo stick (e.g., Hewitt et al. 2009). Some increments may consist of trash, twigs, and other materials which will be discarded during the ISM sample preparation process (e.g., drying, sieving) before submitting the incremental samples (ISs) for lab analysis. As a result, each bulk material ISM increment may not be of same mass of the contaminated material.

A typical IS replicate of specified number of increments (e.g., 36, 64) is collected using the sample support of specified radius (e.g., 0.01, 0.05 units). The size of the desired sample support is determined based on the CSM and particle size distributions (also see Section A.6 for details). Using the selected increment collection location as center, all points within the circle of radius 0.05 (chosen tool, sample support) units are included in that increment. Average (mass) of all points in that sample support constitutes an increment. An increment sampling location without any points represents an inaccessible location. When an increment lands on an empty location,

the “MIS Module” of Scout 1.1 moves to the nearest neighboring accessible location to collect an increment of specified sample support (field crew also moves to the nearest sampling location when the chosen sampling location is inaccessible). Using one of the three sampling patterns (e.g., simple random sampling), 36 (or 64) increments of same sample support (mass) are collected in a similar manner. Results for 3 and 5 replicates based on 36 and 64 increments are presented in Section A.5.

As with M-1 and M-2 simulations, the “true” mean must be defined for evaluation of statistical performance from the repeated simulated sampling of the DU (Monte Carlo simulations). For the M-1 and M-2 cases the “true” mean is defined as the average of all increments (i.e., grid nodes). For the M-3 cases the “true” mean is defined as the average of all discrete point values (all particulates with measurable concentration values). All empty spaces can be viewed as representing inaccessible locations and/or trash that will be sieved out.

A.1.2.2 Criticism of the M-3 simulations

The primary criticism of the M-3 simulations focuses on whether the simulations are sampling from the same population as the one from which the “true” mean is being estimated. Two aspects of this concern are described below.

Inaccessible locations. A DU with inaccessible locations still has soil in those locations, and the soil has some characteristics, including concentration levels of the analyte of interest. These inaccessible locations can be handled several different ways in practice. The sampling team can try to avoid placing proposed sampling locations in inaccessible locations, they can ignore those locations if selected, or they can develop a scheme, much like in the simulations in Section A.5, to take nearby soils if the exact location is inaccessible.

It is fairly intuitive that if the sampling team gathers soils from areas near the inaccessible sample locations, they won’t get exactly the same results as for the locations that were selected (by simple random sampling or other random within grid or other method), but they will get something nearby that will generally be similar to the soils in the inaccessible location. Of course, there are times when this will not be quite true. For example, if the contamination is due to aerial dispersion and there is a building that has been in place since before the beginning of the period of contamination, the soils under that building are not likely to be well represented by the soils near the building. Nonetheless, depending on the CSM, it is often a reasonable sampling approach to collect nearby soils with the expectation that they will provide an acceptable surrogate for the actual selected sampling location.

The intent of the sampling exercise discussed here is to determine the average concentration of a particular analyte across the DU. The simulations in this appendix all have that basic goal in mind. The nice thing about simulations is that, unlike in field studies, we know the “truth.” That is, the actual mean concentration across the DU is known in these simulations. With this knowledge, we can look at the outcomes of various simulated approaches, compare them to the known characteristics of the DU, and determine how well the sampling method works for each simulated DU.

To calculate the true mean of the DU, the approach is usually simple. If the simulations are from a probability distribution (PD), the mean is defined by that distribution. If the simulations are from a map with different concentration values assigned to locations on the map, then the concentrations of all locations should be summed and divided by the total number of locations on the map.

The challenge comes when discrete points are spread over a map and not all points on a map are of equal size or there are areas with multiple points or no points (i.e., the points are not distributed evenly and completely across the map). In these cases, even though it is a simulation, the true mean of the entire DU is not actually known, and some method has to be agreed on for estimating it. Focusing only on the maps of DUs with areas that are inaccessible for sampling, an approach to defining the true mean must be defined. Is it reasonable to include only those locations for which the concentration is known and call that the true mean? To do so would be to ignore that there are many areas of the site where the true concentrations are not known. It is certainly possible to estimate the mean at the site by just including the known concentration values. But, if the true mean is estimated that way, then the sampling method would have to use the same approach (i.e., avoid the inaccessible areas) to have results that can be compared to the estimated true mean. If instead, the concentrations across the inaccessible areas are included in the estimate of the true mean (by using the concentrations of their nearest neighbors as reasonable surrogates for them, for example), then the sampling methodology will again have to use a similar approach for the results to be compared meaningfully to the estimate of the true mean. The project team must determine which population is actually of interest (the soils across the entire DU or only those soils that are accessible) and then work from that population throughout their simulations.

If the method used to estimate the true mean does not match the simulation sampling method, then it is not appropriate to compare the simulation results to the estimated true mean. In the case where the true mean is estimated using only the known concentration values and the simulations impute values from near neighbors rather than avoid the inaccessible locations, the simulation results will be estimating the bias that is created by having that mismatched logic. Essentially, what that would be looking at is the following:

1. The true mean, estimated only from known concentrations without any input for the areas of the site that are inaccessible, is equivalent to assuming that the mean of the concentrations in the inaccessible areas is equal to the mean of the known concentrations.
2. The simulated mean, calculated using nearest neighbors when inaccessible sampling locations are selected, is equivalent to assuming that the concentrations in the inaccessible areas are most similar to their nearest neighbors.
3. Comparing this estimated true mean to these simulated results would be a way to assess the bias that occurs if the inaccessible soils are assumed to be most similar to their neighboring soils, but in fact they are just the same, on average, as the soils from across the entire DU and not more similar to those soils nearest to them.

For this reason, the results of the simulations in Section A.5 should not be considered indicative of the implications of ISM sampling of bulk materials but rather an indication of why it is very important to be sure that the population of interest is well defined and the sampling strategy is carefully designed to be sure it is capturing the information that it is intended to capture.

Equal Sample Support. To define sample support in a simulation, some type of unit size must be defined. In actual applications sample support is often thought of in terms of mass. The simulations could define sample support as a 2-D area or include a mass characteristic for each location.

The simulations in Section A.5 discuss the use of increments with the "same sample support." They then say that they use the method of including the number of points within a specified distance of that location (2-D area). However, the documentation does not explicitly state what is done with the points found within the specific distance. Ignoring the issue in the previous paragraphs, the sampling process in Section A.5 identifies a location then searches for all discrete points within a specified distance and defines those points to be included in an increment that then is included in a larger ISM sample. In the A.5 simulations, it is assumed that equal 2-D area is equivalent to "same sample support."

As increments are collected across the sites in Section A.5, some increments can have 1 point included and others can have 10, 20, or more points included in one increment. The concentration level of each point is averaged to provide one concentration value for the increment. This average of the point values (irrespective of how many points were used to make the increment average) is then included with the other increments, assumed to be of equal support, into the ISM sample. Yet the method to define the true mean is based on a raw average of all points thrown on the map.

The two-stage averaging that does not keep track of the number of points per increment is an additional factor that creates a bias from the assumed "true" mean. This bias is not a result of ISM—it is only a function of the algorithm applied in the simulation process. The underlying assumption of equal sample support is a noble goal, but as with the previously stated concern, the actual simulation routines produce results that are indicative of the logic used in the simulations and not of ISM sampling of bulk materials.

A.2 PROBABILITY DISTRIBUTIONS (PD-1)

A series of Monte Carlo simulations was run using probability distributions with different CVs. Table A-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, M-2, M-3) presented later in this appendix. With sampling from probability distributions, each increment is an independent, random sample obtained from the same defined distribution (i.e., identically distributed), which is analogous to using simple random for increment collection if applied to a real site. The assumption is that the overall distribution throughout the DU is homogeneous and

can be described by a single population. It is important to note that, while this approach is useful for conveying important concepts about ISM, sampling from a probability distribution is an oversimplification for the following reasons:

Table A-2. Summary of simulation results using lognormal distributions

Statistic	*95UCL >= true mean [Overestimate of Mean]*								*95UCL < true mean [Underestimate of Mean]*							
	Chebyshev UCL				Student's t UCL				Chebyshev UCL				Student's t UCL			
	2 Reps	3 Reps	5 Reps	7 Reps	2 Reps	3 Reps	5 Reps	7 Reps	2 Reps	3 Reps	5 Reps	7 Reps	2 Reps	3 Reps	5 Reps	7 Reps
	m=30, CV=1								m=30, CV=1							
count of simulations	4,571	4,835	4,956	4,981	4,693	4,664	4,689	4,678	429	165	44	19	307	336	311	322
UCL coverage	91%	97%	99%	100%	94%	93%	94%	94%	91%	97%	99%	100%	94%	93%	94%	94%
mean RPD	27%	22%	18%	16%	37%	16%	10%	8%	-4%	-3%	-2%	-1%	-4%	-3%	-2%	-2%
5th %ile RPD	3%	4%	5%	5%	4%	2%	1%	1%	-11%	-7%	-7%	-3%	-11%	-8%	-6%	-5%
50th %ile RPD	22%	21%	17%	15%	31%	14%	9%	7%	-4%	-2%	-1%	-1%	-4%	-2%	-2%	-1%
95th %ile RPD	65%	48%	34%	28%	91%	34%	20%	15%	0%	0%	0%	0%	0%	0%	0%	0%
	m=30, CV=4								m=30, CV=4							
count of simulations	4,346	4,690	4,852	4,909	4,519	4,430	4,333	4,351	654	310	148	91	481	570	667	649
UCL coverage	87%	94%	97%	98%	90%	89%	87%	87%	87%	94%	97%	98%	90%	89%	87%	87%
mean RPD	93%	80%	63%	55%	129%	57%	36%	28%	-13%	-10%	-7%	-6%	-13%	-10%	-8%	-6%
5th %ile RPD	6%	8%	9%	10%	9%	4%	3%	2%	-30%	-23%	-18%	-15%	-30%	-25%	-19%	-17%
50th %ile RPD	65%	59%	50%	44%	90%	41%	27%	22%	-12%	-8%	-6%	-5%	-11%	-8%	-6%	-5%
95th %ile RPD	272%	214%	155%	129%	374%	157%	92%	73%	-1%	-1%	0%	0%	-1%	-1%	-1%	-1%
	m=30, CV=7								m=30, CV=7							
count of simulations	4,171	4,532	4,740	4,820	4,414	4,187	4,101	4,137	829	468	260	180	586	813	899	863
UCL coverage	83%	91%	95%	96%	88%	84%	82%	83%	83%	91%	95%	96%	88%	84%	82%	83%
mean RPD	140%	117%	94%	83%	189%	86%	55%	45%	-18%	-13%	-10%	-8%	-18%	-14%	-11%	-9%
5th %ile RPD	8%	8%	9%	11%	11%	5%	4%	3%	-39%	-31%	-24%	-21%	-41%	-32%	-28%	-23%
50th %ile RPD	82%	73%	65%	59%	111%	54%	36%	30%	-16%	-11%	-8%	-6%	-16%	-12%	-9%	-8%
95th %ile RPD	457%	358%	271%	227%	609%	272%	164%	133%	-2%	-1%	0%	-1%	-1%	-1%	-1%	-1%
	m=100, CV=1								m=100, CV=1							
count of simulations	4,604	4,827	4,946	4,979	4,720	4,690	4,687	4,669	396	173	54	21	280	310	313	331
UCL coverage	92%	97%	99%	100%	94%	94%	94%	93%	92%	97%	99%	100%	94%	94%	94%	93%
mean RPD	27%	23%	18%	16%	38%	16%	10%	8%	-4%	-3%	-2%	-1%	-5%	-3%	-2%	-2%
5th %ile RPD	3%	5%	5%	5%	4%	3%	2%	1%	-12%	-8%	-5%	-3%	-12%	-9%	-6%	-5%
50th %ile RPD	22%	21%	18%	15%	32%	14%	9%	7%	-3%	-2%	-2%	-1%	-4%	-3%	-2%	-1%
95th %ile RPD	66%	49%	34%	28%	93%	35%	20%	15%	0%	0%	0%	0%	0%	0%	0%	0%
	m=100, CV=4								m=100, CV=4							
count of simulations	4,358	4,674	4,858	4,926	4,547	4,435	4,375	4,395	642	326	142	74	453	565	625	605
UCL coverage	87%	93%	97%	99%	91%	89%	88%	88%	87%	93%	97%	99%	91%	89%	88%	88%
mean RPD	95%	79%	64%	55%	130%	57%	36%	28%	-13%	-10%	-6%	-6%	-13%	-10%	-7%	-6%
5th %ile RPD	6%	9%	10%	10%	9%	5%	3%	2%	-30%	-23%	-18%	-18%	-31%	-25%	-18%	-16%
50th %ile RPD	65%	59%	51%	45%	89%	41%	27%	22%	-11%	-8%	-5%	-5%	-12%	-9%	-6%	-5%
95th %ile RPD	280%	211%	157%	129%	380%	155%	95%	73%	-1%	-1%	0%	0%	-1%	-1%	0%	0%
	m=100, CV=7								m=100, CV=7							
count of simulations	4,115	4,509	4,739	4,839	4,362	4,186	4,092	4,119	885	491	261	161	638	814	908	881
UCL coverage	82%	90%	95%	97%	87%	84%	82%	82%	82%	90%	95%	97%	87%	84%	82%	82%
mean RPD	135%	114%	93%	80%	183%	84%	54%	43%	-17%	-13%	-9%	-8%	-17%	-14%	-11%	-9%
5th %ile RPD	7%	8%	9%	9%	10%	5%	3%	3%	-39%	-29%	-22%	-20%	-38%	-31%	-26%	-23%
50th %ile RPD	82%	74%	64%	58%	111%	53%	36%	30%	-15%	-11%	-8%	-6%	-15%	-12%	-9%	-8%
95th %ile RPD	417%	321%	251%	210%	557%	240%	156%	122%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%

1. 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 sampling principles in the glossary). Therefore, it is convenient to think of the population as having a fixed mean concentration but a variance contingent on the sample mass. The simulations with defined distributions do not explore the effect of sample mass (see discussion of sample support in the glossary) on performance metrics. Instead, it is assumed that the specified variance simply reflects the collective sources of heterogeneity.

2. 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 (i.e., mixtures). This may be true even for very small DUs where concentrations from samples collected within a 1-foot 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.

2. Only lognormal probability distributions are defined. Alternative positively skewed probability distributions were not explored. In general, because lognormal distributions give greater weight to results in the upper tail than alternative choices (e.g., gamma or Weibull distribution), the standard error for the mean and the corresponding UCLs tends be higher than that of comparable distributions with the same population mean and variance.

A.2.1 Methods

Monte Carlo analysis was used to repeatedly apply a specified sampling design (number of increments and ISs) to a DU scenario. Typically between 5,000 and 30,000 trials were used. The large number of trials can be expected to yield relatively stable (i.e., reproducible) results. Each trial represents a complete sampling event (i.e., "n" increments and "r" replicates) and yields an estimate of the population mean, the standard error of the mean, and the UCL. Collectively, the results yield a distribution of 95% UCLs that can be used to calculate the performance metrics. For example, ideally, the sampling method and UCL calculation yield a probability distribution of UCLs with a 5^{th} percentile equal to (or greater than) the true population mean. This would mean that one can expect that the sampling design applied to this type of population will achieve the desired coverage (or percentage of exceedences of the true mean) of 95%. Table A-2 provides examples of simulation experiments with coverages that vary from approximately 80% to 100%.

Multiple ISM samples (i.e., replicates) must be collected to calculate the standard error and UCL. The expected small sample sizes (e.g., three to seven replicates) for most implementations of ISM preclude the use of bootstrap resampling techniques to calculate a UCL; therefore, simulations were performed using only the Student's-*t* and Cheybshev UCL methods, which are based on the sample size, sample mean, and variance (see equations at the end of this appendix).

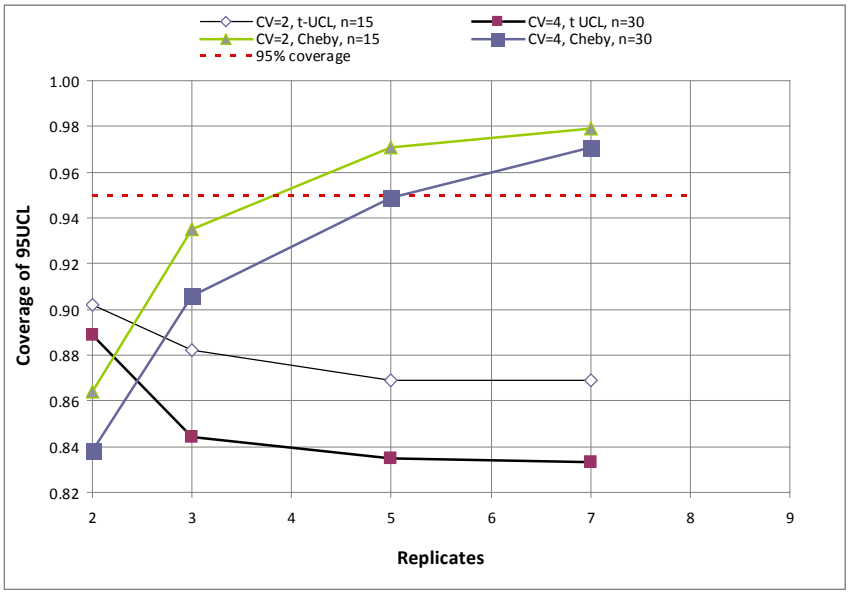
Because the distribution of sample means tends to exhibit less skew than the population due to the central limit theorem, the performance of the Student's-*t* UCL can vary. The Student's-*t* can be expected to yield the most reliable performance metrics for populations with a low (e.g., ≤1) CV. By contrast, the Chebyshev generally yields higher UCLs with higher coverage but also higher RPDs.

Generally, sampling designs were varied 15–100 increments and 2–7 replicate ISM samples. The mean of the distribution represents the population mean and is used to calculate the bias and relative percent difference metrics.

The number of replicates is used to represent the degrees of freedom in UCL calculations using ISM.

A.2.2 Results

Figure A-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 95UCL		Chebyshev 95UCL	
			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 A-1. Examples of simulation results using lognormal probability distributions with CV equal to 2 and 4, increments of 15 and 30, replicates ranging 2–7, and two 95% UCL calculation methods (Cheby = Chebyshev; *t*-UCL = Student's-*t*).

These examples are useful for illustrating the following general patterns that emerge from the simulation experiments with lognormal distributions:

1. 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 (i.e., *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 2 or more to the calculation of the UCL compared to sampling designs with three or more replicates.

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

2. 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.

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%
3	15	5	95%				
	30	4	95%				
	50	4	96%				
	100	3	95%				

The coverage of the Student's-*t* UCL does generally not achieve 95% and does not increase with increasing samples sizes (increments and replicates) within a practical range.

3. The RPD between the 95% UCL and the population mean is generally greater for the Chebyshev than the Student's-*t*, particularly for trials in which the 95% UCL actually exceeds the population mean. Therefore, the tradeoff with the Chebyshev UCL is that it achieves more reliable coverage but also higher UCLs. See Section 4 for a side-by-side comparison of the range of 95% UCL RPDs for different scenarios for both the Chebyshev and Student's-*t* UCL.

4. The simulations with lognormal distributions yield unbiased estimates of the mean.

A.3 SPATIAL AUTOCORRELATION MAPS (M-1)

For most sites, contaminants in soil exhibit some degree of spatial relationship, meaning that often variance in the concentration reduces as the distance between sample locations decreases. It is well established that strong spatial relationships can reduce the effective sample size of a data set because each sample provides some redundant information (Cressie 1993). In statistical terms, this redundancy violates the assumption that observations are independent. ISM confidence intervals 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, e.g., simple random sampling. The issue of spatial relationships applies to discrete as well as ISM sampling.

A.3.1 Methods

Simulations were run to evaluate the effect of spatial autocorrelation on the performance of ISM. Figure A-2 shows a map generated from a real data set of more than 200 observations. The sample results were interpolated with inverse distance weighting techniques to yield a 2-D surface of concentrations. Such spatial "smoothing" is likely to underestimate the small-scale heterogeneity (DH) in concentrations that exists at most sites. Therefore, 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; therefore, the simulations with quadrants represent an overall four-fold 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.

A.3.2 Results

Table A-3 summarizes the simulation results with 1000 Monte Carlo trials using 30 increments and 3, 5, and 7 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 3 replicates, the Chebyshev yields 96%–97% coverage, whereas the 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 2 with systematic grid sampling. Introducing spatial autocorrelation tends to result in an improvement in the RPD metric. This is most likely because the autocorrelation affects the correlation between the sample mean and variance. For nonnormal 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.

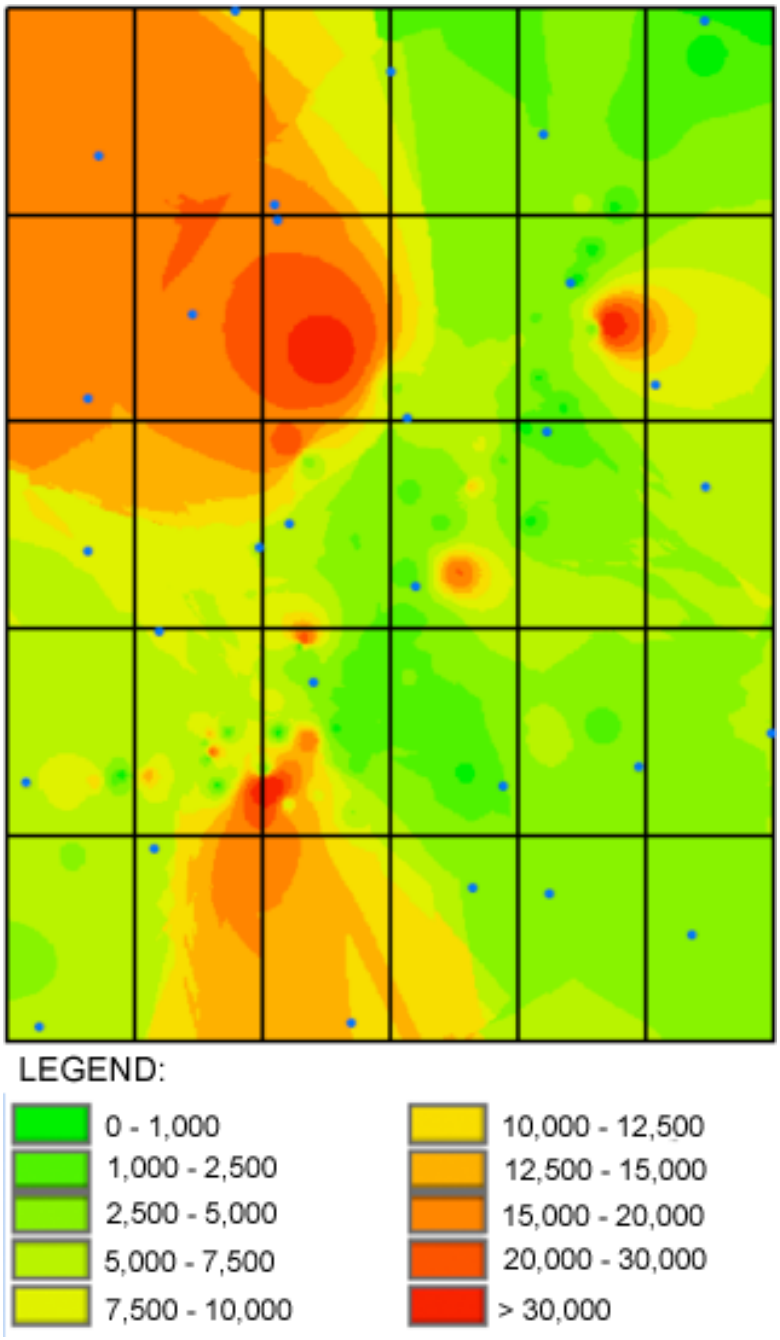


Figure A-2. Example of a map with high spatial autocorrelation (Moran's I z-score = 3.8). Throughout the entire DU (all grid cells combined), the population mean is 8564 and standard deviation is 6507 (CV = 0.7).

Table A-3. Summary of simulation results for a site with high spatial autocorrelation (see map in Figure A-2)

Decision Unit: Map with High Spatial Autocorrelation

Population Parameters

mean:	8,564	**median:**	6,476
SD:	6,087	**min:**	170
CV:	0.71	**max:**	57,378

Sampling: Simple Random Sampling (i.e., mimics no spatial autocorrelation)
Trials: 1,000
Increments: 30

95UCL 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 95UCL 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

Decision Unit: Map with High Spatial Autocorrelation

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

95UCL 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 95UCL 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; this 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.

A.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 (Hathaway and Pulsipher 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 A-3 show each of these sites in 2-D 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.

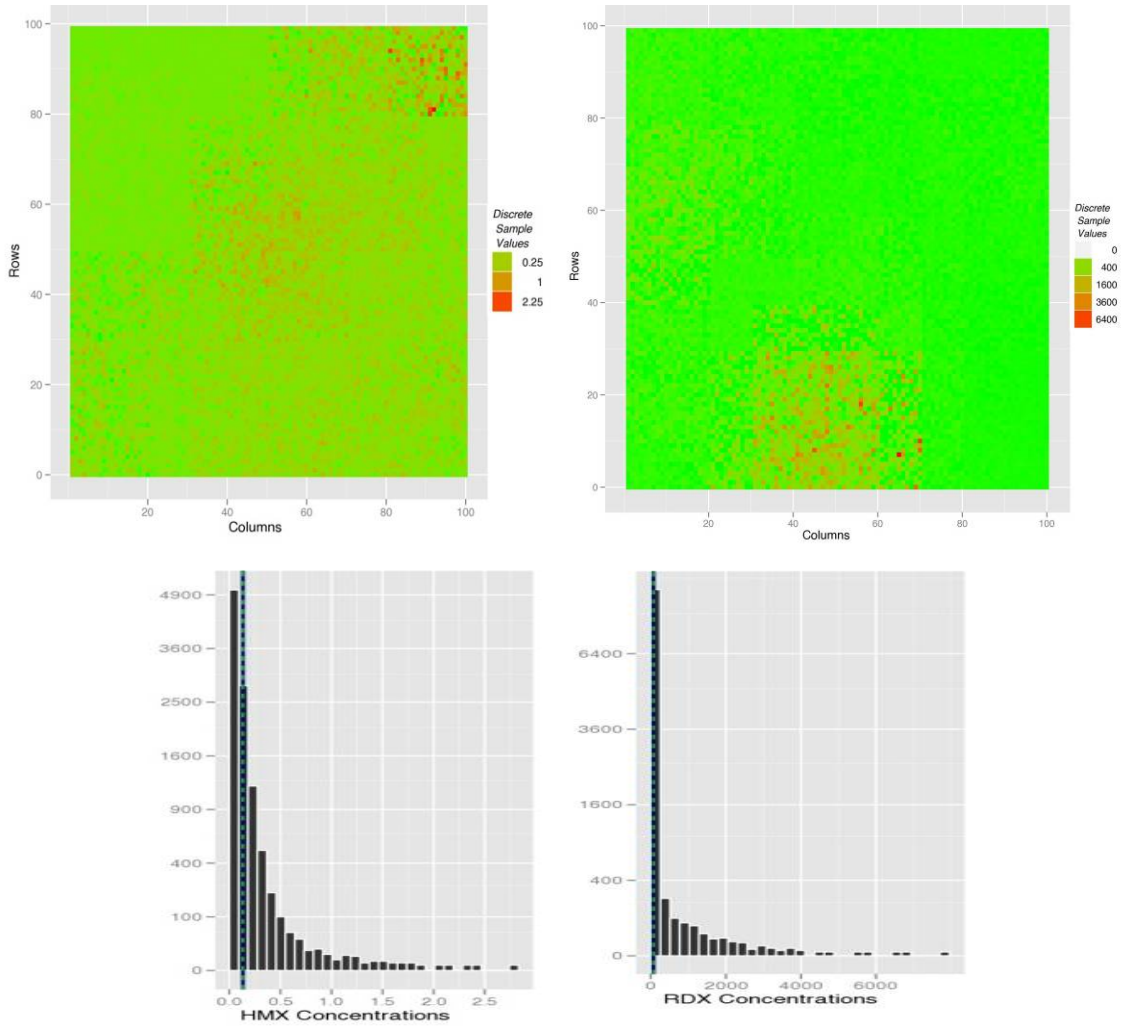


Figure A-3. Spatial distributions and histograms of concentrations for two simulated sites.[6]

A.4.1 Descriptions of Decision Units

Hathaway and Pulsipher (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 A-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 project led by Jenkins and Hewitt (Jenkins et al. 2004, 2006; Hewitt et al. 2005). Each grid value (increment) in Figure A-3 represents the agglomeration of the bulk material from that area with reported values of constituent levels in units of milligram per kilogram (or parts per million). Thus, as with the simulations with lognormal distributions (PD-1), FE and GSE were not explicitly used in

[6] 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 (i.e., "true mean") from which performance metrics are determined.

simulating these sites. These errors are implicitly accounted for in the modeled small-scale (local) spatial variability.

A.4.1.1 HMX decision unit (M2-A)

The HMX concentrations (mg/kg) shown in Figure A-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, standard deviation of 0.15, and maximum of approximately 2.3 mg/kg.

A.4.1.2 RDX decision unit (M2-B)

The RDX concentrations (mg/kg) shown in Figure A-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 spatial heterogeneity with a CV of approximately 4.5 (standard deviation = 319 mg/kg; mean = 71.4 mg/kg).

A.4.2 ISM Sampling Patterns

Figures A-4–A-7 show the 64 different IS patterns that are evaluated and summarized in the results section. For all four figures, each row of plots represents a different number of replicates gathered from the DU (2, 3, 4, and 5), and each column of plots identifies a different number of increments per replicate (16, 30, 49, and 100). Figures A-4 and A-5 show the standard IS procedure with replicate ISs over the entire DU for systematic and random grid sampling. respectively. Figures A-6 and A-7 represent the grouped IS methods for systematic and random grid sampling, respectively. These figures show the general structure for each of the evaluated patterns but represent only an example of one random selection for each pattern. Figure A-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.

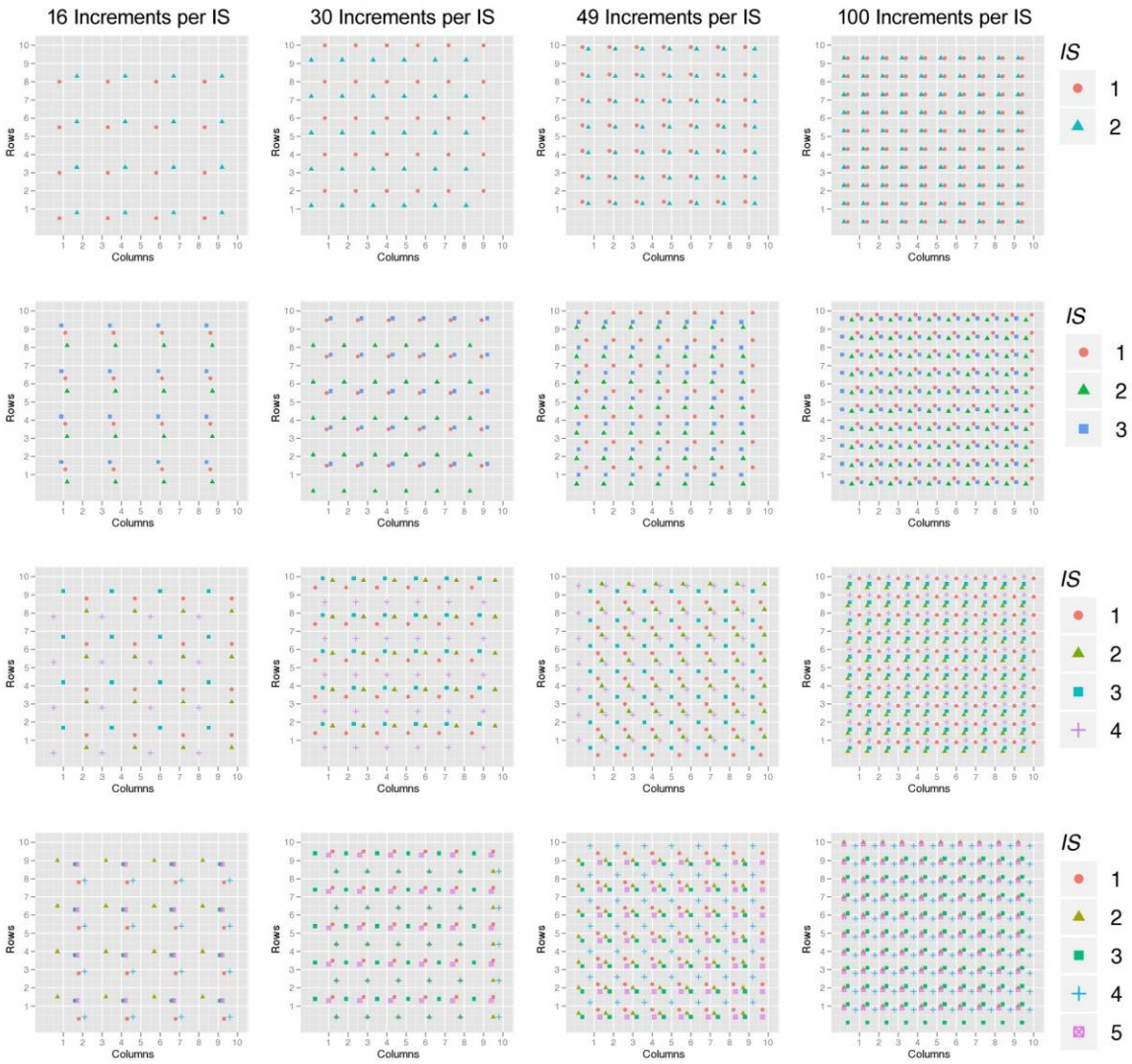


Figure A-4. Standard incremental sampling using a systematic grid sampling approach. Each column represents a differing number of increment per IS, and each row depicts the differing number of ISs that were gathered.

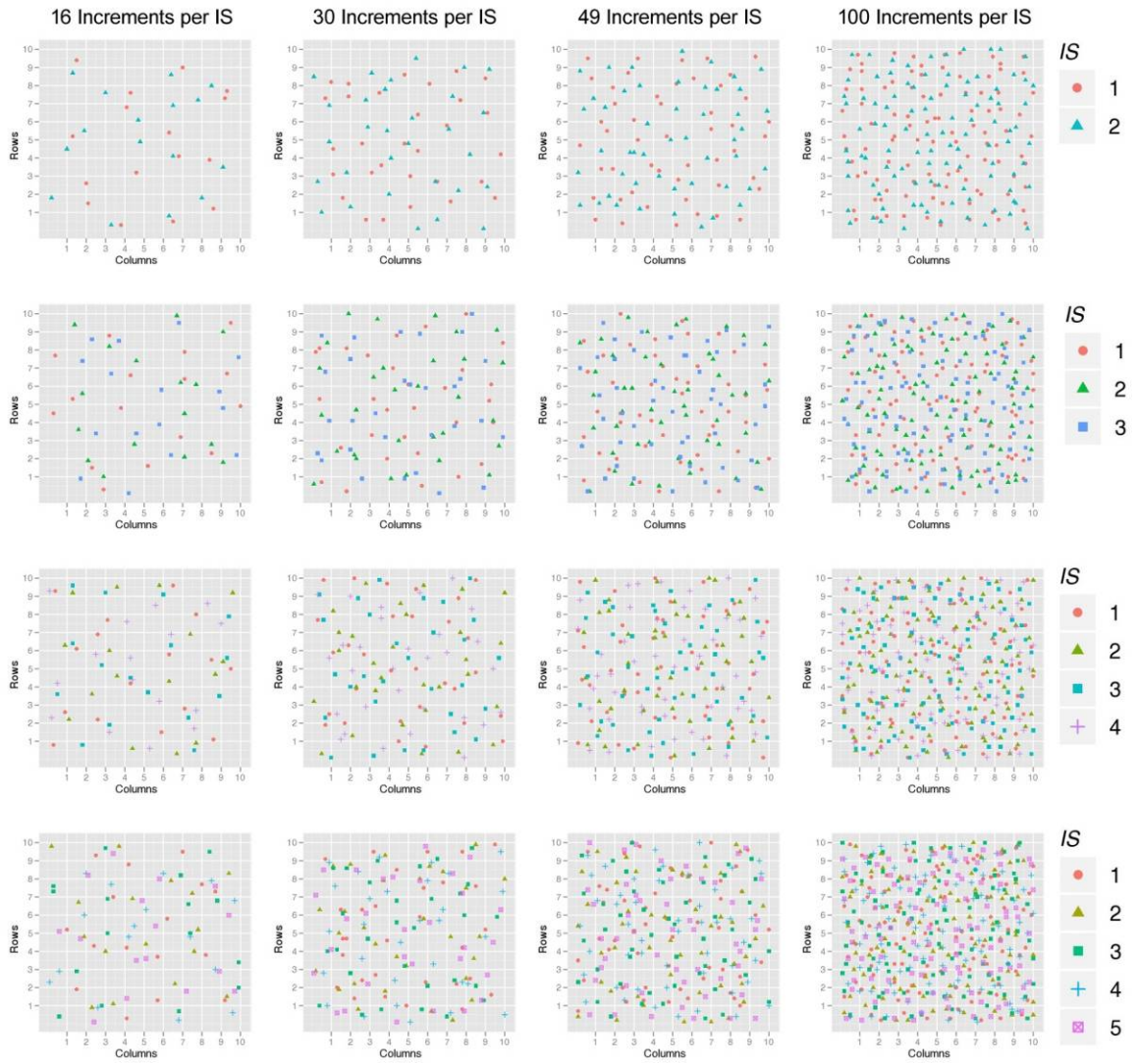


Figure A-5. Standard incremental sampling using a random grid sampling approach. Each column represents a differing number of increment per IS, and each row depicts the differing number of ISs that were gathered.

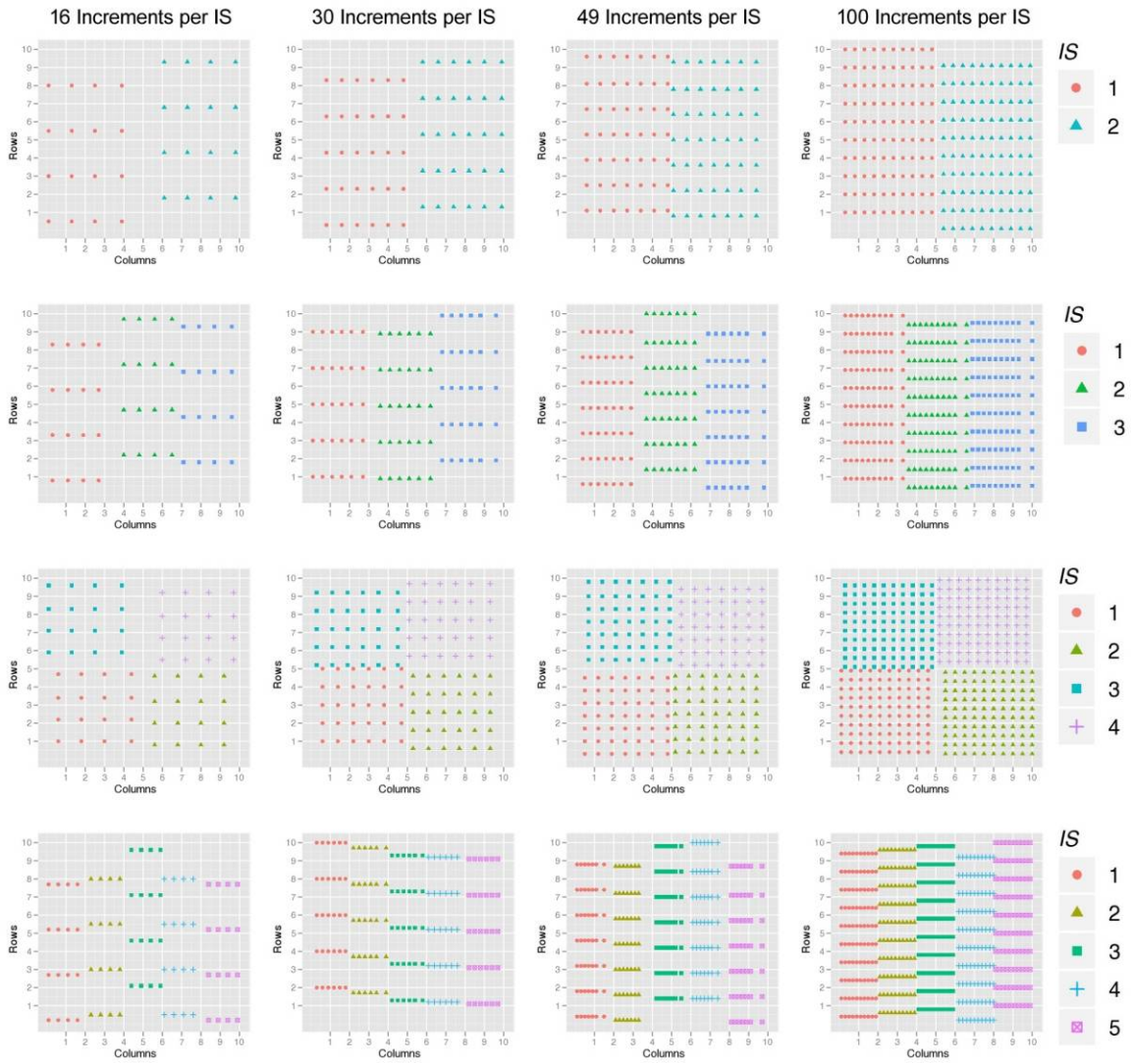


Figure A-6. Grouped incremental sampling using a systematic grid sampling approach. Each column represents a differing number of increment per IS, and each row depicts the differing number of ISs that were gathered.

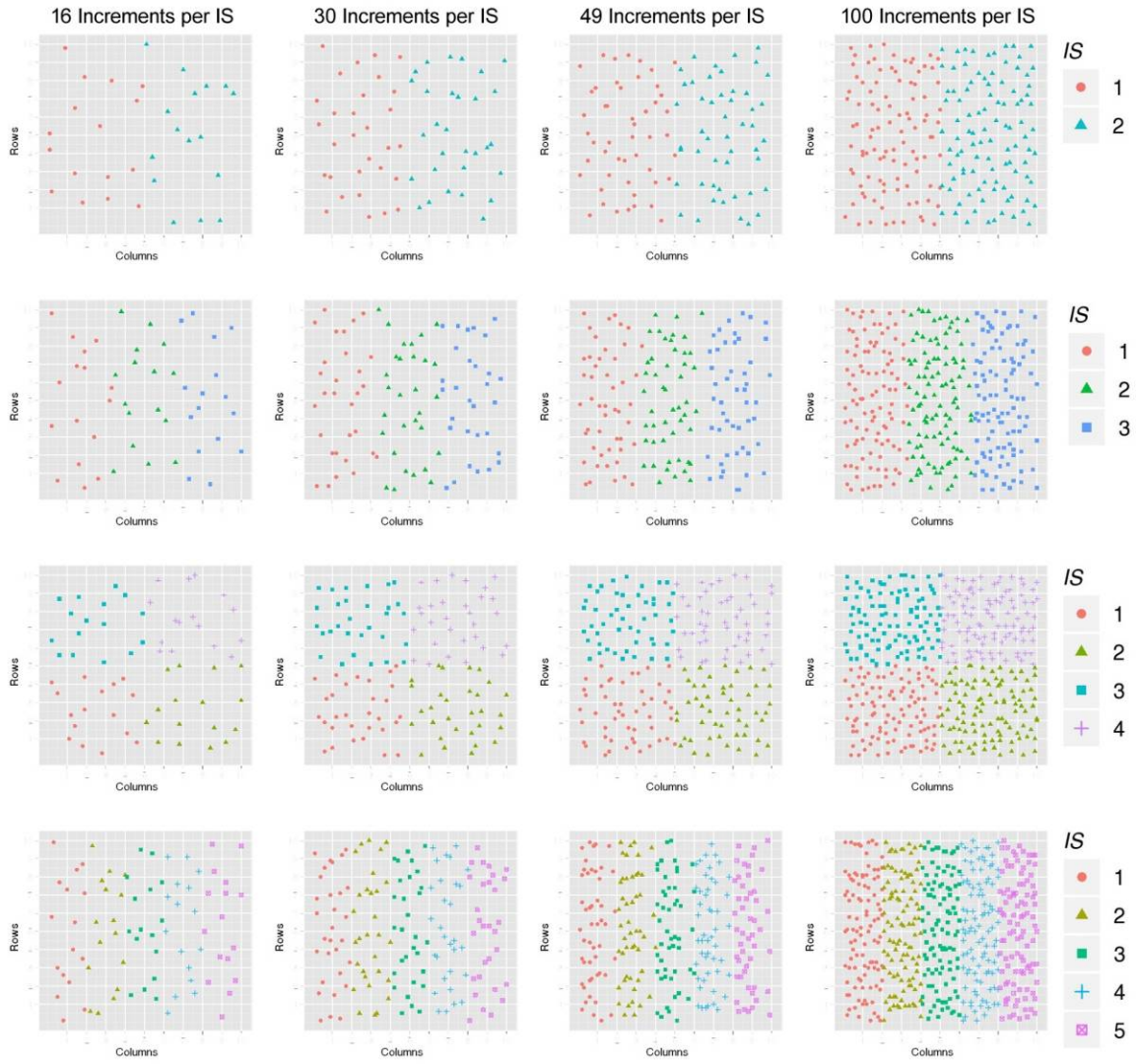


Figure A-7. Grouped incremental sampling using a random grid sampling approach. Each column represents a differing number of increment per IS, and each row depicts the differing number of ISs that were gathered.

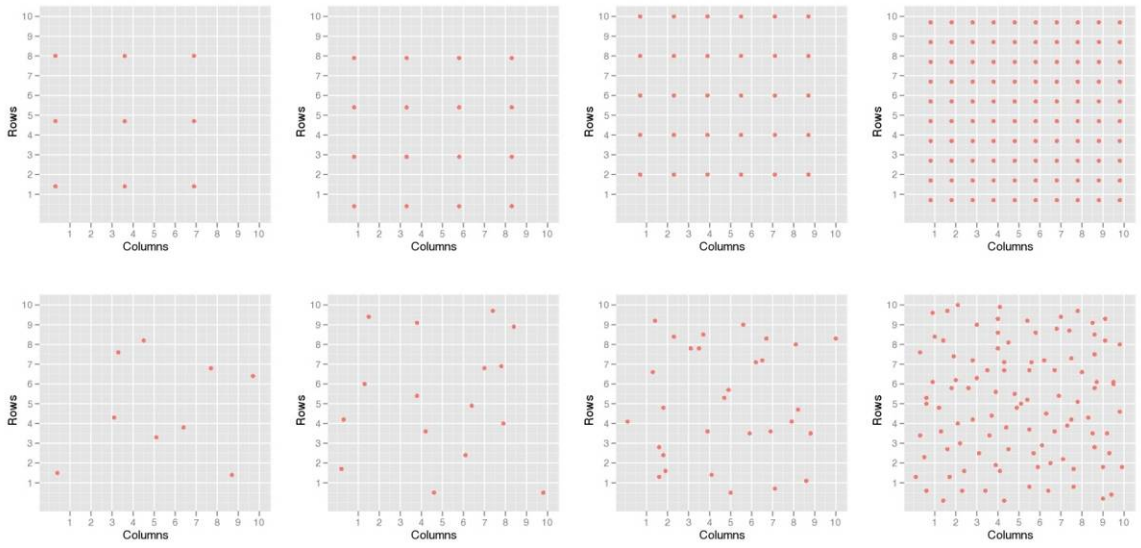


Figure A-8. Discrete sampling using a systematic grid (top row) and random grid (bottom row) sampling approaches. Each column represents a differing number of increments or discrete samples (from left to right 9, 16, 30, and 100 samples per evaluation).

A.4.3 Results Using Discrete Sampling

Table A-4 shows a few of the 2000 iterations from the UCL calculations based on using the mean and standard error calculated from nine systematic grid discrete samples (see upper left plot in Figure A-8) from a DU. These values represent absolute concentrations (e.g., mg/kg). 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 2000 iterations (i.e., 5%). Figure A-9 shows a histogram of 2000 UCL values from one simulation scenario where the y-axis represents the percentage of 2000 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%. The simulation results provide an example demonstrating how one of the performance metrics (coverage of the UCL) may indicate that an ISM sampling design is unlikely to yield reliable results.

Table A-4. Example of mean and 95% UCL calculations for each iteration of a simulation

Mean	UCL
0.61	0.76
0.72	0.94
1.01	1.46
0.79	1.18
⋮	⋮
0.81	1.02

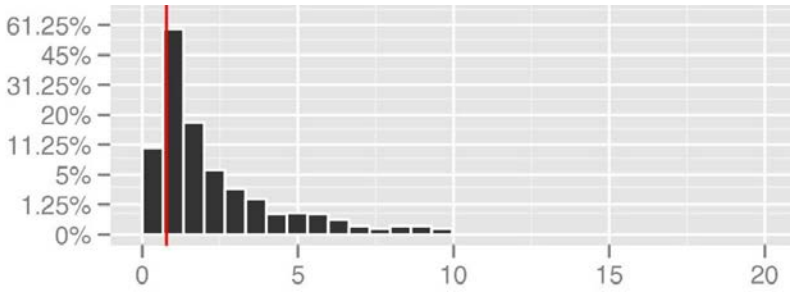


Figure A-9. Histogram of the calculated UCL values using a simulated data set with 2000 iterations.[7]

The discrete sampling examples were restricted to calculations using Student's-*t* UCL and Chebyshev UCL. 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 spatial 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.

A.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 IS patterns were used. Differences in results for these sampling routines were within the range of simulation (stochastic) error. Figure A-10 shows an example of the equal coverage for both M2-A and M2-B using the three different standard IS 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.

[7] For display purposes the y-axis is in terms of percentage of 2000 and is distorted (not evenly spaced between ticks) to highlight the low count bins. The red line identifies the true mean of 0.776.

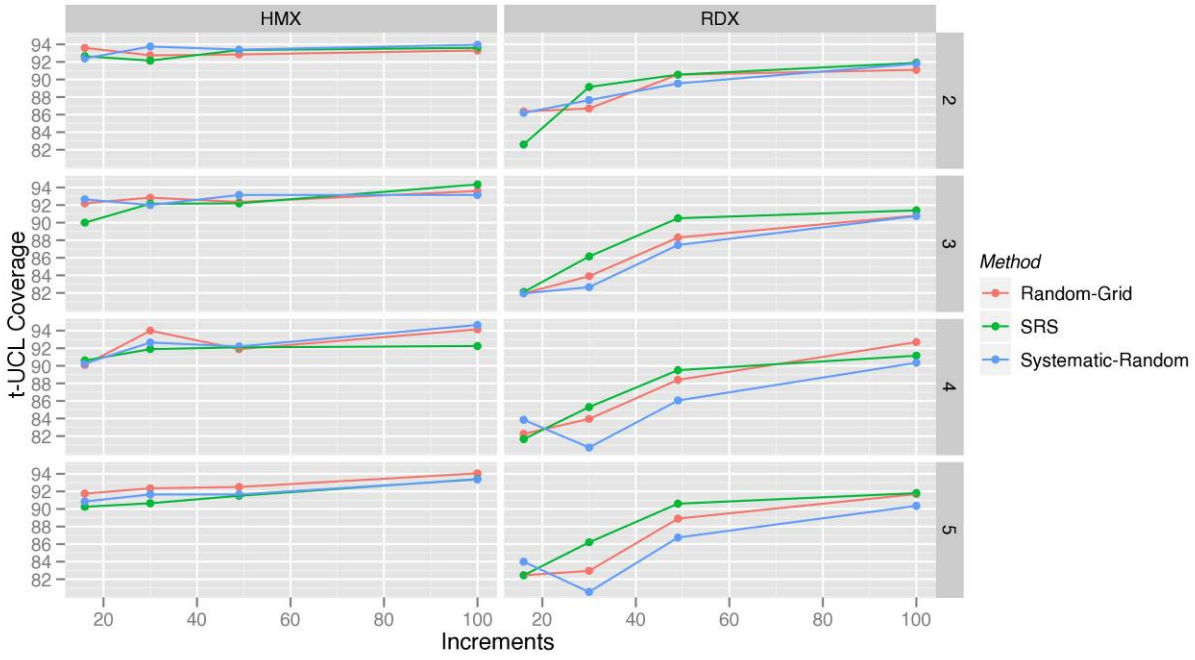


Figure A-10. A coverage plot comparing systematic grid (with random start), random grid, and simple random sampling for the RDX DU (M2-A) and HMX DU (M2-B) when 2, 3, 4, or 5 ISs are collected from the DU.

The tables shown in each section will be separated into the three general sampling patterns—standard IS, grouped IS, discrete sampling. Each table summarizes the results from 2000 iterations. The first two columns are different for the IS and discrete summary tables. For the IS summary tables, the first column identifies the number of ISs sampled from within the DU, and the second column shows the number of increments in each IS. 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 standard deviation. 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 of the UCL 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 of the UCL from the true mean for those UCL values that were below the true mean.

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 IS for the IS designs and total for discrete designs). Each colored line represents a different sampling pattern with a separate plot for the discrete, grouped IS, and standard IS. 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 standard deviation for each scenario. For the IS sampling patterns df is the number of IS replicates gathered from the site minus 1. 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 the 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). 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 et al. 2007) and Visual Sample Plan (Matzke et al. 2007) for use in environmental studies. There are a variety of choices depending on site-specific needs.

For the proposed IS 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 standard error 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 standard error. 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 IS 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 increases the most.

A.4.4.1 Results for RDX (M2-A)

For the RDX (10 m × 10 m DU) simulations, Tables A-5 through A-7 show the summaries from the evaluated simulations. The coverage, bias, number of increments, and number of ISs are used to create the coverage plot shown in Figure A-11. Figure A-12 shows the panel of *t*-UCL histograms for all 40 sampling patterns evaluated on the RDX 10 m × 10 m DU.

Table A-5. Discrete summary: RDX decision unit (M2-A)

Grid sampling type	Number of increments	Chebyshev UCL coverage	*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 A-6. Standard ISM summary: RDX decision unit (M2-A)

Number of ISs	Number of increments	Chebyshev UCL coverage	*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 A-7. Grouped ISM summary: RDX decision unit (M2-A)

Number of ISs	Number of increments	Chebyshev UCL coverage	*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		

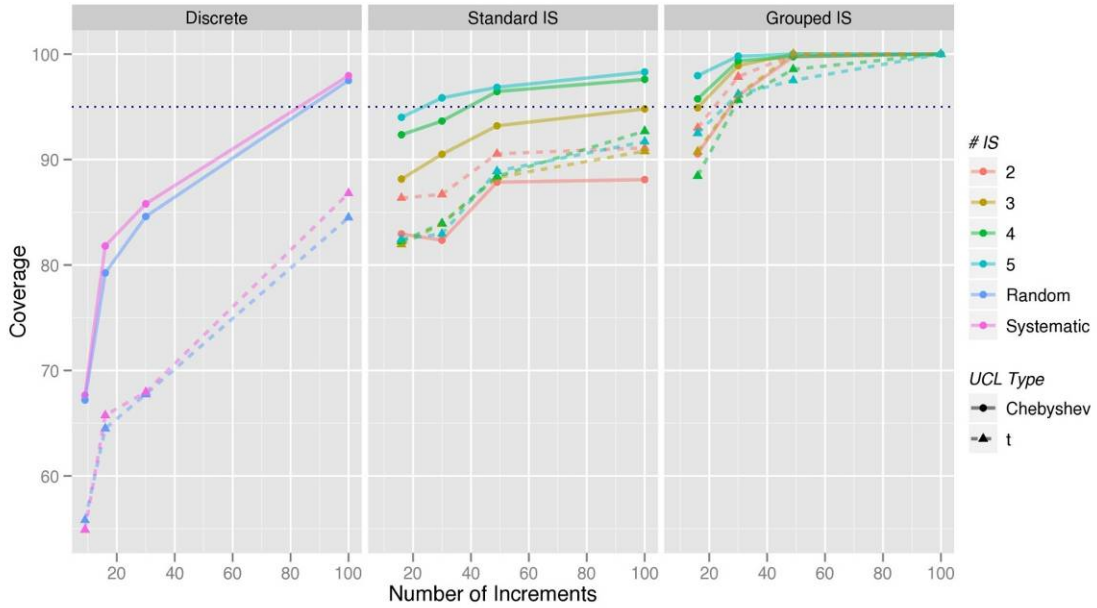


Figure A-11. Plot of the coverage statistics for each of the simulated sampling patterns as applied to the RDX DU.

(Note: The different sampling patterns are displayed within the plot as well as UCL type).

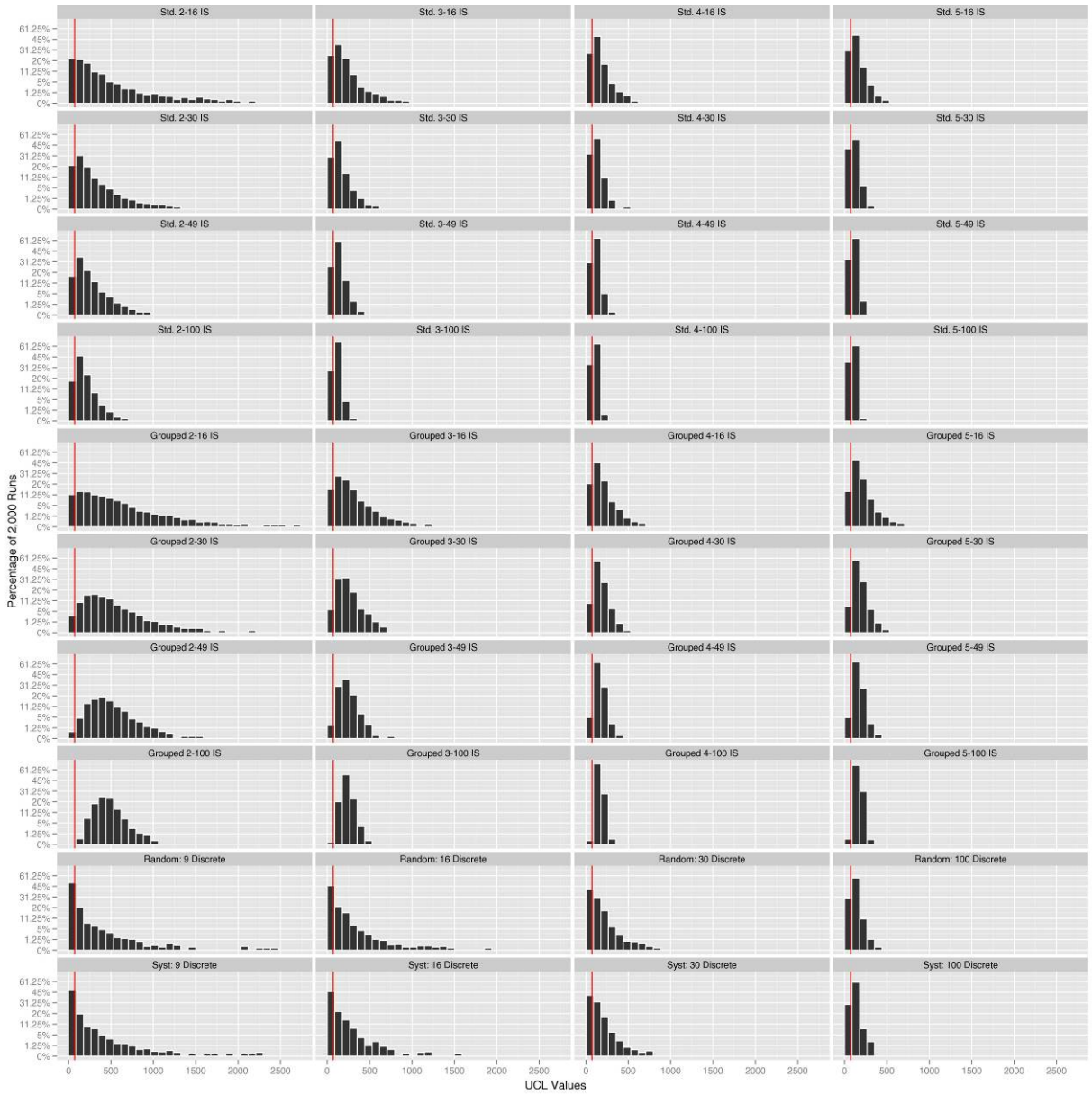


Figure A-12. Panel of histograms of the distribution of *t*-UCL values for the 2000 simulations. (Note: The red line identifies the true mean. The y-axis identifies the percent of 2000 simulations in each bin and is distorted to show the percentage in the low count bins.)

This site had the strongest small- and large-scale spatial heterogeneity of the two DUs evaluated with a CV of 4.47. The mean is 71.36 with a standard deviation of 319.1. The coverage results for the standard IS perform reasonably well for the IS designs of 100 increments per IS. The grouped IS patterns were above the designed criteria of 95% for all but the IS composed of 16 increments. For this DU the grouped ISs are the only patterns that consistently met or exceeded the designed 95% coverage but did have more bias in the mean than the standard IS or discrete methods.

Figure A-13 shows the distribution histograms for the 2000 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 IS, the distribution of means becomes more normally distributed. Both the grouped and standard IS designs provide unbiased estimates of the mean (71.36) and have virtually identical distributions.

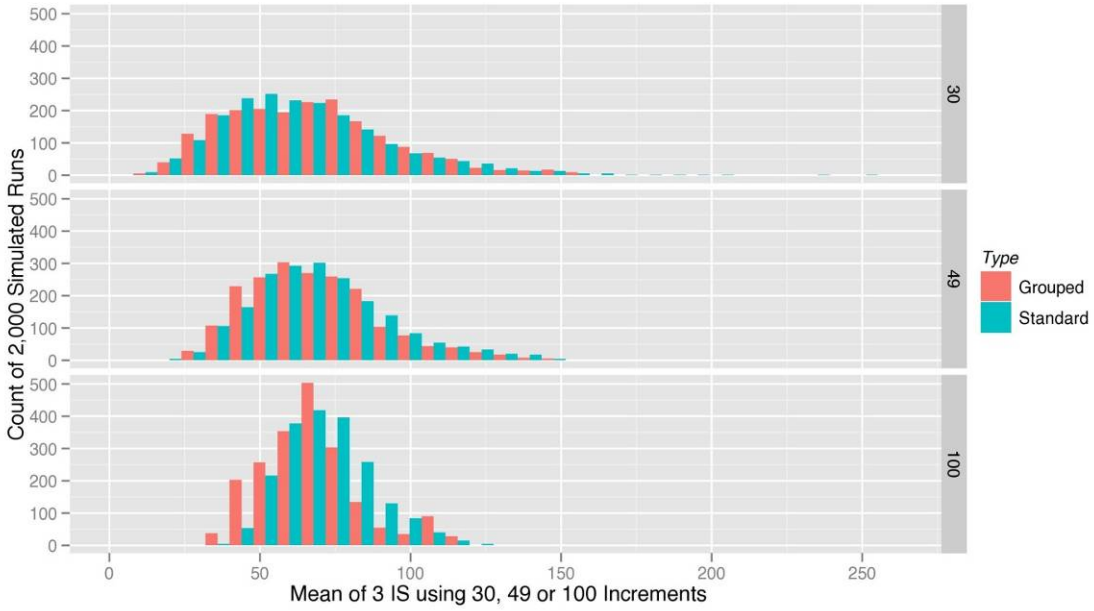


Figure A-13. A comparison of the distribution of means for grouped and standard IS designs using the RDX DU. (Note: Results are similar for all other DUs).

A.4.4.2 Results for HMX (M2-B)

For the HMX (10 m × 10 m DU) simulations, Tables A-8 through A-10 show the summaries from the evaluated simulations. The coverage, bias, number of increments, and number of ISs are used to create the coverage plot shown in Figure A-14. Figure A-15 shows the panel of *t*-UCL histograms for all 40 sampling patterns evaluated on the HMX 10 m × 10 m DU.

Table A-8. Discrete summary: HMX decision unit (M2-B)

Grid sampling type	Number of increments	Chebyshev UCL coverage	*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 A-9. Standard ISM summary: HMX decision unit (M2-B)

Number of ISs	Number of increments	Chebyshev UCL coverage	*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 A-10. Grouped ISM summary: HMX decision unit (M2-B)

Number of ISs	Number of increments	Chebyshev UCL coverage	*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

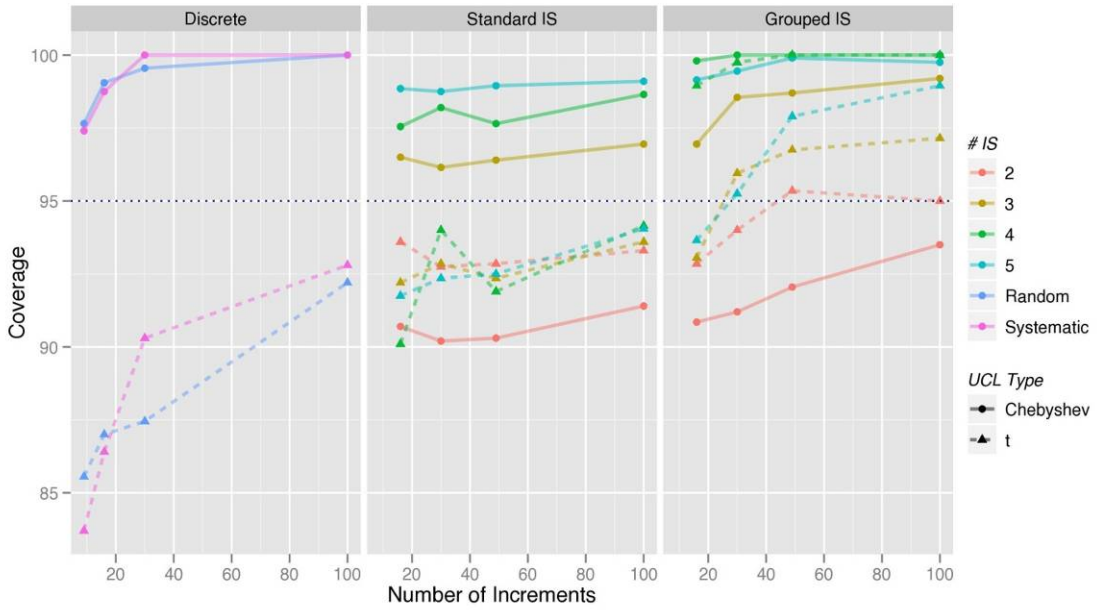


Figure A-14. Plot of the coverage statistics for each of the simulated sampling patterns as applied to the HMX decision unit.
(Note: The different sampling patterns are displayed within the plot as well as UCL type).

This DU has some strong spatial 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 a standard deviation of 0.146. When three or more replicates are used, the coverage results for the grouped IS patterns were near or above the designed criteria of 95% for all but the IS composed of 16 increments. The standard IS performed reasonably well for the 100-increment standard IS design.

Specific observations from these simulations are noted below and support the consensus points listed in Table A-1:

- The mean concentration estimates for grouped ISM and standard ISM sampling have the same expectation and distribution (see Figure A-13).
- The grouped ISM methods have equivalent or greater coverage than standard IS when the same number of ISs and increments are used.
- The RPD of the UCLs for grouped ISM is generally higher than that of standard IS.
- Grouped IS, 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 A-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 (e.g., 50–100) instead of the number of replicates (e.g., 3–5).

Figure A-15. Panel of histograms of the distribution of *t*-UCL values for the 2000 simulations. (Note: The red line identifies the true mean. The y-axis identifies the percent of 2000 simulations in each bin and is distorted to show the percentage in the low count bins.)

A.5 BULK MATERIAL SAMPLING MAPS (M3-A TO M3-C)

In this section, sampling (e.g., ISM) is conducted on DUs consisting of heterogeneous bulk material with CH and DH; appropriate increment sample support(SS)/mass and increments collection sampling patterns are used to address GSE and FE. Increments of predetermined SS (e.g., radius of 0.01, 0.05 units) are collected, and all points contained in that SS make that increment. This operation addresses small-scale DH and GSE. Since FE cannot be eliminated (addressed) completely, it is essential to address GSE to obtain an unbiased estimate of the DU mean. It is also shown that the use of simple random sampling is needed to obtain a representative ISM sample yielding an unbiased estimate of the DU mean, an observation supported by statistical sampling theory (Cochran 1977; Elder, Thompson, and Myers 1980).

A.5.1 DU Generation and ISM Implementation Method

Simulation experiments with three additional maps (scenarios M3-A to M3-C) were performed on hypothetical DUs consisting of bulk material particulates with CH and DH. Section A.1.2 provides a detailed description of the generation of M-3 DUs and implementation of ISM on those DUs. The simulations address the following concepts relative to bulk material sampling and sample support:

- DUs can consist of bulk material (surface soils) with varying degrees of heterogeneities.
- ISM sampling designs can be described in terms of Gy sampling principles (see Section A.6.7).
- Simulations with ISM yield performance metrics, including bias in the mean, standard deviation of the relative bias, and coverage of the 95% UCL.

Issues related to the optimal increment sampling pattern and optimal number of increments and replicates needed to collect representative ISM samples from heterogeneous bulk material DUs are also discussed. When sampling for bulk material, it is not possible to collect single particles, creating the potential to introduce GSE. Presence of GSE and FE in bulk material tends to yield biased samples. Gy proposed the use of incremental sampling to address GSE and then combining all increments to address FE. A key concept associated with bulk material increment sampling is that each increment is associated with a specific sampling location of a DU represented by a 2-D map (e.g., Figures A-16, A-18), and a typical collected IS of specified SS consists of all contaminated particles as well as other potentially uncontaminated items (including trash, twigs, pebbles, dead creatures, etc.) found at that location using an appropriate sampling tool (e.g., pogo stick). In bulk material increment sampling, one is collecting increments of “equal mass” with SS of same size. Each increment potentially consists of many contaminated particles as well as other uncontaminated items which are discarded during ISM sample preparation process (e.g., drying, sieving). Simulations were implemented using software called MIS (Singh, Maichle, and Armbya 2009), which generates 2-D DUs representing surface bulk materials (e.g., soils) and implements alternative ISM sampling designs. Gy sampling concepts, terms, and equations relevant to ISM are defined at the end of this appendix (see Section A.6.1).

In all simulation experiments (PD, M-1, M-2, and M-3) considered in this document, a concentration can be associated with the sample, but the scenarios represent very different perspectives regarding the heterogeneities (and sampling errors) that are captured by the calculation of variance. If each sampling location is represented by a single particle (with some concentration), GSE is not present because each location of the DU consists of one and only one data value, analogous to sampling a single item from a batch of discrete items (Smith 2006). In addition, for DUs with analytical results generated using a population described by a single distribution (e.g., some of the scenarios presented in prior sections), small-scale and large-scale DH does not factor into the results. The scenarios presented in this section are intended to reinforce the following observations regarding bulk material sampling:

- When there is minimal small-scale DH (heterogeneity in particle distribution), there is negligible GSE.
- An important performance metric with ISM sampling is FE in the mean estimate. If GSE is not present (e.g., as in M-1 and M-2 DUs), bias or FE in the mean estimate is negligible using all sampling patterns (e.g., serpentine, simple random, random within grids) and lowest with simple random sampling.
- Bias or FE in the mean estimate cannot be reduced by increasing the number of ISM replicates.
- For heterogeneous DUs with CH and DH, one can obtain an unbiased estimate of DUs mean provided increments are collected using appropriate sample support following simple random sampling pattern. These facts and observations are illustrated by bulk material DU example used in DU M3-A (see Section A.5.2).
- ISM cannot identify spatial or temporal patterns present in any DU (especially if the DU is not divided into multiple SUs), as shown in Example M3-C (see Section A.5.2) using a large real radium-226 data set of over 15,000 points

For homogeneous bulk material DUs with all contaminant particulates of same size and shape which are distributed evenly (e.g., one and only one particulate) throughout the DU (concentration distribution can be highly skewed), all sampling patterns yield unbiased estimates DU mean. For such homogeneous DUs, the size of the sample support does not matter much in reducing the bias in the mean estimate. These observations are illustrated by using a homogeneous bulk material DU in Example M3-B (see Section A.5.2). Just as heterogeneities are difficult to quantify in practice, they are difficult to represent in maps used to define hypothetical DU scenarios. For example, in practice, not all locations in a DU can be sampled due to obstacles such as trees, buildings, boulders, and water; therefore, a truly random sampling design is difficult to implement in practice. For the maps used to represent DUs in this section, accessible sampling locations are represented by points/particulates and inaccessible locations are represented by empty spaces. When an empty space (inaccessible location) is encountered, an increment of the same SS is collected from a neighboring location. These issues are discussed earlier in M-3 simulation section of Section A.1.2. Simulated ISM increments located at empty spaces yield nonzero results by drawing values from a local neighborhood of results. Therefore, the process used to simulate a DU map (e.g., smoothing/interpolation) has important implications as far as the extent to which the modeling results may inform real-world conditions.

In contrast to the M-3 maps, the M-1 and M-2 maps do not include areas that are inaccessible because the maps were generated in a way that interpolated between the original points. Thus, the true means of those DUs actually represent estimated means. Those estimated "true" means are then estimated using ISM. It is a well-known fact that a sample (e.g., consisting of 36 increments) obtained using simple random sampling from a discrete data set (e.g., PD DUs or M-2 DUs) yields an unbiased estimate of the mean of the population represented by that data set.

A.5.2 Simulation Results for Scenario M3-A

Figure A-16 shows a hypothetical simulated heterogeneous target DU with CH and DH. This DU represents a typical training target scenario in which the deposition (and density) of particles (with contaminant loading) decreases as one moves farther from the center of the target. Therefore, the concentrations are highest in the center—the mean concentration around the target central area is about 614 mg/kg, compared to a mean outside the target area (i.e., background conditions) of approximately 3 mg/kg. The overall mean in the DU is approximately 492 mg/kg. Tables A-11 and A-12 give summary statistics for population parameters.

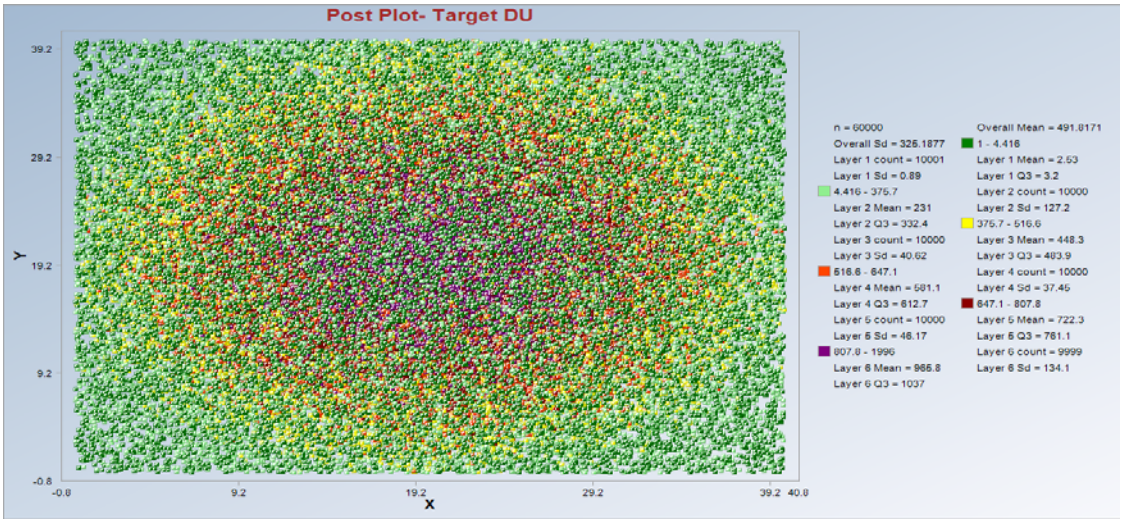


Figure A-16. Post plot for Scenario M3-A representing a shooting range with a central target area (mean concentration = 492 mg/kg).

Table A-11. Summary of population parameters for the DU given by scenario M3-A, entire area and individual quadrants

Statistic	DU	Quadrant 1	Quadrant 2	Quadrant 3	Quadrant 4
Percent of total	100%	25.13%	24.93%	24.95%	25.00%
Number of cells	60,000	15,077	14,957	14,969	14,997
Arithmetic mean	491.8	493.6	491.9	492.7	489.1
Median	516.6	519.1	516.5	516.8	513.6
Minimum	1	1	1	1	1
Maximum	1,996	1,996	1,627	1,732	1,802
Standard deviation	325.2	326.8	324.3	325.5	324.2
CV (RSD)	0.661	0.662	0.659	0.661	0.663

Table A-12. Summary of population parameters for target area and background area in the DU given by scenario M3-A

Statistic	Target	Background
Percent of total	80%	20%
Number of cells	48,000	12,000
Mean	614	3.1
Median	593.3	2.7
Minimum	1	1
Maximum	1,996	15.7
Standard deviation	240	1.6
CV (RSD)	0.39	0.52

The points shown in Figure A-16 represent particulates present at accessible sampling locations. Some empty locations have zero particles and represent inaccessible locations. The mean concentration of the contaminants in the particulates is not the same at all accessible sampling locations, implying that large-scale DH is also present in the DU (i.e., a nonstationary process). Figure A-17 shows the probability distribution of particulates throughout the DU represented graphically on a normal distribution quantile-quantile (Q-Q) plot. The DU is heterogeneous with a CV of 0.86, which is relatively low. This example is a good reminder that a DU can exhibit a range of heterogeneities and yet the distribution need not have a high CV. CV is calculated based on concentration values, and heterogeneities are present in bulk material within the DU. DUs used in these sections represent bulk material DUs with one and only one point at each sampling location. There is no GSE present in such DUs, and all sampling patterns resulted in unbiased estimate of DU (data set) mean.

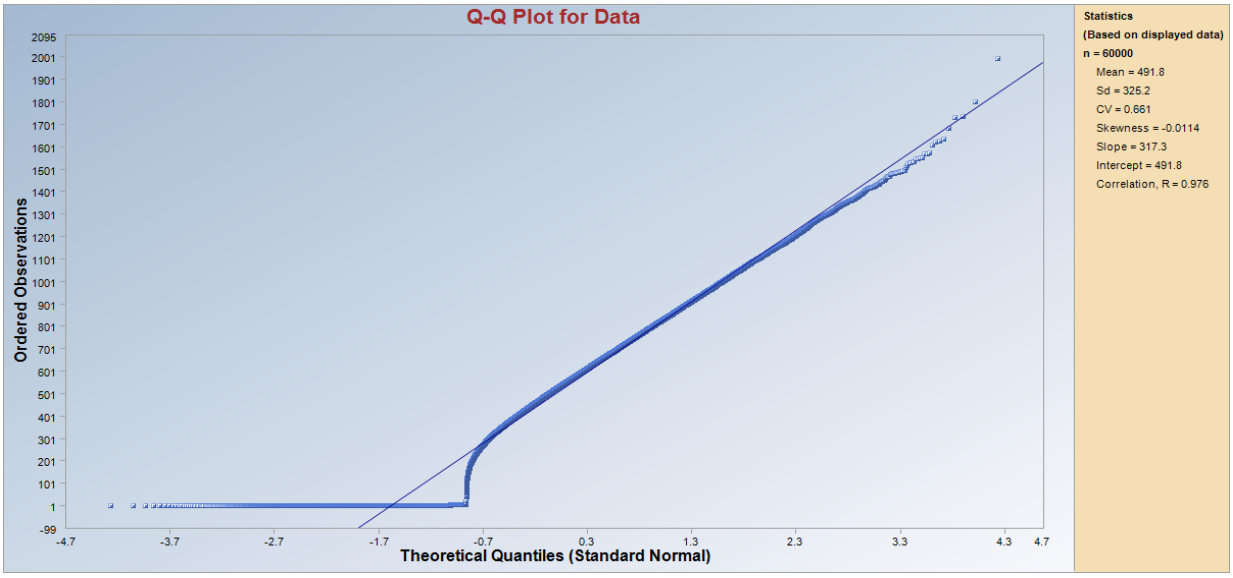


Figure A-17. Scenario M3-A normal distribution Q-Q plot of concentrations of particulates.

The M3-A DU represents a DU with both CH and DH. Being a target, the center of the DU consists of more bulk material particles than other areas of the DU, giving rise to small-scale DH

and resulting in GSE. The GSE (and therefore FE) can be addressed by collecting and combining increments of adequate sample support. Being a heterogeneous DU, the use of an appropriate SS helps in collecting a representative sample that yields unbiased estimate of DU mean. Simulations were conducted using 36 and 64 increments each of SS of 0.05 units collected using systematic random grid sampling pattern and simple random sampling pattern. Tables A-13 and A-14 provide summary statistics relevant to various performance metrics.

Table A-13. Summary statistics of ISM applied to M3-A using 36 increments, each with a sample support of 0.05 units

Statistics	3 Replicates		5 Replicates	
	Systematic (with random start)	Simple random	Systematic random	Simple random
Minimum	337.5	405.5	348.5	423.6
Maximum	457.0	558.4	439.3	559.2
Sample mean	395.7	476.1	395.9	476.5
Bias	–96.1	–15.7	–95.9	–15.4
Relative bias (FE)	0.20	0.032	0.20	0.031
Student's-*t*-UCL95 coverage	14.3%	91%	0.3%	91%
Student's-*t*-UCL95 (average)	450.9	555.7	430.3	524.2
Chebyshev UCL95 coverage	33.7%	95.3%	17.7%	99.3%
Chebyshev UCL95 (average)	478.1	594.9	466.1	574.1
Root mean square error (RMSE)	98.7	31.9	97.4	26.4
Standard deviation of FE	0.200	0.065	0.198	0.054
CV Bar	0.083	0.099	0.091	0.106

Table A-14. Summary statistics of ISM applied to M3-A using 64 increments, each with a sample support of 0.05 units

Statistics	3 Replicates		5 Replicates	
	Systematic random	Simple random	Systematic random	Simple random
Minimum	345.8	403.0	348.2	431.4
Maximum	433.2	552.0	421.7	521.7
Sample mean	390.7	478.1	389.6	476.1
Bias	–101.2	–13.8	–102.2	–15.8
Relative bias (FE)	0.21	0.028	0.21	0.032
Student's-*t*-UCL95 coverage	1.7%	89.7%	0%	82.7%
Student's-*t*-UCL95 (average)	433.4	536.5	414.5	510.7
Chebyshev UCL95 coverage	16.3%	96%	1.7%	97.7%
Chebyshev UCL95 (average)	454.5	565.3	440.4	546.8
RMSE	102.5	27.2	103.1	23.1
Standard deviation of FE	0.208	0.055	0.210	0.047
CV Bar	0.065	0.073	0.067	0.076

Results summarized in Tables A-13 and A-14 provide the following insights:

- ISM collected using simple random sampling (with 3 or 5 replicates) with appropriate SS resulted in an unbiased estimate of DU mean, an observation supported by statistical theory.

- ISM based on a systematic random grid pattern (and also serpentine pattern shown in Figure A-21) yields an estimate of the DU mean with relative bias of 19.5% using 36 increments (3 and 5 replicates) and relative bias of about 21% using 64 increments (3 and 5 replicates); ISM based on an simple random sampling pattern yields DU means with relative bias of only 3% using 36 increments (3 and 5 replicates) and 2.8% to 3.2% using 64 increments (3 and 5 replicates, respectively). The observation that simple random sampling tends to yield unbiased estimates of DU mean is supported by statistical theory.

- SD (FE) based on a systematic random grid pattern is about 0.20 using 36 increments (with 3 and 5 replicates) and 0.21 using 64 increments (with 3 and 5 replicates). SD (FE) based on simple random sampling is 0.06 (36 increments with 3 and 5 replicates) and 0.05 (64 increments with 3 and 5 replicates).

- The coverage is directly related to bias and FE. Since bias and FE in the mean estimate based on a systematic random grid pattern (and also serpentine pattern, not included here) is high (for 36 as well as 64 increments), the associated coverages provided by both UCL methods are poor (i.e., much lower than the nominal 95% coverage).

- The coverage by a UCL tends to improve (come closer to 95%) as FE decreases. For the Student's-*t*-UCL95, coverage decreases marginally (from 91% to 89%) when increments are increased from 36 to 64 (using simple random sampling). For the Chebyshev UCL95, results vary within the margin of stochastic error of the simulation, but all of the coverages exceed the nominal 95% target. It is well known that Chebyshev UCL95 method tends to provide conservative values of UCLs (USEPA 2010b), especially when data are mildly skewed with CV <1.

When selecting a sampling design, it is important to weigh the advantages and disadvantages of various options. These simulations demonstrate how the coverage of the 95% UCL may actually decrease when a greater number of increments (using the same sample support per increment) are collected from a DU with moderate CH and DH. However, while the impact on the statistical concept of coverage is marginal, it is likely offset by the considerable improvement in the spatial coverage afforded by nearly doubling the density of the sampling network within the DU.

Is should be pointed out that for this DU, simulations were performed by collecting increments consisting of a single point (and also using SS of 0.01 units) from each selected sampling location. These simulations resulted in biased estimates of DU mean using all three sampling patterns. These observations reiterate the importance of using adequate SS in collecting incremental samples. Typically appropriate SS is calculated based on the particle size and other properties of the sampling medium (Pitard 1993). Section A.6 provides some details about selecting an appropriate SS size.

The above statements are verified further by using a real raidum-226 DU (where GSE is already addressed just like in M-1 and M-2 simulations) with large-scale DH (Scenario M3-C) and a hypothetical homogenous DU with a constant mean concentration within the DU (Scenario M3-B).

A.5.3 Simulation Results for Scenario M3-B

The DU shown in Figure A-18 represents a homogeneous DU with a mean concentration of about 200 units at each sampling location. The particulates are evenly distributed throughout the DU without any spatial patterns (i.e., a stationary process) with the same mean concentration of 200 mg/kg at each location within the DU. The concentrations of the contaminant present in bulk material follow a skewed gamma distribution with scale parameter of 1000 and shape parameter of 0.2. Figure A-19 shows the normal distribution Q-Q plot.

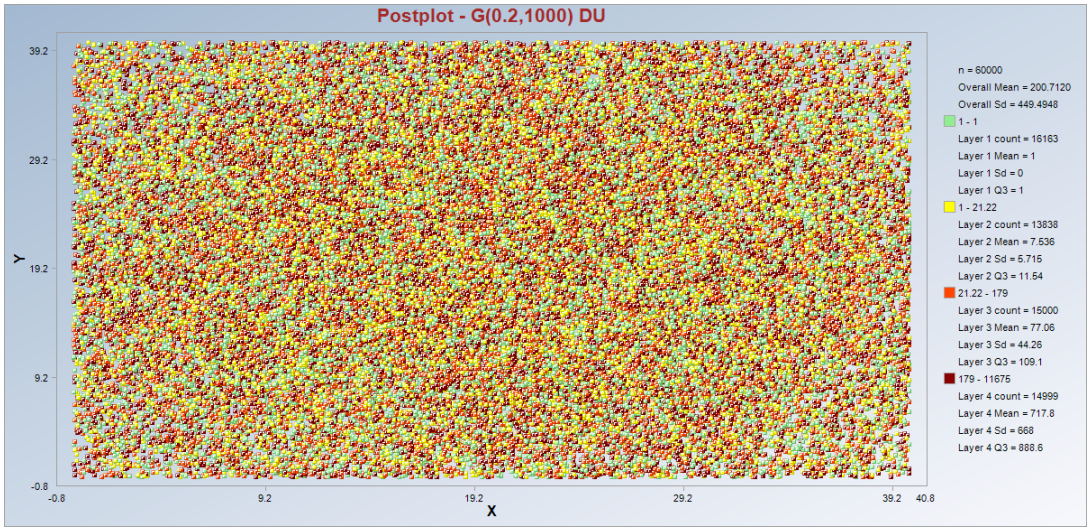


Figure A-18. Post plot for Scenario M3-B representing a homogeneous DU with mean of 200 mg/kg. Concentrations follow a gamma distribution (shape = 0.2, scale =1,000).

This kind of DU can be generated by applying a spatial interpolation method (e.g., kriging, inverse distance weighting [IDW]) to a data set generated from discrete samples. A similar concept was demonstrated in the generation of the map used in Scenario M1, which was generated using IDW. This smoothing process is equivalent to assuming that all particles in the DU are of the same size, shape, and mass and that those particulates are evenly and homogeneously distributed throughout the DU *with one particle per sample location*. Although the particles are homogeneously distributed within the DU, the concentrations of a contaminant can vary from location to location.

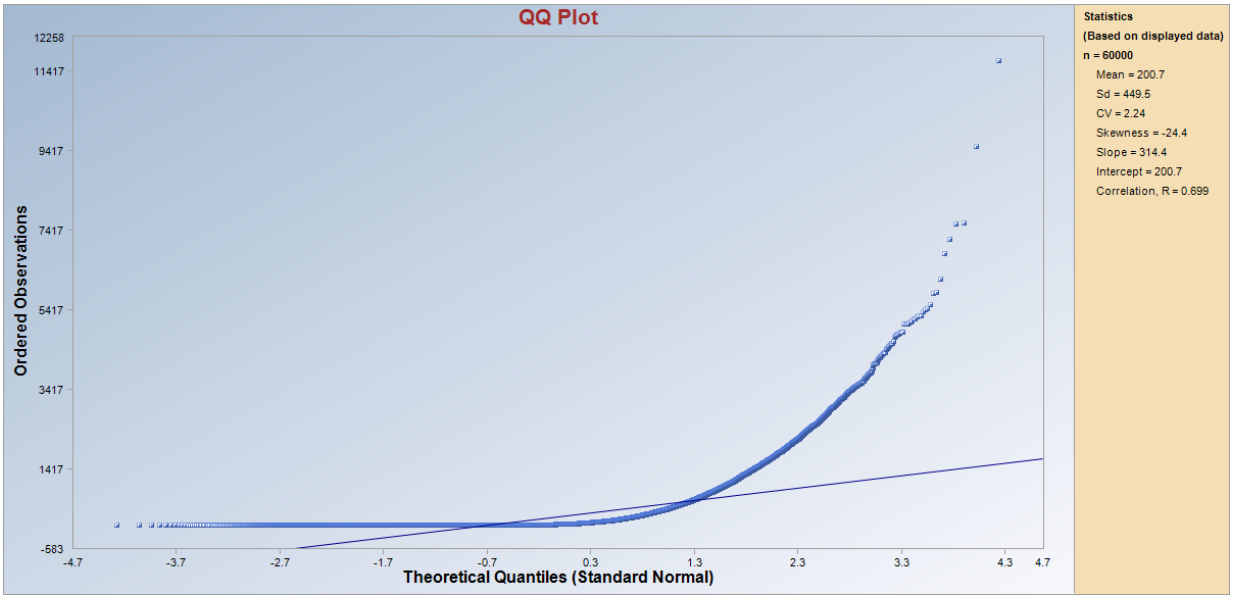


Figure A-19. Scenario M3-B normal distribution Q-Q plot of concentrations.

In the absence of material CH, small-scale DH, and GSE, any random sampling method (e.g., discrete, composite, or ISM) yields a fairly representative, unbiased estimate of the DU mean. The smoothing process does not address large-scale (long-range) DH that may arise due to contamination patterns potentially present in the DU. The presence of large-scale DH implies that the mean concentrations of the contaminant at different locations can be different. As mentioned previously, ISM is not expected to address large-scale (long-range) DH. ISM masks large-scale distributional (e.g., present due to spatial/temporal patterns) heterogeneity present in a DU.

With DUs that are homogeneous with respect to particles as well as with respect to concentration contents of those particles, although any random sampling scheme (discrete or ISM) yields an unbiased estimate of DU mean, this fact does not guarantee that the UCL provides adequate coverage for the mean. If the concentration distribution is highly skewed, the distribution of sample means may also be asymmetric, and the uncertainty in the overall mean of replicate ISM samples can be high (with any sampling method). Specifically, the coverage provided by Student's-*t* 95% UCL could be less than 95%.

Due to homogeneity of particles present in the DU, the size of the SS (mass) does not matter much in reducing the bias in the mean estimate. The simulations demonstrated with M3-B illustrate how increasing the SS (e.g., 0.05 units instead of a single point) introduces a marginal change (<2%) in FE and SD (FE). This observation reiterates the importance of using an appropriate SS for heterogeneous bulk material DUs.

Table A-15 summarizes the population parameters for the M3-B scenario and Tables A-16 and A-17 summarize the simulation results with alternative ISM sampling methods.

Table A-15. Summary statistics (population parameters) for the DU given by scenario M3-B, entire area and individual quadrants

Statistic	DU	Quadrant 1	Quadrant 2	Quadrant 3	Quadrant 4
Percent of total	100%	25.0%	24.9%	25.2%	24.8%
Number of cells	60,000	15,026	14,963	15,115	14,896
Arithmetic mean	200.7	203	197.5	202.5	199.8
Median	21.2	21.2	20.5	21.6	21.9
Minimum	1	1	1	1	1
Maximum	11,675	7,568	6,179	11,675	5,857
Standard deviation	449.5	445.9	438.3	465.6	447.7
CV (RSD)	2.24	2.27	2.22	2.30	2.24

Table A-16. Summary statistics of ISM applied to M3-B using 36 increments, each with no sample support (i.e., single points)

Statistics	3 Replicates			5 Replicates		
	Serpentine	Systematic random	Simple random	Serpentine	Systematic random	Simple random
Minimum	79.7	105.4	101.0	113.9	118.0	121.5
Maximum	325.0	321.3	350.6	339.1	309.8	323.5
Sample mean	202.2	200.9	201.1	203.5	201.7	202.8
Bias	1.5	0.14	0.43	2.8	1.0	2.1
Relative bias (FE)	<0.01	<0.01	<0.01	0.01	<0.01	0.01
Student's-*t*-UCL95 coverage	92.7%	93.7%	91.3%	93.0%	92.7%	92.0%
Student's-*t*-UCL95 (average)	314.8	312.5	310.4	272.9	267.1	267
Chebyshev UCL95 coverage	96%	98.7%	95.3%	99%	99%	99%
Chebyshev UCL95 (average)	370.2	367.6	364.2	345.4	335.3	334.2
RMSE	42.5	42.6	44.1	35.8	33.9	36.2
SD of FE	0.211	0.212	0.220	0.178	0.169	0.180
CV Bar	0.330	0.331	0.321	0.357	0.340	0.333

Table A-17. Summary statistics of ISM applied to M3-B using 64 increments, each with no sample support (i.e., single points)

Statistics	3 Replicates			5 Replicates		
	Serpentine	Systematic random	Simple random	Serpentine	Systematic random	Simple random
Minimum	107.1	108.0	122.9	138.6	136.0	133.9
Maximum	320.2	291.0	305.3	274.1	289.1	275.3
Sample mean	202.7	200.3	205.1	205.2	200.7	202.4
Bias	2.03	-0.42	4.41	4.49	0.003	1.70
Relative bias (FE)	0.01	<0.01	0.02	0.02	<0.01	<0.01
Student's-*t*-UCL95 coverage	94%	93.7%	94%	95.3%	89.3%	93%
Student's-*t*-UCL95 (average)	290.1	283.7	290.9	255.8	251.1	253
Chebyshev UCL95 coverage	98.3%	97%	96%	98.3%	98.3%	99.3%
Chebyshev UCL95 (average)	333.2	324.8	333.2	308.7	303.8	305.8
RMSE	34.4	32.0	33.6	25.0	27.4	26.0
SD of FE	0.171	0.159	0.168	0.125	0.136	0.130
CV Bar	0.26	0.25	0.25	0.26	0.26	0.26

A.5.4 Simulation Results for Scenario M3-C

The DU in Figure A-20 shows spatial patterns in radium-226 based on a data set of 15,356 discrete readings (kindly provided by Dr. Robert Johnson of Argonne National Laboratory). Samples were collected using walk-over gamma detectors with a sample support of approximately 1 m^2 and 6 deep. This process yields a single observation (point) per sampling location; consequently, small-scale DH and GSE are already addressed. Since data are collected using a gamma detector (without collecting physical bulk material samples), it is implicitly assumed that bulk material CH is not present. Therefore, without appreciable CH, this scenario represents a compositionally homogeneous bulk material DU.

The mean background measurement and regulatory action level are 800 and 1800 cpm, respectively. Large-scale DH is present as the average (mean) reading is not the same at all locations in the DU. In addition, there are several pockets of highly elevated radium-226 readings (significantly exceeding the action level), as can be seen in the post plot of the site shown in Figure A-20.

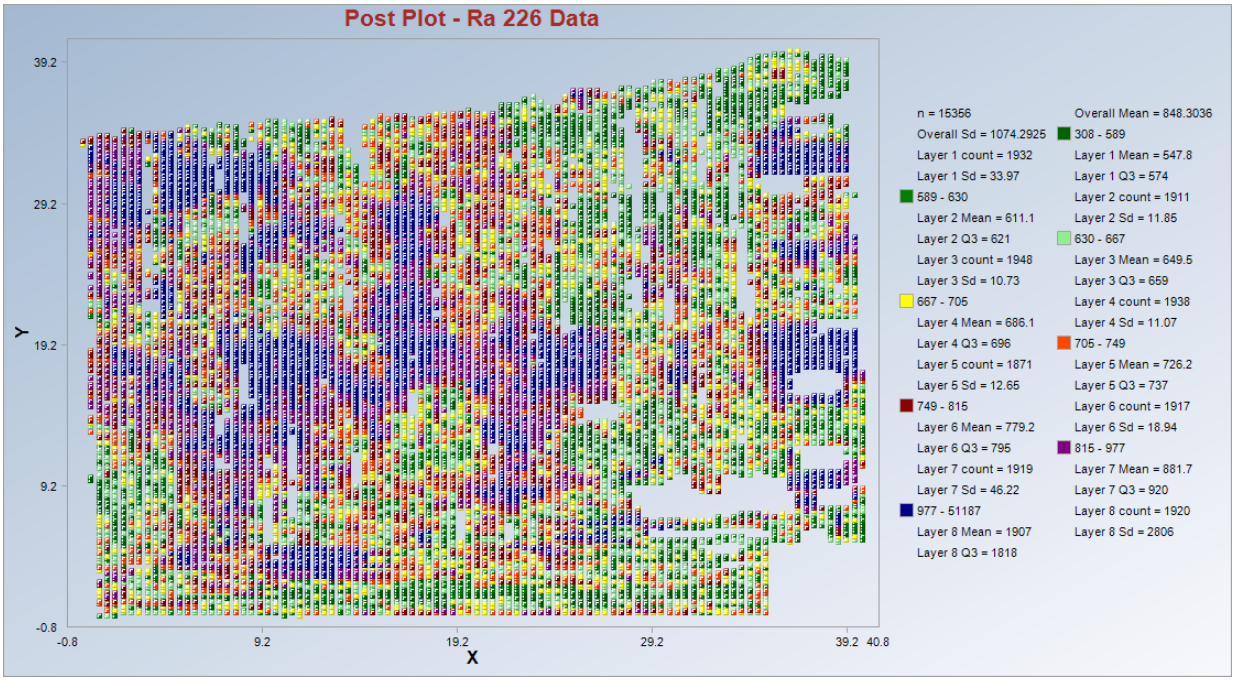


Figure A-20. Post plot of radium-226 readings obtained from a real site (Scenario M3-C).

The (true) DU mean and standard deviation are 848 and 1074 cpm, respectively (CV = 1.3). Since small-scale DH and material CH have already been addressed, this DU scenario with specified (x,y) coordinates is similar to that of the M-1 map. As expected, in the absence of GSE, ISM yields an unbiased estimate of the DU mean using the three sampling patterns; however, simple random sampling yields the least bias in the mean, consistent with statistical theory.

This example demonstrates that the use of ISM correctly accounts for the proportion of the DU with areas of high and low values; however, ISM does not identify the spatial resolution that allows for the identification of hot spots and large-scale DH. If pockets with elevated radium-226 at a scale smaller than the DU are of interest, then the CSM should be revisited to determine whether the DU should be further subdivided. The use of modified ISM described below is better suited to identify large-scale DH. Table A-18 summarizes population parameters for the DU. Tables A-19 to A-22 summarize simulation results for standard ISM, while Tables A-23 and A-24 summarize results for modified ISM.

In practice, it is not realistic to expect/assume that a DU is homogeneous with respect to bulk particulates present in the DU, as the occurrence of particulate CH is inevitable at environmental sites. It is also not practical to assume that heterogeneities (CH, DH—small scale and large scale) are not present in an environmental SU/DU.

Table A-18. Summary statistics (population parameters) for the DU given by scenario M3-C, entire area and individual quadrants

Statistic	DU	Quadrant 1	Quadrant 2	Quadrant 3	Quadrant 4
Percent of total	100%	24.1%	21.8%	28.9%	24.8%
Number of cells	15,356	3,699	3,347	4,442	3,813
Arithmetic mean	848	748	911	898	837
Median	705	663	743	731	689
Minimum	308	432	437	308	393
Maximum	51,187	26,748	20,712	42,171	51,187
Standard deviation	1,074	714.8	892	1,293	1,221
CV (RSD)	1.27	0.96	0.98	1.44	1.46

The DU mean is 848 cpm; however, several patches of elevated radium-226 are present in each of the four quadrants of the DU. The regulatory not-to-exceed threshold is 1800 cpm.

ISM samples based on increments consisting of even single points yield a fairly unbiased estimate of the DU mean, as shown in Tables A-19 (36 increments) and A-21 (64 increments). ISM results based on SS of 0.05 units are shown in Tables A-20 (36 increments) and A-22 (64 increments). The use of increased SS reduces the FE in the mean by less than 1%.

Table A-19. Summary statistics of ISM applied to M3-C using 36 increments, each with no sample support (i.e., single points)

Statistics	3 Replicates			5 Replicates		
	Serpentine	Systematic random	Simple random	Serpentine	Systematic random	Simple random
Minimum	674	710	723	699	730	732
Maximum	1374	1626	1573	1245	1169	1391
Sample mean	833	829	858	841	828	852
Bias	–15.1	–19.7	9.3	–7.1	–20.3	3.7
Relative bias (FE)	0.018	0.023	0.011	<0.01	0.024	<0.01
Student's-*t*-UCL95 coverage	78%	70%	80%	76.5%	66%	74.8%
Student's-*t*-UCL95 (average)	1,053	1,004	1,068	988	935	978
Chebyshev UCL95 coverage	85.7%	81.7%	87.7%	90.5%	86%	91.8%
Chebyshev UCL95 (average)	1161	1090	1171	1141	1047	1110
RMSE	109.1	117.1	117.8	88.7	71.4	85.7
SD of FE	0.129	0.138	0.139	0.105	0.084	0.101
CV Bar (mean of CVs)	0.14	0.11	0.13	0.17	0.13	0.15

Table A-20. Summary statistics of ISM applied to M3-C using 36 increments, each with a sample support of 0.05 units

Statistics	3 Replicates		5 Replicates	
	Systematic random	Simple random	Systematic random	Simple random
Minimum	713	718	733	727
Maximum	1336	1349	1108	1353
Sample mean	844	845	828	846
Bias	–4.8	–3.8	–20.7	–2.7
Relative bias (FE)	<0.01	<0.01	0.02	<0.01
Student's-*t*-UCL95 coverage	73.7%	80%	66%	79.5%
Student's-*t*-UCL95 (average)	1,039	1,040	932	966
Chebyshev UCL95 coverage	81.3%	89.3%	84.8%	94.5%
Chebyshev UCL95 (average)	1135	1136	1041	1092
RMSE	106.2	100.6	70.1	76.4
SD of FE	0.125	0.119	0.082	0.090
CV Bar	0.123	0.125	0.126	0.141

Table A-21. Summary statistics of ISM applied to M3-C using 64 increments, each with no sample support (i.e., single points)

Statistics	3 Replicates			5 Replicates		
	Serpentine	Systematic random	Simple random	Serpentine	Systematic random	Simple random
Minimum	747.5	724.9	749.3	748.9	743.3	757.2
Maximum	1210	1104	1400	1068	1098	1047
Sample mean	836	823	857	835	831	848
Bias	–12.1	–25.7	8.9	–13.5	–17.7	–0.4
Relative bias (FE)	0.014	0.030	0.010	0.0160	0.021	<0.01
Student's-*t*-UCL95 coverage	73%	71.3%	84.3%	69.3%	68.3%	81%
Student's-*t*-UCL95 (average)	986	935	1030	932	920	950
Chebyshev UCL95 coverage	83%	80%	90.7%	88.3%	87.8%	95%
Chebyshev UCL95 (average)	1060	990	1115	1033	1013	1057
RMSE	78.3	63.9	85.6	57.1	54.7	57.0
SD of FE	0.092	0.075	0.101	0.067	0.065	0.067
CV Bar	0.10	0.08	0.11	0.12	0.11	0.12

Table A-22. Summary statistics of ISM applied to M3-C using 64 increments, each with a sample support of 0.05 units

Statistics	3 Replicates		5 Replicates	
	Systematic random	Simple random	Systematic random	Simple random
Minimum	736	737	749	748
Maximum	1126	1415	1035	1126
Sample Mean	830	851	831	851
Bias	–18.6	2.9	–17.3	3.1
Relative Bias (FE)	0.02	<0.01	0.02	<0.01
Student's-*t*-UCL95 coverage	76%	79.7%	68%	81%
Student's-*t*-UCL95 (average)	961	1017	914	955
Chebyshev UCL95 coverage	82.7%	89.3%	90.3%	93.8%
Chebyshev UCL95 (average)	1027	1099	1001	1063
RMSE	70.9	85.5	52.4	61.3
SD of FE	0.084	0.100	0.062	0.072
CV Bar	0.089	0.108	0.101	0.122

Results summarized in Tables A-19 and A-22 provide the following insights:

- For smoothed DUs (without small-scale DH and GSE), all sampling patterns yield fairly unbiased estimates of the DU mean as supported by sampling theory (Cochran 1977). In the present case, FE is <3% for all sampling methods.
- The size of the SS (mass) does not matter as GSE is not present. In other words, increasing SS does not decrease the bias in the mean estimate.
- Being a homogeneous DU without material CH and small-scale DH, the bias does not decrease significantly with increased number of increments and sample support (e.g., Tables A-19 and A-21 for 36 increments; Tables A-20 and A-22 for 64 increments).
- Due to the presence of large-scale DH, the variability is moderate throughout the DU (CV = 1.3), and as a result, the *t*-UCL95 does not provide the specified 95% coverage for the DU mean.
- Chebyshev UCL95 does not provide 95% coverage when only three replicates are used.
- Based on ISM replicate data alone, it is difficult to identify pockets of elevated radium-226.

A.5.5 Simulation Results for Scenario M3-C with Modified ISM (Quadrant Subdivision)

In the modified ISM, a DU is partitioned into several (≥3 to compute statistics) *fairly homogeneous subparts*, and ISM is applied to each of the subparts. Subdivision of a DU into fairly homogeneous subparts should be predicated on the CSM and/or results of a pilot study to extract information about the contamination distribution patterns. For simplicity in this example, equally sized quadrants are used.

In the modified ISM, at least one ISM replicate is collected from each subpart. The modified ISM may require more sampling effort than ISM applied to the entire DU (depending on the number

of increments and replicates). The additional field effort is justified weighed against the potential to increase the information content of ISM and the needs of the site characterization.

This type of subdivision of a DU into subparts is also recommended by Pitard (1993). Unlike the reduced variability obtained from multiincrement sampling (MIS), the modified ISM provides a better estimate of overall DU variability, which can be used to compute a more accurate and rigorous 95% UCL by using direct statistical methods. The use of modified ISM may be particularly helpful under the following conditions: DU with large-scale DH and DU with mean concentration of the contaminant near the regulatory action level.

If the subparts have different areas, an area-weighted procedure can be used to estimate the DU mean and associated 95% UCL (see Section 4). Large DUs may require a greater number of subparts. For large DUs with many subparts, block kriging (Cressie 1993) may be used to characterize the DU contamination distribution.

Simulations using modified ISM were applied to Scenario M3-C DU shown in Figure A-20, with population parameters (including quadrants) summarized in Table A-18. Relevant statistics based on the modified ISM are summarized in Tables A-23 and A-24.

Table A-23. Summary statistics of modified ISM applied to M3-C using 1 ISM per quadrant, each with no sample support (i.e., single points)

Statistics	25 Increments			36 Increments		
	Serpentine	Systematic random	Simple random	Serpentine	Systematic random	Simple random
Minimum	722	719	707	728	740	741
Maximum	1396	1486	1608	1200	1251	1284
DU mean	829	851	858	830	842	857
Mean Q1	744	746	755	756	746	754
Mean Q2	860	900	905	854	893	910
Mean Q3	889	909	910	884	890	920
Mean Q4	825	848	861	826	840	846
Bias	–19.0	2.3	9.4	–18.5	–6.0	9.2
Relative bias (FE)	0.02	<0.01	0.01	0.02	<0.01	0.01
Student's-*t*-UCL95 coverage	74%	81%	87.8%	75.8%	84.8%	88.3%
Student's-*t*-UCL95 (average)	983	1,036	1,068	9,78	1,000	1,043
Chebyshev UCL95 coverage	90%	96.3%	96.3%	97%	96.3%	97%
Chebyshev UCL95 (average)	1114	1194	1247	1104	1134	1200
RMSE	101	115	123	83	84	93
SD FE	0.119	0.136	0.145	0.098	0.099	0.110
CV Bar	0.15	0.17	0.19	0.14	0.15	0.17

Table A-24. Summary statistics of modified ISM applied to M3-C using 1 ISM per quadrant, each sample support of 0.05 units

Statistics	25 Increments		36 Increments	
	Systematic random	Simple random	Systematic random	Simple random
Minimum	735	704	736	740
Maximum	1437	1337	1224	1300
DU mean	845	857	844	850
Mean Q1	755	748	754	749
Mean Q2	891	933	890	918
Mean Q3	910	909	888	892
Mean Q4	825	840	845	841
Bias	–3.0	9.1	–3.9	1.4
Relative bias (FE)	<0.01	0.01	<0.01	<0.01
Student's-*t*-UCL95 coverage	83.3%	85.8	84.8	88%
Student's-*t*-UCL95 (average)	1015	1059	1005	1022
Chebyshev UCL95 coverage	96%	98.25%	96%	98.8
Chebyshev UCL95 (average)	1159	1231	1142	1168
RMSE	98.3	105.6	86.5	88.5
SD of FE	0.116	0.124	0.102	0.104
CV Bar	0.158	0.185	0.151	0.162

A.6 GLOSSARY OF TERMS AND CONCEPTS

A.6.1 Gy Sampling Principles Applied to Bulk Material

The use of Gy's incremental sampling approach to collect samples from environmental bulk materials (e.g., soils, sediments, liquids) is a relatively new concept in contaminated site investigation. Important concepts underlying Gy's sampling principles (and ISM) are reflected by addressing the following questions:

- What are the differences between bulk material sampling and traditional random sampling of discrete items (e.g., individuals in a room, discrete data sets consisting of 100 distinct points, or even 10,000 distinct points)?
- How can Gy's sampling principles be used to collect representative (unbiased) samples from environmental bulk materials?
- Which heterogeneities are sources of sampling error, identified by Gy's sampling theory, are addressed by ISM?
- What is the difference between the heterogeneities present in the bulk material to be sampled and the variability in the analytical results obtained by chemical analysis of collected samples?

The variability observed in measured concentrations within a DU can be directly attributed to heterogeneities in the environmental media that is sampled. This concept applies to individual ISM samples (estimates of the mean) and discrete samples alike. Uncertainty in parameter

estimates and corresponding decision errors are closely tied to the underlying sources of variability and the sampling methodology used to obtain a representative sample (e.g., simple random sampling, serpentine, and systematic random within grid).

Gy's field sampling equation (Pitard 1993, Smith 2006) given below (1) represents the main formula used to compute sample mass, M_s, needed to obtain a representative/unbiased estimate of mean concentration, C_{DU} of the contaminant present in the bulk material of mass, M_L, contained in a DU (lot)

$$s_{FE}^2 = \left(\frac{1}{M_s} - \frac{1}{M_L} \right) clfgd^3 \ , \qquad (1)$$

The details of the parameters used in equation (1) can be found in Pitard (1993). In mining projects, a detailed investigation is conducted a priori to estimate the parameters used in equation (1), so that an adequate amount of mass of the bulk material can be collected to obtain a representative/unbiased estimate of the mean of the lot.

Ideally, the sample mass, M_s, is calculated using Gy's experimental equation (1). Based on a CSM, a pilot study, and/or information from similar sites, an initial estimate, s_{FE}^2, of the variance of the FE, σ_{FE}^2, needs to be estimated or computed beforehand to determine M_s needed to compute a representative estimate of the DU mean concentration, C_{DU}.

Statistical terminology and quantifiable measures (bias, FE, standard deviation of FE, RMSE, 95% UCL) used to assess the performance (accuracy, precision) of the mean estimate, C_{IS}, obtained using ISM are described as follows. Here E represents the expected value operator, *Var* represents the variance operator, and Abs(x) represents absolute value of the quantity, x.

M_L	=	mass of the DU consisting of particulate material (surface soils)
M_s	=	mass of the sampled material collected from the DU
CDU (= μ)	=	mean of the contaminant (e.g., uranium) present in the bulk material of the DU
m	=	number of increments in an ISM replicate; typical values of m = 36, 50, 64, 100
r	=	number of replicates of ISM; r ≥ 1
SS	=	sample support

Sampled mass $M_s = m \times ISM_{incr}$, where ISM_{incr} represents increment mass (same for all increments).

$C_{IS\ (=\bar{x})}$ = an ISM estimate of unknown DU mean, C_{DU}

C_{IS} is computed using r (≥1) replicates from the DU.

Let $\bar{X}_1, \bar{X}_2, ..., \bar{X}_r$ represent the analytical values of the *r* ISM replicates, each replicate made of *m* increments. Note that each analytical result based on an ISM replicate represents an estimate of

the DU mean. The ISM mean estimate, C_{IS}, of the DU mean, C_{DU}, based on r replicates is given in equation (2).

$$C_{IS} = \bar{X} = \frac{\bar{X}_1 + \bar{X}_2 + \ldots + \bar{X}_r}{r} \qquad (2)$$

A.6.2 Bulk Material Sampling

A sampling methodology is considered unbiased/correct if all elements/items present in the SU have exactly the same probability/chance of being selected in the sample. In practice, bulk materials (e.g., soils, sediments, ores, mining waste, liquids) are heterogeneous in nature as they are made of particles/molecules of various density, moisture content, shapes, and sizes with many nondistinct particles present at each location of the DU. Bulk material present in a DU cannot be viewed as a set of distinct objects (e.g., people in a meeting room). It is not possible to select items (molecules, particulates) one by one from a batch (DU) of bulk material in an unbiased manner using simple random sampling. This method introduces GSE (and therefore FE) in the collected bulk material sample. Instead, a group of particles making an increment of specified/practical mass is collected using an appropriate sampling tool (Gerlach and Nocerino 2003). Sampling bias or nonrepresentativeness is introduced due to the distribution (e.g., segregation or grouping) of material particulates contained in the DU. Elder, Thompson, and Myers (1980) showed that even when sampling from bulk materials (consisting of many coexisting and nondistinct particles), the use of simple random sampling yields unbiased estimates of the mean with minimum variance (Smith 2006). The simulation experiments reported in this appendix support Elder, Thompson, and Myers' findings.

A.6.3 Sample Representativeness and Bias

A representative sample of bulk material is one that has the same properties as the bulk material contained in the population. In the case of site investigations, the sample may be obtained through discrete, composite, or ISM sampling and the population is the mass of contaminant in the environmental medium throughout the DU. A representative bulk material sample has all particulates (e.g., size, shape, concentrations) in the *same proportion* as the particulates present in the DU. When a representative sample is collected from the DU, we are more likely to obtain an *unbiased* estimate of the DU mean.

Sampling bias (nonrepresentativeness) is introduced due to composition (e.g., size, shape) and distribution (e.g., segregation, grouping, and spatial patterns) of particulates present in the DU. The distribution and composition of particulates in a sampling DU affect the representativeness of the collected sample. For example, denser particles might have settled at the bottom (e.g., stockpiles), or new contaminants might have settled onto the surface of the DU. Contaminant particulates might be heavily deposited in one part of the DU, resulting in a subarea with elevated concentrations (e.g., spill area or training target location) or clumps with concentrated mass of contaminant (e.g., pesticides in a farm field).

It should be noted that contrary to the common belief, the use of structured sampling patterns such as the serpentine pattern (Figure A-21) and systematic random within grid pattern

(Figure A-22) may not yield an unbiased estimate of DU mean unless the mass/particulates are evenly, uniformly, or homogenously distributed throughout the DU. In the absence of any knowledge about uneven and/or segregated distribution of particles in the DU, the simple random sampling (incremental or discrete) approach is the best approach to collect representative samples from heterogeneous DUs (from the perspective of sample bias). These issues are illustrated by performing simulation experiments on hypothetical heterogeneous target DU considered in this appendix.

A.6.4 Heterogeneities

Heterogeneities in a DU yield sampling errors. In terms of Gy's principles, these errors include FE and GSE, and they apply to all sampling methods (e.g., discrete, composite, and ISM). As a result, measurements of concentrations based on those collected sample(s) may not be representative (i.e., yield an unbiased estimate of the true mean concentration of the contaminant present in the DU). The magnitude of the heterogeneities can be expected to vary among DUs. The terms "homogeneous DU" and "heterogeneous DU" are used in this document to qualitatively describe the relative magnitudes of compositional and distributional heterogeneities (CH and DH).

A.6.4.1 Compositional heterogeneity (CH)

CH is present when particles differ in their size, shape, density, and contaminant loadings (i.e., mass per volume). Different particles consist of different amounts of mass and concentrations. For example, finer particles can be expected to contain greater contaminant masses and concentrations. Presence of CH gives rise to FE (i.e., bias or relative bias in the mean). For a heterogeneous DU with CH, mean estimates are typically biased, and FE (relative bias) cannot be completely eliminated through improvements in sampling design. Presence of CH introduces DH; therefore, as CH increases DH also increases. Gy's sampling theory suggests that FE can be addressed (i.e., reduced but not eliminated) by collecting and combining several increments, each of the same mass (Pitard 1993). This is a fundamental motivation for applying ISM.

A.6.4.2 Distributional heterogeneity (DH)

Uneven, nonrandom distributions (scatter, groupings, deposition) of particles and contamination across a DU yield both small-scale and large-scale (e.g., spatial/temporal) DH. Some areas of the DU may consist of significantly more particles than other areas of the DU (e.g., Figure A-16). There could be isolated area(s) within the DU consisting of more particles (of varying shape, size, orientation) with higher contaminant loadings (e.g., hot spots).

A.6.4.3 Small-scale (short-range) DH

On a smaller scale, DH results due to intrinsic properties of particles such as density, size, shape, and orientation. Both random and nonrandom distribution of particulates can result in small-scale DH. Since it is not possible to collect particles one by one at random when sampling bulk material, small-scale DH results in GSE. GSE also arises when particles preferentially settle (i.e.,

finer particles settle at the bottom and larger particles rise to the top of a column). GSE is one of the main contributors to bias in the mean (Smith 2006).

Small-scale (or short-range) DH is common at most environmental sites. For DUs with small-scale DH (and also large-scale DH), it is likely that some parts of the DU consist of a larger number of particulates than other parts. The small-scale DH may cause large variation in analytical results (both for discrete samples and ISM). These observations have resulted in frequent use of lognormal distribution to model data sets originating from environmental studies. With ISM, small-scale DH and GSE are addressed at the sampling locations by collecting increments of appropriate (and practical) *sample support*.

A.6.4.4 Sample support

SS can be thought of as a portion of soil mass collected from a sampling location of the DU that yields a single incremental sample (pooled across locations). The concept of SS was introduced to address small-scale DH and to minimize GSE. Soil samples with appropriate and practical SS obtained using proper sampling tools (e.g., Gerlach and Nocerino 2003) tend to yield more accurate analytical results. Examples illustrating the influence of the size of SS on the unbiasedness of the estimate of DU mean obtained using an ISM sample were explored, but conclusions are beyond the scope of this document.

The size of the SS and sampling tools used play an important role in collecting representative samples. For example, if one uses large SS exceeding the extent of small-scale DH to collect increments, then the resulting incremental samples may not be representative of the bulk material contained in the DU. Consequently, the analytical result for the composited sample may not represent an unbiased estimate of the mean concentration.

The appropriate size of the SS can be determined by evaluating the particle size distribution of the particles present in the environmental DU under investigation. Further details regarding these concepts can be found in the environmental sampling literature (e.g., Pitard 1993).

A.6.4.5 Fundamental error (FE)

FE is a random quantity that represents the relative difference (relative bias) between the true population parameter (e.g., mean) and the estimate of the mean obtained from sampling of bulk material. The magnitude of the FE (and its variance) is a function of CH and the mass of the collected sample (M_s).

Due to the varying composition (and distribution) of bulk material particles contained in the DU, each individual increment with the same sample support yields a different estimate of the mean. FE is addressed (i.e., reduced but not eliminated) by obtaining adequate sample mass (e.g., by collecting and combining many increments). Sampling pattern (e.g., simple random sampling) used to collect increments also plays an important role in addressing FE and reducing bias in the mean estimate; this is especially true when a DU is moderately to highly heterogeneous.

A.6.4.6 Large-scale (long-range) DH

Large-scale DH occurs due to nonrandom spatial and/or temporal distribution of bulk material particulates in the DU. The DU usage, natural, and temporal factors (e.g., floods, drought, weather patterns, matrix interactions, and multiple uses over time) contribute to large-scale DH. For DUs with large-scale DH, the mean concentration is not the same at all locations of the DU. In statistical terms, this variation leads to a DU with multiple populations, or ISs that are not "identically distributed." Large variability in analytical results among replicate ISM samples does not necessarily mean that the DU is heterogeneous. For example, DUs with lognormally distributed analytical results considered in other sections of this appendix represent homogeneous DUs with large variabilities (CV = 4); the DU considered in Scenario M3-D of this appendix also represents a homogeneous DU with large variability.

Large-scale DH cannot be addressed (or identified) by ISM. When large-scale DH is present, the objective should be delineation. To some extent, this can be achieved using a modified version of ISM in which the DU is subdivided into different SUs (see Section 4).

A.6.5 Homogeneity

There seems to be some confusion about the homogeneity of a DU. In practice, it is not likely to encounter homogeneous environmental DUs; however, for clarification sake, this concept is discussed in some detail here. To illustrate the differences in the various DUs considered in this appendix (including other sections of the appendix), two types of homogeneities are described: a DU can be homogeneous with respect to (a) the bulk material particulates and (b) contaminant concentration present in the bulk material.

A.6.5.1 Homogeneous bulk material

Homogeneous bulk material consists of particulates of same size, shape, and general appearance; however, the concentration of the contaminant present in those particles can differ significantly.

A.6.5.2 Homogeneous contaminant concentration

In a homogeneous bulk material DU, all particulates are assumed to be roughly of same size and shape (e.g., silt and clay DU, soils at a golf course, spill area), and those particulates are uniformly and evenly distributed throughout the DU. This assumption implies that CH and small-scale DH are negligible. However, large-scale DH can be present in such bulk material homogenous DUs; that is, the mean concentration of the contaminant can be different at various sampling locations within the DU.

It should be noted that, in practice, the occurrence of such bulk material homogeneous DU is highly unlikely. This concept is discussed here to illustrate the differences between smoothed DUs (with one point at each location) obtained via spatial interpolations on gridded data and simulated heterogeneous bulk material DUs potentially consisting of multiple nondistinct particles (points) at various sampling locations. Occurrence of multiple points (particulates) at sampling locations gives rise to GSE (and also FE), whereas in smoothed DUs (e.g., Scenario

M3-B), GSE (and, therefore, FE) is already addressed, as one and only one distinct particle is present at each location of the DU.

The following statements can be made about homogeneous bulk material DUs:

- Since all particulates are of the same size and shape and are evenly distributed within the DU, material CH and GSE are minimal; therefore, FE (relative bias) is also minimal. As a result, ISM on such DUs yields a fairly accurate (unbiased) estimate of the arithmetic mean (AM) of the contaminant present in the DU. Several examples in support of these statements are considered in other sections (dealing with lognormal distribution) of this appendix.

- Depending on the contamination variability (e.g., low, high) and patterns (e.g., spatial, temporal, plumes) within the DU (e.g., spill area DU), the mean contaminant loadings of particulates at various sampling locations can be significantly different, giving rise to large-scale DH. Typically, DUs with large-scale DH have higher variability and skewness. For such DUs with higher variability, ISM-based *t*-95% UCL (Student's-*t*-statistic-based 95% UCL) does not provide the specified 95% coverage to DU mean (e.g., Singh, Singh, and Engelhardt 1997). For such DUs, a 95% UCL based on Chebyshev inequality may be used in risk assessment applications (Singh, Singh, and Iaci 2002, USEPA 2010b) to address uncertainties associated with ISM replicates. Scenario M3-D considered later in this section illustrates the issues described in this paragraph in the context of ISM.

- For bulk material homogeneous DUs, the ISM yields an unbiased estimate of DU mean provided increments are collected using a simple random sampling pattern; however, spatial/ temporal patterns (including hot spots) potentially present in the DU could not be identified.

A.6.6 Sampling Patterns

Schematics of the sampling patterns considered in the simulation experiments are given as follows. Figure A-21 shows the serpentine pattern, Figure A-22 shows the systematic random within grid pattern with 16 increments (one from each grid), and Figure A-23 shows the simple random sampling pattern with 16 random increments.

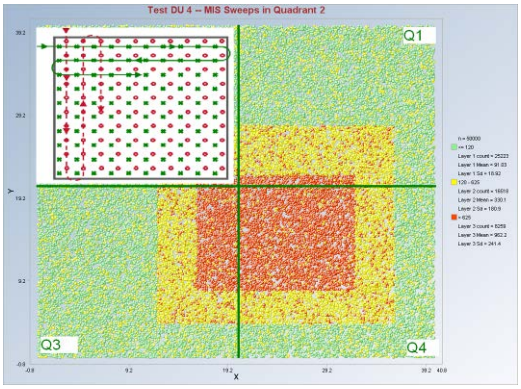


Figure A-21. Serpentine pattern applied to one quadrant of a DU.

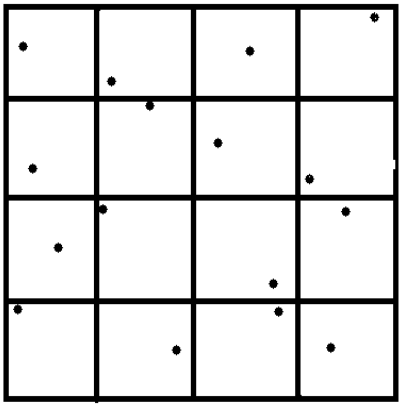


Figure A-22. Systematic random grid pattern with a random start.

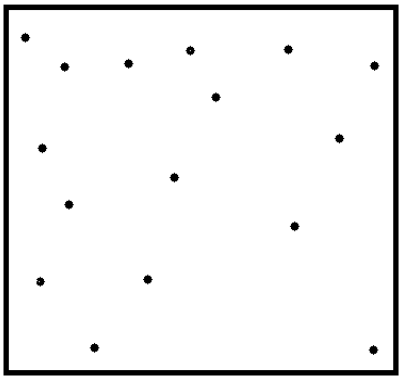


Figure A-23. Simple random sampling pattern.

The simple random sampling pattern yields an estimate that is consistent. As a statistical term, consistency implies that performance measures computed using the simple random sampling (e.g., MSE, RMSE, and SD [FE]) decrease as the sample size (here the number of replicates) increases. The coverage provided by a *t*-95% UCL based on simple random sampling decreases as the sample size increases, and the coverage provided by a Chebyshev 95% UCL increases as the sample size increases. However, based on simple random sampling, mean and standard deviation being consistent estimates, both *t*-UCL and Chebyshev UCL decrease as the sample size (number of replicates for ISM) increases.

Some known properties of sampling patterns and UCLs are noted below:

- For samples (e.g., replicates based on increments collected using simple random sampling) collected using a simple random sampling pattern, the properties of Student's-*t*-statistic-based 95% UCL (Student's-*t*-95% UCL) and Chebyshev inequality–based 95% UCL (Chebyshev 95% UCL) are well established. Specifically, the coverage provided by *t*-95% UCL based on an simple random sampling is nonincreasing as the sample size (e.g., replicates) increases, and the coverage provided by Chebyshev 95% UCL is nondecreasing as the sample size (replicates) increases (e.g., USEPA 2010c; Singh, Singh, and Engelhardt 1997; Singh, Singh, and Iaci 2002, Dudewicz and Mishra 1988).

- For normally distributed data sets, *t*-95% UCL based on a simple random sampling provides approximately 95% coverage for the DU, and Chebyshev UCL 95 tends to provide higher coverage for the DU mean than the nominal 95%.

- For moderately skewed to highly skewed data, *t*-95% UCL fails to provide 95% coverage for the DU mean. For such data sets, the use of Chebyshev 95% UCL is preferred to address uncertainties associated with the estimate of DU mean (Singh, Singh, and Engelhardt 1997; Singh, Singh, and Iaci 2002; USEPA 2010b).

- For serpentine and systematic random sampling patterns, the properties of *t*-95% UCL and Chebyshev 95% UCL are not well established. However, it is noted that for heterogeneous DUs, in addition to yielding biased estimates of the DU mean, the use of serpentine and systematic random within grid sampling patterns tends to yield ISM replicates with lower variability (e.g., Singh, Singh, and Murphy 2009). Therefore, the coverage provided by a 95% UCL (e.g., *t*-95% UCL and Chebyshev 95% UCL) of mean based on ISM increments collected using the serpentine and systematic random within grid sampling patterns is lower than the nominal 0.95 coverage.

Note: If a *t*-95% UCL based on r replicates (increments collected using simple random sampling) does not provide the nominal 95% coverage to DU mean, then a *t*-95% UCL based on a higher (>r) number of replicates also does not provide the 95% coverage to DU mean.

Documents dealing with MIS methodology (e.g., Hewitt et al. 2007, 2009; USACE 2009) suggest the use of serpentine pattern (Figure A-21) or a systematic random grid pattern (Figure A-22) to collect increments making an ISM sample. The use of these patterns is suggested since these sampling patterns are easier to implement in the field; however, statistical sampling theory suggests that data based on simple random sampling (Figure A-23) yield unbiased (representative) estimates. Therefore, increments (of equal mass) should be collected in a completely random/ unbiased manner from the DU consisting of the bulk material. The use of simple random sampling gives each location of the DU an equal chance of being selected in the sample (discrete or ISM) used to estimate the DU mean. These three sampling patterns were evaluated by Singh, Singh, and Murphy (2009). Based on simulation experiments, as expected, they observed that relative bias (FE) in an estimate of DU mean is the least when increments are collected in a random manner, an observation also supported by statistical sampling theory (Elder, Thompson, and Myers 1980).

A.6.7 MIS Software to Generate Homogeneous and Heterogeneous DUs

The homogeneous and heterogeneous DUs in the M3 series of maps considered in this appendix are generated using the DU generation program, MIS (Singh, Maichle, and Armbya 2009). The detailed description of the DU simulation procedure used to generate homogeneous and heterogeneous DUs can be found in Singh, Singh, and Murphy (2009).

The MIS software package is used to collect ISM samples using the three sampling patterns: simple random sampling, serpentine (Hewitt et al. 2007, 2009) sampling, and systematic random within grid. A schematic of the sampling patterns used is given in Figures A-21 to A-23. For systematic random sampling pattern, increments are collected randomly from each grid (one increment per grid).

In a bulk material DU, not all sample locations are accessible due to the presence of bushes, trees, rocks, boulders, trash, and building structures. These locations are represented by empty spots in a simulated DU. While sampling, if one comes across an inaccessible spot (empty space), an increment of the same sample support is collected from a neighboring accessible spot (as commonly done in field).

The ISM sampling process incorporated in the MIS software mimics the sampling process used by CRREL scientists (e.g., Hewitt et al. 2007, 2009). Specifically, all points (material) collected from all increments making an ISM sample are put into a bag. The average of that bag (points in the ISM) represents the analytical value of material (particles) contained in that ISM replicate. The MIS program collects r (≥1, specified by the user) ISM replicates based up 16, 25, 36, 64, and 100 increments. The MIS program can collect ISM samples with increments based up the user-specified sample support (e.g., of radius 0.05 units). For a sample support of 0.05 units, all points (particles) included in a circle of radius 0.05 units are included in each increment making the ISM sample.

For heterogeneous bulk material DUs, not all sampling locations have exactly the same mass or the same number of particulates. Similarly, in simulated hypothetical DUs modeling heterogeneities (CH and small-scale DH), not all increments of the same sample support (radius 0.05 units) consist of the same number of particles/points.

A.7 EQUATIONS FOR PERFORMANCE METRICS

A.7.1 Estimation of FE and Var (FE)

Let C_{ISk}, $k = 1,2,\ldots niter$ represent the ISM mean for the k^{th} iteration (replicate) based on m increments. An estimate of the DU mean, C_{DU}, based on *niter* iterations is given by

$$\hat{C}_{DU} = \overline{C}_{IS} = \sum_{k=1}^{niter} C_{ISk} \,/\, niter$$

Bias can then be expressed as

$$Bias(\hat{C}_{DU}) = Bias = E(\hat{C}_{DU} - C_{DU}) = \sum_{k=1}^{niter} (C_{ISk} - C_{DU}) / (niter)$$

FE in the estimate, $\hat{C}_{DU}$, also referred to a relative bias, is given by

$$FE = |Bias| \,/\, C_{DU}$$

Measures of variance in ISM replicate means, *Var(C_{IS})*, and mean squared error (MSE) are given by

$$Var\ (C_{IS}) = \sum_{k=1}^{niter} (C_{ISk} - \bar{C}_{IS})^2 / (niter - 1)$$

$$\text{MSE} = \sum_{k=1}^{niter} (C_{ISk} - \bar{C}_{IS})^2 / (niter - 1) + Bias^2$$

Finally, an estimate of variance in FE, *Var(FE)*, is given by

$$S_{FE}^2 = \text{MSE} / C_{DU}^2$$

A.7.2 Computation of Coefficient of Variation (CV) of ISM Replicates

The use of CV based on r replicates (e.g., r = 1, 3, 5,...,10) had been proposed as a performance measure associated with the ISM mean, C_{IS}. The MIS software also computes CV; however, CV results are not presented in this section of the appendix, as they are discussed in other sections of this simulation appendix. A brief description of the computation of CV is described as follows.

For the k^{th} iteration, compute CV based on *r* ISM replicates given as follows:

$$CV_k = C_{ISk} / sd_{ISk;} k = 1, 2, ...niter$$

Here, C_{ISk} and sd_{ISk} are ISM mean and ISM standard deviation for the k^{th} iteration and are computed using equations noted in A.7.1.

An estimate of the CV associated with ISM replicates is given by

$$\bar{CV} = CV\_Bar = \sum_{k=1}^{niter} CV_k / niter$$

An estimate of CV associated with the DU, *DU_CV*, can also be computed as follows:

Estimate of DU_CV = est(DU_CV) = sqrt(m) × CV_Bar. where *m* represents the number of increments

$$\text{Bias in DU\_CV} = (\text{DU\_CV} - \text{sqrt(m)} \times \text{CV\_Bar})$$

A.7.3 Performance Metric Equations

Statistical formulae used to compute performance measures are given below:

$$\text{Bias } (C_{IS}) = \text{Bias in estimate of } C_{DU} = E\ (C_{IS} - C_{DU})$$

The estimate C_{IS} of C_{DU} is unbiased when Bias (C_{IS}) = 0

Relative bias (C_{IS}) = Abs (E ($C_{IS} - C_{DU}$)) / C_{DU} = |E ($C_{IS} - C_{DU}$)| / C_{DU}

FE in mean estimate = FE (C_{IS}) = Relative Bias (C_{IS}) = |E ($C_{IS} - C_{DU}$)| / C_{DU}

From the above, it is noted that reduction in FE and its variance are directly related to reduction in the bias in the estimate of DU mean.

MSE and RMSE associated with the ISM mean estimate C_{IS} are given by

$$MSE(C_{IS}) = E(C_{IS} - C_{DU})^2 = Var(C_{IS}) + (Bias(C_{IS}))^2;$$

$$RMSE\ (C_{IS}) = \sqrt{MSE(C_{IS})};$$

$$S_{FE}^2 = Var\ (FE) = \text{MSE} / C_{DU}^2 \text{; and}$$

Standard Deviation (FE) = Sd (FE) = sqrt (MSE) / C_{DU}

The *(1 – α) 100%* Chebyshev UCL of the mean based on *r* ISM replicates is given by

$$UCL = \bar{X} + \left(\sqrt{1 - \frac{1}{\alpha}}\right) * \frac{S_{\bar{X}}}{\sqrt{r}}, \text{where } S_{\bar{X}} = sd\left(\bar{X}_1, \bar{X}_2, ..., \bar{X}_r\right); \text{ and}$$

Student's-*t*-statistic based *(1 – α) 100%* UCL of mean is given by

$$UCL = \bar{X} + t_{(1-\alpha)*(r-1)} * \frac{S_{\bar{X}}}{\sqrt{r}}$$

A.8 REFERENCES

ADEC (Alaska Department of Environmental Conservation). 2009. *Draft Guidance on MULTI INCREMENT Soil Sampling*. Division of Spill Preventions and Response, Contaminated Sites Program. www.itrcweb.org/ism-1/references/multi_increment.pdf.

Cochran, W. G. 1977. *Sampling Techniques*. New York: Wiley.

Cressie, N. 1993. *Statistics for Spatial Data*. New York: Wiley.

Dudewicz, E. J., and S. N. Mishra. 1988. *Modern Mathematical Statistics*. New York:Wiley.

Elder, R. S., W. O. Thompson, and R. H. Myers. 1980. "Properties of Composite Sampling Procedures," *Technometrics* **22**: 179–86.

Gerlach, R. W., and J. M. Nocerino. 2003. *Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples*. EPA/600/R-03/027. www.itrcweb.org/ism-1/references/guidancerl.pdf.

Gy, P. M. 1992. *Sampling for Heterogeneous and Dynamic Material Systems: Theories of Heterogeneity, Sampling, and Homogenizing.* Amsterdam: Elsevier.

Hathaway, J. E., and B. A. Pulsipher. 2010. "Evaluating the Performance of Incremental Sampling." PNNL-19683. Richland, Wash.: Pacific Northwest National Laboratory. www.itrcweb.org/ism-1/references/Final_ISsimulationStudy_082011.pdf.

Hewitt, A. D., T. F. Jenkins, C. A. Ramsey, K. L. Bjella, T. A. Ranney, and N. M. Perron. 2005. *Estimating Energetic Residue Loading on Military Artillery Ranges: Large Decision Units*. ERDC/CRREL TR-05-7. Hanover, N.H.: U.S. Army Engineer Research and Development Center.

Hewitt, A. D., T. F. Jenkins, M. E. Walsh, M. R. Walsh, S. R. Bigl, and C. A. Ramsey. 2007. *Protocols for Collection of Surface Soil Samples at Military Training and Testing Ranges for the Characterization of Munitions Constituents*. ERDC/CRREL TR-07-10. Hanover, N.H.: U.S. Army Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory. www.itrcweb.org/ism-1/references/GetTRDoc.pdf.

Hewitt, A. D., T. F. Jenkins, M. E. Walsh, M. R. Walsh, and S. Brochu. 2009. *ESTCP Cost and Performance Report: Validation of Sampling Protocol and the Promulgation of Method Modification for the Characterization of Energetic Residues on Military Testing and Training Ranges*. ESTCP/CRREL ER-0628. Hanover, N.H.: U.S. Army Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory in partnership with Defense Research and Development Canada-Valcartier, Quebec, Canada. www.serdp.org/Program-Areas/Environmental-Restoration/Contaminants-on-Ranges/Identifying-and-Evaluating-Sources/ER-200628.

Jenkins, T. F., A. D. Hewitt, T. A. Ranney, C. A. Ramsey, D. Lambert, K. L. Bjella, and N. Perron. 2004. *Sampling Strategies Near a Low-Order Detonation and a Target at an Artillery Impact Area*. ERDC/CRREL TR-04-14. Hanover, N.H.: U.S. Army Engineer Research and Development Center.

Jenkins, T. F., A. D. Hewitt, C. L. Grant, S. Thiboutot, G. Ampleman, M. E. Walsh, T. A. Ranney, C. A. Ramsey, A. J. Palazzo, and J. C. Pennington. 2006. "Identity and Distribution of Residues of Energetic Compounds at Military Live-Fire Training Ranges," *Chemosphere* **63**: 1280–90.

Matzke B. D., L. L. Nuffer, R. O. Gilbert, J. E. Hathaway, L. H. Sego, B. Roberts, J. E. Wilson, S. T. Dowson, N. L. Hassig, C. J. Murray, B. A. Pulsipher, and S. McKenna. 2007. *Visual Sample Plan Version 5.0 User's Guide*. PNNL-16939. Richland, Wash.: Pacific Northwest National Laboratory. www.itrcweb.org/ism-1/references/pnnl16939.pdf.

Pitard, F. 1993. *Pierre Gy's Sampling Theory and Sampling Practice: Heterogeneity, Sampling Correctness, and Statistical Process Control*, 2nd ed. Boca Raton, Fla.: CRC Press.

Singh, A., R. Maichle, and N. Armbya. 2009. *A Program to Generate Homogeneous and Heterogeneous DUs and to Perform MULTI INCREMNT SAMPLING Incorporated in EPA Software Scout 1.1*. Las Vegas: Lockheed Martin.

Singh, A., R. Maichle, A. K. Singh, S. Lee, and N. Armbya. 2007. *ProUCL Version 4.00. 02 User Guide*. EPA/600/R-07/038. www.itrcweb.org/ism-1/references/ProUCL_v4.00.02_User_Guide.pdf.

Singh, A., Singh, A. K., and R. J. Iaci. 2002. *Estimation of the Exposure Point Concentration Term Using a Gamma Distribution*. Technology Support Center Issue Paper. EPA/600/R-02/084. www.itrcweb.org/ism-1/references/289cmb02.pdf.

Singh, A., A. K. Singh, and D. Murphy. 2009. *Final Report: Statistical Evaluation of Multi-Increment Sampling (MIS) or Increment Sampling (IS) Approach and Its Applicability in Addressing Various Project Objectives of Environmental Investigations.* Former Nansemond Ordnance Depot, Va.

Singh, A. K., A. Singh, and M. Engelhardt. 1997. *The Lognormal Distribution in Environmental Applications*. Technology Support Center Issue Paper. EPA/600/R-97/006. www.itrcweb.org/ism-1/references/Singh1997.pdf.

Smith, P. L. 2006. *A Primer for Sampling Solids, Liquids, and Gases.* Philadelphia: American Statistical Association and Society for Industrial Mathematics.

USACE. 2009. *Implementation of Incremental Sampling (IS) of Soil for the Military Munitions Response Program.* Interim Guidance Document (IGD) 09-02. Environmental and Munitions Center of Expertise. www.itrcweb.org/ism-1/references/IGD%209-02v2.pdf.

USEPA (U.S. Environmental Protection Agency). 2008. *Quick Overview Scout 2008*. www.epa.gov/esd/databases/scout/abstract.htm.

USEPA 2009. *Scout 2008 Version 1.0 User Guide.* EPA/600/R-08/038.

USEPA. 2010a. *ProUCL 4.00.05 Software for Calculating Upper Confidence Limits (UCLs).* Office of Research and Development. www.epa.gov/esd/tsc/software.htm.

USEPA. 2010b. *ProUCL 4.00.05 Technical Guide* (Draft). EPA/600/R-07/041. Office of Research and Development. www.epa.gov/esd/tsc/TSC_form.htm.

USEPA. n.d. "Scout 1.1. A Robust Statistical Package." Submitted to the National Exposure Research Laboratory, Las Vegas.

USEPA. n.d. "Scout 2008, Version 1.00.01." www.epa.gov/esd/databases/scout/abstract.htm#Scout2008v101.

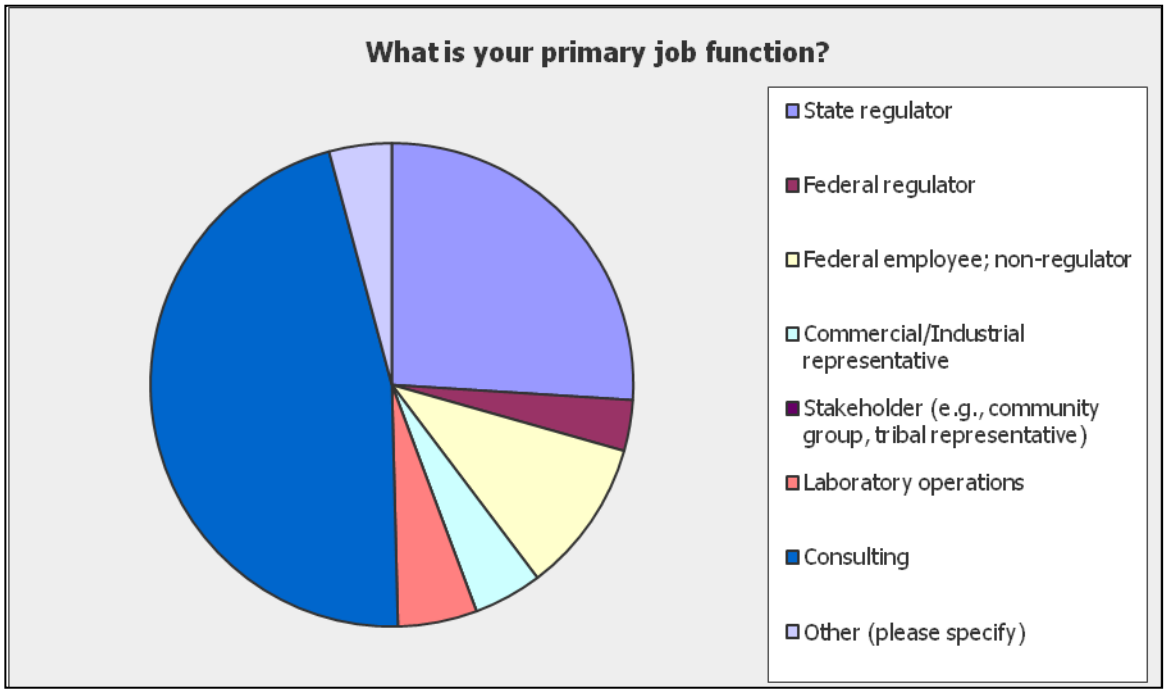
Appendix B

August 2009 Survey Results

AUGUST 2009 SURVEY RESULTS

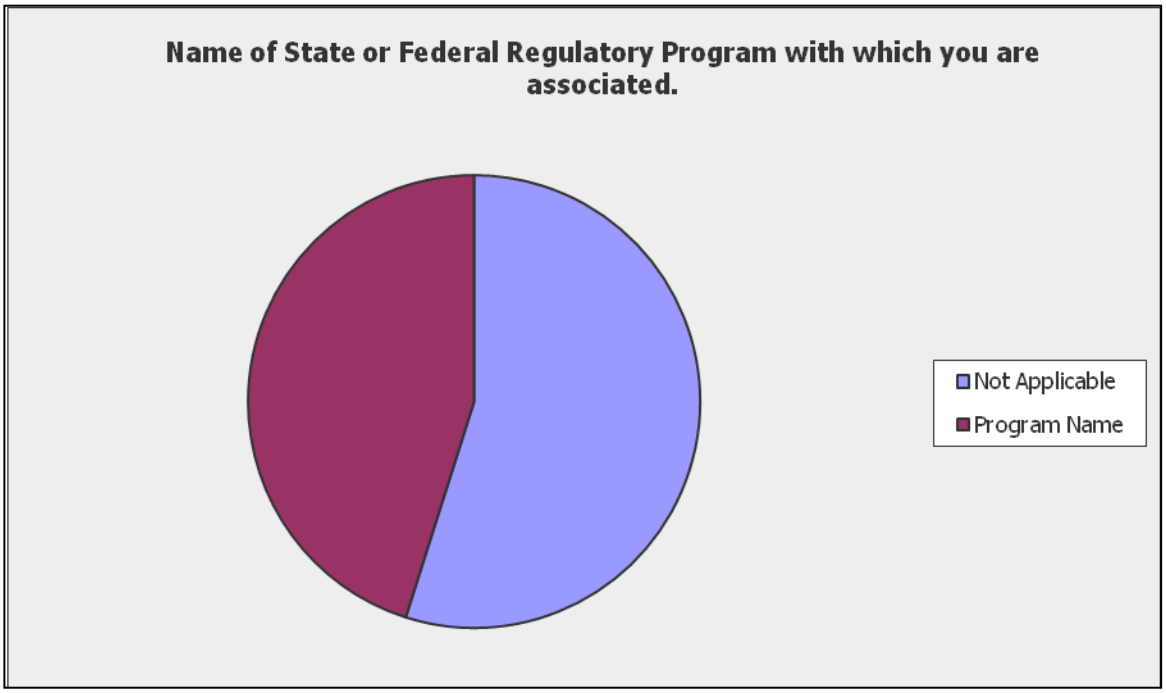
Question 1.

What is your primary job function?		
Answer Options	**Response Percent**	**Response Count**
State regulator	26.0%	68
Federal regulator	3.4%	9
Federal employee; non-regulator	10.3%	27
Commercial/Industrial representative	4.6%	12
Stakeholder (e.g., community group, tribal representative)	0.0%	0
Laboratory operations	5.3%	14
Consulting	46.2%	121
Other (please specify)	4.2%	11
answered question		**262**
skipped question		**0**



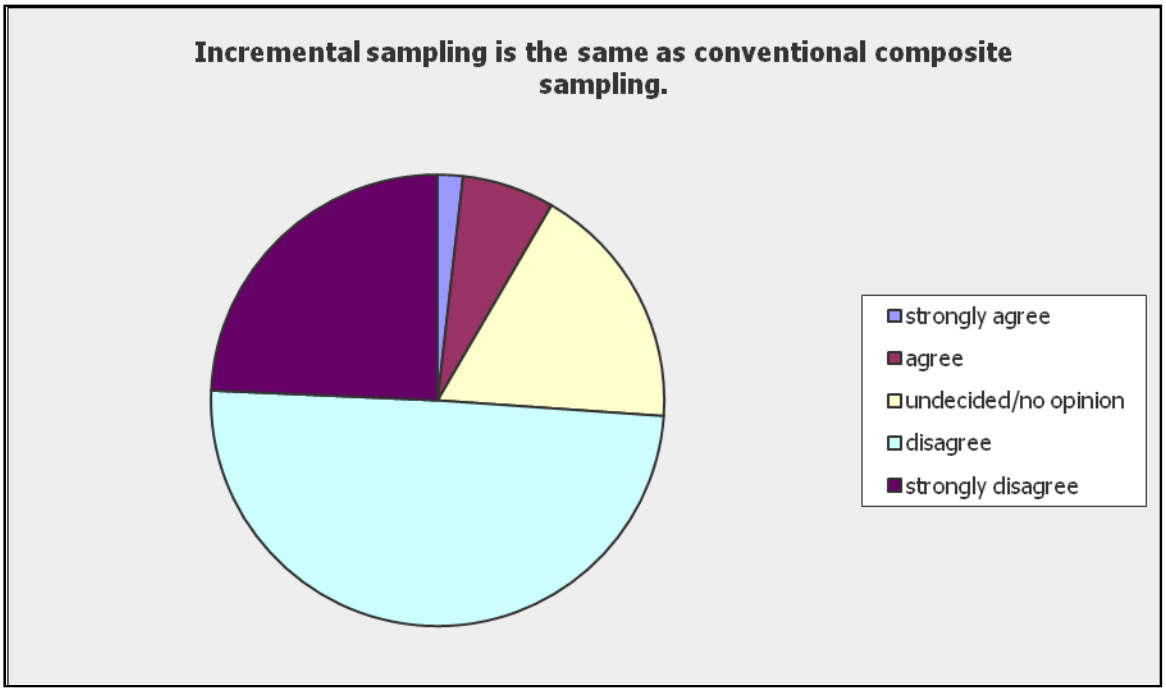
Question 2.

Name of State or Federal Regulatory Program with which you are associated.		
Answer Options	**Response Percent**	**Response Count**
Not Applicable	54.9%	140
Program Name	45.1%	115
	answered question	**255**
	skipped question	**7**



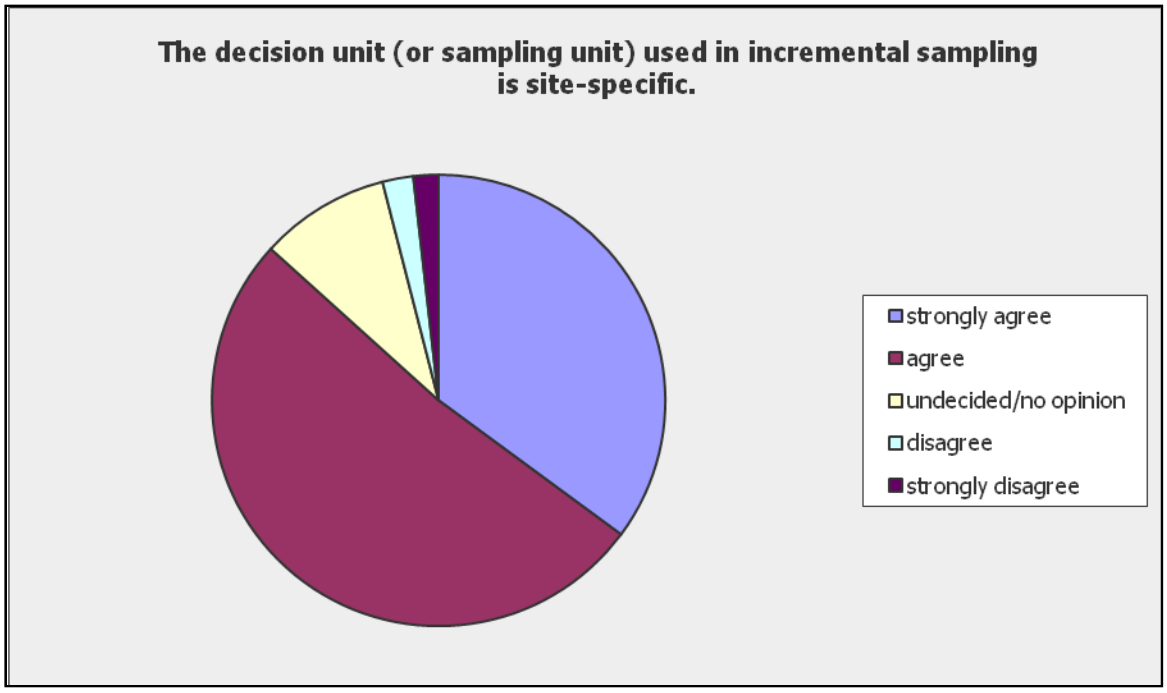
Question 3.

Incremental sampling is the same as conventional composite sampling.		
Answer Options	**Response Percent**	**Response Count**
strongly agree	1.8%	4
agree	6.6%	15
undecided/no opinion	17.7%	40
disagree	49.6%	112
strongly disagree	24.3%	55
answered question		**226**
skipped question		**36**



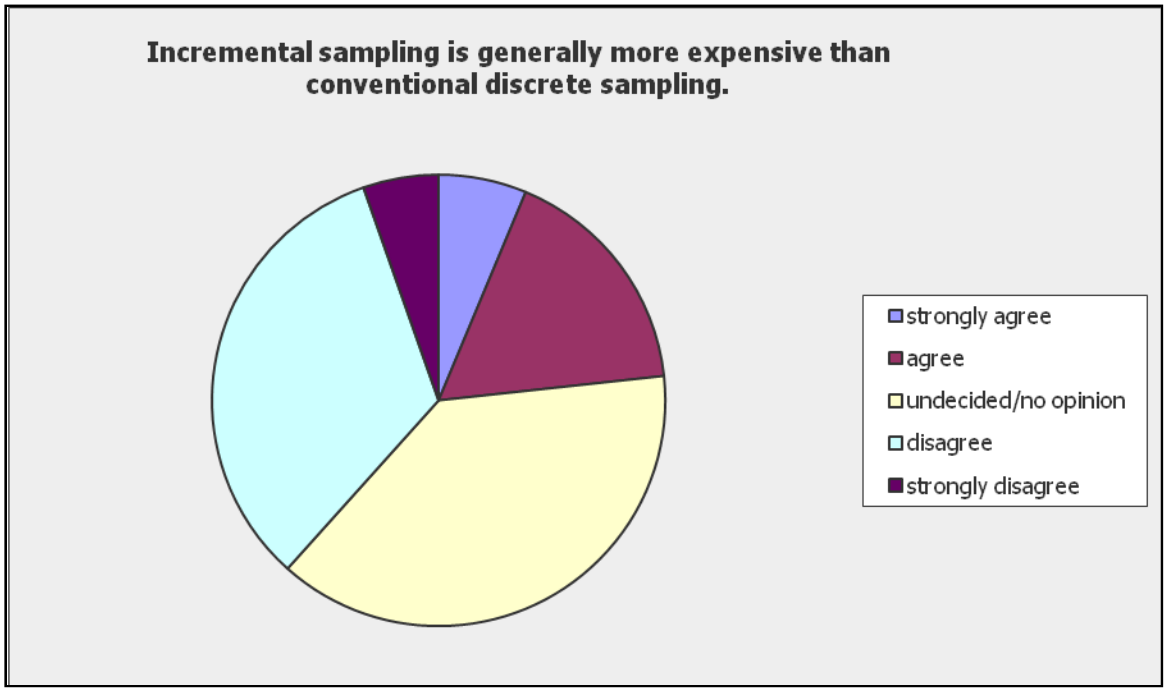
Question 4.

The decision unit (or sampling unit) used in incremental sampling is site-specific.		
Answer Options	**Response Percent**	**Response Count**
strongly agree	35.1%	79
agree	51.6%	116
undecided/no opinion	9.3%	21
disagree	2.2%	5
strongly disagree	1.8%	4
	answered question	**225**
	skipped question	**37**



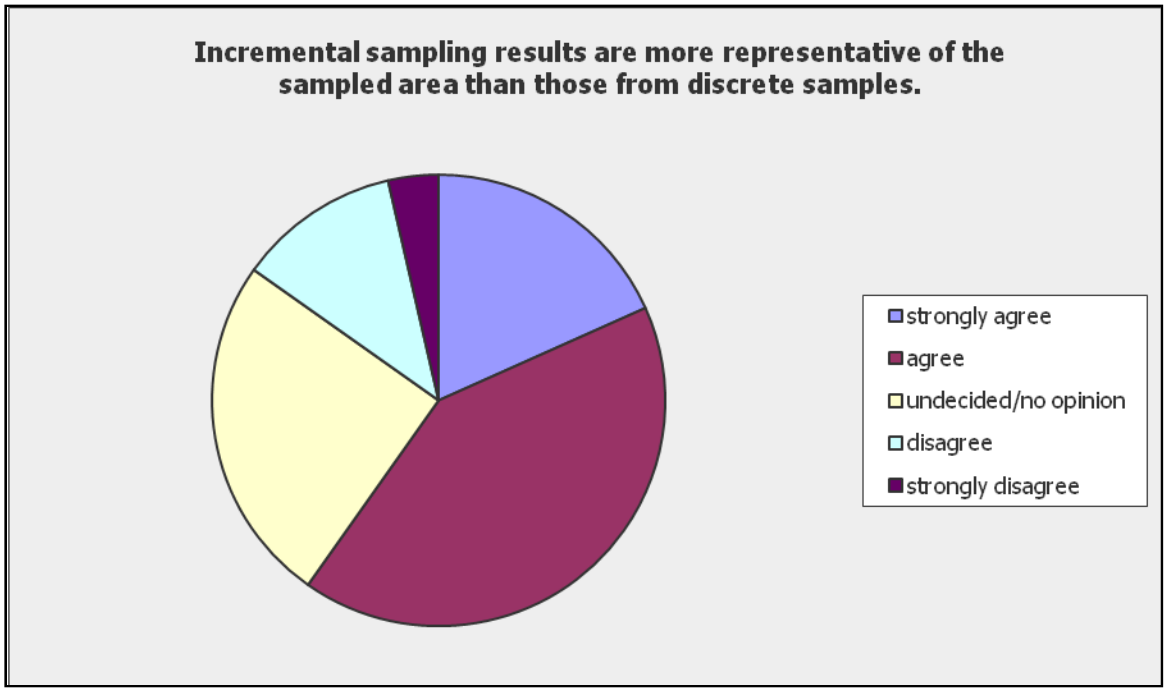
Question 5.

Incremental sampling is generally more expensive than conventional discrete sampling.		
Answer Options	**Response Percent**	**Response Count**
strongly agree	6.3%	14
agree	17.0%	38
undecided/no opinion	38.4%	86
disagree	33.0%	74
strongly disagree	5.4%	12
	answered question	**224**
	skipped question	**38**



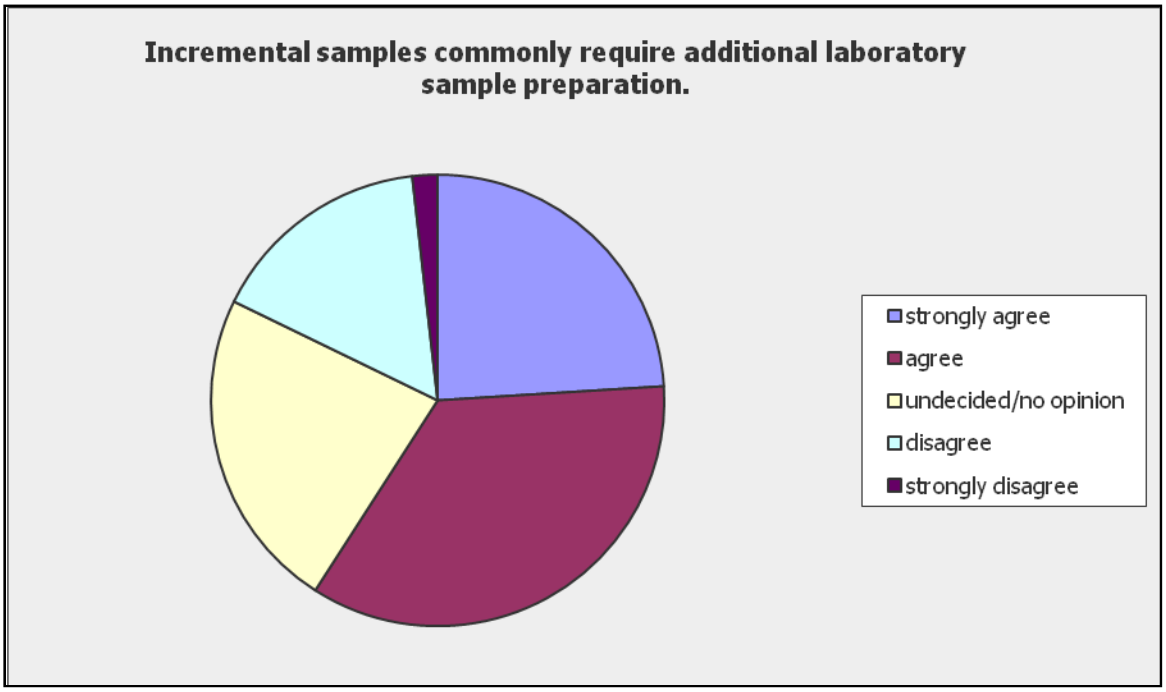
Question 6.

Incremental sampling results are more representative of the sampled area than those from discrete samples.		
Answer Options	**Response Percent**	**Response Count**
strongly agree	18.3%	41
agree	41.5%	93
undecided/no opinion	25.0%	56
disagree	11.6%	26
strongly disagree	3.6%	8
	answered question	**224**
	skipped question	**38**



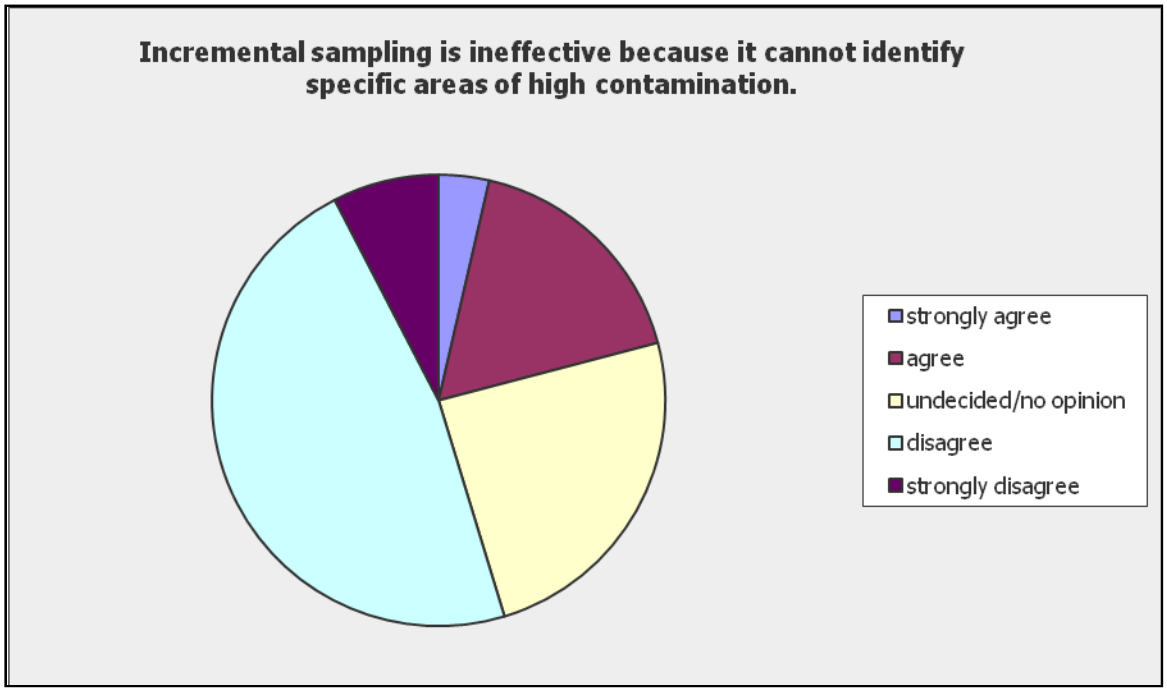
Question 7.

Incremental samples commonly require additional laboratory sample preparation.		
Answer Options	**Response Percent**	**Response Count**
strongly agree	24.0%	54
agree	35.1%	79
undecided/no opinion	23.1%	52
disagree	16.0%	36
strongly disagree	1.8%	4
	answered question	**225**
	skipped question	**37**



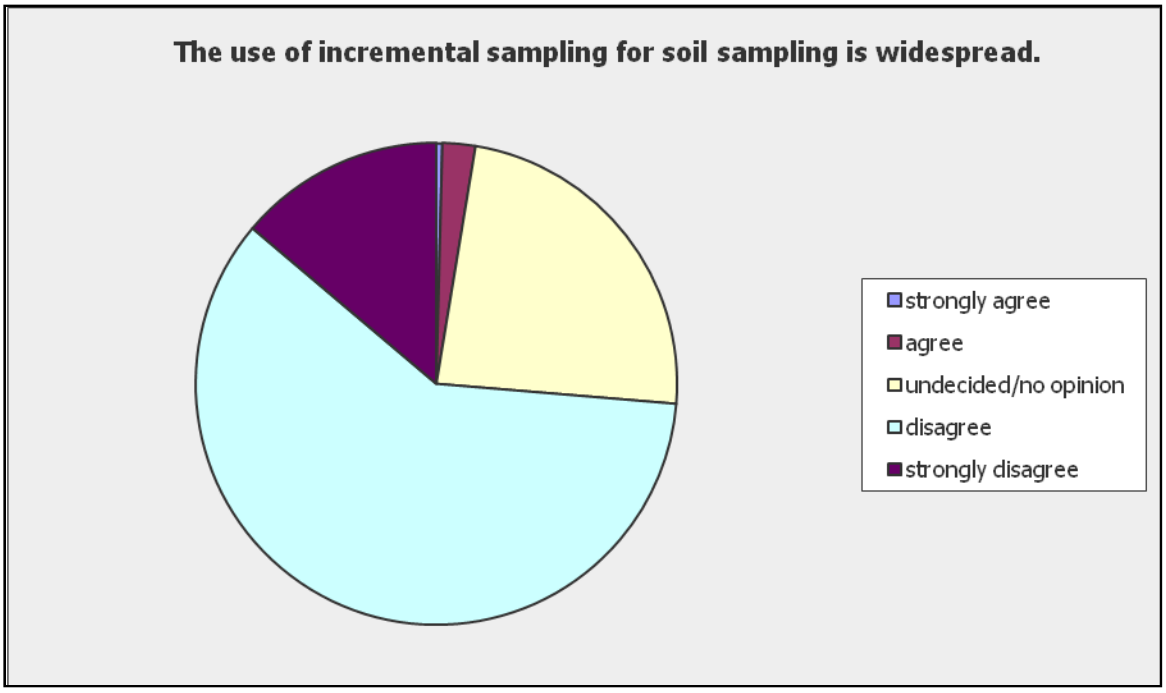
Question 8.

Incremental sampling is ineffective because it cannot identify specific areas of high contamination.		
Answer Options	**Response Percent**	**Response Count**
strongly agree	3.6%	8
agree	17.3%	39
undecided/no opinion	24.4%	55
disagree	47.1%	106
strongly disagree	7.6%	17
	answered question	**225**
	skipped question	**37**



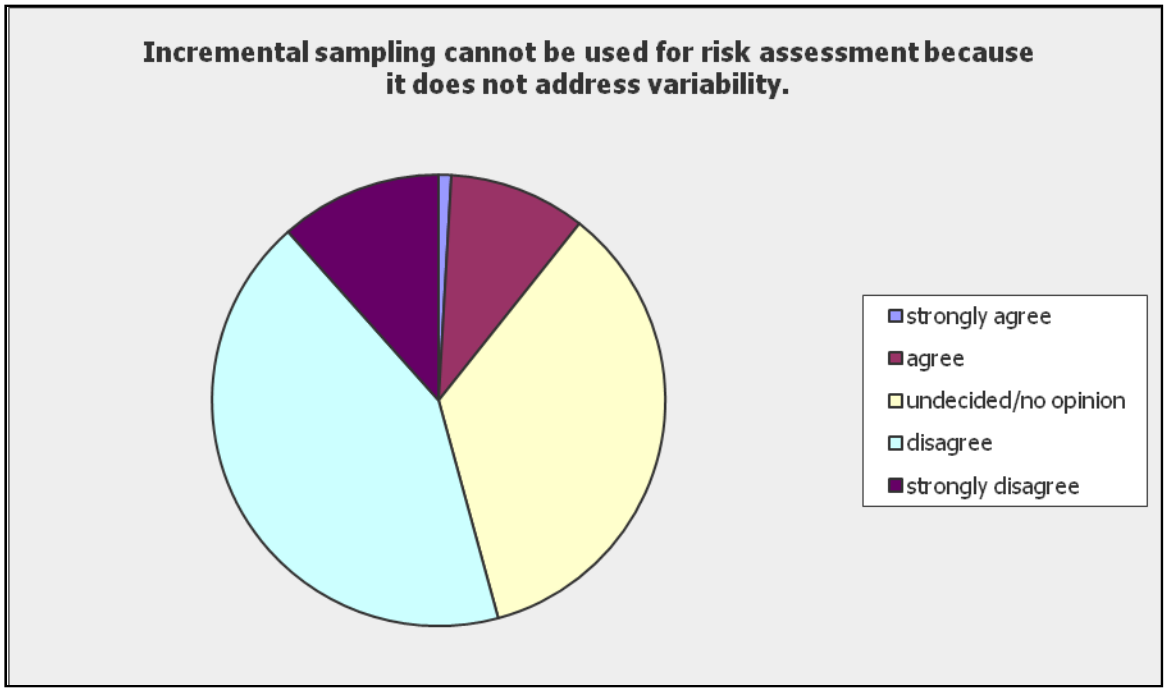
Question 9.

The use of incremental sampling for soil sampling is widespread.		
Answer Options	**Response Percent**	**Response Count**
strongly agree	0.4%	1
agree	2.2%	5
undecided/no opinion	23.7%	53
disagree	59.8%	134
strongly disagree	13.8%	31
answered question		**224**
skipped question		**38**



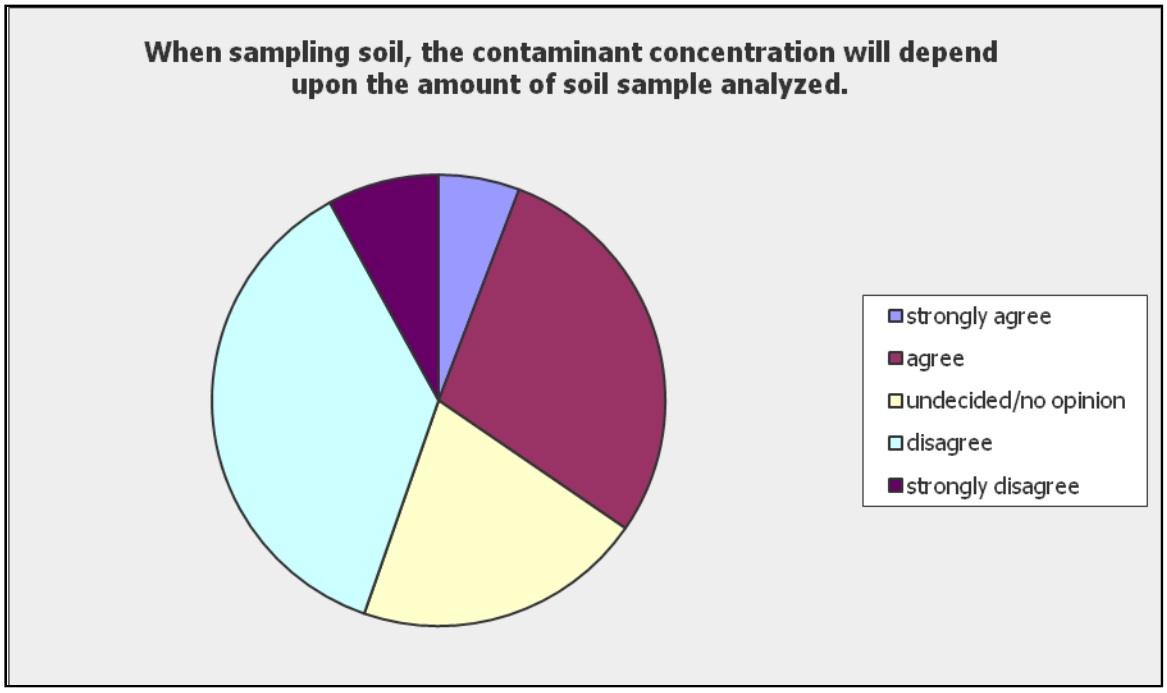
Question 10.

Incremental sampling cannot be used for risk assessment because it does not address variability.		
Answer Options	**Response Percent**	**Response Count**
strongly agree	0.9%	2
agree	9.8%	22
undecided/no opinion	35.1%	79
disagree	42.7%	96
strongly disagree	11.6%	26
answered question		**225**
skipped question		**37**



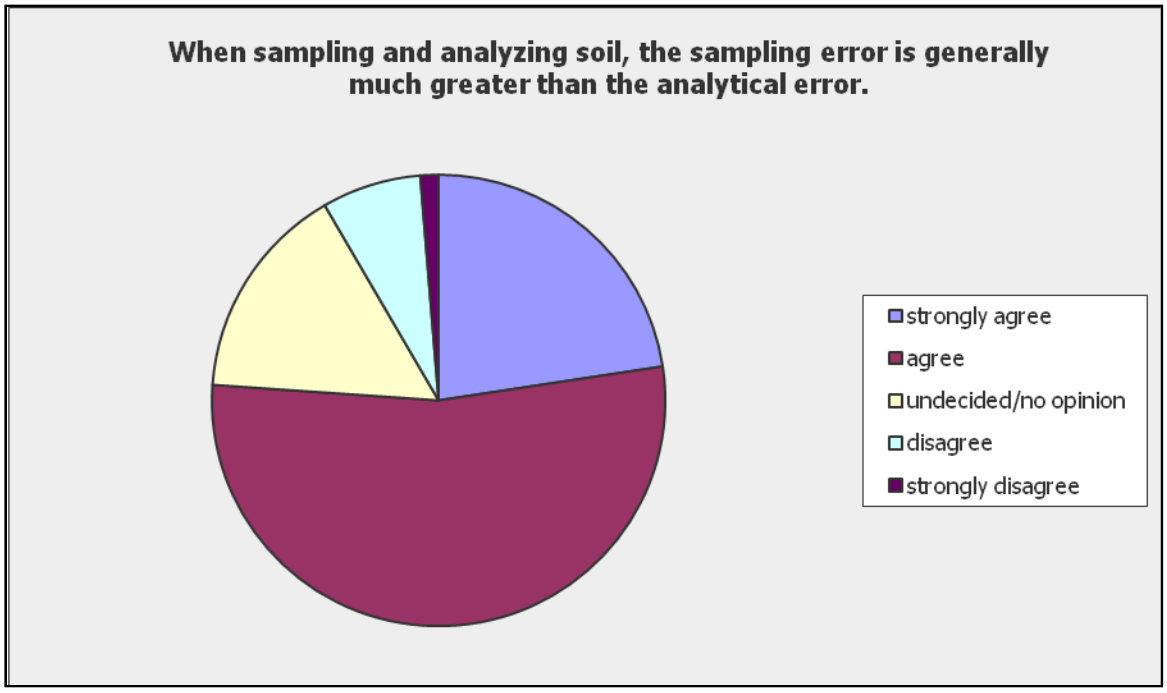
Question 11.

When sampling soil, the contaminant concentration will depend upon the amount of soil sample analyzed.		
Answer Options	**Response Percent**	**Response Count**
strongly agree	5.8%	13
agree	28.8%	65
undecided/no opinion	20.8%	47
disagree	36.7%	83
strongly disagree	8.0%	18
answered question		**226**
skipped question		**36**



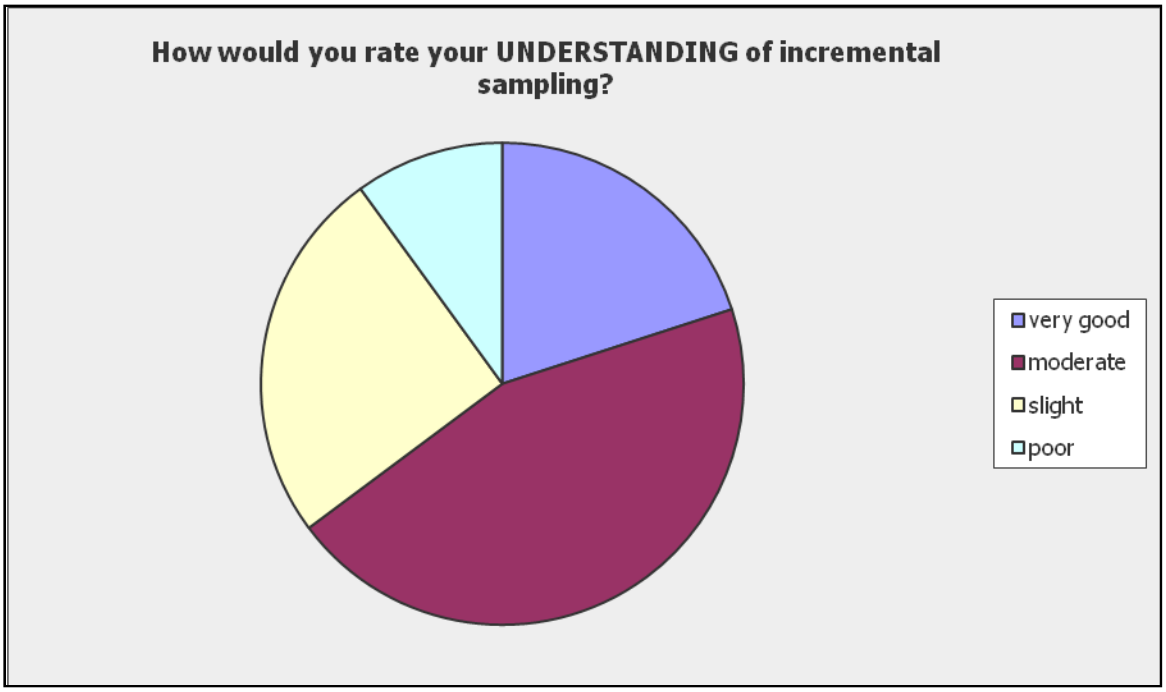
Question 12.

When sampling and analyzing soil, the sampling error is generally much greater than the analytical error.		
Answer Options	**Response Percent**	**Response Count**
strongly agree	22.6%	51
agree	53.5%	121
undecided/no opinion	15.5%	35
disagree	7.1%	16
strongly disagree	1.3%	3
	answered question	**226**
	skipped question	**36**



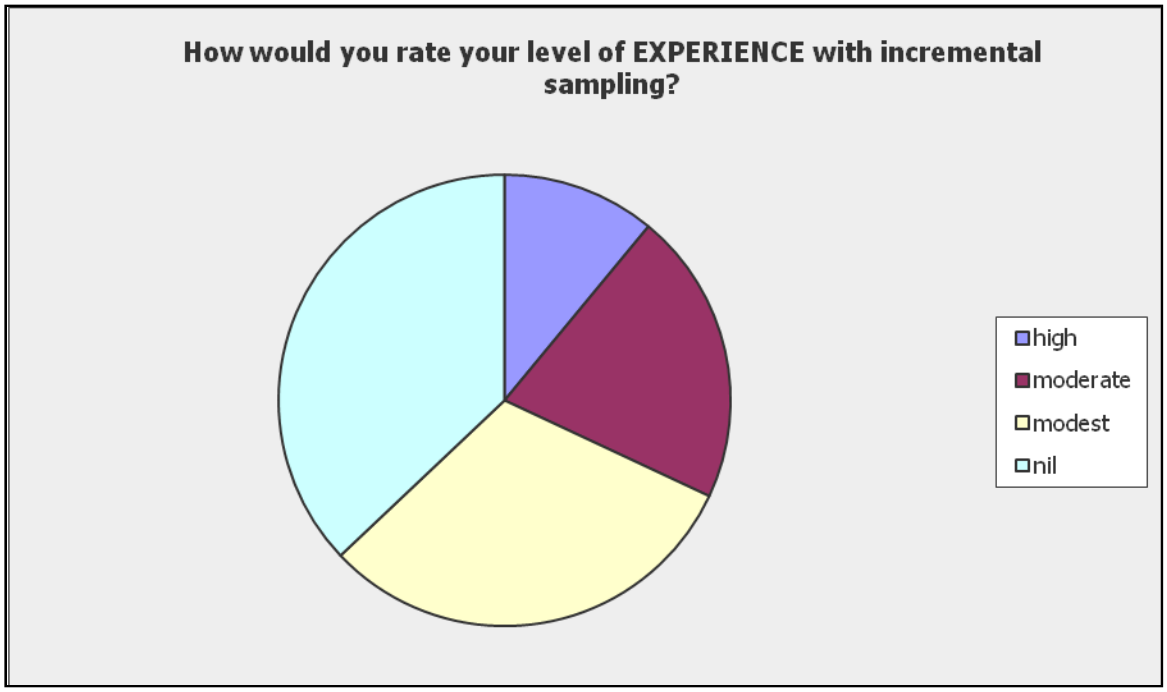
Question 13.

How would you rate your UNDERSTANDING of incremental sampling?		
Answer Options	**Response Percent**	**Response Count**
very good	20.0%	42
moderate	44.8%	94
slight	25.2%	53
poor	10.0%	21
answered question		**210**
skipped question		**52**



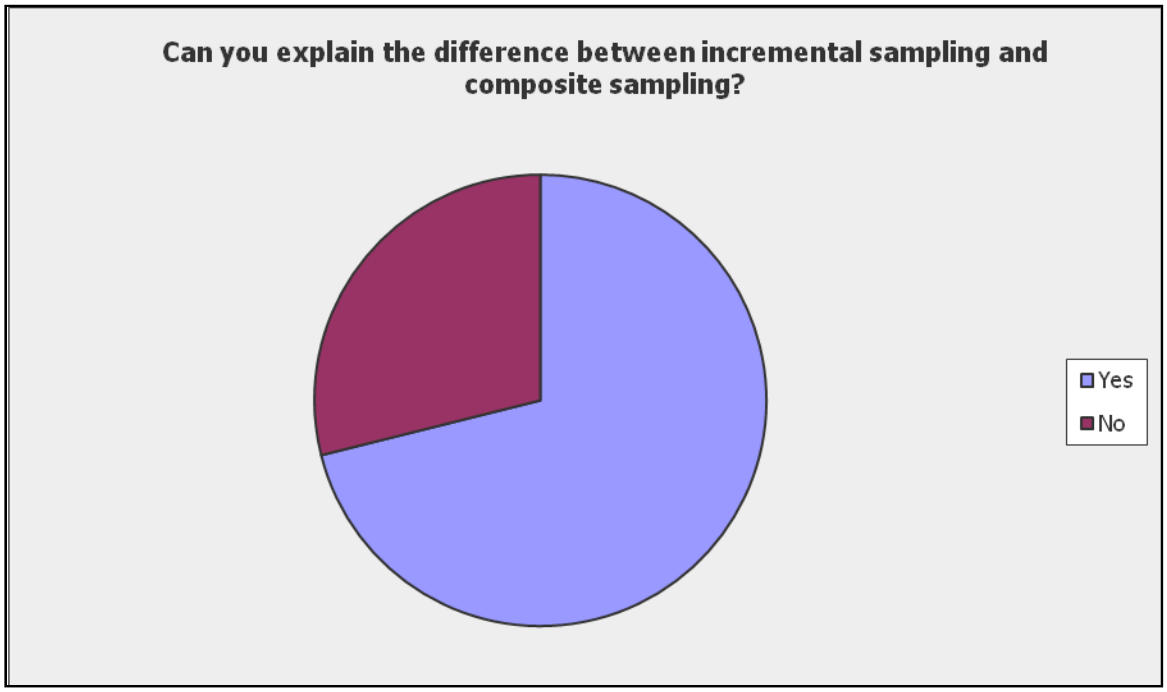
Question 14.

How would you rate your level of EXPERIENCE with incremental sampling?		
Answer Options	**Response Percent**	**Response Count**
high	11.0%	23
moderate	21.0%	44
modest	31.0%	65
nil	37.1%	78
answered question		**210**
skipped question		**52**



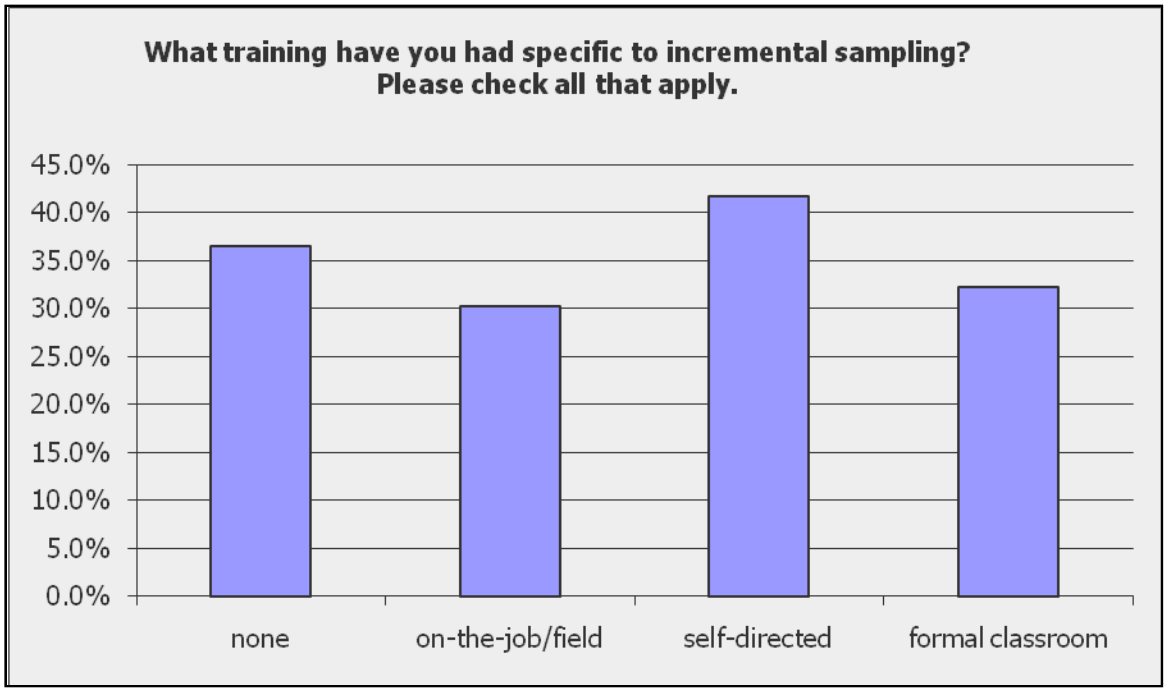
Question 15.

Can you explain the difference between incremental sampling and composite sampling?		
Answer Options	**Response Percent**	**Response Count**
Yes	71.1%	150
No	28.9%	61
answered question		**211**
skipped question		**51**



Question 16.

What training have you had specific to incremental sampling? Please check all that apply.		
Answer Options	**Response Percent**	**Response Count**
none	36.5%	77
on-the-job/field	30.3%	64
self-directed	41.7%	88
formal classroom	32.2%	68
answered question		**211**
skipped question		**51**



Question 17.

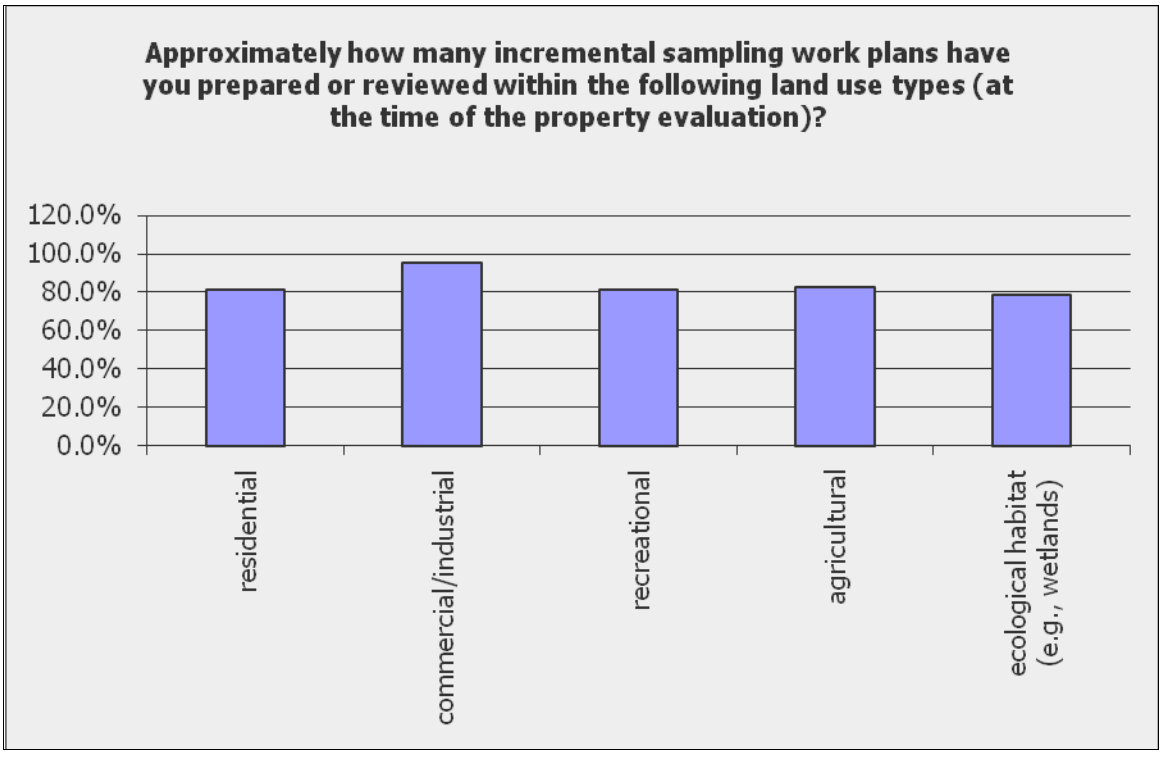
Approximately how many days of FIELD training have you had related specifically to incremental sampling?	
Answer Options	**Response Count**
	194
answered question	**194**
skipped question	**68**

Question 18.

Approximately how many days of CLASSROOM training (including conference seminars) have you had related specifically to incremental sampling?	
Answer Options	**Response Count**
	198
answered question	**198**
skipped question	**64**

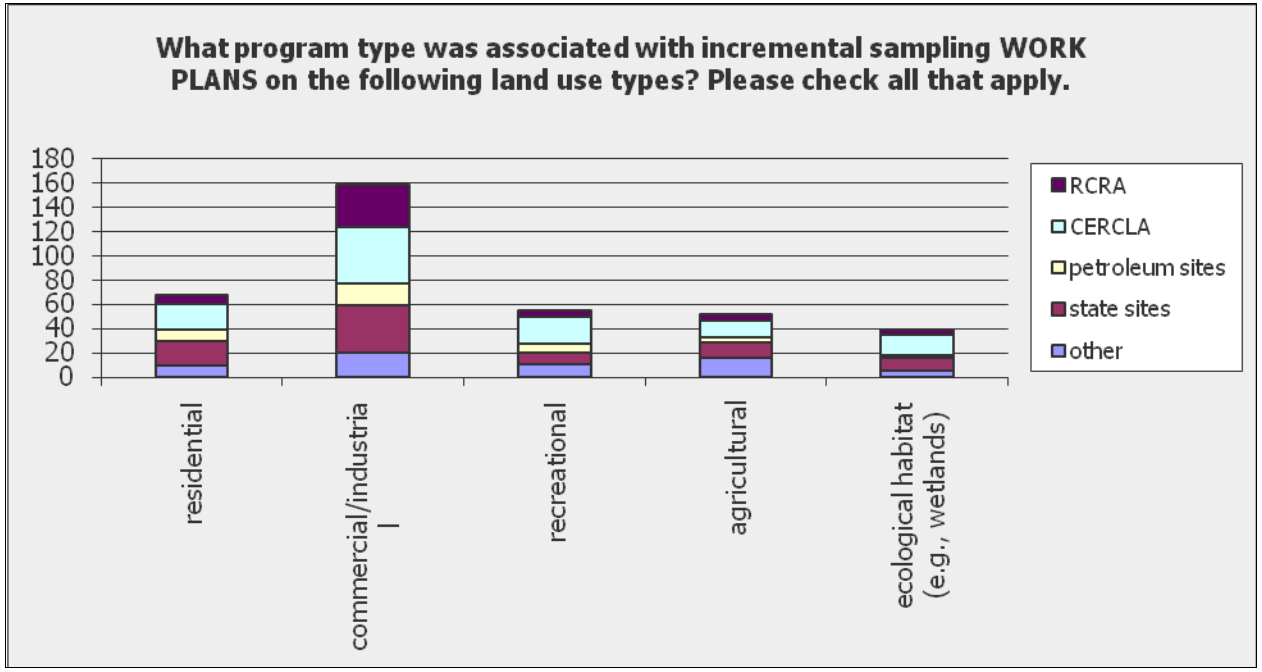
Question 19.

Approximately how many incremental sampling work plans have you prepared or reviewed within the following land use types (at the time of the property evaluation)?		
Answer Options	**Response Percent**	**Response Count**
residential	81.4%	158
commercial/industrial	95.4%	185
recreational	81.4%	158
agricultural	82.5%	160
ecological habitat (e.g., wetlands)	78.9%	153
answered question		**194**
skipped question		**68**



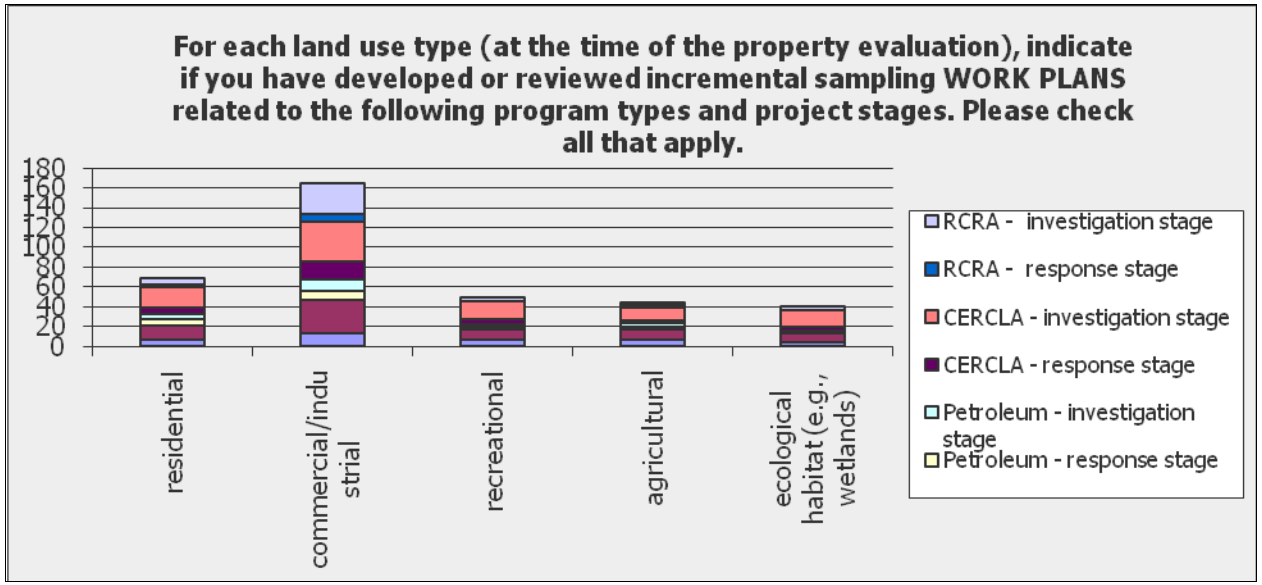
Question 20.

What program type was associated with incremental sampling WORK PLANS on the following land use types? Please check all that apply.						
Answer Options	**RCRA**	**CERCLA**	**petroleum sites**	**state sites**	**other**	**Response Count**
Residential	7	21	10	20	9	38
commercial/industrial	35	46	18	39	20	90
Recreational	6	22	7	10	10	32
Agricultural	5	14	4	13	15	33
ecological habitat (e.g., wetlands)	5	16	3	10	5	26
					answered question	**105**
					skipped question	**157**



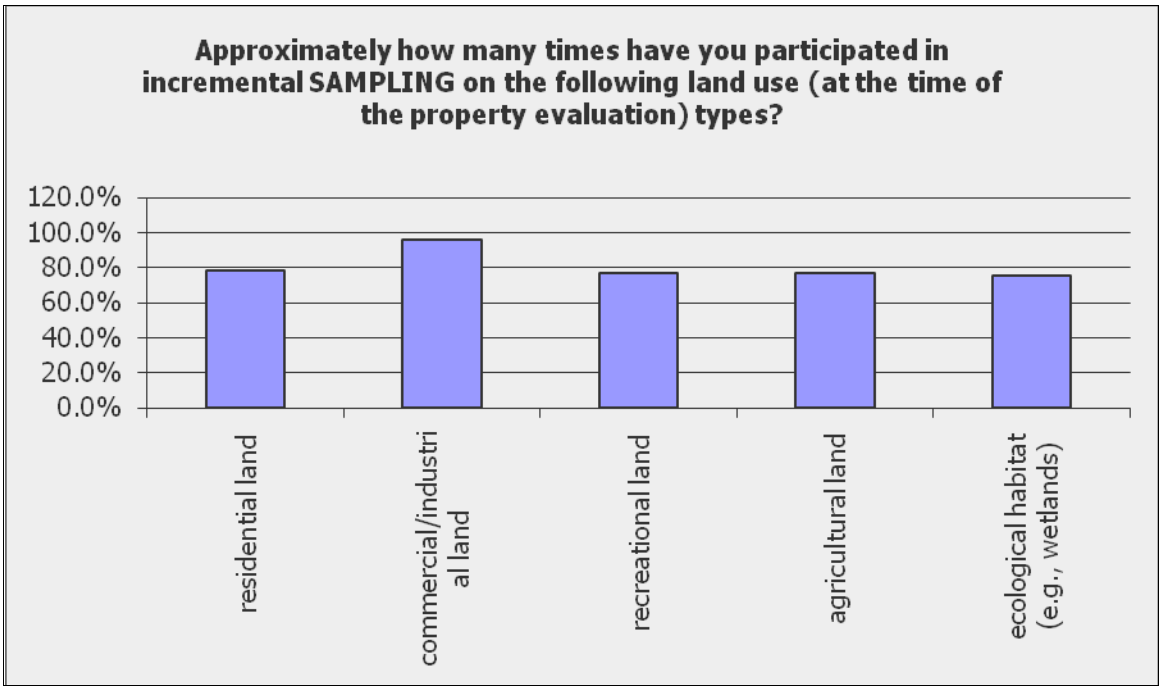
Question 21.

For each land use type (at the time of the property evaluation), indicate if you have developed or reviewed incremental sampling WORK PLANS related to the following program types and project stages. Please check all that apply.									
Answer Options	**RCRA - investigation stage**	**RCRA - response stage**	**CERCLA - investigation stage**	**CERCLA - response stage**	**Petroleum - investigation stage**	**Petroleum - response stage**	**State - investigation stage**	**State - response stage**	**Response Count**
Residential	7	2	21	6	6	6	15	6	32
commercial/ industrial	30	8	41	17	12	9	34	13	79
Recreational	3	1	18	5	2	3	11	6	26
Agricultural	3	2	13	3	4	2	10	7	24
ecological habitat (e.g., wetlands)	4	0	17	3	2	1	9	4	24
								answered question	**91**
								skipped question	**171**



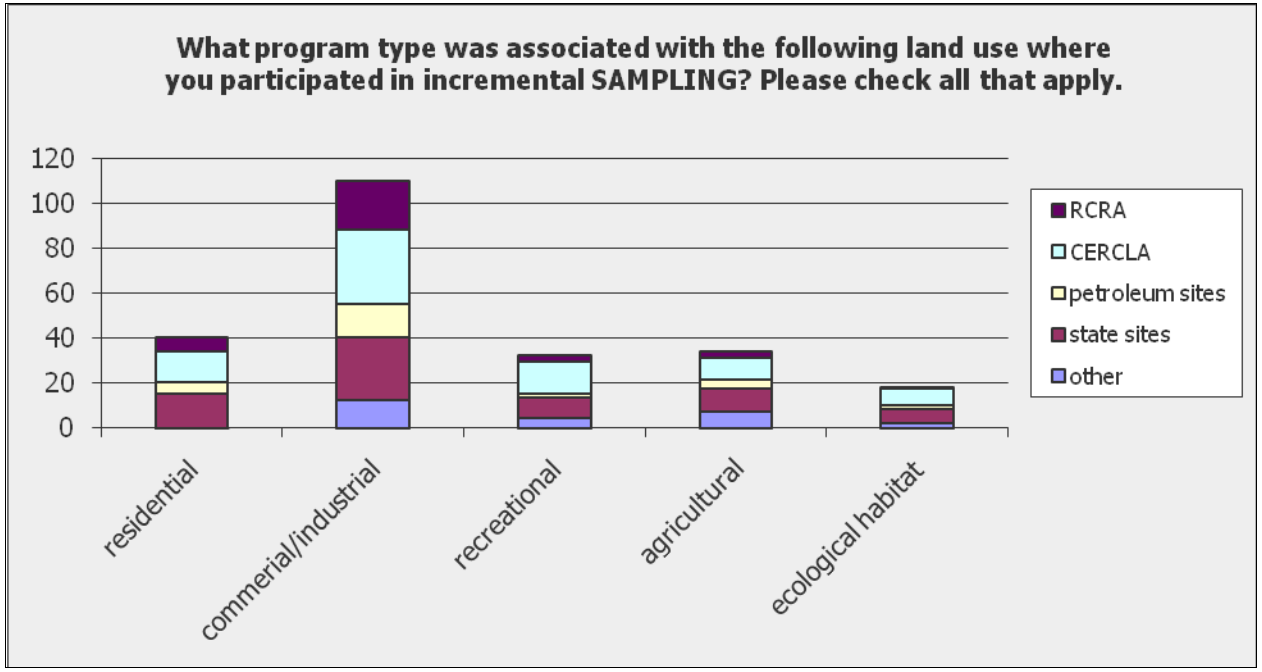
Question 22.

Approximately how many times have you participated in incremental SAMPLING on the following land use (at the time of the property evaluation) types?		
Answer Options	**Response Percent**	**Response Count**
residential land	78.4%	138
commercial/industrial land	96.0%	169
recreational land	76.7%	135
agricultural land	76.7%	135
ecological habitat (e.g., wetlands)	75.6%	133
	answered question	**176**
	skipped question	**86**



Question 23.

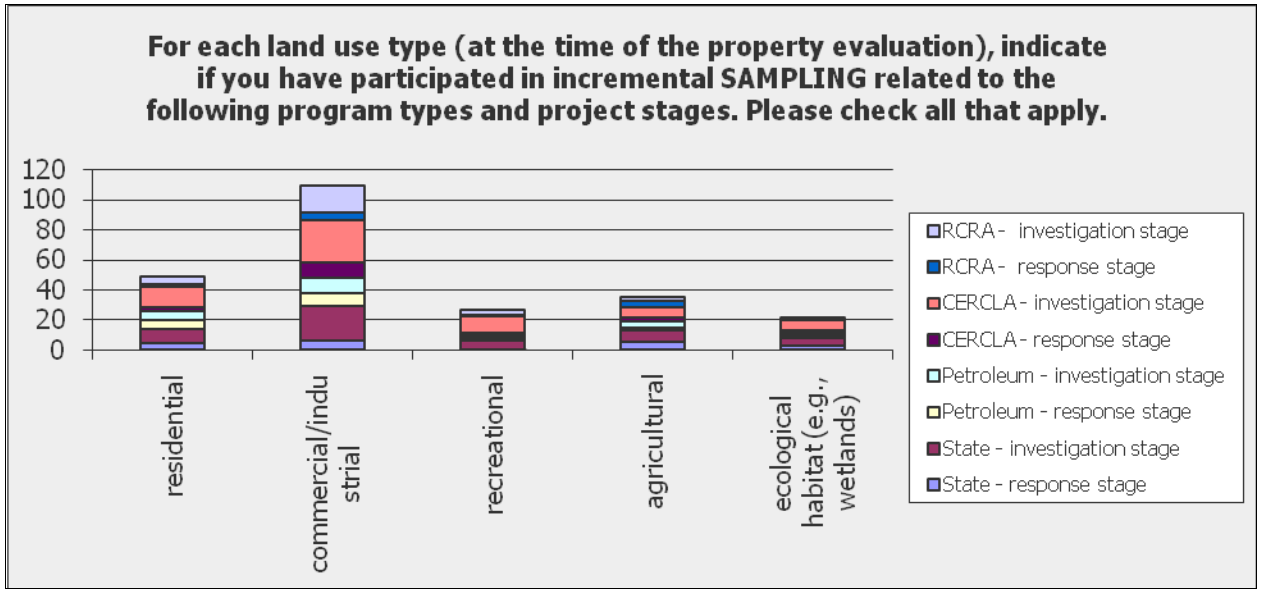
What program type was associated with the following land use where you participated in incremental SAMPLING? Please check all that apply.						
Answer Options	**RCRA**	**CERCLA**	**petroleum sites**	**state sites**	**other**	**Response Count**
Residential	6	14	5	14	1	23
commercial/industrial	22	33	15	28	12	64
Recreational	3	14	2	9	4	21
Agricultural	3	10	4	10	7	21
ecological habitat	1	7	2	6	2	11
					answered question	**76**
					skipped question	**186**



Question 24.

For each land use type (at the time of the property evaluation), indicate if you have participated in incremental SAMPLING related to the following program types and project stages. Please check all that apply.

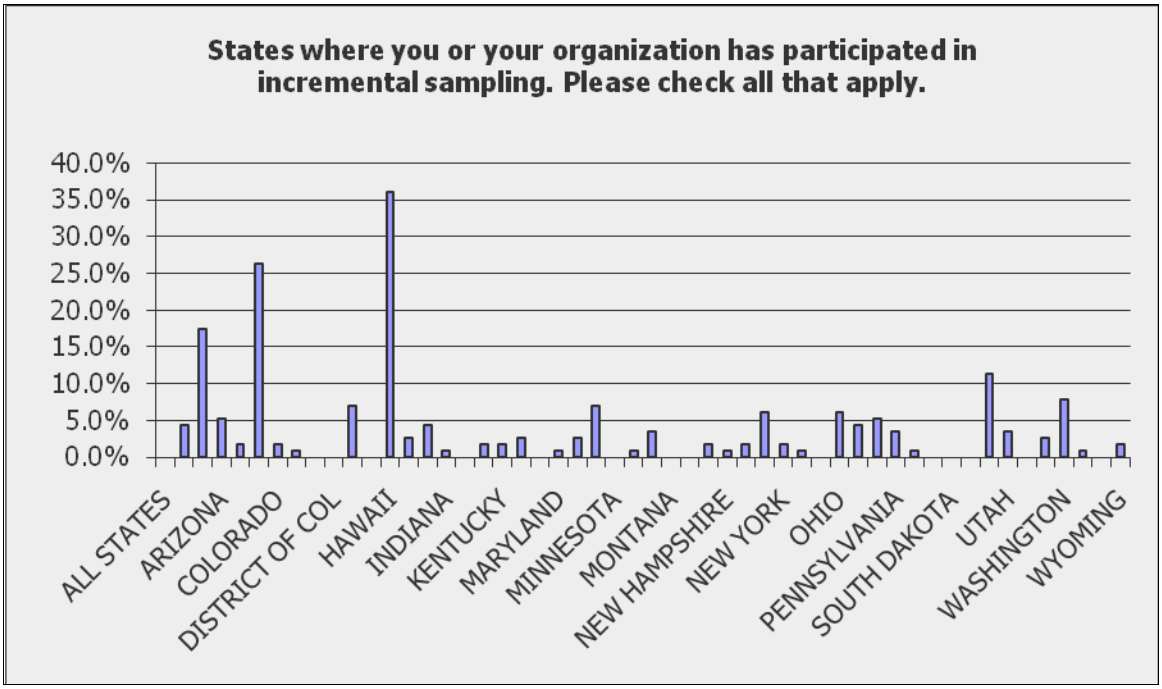
Answer Options	RCRA - investigation stage	RCRA - response stage	CERCLA - investigation stage	CERCLA - response stage	Petroleum - investigation stage	Petroleum - response stage	State - investigation stage	State - response stage	Response Count
Residential	5	2	13	3	6	6	9	5	21
commercial/ industrial	18	5	28	11	10	8	23	7	55
Recreational	3	1	11	2	2	1	5	2	13
Agricultural	3	4	7	3	4	2	7	6	16
ecological habitat (e.g., wetlands)	2	0	7	1	2	2	5	3	12
								answered question	**65**
								skipped question	**197**



Question 25.

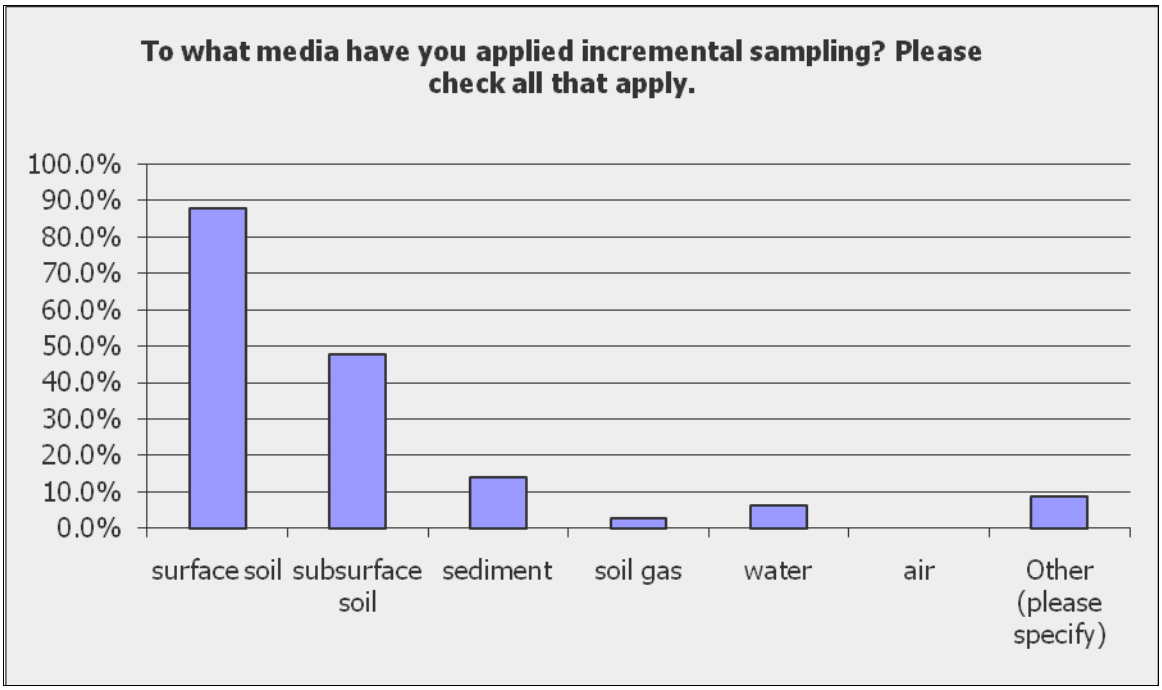
States where you or your organization has participated in incremental sampling. Please check all that apply.		
Answer Options	**Response Percent**	**Response Count**
ALL STATES	0.0%	0
ALABAMA	4.4%	5
ALASKA	17.5%	20
ARIZONA	5.3%	6
ARKANSAS	1.8%	2
CALIFORNIA	26.3%	30
COLORADO	1.8%	2
CONNECTICUT	0.9%	1
DELAWARE	0.0%	0
DISTRICT OF COL	0.0%	0
FLORIDA	7.0%	8
GEORGIA	0.0%	0
HAWAII	36.0%	41
IDAHO	2.6%	3
ILLINOIS	4.4%	5
INDIANA	0.9%	1
IOWA	0.0%	0
KANSAS	1.8%	2
KENTUCKY	1.8%	2
LOUISIANA	2.6%	3
MAINE	0.0%	0
MARYLAND	0.9%	1
MASSACHUSETTS	2.6%	3
MICHIGAN	7.0%	8
MINNESOTA	0.0%	0
MISSISSIPPI	0.9%	1
MISSOURI	3.5%	4
MONTANA	0.0%	0
NEBRASKA	0.0%	0
NEVADA	1.8%	2
NEW HAMPSHIRE	0.9%	1
NEW JERSEY	1.8%	2
NEW MEXICO	6.1%	7
NEW YORK	1.8%	2
NORTH CAROLINA	0.9%	1
NORTH DAKOTA	0.0%	0
OHIO	6.1%	7
OKLAHOMA	4.4%	5
OREGON	5.3%	6
PENNSYLVANIA	3.5%	4
RHODE ISLAND	0.9%	1
SOUTH CAROLINA	0.0%	0
SOUTH DAKOTA	0.0%	0
TENNESSEE	0.0%	0

TEXAS	11.4%	13
UTAH	3.5%	4
VERMONT	0.0%	0
VIRGINIA	2.6%	3
WASHINGTON	7.9%	9
WEST VIRGINIA	0.9%	1
WISCONSIN	0.0%	0
WYOMING	1.8%	2
	answered question	**114**
	skipped question	**148**



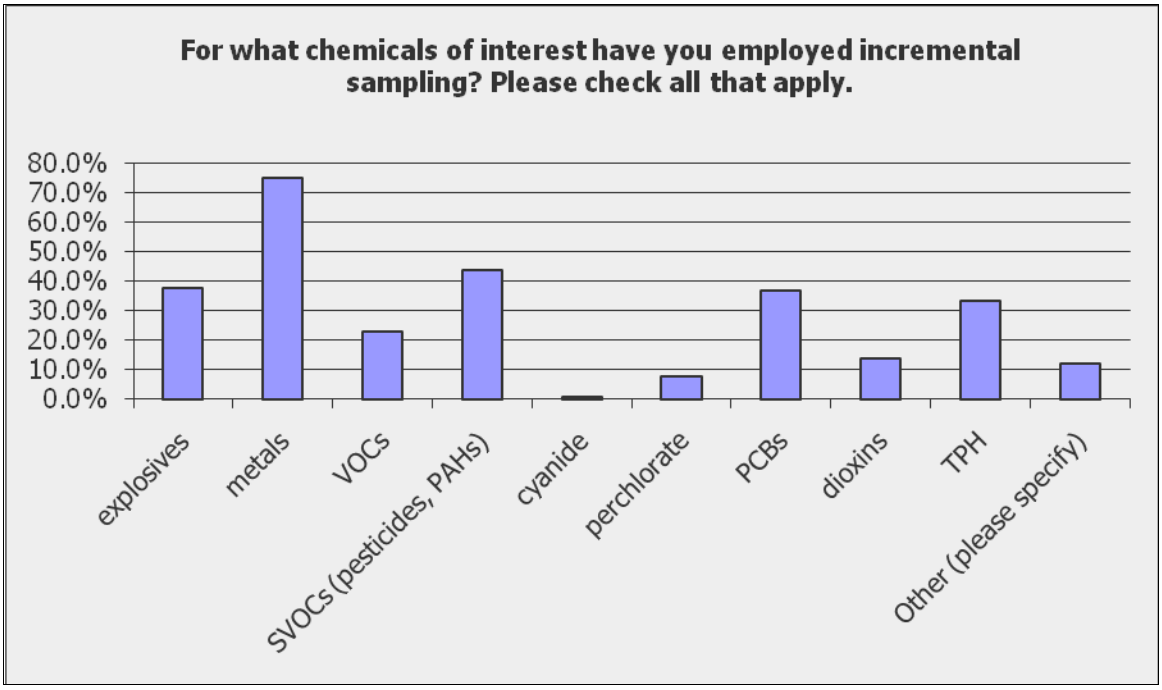
Question 26.

To what media have you applied incremental sampling? Please check all that apply.		
Answer Options	**Response Percent**	**Response Count**
surface soil	87.8%	101
subsurface soil	47.8%	55
sediment	13.9%	16
soil gas	2.6%	3
water	6.1%	7
air	0.0%	0
Other (please specify)	8.7%	10
answered question		**115**
skipped question		**147**



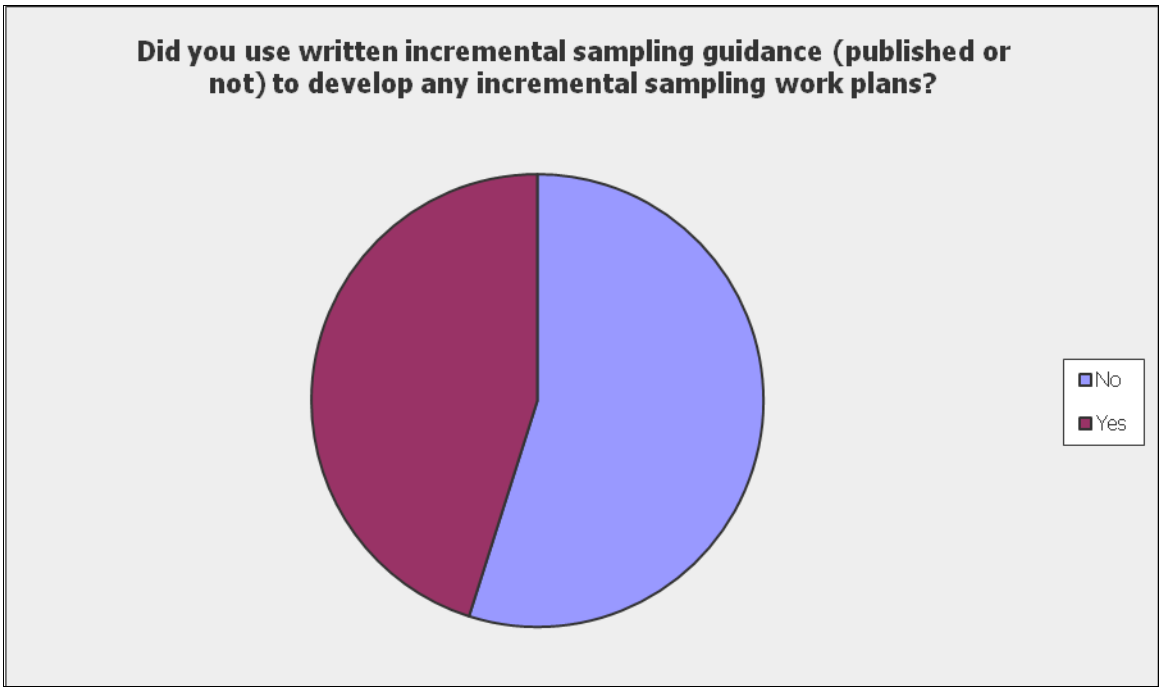
Question 27.

For what chemicals of interest have you employed incremental sampling? Please check all that apply.		
Answer Options	**Response Percent**	**Response Count**
explosives	37.6%	44
metals	75.2%	88
VOCs	23.1%	27
SVOCs (pesticides, PAHs)	43.6%	51
cyanide	0.9%	1
perchlorate	7.7%	9
PCBs	36.8%	43
dioxins	13.7%	16
TPH	33.3%	39
Other (please specify)	12.0%	14
answered question		**117**
skipped question		**145**



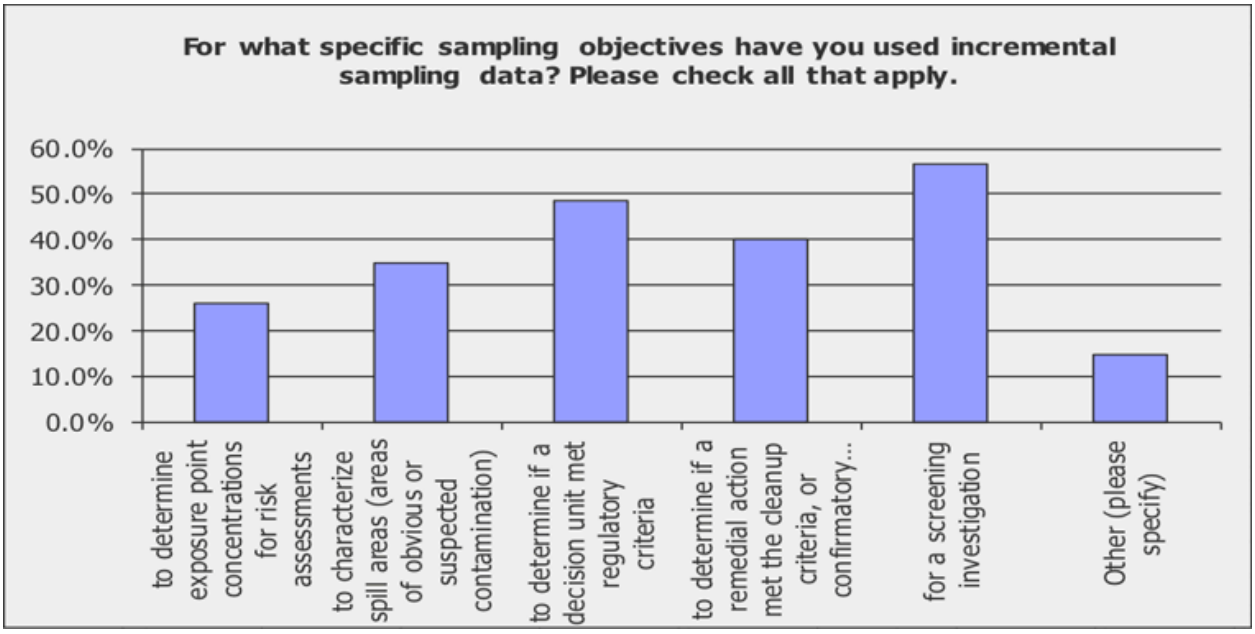
Question 28.

Did you use written incremental sampling guidance (published or not) to develop any incremental sampling work plans?		
Answer Options	**Response Percent**	**Response Count**
No	54.9%	73
Yes	45.1%	60
if yes, please list		53
answered question		**133**
skipped question		**129**



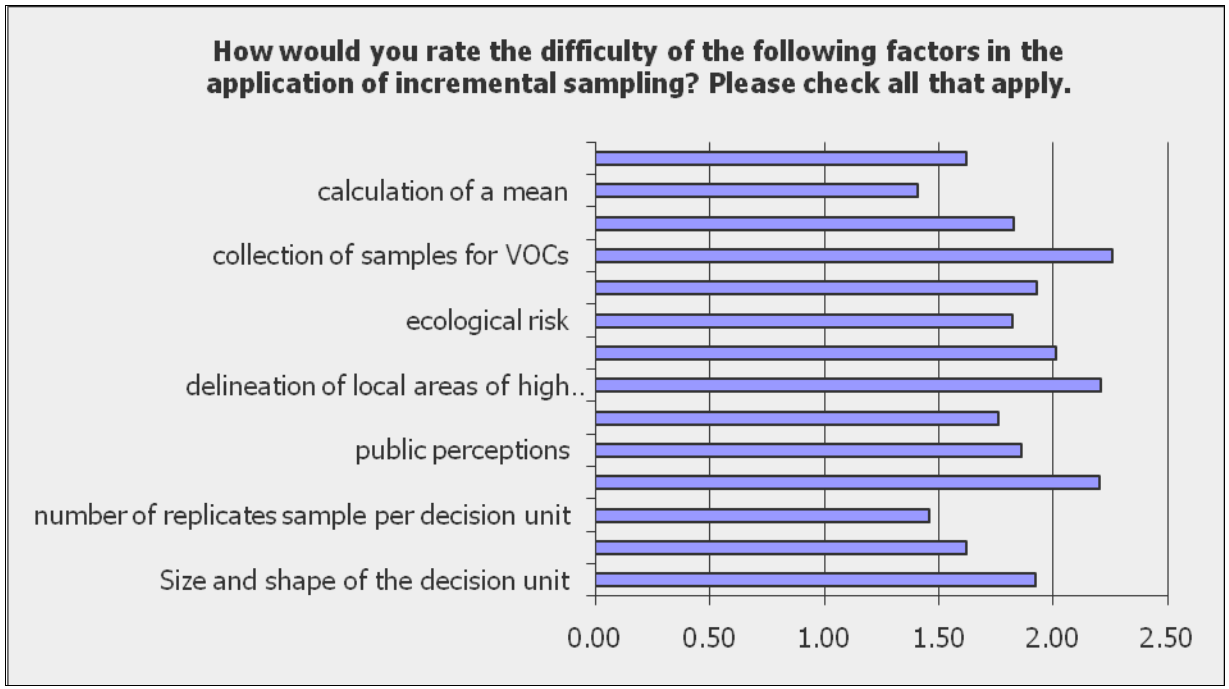
Question 29.

For what specific sampling objectives have you used incremental sampling data? Please check all that apply.		
Answer Options	**Response Percent**	**Response Count**
to determine exposure point concentrations for risk assessments	26.1%	30
to characterize spill areas (areas of obvious or suspected contamination)	34.8%	40
to determine if a decision unit met regulatory criteria	48.7%	56
to determine if a remedial action met the cleanup criteria, or confirmatory sampling	40.0%	46
for a screening investigation	56.5%	65
Other (please specify)	14.8%	17
	answered question	**115**
	skipped question	**147**



Question 30.

How would you rate the difficulty of the following factors in the application of incremental sampling? Please check all that apply.					
Answer Options	**not difficult**	**moderately difficult**	**very difficult**	**Rating Average**	**Response Count**
Size and shape of the decision unit	37	74	26	1.92	137
number of increments per decision unit	61	66	9	1.62	136
number of replicates sample per decision unit	76	54	4	1.46	134
regulatory issues or acceptance	23	62	50	2.20	135
public perceptions	41	57	24	1.86	122
laboratory capabilities	53	62	20	1.76	135
delineation of local areas of high concentration	19	59	45	2.21	123
delineation of the extent of the release	31	63	32	2.01	126
ecological risk	43	50	22	1.82	115
collection of subsurface samples	38	57	29	1.93	124
collection of samples for VOCs	20	40	48	2.26	108
logistical aspects (e.g., methanol transport)	36	56	17	1.83	109
calculation of a mean	79	30	9	1.41	118
calculation of a 95% UCL of the mean	62	39	17	1.62	118
				answered question	**140**
				skipped question	**122**

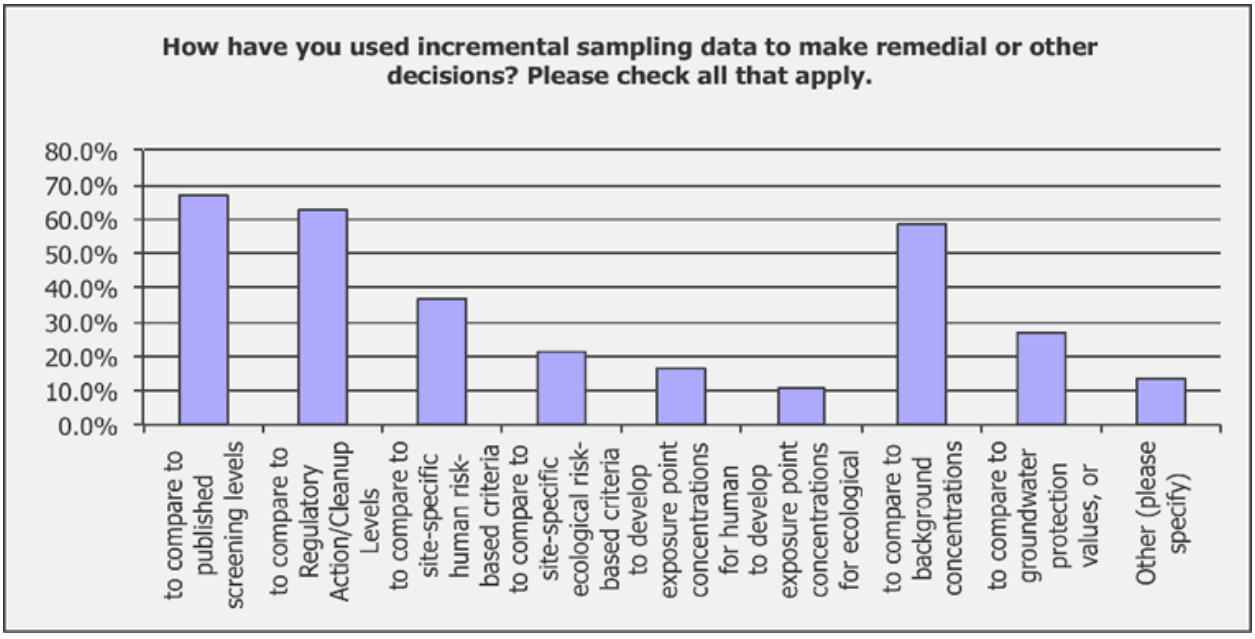

How would you rate the difficulty of the following factors in the application of incremental sampling? Please check all that apply.
calculation of a mean
collection of samples for VOCs
ecological risk
delineation of local areas of high..
public perceptions
number of replicates sample per decision unit
Size and shape of the decision unit
0.00
0.50
1.00
1.50
2.00
2.50

Question 31.

If there are other significant factors that make the application of incremental sampling difficult, please list and evaluate them.	
Answer Options	**Response Count**
	34
answered question	**34**
skipped question	**228**

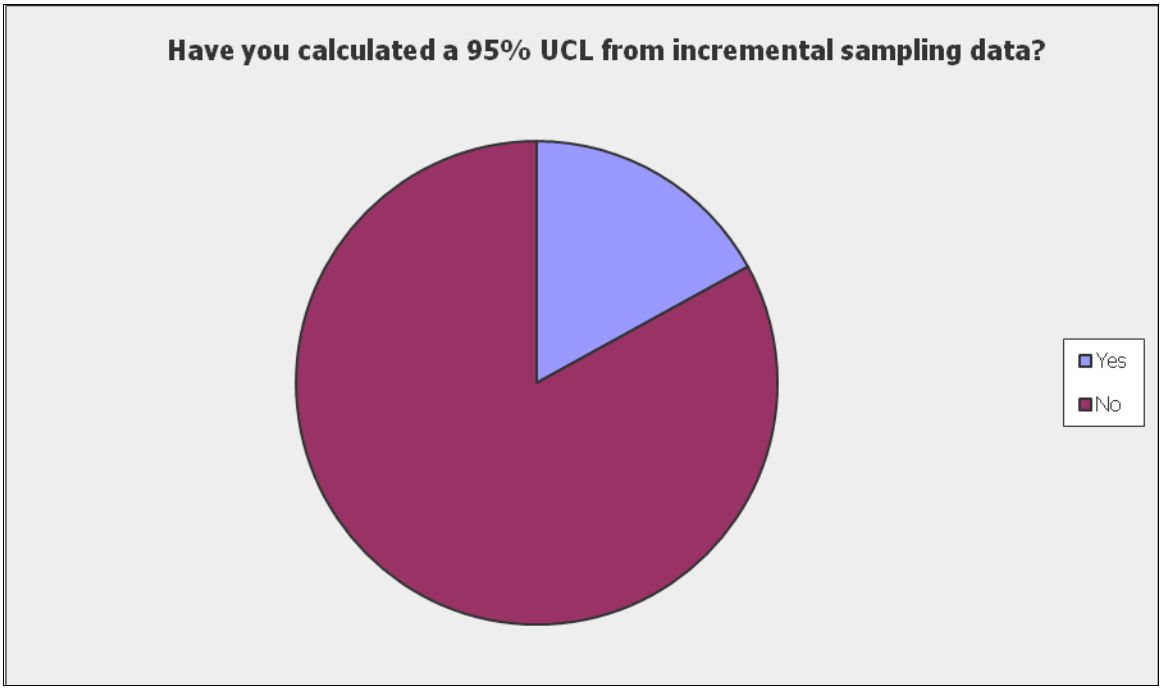
Question 32.

How have you used incremental sampling data to make remedial or other decisions? Please check all that apply.		
Answer Options	**Response Percent**	**Response Count**
to compare to published screening levels	67.3%	70
to compare to Regulatory Action/Cleanup Levels	62.5%	65
to compare to site-specific human risk-based criteria	36.5%	38
to compare to site-specific ecological risk-based criteria	21.2%	22
to develop exposure point concentrations for human receptors	16.3%	17
to develop exposure point concentrations for ecological receptors	10.6%	11
to compare to background concentrations	58.7%	61
to compare to groundwater protection values, or evaluate potential leaching hazards (soil leaching)	26.9%	28
Other (please specify)	13.5%	14
answered question		**104**
skipped question		**158**



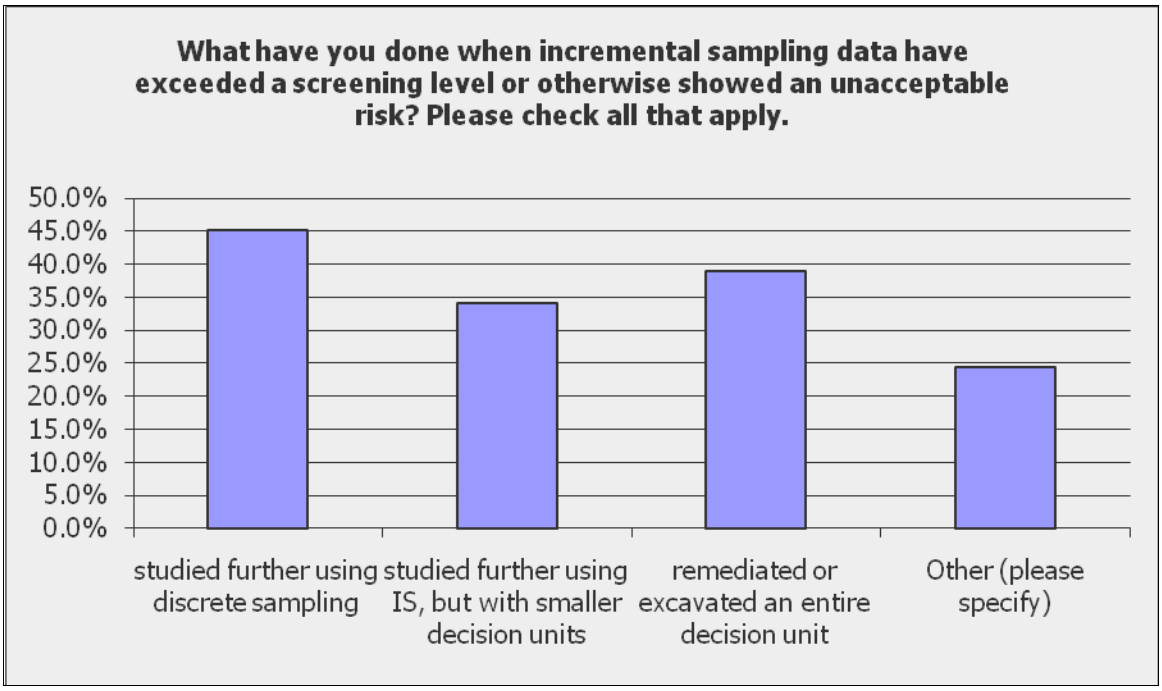
Question 33.

Have you calculated a 95% UCL from incremental sampling data?		
Answer Options	**Response Percent**	**Response Count**
Yes	17.0%	27
No	83.0%	132
If you are willing to share your calculation methodology, please provide contact information (name, phone, or e-mail)		5
answered question		**159**
skipped question		**103**



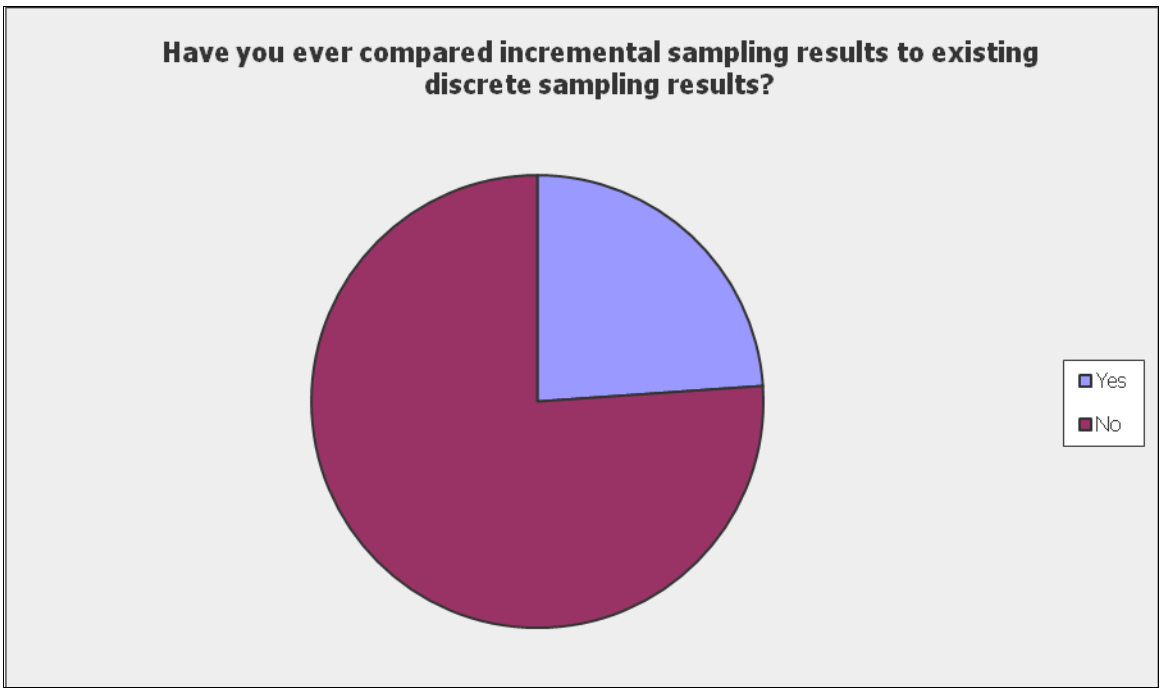
Question 34.

What have you done when incremental sampling data have exceeded a screening level or otherwise showed an unacceptable risk? Please check all that apply.		
Answer Options	**Response Percent**	**Response Count**
studied further using discrete sampling	45.1%	37
studied further using IS, but with smaller decision units	34.1%	28
remediated or excavated an entire decision unit	39.0%	32
Other (please specify)	24.4%	20
answered question		**82**
skipped question		**180**



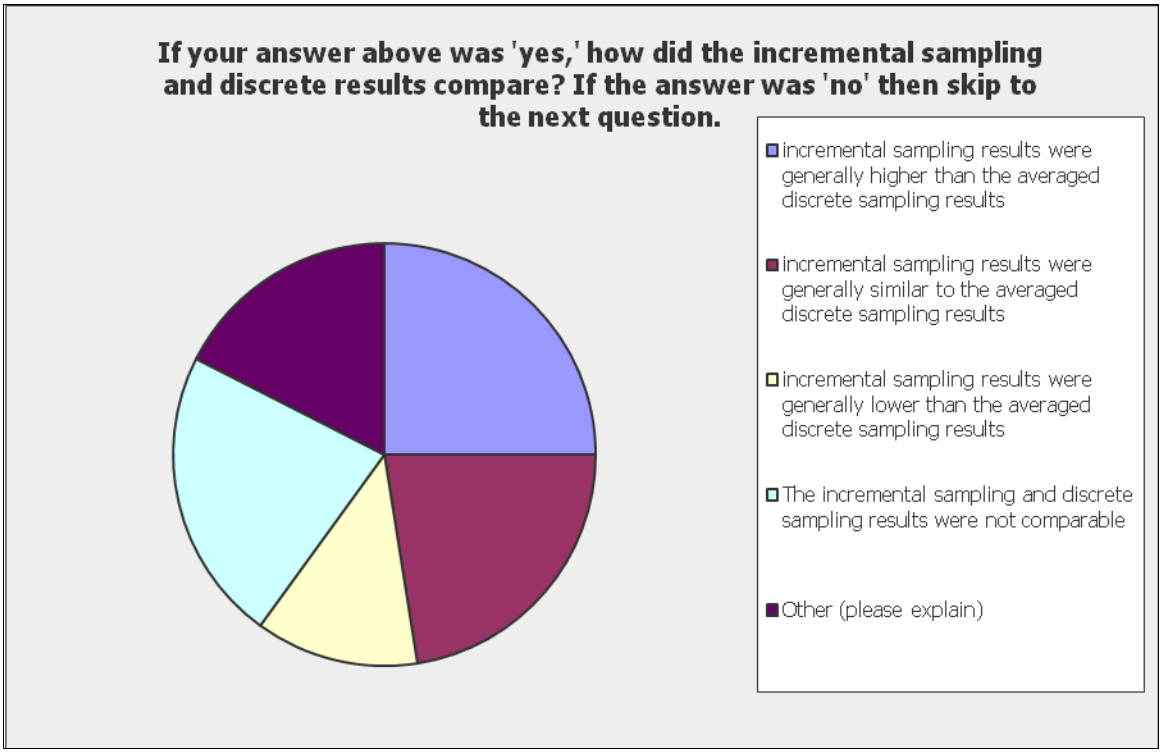
Question 35.

Have you ever compared incremental sampling results to existing discrete sampling results?		
Answer Options	**Response Percent**	**Response Count**
Yes	23.9%	39
No	76.1%	124
answered question		**163**
skipped question		**99**



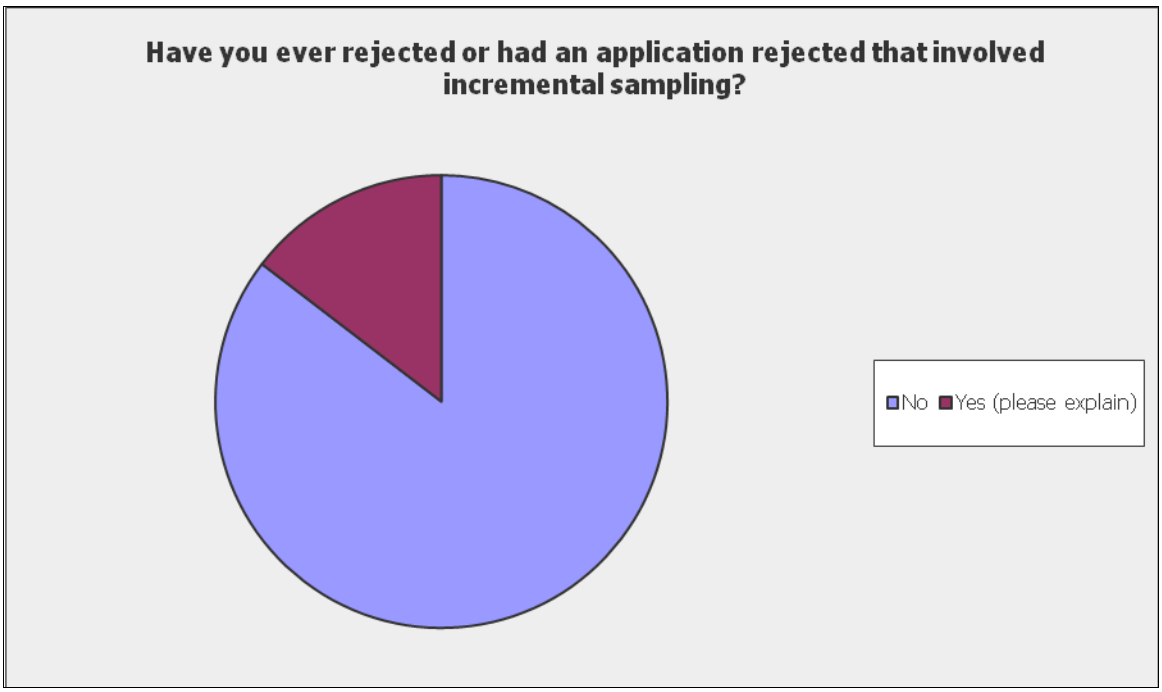
Question 36.

If your answer above was 'yes,' how did the incremental sampling and discrete results compare? If the answer was 'no' then skip to the next question.		
Answer Options	**Response Percent**	**Response Count**
incremental sampling results were generally higher than the averaged discrete sampling results	25.0%	10
incremental sampling results were generally similar to the averaged discrete sampling results	22.5%	9
incremental sampling results were generally lower than the averaged discrete sampling results	12.5%	5
The incremental sampling and discrete sampling results were not comparable	22.5%	9
Other (please explain)	17.5%	7
answered question		**40**
skipped question		**222**



Question 37.

Have you ever rejected or had an application rejected that involved incremental sampling?		
Answer Options	**Response Percent**	**Response Count**
No	85.4%	129
Yes (please explain)	14.6%	22
answered question		**151**
skipped question		**111**

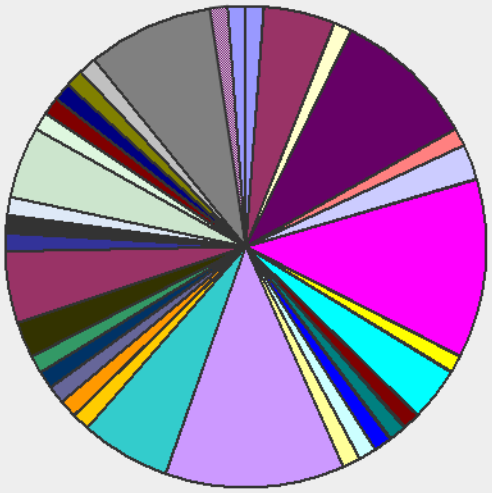


Question 38.

What state or EPA region do you represent? Select one only.		
Answer Options	**Response Percent**	**Response Count**
ALABAMA	1.2%	1
ALASKA	4.8%	4
ARIZONA	1.2%	1
ARKANSAS	0.0%	0
CALIFORNIA	9.6%	8
COLORADO	1.2%	1
CONNECTICUT	0.0%	0
DELAWARE	2.4%	2
DISTRICT OF COLUMBIA	0.0%	0
FLORIDA	12.0%	10
GEORGIA	1.2%	1
HAWAII	3.6%	3
IDAHO	0.0%	0
ILLINOIS	1.2%	1
INDIANA	1.2%	1
IOWA	1.2%	1
KANSAS	0.0%	0
KENTUCKY	1.2%	1
LOUISIANA	0.0%	0
MAINE	1.2%	1
MARYLAND	0.0%	0
MASSACHUSETTS	0.0%	0
MICHIGAN	12.0%	10
MINNESOTA	0.0%	0
MISSISSIPPI	0.0%	0
MISSOURI	6.0%	5
MONTANA	0.0%	0
NEBRASKA	1.2%	1
NEVADA	1.2%	1
NEW HAMPSHIRE	0.0%	0
NEW JERSEY	1.2%	1
NEW MEXICO	0.0%	0
NEW YORK	1.2%	1
NORTH CAROLINA	1.2%	1
NORTH DAKOTA	0.0%	0
OHIO	2.4%	2
OKLAHOMA	0.0%	0
OREGON	4.8%	4
PENNSYLVANIA	1.2%	1
RHODE ISLAND	1.2%	1
SOUTH CAROLINA	0.0%	0
SOUTH DAKOTA	1.2%	1

Answer Options	Response Percent	Response Count
TENNESSEE	0.0%	0
TEXAS	4.8%	4
UTAH	1.2%	1
VERMONT	0.0%	0
VIRGINIA	0.0%	0
WASHINGTON	1.2%	1
WEST VIRGINIA	0.0%	0
WISCONSIN	1.2%	1
WYOMING	1.2%	1
Region 1	0.0%	0
Region 2	0.0%	0
Region 3	1.2%	1
Region 4	8.4%	7
Region 5	0.0%	0
Region 6	0.0%	0
Region 7	0.0%	0
Region 8	0.0%	0
Region 9	1.2%	1
Region 10	1.2%	1
answered question		**83**
skipped question		**179**

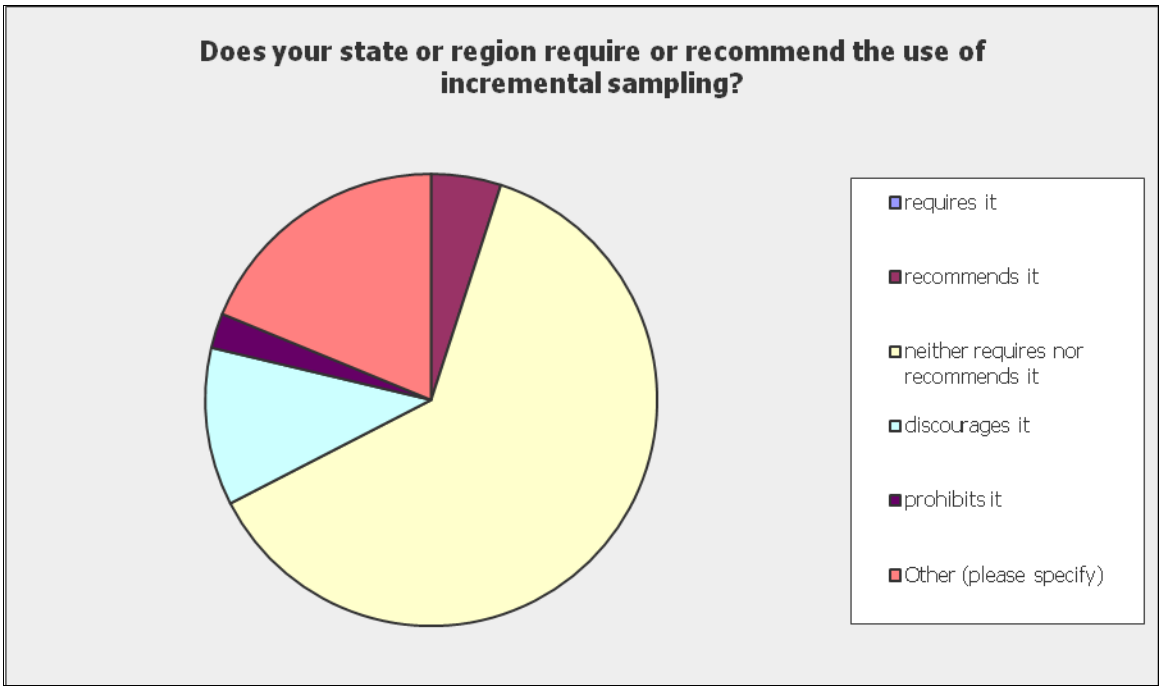
What state or EPA region do you represent? Select one only.



- ALABAMA
- ALASKA
- ARIZONA
- ARKANSAS
- CALIFORNIA
- COLORADO
- CONNECTICUT
- DELAWARE
- DISTRICT OF COLUMBIA
- FLORIDA
- GEORGIA
- HAWAII
- IDAHO
- ILLINOIS
- INDIANA
- IOWA
- KANSAS
- KENTUCKY
- LOUISIANA
- MAINE
- MARYLAND
- MASSACHUSETTS
- MICHIGAN
- MINNESOTA
- MISSISSIPPI
- MISSOURI
- MONTANA
- NEBRASKA
- NEVADA
- NEW HAMPSHIRE
- NEW JERSEY
- NEW MEXICO
- NEW YORK
- NORTH CAROLINA
- NORTH DAKOTA
- OHIO
- OKLAHOMA
- OREGON
- PENNSYLVANIA
- RHODE ISLAND
- SOUTH CAROLINA
- SOUTH DAKOTA
- Answer Options
- TENNESSEE
- TEXAS
- UTAH
- VERMONT
- VIRGINIA
- WASHINGTON
- WEST VIRGINIA
- WISCONSIN
- WYOMING
- Region 1
- Region 2
- Region 3
- Region 4
- Region 5
- Region 6
- Region 7
- Region 8
- Region 9
- Region 10

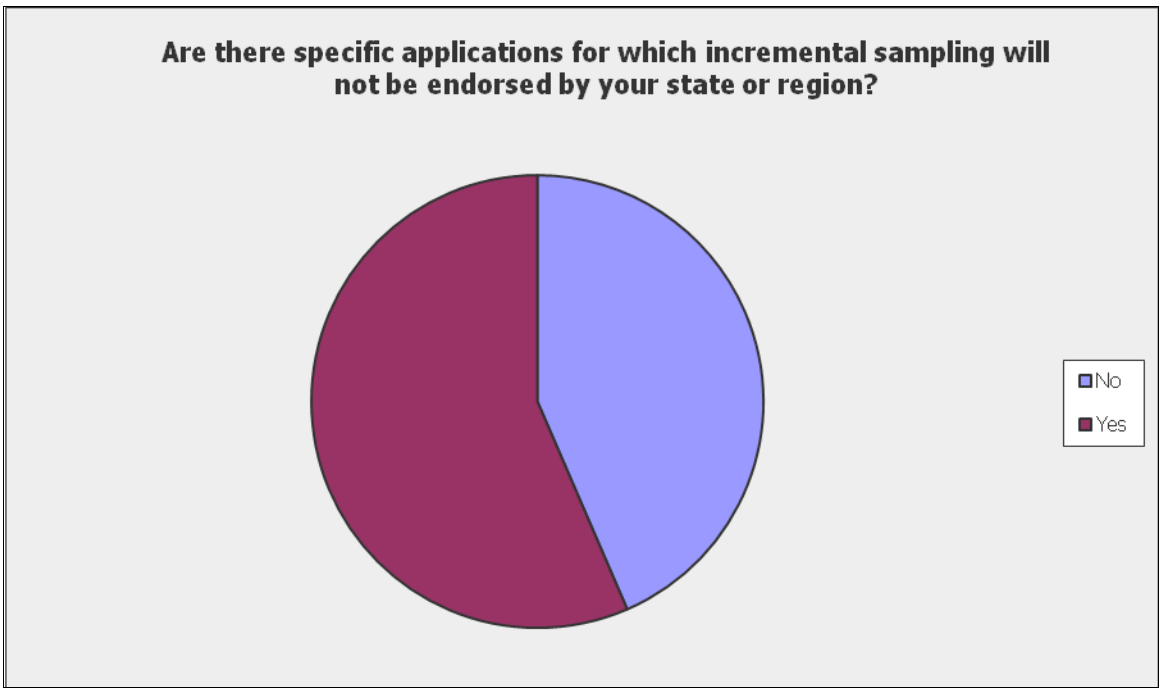
Question 39.

Does your state or region require or recommend the use of incremental sampling?		
Answer Options	**Response Percent**	**Response Count**
requires it	0.0%	0
recommends it	5.0%	4
neither requires nor recommends it	62.5%	50
discourages it	11.3%	9
prohibits it	2.5%	2
Other (please specify)	18.8%	15
answered question		**80**
skipped question		**182**



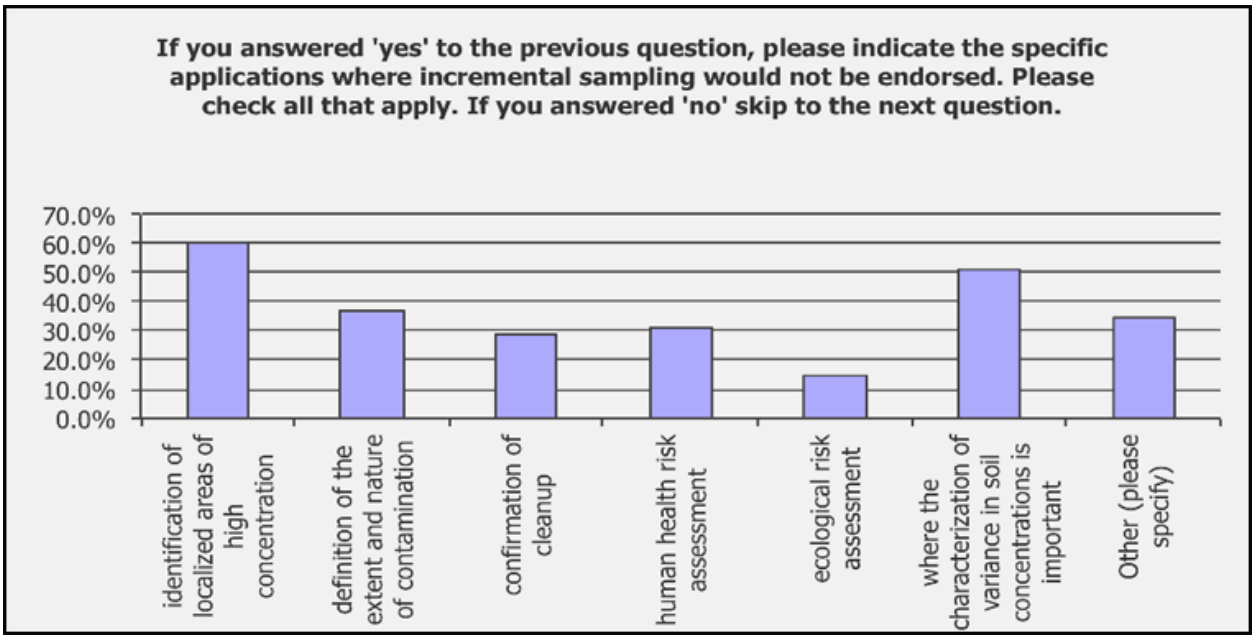
Question 40.

Are there specific applications for which incremental sampling will not be endorsed by your state or region?		
Answer Options	**Response Percent**	**Response Count**
No	43.5%	27
Yes	56.5%	35
answered question		**62**
skipped question		**200**



Question 41.

If you answered 'yes' to the previous question, please indicate the specific applications where incremental sampling would not be endorsed. Please check all that apply. If you answered 'no' skip to the next question.		
Answer Options	**Response Percent**	**Response Count**
identification of localized areas of high concentration	60.0%	21
definition of the extent and nature of contamination	37.1%	13
confirmation of cleanup	28.6%	10
human health risk assessment	31.4%	11
ecological risk assessment	14.3%	5
where the characterization of variance in soil concentrations is important	51.4%	18
Other (please specify)	34.3%	12
	answered question	**35**
	skipped question	**227**

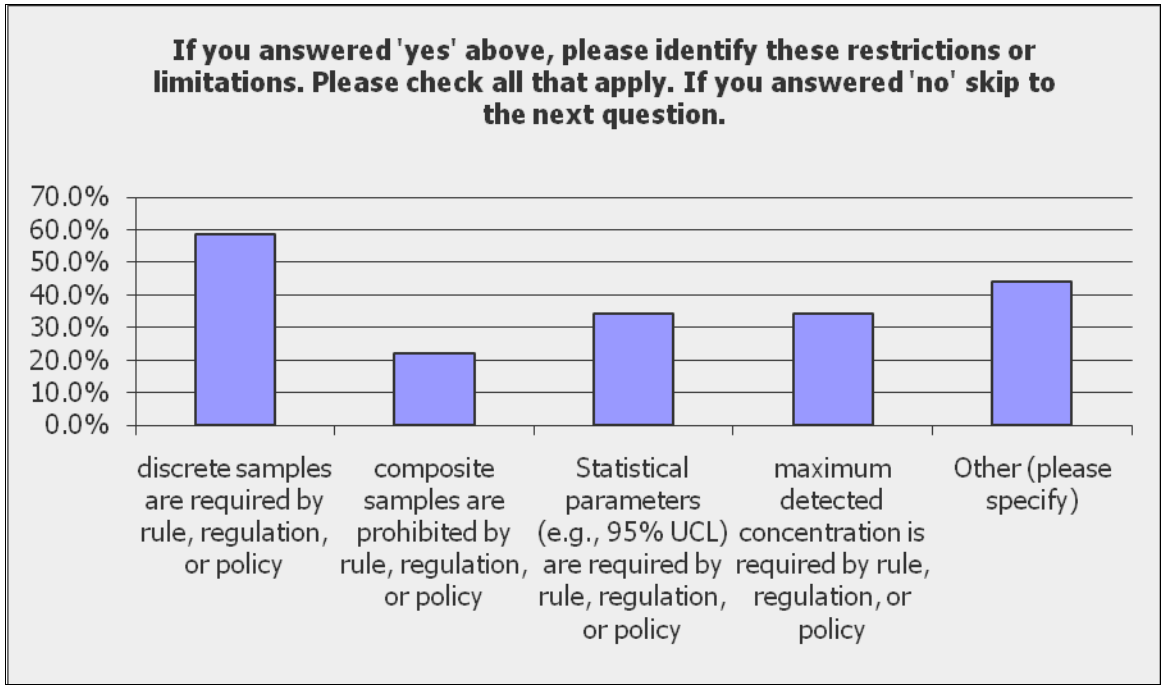


Question 42.

Based on your experience, are there restrictions or limitations that prevent using incremental sampling in your state or region?		
Answer Options	**Response Percent**	**Response Count**
Yes	56.5%	39
No	43.5%	30
answered question		**69**
skipped question		**193**

Question 43.

If you answered 'yes' above, please identify these restrictions or limitations. Please check all that apply. If you answered 'no' skip to the next question.		
Answer Options	**Response Percent**	**Response Count**
discrete samples are required by rule, regulation, or policy	58.5%	24
composite samples are prohibited by rule, regulation, or policy	22.0%	9
Statistical parameters (e.g., 95% UCL) are required by rule, regulation, or policy	34.1%	14
maximum detected concentration is required by rule, regulation, or policy	34.1%	14
Other (please specify)	43.9%	18
answered question		**41**
skipped question		**221**

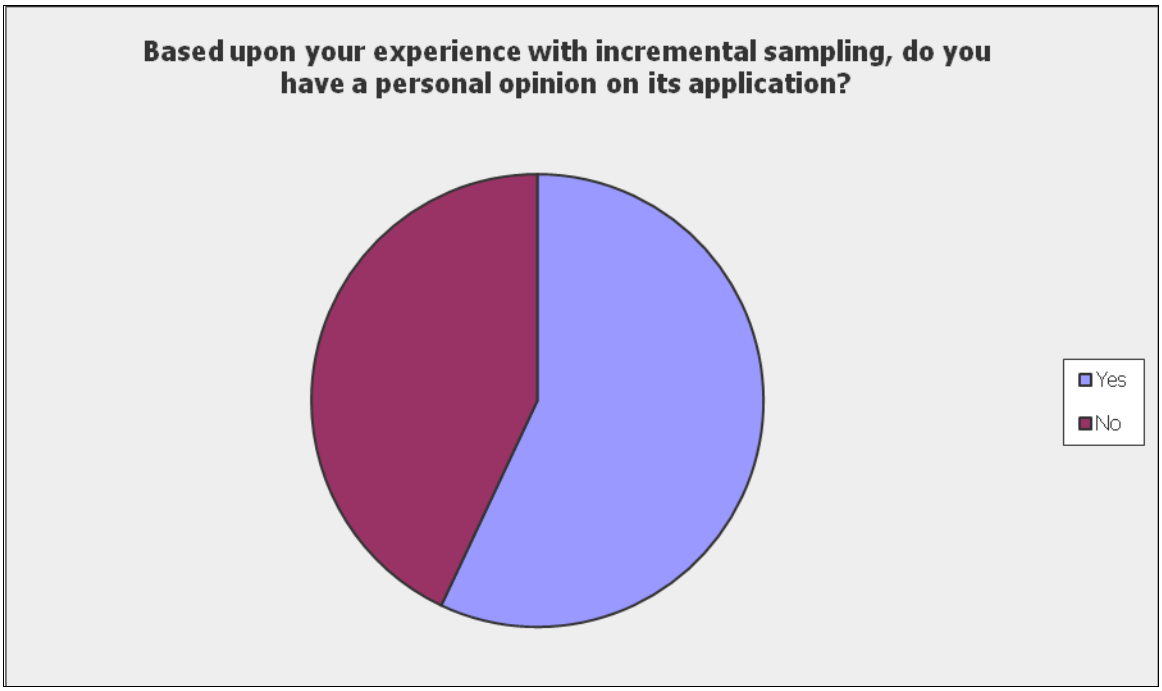


Question 44.

If your state or agency has specific codes, regulations, or rules that restrict or limit use of ISM data please provide the specific citations. If your agency has specific policies or guidance, please provide a URL link to the policy or guidance. If you would like to provide this information at a later time, please send it to: rick.galloway@state.de.us.	
Answer Options	**Response Count**
	10
answered question	**10**
skipped question	**252**

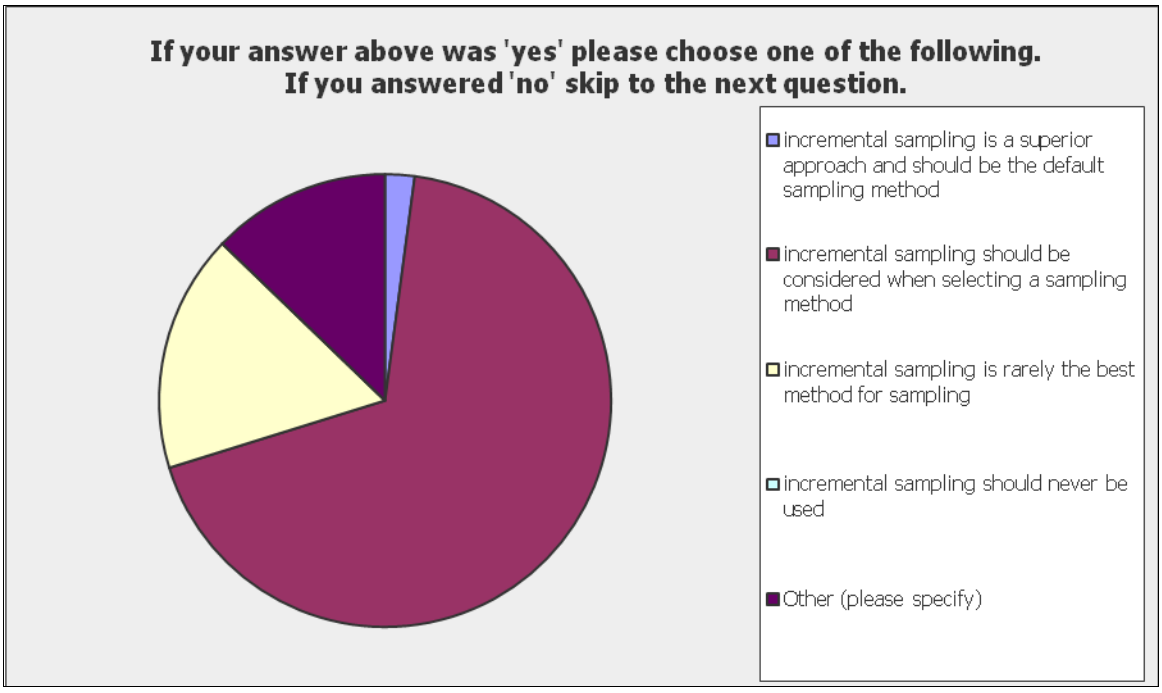
Question 45.

Based upon your experience with incremental sampling, do you have a personal opinion on its application?		
Answer Options	**Response Percent**	**Response Count**
Yes	57.0%	45
No	43.0%	34
answered question		**79**
skipped question		**183**



Question 46.

If your answer above was 'yes' please choose one of the following. If you answered 'no' skip to the next question.		
Answer Options	**Response Percent**	**Response Count**
incremental sampling is a superior approach and should be the default sampling method	2.1%	1
incremental sampling should be considered when selecting a sampling method	68.1%	32
incremental sampling is rarely the best method for sampling	17.0%	8
incremental sampling should never be used	0.0%	0
Other (please specify)	12.8%	6
answered question		**47**
skipped question		**215**



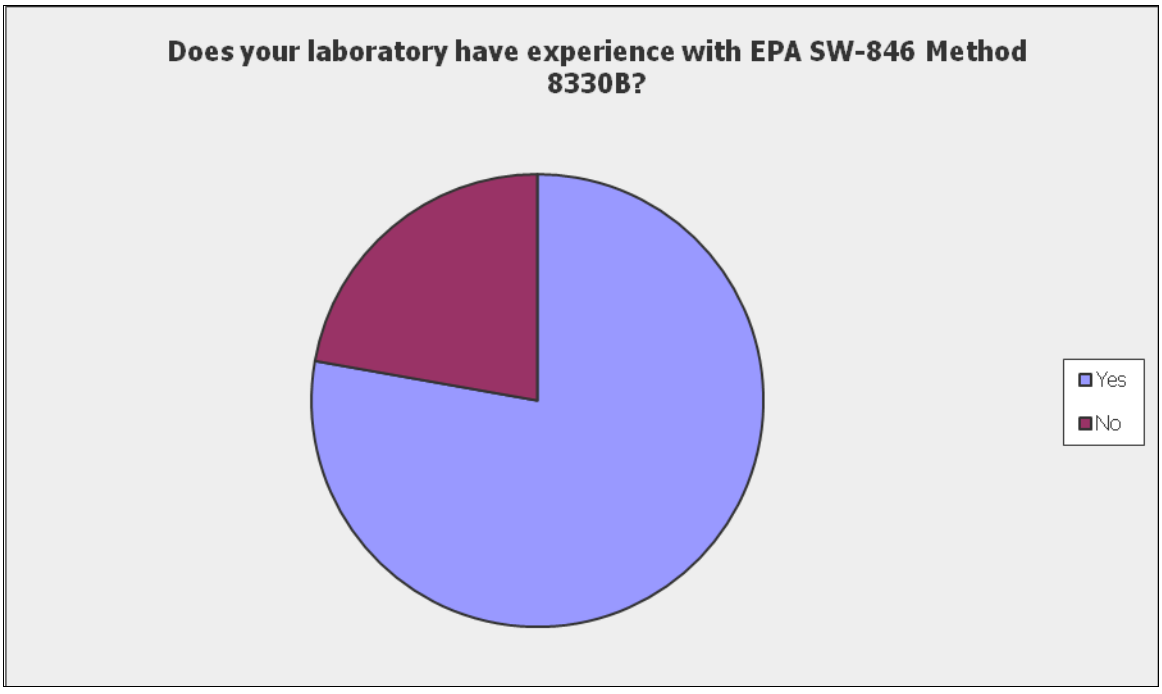
Question 47.

What states do you serve? Please check all that apply.		
Answer Options	**Response Percent**	**Response Count**
ENTIRE United States	30.8%	8
ALABAMA	11.5%	3
ALASKA	7.7%	2
ARIZONA	3.8%	1
ARKANSAS	3.8%	1
CALIFORNIA	23.1%	6
COLORADO	3.8%	1
CONNECTICUT	0.0%	0
DELAWARE	0.0%	0
DISTRICT OF COLUMBIA	0.0%	0
FLORIDA	15.4%	4
GEORGIA	7.7%	2
HAWAII	23.1%	6
IDAHO	3.8%	1
ILLINOIS	0.0%	0
INDIANA	0.0%	0
IOWA	0.0%	0
KANSAS	0.0%	0
KENTUCKY	7.7%	2
LOUISIANA	3.8%	1
MAINE	0.0%	0
MARYLAND	0.0%	0
MASSACHUSETTS	0.0%	0
MICHIGAN	3.8%	1
MINNESOTA	0.0%	0
MISSISSIPPI	7.7%	2
MISSOURI	0.0%	0
MONTANA	0.0%	0
NEBRASKA	3.8%	1
NEVADA	7.7%	2
NEW HAMPSHIRE	0.0%	0
NEW JERSEY	0.0%	0
NEW MEXICO	0.0%	0
NEW YORK	0.0%	0
NORTH CAROLINA	11.5%	3
NORTH DAKOTA	0.0%	0
OHIO	0.0%	0
OKLAHOMA	3.8%	1
OREGON	3.8%	1
PENNSYLVANIA	0.0%	0
RHODE ISLAND	0.0%	0
SOUTH CAROLINA	7.7%	2
SOUTH DAKOTA	0.0%	0
TENNESSEE	7.7%	2

TEXAS	7.7%	2
UTAH	3.8%	1
VERMONT	0.0%	0
VIRGINIA	0.0%	0
WASHINGTON	0.0%	0
WEST VIRGINIA	0.0%	0
WISCONSIN	3.8%	1
WYOMING	3.8%	1
	answered question	**26**
	skipped question	**236**

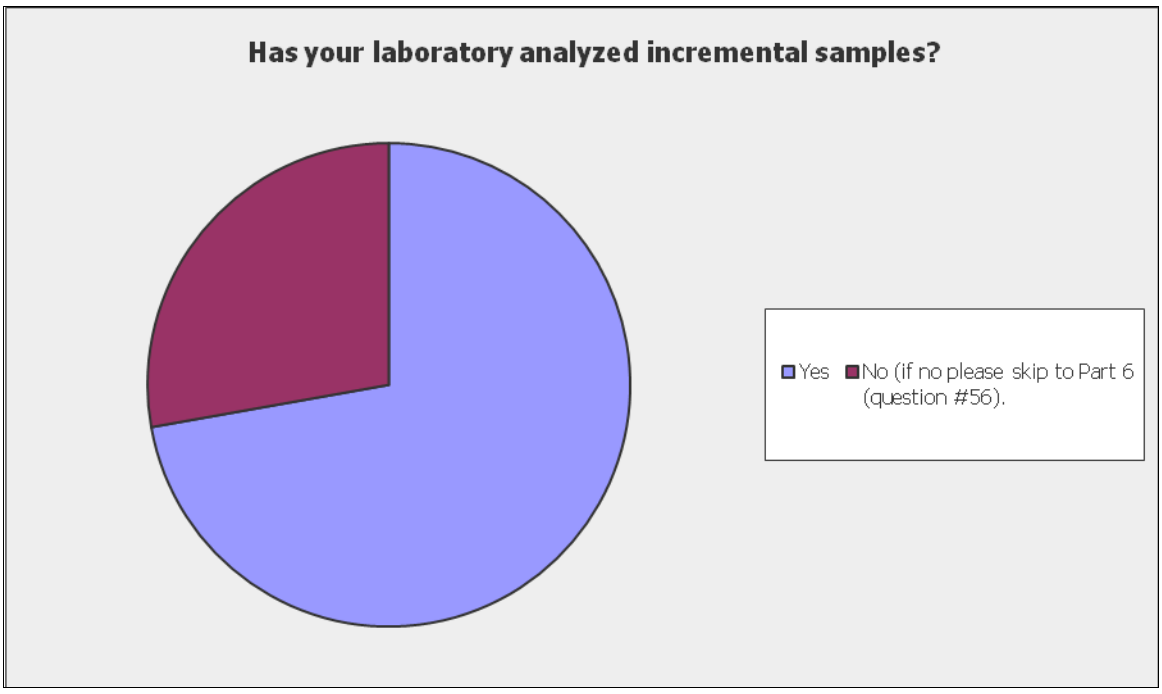
Question 48.

Does your laboratory have experience with EPA SW-846 Method 8330B?		
Answer Options	**Response Percent**	**Response Count**
Yes	77.8%	14
No	22.2%	4
	answered question	**18**
	skipped question	**244**



Question 49.

Has your laboratory analyzed incremental samples?		
Answer Options	**Response Percent**	**Response Count**
Yes	72.2%	13
No (if no please skip to Part 6 (question #56).	27.8%	5
answered question		**18**
skipped question		**244**

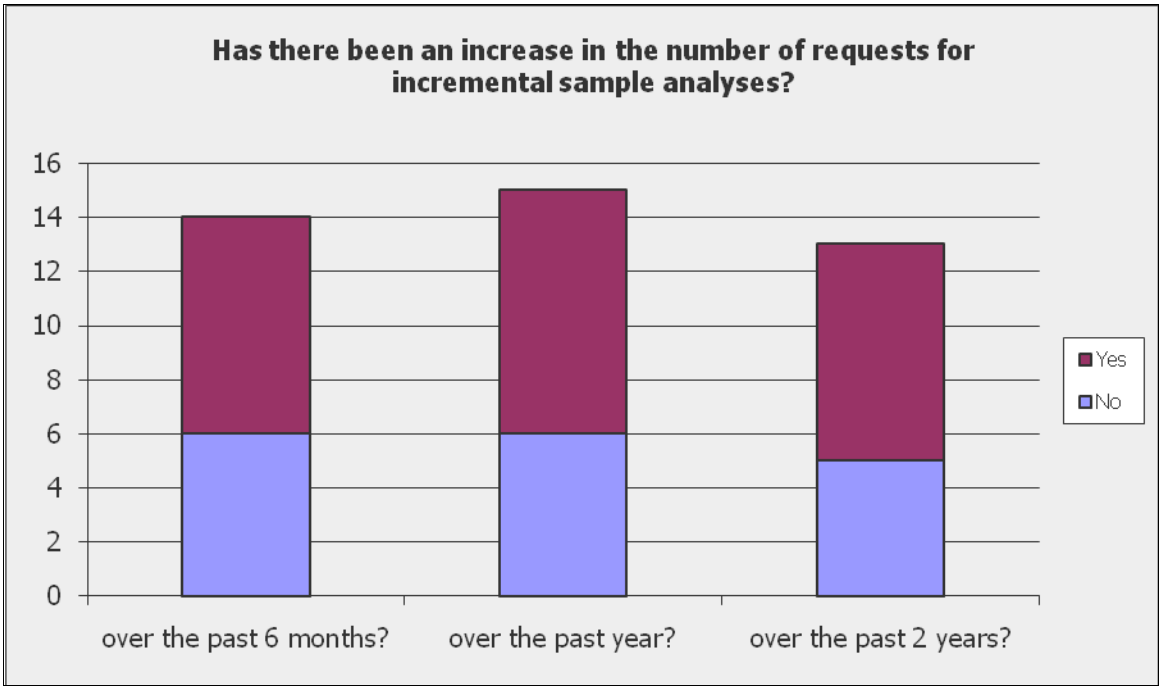


Question 50.

For approximately how many incremental sampling projects have you provided services?	
Answer Options	**Response Count**
	11
answered question	**11**
skipped question	**251**

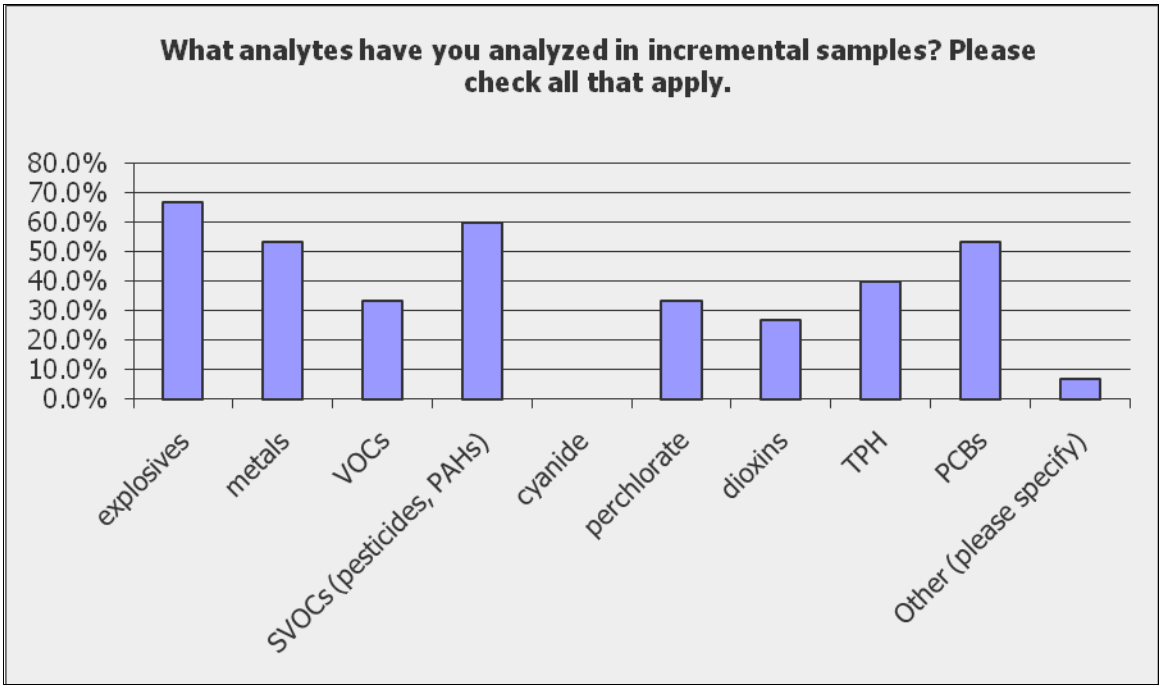
Question 51.

Has there been an increase in the number of requests for incremental sample analyses?			
Answer Options	**Yes**	**No**	**Response Count**
over the past 6 months?	8	6	14
over the past year?	9	6	15
over the past 2 years?	8	5	13
		answered question	**15**
		skipped question	**247**



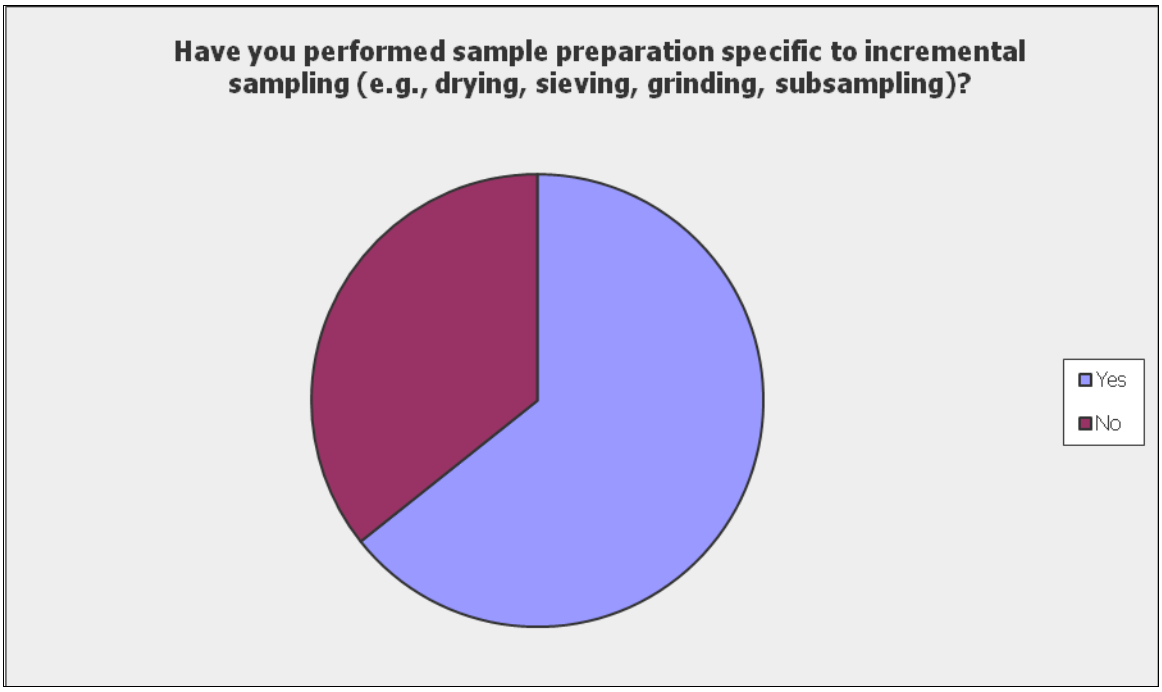
Question 52.

What analytes have you analyzed in incremental samples? Please check all that apply.		
Answer Options	**Response Percent**	**Response Count**
explosives	66.7%	10
metals	53.3%	8
VOCs	33.3%	5
SVOCs (pesticides, PAHs)	60.0%	9
cyanide	0.0%	0
perchlorate	33.3%	5
dioxins	26.7%	4
TPH	40.0%	6
PCBs	53.3%	8
Other (please specify)	6.7%	1
answered question		**15**
skipped question		**247**



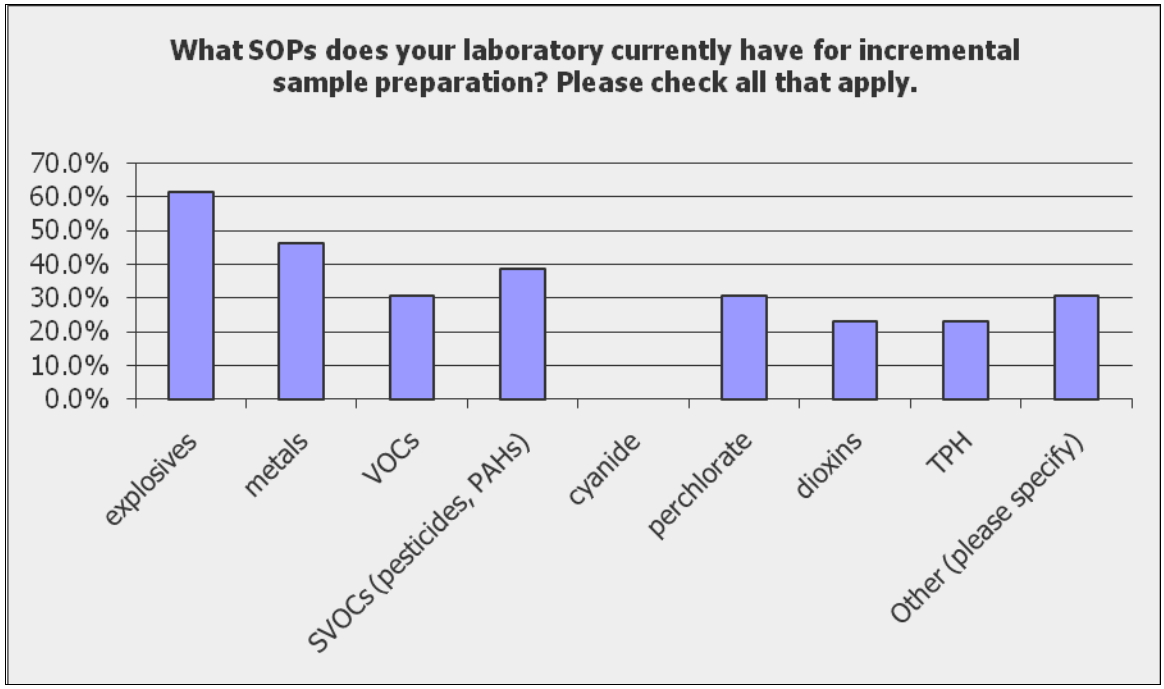
Question 53.

Have you performed sample preparation specific to incremental sampling (e.g., drying, sieving, grinding, subsampling)?		
Answer Options	**Response Percent**	**Response Count**
Yes	64.3%	9
No	35.7%	5
answered question		**14**
skipped question		**248**



Question 54.

What SOPs does your laboratory currently have for incremental sample preparation? Please check all that apply.		
Answer Options	**Response Percent**	**Response Count**
explosives	61.5%	8
metals	46.2%	6
VOCs	30.8%	4
SVOCs (pesticides, PAHs)	38.5%	5
cyanide	0.0%	0
perchlorate	30.8%	4
dioxins	23.1%	3
TPH	23.1%	3
Other (please specify)	30.8%	4
	answered question	**13**
	skipped question	**249**

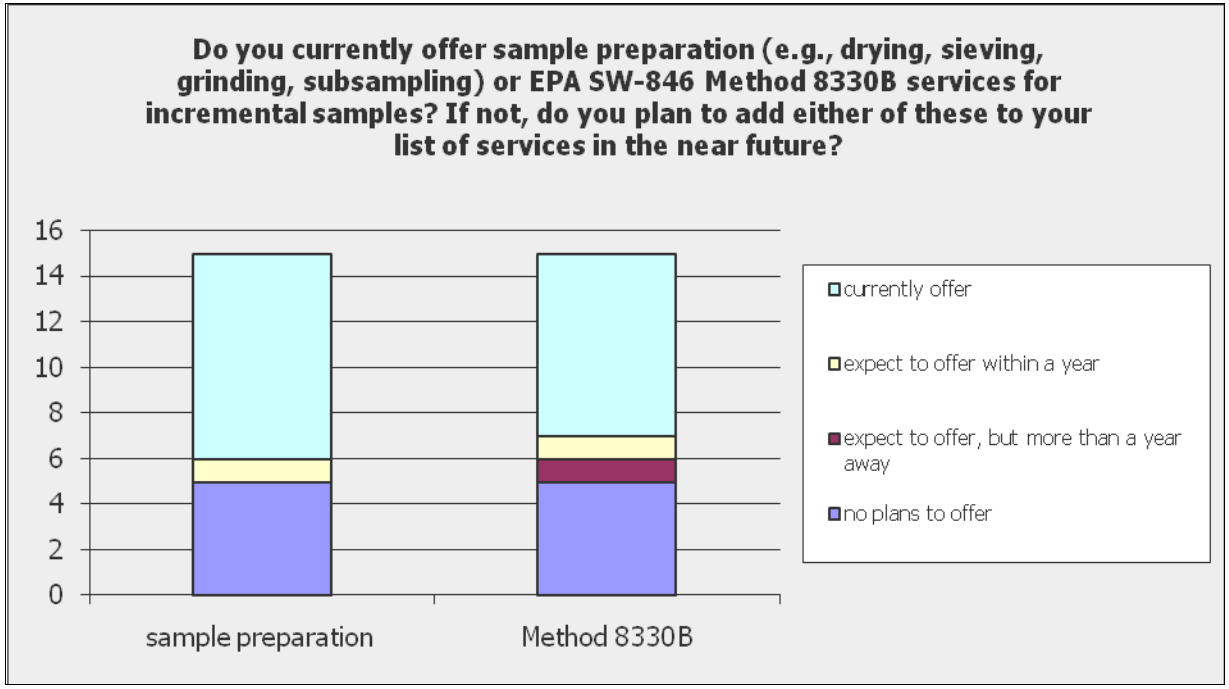


Question 55.

Are you aware of any laboratory certification specific to incremental sampling methodology?		
Answer Options	**Response Percent**	**Response Count**
Yes	42.9%	6
No	57.1%	8
	answered question	**14**
	skipped question	**248**

Question 56.

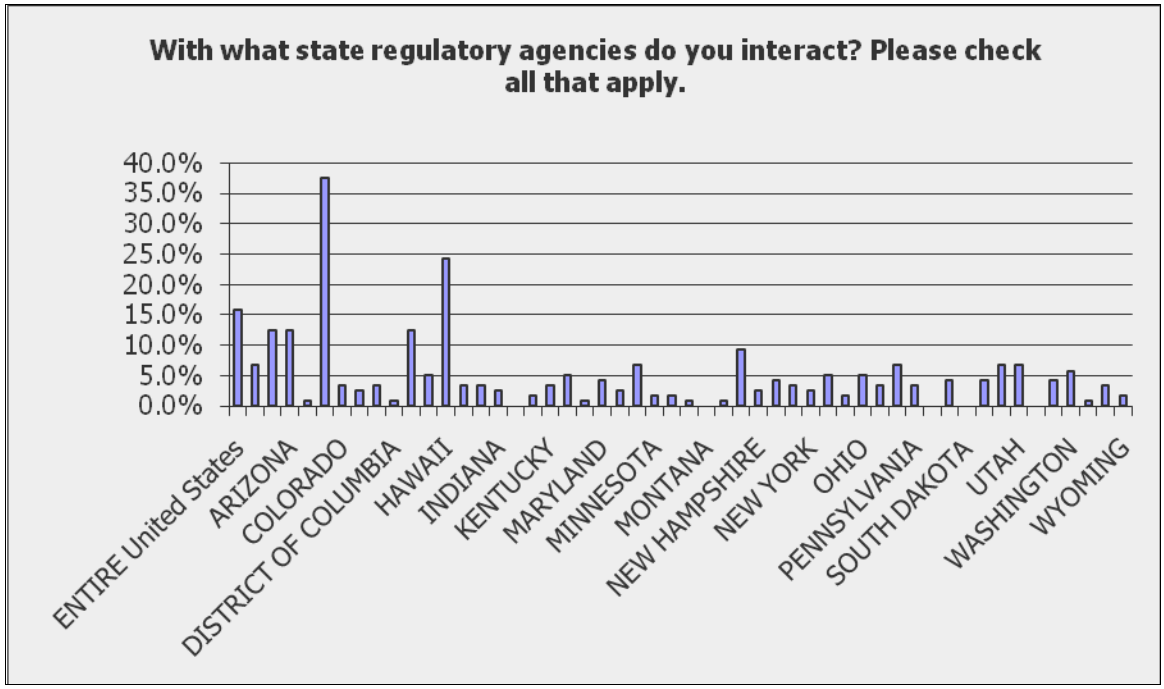
Do you currently offer sample preparation (e.g., drying, sieving, grinding, subsampling) or EPA SW-846 Method 8330B services for incremental samples? If not, do you plan to add either of these to your list of services in the near future?					
Answer Options	**currently offer**	**expect to offer within a year**	**expect to offer, but more than a year away**	**no plans to offer**	**Response Count**
sample preparation	9	1	0	5	15
Method 8330B	8	1	1	5	15
				answered question	**16**
				skipped question	**246**



Question 57.

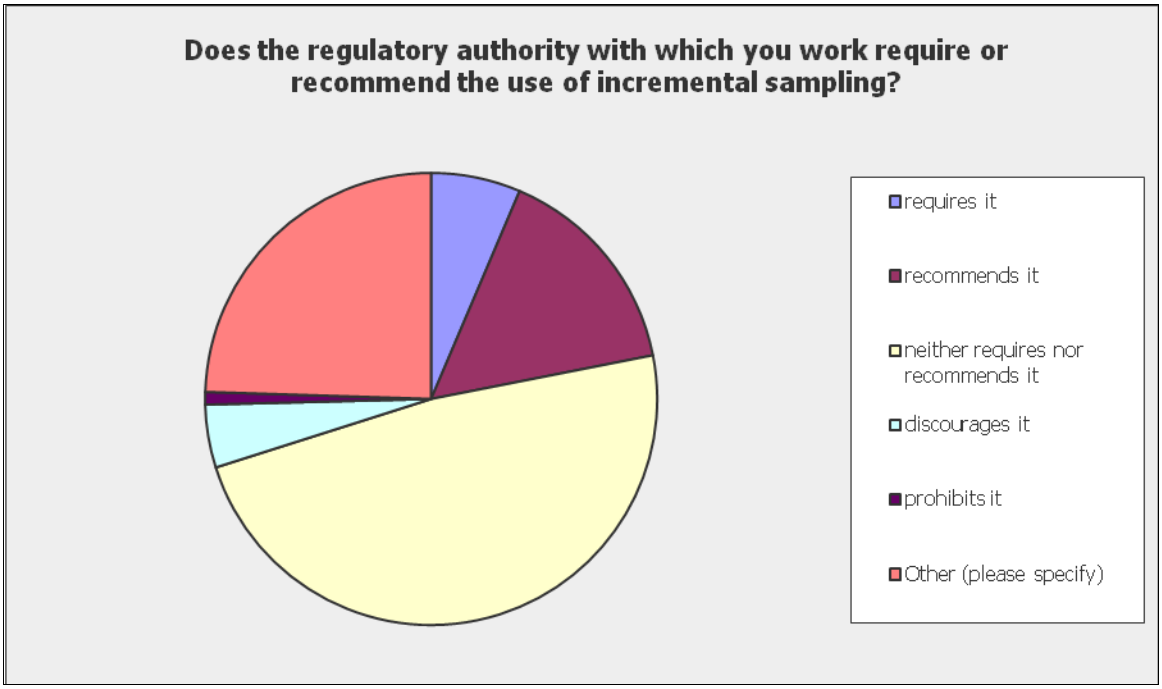
With what state regulatory agencies do you interact? Please check all that apply.		
Answer Options	**Response Percent**	**Response Count**
ENTIRE United States	15.8%	19
ALABAMA	6.7%	8
ALASKA	12.5%	15
ARIZONA	12.5%	15
ARKANSAS	0.8%	1
CALIFORNIA	37.5%	45
COLORADO	3.3%	4
CONNECTICUT	2.5%	3
DELAWARE	3.3%	4
DISTRICT OF COLUMBIA	0.8%	1
FLORIDA	12.5%	15
GEORGIA	5.0%	6
HAWAII	24.2%	29
IDAHO	3.3%	4
ILLINOIS	3.3%	4
INDIANA	2.5%	3
IOWA	0.0%	0
KANSAS	1.7%	2
KENTUCKY	3.3%	4
LOUISIANA	5.0%	6
MAINE	0.8%	1
MARYLAND	4.2%	5
MASSACHUSETTS	2.5%	3
MICHIGAN	6.7%	8
MINNESOTA	1.7%	2
MISSISSIPPI	1.7%	2
MISSOURI	0.8%	1
MONTANA	0.0%	0
NEBRASKA	0.8%	1
NEVADA	9.2%	11
NEW HAMPSHIRE	2.5%	3
NEW JERSEY	4.2%	5
NEW MEXICO	3.3%	4
NEW YORK	2.5%	3
NORTH CAROLINA	5.0%	6
NORTH DAKOTA	1.7%	2
OHIO	5.0%	6
OKLAHOMA	3.3%	4
OREGON	6.7%	8
PENNSYLVANIA	3.3%	4
RHODE ISLAND	0.0%	0
SOUTH CAROLINA	4.2%	5
SOUTH DAKOTA	0.0%	0
`TENNESSEE	4.2%	5

TEXAS	6.7%	8
UTAH	6.7%	8
VERMONT	0.0%	0
VIRGINIA	4.2%	5
WASHINGTON	5.8%	7
WEST VIRGINIA	0.8%	1
WISCONSIN	3.3%	4
WYOMING	1.7%	2
	answered question	**120**
	skipped question	**142**



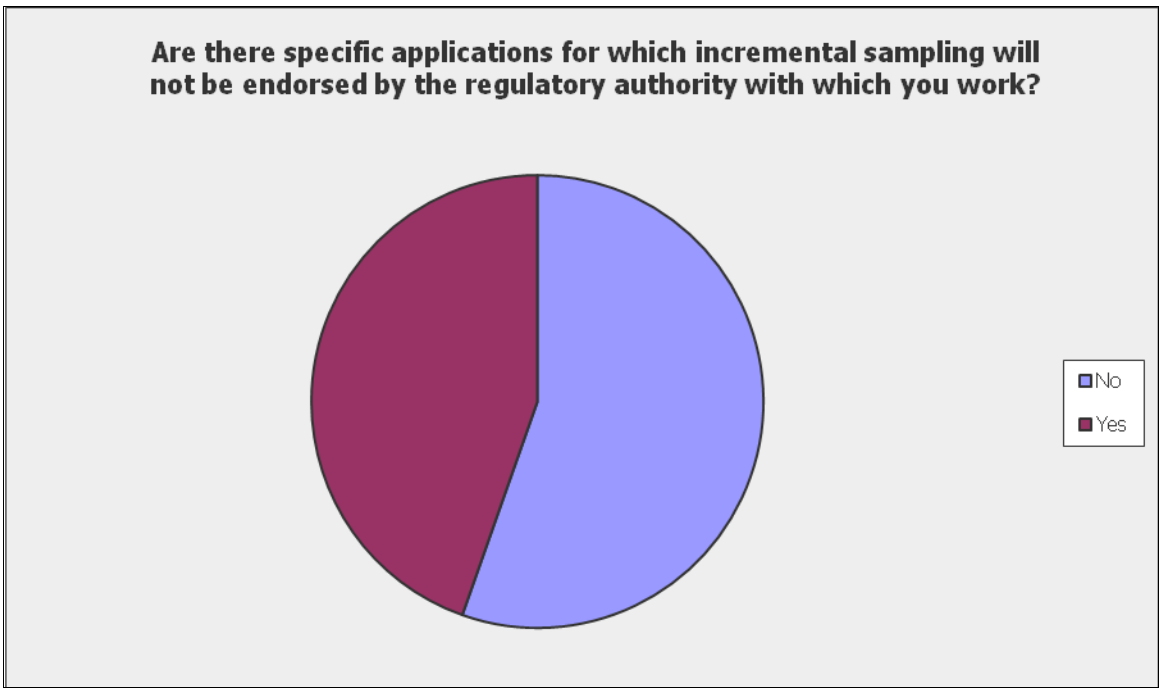
Question 58.

Does the regulatory authority with which you work require or recommend the use of incremental sampling?		
Answer Options	**Response Percent**	**Response Count**
requires it	6.4%	7
recommends it	15.5%	17
neither requires nor recommends it	48.2%	53
discourages it	4.5%	5
prohibits it	0.9%	1
Other (please specify)	24.5%	27
answered question		**110**
skipped question		**152**



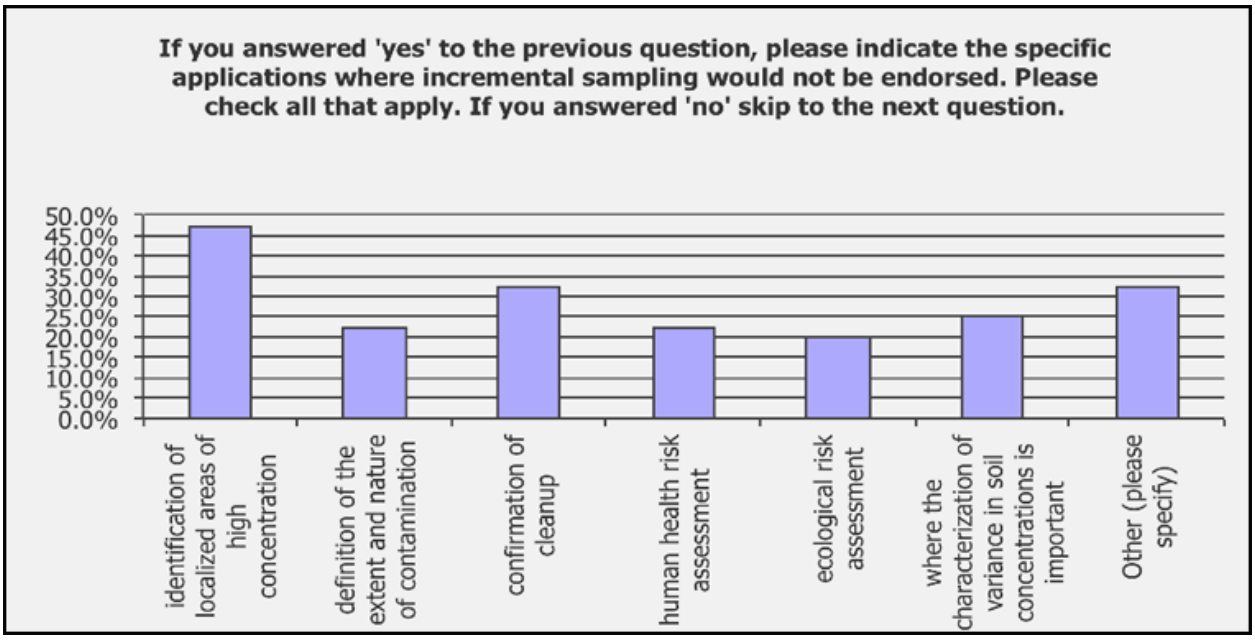
Question 59.

Are there specific applications for which incremental sampling will not be endorsed by the regulatory authority with which you work?		
Answer Options	**Response Percent**	**Response Count**
No	55.4%	46
Yes	44.6%	37
answered question		**83**
skipped question		**179**



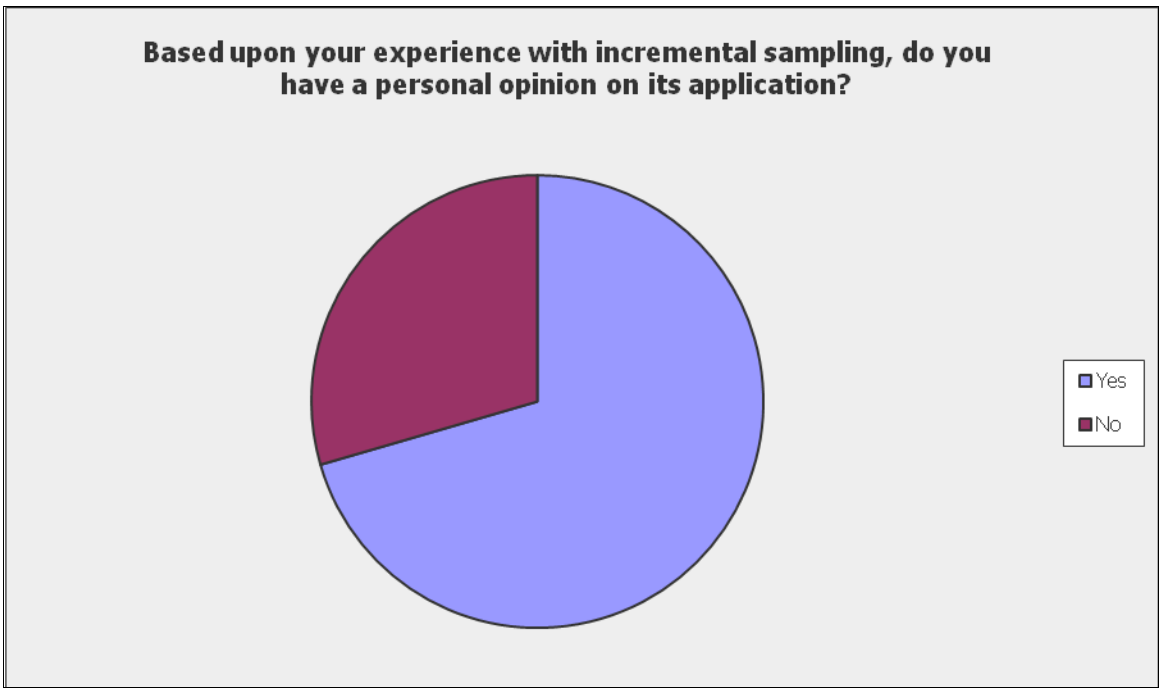
Question 60.

If you answered 'yes' to the previous question, please indicate the specific applications where incremental sampling would not be endorsed. Please check all that apply. If you answered 'no' skip to the next question.		
Answer Options	**Response Percent**	**Response Count**
identification of localized areas of high concentration	47.5%	19
definition of the extent and nature of contamination	22.5%	9
confirmation of cleanup	32.5%	13
human health risk assessment	22.5%	9
ecological risk assessment	20.0%	8
where the characterization of variance in soil concentrations is important	25.0%	10
Other (please specify)	32.5%	13
answered question		**40**
skipped question		**222**



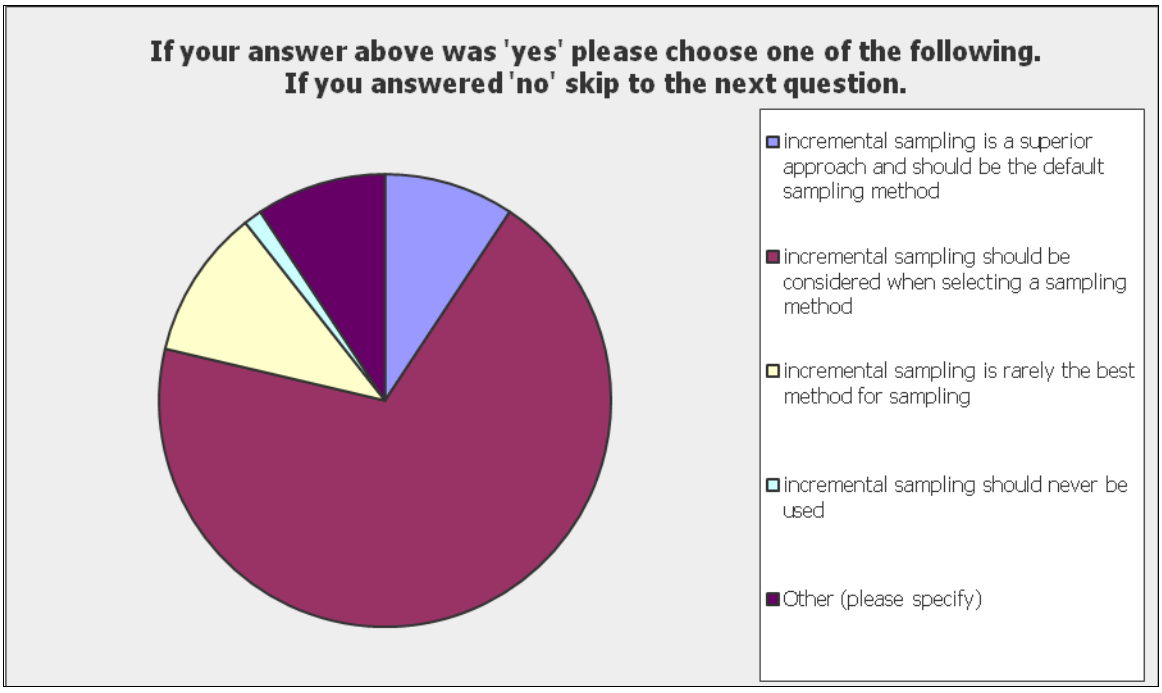
Question 61.

Based upon your experience with incremental sampling, do you have a personal opinion on its application?		
Answer Options	**Response Percent**	**Response Count**
Yes	70.5%	74
No	29.5%	31
answered question		**105**
skipped question		**157**



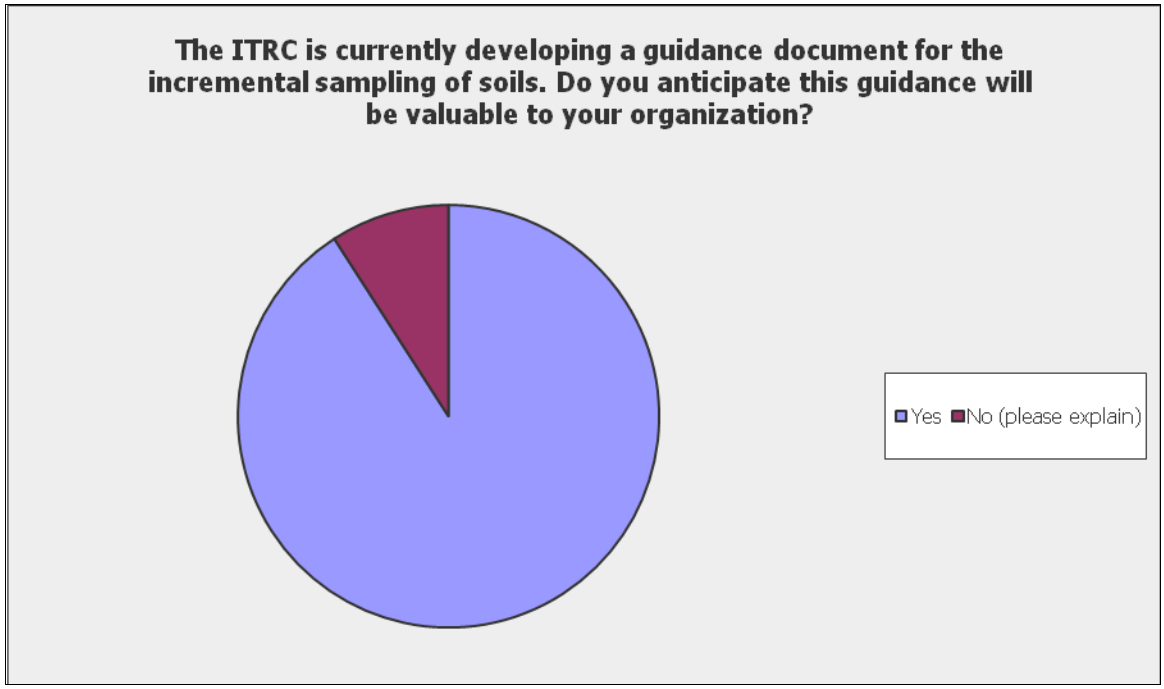
Question 62.

If your answer above was 'yes' please choose one of the following. If you answered 'no' skip to the next question.		
Answer Options	**Response Percent**	**Response Count**
incremental sampling is a superior approach and should be the default sampling method	9.3%	7
incremental sampling should be considered when selecting a sampling method	69.3%	52
incremental sampling is rarely the best method for sampling	10.7%	8
incremental sampling should never be used	1.3%	1
Other (please specify)	9.3%	7
answered question		**75**
skipped question		**187**



Question 63.

The ITRC is currently developing a guidance document for the incremental sampling of soils. Do you anticipate this guidance will be valuable to your organization?		
Answer Options	**Response Percent**	**Response Count**
Yes	90.9%	169
No (please explain)	9.1%	17
	answered question	**186**
	skipped question	**76**



Question 64.

The ITRC document under development will address when to use incremental sampling and when other sampling methods are more appropriate; and how to define decision units. If there are other topics you would like to see discussed, please list them below.	
Answer Options	**Response Count**
	57
answered question	**57**
skipped question	**205**

Question 65.

If you have any comments on this survey in general, or on specific questions, please write them here. If you desire, you may also submit comments via e-mail to rick.galloway@state.de.us.	
Answer Options	**Response Count**
	17
answered question	**17**
skipped question	**245**

Appendix C

Case Studies

CASE STUDIES

C.1 CASE STUDY 1: KURE ATOLL, HAWAII

Site Name: Green Island Landfill and Reburial Pit, Kure Atoll, Hawaii

Contact Name: Roger Brewer, HDOH

Site Location: Kure Atoll is the northernmost island in the Hawaiian Island chain, located approximately 1400 miles northwest of the island Oahu and 56 miles northwest of Midway atoll. The atoll consists of a lagoon encircled by a reef and a single vegetated island, Green Island. Green Island is just under 1.5 miles long and about 0.35 miles wide and has a maximum elevation of 15 feet.

C.1.1 Background and Previous Investigations

This case study summarizes the investigation of a former landfill site on Kure atoll, a remote island in the central Pacific Ocean. A detailed discussion of the investigation is presented in the report *Evaluation of Green Island Landfill and Reburial Pit, Former U.S. Coast Guard LORAN Station Kure* (USCG 2009). A copy of the report is available from the HDOH Hazard Evaluation and Emergency Response Office.

Kure atoll is the northernmost island in the Hawaiian island chain, located approximately 1400 miles northwest of the island of Oahu and 56 miles northwest of Midway atoll (see Figure C.1-1). The atoll consists of a lagoon encircled by a reef and a single, vegetated island (Green Island, Figure C.1-2). Green Island is just under 1.5 miles long and about 0.35 miles wide and has a maximum elevation of around 15 feet. The island is not inhabited on a permanent basis although it is visited periodically by marine research groups.

A USCG station was located on the atoll from the 1960s through the 1990s. When the station was operating, a small, approximately ½-half acre area on the southwest corner of the island was used to dispose of old electrical components and scrap metal (e.g., capacitors, batteries, and transformers, see Figure C.1-2). Debris and approximately 700 yd^3 of PCB-contaminated soil were removed from the site in 1993. Discrete, confirmation soil samples identified concentrations of PCBs as high as 170 mg/kg within the former landfill footprint (see Figure C.1-3). Soil, sediment, and biota samples collected in the surrounding area indicated that PCB contamination was primarily restricted to the landfill site.

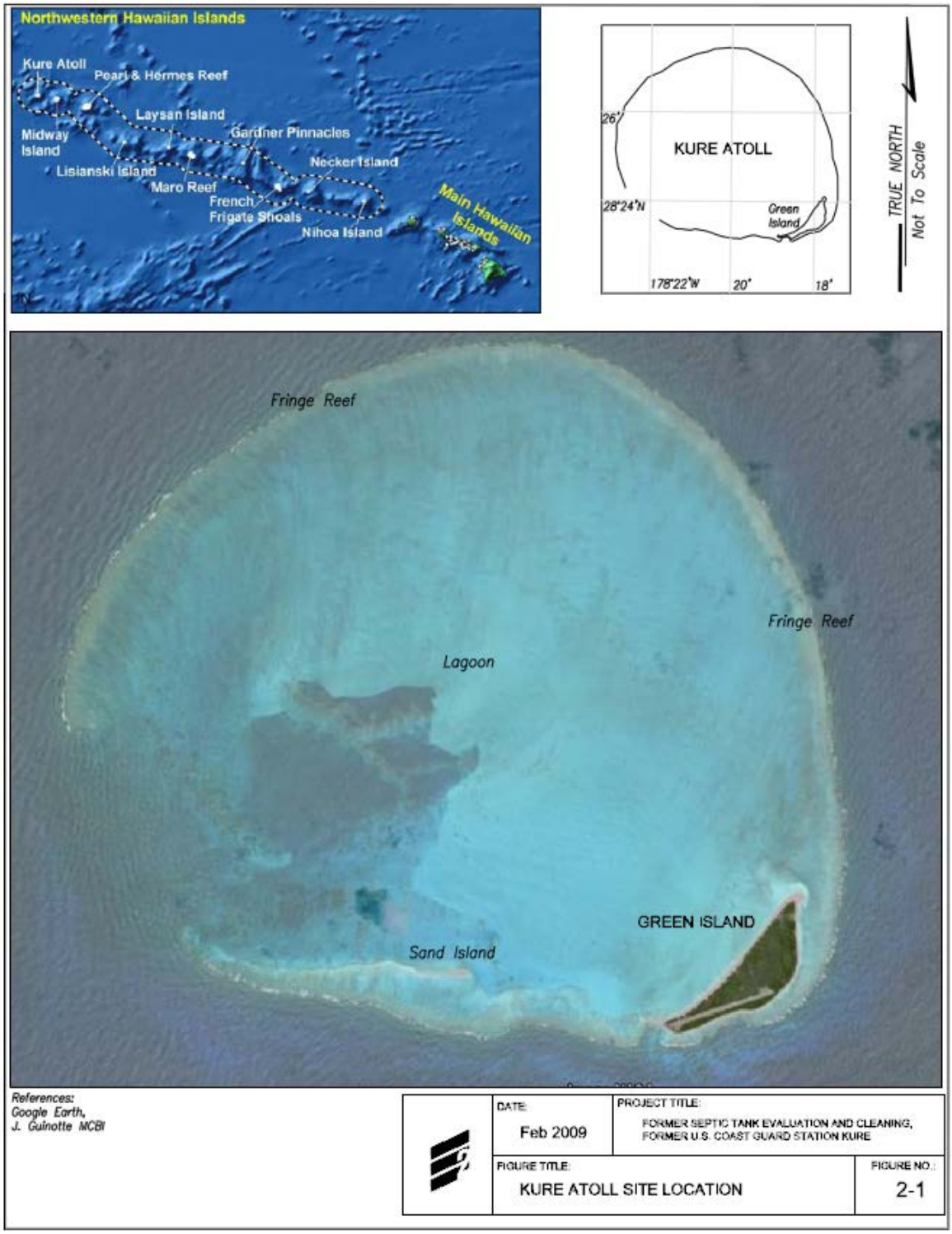


Figure C.1-1. Kure Atoll location map. *Source*: USCG 2009, Figure 2-1.

Figure C.1-2. Green Island map showing location of former landfill area.
Source: USCG 2009, Figure 2-2.

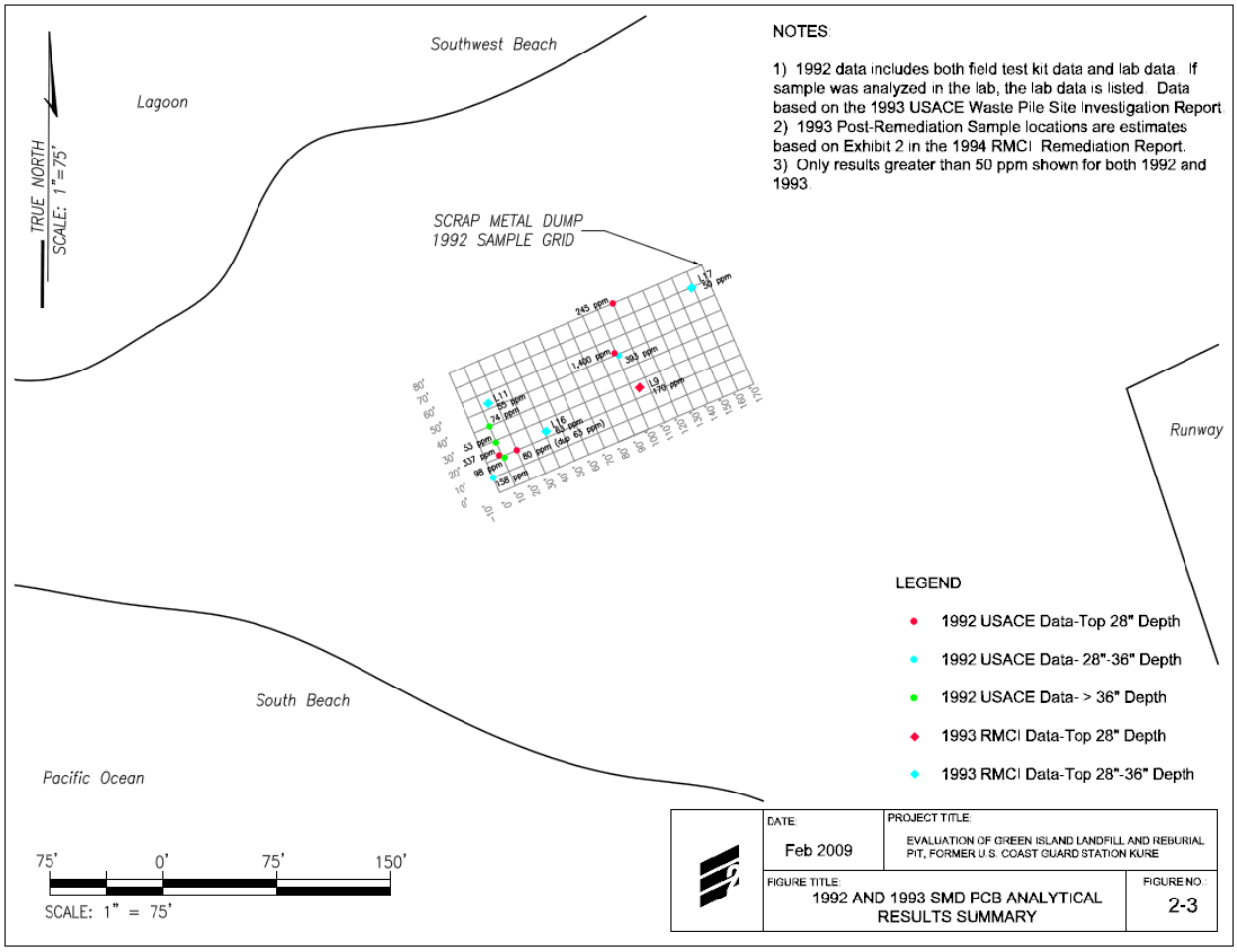


Figure C.1-3. Summary of 1992 and 1993 soil sample PCB data. *Source*: USCG 2009, Figure 2-3.

C.1.2 DU-IS Investigation (2008)

C.1.2.1 DU-IS Investigation Approach

A follow-up study of the former landfill area was carried out in 2008. The investigation focused on the use of decision unit and incremental sampling investigation strategies published by the HDOH Office of Hazard Evaluation and Emergency Response (HDOH 2008b). Note that incremental soil samples are referred to as "multiincrement" soil samples in the HDOH guidance.

C.1.2.2 DU Designation and Investigation Objectives

The footprint of the former landfill area was designated as a spill area DU, based on the past history of the site and the approximate extent of PCB-contaminated soil identified in the earlier investigations. An 80 × 180 foot DU was established, covering an area of approximately 15,000 ft^2. The targeted depth interval of the DU was 3 feet, although in some cases samples were collected to a depth of 5 feet. The total volume of the soil incorporated by the DU was approximately 2,700 yd^3.

The objective of the investigation was to estimate the representative (i.e., mean) concentration of PCBs for the designated DU mass of soil. Identification of the maximum concentration of PCBs for any given aliquot mass of soil within the DU or sample-size hot spots was determined not to be feasible or, more importantly, necessary. The area and volume of the DU were considered to be small enough for evaluation of potential risks to ecological and human receptors. Risk-based decisions on the need for additional remedial actions at the site would be made for the mass of soil incorporated within the spill area DU as a whole. Incorporating these objectives into the design of the investigation was intended to help minimize the need for additional, follow-up investigation and to avoid confusion over the need to investigate and address smaller, sample-size hot spots within the DU as a whole.

C.1.2.3 Landfill DU Characterization

As part of the site investigation, USCG took the opportunity to evaluate the potential advantage and limitations of incremental sampling methodology (ISM) over traditional, discrete sampling approaches. More than 600 discrete samples were collected from within the landfill footprint. Splits of the discrete samples were combined and used to prepare IS samples for targeted areas and depth intervals.

A 10-foot spaced sampling grid was initially established across the entire landfill footprint (see Figure C.1-4). Three depth intervals were targeted for characterization: 0–4 inches (152 samples), 28–36 inches (128 samples), and 36–60 inches (128 samples). A split sample or increment was randomly collected from each discrete sample. Increments for targeted areas and depth intervals were combined into a single ISM sample for that interval. Triplicate ISM samples were prepared for the 36–60 inch interval, for a total of five ISM samples for the DU as a whole.

Each ISM sample was air-dried and passed through a 2 mm (#10) sieve to remove larger particles. An aliquot was prepared by collecting and combining thirty 1 g increments of soil from a sample. The aliquot was tested for PCBs using a RaPID Assay Immunoassay field kit. Splits of discrete samples submitted to a laboratory for gas chromatograph analysis indicated good correlation with the immunoassay field kit data.

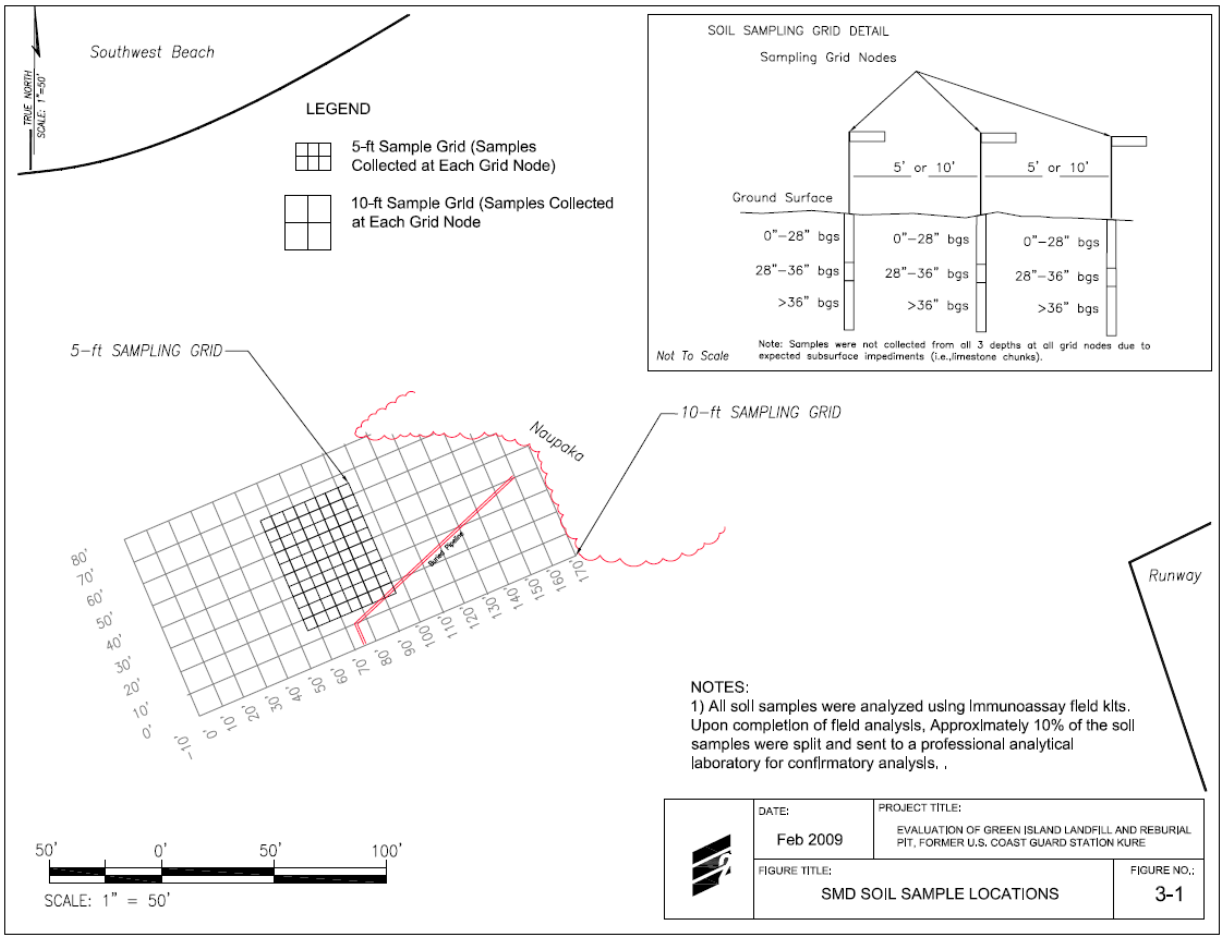


Figure C.1-4. Ten-foot and five-foot sampling grids used in 2008 ISM study.
Source: USCG 2009, Figure 3-1.

C.1.2.4 Landfill DU ISM Results

Reported concentrations of PCBs in the 0–4 inch, 28–36 inch, and 36–60 inch interval ISM samples were 0.35, 2.5, and 40 mg/kg, respectively (see Figure C.1-5). Reported concentrations of PCBs in the two replicate samples collected from the lowermost interval were 36 and 34 mg/kg. Triplicate data suggested a minimal degree of combined field and laboratory error in the samples. This is not surprising, given the large number of increments (i.e., 128–152) collected from each interval.

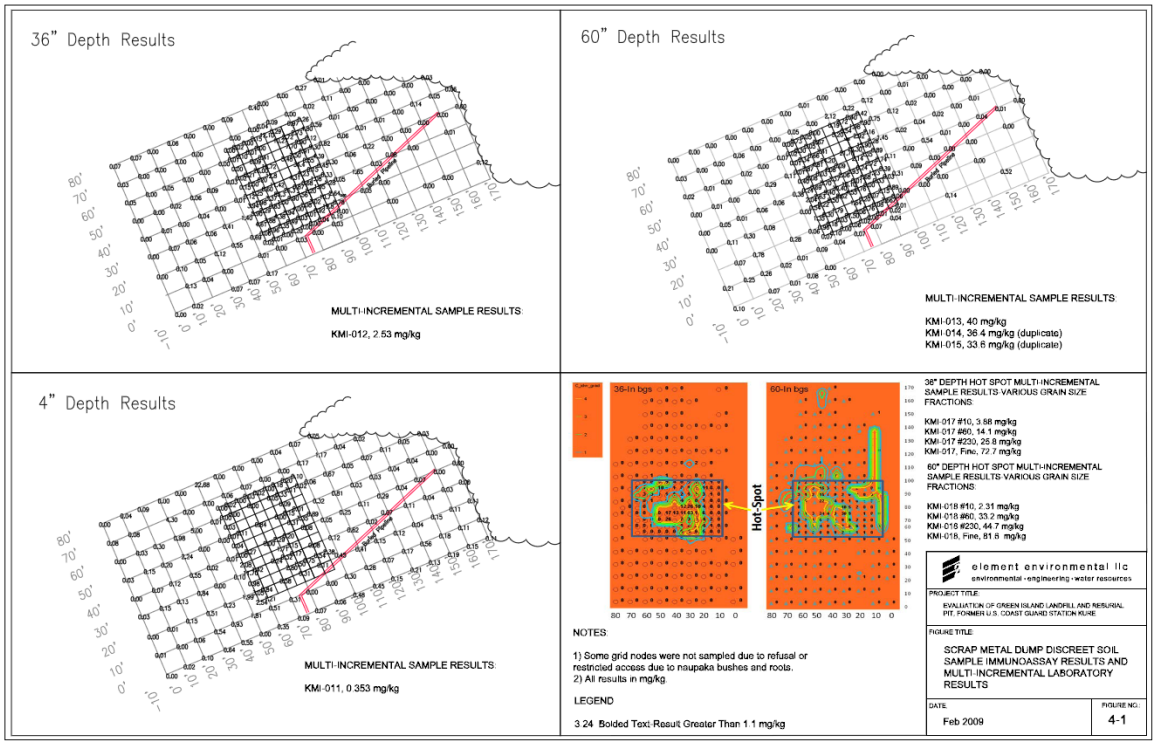


Figure C.1-5. Summary of ISM investigation results. *Source*: USCG 2009, Figure 4-1.

PCB data for each of the three targeted intervals indicated contamination above the USEPA Regional Screening Level of 0.22 mg/kg (USEPA 2008); the lower two intervals also exceeded the HDOH soil action level of 1.1 mg/kg (HDOH 2008a). Both of these screening levels are based on continuous, long-term human occupation of an area and are not necessarily applicable to current conditions on the remote, uninhabited atoll. Potential erosion of the former landfill area and dispersal of PCBs into adjacent aquatic habitats is considered to be the primary hazard posed by the contaminated soil. Reported concentrations of PCBs exceeded the marine sediment probably effects level of 0.709 mg/kg (CCME 2001, as referenced in Buchman 2008).

C.1.2.5 Use of Discrete Data to Identify Localized Spill Area

Further characterization of the landfill DU was warranted based on the initial, incremental sample PCB sample data. As part of the study, each of the 600+ discrete samples collected from the targeted intervals of the DU was tested in the field for PCBs. Tables C.1-1a through C.1-1c present a summary of the discrete data results. The discrete data indicate the presence of an approximately 3000 ft^2 concentrated area of PCB-contaminated soil at depth in the center of the former landfill footprint (see Figures C.1-5 and C.1-6 and Tables C.1-1a through C.1-1c). This area was targeted for a more detailed ISM investigation. Additional soil samples were collected from a 5-foot grid established across the central spill area and for the 28–36 and 36–60 inches depth intervals (refer to Figure C.1-5). Seventy-four discrete samples were collected from each interval. ISM samples were prepared from splits of the discrete samples in the same manner as described above. Sieves were used to separate the soil samples into four size fractions for analysis, >2 mm, ≤2–>0.25 mm, <0.25–≥0.063 mm, and <0.063 mm.

Table C.1-1a. Landfill DU PCB discrete data summary, 0–4 inch depth interval

(USCG 2009, Table B-1)

Y-Axis	X-Axis 0	10	15	20	25	30	35	40	45	50	55	60	65	70	80
0	0.00	0.01		0.00		0.03		0.00		0.03		0.00		0.08	0.00
10	0.02	0.00		0.01		0.15		0.07		0.15		0.03		0.08	0.04
20	0.10	0.06		0.15		0.13		0.12		0.20		0.30			0.09
30	0.64	0.54		0.51		0.84		0.55		2.24		0.98		4.30	0.00
40	0.55	0.93		0.23		0.00		0.02		0.00		5.00		0.00	0.00
50	0.00	1.81		1.99		0.10		0.01		0.00		0.01		0.00	22.68
55			2.54	3.05	0.98	11.7	2.08	0.00	0.04	0.01	0.03	0.00	0.00		
60	0.35	0.51	1.21	0.84		1.42		5.27		0.11		0.00	0.02	0.06	0.00
65													0.00		
70	0.09	0.31		0.58		0.24		0.00		0.84		0.00	0.03	0.07	0.00
75					2.80	0.32	1.71	2.29					0.02		
80	0.07	0.00		0.31	1.50	0.17	3.15	0.20		0.00		0.05	0.00	0.00	0.00
85					0.75				0.31	0.15	0.09	0.33	0.18		
90	0.06	0.00		0.11	0.54	0.08		0.08		0.01		3.11	0.20	0.00	0.04
95					2.38							0.02	0.10		
100	0.48	0.30		0.45		0.12		0.05		0.32		0.06		0.17	0.07
110	0.45	0.28		0.41		0.62		0.25		0.81		0.67		1.17	0.05
120	0.15	0.07		0.15		0.02		0.11		0.03		0.05		0.04	
130	0.21	0.17		0.12		0.07		0.29		0.03		0.07		0.02	
140	0.13	0.56		0.04		0.11		0.04		0.04		0.11			
150	0.19	0.18		0.15		0.04		0.07		0.04		0.05			
160	0.15	0.11		0.11		0.03		0.00		0.00		0.03			
170	0.11	0.12		0.11		0.00		0.00		0.00	0.11	0.12		0.11	

Table C.1-1b. Landfill DU PCB discrete data summary, 28–36 inch depth interval

(USCG 2009, Table B-2)

Y-Axis	X-Axis 0	10	15	20	25	30	35	40	45	50	55	60	65	70	80
0	0.00	0.01		0.00		0.03		0.00		0.03		0.00		0.08	0.00
10	0.02	0.00		0.01		0.15		0.07		0.15		0.03		0.08	0.04
20	0.10	0.06		0.15		0.13		0.12		0.20		0.30			0.09
30	0.64	0.54		0.51		0.84		0.55		2.24		0.98		4.30	0.00
40	0.55	0.93		0.23		0.00		0.02		0.00		5.00		0.00	0.00
50	0.00	1.81		1.99		0.10		0.01		0.00		0.01		0.00	22.68
55			2.54	3.05	0.98	11.7	2.08	0.00	0.04	0.01	0.03	0.00	0.00		
60	0.35	0.51	1.21	0.84		1.42		5.27		0.11		0.00	0.02	0.06	0.00
65													0.00		
70	0.09	0.31		0.58		0.24		0.00		0.84		0.00	0.03	0.07	0.00
75					2.80	0.32	1.71	2.29					0.02		
80	0.07	0.00		0.31	1.50	0.17	3.15	0.20		0.00		0.05	0.00	0.00	0.00
85					0.75				0.31	0.15	0.09	0.33	0.18		
90	0.06	0.00		0.11	0.54	0.08		0.08		0.01		3.11	0.20	0.00	0.04
95					2.38							0.02	0.10		
100	0.48	0.30		0.45		0.12		0.05		0.32		0.06		0.17	0.07
110	0.45	0.28		0.41		0.62		0.25		0.81		0.67		1.17	0.05
120	0.15	0.07		0.15		0.02		0.11		0.03		0.05		0.04	
130	0.21	0.17		0.12		0.07		0.29		0.03		0.07		0.02	
140	0.13	0.56		0.04		0.11		0.04		0.04		0.11			
150	0.19	0.18		0.15		0.04		0.07		0.04		0.05			
160	0.15	0.11		0.11		0.03		0.00		0.00		0.03			
170	0.11	0.12		0.11		0.00		0.00		0.00	0.11	0.12		0.11	

Table C.1-1c. Landfill DU PCB discrete data summary, 36–60 inch depth interval

(USCG 2009, Table B-3)

Y-Axis	X-Axis 0	10	15	20	25	30	35	40	45	50	55	60	65	70	80
0	0.00	0.01		0.00		0.03		0.00		0.03		0.00		0.08	0.00
10	0.02	0.00		0.01		0.15		0.07		0.15		0.03		0.08	0.04
20	0.10	0.06		0.15		0.13		0.12		0.20		0.30			0.09
30	0.64	0.54		0.51		0.84		0.55		2.24		0.98		4.30	0.00
40	0.55	0.93		0.23		0.00		0.02		0.00		5.00		0.00	0.00
50	0.00	1.81		1.99		0.10		0.01		0.00		0.01		0.00	22.68
55			2.54	3.05	0.98	11.7	2.08	0.00	0.04	0.01	0.03	0.00	0.00		
60	0.35	0.51	1.21	0.84		1.42		5.27		0.11		0.00	0.02	0.06	0.00
65													0.00		
70	0.09	0.31		0.58		0.24		0.00		0.84		0.00	0.03	0.07	0.00
75					2.80	0.32	1.71	2.29					0.02		
80	0.07	0.00		0.31	1.50	0.17	3.15	0.20		0.00		0.05	0.00	0.00	0.00
85					0.75				0.31	0.15	0.09	0.33	0.18		
90	0.06	0.00		0.11	0.54	0.08		0.08		0.01		3.11	0.20	0.00	0.04
95					2.38							0.02	0.10		
100	0.48	0.30		0.45		0.12		0.05		0.32		0.06		0.17	0.07
110	0.45	0.28		0.41		0.62		0.25		0.81		0.67		1.17	0.05
120	0.15	0.07		0.15		0.02		0.11		0.03		0.05		0.04	
130	0.21	0.17		0.12		0.07		0.29		0.03		0.07		0.02	
140	0.13	0.56		0.04		0.11		0.04		0.04		0.11			
150	0.19	0.18		0.15		0.04		0.07		0.04		0.05			
160	0.15	0.11		0.11		0.03		0.00		0.00		0.03			
170	0.11	0.12		0.11		0.00		0.00		0.00	0.11	0.12		0.11	

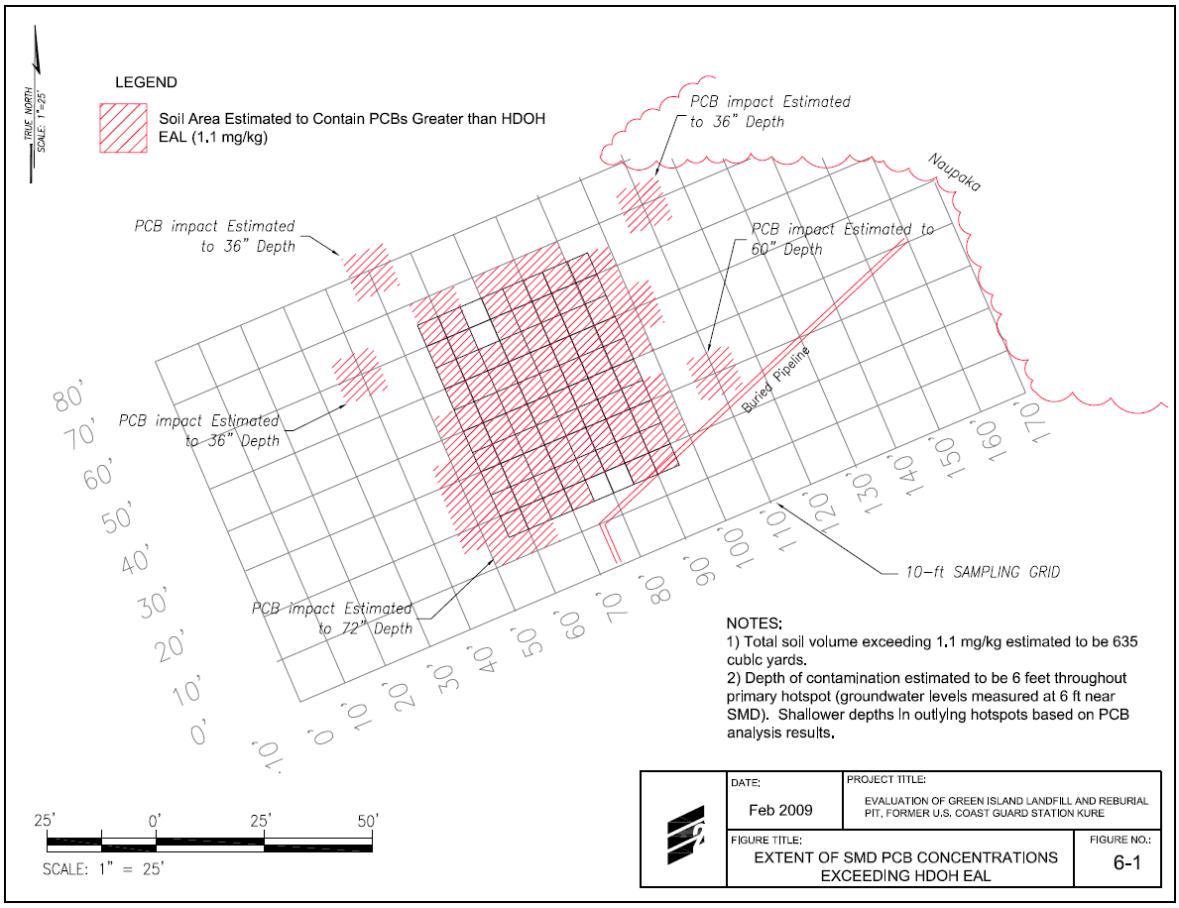


Figure C.1-6. Identification of central spill area using discrete sample data. Isolated sample points >1.1 mg/kg PCBs indicate heterogeneity of PCB distribution in area outside central spill area at the scale of a laboratory aliquot (i.e., 5 g). See also Figure C.1-5.
Source: Modified from USCG 2009, Figure 6-1.

Elevated concentrations of PCBs were also reported in four isolated, discrete sample points, outside of the central spill area (Figure C.1-6). It is important to recognize that the presence of elevated PCBs at these sample points indicate *heterogeneity* of PCB distribution in area outside of central spill area at the scale of a laboratory aliquot (i.e., 5 g). These sample point locations *do not* represent plottable spill areas and cannot be treated as such (e.g., potentially excavated and removed as part of a future cleanup action (refer to HDOH 2008b, Section 4.3.5). Instead, the presence of elevated PCBs in four single-sample points outside of the central spill area more likely indicates that the reported concentration of PCBs, in any given discrete sample collected within this area is likely to exceed 1.1 mg/kg a small percentage of the time. Removal of soil around the four, isolated sample points that happened to be identified during the study would not remove all sample-size spots above this screening level outside of the central spill area or reduce the overall mean concentration of PCBs in the soil for this area.

C.1.2.6 Targeted Spill Area DU ISM Results

Figure C.1-5 and Table C.1-2 summarize the reported concentration of PCBs in ISM samples collected from the 28–36 and 36–60 inch depth intervals of the central spill area DU. The concentration of PCBs was highest for the fine soil fraction (<0.063mm), although the mass of PCBs is present in the ≤2 mm to >0.25 mm soil fraction, which makes up 75%–80% of the samples from both intervals. Based on a weighted average of the individual size-fraction data, the total concentrations of PCBs in the <2 mm soil fraction from the two targeted intervals are 15 mg/kg and 33 mg/kg, respectively.

Table C.1-2. Incremental soil sample results for central spill area

DU	Sample interval	Grain size fraction (mm)	Sample weight (g)	Percentage of sample (%)	PCB concentration (mg/kg)
17	28–36 inches	>2 mm	110	9	3.88
		≤2 to >0.25 mm	975	83	14.1
		<0.25 to ≥0.063 mm	95	8	25.8
		<0.063 mm	1.5	0.1	72.7
		Weighted average (<2 mm):			15
18	28–36 inches	>2 mm	220	19	2.31
		≤2 to >0.25 mm	880	74	33.2
		<0.25 to ≥0.063 mm	82.5	7	44.7
		<0.063 mm	2.5	0.2	81.6
		Weighted average (<2 mm):			33

C.1.3 Comparison of ISM and Discrete Soil Data

Tables C.1-1a, C.1-1b and C.1-1c provide a summary of the discrete sample data for PCBs for each targeted interval. A statistical evaluation of discrete vs. ISM sample data is currently under way. One objective of the review is to compare estimates of the mean concentration of PCBs in the DU soil based on a specific number of discrete samples vs. one to three multiincrement samples drawn from the same data set. Examples of the types of questions to be addressed in the evaluation include the following:

- When is the mean estimated by the ISM samples better than the mean estimated from a randomly selected set of discrete data?
- Is the mean estimated by a 30-point ISM sample (or set of triplicates) better than 5 discrete samples (or 10, 15, 20 samples, etc.)?

Specific information that will be determined from the data set is as follows:

- best estimate of “true” mean PCB concentration based on all 93 discrete data points (i.e., adjusted with respect to data variability)
- range of mean estimated by any randomly collected, 30-point ISM sample

- range of mean estimated by any randomly collected set of 30-point, triplicate ISM samples (adjusted with respect to data variability)
- range of mean estimated from any randomly collected set of "X" (e.g., 5, 10, 15, 20, 25, 30) number discrete samples (adjusted with respect to data variability)

For the Kure Atoll data set, the objective was to determine the equivalent number of discrete samples to a triplicate set of 30- to 50-point ISM samples. This information will help to determine the cost-effectiveness of collecting ISM samples over discrete samples.

C.1.4 References

Buchman, M. F. 2008. *NOAA Screening Quick Reference Tables.* NOAA OR&R Report 08-1. Seattle: Office of Response and Restoration Division, National Oceanic and Atmospheric Administration. www.itrcweb.org/ism-1/references/122_NEW-SQuiRTs.pdf.

CCME (Canadian Council of Ministers of the Environment). 2001. *Reference Method for the Canada-Wide Standard for Petroleum Hydrocarbons in Soil: Tier 1 Method*. www.itrcweb.org/ism-1/references/final_phc_method_rvsd_e.pdf.

HDOH (Hawaii Department of Health). 2008a. *Evaluation of Environmental Hazards at Sites with Contaminated Soil and Groundwater*. Office of Hazard Evaluation and Emergency Response. http://hawaii.gov/health/environmental/hazard/docs/ealvolume1mar2009.pdf.

HDOH. 2008b. *Technical Guidance Manual*, in preparation. Office of Hazard Evaluation and Emergency Response. www.hawaiidoh.org.

USCG (U.S. Coast Guard). 2009. *Evaluation of Green Island Landfill and Reburial Pit, Former U.S. Coast Guard LORAN Station Kure*. Prepared by Element Environmental for Civil Engineering Unit, Honolulu.

USEPA (U.S. Environmental Protection Agency). 2008. "Regional Screening Levels for Chemical Contaminants at Superfund Sites." www.epa.gov/region09/waste/sfund/prg or www.epa.gov/reg3hwmd/risk/human/rb-concentration_table/index.htm.

C.2 CASE STUDY 2: PETROLEUM CONTAMINATED SOIL STOCKPILE

Site Name: Petroleum Contaminated Soil Stockpile, Prince of Wales Island, Alaska

Contact Name: Earl Crapps, ADEC

Site Location: The site is located on the Prince of Wales Island near Craig, Alaska. Craig is on a small island off the west coast of Prince of Wales Island and is connected by a short causeway. It is 56 air miles northwest of Ketchikan and 220 miles south of Juneau.

C.2.1 Purpose

The purpose of this project was to test the protocols in the Alaska Department of Environmental Conservation multiincrement (MI) sampling guidance.

C.2.2 Location

A small soil stockpile in a rock quarry on Prince of Wales Island near Craig was sampled (see Figure C.2-1). Craig is located on a small island off the west coast of Prince of Wales Island and is connected by a short causeway. It is 56 air miles northwest of Ketchikan and 220 miles south of Juneau.

Figure C.2-1. Map and aerial photograph of the site location.

C.2.3 Synopsis

During the 2006 excavation and removal of an underground heating oil tank, discrete samples were collected which documented diesel range organics (DRO) at 300–900 mg/kg. Stockpile tilling and fertilizing were conducted by the responsible party several times after the soil was moved from its original location in May 2006.

ADEC personnel sampled the stockpile on May 24–25, 2007. MI bulk samples were collected from 90 different locations in the 12–15 yd^3 stockpile. Subsamples were sieved to 2 mm and placed in sample jars for laboratory analysis. Fundamental error (FE), relative standard deviation (RSD), and the 95% UCL of the mean were determined following receipt of analytical results; all calculations were within acceptable parameters. The average DRO concentration was below the Method 2 migration-to-groundwater cleanup level (230 mg/kg).

C.2.4 Field Sampling Procedures

Tools and Materials

- Internet random number generator
- garden shovel
- 20-penny galvanized nails
- hand spade
- stainless steel spoons
- 2-gal zip-lock bags
- colored nylon twine
- 50-foot flexible tape
- 12-foot tape
- stainless steel ruler
- hand calculator
- leather gloves
- disposable latex sampling gloves
- field notebook
- digital camera

Although the edges of the stockpile were not clearly delineated, the stockpile dimensions measured approximately 33 × 13 × 1 feet deep. A 30-cell grid (10 cells long, 3 wide) was constructed using 20-penny nails for stakes and colored twine to form the grid pattern. Each cell measured approximately 40 inches long × 52 inches wide (see Figure C.2-2).

21	22	23	24	25	26	27	28	29	30
11	12	13	14	15	16	17	18	19	20
1	2	3	4	5	6	7	8	9	10

Overall dimensions: 33 × 13 feet
Individual cell dimensions: 52 × 40 inches
Depth: 12 inches

NORTH

Figure C.2-2. Design and construction of the grid.

Random planar and depth coordinates were determined after the cell dimensions were established using an online random number generator. Thirty length coordinates were determined by setting the minimum and maximum numbers in the random number generator between 0 and 40. Thirty width coordinates were determined by setting the minimum and maximum numbers between 0 and 52. Thirty depth coordinates were determined by setting the minimum and maximum numbers between 6 and 12. This method ensured that the top 6 inches of soil would not be sampled, as dictated by the MI sampling guidance.

Sampling locations were determined by assigning X- and Y-axes to the grid. Length was measured along the X-axis beginning at the southwest corner of the cell, followed by a Y-axis, or perpendicular measurement, to determine the width coordinate. A 20-penny nail was pushed into the soil at each coordinate to establish the primary sampling location. For example, the random coordinates for cell #2 were 33 inches along the length (X-axis) and then 18 inches to the north (Y-axis). Beginning at the southwest corner of each cell, this process was repeated until the 30 primary sampling locations were established.

A garden shovel was used to dig the holes to the approximate depth once all planar coordinates were determined. A small hand spade and 12-inch ruler were used to obtain the exact depth at each location and to clean away any soil that may have sloughed from the sidewalls.

Using a stainless steel spoon, three tablespoons of soil (~60 g per increment) were collected from the proper depth at each location and placed in zip-lock bags. If the hole was overexcavated, the sample was taken from the sidewall at the proper depth. This process was repeated until all 30 primary bulk sample increments were collected.

Duplicate and triplicate bulk samples were collected at the same depth as the primary sample within each cell using the procedures described in Figure C.2-3. Sample locations were determined by stepping out approximately one-half the distance of the cell length and width from the primary sample hole. The step-out direction varied depending on the location of the primary sample hole within the cell. For example, if the primary hole was near the far corner to the right, as shown by Figure C.2-3, step-out directions were to the left (duplicate) or down (triplicate). This method ensured an independent and systematic random approach within each cell.

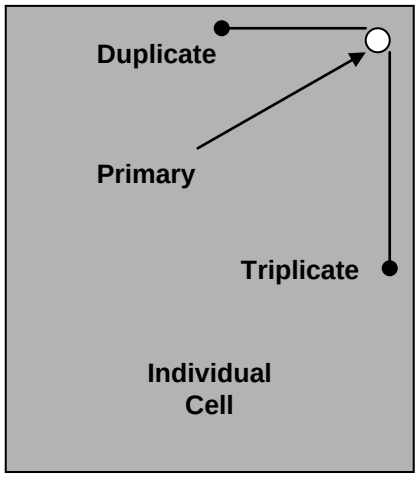


Figure C.2-3. Depiction of the primary, duplicate, and triplicate increments.

Bulk soil samples, weighing approximately 1.8 kg each, were doubled-bagged, sealed, and taped for shipment (see Figure C.2-4). Samples were not cooled because transit time back to Juneau was minimal. After the samples arrived in Juneau, they were refrigerated until the time of subsampling (see Figure C.2-5).

Figure C.2-4. Photograph showing increment collection.

Figure C.2-5. Photograph of soil sifted through a #10 sieve.

C.2.5 Subsampling Procedures

Tools and Materials

- # 10 sieve
- stainless steel trays lined with aluminum foil
- stainless steel spatula
- 4-ounce amber sample jars
- bench scale
- wire brush
- liquid soap
- 12-inch ruler
- disposable latex sampling gloves
- notebook and digital camera

Subsampling was conducted in the Department of Transportation and Public Facilities Materials Lab in Juneau, as shown in Figure C.2-6. Six subsamples were collected, including a duplicate for each of the three subsamples for an additional comparative metric.

Figure C.2-6. Photograph of the subsampling process.

Bulk samples were sieved to 2 mm using a #10 sieve. Following an initial attempt at sieving the wet soil (sample HWL 1-1), the bulk samples were placed on trays and dried at room temperature for 30 hours prior to subsampling.

Before sieving, the bulk soil samples each weighed approximately 1.8 kg. Less than half of the bulk sample was removed during the sieving process, leaving approximately 1 kg of soil for use during subsampling.

Sieved soil was spread onto a foil-lined tray with dimensions of about 7 × 10 × 3/8 inches thick. The soil was then evenly divided into a 30-square grid. About 1.5 g was collected from a minimum of two locations in each square using a small spatula to ensure that fine particles were not missed.

The 30 subsample increments, weighing approximately 45 g total, were placed in a labeled, wide-mouth sample jar placed on a bench scale. The process was repeated for the remaining two bulk samples and their duplicates; the spatula was cleaned with soap and water between each bulk subsampling event.

C.2.6 Results

Table C.2-1. Laboratory results (DRO by AK 102)

#1 Samples		#2 Samples	
HWL 1-1	130 mg/kg	HWL 1-2	87 mg/kg
HWL 2-1	160 mg/kg	HWL 2-2	140 mg/kg
HWL 3-1	110 mg/kg	HWL 3-2	110 mg/kg
Mean	133.33 mg/kg	Mean	112.33 mg/kg
Standard deviation	25.17	Standard deviation	26.58

Table C.2-2. Fundamental error (based on mass analyzed by the lab)

$$FE = \sqrt{\frac{20(d^3)}{m}}$$

d = particle size (0.2 cm for all samples)
m = sample mass

Sample 1-1	Sample 1-2	Sample 2-1	Sample 2-2	Sample 3-1	Sample 3-2
m = 29.98 g	m = 30.04 g	m = 30.05 g	m = 30.01 g	m= 30.03 g	m = 30.02 g
FE = 0.07	FE = 0.07	FE = 0.07	FE = 0.07	FE = 0.07	FE = 0.07

Table C.2-3. Relative standard deviation

$$RSD = \frac{100s}{\bar{x}}$$

#1 Samples	#2 Samples
$RSD = \frac{100(25.17)}{133.33}$	$RSD = \frac{100(26.58)}{112.33}$
$RSD = 18.9\%$	$RSD = 23.7\%$

Table C.2-4. 95% Upper confidence limit

$$95\%UCL = \bar{x} + \frac{ts}{\sqrt{n}}$$

#1 Samples	#2 Samples
$95\%UCL = 133.33 + \frac{(2.92)(25.17)}{\sqrt{3}}$	$95\%UCL = 112.33 + \frac{(2.92)(26.58)}{\sqrt{3}}$
$95\%UCL = 176$	$95\%UCL = 157$

C.2.7 Quality Control Review

- Field QC protocols were violated because samples were not cooled for shipping.

- Subsampling inconsistency occurred because one subsample was collected wet and the other five subsamples were collected dry.
- Laboratory samples were prepared and analyzed for DRO according to Method AK102. The Laboratory Data Review Checklist was completed for the lab data. All data requirements were met except the 14-day hold time; sample temperatures were thus exceeded due to subsampling challenges and shipping problems.[8]

C.2.8 Discussion

Although the stockpile was shallow, it was compacted and difficult to excavate by hand. Field sampling was therefore labor-intensive, requiring approximately 15 person-hours to complete.

Data quality may have been affected by three factors:

- Bulk samples were not cooled for shipping; hydrocarbon degradation due to an increase in microbial activity may have occurred.
- The initial attempt at subsampling was challenging due to high soil moisture content, which caused clumping and clogged the #10 sieve. The next five subsamples were collected after first air-drying the remaining bulk samples; however, data comparability is assumed because of the requirement to report on a dry-weight basis.
- The 14-day holding time and sample temperatures were exceeded. The increased microbial activity due to elevated temperatures may have biased sample results low.

FE is a result of not representing proportional concentrations of all particles in the population. Adequate mass (30 g) and a maximum particle size of 2 mm control FE. As expected, the FE for each of the samples was well below the required 15% since the particle size was ≤2 mm and the sample masses were >30 g.

RSD is a measure of data precision and is used as a QC measure to assess the MI sampling procedure and the mean concentration of the DU. The RSD calculations were 18.9% for Samples #1 and 23.7% for Samples #2. The RSD limit for a normal distribution is about 30%; therefore, one can be confident that the MI sampling results are representative.

The 95% UCL for Samples #1 was 176 and for Samples #2 was 157, indicating that the DRO cleanup level of 230 mg/kg has been met.

C.2.9 Conclusions and Recommendations

- QC problems could cause ADEC to reject the data under some circumstances, such as closing the site to a human health–based threshold.

[8] Samples were temporarily misplaced in Seattle, returned to Juneau, and then repackaged and sent to the analytical lab in Colorado.

- The random number generator worked well to establish 3-D, independent sampling coordinates. A simpler method, and equally effective, would be to generate a random location for the first cell and apply that coordinate to all other cells.

- Even though the stockpile was shallow and had been periodically mixed, the MI sampling guidance was strictly followed to ensure that the top 6 inches was not sampled. This should be standard practice, even for shallow, well-tilled stockpiles.

- While 20-penny galvanized nails worked to establish the field grid, they pulled out too easily; wooden stakes would have performed better. For large DUs, cell corner stakes would be sufficient rather than delineating the entire grid with twine.

- To minimize field time, QC for properly designed MI sampling could possibly be reduced for low-risk petroleum sites where concentrations are expected to be well below levels that may be a human health concern. Examples are direct contact and inhalation where migration to groundwater is not a concern or where groundwater is already being monitored. The merits of this recommendation will be evaluated at the end of the 2007 field season.

- Proper sampling oversight can best be achieved by the third-party contractor directly employed by the responsible party. For this reason, a contractor may wish to conduct subsampling in a controlled environment prior to shipment to the selected laboratory. The merits of this recommendation will be evaluated at the end of the 2007 field season.

- Sieving wet soil is problematic. Although holding times and temperatures should be maintained to the extent practicable, contaminants such as weathered diesel are not expected to significantly degrade. Air-drying prior to sieving may therefore be justified for DRO and residual-range organics in some cases, particularly at lower-risk sites. If volatile contaminants are a concern, separate samples should be collected according to procedures in the guidance. The merits of air-drying prior to sieving will be evaluated at the end of the 2007 field season.

- At sites where the action level is human health direct contact or inhalation, where migration to groundwater is a significant issue, or where another exposure pathway is a potentially significant concern, splitting each increment subsample as an additional laboratory QC measure may be prudent. FE, RSD, and 95% UCL calculations can be independently performed on the two data sets; archived lab samples could be evaluated if there are significant differences. The merits of this recommendation will be evaluated at the end of the 2007 field season.

C.3 CASE STUDY 3: FORMER GOLF COURSE FIELD DEMONSTRATION OF ISM

Site Name: Former Golf Course

Contact Names: Kelly Black, Neptune and Company, Inc.; Deana Crumbling, USEPA; Ligia Mora-Applegate, Florida Department of Environmental Protection; Mark Malinowski, California

Department of Toxic Substance Control; Phil Goodrum, Cardno ENTRIX; Keith Tolson, Geosyntec Consultants; Ed Corl, NAVFAC Laboratory Quality and Accreditation Office; Hugh Rieck, USACE; Steve Roberts, University of Florida; Leah Stuchal, University of Florida; Richard Lewis, Conestoga-Rovers & Associates, Inc.

Site Location: Florida

C.3.1 Background

The ITRC ISM Team identified this site for a field demonstration of ISM. The site was a former golf course where both fertilizers and herbicides containing arsenic were applied.

C.3.2 Site Investigation

This former golf course will become a residential development. While it was an active golf course, arsenic was applied in two ways. MSMA was used as an herbicide to stunt the growth of unwanted plant life, mostly on the fairways. Also, arsenic-rich fertilizer was used frequently on the course. Fertilizer was used more heavily on the tee boxes and greens than on the fairways. The contaminant of concern (COC) is arsenic and soils are the media of concern. Preliminary characterization showed that arsenic is the only COC, and that it ranges from about 0 to nearly 100 mg/kg in some areas, with significant contamination limited to the top 6 inches of soil.

Beyond characterization and collection of data suitable for human health risk assessment, investigation of this site was also used as a field demonstration of various theories relating to ISM; therefore, several alternative sampling designs (incremental and discrete) were implemented concurrently, allowing comparison of their efficacy.

Three-quarter-acre DUs were identified for investigation as representative exposure units for a human health risk assessment. In each DU, both discrete and ISM samples were collected. DU 1 was an area where previous remediation had been conducted, and the soils were expected to be quite homogeneous in regards to arsenic contamination. DU 2 and DU 3 were selected based on an expectation that they might have more elevated levels of arsenic. Similar approaches were taken for DUs 2 and 3, so in the interest of brevity, only results from DU 1 and DU 2 will be presented herein.

C.3.2.1 DU 1

This DU was a 105 × 105 foot square that comprised one-quarter of an acre. It was investigated with three different sampling approaches:

- A grid was placed on the site with each grid cell being 17.5 × 21 feet such that there were 30 cells covering the site. A systematic random sampling approach was used to collect ISM samples composed of 30 increments. Three such ISM samples were collected.
- A grid was placed on the site with each grid cell being 10.5 × 10.5 feet such that there were 100 cells covering the site. A systematic random sampling approach was used to collect ISM samples composed of 100 increments. Three such ISM samples were collected.

- Ten discrete samples were collected using simple random sampling (i.e., the locations of the 10 samples were randomly allocated across the site).

The discrete samples were collected identically to the increments; thus, the volume of the ISM samples was roughly 30 or 100 times the volume of each discrete sample. Each sample or increment was expected to be representative of the soils in the top 6 inches bgs. Data from each sampling approach were analyzed, and a 95% UCL was calculated for each. For the discrete samples, a 95% UCL can be collected directly from the set of *n* observations; it is not necessary to repeat the discrete sampling protocol multiple times to calculate a 95% UCL. For the ISM approach, a 95% UCL can be calculated because three replicate ISM samples (each based on 30 or 100 increments) were collected. As explained in Section 4 and Appendix A, while both discrete and ISM sampling may be expected to yield unbiased estimates of the mean for most sampling protocols, they represent different distributions with different standard deviations (SDs). Therefore, the methods can be expected to yield similar estimates of the mean but different confidence limits for the estimate of the mean. The 95% UCLs were compared to the Florida Department of Environmental Protection (FDEP) cleanup level of 2.1 mg/kg arsenic in soil to determine whether the site presents an unacceptable human health risk.

For the ISM approach with 30 increments, concentrations among the three replicates ranged 1.8–1.9 mg/kg with an arithmetic mean and SD of 1.8 and 0.08 mg/kg, respectively. For the ISM approach with 100 increments, concentrations among the three replicates were all roughly 1.7 mg/kg with an arithmetic mean and SD of 1.7 and 0.03 mg/kg, respectively. The 95% UCLs calculated using either Student's-*t* or Chebyshev yielded approximately the same result (rounded to two significant figures). The 95% UCLs were 2.0 and 1.8 mg/kg for the 30- and 100-increment samples, respectively. Since the upper-bound estimates of the mean are both below the action level of 2.1 mg/kg, either ISM sampling design would have provided evidence that arsenic at this site does not pose an unacceptable risk and that the site could be left in its current condition for the impending residential development.

For the $n = 10$ discrete samples collected from DU 1, arsenic concentrations ranged 0.7–5.4 mg/kg with an arithmetic mean of 2.0 mg/kg, SD of 1.4 mg/kg, and coefficient of variation of 0.7, which indicates that the data exhibit low skew. The data are not normally distributed, so a bootstrap technique was used to calculate the UCL. The 95% UCL using a bias-corrected accelerated bootstrap is 3.0 mg/kg. That level is above the threshold of interest and is considered an indication that the arsenic in soil at this site might cause an unacceptable risk for residents.

It is interesting that the data collected via discrete samples and the data collected via ISM lead to different results for this DU. In one case, the data show no unacceptable human health risk due to arsenic at this DU. In the other case, the data show that there is, indeed, an unacceptable risk due to arsenic at this site. In addition, for the ISM approach, the decision to collect three replicates allowed for an evaluation of the confidence in the estimate of the mean. Since all of the individual ISM results were within approximately 10%–20% of the action limit, any single result may have introduced uncertainty about the level of protectiveness of the risk assessment. Demonstrating that three individual ISM results and the corresponding 95% UCL are all below the action level provides stronger evidence that arsenic does not pose an unacceptable risk for

DU 1. The ISM samples that are based on 90 (3 × 30 increments) or 300 (3 × 100 increments) sample locations achieve better spatial coverage of the site than the 10 discrete samples, but both types of sampling approaches yield an unbiased estimate of the mean. It is important to recognize that any of these sampling approaches might be considered reasonable for this site, yet they lead to different conclusions and may even lead to different decisions regarding the need for remediation.

C.3.2.2 DU 2

This DU was a 52.5 × 210 foot rectangle that composed one-quarter of an acre. This DU was investigated with four different sampling approaches:

- A grid was placed on the site with each grid cell being 17.5 × 21 feet such that there were 30 cells covering the site. A systematic random sampling approach was used to collect samples composed of 30 increments. Three such ISM samples were collected.

- Thirty discrete samples were collected from immediately adjacent to the 30 systematic random sample locations used in the first of the 30-increment ISM samples collected.

- A grid was placed on the site with each grid cell being 10.5 × 10.5 feet such that there were 100 cells covering the site. A systematic random sampling approach was used to collect samples composed of 100 increments. Three such ISM samples were collected.

- The 100-cell grid was divided into four equal-sized quadrants. These quadrants were set by putting a big cross through the middle of the rectangular shape of the DU and allocating one corner to each quadrant. Five ISM samples with 25 increments each were collected from each quadrant. Then the quadrants were redrawn based on the CSM and prior information regarding expected arsenic concentrations. Specifically, one quadrant covered only the area of the former green (expected to have higher arsenic than other quadrants that included portions of the fairway), the next quadrant included a small portion of the green, and the third and fourth quadrants were composed solely of fairway.

Data from each sampling approach were analyzed, and the mean, SD, standard error (SE) of the mean, and 95% UCL were calculated for each. The 95% UCLs were compared to the FDEP cleanup level of 2.1 mg/kg arsenic in soil. In all cases, the 95% UCLs for DU 3 exceeded the threshold value of 2.1 mg/kg; however, the 95% UCLs for some of the quadrants did not exceed this threshold.

Figures C.3-1 and C.3-2 represent the data from DU 2 in box plots. The black circles on these box plots show the actual data points. The thick black line across the middle of each box is the median result. The thinner black lines above and below that are the 75^{th} and 25^{th} quantiles, respectively. That is, they represent the range in which the middle 50% of the data fall. The middle red line is the mean of the data, and the red lines above and below it represent the upper and lower confidence limits on that estimate of the mean.

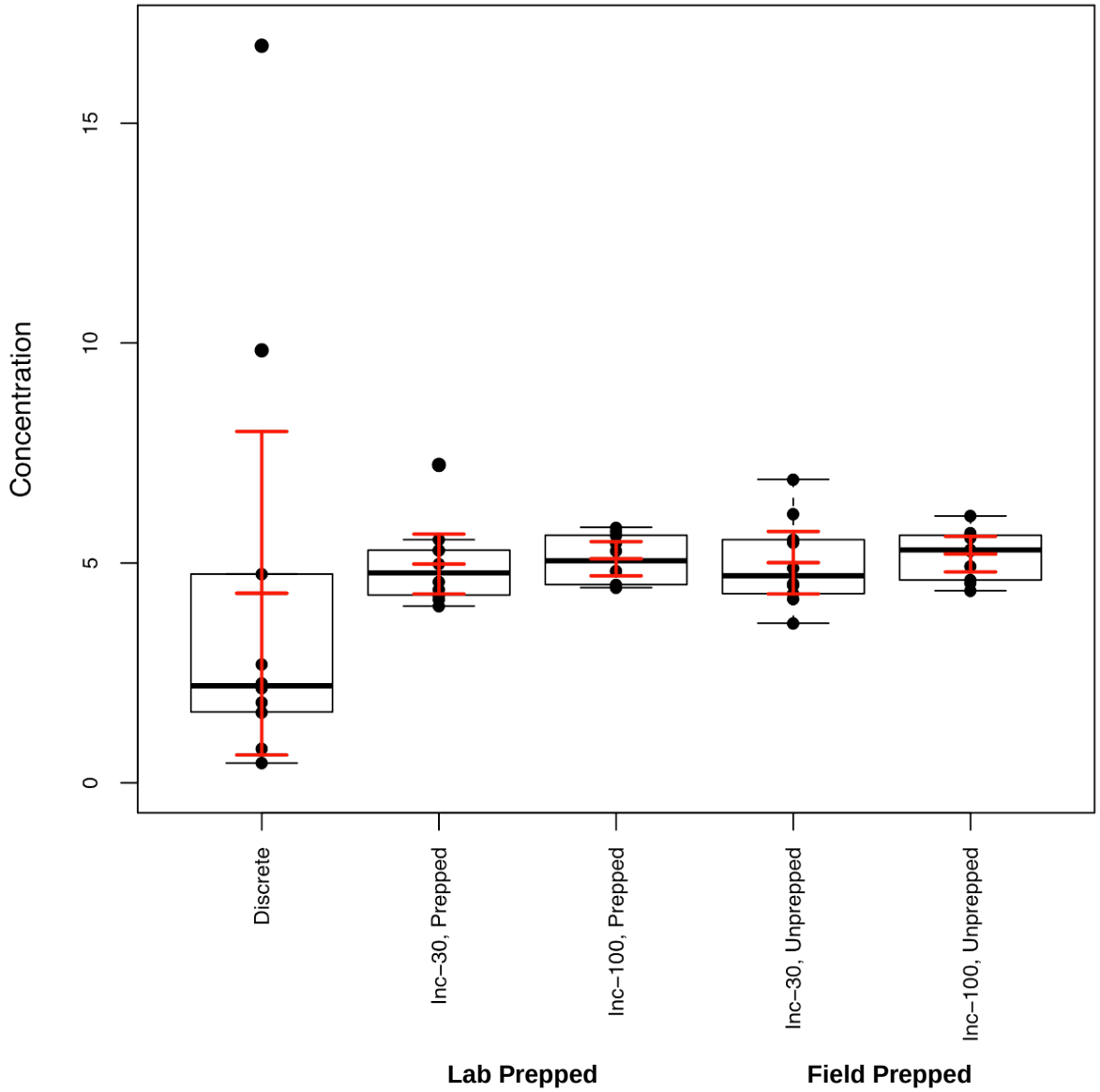


Figure C.3-1. Discrete and ISM results for DU 2.

Figure C.3-1 shows side-by-side box plots of the results of the first three (full-DU) sampling approaches. The first box represents the discrete data. The second and third boxes show the "lab-prepped" sample results, that is, the results from the samples that were carefully dried, sieved, homogenized and subsampled in a laboratory in accordance with the methodologies presented in EPA Method 8330B. The "field prepped" results shown in the final two boxes come from the same ISM samples, but subsampling of the material was performed in the field prior to shipping the main ISM sample to the laboratory for processing. The purpose of doing the field subsampling and then careful lab subsampling was to determine whether the samples that were processed in the lab would have less variability than those subsampled in the field. In Figure C.3-1 it is easy to see that there is very little difference between the results for these two types of subsampling protocols; however, it is important to note that the material being sampled here was easily mixed and that due to the nature of the contamination and the soil type, it is not surprising

that field-homogenizing was nearly as effective as the more stringent USEPA SW-846 Method 8330B–type preparation. Any extrapolation of this particular finding beyond such a simple matrix and contaminant should be considered with great caution. In other situations, it is very likely that there may be a pronounced decrease in the variability between results when a thorough homogenization protocol is used.

It is also interesting to note, and apparent on Figure C.3-1, that there was not a notable improvement in the results between samples that contained 30 increments and those that contained 100 increments. The simulations performed by the ITRC ISM Team and presented in Section 4 of this document support this finding, and in fact show that only in cases with strongly skewed or variable data is there much value in collecting more than 30 increments per sample.

Finally, in Figure C.3-1 shows that the discrete data behave exactly as expected in comparison to the ISM data. Due to the smaller sample support for the discrete data, they are expected to be much more variable than ISM data. ISM physically averages over 30 or 100 samples, thus making each result essentially an average of many single discrete samples. While the means from the different sampling approaches shown in this figure do not significantly disagree, it is very clear that the discrete samples span a much wider concentration range and are more variable than the ISM results. This is a finding that matches the theory behind ISM, is borne out in the simulation studies, and can generally be expected to be true for most types of environmental investigations. Accordingly, one would anticipate that the magnitude of the UCL generated with discrete sampling with typical samples sizes (e.g., n = 10 to 30) would be greater than UCLs generated with ISM sampling.

Figure C.3-2 shows the results (in box plot form) of the quadrant sampling. Ideally, a DU would be composed of largely homogeneous media (at least in regards to the parameters of interest). If the CSM is not convincing, or if there is some reason to believe that the DU may have gross spatial heterogeneity (i.e., different concentrations of the chemical of interest in different areas of the DU), then partitioning the DU and taking separate ISM samples in each partition might be a useful strategy. For this former golf course, there was reason to believe that the greens and tees would have different concentrations of arsenic than the fairways, so partitioning into quadrants was employed.

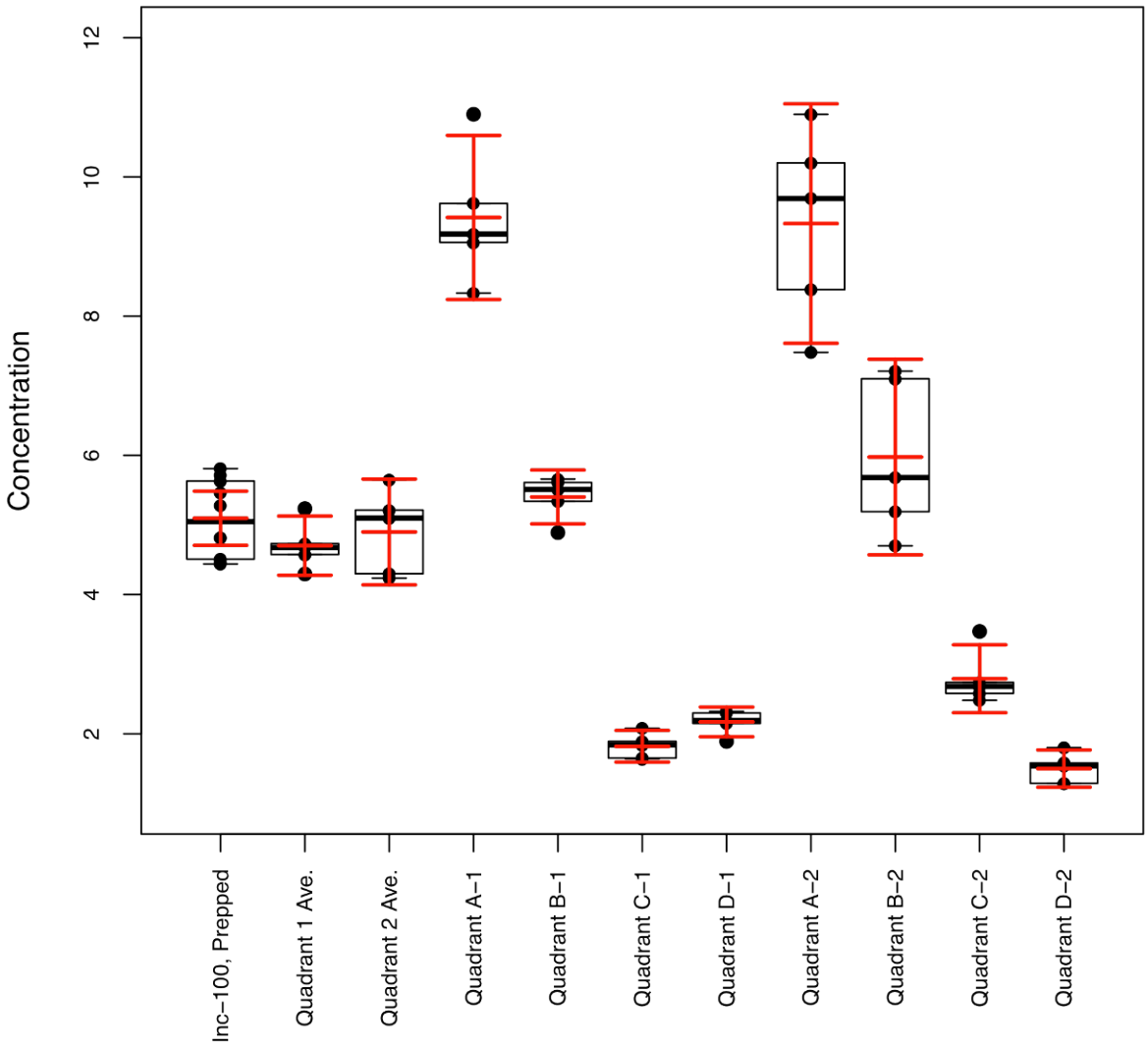


Figure C.3-2: ISM results for DU 2 by quadrant.

In Figure C.3-2, the boxes representing samples A-1 through D-1 show the results from the samples based on the original quadrants selected purely by breaking the 100-cell grid into four conveniently shaped sections without any recourse to prior knowledge or expectations for the site. The final four boxes representing samples A-2 through D-2 show the results for the samples from the quadrant configuration based on the CSM and our prior knowledge of the site. It is evident that, indeed, beginning with the quadrant placed on the green (1) and moving out to the fairway (4), there is a clear and significant difference in concentrations of arsenic trending down with distance from the green. It is interesting that data presented in the box plots for quadrants A-1 through D-1 would certainly have provided some interesting conversation among the project team if we did not actually already have a reason to suspect there were spatial differences across this DU.

Based on these results, it is likely that remediation would be considered necessary throughout DU-2. However, the project team might also decide to revisit the CSM and conduct additional sampling to better define the areas that would require remediation.

C.3.3 Lessons Learned

The information presented in this case study is only a small portion of the interesting information learned by implementing ISM during this field demonstration. Other aspects that will be presented in a final report on this study include the time required for various aspects of the sampling; cost comparisons (based on time, resources, and analytical costs, including sample preparation); Monte Carlo simulations from the discrete and ISM data to test theories on the value of the information; evaluation of the observed data distributions and their similarity to the simulated distributions used in Section 4; and comparison of results from ground vs. unground samples.

From the analyses presented herein, there are a few important ideas to consider:

- In cases where the concentration of the COC is near the threshold of interest, it is prudent to be aware that any conclusions made about the site are based on a sample of data, and even if collected in a careful and appropriate manner, it may or may not lead to the same conclusion that would be reached based on another sample of the data. The expected variability in sample results is a primary reason why a common DQO is to collect sufficient data to calculate a UCL for a parameter estimate.

- Partitioning DUs into subareas may provide an opportunity to discern spatial differences that would not be apparent if ISM samples were collected from the entire DU as a whole.

- Discrete sampling is generally expected to yield a distribution of results with approximately the same arithmetic mean but higher SD, SE, and 95% UCL than ISM sampling of the same DU.

- For this site, there was no added benefit to increasing the number of increments from 30 to 100 per ISM sample. For locations in which the sample mean and corresponding 95% UCL are close to a decision threshold, increasing the number of increments can reduce the SE (and corresponding UCL) enough to alter the decision. The challenge for most sites, particularly in the absence of pilot data, is that a risk assessor typically lacks a priori knowledge about how close the population mean may be to a decision threshold.

C.4 CASE STUDY 4: HAWAIIAN HOMELANDS DEVELOPMENT, KAPOLEI, OAHU, HAWAII

Site Name: Hawaiian Homelands Development, Kapolei, Oahu, Hawaii

Contact Name: Roger Brewer, HDOH

Site Location: The East Kapolei Affordable Housing Project property is located in East Kapolei, Kapolei, Oahu, Hawaii.

C.4.1 Background and Previous Investigations

This case study summarizes the investigation of a 400-acre, former sugarcane field and a ½-acre pesticide mixing area located within the field. The area was being developed for residential and commercial use. The primary COCs were arsenic, PCP, dioxins (associated with past use of PCP), and triazine herbicides, each used in the past for weed control. A detailed discussion of the sugarcane field investigation is provided in the report *East Kapolei Affordable Housing Project Kapolei, Oahu, Hawaii, Final Site Assessment Report* (TTEMI 2007). A summary of the pesticide mixing area investigation is provided in the report *Site Investigation Report and Environmental Hazard Evaluation, East Kapolei II Pesticide Mixing and Loading Site* (ESTC 2007, 2010).

DU and ISM (ISM, referred to as "multiincrement sampling" or "MIS" in the reports) investigation approaches were used to investigate the site. The pesticide mixing area, where heavy pesticide contamination had been previously identified, was investigated separately from the field area. This approach allowed the field area to be cleared for development early in the process. Except as noted, 30- to 50-increment ISM samples were collected from DUs and subsampled in the laboratory for preparation of aliquots and analysis.

C.4.1.1 Field Area Investigation

The field was investigated through the characterization of 59 hypothetical, residential lot-size DUs (5000 ft^2) randomly located within the 400-acre field (Figure C.4-1, TTEMI 2007). Using the terminology proposed in the ITRC ISM document, the entire field could be alternatively considered to be the DU, and in the individual lots, DUs as SUs. Previous suggestions to characterize the field using a similar number of discrete samples were rejected due to poor coverage of individual DU areas (e.g., a single discrete sample at each of the 59 target locations vs. a 30-point ISM sample).

Testing a minimum of 59 hypothetical lots ensured that contaminant levels in 95% of the lots not tested were no higher than in the most contaminated DU identified (HDOH 2008). After locating the center point for a DU in the field, a 5000 ft^2 area was marked off and a 40-point ISM sample collected (total 59 ISM samples and 2360 increments; see Figure C.4-1). Triplicate samples were collected in 10% (six) of the DUs. Samples were collected over a 5-day period. Reported concentrations of targeted contaminants in all the DUs were below environmental action levels for residential use and the fields were cleared for development (e.g., maximum 100 ng/kg toxicity equivalent [TEQ] dioxins).

C.4.1.2 Pesticide Mixing Area Investigation

The pesticide mixing area was ringed with 33 1000 ft^2 and 5000 ft^2 DUs to verify that the boundary of heavy contamination had been adequately identified, based on previous discrete

sample investigations (Figure C.4-2, ESTC 2007). A 0–6 inch surface ISM sample was collected from each DU (total 33 samples and 990 increments, plus replicates).

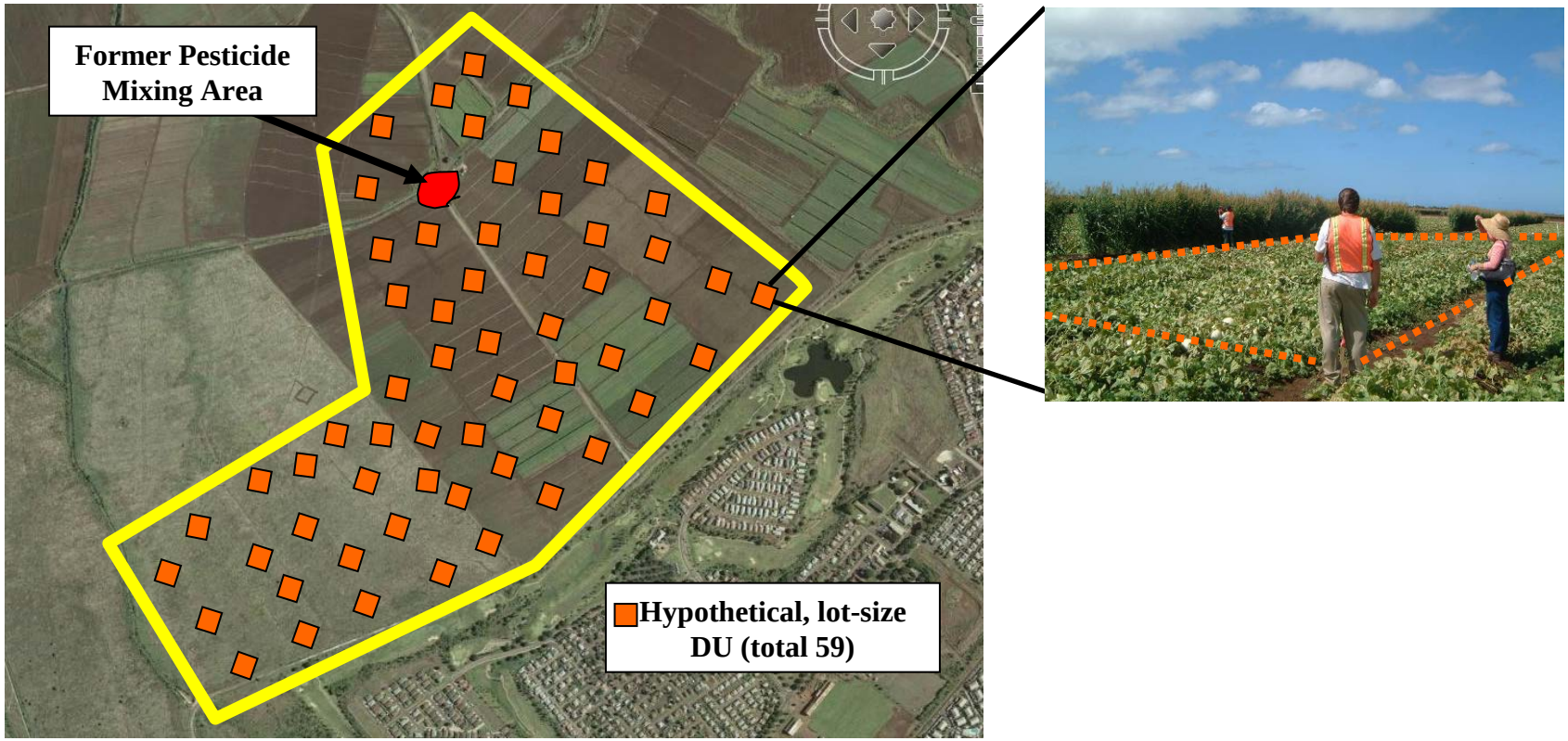


Figure C.4-1. ISM investigation of a 400-acre former sugarcane field and 59 hypothetical, lot-size (5000 ft^2) DUs characterized.

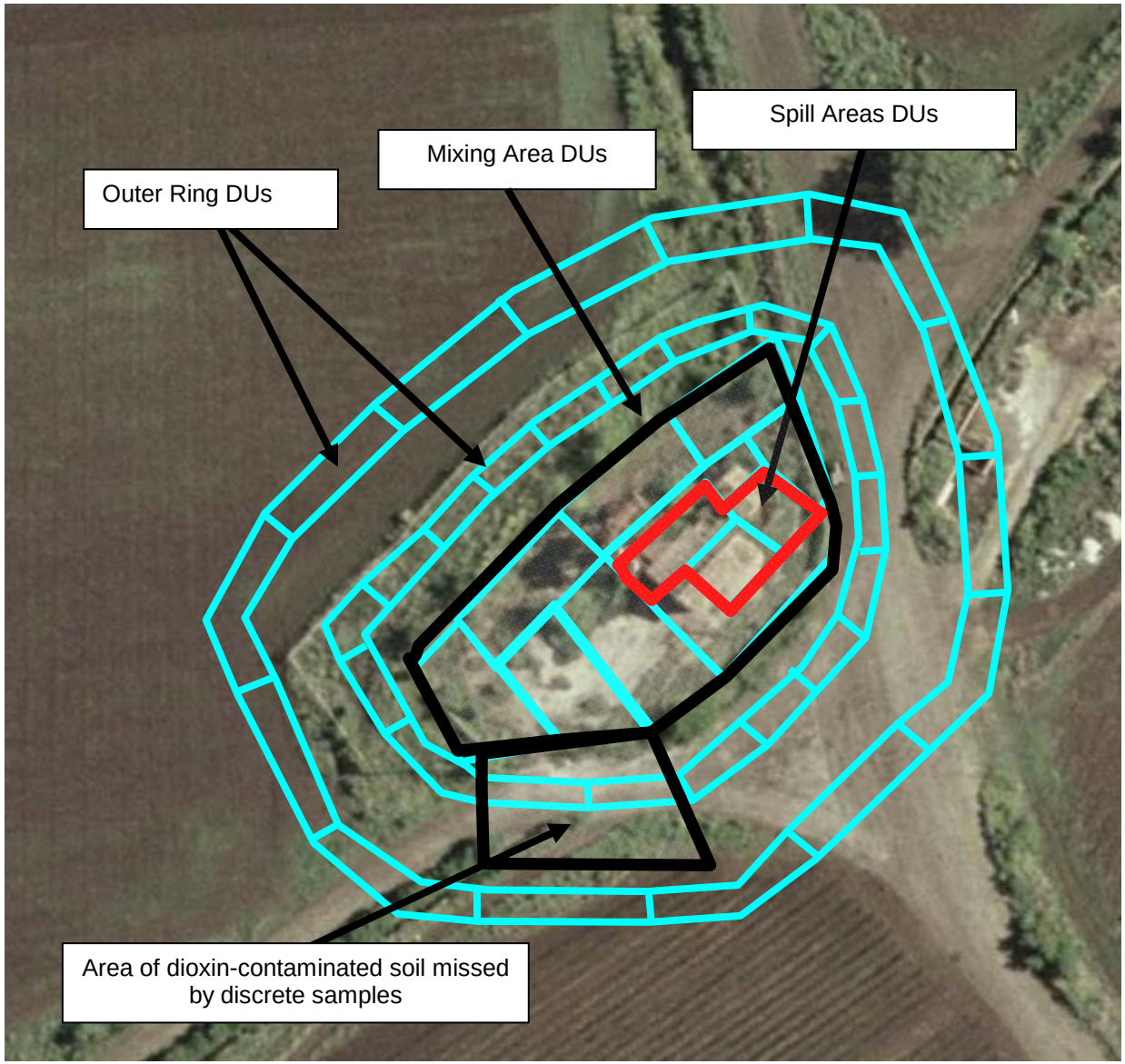


Figure C.4-2. ISM investigation of a ½-acre pesticide mixing area DU within the former sugarcane field.

The interior of the mixing area was divided into 15 DUs (ESTC 2010). A suspected area of especially heavy contamination was subdivided into three small spill-area DUs (<2000 ft^2), with the remaining area divided into 12 DUs equal to or less than the default, residential-lot exposure area of 5000 ft^2. Each of the 15 DUs was subdivided into up to four "sampling unit" (SU) layers to investigate the vertical distribution of contaminants. An ISM sample was collected within each surface and subsurface SU (total 31 SUs), with triplicates collected in three units. Twenty direct-push borings were installed in the three spill-area DUs to characterize subsurface contamination (i.e., one 20-increment ISM sample per subsurface SU layer). Subsurface SUs in the outer DUs were accessed and sampled by trenching.

A total of 64 ISM samples composed of 2000+ increments, plus replicates, was collected. Significant dioxin contamination was identified in all 15 DUs (maximum 650,000 ng/kg TEQ dioxins) and heavy triazine contamination within the targeted spill areas (see Figure C.4-2). Both the lateral and vertical extent of contamination was significantly greater than estimated based on earlier, discrete sample data, increasing the volume of contaminated soil by a factor of at least 3. The ring DU ISM samples also identified a 15,000 ft^2 area of dioxin-contaminated soil on the south side of the mixing area that was likewise missed by earlier discrete samples (see Figure C.4-2).

In 2009, the USEPA collected 83 surface and subsurface discrete samples around the perimeter of the mixing area to confirm that the extent of contamination had been adequately identified (USEPA 2009, unpublished). The discrete samples similarly suggested that contamination around the mixing area was below target action levels. The samples failed to identify the outer area of contamination identified in ring DUs to the south, however. The investigations confirm that ISM samples, essentially very good "composite" samples with additional lab requirements, are better able to capture small hot spots and overall contaminant heterogeneity within a targeted area.

C.4.2 References

ESTC (EnviroServices & Training Center, LLC). 2007. *Site Investigation Report and Environmental Hazard Evaluation, East Kapolei II Pesticide Mixing and Loading Site.*

ESTC. 2010. *Site Investigation Report and Environmental Hazard Evaluation, East Kapolei II Pesticide Mixing and Loading Site.* http://hawaii.gov/dhhl/publications/ekii/SIR-EHE,%20Mar2010.pdf.

HDOH. 2008. *Technical Guidance Manual*, in preparation. Office of Hazard Evaluation and Emergency Response. www.hawaiidoh.org.

TTEMI (Tetra Tech EM, Inc.). 2007. *East Kapolei Affordable Housing Project Kapolei, Oahu, Hawaii, Final Site Assessment Report.* Prepared for the Hawaii Department of Health Office of Hazard Evaluation and Emergency Response.

Appendix D

ISM Team Contacts

ISM TEAM CONTACTS

Ligia Mora-Applegate, ISM Team Co-Lead
Florida Dept. of Environmental Protection
850-245-8992
ligia.mora-applegate@dep.state.fl.us

Mark Malinowski, ISM Team Co-Lead
California Dept. of Toxic Substance Control
916-255-3717
mmalinow@dtsc.ca.gov

Angela Sederquist, ITRC Program Advisor
Booz Allen Hamilton
210-932-5668
sederquist_angela@bah.com

Diane Anderson
Agriculture and Priority Pollutants Laboratory, Inc.
559-275-2175
danderson@applinc.com

Matthew Baltusis
Michigan Dept. of Environmental Quality
517-335-3140
baltusism@michigan.gov

Alfonso Benavides
Texas A&M University
979-820-4412
benavidesi.a@gmail.com

Nancy Bettinger
Massachusetts Dept. of Environmental Protection
617-556-1159
nancy.bettinger@state.ma.us

Kelly Black
Neptune and Company, Inc.
720-746-1803
kblack@neptuneinc.org

Robin Boyd
AECOM Environment
808-356-5376
robin.boyd@aecom.com

Roger Brewer
Hawaii Dept. of Health
808-586-4328
roger.brewer@doh.hawaii.gov

Mark Bruce
TestAmerica, Inc.
330-966-7267
mark.bruce@testamericainc.com

Matt Butcher
ARCADIS
206-399-4426
matthew.butcher@arcadis-us.com

Anna Butler
USACE
912-652-5515
anna.h.butler@usace.army.mil

Frank Camera
New Jersey Dept. of Environmental Protection
609-633-7840
frank.camera@dep.state.nj.us

Janie Carrig
USACE–Omaha District
402-995-2283
janie.e.carrig@usace.army.mil

Michelle Caruso
Tetra Tech, Inc.
973-630-8128
michelle.caruso@tetratech.com

Michael Chacon
Pueblo de San Ildefonso
505-455-4122
mchacon@sanipueblo.org

Arnab Chakrabarti
Geosyntec Consultants
510-285-2755
achakrabarti@geosyntec.com

Eric Cheng
USACE
502-315-7443
eric.s.cheng@usace.army.mil

Jay Clausen
U.S. Army Cold Regions Research and Engineering Laboratory
603-646-4597
Jay.L.Clausen@us.army.mil

Jennifer Corack
Navy and Marine Corps Public Health Center
757-953-0950
jennifer.corack@med.navy.mil

Ed Corl
NAVFAC Laboratory Quality and Accreditation Office
757-322-4768
william.corl@navy.mil

Earl Crapps
Alaska Dept. of Environmental Conservation
907-269-7691
earl.crapps@alaska.gov

Deana Crumbling
USEPA
703-603-0643
crumbling.deana@epa.gov

Annette Deitz
Oregon Dept. of Environmental Quality
503-229-6258
DIETZ.Annette@deq.state.or.us

Jan Dunker
U.S. Army Corps of Engineers
402-697-2566
Jan.W.Dunker@usace.army.mil

John Esparza
USACE
916-557-7451
john.esparza@usace.army.mil

Timothy Fitzpatrick
Florida Dept. of Environmental Protection
850-245-8083
timothy.fitzpatrick@dep.state.fl.us

Tim Fredrick
USEPA Region 4
404-562-8598
frederick.tim@epa.gov

Jon Gabry
Tetra Tech, Inc.
973-630-8368
jon.gabry@tetratech.com

Rick Galloway
Delaware Dept. of Natural Resources and Environmental Control
302-395-2614
rick.galloway@state.de.us

Edward Gilbert
Environmental Response Team
702-784-8018
gilbert.edward@epa.gov

John Gillette
AFCEE
210-969-8440
john.gillette.1@us.af.mil

Philip Goodrum
Cardno ENTRIX
315-396-6655
philip.goodrum@erm.com

Adrienne Gossman
Louisiana Dept. of Environmental Quality
504-736-7763
adrienne.gossman@la.gov

Joe Guarnaccia
BASF Corporation
732-762-4743
joseph.guarnaccia@basf.com

Paul Hadley
California Dept. of Toxic Substance Control
916-324-3823
phadley@dtsc.ca.gov

Jeanene Hanley
Arizona Dept. Environmental Quality
602-771-4314
jph@azdeq.gov

Bryan Harre
NAVFAC ESC
805 982-1795
bryan.harre@navy.mil

John Hathaway
Pacific Northwest National Laboratory
509-372-4970
john.hathaway@pnnl.gov

Marvin Heskett
TestAmerica, Inc.
808-486-5227
marvin.heskett@testamericainc.com

Keith Hoddinott
U.S. Army Public Health Command
410-436-5209
keith.hoddinott@us.army.mil

Julie Hoskin
Arizona Dept. of Environmental Quality
602-771-4866
hoskin.julie@azdeq.gov

Paul Jurena
AFCEE/TDV
210-395-8425
paul.jurena.ctr@us.af.mil

Guy Kaminski
ARCADIS
813-353-5788
guy.kaminski@arcadis-us.com

Derek Kinder
USACE–Louisville District
502-315-6393
derek.s.kinder@usace.army.mil

Robert Kirgan
USAEC
201-295-2092
Robert.Kirgan@us.army.mil

Richard Lewis
Conestoga-Rovers & Associates, Inc.
239-936-0789
rlewis@craworld.com

Peter Lorey
USACE
716-879-4158
Peter.M.Lorey@usace.army.mil

Pornteera Manakun
ERM
pornteera.mahakun@erm.com

Joseph McElhaney
ExxonMobil
281-654-6886
jdmcelh@gmail.com

Kari Meier
USACE
502-315-6316
kari.l.meier@us.army.mil

Anita Meyer
USACE
402-697-2585
anita.k.meyer@usace.army.mil

Brian Nagy
TestAmerica, Inc.
808-486-5227
brian.nagy@testamericainc.com

Neal Navarro
USACE
916-557-6948
neal.navarro@usace.army.mil

Dee O'Neill
Columbia Analytical Services
360-577-7222
doneill@caslab.com

Osaguona Ogbebor
CH2M HILL
937-220-2904
oogbebor@ch2m.com

Katherine Owens
Eastern Idaho Community Action Partnership
208-521-3696
paragon@ida.net

Jeffrey Patterson
Texas Commission on Environmental Quality
512-239-2489
jepatter@tceq.state.tx.us

Mike Pfister
California EPA, Dept. of Toxic Substances Control
559-297-3958
mpfister@dtsc.ca.gov

Hugh Rieck
USACE
402-697-2660
hugh.j.rieck@usace.army.mil

Stephen Roberts
University of Florida
352-294-4514
smroberts@ufl.edu

Randall Ryti
Neptune and Company, Inc.
505-662-2121
rryti@neptuneinc.org

Christopher Saranko
Environmental Planning Specialists, Inc.
770-862-6558
csaranko@envplanning.com

George Shaw
W.L. Gore & Associates, Inc.
410-506-4776
gshaw@wlgore.com

Lizanne Simmons
Kleinfelder, Inc.
858-320-2267
lsimmons@kleinfelder.com

Anita Singh
Lockheed Martin
702-563-9906
asingh428@gmail.com

Michael Stroh
Missouri Dept. of Natural Resources
573-522-9902
michael.stroh@dnr.mo.gov

Leah Stuchal
University of Florida
352-264-4641
lstuchal@ufl.edu

Keith Tolson
Geosyntec Consultants
813-558-0990
ktolson@geosyntec.com

Brandon Swope
SPAWAR Systems Center Pacific
619-553-2761
brandon.swope@navy.mil

Deborah Walker
U.S. Army Engineering & Support Center
256-895-1796
deborah.d.walker@usace.army.mil

William Sy
USEPA Region 2
732-632-4766
sy.william@epa.gov

Gary Zimmerman
Golder Associates
425-883-0777
gzimmerman@golder.com

Appendix E

Glossary

GLOSSARY

accuracy – the degree of closeness of measurements of a quantity to its actual (true) value. Together precision and bias determine accuracy.

action level – the generic term applied to any numerical concentration value which 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.

arithmetic mean – the sum of *x* measurements divided by the number of equally weighted measurements.

average – see *arithmetic mean*.

bias – the tendency for a measurement to consistently over- or underestimate the actual (true) value. Together precision and bias determine accuracy.

colocated sample – a QC check sometimes performed in traditional sampling. In discrete sampling plans they 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 colocated samples usually uses the relative percent difference (RPD) as the measure.

composite sample – a sample composed of two or more increments, which generally undergoes some preparation procedures designed to reduce the variance in the errors associated obtaining a measurement from the combined sample. An ISM sample is a composite sample whose collection and preparation steps are designed using the general suggestions of Gy's sampling theory. Traditional composite samples generally do not consist of a large volume and a large number of increments and do not undergo the same preparation and subsampling steps suggested by Gy's sampling theory.

compositional heterogeneity – the heterogeneity arising from the composition of each particle within a decision unit.

coverage – for statisticians, the probability that a confidence interval encloses or captures the true population parameter. For example, a calculated 95% UCL 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 objective (DQO) – a qualitative and quantitative statement derived from the DQO process that clarifies study technical and quality objectives, defines the appropriate type of data, and specifies tolerable levels of potential decision errors that will be used as the basis for establishing the quality and quantity of data needed to support decisions.

decision – a determination made about a potential contaminant(s) of concern (from survey through remediation) or a determination for a volume of media using environmental sample data.

decision error – the likelihood of making an incorrect decision based on site-specific information including measurement from samples collected within the DU.

decision mechanism – an algorithm or protocol that results in the decision about a potential contaminant of concern or for a decision for a volume of media.

decision unit (DU) – the smallest volume 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 (SUs).

disaggregation – the act of breaking the soil clumps into individual small particles but keeping the small pebbles and hard crystalline particles intact.

distributional heterogeneity – the heterogeneity arising from the way particles are distributed within the decision unit or sample. For example, if heavy particles settling to the bottom of a sample results in less distributional heterogeneity.

energetics – explosives and propellant residues as specified in USEPA SW-846 Method 8330B.

exposure point concentration (EPC) – The value, based on either a statistical derivation of measured data or modeled data, that represents an estimate of the chemical or radionuclide concentration available from a particular medium or route of exposure.

exposure unit (or exposure area) – for purposes of risk assessment, a defined area throughout which a potential receptor may be exposed to a contaminant. The receptor is assumed to move randomly across the area, being exposed equally to all parts of the area. The assumption of equal exposure to any and all parts of the exposure area is a reasonable approach (USEPA 1992) that allows a spatially averaged soil concentration to be used to estimate the true average concentration contacted over time.

field replicate samples – collected following the same the process within the DU but from a different set of locations. The manner in which the replicate is collected is determined during systematic planning. The purpose of the collection of replicates is to provide multiple estimates of the mean.

fundamental error – the error that results from the compositional heterogeneity of the sampling unit and the mass of the sample collected. Fundamental error is always present and can be estimated before sampling.

grand mean –the arithmetic mean of ISM replicates from the same DU.

grinding – a generic term for soil disaggregation or milling. The grinding type or equipment must be specified to select a particular laboratory process.

grouping and segregation error – the error that results from the distributional heterogeneity of the sampling unit.

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 by soil sampling results. The specific area and magnitude of contamination constituting a hot spot should be agreed on during systematic project planning.

heterogeneity – the condition of being nonuniform because of dissimilar or diverse constituents. For soils examined at the spatial scale of containerized samples, “heterogeneous” describes soils composed of different materials, such as different minerals or organic carbon. Even if all soil particles are made of the same material, soil may still labeled heterogeneous if the material is present in different forms, such as different particle sizes. These types of heterogeneity are called “compositional heterogeneity.” Soils may also be heterogeneous on larger spatial scales. A prime example is differences in contaminant concentrations from one

location to another within some area or volume of soil. This condition is called "distributional heterogeneity."

increment – a portion of the sampling unit that is collected with a single operation of a sampling device and combined with other increments to form an incremental sample.

increment delimitation error – the error that results from incorrect shape of the volume of material extracted from the sampling unit to form the sample.

increment extraction error – the error that results from incorrectly extracting the increment from the sampling unit.

incremental sample – a collection of increments collected from a single sampling unit, which are combined, processed, and analyzed to estimate the mean concentration in that sampling unit.

laboratory replicate sample – a sample that is split into subsamples at the lab. Each subsample is then analyzed and the results compared. They are used to test the precision of the laboratory procedures.

long-range heterogeneity fluctuation error (CE_2) – the error generated by nonrandom local trends within the population.

milling – complete particle size reduction of all soil components including hard crystalline materials to a defined maximum particle size (e.g. <250 µm or <75 µm).

periodic heterogeneity fluctuation error (CE_3) – the error generated by cyclic nonrandom phenomena.

precision –a measure of reproducibility. Together precision and bias determine accuracy.

preparation error – the sum of the errors that occur during sample transfer and preparation processes.

relative standard deviation – the arithmetic standard deviation divided by the arithmetic mean. Also called the "coefficient of variation" (CV).

replicate (duplicate) sample – one of the two or more samples or subsamples obtained separately at the same time by the same sampling procedure or subsampling procedure.

representativeness – description of the degree to which an estimate 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).

sample – for statisticians, a set of observations collected from a population (i.e., a set of ISM samples). For field investigators, it is the mass/volume of material obtained from a sampling unit (i.e., consisting of multiple increments). For laboratory technicians, the sample is all the material delivered to the laboratory in a container collected by the field crew.

sample support – the size (mass or volume), shape, and orientation of that portion of the sampling unit that is sampled.

sampling error – anything during sample collection and handling that causes the measured properties of sample to deviate from the targeted properties of the population. The population of interest is defined in accord with the decision to be made from the data.

sampling unit – user-defined volume of soil (or other media) from which increments are collected to determine an estimate of the mean concentration for that volume of soil (or other media).

soil – the fragmented material in the surface of the earth formed as the result of the complex interaction of the rock surface with atmospheric and mechanical factors, consisting of rock and mineral particles mixed with decayed organic matter (humus), excluding only the top few inches with its organic content (known as "topsoil"). Soils can represent the parent rock types that lie below (e.g., residual soils) or can be formed by various erosional processes of water and wind far away from the parent rock materials (e.g., transported soil) or transported in bulk by anthropogenic or artificial mechanisms.

specimen – portion of the sampling unit collected prior to taking into account the suggested sampling, preparation, and subsampling activities of sampling theory intended to produce sufficiently representative estimates in specified volumes of media.

standard deviation – measure of the dispersion of imprecision of a sample or population distribution expressed as the positive square root of the variance and that has the same unit of measurement as the mean.

statistic – function of the sample measurements; for example, the sample mean or standard deviation. A statistic usually but not necessarily serves as an estimate of a population parameter. A summary value calculated from a sample of observations.

subsample – the structured composite of the increments collected from an SU sample.

Appendix F

Acronyms

ACRONYMS

ADEC	Alaska Department of Conservation
AE	analytical error
AM	arithmetic mean
AST	aboveground storage tank
ASTM	ASTM International, formerly American Standards and Testing Society
bgs	below ground surface
BTEX	benzene, toluene, ethylbenzene, and xylenes
CE_2	long-range heterogeneity fluctuation error
CE_3	periodic heterogeneity fluctuation error
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CFR	Code of Federal Regulations
CH	compositional heterogeneity
CLT	central limit theorem
COC	chemical of concern
COPC	chemical of potential concern
cpm	counts per minute
CRREL	U.S. Army Cold Regions Research and Engineering Laboratory
CSM	conceptual site model
CV	coefficient of variation
DE	delimitation error
DEC	Department of Conservation
df	degrees of freedom
DH	distributional heterogeneity
DL	detection level
DOD	U.S. Department of Defense
DOE	U.S. Department of Energy
DQI	data quality indicator
DQO	data quality objective
DRO	diesel-range organics
DU	decision unit
EE	extraction error
ECOS	Environmental Council of the States
EDQW	Environmental Data Quality Working Group
Eh	oxidation-reduction potential
ELAP	Environmental Laboratory Accreditation Program
EPC	exposure point concentration
ERIS	Environmental Research of the States
ESTCP	Environmental Security Technology Certification Program
FDEP	Florida Department of Environmental Protection
FE	fundamental error
FOT	fields of testing
GOF	goodness-of-fit
GPS	Global Positioning System

GSD	geometric standard deviation
GSE	grouping and segregation error
HDOH	Hawaii Department of Health
HMX	octahydro-1,3,5,7-tetranitro-1,3,5,7-tetrazocine
HTRW	Hazardous, Toxic and Radioactive Waste
IATA DGR	International Air Transport Association Danger Goods Regulations
IDW	inverse distance weighting
IS	incremental sample, incremental sampling
ISM	incremental sampling methodology
ISO	International Organization for Standardization
ITRC	Interstate Technology & Regulatory Council
LCS	laboratory control sample, laboratory control spike
LOD	limit of detection
LOQ	limit of quantitation
M	maps
MCA	Monte Carlo analysis
MCL	maximum contaminant level
MDI	method data indicator
MDL	method detection limit
MI	multiincrement
MIS	multiincrement sampling
MQI	method quality indicator
M_s	mass of the collected sample
MS	matrix spike
MSD	matrix spike duplicate
MSE	mean squared error
MSMA	monosodium methanearsonate
NELAP	National Environmental Laboratory Accreditation Program
NPL	National Priorities List
OE	overall estimation error
PAH	polyaromatic hydrocarbon
PCB	polychlorinated biphenyl
PCP	pentachlorophenol
PD	probability distribution
PE	preparation error
POC	point of contact
PQL	practical quantitation limits
QA	quality assurance
QAPP	Quality Assurance Project Plan
QC	quality control
QSM	Quality Systems Manual
RCRA	Resource Conservation and Recovery Act
RDX	1,3,5-trinitroperhydro-1,3,5-triazine
RL	reporting limit
RMSE	root mean square error

RPD	relative percent difference
RRO	residual-range organic
RSD	relative standard deviation
RSL	Regional Screening Level
SAP	sampling and analysis plan
SD	standard deviation
SE	sampling error, standard error
SERDP	Strategic Environmental Research and Development Program
SIM	selective ion monitoring
SOP	standard operating procedure
SRS	simple random sampling
SS	sample support
SU	sampling unit
SVOC	semivolatile organic compound
TE	total sampling error
TEQ	toxicity equivalent
TPH	total petroleum hydrocarbons
TPP	technical project planning
UCL	upper confidence limit
USACE	U.S. Army Corps of Engineers
USCG	U.S. Coast Guard
USEPA	U.S. Environmental Protection Agency
UST	underground storage tank
UTL	upper tolerance limit
UV	ultraviolet
VOC	volatile organic compound
WRS	Wilcoxon rank sum
XRF	X-ray fluorescence

Appendix G

Hyperlinks

NOTE: “Hyperlinks”

This guidance was developed as Web-based document. The blocks of information presented online as “Hyperlinks” are contained here in Appendix G.

HYPERLINKS

Hyperlink 1. Estimates of the Mean in Risk Assessment

Often, a small set of discrete samples is used to represent an volume of soil and to determine the mean concentration of the volume. However, since short- and microscale heterogeneities are typically not addressed, the variability in a discrete soil data set is often high, causing considerable uncertainty in whether the calculated mean is close to the true mean. To accommodate the uncertainty in estimating the mean from a small group of samples, a UCL on the mean is used to add conservatism to the estimate. The greater the variability in the data set, the wider the distance between the calculated mean and the UCL. USEPA risk assessment guidance (USEPA 1992) recommends the 95% UCL because it accounts for uncertainties due to limited sampling data and provides reasonable confidence that site means will not be underestimated.

Risk assessment is interested in the mean concentration over an exposure unit. *Risk Assessment Guidance for Superfund* (USEPA 1989b) states that the concentration term in the exposure equation is the mean concentration contacted at the exposure point or points over the exposure period. This point is reiterated in *Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites* (USEPA 2002a), which states that the concentration, commonly termed the "exposure point concentration" (EPC), is to be a conservative estimate of the mean chemical concentration in an environmental medium.

To adequately determine a mean concentration, more discrete samples would be needed than are commonly collected for discrete data sets. To accommodate the uncertainty in estimating the mean from a small group of samples, the 95% UCL on the mean is used to add conservatism. Too few sampling points create a wide interval between the calculated mean and the UCL. USEPA guidance states that data sets with fewer than 10 samples per exposure area provide poor estimates of the mean concentration (i.e., there is a large difference between the calculated sample mean and the 95% UCL), while data sets with 20–30 samples provide fairly consistent estimates of the mean (i.e., the 95% UCL is closer to the calculated sample mean). In general, the distance between the UCL and the calculated mean decreases as more samples are included in the calculation (USEPA 1992).

As is discussed throughout this document, ISM offers denser sampling coverage. ISM meets risk assessment goals for reliable estimation of the true DU mean, usually with more precision and less bias than for typical discrete sample data sets with far fewer sampling points.

Hyperlink 2. Nonrepresentative Data, Sampling Errors, and Decision Making

Consider an experiment where it is known that the true mean concentration of a large soil mass is 12.3 mg/kg. Assume an analysis was performed on a 1 g aliquot from a sample jar and the result was 12.3 mg/kg. The actual mass of contaminant present in the 1 g analytical subsample is 0.0123 mg. Now imagine that a particular 1 g analytical subsample from the same sample jar

captured a highly concentrated "nugget" made of organic carbon (red dot). The organic carbon nugget contains 0.1000 mg of contaminant mass in addition to the baseline 0.0123 mg. The contaminant mass in that 1 g subsample would be 0.1123 mg. When analyzed, the reported sample result would be 0.1123 mg/g, or 112.3 mg/kg, an order of magnitude different from the first result. The value of 112.3 mg/kg would be an erroneous result (a sampling error) caused by within-sample heterogeneity. It is not an analytical error because the analysis correctly measured the 0.1123 mg of contaminant present in that 1 g of soil. It is a subsampling error because the subsample did not contain the contaminant in the same proportions as the large soil mass.

Another factor is the concept of sample support is illustrated graphically in Figure H2-1. Imagine it was possible to extract and analyze 1 kg instead of 1 g of soil. Imagine also that the nugget was present in the kilogram of soil being analyzed. The baseline amount of contaminant present in the 1 kg analytical sample is 12.3 mg. Adding the mass of the 0.1000 mg nugget gives 12.4 mg, so the reported concentration is 12.4 mg/kg, quite close to the true concentration of 12.3 mg/kg. This scenario illustrates how strongly analytical subsample support affects reported concentrations. A nugget can cause much larger sampling errors when analytical subsamples are small.

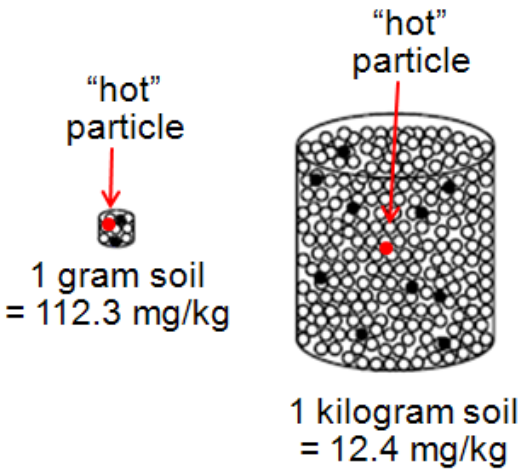


Figure H2-1. Illustration of the relative effect of a concentrated nugget on a 1 g sample vs. a 1 kg sample.

Hyperlink 3. Contaminant Behavior in Soil

Contaminants occur in many forms, including gases and vapors, liquids, single atoms or molecules, microscopic particles, and relatively large pellets or fibers. Some chemicals evaporate easily; others stay in the same place for a long time. The longer contaminants have been in the environment, the more likely they are to exist in forms that interact with soil particles. Some contaminants (e.g., metals) can be present in the environment as cations or anions. Ions with an overall positive charge (cations) adhere strongly to soil particles, especially particles made of clay minerals that carry negative charges. Polyatomic ions such as chromium and arsenic, on the other hand, have an overall negative charge (anions) and so may adhere weakly or experience electrostatic repulsion from clay minerals. When repulsive forces outweigh attractive forces, polyatomic ions become mobile in soil and groundwater, especially under conditions of changing acidity/alkalinity (pH) and oxidation-reduction potential (Eh). There are also weak atomic forces (i.e., Van der Waals interactions) that cause contaminants to adhere to the surface of soil particles even if no ionic charge attraction is involved. Van der Waals attraction is an important adsorption mechanism for organic contaminants.

Further complicating the picture, contaminants may be present as polar/nonpolar molecules (e.g., organic contaminants). "Polar" means that part of a molecule carries a partial negative charge, while another part of the same molecule has a partial positive charge. Many organic contaminants are hydrophobic (nonpolar), meaning that they do not dissolve well in water.

Organic contaminants may readily infiltrate the interstitial space (i.e., nooks and crannies) of organic soil matter and bind within the large hydrophobic surface area it provides. These organic compounds will tend to stay bound in the soil organic material rather than migrate in aqueous phase such as rainwater infiltration. Depending on the geochemical makeup of the soil, the organic matter content, the ambient conditions, and the physical and chemical properties of a particular contaminant, its molecules may bond loosely or tightly to soil particles.

Hyperlink 4. The Perils of Sampling Particulate Materials and Effects on Analytical Results

Figure H4-1 illustrates adsorption of contaminant to mineral grains Since iron (Fe) hydroxide is in a particulate form within the soil matrix, the contaminants sorbed to the iron hydroxide particle also behave as particulates. The light-colored material covering certain grains is arsenic. The concentration of a contaminant such as arsenic depends on how many particles of arsenic-laden iron complex are present within a "cleaner" silicate matrix. For the soil particle in Figure H4-1, the total mass of arsenic on the iron particles is miniscule (on the order of 0.005 µg), but the amount of total, mostly nonarsenic bearing, soil matrix is also very small (approximately 1 µg). Therefore, on a concentration basis (mass of contaminant per mass of soil) expressed in typical units, an analytical result would come out very high (approximately 5000 mg/kg). Figure H4-1 illustrates the important point that concentration is not the actual measure of exposure. The mass, the actual amount of contaminant that is present, is the measure of exposure. If a receptor ingested the 1 µg soil particle in Figure H4-1, the actual mass of arsenic ingested is only 0.005 µg. The concentration value for the particle, 5000 mg/kg, however, makes the exposure appear much worse than it really is.

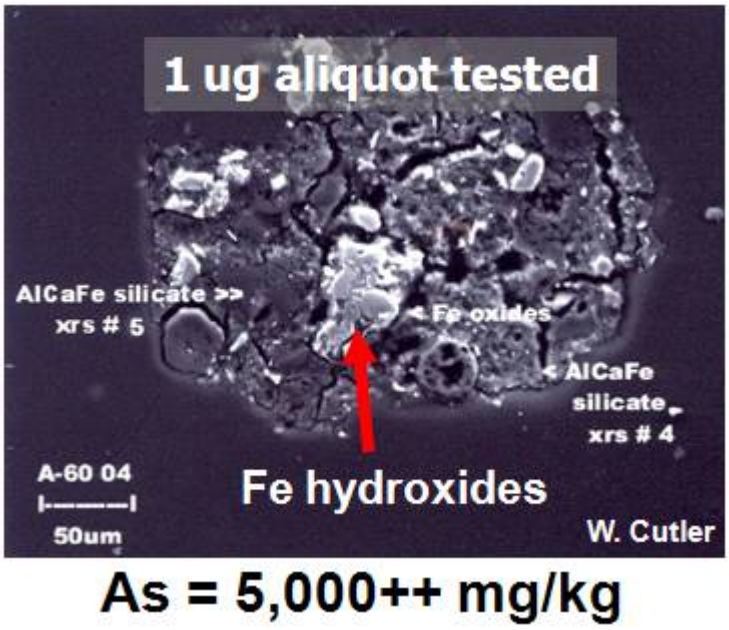


Figure H4-1. Arsenic-laden iron hydroxide particles within a "clean" silicate matrix. The total masses of the soil particle and arsenic are 1 µg and 0.005 µg, respectively. Photograph courtesy of Roger Brewer, HDOH.

Concentration can be very high while the actual *amount* of the contaminant is low. The reported concentration depends on the mass of the sample that is analyzed and how it was prepared prior to analysis.

Typically, sands bind less contaminant mass than other soil types. Their larger grain size limits their surface area, and the major mineral component (silica) in sand does not have the sorption ability of more clayey soils. Both organic and inorganic contaminants that do adhere to sand particles are relatively easily removed.

Organic carbon is another important component of soils that plays a large role in giving soil structure and in binding contaminants. A mature compost pile is an excellent example of soil material that is very high in organic carbon. Organic carbon aggregates absorb and concentrate hydrophobic organic contaminants into their matrix. Organic carbon is the primary food for soil microorganisms, and as it is cycled through the microbial community it ages until it turns into humic and fulvic acids. These organic acids are very persistent in soil and have a large surface area. Their molecules include many oxygen atoms carrying partial negative charges that attract contaminant molecules or atoms with an overall positive charge or with positively charged atomic groups. Organic carbon can be a very important soil component that sequesters contaminants.

The bottom line is that the measured concentration of a soil sample greatly depends on whether many or few contaminant-laden particles are present in the analytical subsample. A representative soil sample will have the same proportion of these particles as the targeted soil population. The smaller the sample or subsample, the more difficult it is to get the same proportion as in the parent material. As shown in Figure H4-2, a small sample (Sample A) taken from particulate material can miss concentrated particles that are scattered throughout the original matrix. This scenario leads to an analytical result that is less than the concentration of the original. On the flip side, there is also the chance that a small sample could capture contaminated particles (Sample C), but it is likely that the proportion of "hot" to "clean" particles will be different from the original. Because the ratio for Sample C is skewed in favor of the "hot" particles, Sample C would have an analytical result that is higher than the true concentration of the parent material. A larger sample (Sample B) has a much better chance of accurately representing the original.

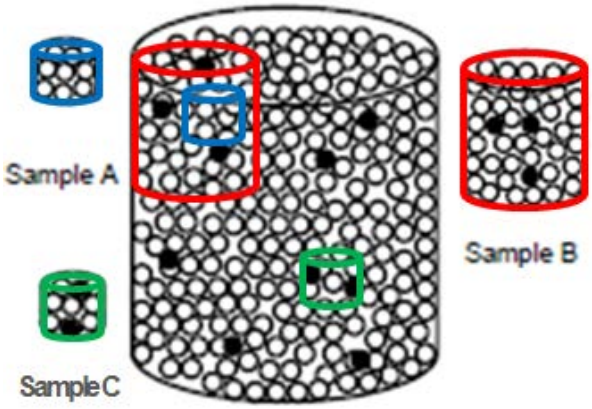


Figure H4-2. Illustration of the difficulties of collecting representative samples from particulate parent material.

Hyperlink 5. Sample Support

Defining the soil volume of interest requires knowing the spatial dimensions of a DU, including its area, depth, shape, and orientation in space. These spatial characteristics are collectively called the "support." Since these characteristics apply to the target for decision making, it is called a "decision support." The same spatial characteristics apply to samples taken from the decision support and are referred to as the "sample support." "Sample support" is the more general term, although the term "subsample support" might also be used as appropriate.

A criterion of sample representativeness is that the sample support must mirror the decision support. For example, if the depth of the DU is 12 inches, then the sample support depth should be 12 inches. If the area of the decision support is 5000 ft^2, then the samples should come from that same 5000 ft^2 area. The decision support is the basis for identifying the dimensions of DUs in ISM designs. The area and shape of the DU can be blanketed with increments so that the

sample mirrors the same dimensions of the DU. When a single sample is taken from the center of an area with discrete sampling, there is no natural sense of where the boundaries of the DU are, and there is no reason to assume that the sample is physically representative of some defined area. With ISM, the incremental sample is designed to closely match the DU.

Hyperlink 6. Why Can Spatially Close Samples Be So Different?

The nature and extent of contamination gets more complicated as time progresses and transport mechanisms change the distribution. Over the years, contaminants can be moved around by wind, flowing water, and disturbance of the soil during construction or landscaping activities. Even ecological receptors may change the pattern of contamination. For example, large and small animals and insects burrow or otherwise disturb soil. By their activities, they can take contaminated surface soil down vertically and bring cleaner subsurface soil to the surface or vice versa. It is not hard to imagine that the concentration of a small scoop of soil from one spot can differ greatly from the concentration of another small scoop taken yards, feet, or even just inches away.

Hyperlink 7. Field Study on Short-Scale Heterogeneity

The results of a field study presented in Figure H7-1 provide an example of short-scale heterogeneity for arsenic concentrations in a residential yard. To ascertain the degree of short-scale heterogeneity, linear groups of samples were taken, with samples located 1–2 feet apart. Within-sample heterogeneity was controlled so that the data results were true measures of short-scale heterogeneity. Figure H7-1 shows some of the data from that study. Sample locations are 1 foot apart.

Figure H7-1. Part of a variability study of short-scale heterogeneity for arsenic concentrations (mg/kg) in a residential yard.
Source: Unpublished data contributed by Deana Crumbling, USEPA.

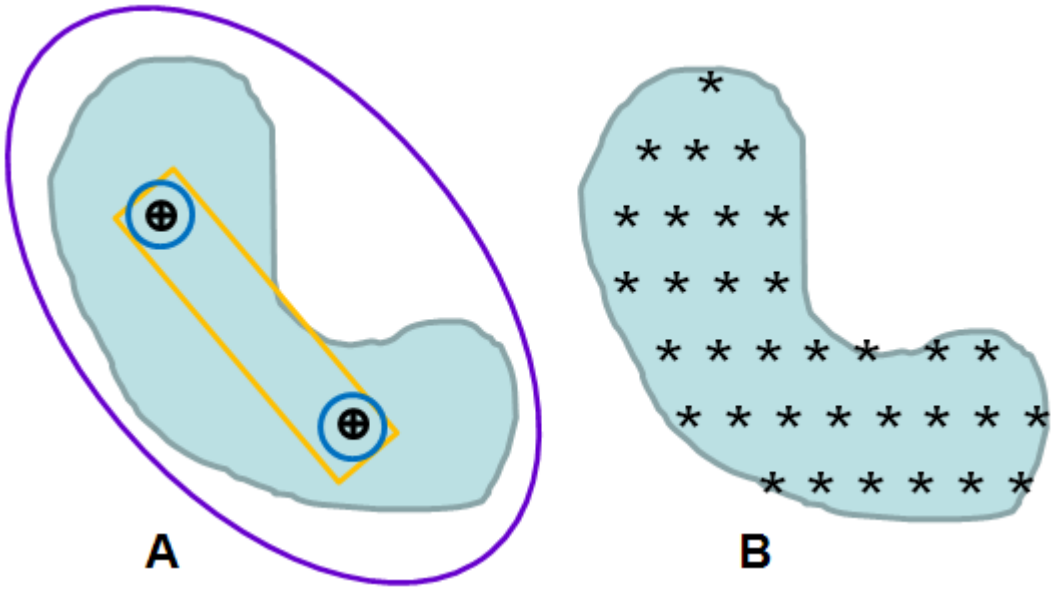
Hyperlink 8. Contaminant Analysis Does Not Actually Measure Concentration

In the process of soil analysis for contaminants, "concentration" itself is *not* actually "measured." Concentration is a derived property in that two other properties are measured and then concentration is calculated. Concentration is the mass of an analyte (which is measured by the analytical instrument) divided by the mass of the analytical subsample containing the analyte (which is measured when the analytical sample is weighed). The same mass of contaminant can be present in two different analytical samples, but if the mass of one of the samples is larger than the other, the larger one will be reported with a lower concentration than the smaller one. Contaminant concentration is a convenient measure, but it is actually the mass of contaminant that governs exposure severity and contaminant transport.

Hyperlink 9. What Volume Does the Sample Represent?

With a limited number of discrete samples, the reasonableness of data extrapolation can be questioned. ISM addresses these problems. The spacing of increment locations and the shape of DUs are designed during up-front systematic planning. Consider Figure H9-1 and suppose the

kidney-shaped light blue area portrays an area of surface soil for which a decision needs to be made. The area is on the order of 1/8 acre. In Scenario A, only two discrete samples are taken. In Scenario B, ISM is used. In Scenario A, the representativeness of the two discrete samples is unknown. Theoretically, one could say that each represents 1/16 acre. However, there is no objective reason to assume those samples represent the spatial area of interest. A practitioner could just as well say the two samples represent the rectangular area outlined in gold as say they

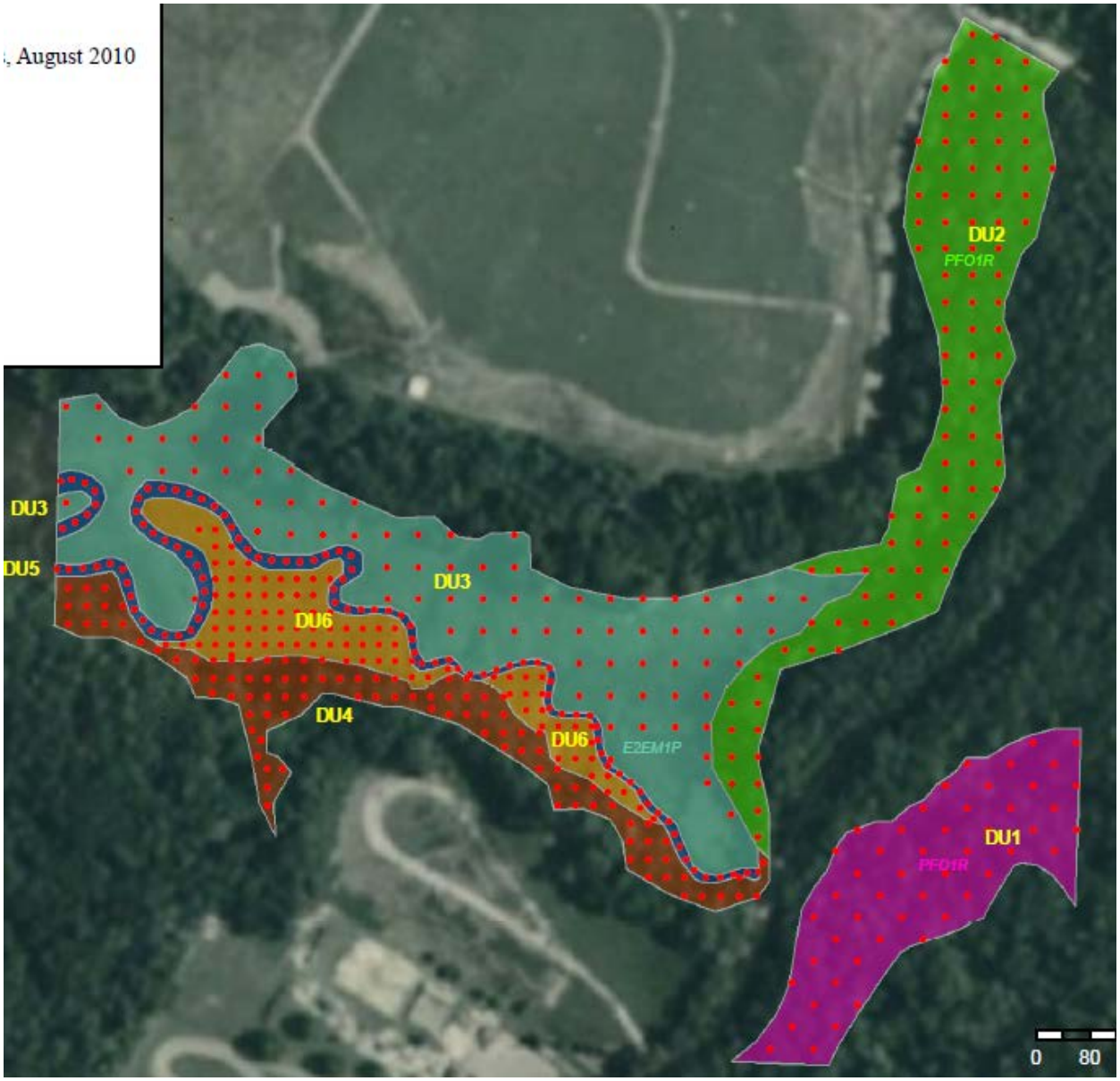


represent the oval area outlined in purple. If the two results are very different and significant short-scale heterogeneity is suspected or found, a cautious practitioner may feel that each sample result can only be trusted to represent a small area around its location (i.e., the small blue circles around the sampling points). Without more data, there is no basis for deciding how far extrapolation of sample results should go.

Figure H9-1. Discrete and ISM sampling approaches for an irregularly shaped DU.

In contrast, notice the intuitive appeal of Scenario B. Increments were collected from all over the area of interest and conform to the spatial boundaries of the area. When the increments are physically combined and averaged into the incremental sample, it is obvious that that single incremental sample represents the targeted area.

A real-world example is shown in Figure H9-2. For this site, DUs were selected to encompass particular hydric and soil type regimes. The proposed locations and density of increments are shown for each DU.



DU1 – 2.32 acre forested tidal wetland characterized by a predominately flat topography

DU2 – 3.49 acre forested wetland with tidal tributary associated with Bailey Creek moving through, characterized by moderately sloping topography

DU3 – 5.203 acre brackish tidal marsh north of Bailey Creek, flat topography with saltmarsh cordgrass, saltmeadow grasses and big cordgrass in higher elevation areas

DU4 – 1.62 acre forested upland buffer with steep changes in elevation

DU5 – 0.88 acre stream bed of Bailey Creek

DU6 – 1.346 acre brackish tidal marsh south of Bailey Creek, flat topography with saltmarsh cordgrass, saltmeadow grasses and big cordgrass in higher elevation areas

Figure H9-2. Example of multiple irregularly shaped DUs.

Hyperlink 10. Representativeness, Averages, and Populations

The term "representative" as used in this document has been previously defined. However, the term has been used in a variety of ways in other guidance documents. For example, other guidance has linked the concept of representativeness to a mean for a population (the "universe or whole") of interest. According to the Resource Conservation and Recovery Act (RCRA), a representative sample is defined in terms of an average: "Representative sample means 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 the whole" (Title 40 Code of Federal Regulations Part 260.10 [40 CFR §260.10], http://cfr.vlex.com/vid/260-10-definitions-19819125). In addition, the ASTM International (ASTM) Standard D 6044-96 defines 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 2009). Although it is nearly impossible for a single discrete soil sample to fulfill this expectation, an ISM sample, which more closely resembles a statistical sample, can.

A vital prerequisite for developing a sampling design is to unambiguously define the "population" and the characteristics of interest to be targeted by the decision-making process. A "population" is the entire set or total membership of entities that display some characteristic of interest. For soils, the population of interest is defined by location, spatial dimensions, and characteristics such as particle size. The population is the "whole" from which samples are taken to measure properties of interest. The population is what the samples are to represent.

Hyperlink 11. Testing for Nutrient Status in Your Yard

Suppose a homeowner wanted to know whether to apply lawn fertilizer. One approach would be to collect a small sample of soil from the center of the front yard and then put about a thimble-full into a nutrient test kit. A result of low nitrogen might then lead to the decision to apply fertilizer. But should it be applied to the entire front and back yards? How big an area did that one sample represent? A better approach would be to take samples from all over the front and back yard and mix them together before taking the thimble-full for the analysis step. In fact, that is the procedure recommended by the directions that come with soil test kits. A composite sample is preferable because (a) there may be different types of soil over the lawn area; (b) previous fertilizer applications were not distributed evenly, resulting in different nutrient levels in different parts of the yard; and (c) different plants and trees use nitrogen at different rates. A single sample provides information about only that location. It does not necessarily represent conditions in the rest of the yard. Chances are the yard is heterogeneous; one part is different from another part. Heterogeneity occurs at many different spatial scales, from microscopic, to the size of a yard, to thousands of times the size of a yard. Heterogeneity is the reason it is unwise to take just one or two soil samples and trust that the results apply to all the soil in an entire yard.

Hyperlink 12. Within-Sample Heterogeneity

The nonuniform distribution of particles (within-sample heterogeneity) poses a challenge for obtaining a representative subsample for analysis. There are three things that can be done in the laboratory to address this: increase mass of subsample analyzed, reduce the size of particles in the sample (controlling fundamental error), or increase the number of increments that compose the analytical subsample (controlling grouping and segregation error) by using the techniques described in Section 6. Fundamental and grouping errors are discussed in Section 2.5.1.

Obtaining a representative subsample in the face of within-sample heterogeneity involves additional labor, time, equipment, and cost. But the alternative can be data that do not stand up to scrutiny, that cannot be defended as reliable, and that lead to erroneous decisions about risk and remedial design.

The exact procedures selected to prepare samples always depend on the numerous project variables related to soil type, contaminants of interest, staffing, budget, availability of equipment, desired workflow, number of samples, subsequent sample preservation or preparation steps, etc. Details of these procedures are provided in Sections 5 and 6.

Increasing the mass of soil that is digested or extracted manages within-sample heterogeneity by reducing FE. Increasing the mass increases the likelihood that the same ratio of particle types will be present in the analytical sample as in the original sample (see Figure H4-1). It also reduces the "nugget effect," which produces very high concentration results when a "hot" particle is captured in a very small volume of "clean" soil particles. Because the soil mass used for typical metals analysis is very small (0.5–2 g, depending on the laboratory), the nugget effect can be particularly pronounced when samples are analyzed for metals. If it is not practical to increase the mass digested for metals analysis, then careful attention must be paid to sample preparation. Because analyses for organics typically involve 10–30 g of soil, the nugget effect is less of a problem than for metals; however, it can still occur. Therefore, thought needs to be given to what sample preparation is necessary to ensure an analytical sample that gives a concentration result representative of the matrix in the field.

The other way to address within-sample heterogeneity through reduction of FE is reduction of particle sizes through milling as discussed in Section 6.

Hyperlink 13. Measuring and Interpreting Sampling Variability

Measuring sampling variability (error) is not difficult. The necessary measurements are done as part of routine QC for environmental contaminant analysis. For the sake of this explanation, suppose the analytical portion of the sampling and analysis process has negligible error (i.e., is very precise). This is rarely true, but it is stipulated here for the sake of simplifying the illustration. In real life, the amount of variability on the analytical side can be determined from laboratory QC, and the analytical and sampling errors can be separated. Here we want to assume that the measurement variation comes only from subsampling error, as caused by within-sample heterogeneity.

Further suppose that the initial result of a laboratory duplicate pair was 12.3 mg/kg and the result for the duplicate analytical subsample gave a concentration result of 9.6 mg/kg. One way to measure this variation is by the RPD, which is calculated here as the difference between the two results (determined by subtraction) divided by the average of the two results, so that RPD between 12.3 and 9.6 is 25%. An RPD is the most common measure of precision when duplicates are involved. How do we know what an acceptable RPD is? Many times this is set arbitrarily at the beginning of a project. But there is another way to look at it.

An investigator may ask whether the level of data variability indicated by an RPD of 25% (using the value in this example) could cause a decision error. As described in other sections, whether a decision error is likely depends on (a) what the decision threshold is, (b) what the true mean is, (c) how much variability is present, and (d) the strategy for making decisions (i.e., whether decisions are based on a single sample result *or* on multiple results using the calculated mean or a UCL on the mean). Suppose for this example that the decision threshold is 100 mg/kg and the true mean is 12.3 mg/kg. If the variability present in the subsampling process is an RPD of ±25%, repeated analyses of subsamples will produce results that vary mostly between 9.6 and 15.8 mg/kg (although some individual results could fall outside these boundaries). Can that level of data variability cause a decision error if a decision were to be made on the result of a single analysis?

If the RPD were 25% around a true mean of 12.3 mg/kg and the decision threshold is 100 mg/kg, it is unlikely (although not impossible) to generate a decision error because there is little chance that any single result could be higher than 100 mg/kg. On the other hand, if the decision threshold is 15 mg/kg (rather than 100 mg/kg), the true mean is 12.3 mg/kg, and the RPD is 25%, it is quite possible for any single result to be higher than the action level and cause a decision error. If the consequences of a decision error were dire, an investigator might guard against the error by making multiple analyses and calculating the average of those analytical results. Alternatively, the investigator could decide to reduce within-sample heterogeneity, and thus the RPD, by reducing the sample particle size (e.g., milling the sample) so that any single subsample analysis is more likely to give results closer to the true mean of the sample. For a threshold of 15 mg/kg and a true mean of 12.3 mg/kg, the RPD needs to be reduced (i.e., the precision needs to be improved) to 20% or better.

Hyperlink 14. Select Gy Sampling Theory Equations

According to Gy's sampling theory, the overall estimation error (OE) is the sum of the total sampling error (TE) and the analytical error (AE):

$$OE = TE + AE$$

For each stage in the sampling and analytical process, TE is composed of a sampling error (SE) and a preparation error (PE), thus:

$$TE = SE + PE$$

The different sampling and preparation stages can be thought of as an initial sample event followed by subsequent subsampling events. For a typical environmental analytical process, there are two sampling and preparation stages, one occurring in the field and one occurring in the laboratory.

If there is subsampling in the field, or if more than one subsampling stage occurs in the laboratory, then each stage contributes a total sampling error component to the overall estimation error.

As described in Section 2.5, Gy recognizes seven basic sampling errors that comprise the total sampling error:

1. fundamental error (FE)
2. grouping and segregation error (GSE)
3. long-range heterogeneity fluctuation error (CE_2)
4. periodic heterogeneity fluctuation error (CE_3)
5. increment delimitation error (DE)
6. increment extraction error (EE)
7. preparation error (PE)

Thus:

$$TE = FE + GSE + CE_2 + CE_3 + DE + EE + PE$$

Note: It is actually the variances of the errors that are additive in the above equations, rather than the errors themselves. The total sampling error is a measure of how well one has controlled the various errors described above.

The errors are all derived mathematically by Gy and can be found in *Sampling for Analytical Purposes* (Gy 1998) and *Pierre Gy's Sampling Theory and Sampling Practice: Heterogeneity, Sampling Correctness, and Statistical Process Control* (Pitard 1993). To simplify this document for the average practitioner and regulator, the mathematical derivations have been omitted.

The overall estimation error can be determined through the collection of replicate samples. The differences between the field replicates (i.e., coefficient of variation [CV]) are an estimate of the OE. When laboratory replicates are analyzed, the TE for the analytical stages occurring in the laboratory can be estimated, and according to the formula above, the TE can be estimated by subtracting the total analytical error from the OE.

To correctly collect samples, as defined by Gy, all these errors must be addressed. In practice, the focus is usually on FE and GSE; however, the other errors can be important if correct sampling procedures are not used. The FE can be minimized by collecting sufficient mass of sample, and the GSE and SE can be minimized by collecting numerous increments.

The effects of sample mass and particle size are shown in the following equation for variance of the FE:

$$s^2_{FE} = c\beta fgd^3/M_s$$

where

M_s = mass of the sample
c = constitution parameter
β = dimensionless liberation factor
f = dimensionless shape factor
g = dimensionless size range factor
d = diameter of the mesh opening that retains no more than 5% of the sample

It is apparent from this relationship that the mass of sample necessary to minimize the FE is primarily controlled by the largest particle size of the population being sampled since this term is raised to the third power. The other factors can be thought of as constants since they do not have great variability in their values.

The constitution parameter, c, depends on the amount of the analyte of interest, a, in the lot and the mean density of the lot. If the amount of a in the lot is small, a <<1, then an approximation for c is given by $c = \delta_M/a_L$, where δ_M is the mean density of the lot and a_L is the decimal fraction of a in the lot.

The number of increments is a more complex derivation and is related to the magnitude of heterogeneity present at the site. If the total sampling error is high and the FE has been appropriately minimized, then it may be that the GSE is not being properly controlled.

One can compare the FE inherent in the usual practice of collecting discrete samples to the FE associated with using an ISM approach.

Typically, during discrete sampling the amount of soil collected in the field is enough to fill the required sample container for the specific analyte. For metals analysis and most organic analyses, the amount of soil in a 4-ounce container is adequate. In the laboratory an aliquot of 1 g is taken for metals analysis, while an aliquot of 30 g is taken for most organic analyses. Thus, there are two sampling stages: the first the field sampling and the second is the subsampling in the laboratory.

By using the following values for the parameters in the equation for the variance of the fundamental error: δ_M == 1.6 g/cm^3 (a typically density for soil), β = 1, f = 0.5, g = 0.25 (for unsieved soils), and d = 0.2 cm (from the definition of soil), one can solve for the variance of the FE. The mass of a discrete soil sample collected in a 4-ounce container would be about 180 g. Using an example concentration for the analyte of interest (action level) of 100 ppm (mg/kg) gives a value for $a_L = 1 \times 10^{-4}$. Thus, $FE_{(field)}$ = 30%.

Applying the same equation and values for the subsampling stage at the laboratory, for metals with a 1 g subsample, then $FE_{(lab)}$ = 400%. The overall variance of the FE is the sum of the

variances of the FE for each sampling stage; $FE_{(total)} = \sqrt{FE^2_{(field)} + FE^2_{(lab)}}$. Thus the total FE = 401%.

Applying the same calculations for organic analysis where the mass of the laboratory subsample is 30 g, one obtains $FE_{(lab)}$ 73% = and $FE_{(total)}$ = 79%.

In comparison, for ISM a 2 kg field sample is typically collected. Applying the same assumptions as above, one obtains a $FE_{(field)}$ of 9%.

If the sample is ground in the laboratory to 60 mesh (d = 0.0251 cm) and a 1 g sample is taken for metals, the $FE_{(lab)}$ = 18% and FE(total) = 20%. If the sample is instead ground to 100 mesh (d = 0.0152 cm), then $FE_{(lab)}$ = 8% and $FE_{(total)}$ = 12%.

For organic analysis if the sample is not ground, then $FE_{(lab)}$ = 73% and $FE_{(total)}$ = 74%. If the sample is ground to 60 mesh, then $FE_{(lab)}$ = 3% and $FE_{(total)}$ = 10%.

One can conclude from this analysis that the conventional practice of taking discrete samples results in a large FE, as is evident from the often-observed differences in the lab results between laboratory duplicates and field duplicates. By contrast, the techniques of ISM address the factors that lead to a large FE and ultimately result in less data variability.

Hyperlink 15. Making Decisions Under Different Concentration Scenarios

Consider a theoretical spatial distribution where a heavily contaminated volume of soil (DU 1) is surrounded by two successively less-contaminated volumes (DU 2 and DU 3) as shown in Figure H15-1.

DU 1—Heavy Contamination. DU 1 in this figure represents heavy contamination, where the true mean exceeds the action level and where further action is warranted according to the risk assessment results. In this DU, the concentration of the contaminant at any given point is well above the action level. Therefore, even a small number of discrete samples are likely to indicate the correct decision ("mean > action level"). Note how far both the mode (the most frequently observed sample result) and the mean are from the action level in histogram 1 on the figure. Of course, a much larger number of discrete samples would be required for this DU if the decision required a mean estimate with a higher degree of precision than that needed by a simple "clean-dirty" decision.

DU 2—Moderate Contamination. The reliability of a decision using a small number of discrete samples changes drastically when the true mean concentration is closer to the action level, as shown in DUs 2 and 3 of the figure. In this situation, the precision with which the mean concentration is estimated, as well as the effects of the right-tail of the distribution, are more important. In histogram 2 of the figure, the mode is below the action level, but the mean is above the action level. Because the mean concentration is greater than the action level, a correct decision for this DU would result in remedial action. However, note that because the mode is

lower than the action level, the majority of discrete samples will likely contain concentrations *below* the mean. Therefore a low-density discrete sampling approach in this DU would usually be expected to mislead the investigator into thinking that the DU is “clean” when it is not. The more samples collected, the more likely it is that one or more will exceed an action level (and even exceed the true mean). So, the more discrete samples collected, the closer the estimate of the mean is to the true mean.

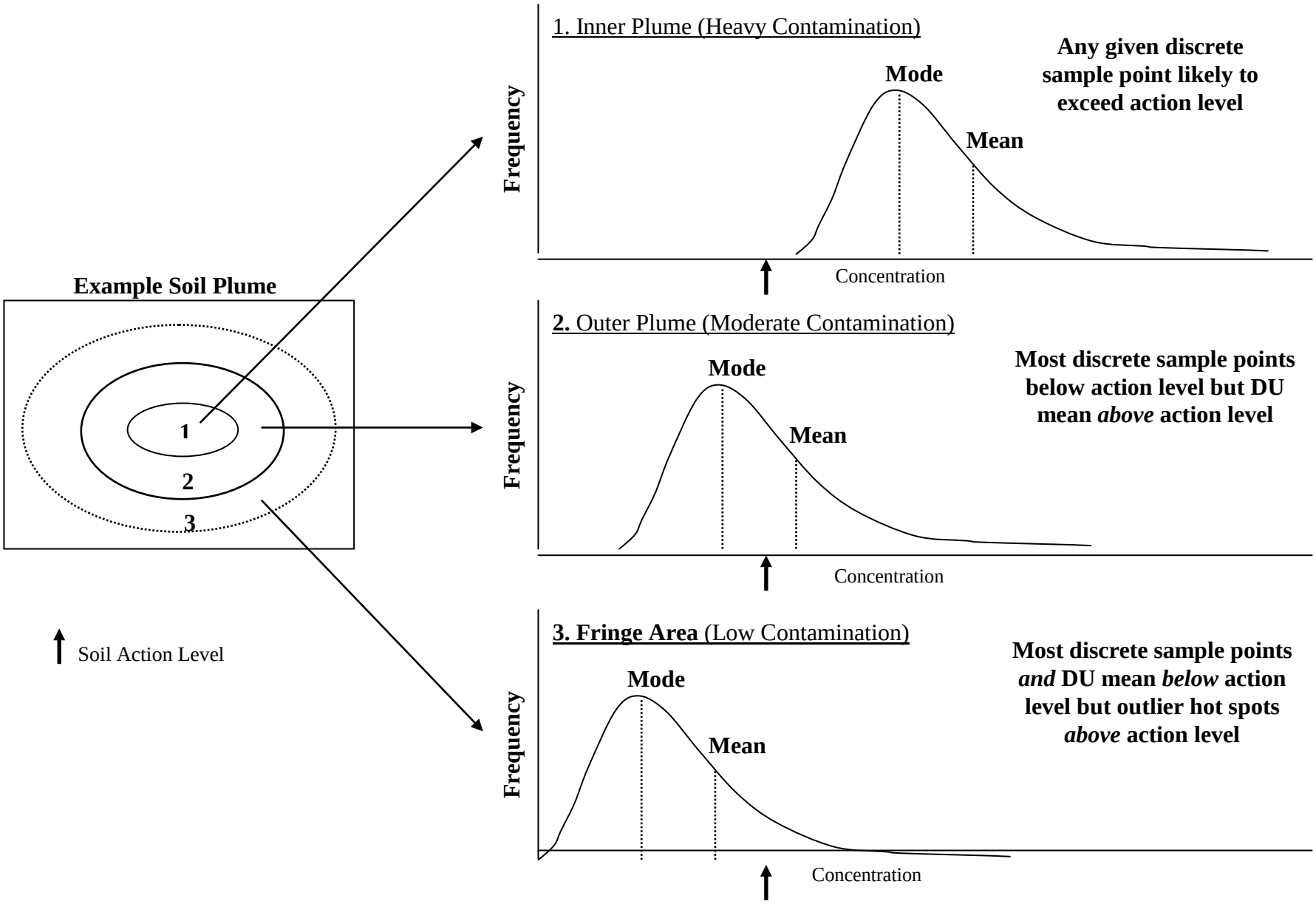


Figure H15-1. Implications of lognormal distribution at varying levels of contaminant concentrations in soil.

DU 3—Light Contamination. DU 3 contains low concentrations. Note that in histogram 3 of the figure, both the mean and the mode fall below the action level. No action would be required in this DU from a risk-based perspective, as this DU would be considered "clean." However, this does not mean that the concentration for every discrete soil sample potentially collected in DU 3 will be below the action level. It is quite possible that, due to the right-tailed lognormal distribution of histogram 3, some discrete samples collected in DU 3 will contain concentrations above the action level. If enough discrete samples are collected, such an outcome is even expected. In this situation, discrete data sets may be characterized by isolated, sample-size hot spots that do not represent meaningfully sized volumes of soil and do not pose a significant risk to human health or the environment.

As is apparent from the histograms in the figure, with enough samples, the pattern of contamination in DU 3 will be very different from that in DU 1, where most sample points will be above the action level, or DU 2, where a distinct clustering of elevated sample points is likely. However, because insufficient numbers of discrete samples are generally relied on, investigators are frequently unable to distinguish which situation is actually present. In the face of this uncertainty, they make the conservative assumption that the mean concentration in a DU will be represented by the maximum encountered concentration. As shown by this example, however, the maximum concentration from a small set of discrete samples can underestimate or overestimate the true mean, sometimes by a large magnitude.

Hyperlink 16. To What Volume of Soil Do Discrete Samples Apply?

While the volume of soil represented by ISM samples is defined by the demarcation of the DUs, to be useful, discrete samples must also apply to some volume of soil in a decision. However, the volume of soil to which discrete samples apply is often ambiguous, determined after samples are collected and analyzed, and subject to the vagaries of the allocation approach used. Vastly different volumes of soil will be determined based on the allocation approach used. Some of the various allocation approaches are provided below.

- The sample applies to some arbitrary volume of soil determined through agreement, regulatory requirement, or professional judgment.
- The sample result is assumed to apply to all the soil until the next "dirty" or "clean" discrete sample is encountered.
- A sample applies to some volume of soil as in a tacit or actually calculated Thiessen polygon. Thiessen polygons are polygons which indicate the "area of influence" of a single point in relation to all other sampled points. They are determined by the perpendicular bisector of the lines between all neighboring points. Thiessen polygons may be calculated; however, it is it more often that the concept of "areas of influence" is tacitly used when discrete samples are employed to determine where samples are located and what their results mean in the assessment of volumes of soil.
- Discrete samples are used to populate iso-concentration maps where groups of samples collectively represent the concentration of contaminants in large volumes of soil based on their concentration, weighting, and the algorithm employed.
- The mean of a number of discrete samples collectively applies to some volume of soil.

Each of these approaches potentially results in different decisions over vastly different volumes of soil when using the same discrete data.

Hyperlink 17: Particle Size Selection

Particle size selection must be addressed during project planning. Selecting the target soil particle size is driven by the project decision(s) to be made on the data. Certain decisions may call for the analysis to be performed on bulk soil. Generally, bulk soil refers to all particle sizes <2 mm. This means preparing the soil such that the ratios of all particles <2 mm in the sample reflect the same ratios as in the field. Then, that same ratio must be achieved for the subsample that will be analyzed. Other project decisions may require targeting a different particle size. For example, exposure assumptions may require knowing the concentration of dust-sized particle (see *Short Sheet: RTW Recommendations for Sampling and Analysis of Soil at Lead Sites* [USEPA 2000c]). That constrains a sample to those particles sizes that represent dust, and sizes larger than that must be removed from the sample. Sample preparation procedures cannot be selected until the target particle size has been identified by upfront systematic project planning.

Hyperlink 18. Calculating the Mass of Sample Needed to Achieve a Specific Fundamental Error

One can estimate the mass of a sampled needed to achieve a specific fundamental error (FE) by applying Gy theory equations that relate the mass of the sample, diameter of the largest particle, and the variance of the fundamental error.

A simplified Gy theory equation useful for this calculation is as follows (Eq.1):

$$s^2_{FE} = Cd^3/M_s \quad (1)$$

where

M_s = mass of the sample
d = diameter of the mesh opening that retains no more than 5% of the sample
s^2_{FE} = variance of the fundamental error
C = sampling constant

Rearranging to solve for the mass of the sample one obtains the following (Eq. 2):

$$M_s = Cd^3/s^2FE \quad (2)$$

A common value for d is 0.2 cm, as the media of interest usually is soil, which is generally defined as those particles <2 mm in diameter. The desired FE is often selected as 15%, as one would like to keep the overall error fairly low. If there are several sampling stages, then each stage contributes an FE to the overall error. Note that each sampling stage also contributes to the other six sampling errors that compose the overall error, so it is desirable to keep these errors to a minimum also.

The sampling constant is often given a value of 22.5. However, one should be aware of the assumptions used in obtaining this value as well as the assumptions used to generate equation (1).

Using these values, the calculated mass of the sample is 8 g.

The most problematic assumption from an environmental perspective in using a sampling constant of 22.5 is that this assumes that the concentration of the analyte of interest is in the percent range. For environmental sampling where the concentrations of interest are on the order of parts per million (ppm), then the sampling constant will be about 200,000.

Using this sampling constant (2×10^5), the calculated mass of the sample would be 71 kg with the same FE and particle diameter as above.

The equation with the sampling constant broken out into its components is as follows (Eq. 3):

$$s^2_{FE} = c\beta fgd^3/M_s \qquad (3)$$

where

M_s = mass of the sample
c = constitution parameter
β = dimensionless liberation factor
f = dimensionless shape factor
g = dimensionless size range factor
d = diameter of the mesh opening that retains no more than 5% of the sample

Guidelines for the values of these parameters are given by Gy (1998):

> The constitution parameter, c, depends upon the amount of the analyte of interest, a, in the lot and the mean density of the lot. If the amount of a in the lot is small, a $<<1$, then an approximation for c is given by $c = \delta_M/a_L$, where δ_M is the mean density of the lot and a_L is the decimal fraction of a in the lot.
>
> The dimensionless liberation parameter, β, can have values from 0 to 1. The parameter is 0 when the components are completely homogenized (an impossible situation) and is 1 when the components are completely liberated. It is best to set $\beta = 1$ if one is not certain of the state of liberation.
>
> The dimensionless shape parameter, f, also can have values from 0 to 0. For a sphere f = 0.52. For most compact particles f has values near 0.5.
>
> The dimensionless size range parameter, g, also can have values from 0 to 1. Some values used in practice are:

- Undifferentiated, unsized materials, mean value g = 0.25
- Undersized material passing through a screen g = 0.40
- Oversize material retained by a screen g = 0.50
- Material sized between two screens g = 0.6/0.75
- Naturally sized materials, e.g., cereal grains g = 0.75
- Uniformly sized objects, e.g., bearing balls g = 1.0

Rearranging Eq. 3 to solve for the mass of the sample and substituting δ_M/a_L for c yields the following (Eq. 4):

$$M_s = (\delta_M/a_L)\beta f g d^3/s^2_{FE} \quad (4)$$

For this example, suppose that a DQO has been established limiting FE to 15%. By using the following values for the parameters in the equation for the variance of the fundamental error: δ_M = 1.6 g/cm^3 (a typically density for soil), β = 1 (as suggest above), f = 0.5 (also as suggested above), g = 0.25 (for unsieved soils), and d = 0.2 cm (from the definition of soil), one can solve for the mass of the sample for anticipated situations (Eqs. 5/6):

$$M_s = (1.6/a_L)(1)(0.5)(0.25)(0.2)^3/s^2_{FE} \quad (5)$$

or

$$M_s = (1.6 \times 10^{-3})/(a_L s^2_{FE}) \quad (6)$$

Therefore, for a desired FE of 15% and $a_L = 1 \times 10^{-6}$, the mass of the sample needs to be 71 kg.

The assumption made in both Eqs. 1 and 3 is that the mass of the sample is much less than the mass of the population or lot, M_L. The equation both are derived from is as follows (Eq. 7):

$$s^2_{FE} = \left(\frac{1}{M_s} - \frac{1}{M_L}\right) c\beta f g d^3 \quad (7)$$

where

M_s = mass of the sample
M_L = mass of the lot
c = constitution parameter
β = dimensionless liberation parameter
f = dimensionless shape parameter
g = dimensionless size range parameter
d = diameter of the largest particle

This assumption is probably true for field sampling but may not be true for laboratory subsampling.

When using any of these equations to determine the mass of the sample, it is important to note that the values obtained are approximate (on the order of magnitude of the mass needed). The actual mass of sample needed to achieve a specific fundamental error may be greater than that calculated.

One can experimentally determine whether the mass calculated is sufficient by analyzing at least 10 replicate samples. If the variance of the results is less than the variance of the desired fundamental error, then the mass of the sample is sufficient. If the variance of the results is greater than desired, then overall sampling and analysis process must be reexamined. Some part of the process—field sampling, laboratory subsampling, processing, analysis, etc.—is not in control and must be corrected to achieve the desired variance.