

Fake Data? The Need for Theory of Sampling Concepts in Environmental Research and Investigations

*R. Brewer^a, C. Ramsey^b, M. Heskett^c, J. Song^d,

^aSenior Geologist/Environmental Scientist, Hawaii Department of Health, 2785 Waimano Home Road, Suite #100, Pearl City, Hawai'i 96782, United States of America. Email: roger.brewer@doh.hawaii.gov

^bPresident, EnviroStat, Inc., P.O. Box 339, Vail, Arizona 85641, United States of America. Email: chuck@envirostat.org

^cSenior Scientist, Element Environmental, 98-030 Hekaha St Unit 9, Aiea, Hawai'i 96701, United States of America. Email: mheskett@e2hi.com

^dProfessor, Institute of Soil Science, Chinese Academy of Sciences, No. 71 East Beijing Road, Nanjing, Peoples Republic of China, Post Code: 210008. Email: jingsong@issas.ac.cn

Abstract: The sampling intensive mining and environmental industries share a common need for representative data but differ in the motivations to accomplish this objective. The desire to obtain representative sample data for “commodities” in the mining industry is driven by anticipated economic gain and the exploitation of natural resources. The desire to obtain representative sample data for “contaminants” in the environmental industry is driven by anticipated social gain and the protection of natural resources. In terms of obtaining reliably representative data, motivation driven by economic gain has thus far been the clear winner. Theory of Sampling concepts are well established and tested in the mining industry. The environmental industry, in contrast, has traditionally been plagued by scientifically unsound sampling practices and data that are not reliably representative of conditions in the field. This has significant implications for topics ranging from the efficient identification and remediation of contaminated industrial lands to the accurate assessment of risk to human health and the environment.

This paper explores the nature and cause of this dichotomy and presents a methodical approach for application of Theory of Sampling concepts to environmental testing of soil, water and air. Much of the problem is tied to a general recognition of compositional and distributional heterogeneity in contaminated media but unawareness of a method to control it or an understanding of the magnitude of potential error. As a result, published regulatory guidance focused on classical sampling and statistical methods appropriate for testing of “finite element” media. A lone exception is testing of indoor air, where concepts of “Decision Units” and sampling methods more appropriate for testing of “infinite element” media have long been employed to control and represent heterogeneity.

The solutions are, in hindsight, relatively simple. Pushback from affected parties and even scientists and environmental agencies can be significant, however. This is primarily due to a lack of training of environmental professionals in the Theory of Sampling and the common absence of clear evidence of erroneous or misleading sample data in the field. Reluctance to change is also tied in some cases to implications regarding liability for past and ongoing projects. The need for more reliable, efficient and science-based methods to assess and address risk posed by environmental contamination is clear, however. Progress will be made by countries like China that are beginning to tackle legacies of early development and are able to learn from the successes as well as the mistakes of countries that have been addressing environmental contamination for several decades. Training of environmental workers as well as pressure from liability-savvy responsible parties, attorneys and financial institutions will continue to force the industry to evolve, to the benefit of the environment as well as stakeholders on all sides.

Introduction

Similar stages of investigation are common to both the mining industry and the environmental industry in their efforts to identify and take action on anomalously high concentrations of chemicals in the environment (Figure 1). The process begins in the mining industry with geologic mapping or ore bodies and ends with extraction and marketing of the targeted “commodity.” The process begins in the environmental industry with preparation of a Conceptual Site Model (CSM) of the extent and magnitude of impacts and ends with destruction or isolation of the targeted “contaminant.” In each case, samples are collected, processed and tested and the resulting data used to make decisions regarding the economic extraction of the commodity or the risk to human health and the environment posed by the contaminant.

The success of both industries relies on the efficient and reliable collection of representative sample data. In the mining industry, extraction and measurement of the true mass of the commodity present in crushed ore allows the accuracy of the data and the sampling method employed to be routinely tested. Economic and business repercussions for data error are severe and quickly realized by those responsible for directing the collection of sample data. Such economic incentives combined with an intense dedication to the science of sampling led to a progressive evolution and improvement in the

collection, processing and testing of samples that ultimately came to be referred to as Gy's Theory of Sampling^{1,2,3,4} (see also 5,6).

A correlative pressure to identify and address sampling error in the environmental industry has been lacking, due in large part to two factors: 1) Lack of routine extraction and direct measurement of the mass of the contaminant present in the targeted media which would directly identify sample error and 2) Oversight and direction of sampling methods by government agencies largely shielded from repercussions of erroneous or misleading data and decision making. Some degree of "third-party" verification of data reliability exists in operations to extract volatile contaminants from soil or to otherwise treat and destroy contaminants in situ. Remediation experts typically over engineer the design of a remedial action at significant added cost in order to address uncertainty in the sample data. Discrepancies in the mass of contaminant extracted and the mass initially predicted by sample data is often attributed to "laboratory error," however, rather than to mistakes in the field.

The disconnection of government agencies from unnecessary legal and economic damage inflicted on regulated entities has hindered serious, scientific review of the nature and source of errors and improvement of sampling methods. Basic concepts of Gy's Theory of Sampling and related sampling methodologies used in the mining industry since the 1950s can, however, help address lapses in the environmental industry and improve the reliability of data used in research and routine investigations.

Traditional Environmental Sampling Methods

Testing of Soil, Water and Subsurface Vapors

Soil and water sampling guidance was initially published by the United States Environmental Protection Agency (USEPA) in the 1980s, following passage of the Resource Conservation and Restoration Act^{7,8,9,10} (see also 11). The Superfund program and other offices within the USEPA were tasked with developing guidance for characterization of the extent and magnitude of contamination in soil, sediment and other media. Concurrent guidance^{12,13,14} for quantitative assessment of health risk emphasized the need to estimate a mean contaminant concentration for a targeted "Exposure Area".

A detailed overview of the evolution of soil and sediment sampling guidance in the environmental industry is provided by Brewer et al^{15,16}. Early environmental guidance for site characterization focused on the collection and individual testing of small, discrete samples of soil and water with the mass or volume of the sample dictated only by the minimum mass required by the laboratory for testing. This typically required the collection of only 100 to 300 grams of soil and a few liters to as little as 40 milliliters of water. This is likely due to experience of the early authors with testing of relatively uniform, industrial wastewater. Concentrations of contaminants within relatively small volumes of water can be assumed to be consistent provided that operations within the manufacturing facility do not change.

Reliance on testing of small-mass or small-volume, discrete samples from a single point characterize contaminant distribution within a targeted media necessarily requires an assumed lack of significant, variability over very short distances and even within the same sample. This assumption is clearly stated in guidance for environmental testing of soil and sediment:

"The implicit assumption that residual contamination is equally likely to be present [at similar concentrations] anywhere within the sampling area is reasonable"⁷.

"To apply this [discrete sampling] method, the following assumptions are required... The shape and size of the contaminated zone must be known at least approximately. Any sample located within the contaminated zone will identify the contamination. These assumptions are not severe and should be met in practice"⁸.

"When there is little distance between points it is [assumed] that there will be little variability between points"¹⁴.

Distributional heterogeneity as well as compositional heterogeneity within samples of soil and other particulate matter is largely ignored at the laboratory. Minimal to no processing of samples is required before collection of a subsample for testing. Subsamples are collected at random from a single point within the sample submitted. The mass of subsample is intentionally minimized to the extent possible in order to minimize reagents and waste and to shrink benchtop automated systems, with as little as half a gram of unground material tested in some cases. Validation of sample data quality only focuses on the reproducibility of the laboratory method itself rather than the representativeness of the subsample tested. Discrepancies between duplicate laboratory subsamples or field samples is normally resolved by simply using the higher concentration.

Warnings were raised within the USEPA regarding the unreliability of individual discrete sample data points for determining the extent of contamination in the field, including¹⁷ (see also 18,19):

"Grab samples' or judgmental samples lack the component of correctness; therefore, they are biased. The so-called grab sample is not really a sample but a specimen of the material that may or may not be representative of the sampling unit. Great care must be exercised when interpreting the meaning of these samples."

The fact that the prescribed sampling approaches for soil might prove in the future to be unreliable was not lost to the authors¹⁴:

“Future changes in EPA policy may invalidate some of the discussions in this (guidance).”

Detailed field and laboratory studies have indeed demonstrated significant and random distributional and compositional heterogeneity in contaminated soil as well as other particulate matter^{15,16}. High variability between sample data hinders the precision of classical statistical tests to estimate a mean contaminant concentration for assessment of risk. In order to address this issue, environmental risk assessment erroneously recommends that high-concentration “outlier” data be ignored²⁰. Error in sample data collected and tested in this manner is, as a result, unknowable. The potential for significant underestimation of the mean is apparent, however, by the common need to over engineer *in situ* treatment systems by experienced remediation experts in order to accommodate an anticipated order-of-magnitude or more actual mass of contaminant present than predicted by discrete sample data.

Guidance documents for testing of groundwater⁹ (e.g., to assess drinking water aquifer risk) and guidance for testing of vapors from volatile chemicals immediately beneath the slabs of buildings (“subslab vapors”)^{21,22,23} (e.g., to assess vapor intrusion risk and impacts to indoor air) similarly rely on an undemonstrated assumption of negligible, distributional heterogeneity over short distances and recommendations for individual testing of widely-spaced small-volume, discrete samples:

“(Groundwater) samples taken in close proximity (e.g., within a few meters)... are highly auto-correlated. ...Dense spatial monitoring... runs the risk of redundant data”⁹;

“The default (indoor air:subslab vapor attenuation factors) assume (that) ...the subsurface (vapor plume) is reasonably homogeneous”²².

Detailed field studies of the nature of random, distributional heterogeneity of contaminants in groundwater have never been carried out to the knowledge of the authors. Such localized variability is inevitable, however, given the inherent heterogeneous nature of soil. The variability of contaminant concentrations in subslab vapor plumes over very short distances is, in contrast, well documented^{24,25}. As stated by Luo et al.²⁶ in reference to the use of single samples to estimate attenuation factors:

“Random sampling of a few locations might not reveal the true range of concentrations... Even if one had precise knowledge of the subslab soil-gas distribution, it is not clear how it would be used to assess pathway significance without knowledge of the vapor entry points to the building and soil-gas entry rates through those points.”

An especially egregious example is USEPA’s “Empirical Database” of the attenuation of intruding, subslab vapors in indoor air²⁷. Subslab vapor and indoor air data were compiled for hundreds of structures. Data for a single, small-volume (typically one liter) sample of subslab vapors was compared to data for indoor air and used to generate an indoor air:subslab attenuation factor for the structure. Apparent attenuation factors calculated from the database were spread over five orders of magnitude, even though the structures involved were reasonably similar. Statistical tests were carried out on the database in an attempt to make sense of the data, with the 95% Upper Confidence Level of the mean selected for use in regulatory guidance. The total error associated with the USEPA database is impossible to quantify and the proposed attenuation factor scientifically invalid. Discrepancies between impacts to indoor air predicted by the default attenuation factor and actual field measurements led some workers to simplistically declare that the physics involved in vapor intrusion was “wrong,” rather than consider that the data collected were in fact unrepresentative.

The potential for localized and random, distributional heterogeneity of contaminants within the targeted area and volume of surface water is recognized in sampling guidance²⁸:

“The probability is small that any body of still water (lake, reservoir, pond, lock, storage pool) is relatively homogeneous with regard to any water-quality characteristic. Therefore, a single sampling point generally is not adequate to describe the physical and chemical properties of the water body, or the distribution and abundance of the inhabiting biological community.”

The collection and individual testing of discrete samples was still called for, however²⁸:

“Still-water samples generally also are collected at multiple locations in the water body and at multiple depths...”

The number of sampling locations selected and the depths where samples will be collected depend on study objectives and the physical, chemical, and biological characteristics of the water body.”

Statistical tests were to be used to estimate mean contaminant concentrations for targeted bodies of groundwater or groundwater extracted from a single well. In the absence of replicate sets of sample data, however, the representativeness of a single set of discrete sample data for a targeted volume of water as a whole will remain uncertain.

Such lapses did not go entirely unrecognized by early environmental professionals. A small number of scientists within the USEPA pushed internally for consideration of sampling methods familiar to the mining industry at contaminated sites soon after initial guidance was published but with limited success^{29, 30 (see 16)}. As discussed by the authors of USEPA guidance for the collection of representative subsamples at the laboratory³⁰:

“It soon became clear that sampling is perhaps the major source of error in the measurement process, and, potentially, sampling [including the collection of laboratory subsamples] is an overwhelming source of error for

heterogeneous particulate materials, such as soils. It was also clear that classical statistical sampling theory was not adequate for such samples... We knew that geostatistics offered some answers, such as the sample support (mass, volume, and orientation) and particle size (diameter) make a difference. That is only common sense... We searched for a non-conventional statistical sampling theory that actually takes into account the nature of particulate materials and, in 1989, we hit "pay dirt," ...the Pierre Gy sampling theory for particulate materials. ...Over the ensuing years, we were only moderately successful at transferring this technology to the environmental community so that it might be implemented."

These and other examples of data error in the environmental industry's reliance on discrete sample data for decision making have caused significant financial and legal hardships for entities tasked with carrying out investigation and remedial actions, including:

- Repeated need to collect additional sample data with no clear end point;
- Significant time delays and cost overruns for site characterization and remediation;
- Liability associated with contamination discovered on supposedly remediated properties; and
- Unnecessarily abandoned properties sites later found to still be contaminated.

With only a few exceptions, this struggle to incorporate sound science into the environmental industry continues thirty years later but in recent times has made progress.

Testing of Indoor Air

An exception to the above noted problems in the environmental industry is traditional method used to test indoor air. Although not specifically described as such, Theory of Sampling type concepts have long been incorporated into testing of indoor air quality. A typical Investigation Question might be: "Does the mean concentration of contaminant 'X' in the air circulating through this room during time period 'Y' exceed screening level or target risk level 'Z'?" The volume targeted for sample collect is calculated as the volume of the indoor space times the measured or estimated indoor air exchange rate times the designated time period. For example, 2,050,000 liters of air are predicted to circulate within a default, single family home with a volume of 244 cubic meters³¹ and a minimum-recommended indoor air exchange rate of 0.35 times per hour³² over a 24-hour time period. Risk is assessed by comparison of the mean concentration of the target compound in the air to a screening level pertinent to the exposure scenario or by incorporating the data into a human health risk assessment.

The collection and testing of single or multiple, instantaneous, discrete samples of indoor air is rarely if ever done. The sampling device is instead placed within the targeted indoor area and allowed to collect a continuous sample throughout the designated exposure period. Air flowing past the device is assumed to be representative of air circulating within the DU as a whole. One of three methods is typically used to collect a sample: 1) Whole air sampling; 2) Sorbent tube sampling and 3) Passive sampling using sorbent materials. The first two methods are referred to as active sampling³³. Whole air sampling involves the collection of air in a sample container over a specified time interval, such as a Summa canister or a Tedlar bag, and analyzing the gas from that container directly³⁴. In the second method, a pump or syringe is used to pull air through a tube containing a sorbent material³⁵. The concentration of the chemical in air is determined by dividing the mass of the chemical sorbed to the material by the volume of air pulled through the tube. Passive sampling involves placing an adsorbent material in the targeted DU area and allowing compounds in the air to sorb to the material over the targeted time period without the use of a pump or a Summa canister³⁶.

This approach to testing of indoor air fits well into Theory of Sampling concepts regarding the collection of a single sample that is reliably representative of a very large volume of infinite element media. The sample is normally collected over a time period of one to five days and during hours when the structure is occupied (e.g., 24 hours a day for residences and during working hours for commercial buildings). The resulting data are assumed to be representative of long-term, indoor air quality. Replicate samplers can be placed in independent areas of the targeted indoor area and/or on different days in order to test the precision of the data collected. Similar concepts of the collection of small samples that are directly representative of very large volumes of environmental media can also be applied to soil, water and even subsurface vapors.

Application of Mining Industry Concepts to the Environmental Industry

Systematic Planning

Detailed guidance on the application of Theory of Sampling concepts to testing of soil, sediment, water and subsurface vapors has been in use by the State of Hawaii for over ten years³⁷. Guidance specific to the investigation of contaminated soil has recently been published by the Chinese Academy of Sciences³⁸. The following presents a summary of these and related guidance documents.

Careful, systematic planning of an environmental investigation is necessary to ensure that the data collected reliably address the investigation objectives. The process centers on a basic but methodical series of questions applicable to any

type of investigation:

- What is the media of interest?
- What is the contaminant(s) of potential concern (COPC)?
- What is the specific Investigation Question?
- What is the Decision Unit(s)?
- How can a representative sample of the DU be collected?

A Decision Unit (DU) is defined as the area and volume of targeted media about which a decision is to be made^{37,38}. In the environmental industry, such decisions typically reflect the assessment of risk or optimization of remedial actions. Efficient designation of DUs for sample collection requires the development of specific Investigation Questions beyond the simplistic determination of the presence or absence of contamination, for example: “Does the concentration of ‘X’ in the media ‘Y’ exceed a target screening level or health risk of ‘Z’?” This is most commonly assessed based on comparison of data representative of the mean concentration of the COPC for the DU to a published or site-specific, risk-based screening level. If an adverse risk is already known, then the Investigation Question might be “What is the location of the main mass of contamination that will require remediation to address risk?”

Initial preparation of the scope of an investigation begins with a review of the site history and available environmental data. This is used to prepare a preliminary Conceptual Site Model (CSM) of suspected conditions, identify COPCs, develop specific investigation questions and determine data needs. The site is then divided into well-thought-out risk- or remediation-based DUs and a detailed SAP prepared. When the data are made available, a thorough data quality review is carried out to confirm that samples were properly collected, processed and tested and that data for replicate samples met data quality criteria. When an acceptable level of confidence in the data is reached, the environmental risk is assessed, the CSM updated as needed and recommendations for additional actions proposed.

A key part of the Systematic Planning process is the development of DU decision statements that specify how sample data will be used for decision making. This is done *prior* to the collection of samples and agreed upon by the field workers, risk assessors and remediation experts involved in the project. An example might include “If Contaminant ‘X’ in soil exceeds Screening Level ‘Z’ then the DU volume of material will be excavated and disposed of in a regulated landfill.” Investigations are intentionally designed in a manner that minimizes the need for remobilization and collection of additional samples. This can be accomplished with proper forethought and planning.

Sampling Error Control

Controlling data error is critical to making reliable decisions. Potential error associated with the collection of samples in the field and subsamples for testing in the laboratory is in part related to the compositional and distributional heterogeneity of the media of interest^{2,3,4,5,6} (Figure 2). Compositional heterogeneity or “Fundamental Error” refers to variability between individual particles and is described in terms of particle size, shape and density (Figure 2a). Other characteristics of importance include whether the contaminant of interest is bound up within individual particles or fully liberated and present as individual nuggets.

Error associated with compositional heterogeneity is controlled through the collection of an adequate mass of material. Compositional heterogeneity is relatively easy to control in testing of air and water due to the smaller particle size but plays an important role in testing of particulate matter. In the case of soil, defined as particles equal to or less than or two millimeters in diameter, a minimum subsample mass for testing of 30 grams is required to reliably address Fundamental Error^{37,38}. Smaller subsample masses can be acceptable for finer-grained material, but testing of a minimum of 10 grams of material is required for any particle-size material that has not been thoroughly ground.

Distributional heterogeneity refers to variability of the mean concentration of a contaminant within the project site as well as within individually targeted DUs and even the sample collected from a DU and submitted for processing and testing (Figure 2b). Division of a project site into DUs in order to isolate heavy contamination and optimize remedial efforts requires a preliminary estimate of distributional heterogeneity based on historical information and available data. This must be carried out at a scale that is cost-beneficial for the project as a whole. Excessively large DUs could lead to unnecessary cost for treatment of otherwise clean material. Excessively small DUs could lead to unnecessary time and effort in the field with no benefit for successful completion of the project.

Distributional heterogeneity, described by Gy as Grouping and Segregation Error, plays an important role in all types of environmental media and is the main cause of data error in research and regulatory field investigations. Other types of sampling error pertinent to environmental investigations include designation of increment collection points (Increment Delimitation Error) and collection of sample increments in the field (Increment Extraction Error); sample processing at the laboratory (Preparation Error) and error associated with analysis of a subsample (Analytical Error). Analytical error is often considered the primary concern for data variability. Analytical error is generally minor, however, in comparison to sample collection and processing error^{1,4}.

Testing of “Finite Element” Versus “Infinite Element” Media

Proper collection of samples in the field is required to obtain representative data. Media targeted for sample collection can be defined as “finite element” or “infinite element”^{38,40}. “Elements” are defined as the individual components that comprise a material, for example particles of minerals and organic material that make up soil and individual molecules that make up liquids and vapors. Finite element media are defined as media composed of elements that can be individually identified and individually selected at random. Examples include room of people, forest of trees, cans of tuna, bags of rice, bags of soil, bottles of water, etc. Data Quality Objectives (DQOs) for sample collection typically include the need to assess variability between individual elements (e.g., people or packages) or estimate “minimum,” “maximum” or other quality control parameters. Data are compiled by are collecting and independently testing individual elements. The elements collected in total represent the “sample.” Traditional statistical tests are used to compile the desired DQOs. The only variable of concern is the number of elements selected. Variability between individual elements is used to assess the overall precision of the estimated parameter values.

Infinite element media are defined as media composed of elements that cannot realistically be individually identified or individually selected at random for testing. Examples include a pile of flour, rice or soil as well as a pool of water or a room of air. Vapor in unsaturated pore space in soil immediately beneath the slab of a building, often the focus of environmental investigations associated with “vapor intrusion,” also represents a form of “infinite element media. The primary DQO associated with testing of these types of media is always determination of the mean or “true” concentration of the specified chemical or other parameter for the DU area and volume of material as a whole. Data are compiled by collecting and combining groups or “increments” of individual elements into single, bulk sample for processing and testing^{37,38}. Variables of concern include the number of increments used to prepare the sample, the method used to collect the increment and the total mass of the resulting sample. Data for independent, replicate samples collected from a portion of the project DUs are used to test the overall precision of the sampling method. The precision of replicate sample data in combination with a review of the final sampling collection and processing methods is used to assess the overall quality and reliability of sample data for decision making.

A historical mistake in the environmental industry, as described above, has been to apply “discrete” sampling methods appropriate to finite element media to testing of infinite element media such as air, water and soil. As a result, the data provided by the laboratory are not reliably representative of the sample provided and the sample provided is not reliability representative of the area and volume of media that it was collected from. This is the crux of the problem with traditional sampling methods employed by the environmental industry.

Collection of Representative Environmental Samples

Testing of Soil and Sediment

Investigation Question and DU Designation. Testing of soil, sediment and other particulate matter for contaminants presents challenges similar to testing of crush ore for commodities. In the case of environmental contamination, the basic Investigation Question might be “Does the mean concentration of contaminant ‘X’ in the targeted area and volume of soil ‘Y’ exceed screening level or risk ‘Z’?”

Decision Units are designated for testing based on either the need to assess risk (“Exposure Area” DUs) or to better isolate areas of high contamination or estimate contaminant mass (“Source Area” DUs) and optimize remedial actions^{37,38}. The size, shape and volume of DUs for a particular investigation will be dependent on the specific questions being asked. For example, a default DU area of a few hundred square meters (m²) to a depth of 15 cm might be adequate to assess risk posed by direct contact with contaminated soil in existing or planned residential areas. Larger DUs sizes might be appropriate for investigation of commercial or industrial sites. Designation of DUs for deeper, subsurface soil is usually based on the need to identify the main mass of contamination and optimize potential remedial actions. This is similar to the identification of an economically minable subsurface body of ore in the mining industry.

Collection of Representative Samples. The State of Hawai’i undertook detailed field studies of Theory of Sampling methodologies for testing of contaminated soil and sediment in 2004, the first state in the United States to do so. Guidance on the use of “Decision Unit” and “Multi Increment® Sample” (DU-MIS) investigation methods was first published in 2009 and updated in 2016. (Multi Increment® is a registered trademark of EnviroStat, Inc.) Additional updates were underway at the time this paper was published. Coordination was subsequently carried out with the Chinese Academy of Sciences, Nanjing Institute of Soil Science, for expansion and use of the guidance in China³⁸ (see also 41).

Error associated with distributional heterogeneity within a targeted DU is addressed through the collection and combination of a large number of small masses of particles throughout the targeted DU into a single sample. Each mass is referred to as an “increment.” For soils and other particulate matter, a field sample composed of at least 50 increments with a minimum mass of one to three kilograms and collected in a systematic, random manner is normally adequate to address distributional

heterogeneity and general sample collection error^{37,38}. A greater number of sample increments and/or a larger sample mass might be required in cases where the contaminant is present as liberated “nuggets” within the particulate matter.

Error in the collection of subsamples at the laboratory is controlled in a similar manner as done in the field. The sample is allowed to air dry and then sieved to isolate the target, particle-size fraction. The sample might or might not be ground, depending on the nature of the Investigation Question being asked. Subsamples for testing are collected either through use of a sectoral splitter (preferred) or manually, following a similar, Multi Increment type collection method as used in the field. Subsamples must meet minimum mass requirements to address Fundamental Error and other error associated with subsample collection and testing, typically 30 grams.

A detailed review of sample collection and processing methods is provided in guidance published by the Hawai'i Department of Health³⁷ and the Chinese Academy of Sciences³⁸. This includes methods for the collection of both surface and subsurface samples, as well as samples from stockpiles. Alternative approaches for the collection and testing of samples to be tested for volatile contaminants are also provided, as is guidance on the collection of replicate samples and step-by-step methods for a review of overall data quality.

Testing of Subslab Vapors

Investigation Question and DU Designation. Contamination of indoor air is sometimes associated with a release of chlorinated solvents and petroleum fuels into soil and groundwater and subsequent upward diffusion and advective migration of vapors into overlying buildings, a process referred to as “vapor intrusion”^{27,31,21,25}. A similar Investigation Question as applied to indoor air is applied to testing of subslab vapors: “Does the mean concentration of contaminant ‘X’ in vapors intruding the building during time period ‘Y’ exceed screening level or risk ‘Z’?”

The volume of subslab vapors intruding a building during the targeted time period is the DU. This volume can be approximated as the predicted vapor entry rate through gaps in the building slab times the time period of concern²⁵. For example, a subslab vapor entry rate of 4.5 liters per minute is assumed for single-family home in cold climates where buildings are heated for much of the year based on the typical annual-average indoor-outdoor air pressure differential and the total area of cracks and gaps in floors. This implies a daily vapor entry rate of 6,480 liters.

Risk is assessed by comparison of the mean concentration of the contaminant in intruding vapors to a risk-based screening level or direct incorporation of the data into an indoor air impact model. Subslab vapor screening levels for vapor intrusion risk are calculated as the screening level for the contaminant in indoor air times the default indoor air:subslab vapor attenuation factor⁴². The attenuation factor reflects the volume of subslab vapors intruding through a building slab during the specified time period divided by the volume of indoor air circulating within the structure during the same time period and including fresh air brought in through the ventilation system (see above discussion). For example, an attenuation factor of 0.0032 is predicted for a single-family home located in a cold climate zone (6,480/2,050,000)²⁵. In theory, a building specific attenuation factor can be calculated by dividing the concentration of a targeted contaminant in indoor air by the mean concentration of the contaminant in vapors intruding through the building slab. While simple in concept, this requires careful collection of a representative sample and knowledge of the specific location of intruding vapors is rarely known. The latter, in particular, is rarely known.

Collection of Representative Samples. A sample representative of a DU volume of vapors at a targeted point beneath the building slab is collected using a “Large Volume Purge (LVP)” sampling methodology^{43,44}. Following this approach, a small-diameter (e.g., 5cm) subslab vapor sampling point is installed in the floor of a building based on the area anticipated to be most heavily contaminated or the area of the most sensitive receptors or, if neither is applicable, in the approximate center of the slab. The objective is to estimate the mean concentration of the targeted contaminant in vapors that could intrude through the point and impact indoor air over the targeted exposure period. Testing of vapors representative of vapor intrusion over an exposure period of several years is not feasible. Initial screening of potential risk can instead be carried out but collection of a sample representative of the volume of vapors predicted to intrude through the point over a period of several days. For example, the volume of vapors predicted to intrude a single family home in a cold climate over a period of five days would be approximately 32,000 liters²⁵. This allows some control of distributional heterogeneity within the subslab vapor plume.

Under ideal circumstances, the entire DU volume of subslab vapors would be drawn from the sampling point and sent in a single container for testing. This again is practicable and a representative sample of the DU volume of air must instead be collected using an LVP method. Under this approach, a vacuum is used to draw the targeted volume of vapors from the sampling point. A continuous sample of the vapor stream is collected by connection of a Summa canister to the purge line. A six-liter canister is normally used. The vacuum is turned off when the target purge volume has been reached and the valve to the Summa canister closed. The canister is then submitted to the laboratory for analysis.

Data for the relatively small volume of vapor collected in the sample can be reasonably be assumed to be representative of the mean concentration of the targeted contaminant in the full volume of vapors purged from the point. As would be expected, field studies have documented an increasing mean concentration of the contaminant in vapors as the purge

volume increases⁴⁵. This reflects the capture and inclusion of small pockets of elevated contamination of contaminants into the purge stream. This is similar to the observed increase in the mean concentration of a contaminant in soil with an increasing number of increments included in a sample^{15,16}.

Testing of Surface Water

Investigation Question and DU Designation. Testing of surface water is similar to testing of indoor air. In both cases the targeted media is contained within a fixed structure, in the first case a basin and in the second case a room or building. The water or air might or might not be flowing into and out of the containment. The contaminant of concern might be present as either individual, molecular-scale particles in a single-phase media, for example vaporized in air or dissolved in water, or as particulate matter in a dual-phase media. In the case of surface water, particulate matter might include suspended sediment or bacteria.

The volume of water held within or moving through a targeted area during the targeted time period is the DU. Designation of one or more DUs within a project area might be desirable based on the locations of specific aquatic habitats, suspect source areas, discharge points or other considerations. Risk is assessed by comparison of the mean concentration of the target compound for the DU volume of water as a whole to a screening level pertinent to the investigation question or by incorporating the data into a human health risk assessment (e.g., bacterial count or concentration of suspended sediments or dissolved-phase contaminants).

Collection of Representative Samples. The above dilemma can be addressed through consideration of sampling methods more appropriate for infinite element media. Under this approach, a single sample of water is prepared by collection and combination of multiple "increments" of water from throughout the targeted DU. Water collected from each point is referred to as an "increment" and the combined increments referred to as a "Multi Increment" sample.

Isokinetic water sampling methods are used to collect depth-integrated samples from water moving more than 0.5 to 1.0 meters per second^{28,46}. Under this approach the sampling device is lowered at a constant rate through the targeted depth interval. The device is designed to ensure that water enters the sampling container at a rate that directly corresponds to the velocity that the water is moving. This allows the sample collected to be representative of the volume of water moving through the point as a whole. Increments of water can be collected from multiple points perpendicular to the flow direction to prepare a Multi Increment sample. Concurrent, replicate samples can be collected from different points along the same transit line to evaluate the precision of the sample data. Samples can be collected at different time periods to assess variability in contaminant concentrations over time and in stream volumes of water.

Non-isokinetic samples can be used to collect Multi Increment samples in still water⁴⁶. An example is the "Aloha Sampler™". (Aloha Sampler™ is a trademark of EnviroStat, Inc.) The sampler consists of a standard one-liter, high density polyethylene sample bottle with two to three holes drilled in the cap. The bottle is immersed horizontally at a consistent rate in a manner that allows water to enter the lower hole and air to escape from the top hole. The sampler lowers and raises the bottle to the target collection depth at multiple points throughout the entire DU. This allows collection of a sample representative of the DU volume of water as a whole.

These approaches allow collection of a representative sample of water in a time- and cost-efficient manner. Data collected are directly relatable to the Investigation Question asked. Independent, replicate samples are collected to test the precision of the sample data.

Testing of Groundwater

Investigation Question and DU Designation. The collection of representative samples of groundwater presents challenges similar to the collection of subsurface soil vapor. Investigation questions applied to groundwater are similar to those applicable to subsurface vapors, for example: "Does the mean concentration of contaminant 'X' in groundwater passing through targeted area during time period 'Y' exceed screening level or risk 'Z'?" As applied to an individual well, the question might be: "Does the mean concentration of contaminant 'X' in groundwater extracted from the well during time period 'Y' exceed screening level or risk 'Z'?"

The concept of "Decision Units" is again not routinely discussed in published guidance. Decision Units for groundwater should be designated based on risk or optimization of remedial actions. For example, the standard exposure duration for assessment of noncancer health hazard and protection of young children is six years⁴⁷. The standard exposure duration for assessment of cancer health risk in adults is 30 years. Under ideal circumstances, the capture zone that predicts the volume of groundwater to be produced by the well over the time period applicable to the COPC would be identified as the DU.

A purely risk-based approach to DU volume designation could be applied for periodic, snapshot monitoring of water quality. For example, an average drinking water ingestion rate of 0.78 liters per day is assumed for children ages 0 to 6 years and an average rate of 2.5 liters per day is assumed for adults. This equates to an annual volume of water consumed by a child over six years of 1,638 liters and over a 30 years by an adult of 23,538 liters. These could be considered to represent

Exposure Volume DUs.

Alternative DU volumes could also be designated for monitoring of remedial efforts (Remedial Monitoring DUs). The Investigation Question for a specific well can now be reworded to state: “Does the mean concentration of contaminant ‘X’ in the targeted DU volume of groundwater extracted from the well exceed screening level or risk ‘Z’?” The next step is to devise a method to collect a representative sample of this water.

Collection of Representative Samples. In the most ideal case, the entire six- to thirty-year production volume of water would be extracted from a production well and submitted to the laboratory for analysis as a single sample. This is of course not practicable. The water could alternatively be extracted, stored in a large tank and a single, Multi Increment sample collected and submitted for analysis. This again is unlikely to be practicable under most circumstances. A second alternative is to install multiple wells within the aquifer and prepare a single, Multi Increment sample by combining increments of water from each well into a single sample. Determination of the number of wells necessary to represent the DU volume of water would require a detailed field study. Replicate sets of wells would need to be installed and used to assess the precision of the data. This approach once more is unlikely to be practicable due to the cost required.

The focus of sample collection must therefore return to the collection of sample data from individual wells. Although not currently done, Large Volume Purge type sampling methods described above for the collection of subsurface vapor samples could be utilized to collect a sample representative of the targeted DU volume of water from a well⁴⁶. For example, risk-based volumes of groundwater corresponding to the amount of water consumed over six years by children (e.g., 1,638 liters) could be periodically extracted from the well for testing. For nonvolatile contaminants, this might include storage of purged water in a large, holding tank and collection of a sample for testing. Replicate samples could be collected to test the precision of the sample collection method and the approach revised as needed. Excess water could be reinjected in the same area of the aquifer, allowed to infiltrate or otherwise treated or disposed of.

The collection of representative samples to be tested for volatile or otherwise unstable contaminants is more challenging. A continuous sample would presumably need to be continuously drawn from the purge train as the water is extracted from the well in the same manner as the collection of a sample during testing of subsurface vapors. Such methods are used for testing of flowing, surface water²⁸ but to the authors’ knowledge have not been adapted for testing of water flowing through a pipeline or sampling train. The precision of the data can be tested through the collection of replicate samples from the same sampling train.

Summary and Discussion

Whether crushed ore, soil, air, water or other type of infinite element media, similar sampling concepts apply. The delay in the incorporation of Theory of Sampling concepts in the environmental industry for testing of subsurface vapors and water as well as soil and other particulate matter is related to several factors. This includes a lack of training of environment regulators; a reluctance to move away from traditionally accepted sampling methods; a lack of routine, third-party verification of sample data reliability; a lack of clear repercussions for erroneous data and the lack of guidance that presents a viable alternative. These problems are slowly beginning to be addressed as field-tested guidance becomes available and demand for more efficient and reliable investigation methods grows.

Environmental risk assessors sometimes mistakenly cling to discrete sample data for small volumes of air or other media in an attempt to assess “acute” health affects posed by the “maximum” concentration of a contaminant in the targeted media. Such an assessment would still require that a DU volume of the media specific to acute risk be designated, for example a few liters representing a single breath of air. This represents a finite element media sampling challenge, with each DU volume representing an individual “element” for which data are required. A large and likely unmanageable number of samples would then be required to predict the maximum concentration of the contaminant in any given DU volume of air within a specified level of confidence. For example, the collection of a minimum of 59 samples would be required to predict the maximum concentration of a contaminant of the contaminant in any given DU volume of air with a 95% confidence level (e.g., one-sided nonparametric tests)¹⁴. This is a different sampling problem and only occasionally a part of environmental investigations⁴⁸.

Caution is still called for, however. A recent guidance document⁴⁹ for what is referred to as “Incremental Sampling Methodology (ISM)” provides a reasonable overview of discrete sample data error and the incorporation of Theory of Sampling concepts in the collection, processing and testing of soil and sediment samples. As critiqued by the Hawaii Department of Health⁵⁰, the document then erroneously regresses to require the use of statistical tests applicable to finite element media to assess risk and calculation a 95% Upper Confidence Level (UCL) for final decision making. This necessarily requires the expensive and unnecessary collection of replicate samples in all DUs. As an alternative, the guidance implies that a 95% UCL based on as few as ten discrete samples can instead be used for decision making, ignoring the unreliability of discrete sample data discusses in earlier sections of the document. Such compromises and mistakes do not properly address Theory of Sampling and must be rigorously opposed.

References

1. F.F. Pitard, *Pierre Gy's Sampling Theory and Sampling Practice*. New York: CRC Press (1993).
2. F.F. Pitard, "Sampling correctness – A comprehensive guideline", *Proceedings of Sampling and Blending Conference (WCSB2)*. Sunshine Coast, Queensland, Australia (2005).
3. F.F. Pitard, "Theoretical, practical and economic difficulties in sampling for trace constituents", *Proceedings of the Fourth World Conference on Sampling and Blending (WCSB4)*. The Southern African Institute of Mining and Metallurgy, Cape Town (2009).
4. F.F. Pitard, *Theory of Sampling and Sampling Practice*. Chapman and Hall/CRC, Boca Raton, Florida, ISBN 9781138476486 (2019).
5. C.A. Ramsey and A.D. Hewitt, "A methodology for assessing sample representativeness", *Environ Forensics* **6**, 71–75 (2005).
6. R.C.A. Minnitt, P.M. Rice and C. Spangenberg, "Part 1: Understanding the components of the fundamental sampling error: A key to good sampling practice", *J. South. Afr. Inst. Min. Metall.* **107**, 505–511 (2007).
7. USEPA, *Verification of PCB Spill Cleanup by Sampling and Analysis*. United States Environmental Protection Agency, Washington, DC, EPA-560/5-85-02, (1985).
8. USEPA, *Field Manual for Grid Sampling of PCB Spill Sites to Verify Cleanups*. United States Environmental Protection Agency, Office of Toxic Substances, Washington, DC, EPA-560/5-86-017 (1986).
9. USEPA, *Low Flow (Minimal Drawdown) Ground-Water Sampling Procedures*. United States Environmental Protection Agency, Office of Research and Development, EPA/540/S-95/504 (1996).
10. USEPA, *Groundwater Sampling*. United States Environmental Protection Agency, Science and Ecosystem Support Division, SESDPROC-301-R4 (2017).
11. USEPA, *RCRA Orientation Manual*. Office of Resource Conservation and Recovery, United States Environmental Protection Agency (2014).
12. USEPA, *Risk Assessment Guidance for Superfund. Volume I, Human Health Evaluation Manual (Part A)*. United States Environmental Protection Agency, Office of Emergency and Remedial Response, EPA/540/1-89/092 (1989a).
13. USEPA, *Risk Assessment Guidance for Superfund. Volume II, Environmental Evaluation Manual*. United States Environmental Protection Agency, Office of Emergency and Remedial Response, EPA/540/1-89/001 (1989b).
14. USEPA, *Methods for Evaluating the Attainment of Cleanup Standards, Volume 1: Soils and Solid Media*. United States Environmental Protection Agency, Office of Policy, Planning, and Evaluation, Washington, DC, EPA/230/02-89/042 (1989c).
15. R. Brewer, J. Peard and M. Heskett, "A critical review of discrete soil sample reliability: Part 1 – Field study results", *Soil and Sediment Contamination* **26**, 1 (2017a). <http://dx.doi.org/10.1080/15320383.2017.1244171>
16. R. Brewer, J. Peard and M. Heskett, "A critical review of discrete soil sample reliability: Part 2 – Implications", *Soil and Sediment Contamination* **26**, 1 (2017b). <http://dx.doi.org/10.1080/15320383.2017.1244172>
17. USEPA, *Preparation of Soil Sampling Protocols: Sampling Techniques and Strategies*. United States Environmental Protection Agency, Office of Research and Development, Washington, DC, EPA/600/R-92/128 (1992).
18. USEPA, *A Rationale for the Assessment of Errors in the Sampling of Soil*. United States Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Washington, DC, EPA/600/4-90/013 (1990).
19. USEPA, *Guidance on Surface Soil Cleanup at Hazardous Waste Sites*. United States Environmental Protection Agency, Office of Emergency and Remedial Response, Washington, DC (EPA 9355.0-91 (2005).
20. USEPA, *ProUCL Version 5.0.00, User Guide*. United States Environmental Protection Agency, Office of Research and Development, Washington, DC, EPA/600/R-07/041 (2013).
21. ITRC, *Vapor Intrusion Pathway: A Practical Guideline*. Interstate Technology Regulatory Council (2007).
22. CAEPA, *Guidance for the Evaluation and Mitigation of Subsurface Vapor Intrusion to Indoor Air*. California Environmental Protection Agency, Department of Toxic Substances Control (2011).
23. CAEPA, *Advisory – Active Soil Gas Investigation*. California Environmental Protection Agency, Department of Toxic Substances Control (2015).
24. T. E. McHugh, T.N. Nickels and S. Brock, "Evaluation of spatial and temporal variability in VOC concentrations at vapor intrusion investigation sites", *Proceeding of Air & Waste Management Association's Vapor Intrusion: Learning from the Challenges*, pp 129-142 (2007).
25. R. Brewer, J. Nagashima, M. Rigby, M. Schmidt and H. O'Neill, "Estimation of generic subsurface attenuation factors for vapor intrusion investigations", *Groundwater Remediation and Monitoring* **34**, 79–92 (2014). doi: 10.1111/gwmr.12086
26. H. Luo, P. Dahlen, P.C. Johnson, T. Peargin and T. Creamer, "Spatial variability of soil-gas concentrations near and beneath a building overlying shallow petroleum hydrocarbon impacted soils", *Groundwater Monitoring & Remediation* **29** (1), 81–91 (2009).

27. USEPA, *EPA's Vapor Intrusion Database: Evaluation and Characterization of Attenuation Factors for Chlorinated Volatile Organic Compounds and Residential Buildings*. United States Environmental Protection Agency, Office of Solid Waste and Emergency Response, EPA 530-R-10-002 (2012a).
28. USGS, *National Field Manual for the Collection of Water-Quality Data, Chapter A4 – Collection of Water Samples*. United States Geological Survey (2006).
29. USEPA, *Correct Sampling Using the Theories of Pierre Gy*. United States Environmental Protection Agency, National Exposure Research Laboratory, Environmental Sciences Division, Technology Support Center, Washington, DC, Fact Sheet 197CMB98.FS-14 (1999).
30. USEPA, *Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples*. United States Environmental Protection Agency, Office of Research and Development, Washington, DC, EPA/600/R-03/027 (2003).
31. USEPA, *User's Guide for Evaluating Subsurface Vapor Intrusion into Buildings*. United States Environmental Protection Agency, Office of Emergency and Remedial Response (2004).
32. ASHRAE, *Ventilation for Acceptable Indoor Air Quality*. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Standard 62.1. Atlanta, GA (2013).
33. B. Hartman, "How to collect reliable soil-gas data for risk-based applications, Part 1: Active soil-gas method", *New England Interstate Water Pollution Control Commission, LUSTLine Bulletin* **42** (2002).
34. Air Toxics, *Guide to Air Sampling and Analysis, Canisters and Tedlar Bags*. Air Toxics, Ltd. (2012).
35. Air Toxics, *Sorbent and Solution Sampling Guide*. Air Toxics, Ltd. (2012).
36. USEPA, *Vertical Distribution of VOCs in Soils from Groundwater to the Surface/Subslab*. United States Environmental Protection Agency, Office of Research and Development, EPA/600/R-09/073 (2009).
37. HDOH, *Technical Guidance Manual (Sections 3-5, Soil and Sediment)*. Hawai'i Department of Health, Office of Hazard Evaluation and Emergency Response (2021).
38. CAS, *Standard for the Use of Decision Unit and Multi Increment Sample Investigation Methods to Assess Soil Quality*. Chinese Academy of Sciences, Nanjing Institute of Soil Science (in final review, 2022).
39. AAFCO, *GOODSamples: Guidance on Obtaining Defensible Samples*. Association of American Feed Control Officials, Champaign, Illinois. (2015).
40. AAFCO, *Good Test Portions: Guidance on Obtaining Defensible Test Portions*. Association of American Feed Control Officials, Champaign, Illinois (2018).
41. R. Brewer, J. Song, J. and C. Ramsey, "Advanced sampling methods to expedite the reliable characterization, remediation and redevelopment of contaminated industrial lands in China", *Proceedings of the 9th World Conference on Sampling and Blending (WCSB9)*. Beijing, China, pp. 292-307 (2019).
42. HDOH, *Evaluation of Environmental Hazards at Sites with Contaminated Soil and Groundwater (Hawaii Edition)*. Hawai'i Department of Health, Office of Hazard Evaluation and Emergency Response (2017a).
43. T.A. McAlary, P.J. Nicholson, L.K. Yik, D.M. Bertrand, and G. Thrupp. "High purge volume sampling - a new paradigm for subslab soil gas monitoring", *Ground Water Monitoring and Remediation* **30** (2), 73–85. (2010).
44. HDOH, *Technical Guidance Manual (Section 7: Indoor and Soil Vapor)*. Hawai'i Department of Health, Office of Hazard Evaluation and Emergency Response (2017b).
45. HDOH, *Field Study of High-Density Passive Sampler and Large-Volume Purge Methods to Characterize Subslab Vapor Plumes*. Hawai'i Department of Health, Office of Hazard Evaluation and Emergency Response (2017c).
46. HDOH, *Technical Guidance Manual (Section 6: Surface Water and Groundwater)*. Hawai'i Department of Health, Office of Hazard Evaluation and Emergency Response (2021).
47. USEPA, *Screening Levels for Chemical Contaminants*. United States Environmental Protection Agency, Superfund Program (2021).
48. C. Ramsey and K. H. Esbensen, "Inferential statistical sampling of hyper-heterogeneous lots with hidden structure: the importance of proper Decision Unit definition", *Spectroscopy Europe* **34** (2) (2022).
49. ITRC, *Incremental Sampling Methodology*. Interstate Technology and Regulatory Council (2020).
50. HDOH, *Critique and Use of ITRC Incremental Sampling Methodology Guidance Document in Hawai'i*. Hawai'i Department of Health, Office of Hazard Evaluation and Emergency Response (November 23, 2020).

Mining Industry	Environmental Industry
1. Geologic Map	1. Site Conceptual Site Model
2. Preliminary Ore Body Identification	2. Preliminary Source Area Identification
3. Mining and Crushing	3. (not applicable; already particulates)
4. Sample Collection (TOS)	4. Sample Collection (discrete)
5. Sample Processing (TOS)	5. Sample Processing (minimal/none)
6. Subsample Collection (TOS)	6. Subsampling Collection (discrete)
7. Data Quality Review <ul style="list-style-type: none"> • Sample collection method; • Field replicate precision; • Laboratory replicate precision; • Analytical precision; • Unadjusted data used for decision making. 	7. Data Quality Review <ul style="list-style-type: none"> • Minimal to no testing of field sample or lab subsample reproducibility; • Validation of laboratory analytical method only; • 95% UCL of mean of discrete samples used for decision making.
8. Economic Feasibility Assessment	8. Environmental Risk Assessment
9. Commodity Extraction and Marketing	9. Contaminant Destruction or Isolation

Figure 1. Comparison of mining industry and environmental industry sampling methodologies.

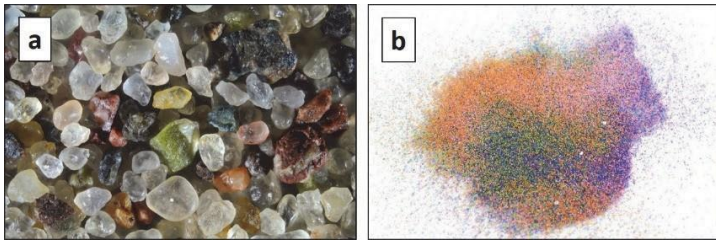


Figure 2. Example of compositional heterogeneity between individual particles (a) and distributional heterogeneity of particle types within a targeted volume of material (b).