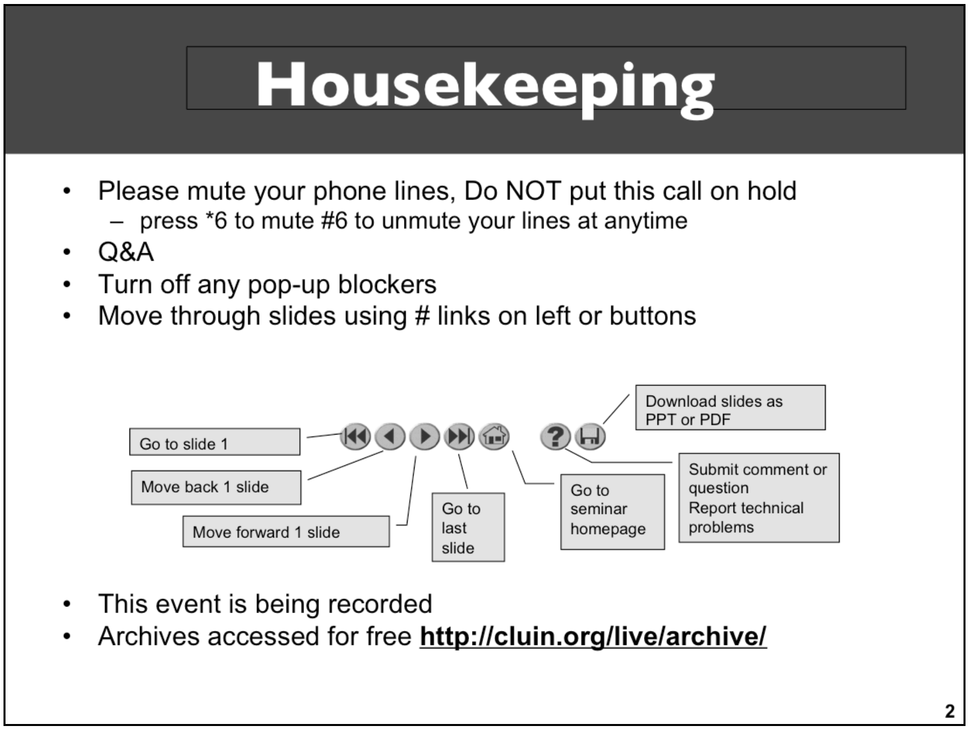
Welcome to the CLU-IN Internet Seminar

Incremental-Composite Sampling Designs for Surface Soil Analyses, Module 2 of 4

Delivered: February 21, 2012, 2:00 PM - 4:00 PM, EST (19:00-21:00 GMT)

Presenters:
Deana Crumbling, EPA Superfund (crumbling.deana@epa.gov)
Robert Johnson, Argonne National Laboratory (rljohnson@anl.gov)
Moderator:
Jean Balent, U.S. EPA, Technology Innovation and Field Services Division (balent.jean@epa.gov)

Visit the Clean Up Information Network online at www.cluin.org



Although I’m sure that some of you have these rules memorized from previous CLU-IN events, let’s run through them quickly for our new participants.

Please mute your phone lines during the seminar to minimize disruption and background noise. If you do not have a mute button, press *6 to mute #6 to unmute your lines at anytime. Also, please do NOT put this call on hold as this may bring delightful, but unwanted background music over the lines and interupt the seminar.

You should note that throughout the seminar, we will ask for your feedback. You do not need to wait for Q&A breaks to ask questions or provide comments. To submit comments/questions and report technical problems, please use the ? Icon at the top of your screen. You can move forward/backward in the slides by using the single arrow buttons (left moves back 1 slide, right moves advances 1 slide). The double arrowed buttons will take you to 1st and last slides respectively. You may also advance to any slide using the numbered links that appear on the left side of your screen. The button with a house icon will take you back to main seminar page which displays our agenda, speaker information, links to the slides and additional resources. Lastly, the button with a computer disc can be used to download and save today’s presentation materials.

With that, please move to slide 3.

Module 2.1

Incremental-Composite Sampling Designs for Surface Soil Analyses

Introduction to 2nd Day

3

Webinar Purpose and Goal

- Multi-increment / Incremental / Composite soil sampling designs are generating lots of interest
- EPA guidance is clear about benefits and roles these designs can play; but more implementation detail needed
- Other information sources now filling that gap
- When done correctly these soil sampling strategies provide significant benefits to soil characterization programs
- Goal is foster understanding incremental and composite sampling strategies, and how to apply them

Overall Webinar Agenda

- 4-part series of 2-hr Clu-In webinars (Feb 20, 21, 24 & 27)
- Soil sampling fundamentals (Day 1)
- Terminology and existing sources of information (Day 1)
- Incremental and composite sampling designs (Day 2)
  - Incremental: Estimating average concentrations for a decision unit
  - Composites: Looking for contamination “hot spots”
  - ICS quality control
- Calculations and using Visual Sample Plan (VSP) (Day 3)
- Several case studies (Days 3 & 4)
- Interspersed periods for Q & A

Today’s Agenda

- Incremental averaging to improve estimates of the concentration mean
- Composite searching to improve “hot spot” detection
  - 15-min Q & A
- Incremental-composite sampling: limitations, caveats and QC
  - 10-min Q & A

Today's Instructors

- **Deana Crumbling,** crumbling.deana@epa.gov
  Office of Superfund Remediation &Technology Innovation
  U.S. Environmental Protection Agency
  Washington, D.C.
  (703) 603-0643

- **Robert Johnson,** rlj@anl.gov
  Environmental Science Division
  Argonne National Laboratory
  Argonne, Illinois
  (630) 252-7004

Does anyone feel that having a little fun during the day is too juvenile, like working for reward by answering questions in quizzes?

We try to liven up a dry, dry…dry, dry topic and a long course to keep you interested and awake.

Speak now or forever hold your peace.

Software Resources and Disclaimer

- Disclaimer: "References to product or service providers are for information purposes only and do not constitute endorsement"
- ITRC Incremental Sampling Methodology (ISM) Web document: http://www.itrcweb.org/ism-1/
- 2 software programs referenced
- For more information on the software programs:
  - Visual Sample Plan (VSP) (http://vsp.pnl.gov/)
  - ProUCL (http://www.epa.gov/esd/tsc/software.htm)

Incremental vs. Composite Sampling
Basic Differences

- Goal of incremental sampling is to find an average concentration over some defined area (a DU).
  - Uses 3 or more independent replicates to obtain a UCL
- Usually the goal of composite sampling is to gain information about contaminant spatial distribution.
  - Accurate or conservative estimate of the concentration for the given area (an SU) not needed
  - Typically not use replicates
- Composite usually involves fewer increments than incremental sampling

Key Take Away Points…

- Incremental and composite sampling designs provide significant advantages for soil characterization
- The difference between ISM and ICS
- Strategies used need to match project goals (i.e., DQOs)
- Systematic planning and conceptual site models (CSMs) are critical to “getting it right”
- CAN find “hot spots” via a composite sampling strategy, BUT “hot spot” must be defined!
- Sample processing and correct subsampling prior to analysis is vital to data quality

Module 2.2

Incremental-Averaging

Improving Estimates of the Concentration Mean

Triad

Systematic Planning

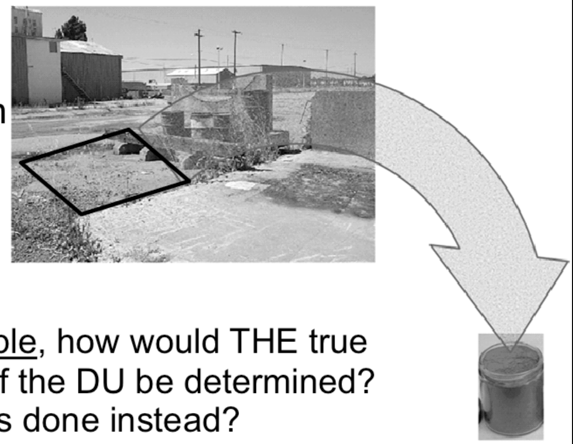
Dynamic Work Strategies

Real-Time Measurements

2 Key Assumptions of Incremental-Averaging

(1) There is a specified volume of soil called the DU

(2) There is a decision that rests on knowing THE concentration of that volume of soil



<u>If it were possible</u>, how would THE true concentration of the DU be determined? What is done instead?

2/21/2012

12

Average Concentration Estimation

- Determining average concentrations over the DU area is often a significant goal:
  1) Developing exposure point concentrations (EPC) for risk assessment purposes
  2) Demonstrating compliance with
     - Area-averaged cleanup goals
     - “Hot spot” criteria
- Fundamental concept:
  – ***How is the decision unit defined???***

Discrete Sampling Approach

- Collect one or more samples from an area, analyze, and use the mathematical average (plus an uncertainty buffer = UCL) to estimate true average
- Heterogeneity introduces uncertainty into how good that estimate is
- More samples create a better estimate
- Statistics (e.g., VSP) can be used to estimate required sample numbers **before sampling**
- Statistics (e.g., ProUCL) can be used to determine how good an estimate is **after sampling**

Issues with Discrete Sampling Approach

- How many samples are enough?
  - Won’t know until after samples are collected
- Statistics (e.g., VSP): more samples than budget
- The statistical approach might assume underlying sample distribution is normal (bell-curved)
  - Assumption affects sample numbers
- Almost all statistical approaches assume NO SPATIAL PATTERNING! (aka, “spatial autocorrelation”)
- Tendency to make decisions on single discrete sample results—DANGEROUS
  - 2 overarching reasons why

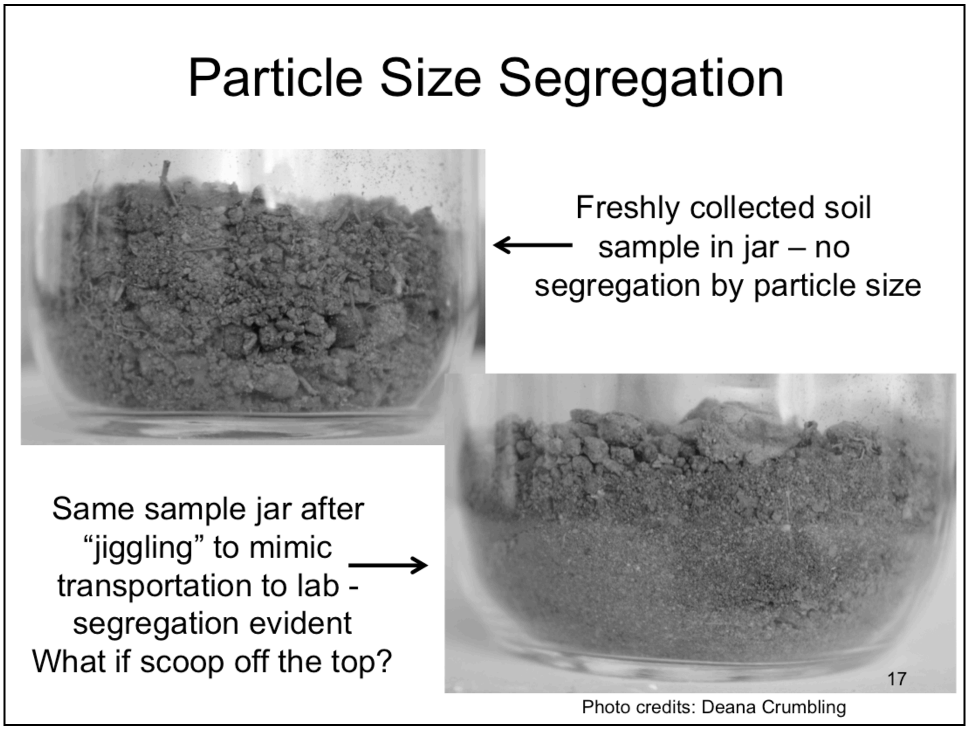
2/21/2012 Clu-In Incremental-Composite Webinar Module 2.2 15

• Generally the budget determines how many samples are collected—not a very scientific way to perform a scientific study.

• Often decisions are made on single data points. For example, if just as single analysis gives a result above an action level, the location where that sample came from may be designated a “hot spot.”

1) Typical Lab Processing

- Labs assume the sample they get is ready for analysis
- May give a “stir” in jar to “mix”—makes particle segregation worse
- Then scoop sample off top
- Lab duplicates often don’t match
  - Indication of variability in subsampling process



Here's what we mean by "particle segregation."

• These photos contrast non-segregated soil with segregated soil

• With shaking or jiggling, larger particles migrate to the top while smaller particles settle downward

• Stirring to "mix" is ineffectual to redistribute particles; often makes segregation worse

• If subsampling involves scooping off the top, could predominately get larger particles; but this depends on another factor (see next slide)

1 Gram of Soil Can Decide Action on Tons

vs.

Photo credits:
Roger Brewer, HDOH

18

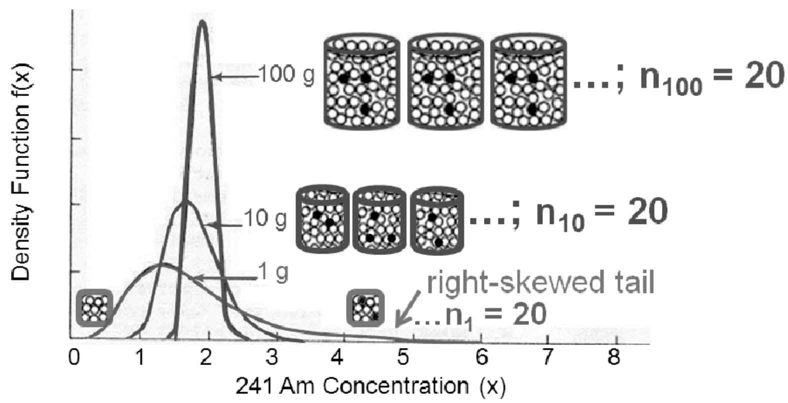
• Based on the results of analyses performed on a few grams of soil, decisions are made about whether contamination is present (and at what level) in tens to hundreds to thousands of tons of soil.

• Although a jar of soil containing 100 or more grams of soil is submitted to the lab, routine metals analysis actually analyzes only 0.5, 1 or sometimes 2 gram of soil (depending on the lab) from that jar.

• Organics analysis typically will analyze from 5 to 30 grams (depending on the lab and the analyte).

Small Subsample Supports Skew Data

Contribute to skewed statistical distributions



ITRC, ISM-1, Section 2.4.1.3 Adapted from DOE study (Gilbert, 1978) 19

Speaker Notes

- This graph plots the data from a study done in the 1970s. It directly measured how different masses of analytical samples (i.e., the sample support) influenced the statistical distribution of the data.
- Measurement units are in nCi/g
- The experiment involved first preparing a large soil sample of about 2 kg from which subsamples of various sizes could be taken. Preparing the large sample involved moderate homogenization efforts involving mild grinding and then sieving to less than 10-mesh.
- A series of 20 subsamples each of different supports were taken from the large prepared sample.
- The subsample supports that were tested included 1-g, 10-g, and 100-g .
- The wider the peak shape, the more variability present in the data set.
- The data set from the 1-g subsamples plots as a statistical distribution that is unsymmetrical and skewed in that the right-hand tail is pulled out.
  - The 1-g tail does not reach the x-axis until nearly 6 (green subsample with more nuggets than the proportion in the large sample).
  - Many samples have low concentrations, reaching down to about 0.25 (blue subsample without any high-load nuggets)
  - The width and shape (a low hump) of the curve mean that repeated subsamplings of the large sample will produce data results that are frequently quite low concentration. But sometimes there will be very high concentration results. This variability is also called imprecision. No single result can be trusted to be close to the true mean.
- In contrast to the 1-g subsamples, the 20 10-g subsamples (purple) showed much less skewing of the right tail.
  - The right-hand tail reaches the x-axis just past 3.
  - The left-hand tail shows fewer samples (than the 1-g data set) with very low results, with the lower range of the distribution ending at about 0.8
  - The width of the 10-g peak is narrower, reflecting less variability in the 10-g data set
- For the 100-g subsamples (red), the statistical distribution is almost symmetrical, with a high tight peak and the right skewing nearly gone.
  - The 100-g curve reaches the x-axis on the right at about 2.5
  - On the left, the 100-g curve runs only down to about 1.4
  - The height and narrowness of the 100-g peak indicates that replicate subsamplings of the same jar produce values that are close to each other (precise), and most likely close to the true mean for the large sample.
- Not only do small sample supports increase variability, they also contribute to data taking a lognormal or gamma statistical distribution.
- So what does this have to do with decision errors?

P.G. Doctor and R.O. Gilbert. 1978. DOE NAEG Report. *Two Studies in Variability for Soil Concentrations: with Aliquot Size and with Distance* [provided in webinar References]

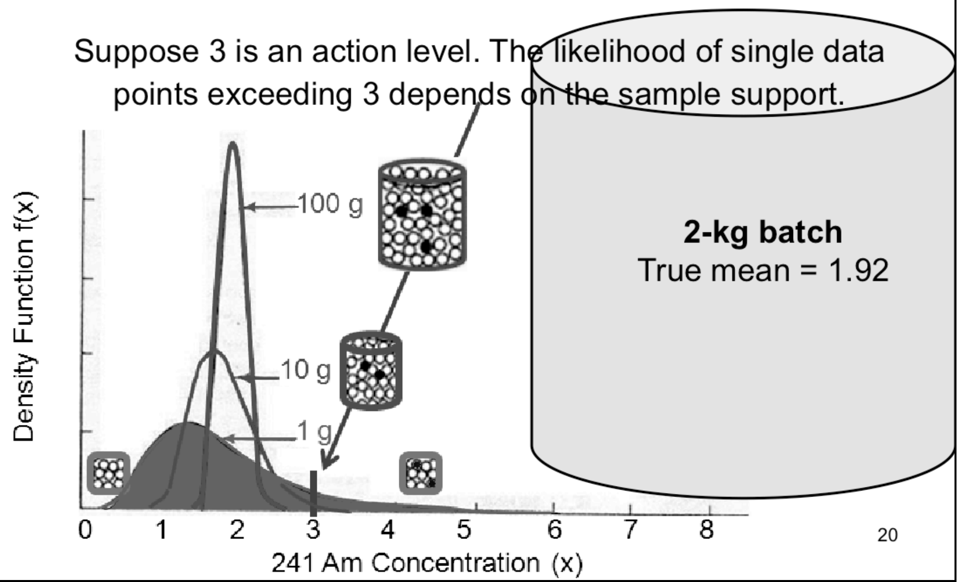
See also Gilbert, Richard O. and Pamela G. Doctor. 1985. *Determining the Number and Size of Soil Aliquots for Assessing Particulate Contaminant Concentrations.* Journal of Environmental Quality Vol 14, No 2, pp. 286-292.

Supplemental Information

See ISM-1 Section 2.4.1.3

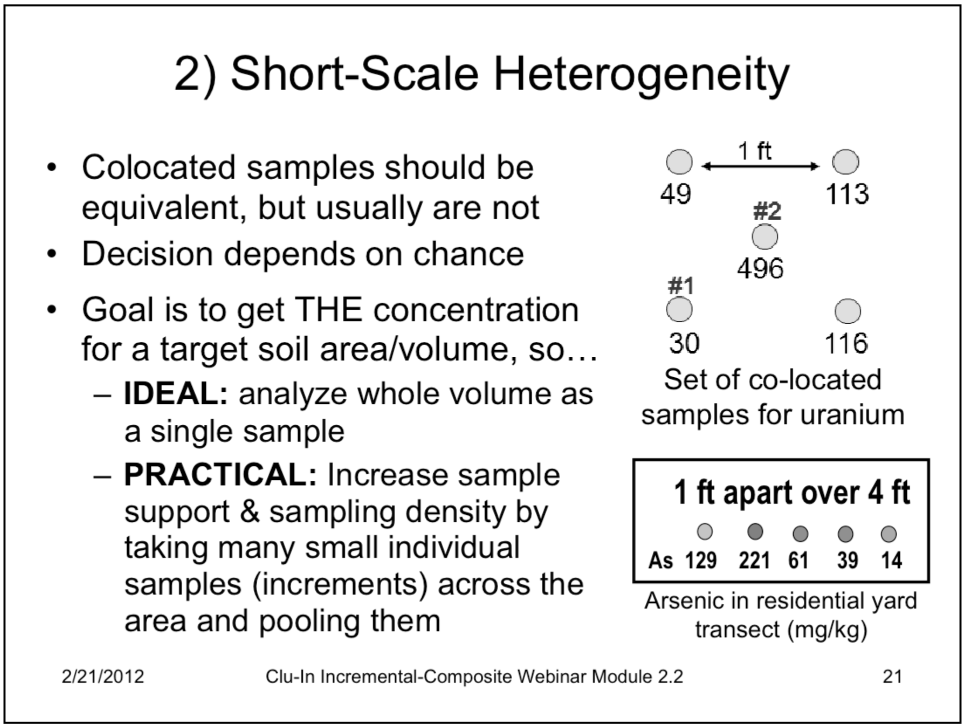
Skewed Data Distributions Can Promote Decision Errors

Suppose 3 is an action level. The likelihood of single data points exceeding 3 depends on the sample support.



- It is known that the true concentration of the large, 2-kg sample is 1.92
- Measurement units are in nCi/g
- Now suppose 3 is an action level, which is shown as the small vertical blue line on the x-axis.
- Therefore, the true concentration of the large sample is below the action level of 3
- Will the subsample that is analyzed lead to the correct conclusion, or lead the data user astray?
- Look again at the curve representing the 1-g subsamples: Even though the true mean is well below 3, the skewed nature of the data means that some of the data results are going to be higher than 3, as exemplified by the green subsample. Yet many of the 1-g subsamples will have concentrations much lower than the true mean, as exemplified by the blue subsample.
- Look at the curve representing the 10-g subsamples (the purple subsample): Only rarely will a result from a 10-g subsample exceed 3.
- In contrast, look at the 100-g curve (red subsample). Since that curve ends around 2.5, it is very, very unlikely that any single data result would be greater than 3.
- Larger subsamples are more likely to provide data results that are close to the true mean, as evidenced by the tighter peaks of the 10- and 100-g subsamples.
- The bottom line is that decisions that are based on a single sample result are more likely to be in error when subsample supports are small.
- As we talked about before, metals analysis typically uses around 1 gram of soil. Deciding that a few high results represent hotspots could well be decision errors due to the skewed distribution of data from small subsamples. This is why areas initially called hotspots sometimes cannot be found upon repeat sampling.
- Sampling errors operate in the other direction too. A sample from a true hotspot might give a data result biased far lower than the true value (blue subsample) and the hotspot would be missed.

Gilbert, Richard O. and Pamela G. Doctor. 1985. Determining the Number and Size of Soil Aliquots for Assessing Particulate Contaminant Concentrations. Journal of Environmental Quality Vol 14, No 2, pp. 286-292.



The same principles apply to short-scale sampling error. Recall that this refers to extrapolating a single data point to a large field area without taking heterogeneity into account. Taking the whole targeted soil volume as a single sample for analysis would provide THE concentration for that volume without any sampling error. Of course, that's not possible. That's why we take samples. The trick is to have enough samples to capture field heterogeneity without breaking the bank. This can be done by taking increments of soil from many locations and pooling them together for a single analysis. This both increases sampling density of the area AND increases the sample support of the field sample—both of which help control sampling error. When increments are pooled for this purpose, it's called incremental sampling.

Incremental/Composite Averaging

- Used to cost-effectively suppress short-scale (collocated) heterogeneity
  - Estimates of the mean less uncertain & closer to true
- Multiple increments contribute to the composite that is analyzed
- Increments systematically distributed over an area equivalent to, or less than, decision requirements
- Effective when the cost of analysis is significantly greater than cost of sample acquisition
- Benefit dependent on sample processing!!

2/21/2012 Clu-In Incremental-Composite Webinar Module 2.2 22

The UCL is a conservative estimate of the mean. The point is not to develop a conservative estimate of the UCL, the point is to develop a conservative estimate of the mean, which is the basis of many decisions (for example, RCRA regs say data should be produced from "representative samples" & a "representative sample" is "…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 (40 CFR 260.10)." .

The UCL is used when the mean is the actual parameter desired, but it is recognized that there is uncertainty in how well the data from the statistical sample accurately represents the statistical population. Data uncertainty in an undesirable thing. The more we can reduce data uncertainty, the more confidence we have in our estimates of population parameters (upon which the decisions are made). Reduced uncertainty in the data set is reflected as a narrowing of the interval between the calculated mean and the UCL. Reducing the UCL means the data is better.

For example, having more samples in the data set lowers the UCL. If we wanted the most conservative UCL we can get, we would limit our data sets to 2 or 3 samples. But we don't because we all know that more data is indeed "better."

How Does Incremental/Composite Averaging Work?

- Physical equivalent of averaging individual sample results
- Tends to “normalize” underlying distribution, allowing simpler statistics to be used
  - Student’s t tests, Student’s t UCL
- A set of composite sample results show less variability than discrete sample counterparts
- Theoretically, the more increments per sample per DU, the lower the variability in sample results
  - Sample processing is a critical factor!!

23

What About Dilution Concerns?

- For area-averaging goals, the concern doesn't apply
  - Goal is to get estimate of average concentration over the DU (i.e., an exposure unit)
  - Pooling increments with proper processing is physical equivalent of mathematical averaging
  - High increment density incorporates high & low concentration areas in actual field proportions
- For hot spot identification, compositing works *against* missing hot spots:
  - Hot spots also an average concept, but over smaller area
  - Compositing actually *increases* likelihood that hot spots will be incorporated into the ICS sample, raising its concentration
  - Higher sample concentration flags area for more investigation

The caveat to all sample analysis is that sample processing must be thorough so that analytical subsamples are representative of the incremental/composite sample.

How Many Increments Total Per DU?

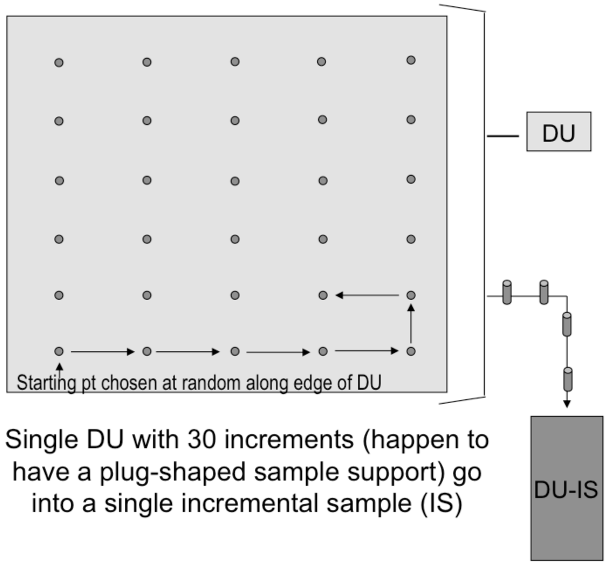
- Can vary depending on size of the area & sampling design
- For incremental samples (ISs) that cover a DU with 1 sample
  - ITRC recommends ~30 as a default for general contaminants and DUs about residential size; statistical simulations also support 30
  - Large areas may need more increments to achieve sufficient density
  - More needed for areas with high spatial heterogeneity: Military energetics & metals on firing ranges need 50-100
  - Generally fewer needed to confirm low or high concentrations far from action level
  - More needed where suspect close to action level
- If 3 replicate ISs per DU used (gives a UCL), density is 90 increments per about ¼-½ acre—generally sufficient

25

Incremental Sampling Methodology (ISM)

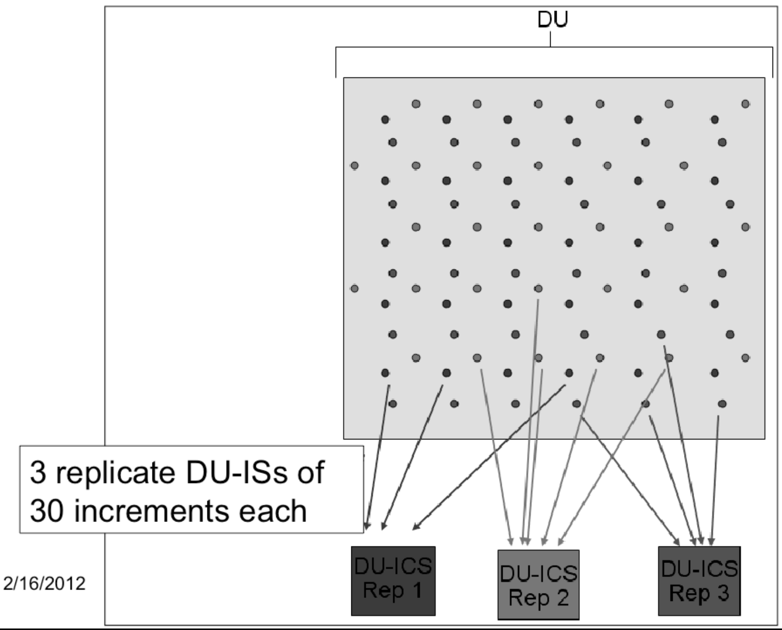
Single incremental sample (IS) covers a decision unit (DU)

Definitive guidance is the ITRC ISM Tech Reg web doc



Single DU with 30 increments (happen to have a plug-shaped sample support) go into a single incremental sample (IS)

Independent Replicate ISs



2/16/2012 27

Justification for Default of 30 Increments

- If not determining # of increments statistically, can use default of 30 increments per ICS per DU.
  - Each replicate also has 30 increments; so if using 3 replicates => total of 90 increments per DU
- Justification/conditions for 30:
  - Central Limit Theorum of classical statistics
  - Field studies & experience (see USEPA guidance next slide)
  - Computer simulations (see ITRC ISM Tech Reg)
  - Expected conditions: ¼- to ½-acre residential lots for generating average & UCL for exposure unit (EU)
    - Provides good sampling density that accommodates moderate heterogeneity

2/21/2012 Clu-In Incremental-Composite Webinar Module 2.2 28

Central Limit Theorum (CLT) states that under certain conditions, the mean of a “sufficiently” large number of independent random variables, each with finite mean and variance, will be approximately normally distributed (Wikipedia, as of 15Feb2012).

Translated to English: Take a population, such as the total number of possible samples in the field. The statistical distribution of that population can be non-normal (such as gamma distributed, lognormally distributed, or nonparametrically distributed). Now, repeatedly sample that population (say 200 times using computer simulations) using a “sufficient” number of samples, and calculate the mean for each repeat sampling event. That will produce a data set of 200 means. Then take that set of 200 means and plot its statistical distribution. That distribution of means will be normal (or close to normal) if a “sufficiently” large number of samples were used to sample the original population. The question is: how many is “sufficient”? It turns out that the sufficient number depends on how non-normal the original population was. The more non-normal the original population, the more samples that need to be taken to get a normal distribution when the means are plotted. But a “rule-of-thumb” that statisticians use is that 30 samples seems to be sufficient for most applications, but more are needed if the variability in the original population is high.

What does this mean for incremental sampling? An incremental sample represents a physical mean of a set of samples, which are actually the increments in our application. If you take 30 increments and make an incremental sample, unless the contamination is highly variable across the DU, most times 30 increments will be enough so that the data set of say, 200 incremental samples (all from the same DU, and each made of 30 increments), will form a normal distribution.

This is important because when we take only 3 replicate incremental samples, there are not enough data points to test what statistical distribution those 3 data points come from. (We need to know that to determine how to calculate the UCL.) But since they are made of 30 increments, we can assume that those 3 incremental samples came from a normal distribution AS LONG AS the contaminant heterogeneity is not too bad across the DU. So if we know that the contaminant heterogeneity is not too bad (for how bad, see the ITRC guidance discussion in Section 4.3.4.1), we can invoke the Central Limit Theorum and calculate a UCL from 3 ISs using the Student’s t-distribution. If the contaminant heterogeneity across the DU is pretty bad, or if we don’t know how bad it is, we should play it safe and not use the Student’s t-distribution to calculate a UCL. Instead we should use the Chebyshev formula to calculate a UCL. A Chebyshev UCL is always more conservative (i.e., higher) than a t-UCL.

ITRC ISM Tech Reg is at http://www.itrcweb.org/ism-1/

Calculating the Concentration Term – EPA 1992

On page 3:

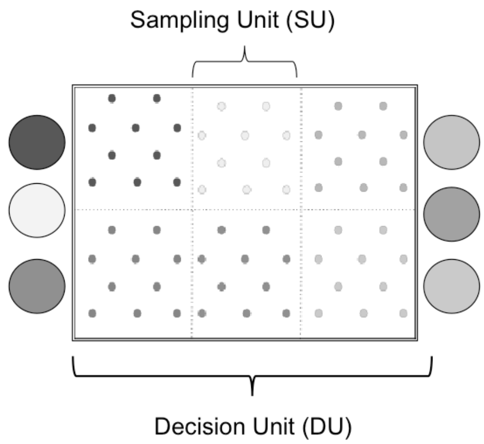
CALCULATING THE UCL

How many samples are necessary to calculate the 95 percent UCL?

Sampling data from Superfund sites have shown 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 sample mean and the 95 percent UCL), while data sets with 10 to 20 samples per exposure are provide somewhat better estimates of the mean, and data sets with 20 to 30 samples provide fairly consistent estimates of the mean (i.e., the 95 percent UCL is close to the sample mean). Remember that, in general, the UCL approaches the true mean as more samples are included in the calculation.

Example of A Non-Overlapping Composite Design

- DU contains 6 SUs
- 6 composite SU samples composed of 10 increments each: 60 total
- All SU samples immediately analyzed
- Increments distributed systematically over adjacent SUs
- Preserves spatial info
- Can calculate UCL



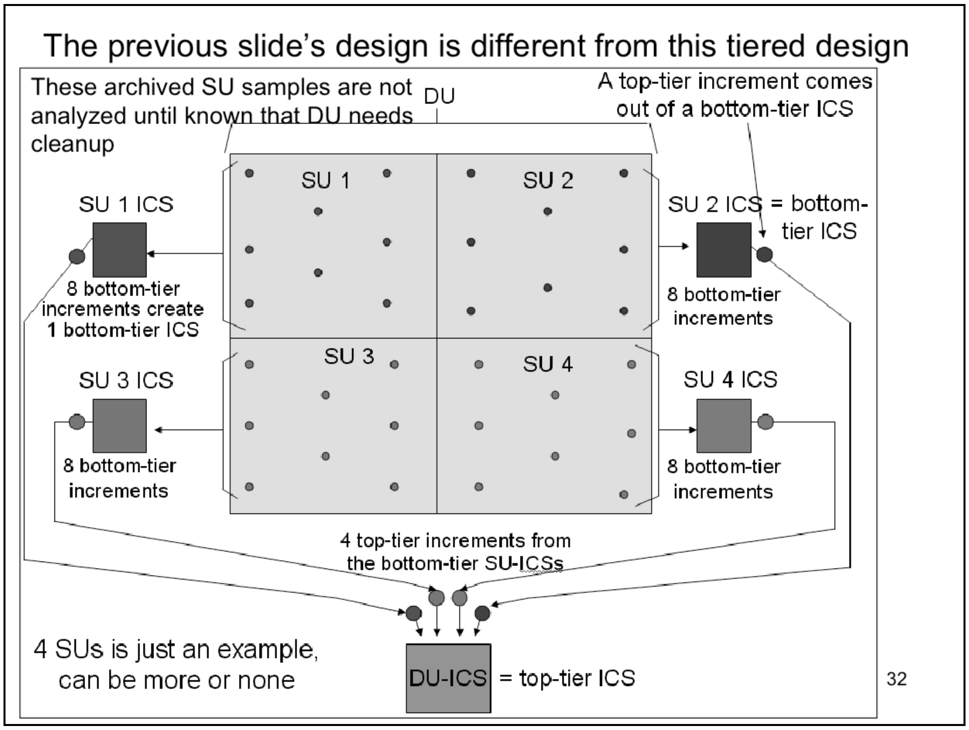
A Non-Overlapping Composite Design
(Determine a DU Mean, UCL & Hot Spots)

- If composites are non-overlapping, similar to discrete samples, but more representative of "area of inference"
- Increased sample support of the composite reduces data variability (reduces the skewed character of data distribution)
- These are not replicates, so data variability > variability between ISM replicates
  - Because spatial variability exists, cannot assume data will be normal!
- If have 8 or more, can calculate UCL with ProUCL
- If <8, safest to use the nonparametric Chebyshev formula

$$UCL = \bar{x} + \left(\sqrt{\frac{1}{\alpha} - 1} \right) \frac{s}{\sqrt{n}}$$

2/21/2012

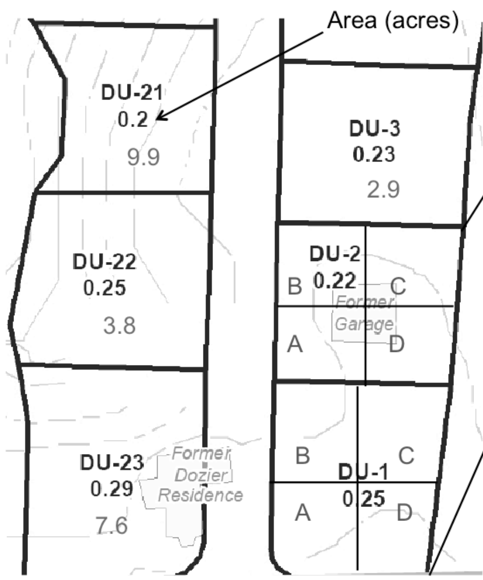
31



This is a tiered ICS design. The 1st DU-ICS sample is formed from a portion from each of the 4 SU-ICS samples. If a UCL is required, 2 replicate DU-incremental samples can be collected. Only 1 of the DU-incremental samples need go through the tiering procedure, since the SU samples are only to indicate where high results are. If done carefully the first time, should not need 3 sets of SU data to indicate high concentration SUs.

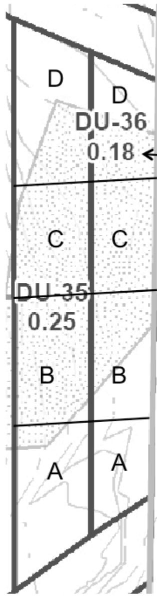
So, initially there would be 3 replicates DU-ICSs from which to calculate a UCL on the mean. If the UCL exceeds, then you can go back & analyze the archived SU samples to determine where the concentrations are high. This contrasts with the design on the previous slide where all SU samples are analyzed from the start and the UCL is calculated from the mean and standard deviation of the SU samples (which are not replicates of each other). For a site with contamination only in 1 or 2 areas, the UCL for the non-overlapping composite design can be expected to be higher than the UCL from the tiered design on this slide, because the UCL for the tiered design is generated from 3 estimates of the DU mean (the 3 replicate DU-ICSs). For a site where either there is no contamination, or the contamination is mostly uniformly distributed across the DU, the UCL might be about the same for both the tiered incremental design and the non-overlapping composite design. When the non-overlapping composite design has more than 3 SUs, its UCL might be lower simply due to the higher n in the UCL equation. (A higher n lowers the UCL when all other inputs are the same.)

Tiered Design (Residential Lot EUs)



Sample ID	DU-1	DU-2
DU-wide (0-2 in)	2.5	7.4
SU-A	10.5	3.2
SU-B	4.6	4.6
SU-C	4.1	5.2
SU-D	28.2	4.1
Math Ave of 4 SUs	**11.8**	**4.5**

Results as dioxin TEQ (ppt)
36 increments per DU
4 SUs per DU
9 increments per SU (0-2 inches)

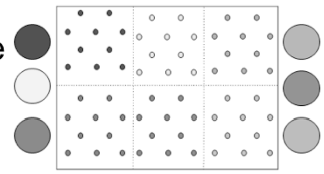


Tiered Design (Characterization Area)

Sample ID	DU-35	DU-36
DU-wide	536	486
SU-A	3974	4026
SU-B	26.7	138
SU-C	7.3	1277
SU-D	13.5	27.2
Math Ave of 4 SUs	**1005**	**1370**

Results as dioxin TEQ (ppt)
36 increments per DU
4 SUs per DU
9 increments per SU (0-2 inches)

Specifics for this Composite Design

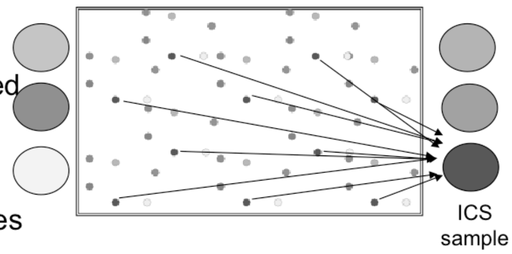


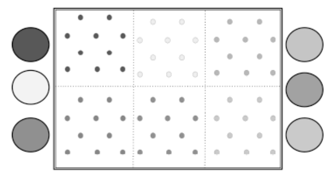
- How many increments per composite
- How many composites
- Key parameters:
  - Degree of spatial heterogeneity & how close to the action level concentrations are expected to be
  - As true mean gets closer to action level, the more increments needed to bring the UCL below the mean (assuming the true mean < action level)
- VSP can be used to estimate these numbers
- VSP covered in Calculations module (Day 3)

A Different Composite Design: Replicate DU Composites

6 replicate composite samples composed of 10 increments each

- Each set of increments distributed over entire DU
  - Increments can be placed randomly or systematic
  - VSP can estimate increments & composite numbers & place samples
- Contrast against previous non-overlapping composite design
- For uniform DUs, UCL may be lower than 30-inc ISs w/ n=3 because of higher n





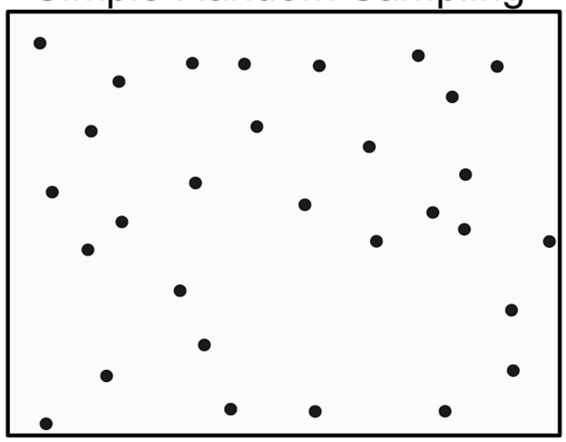
2/21/2012 Clu-In Incremental-Composite Webinar Module 2.2 36

For DUs with variable spatial distributions of contamination, the UCL for this replicate DU composite design may be higher than the corresponding UCL for a 3-replicate 30-increment DU-incremental sampling design. This can be true even though the number of replicate ICSs (n in the UCL equations) would be higher for the replicate DU composite design, which would tend to lower the UCL. However, because the number of increments per DU-composite (10 in this example) is less than the number of increments with a full DU-incremental sampling design (usually around 30), there will probably be more variability in the data set from DU-composite replicates (6 data points in this example), thus increasing the UCL.

36

Quick Review of Random Patterns of Sample Placement (1)

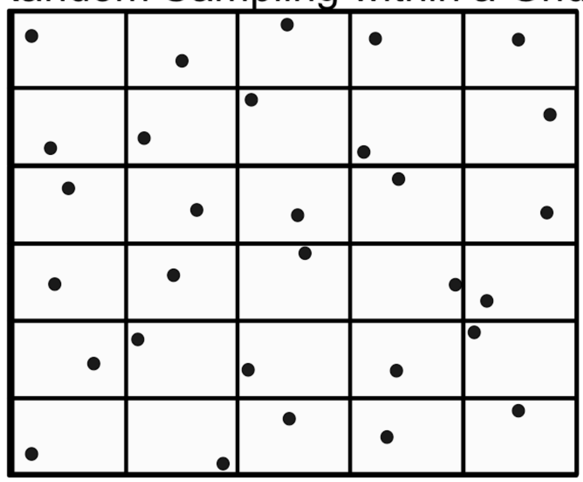
Simple Random Sampling



2/21/2012 37

Quick Review of Random Patterns of Sample Placement (2)

Random Sampling within a Grid

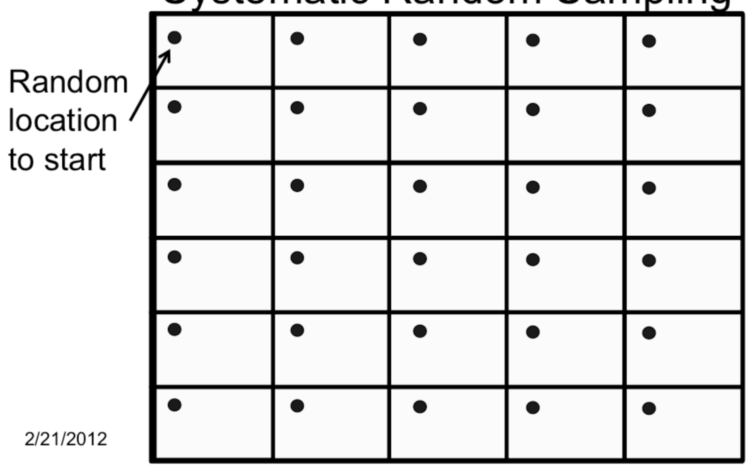


2/21/2012

38

Quick Review of Random Patterns of Sample Placement (3)

Systematic Random Sampling



An Option if Need Both an Estimate of Average AND Spatial Information AND Quick Field Analysis Available

- As collect the increments to form the ICS, screen each increment with a real-time technique
  - Provides information on spatial distribution for reduced cost
  - A great example is XRF for metals
  - Similar techniques available for radioisotopes

Incremental-Averaging Benefits

- Compared to discrete sampling, significantly reduce analytical costs for the same performance
- Significantly improve decision-making performance for the same budget (analytical $ diverted to pay for increased sampling $ that reduces error)
  - Reduce decision-making errors
  - ***Much more likely that “hot spots” will be found and accounted for if composite sampling is properly designed***

What about Risk Assessment?

- Exposure unit (EU) MUST be carefully defined—MUST work with risk assessor when DEVELOPING the sampling design
- The EU is the DU area over which the average will be determined
- USEPA guidance (covered Day 1) makes clear that the EU's average is the exposure point concentration (EPC) value
- 95% UCL is a conservative estimate of the average; is used to be sure that the data set mean is = or > the DU true mean 95% of the time

Module 2.3

Composite Searching

Improving Hot Spot Detection

43

Looking for Problems

- Examples:
  - Does contamination exist at a site?
  - Does contamination exist at a site above levels of potential concern?
  - Are there **"hot spots"** that need to be addressed?
- Again – definitions are fundamental:
  - ***How is the decision unit defined???***
  - ***How is a "hot spot" defined???***
    - Need conc & the area over which that conc applies

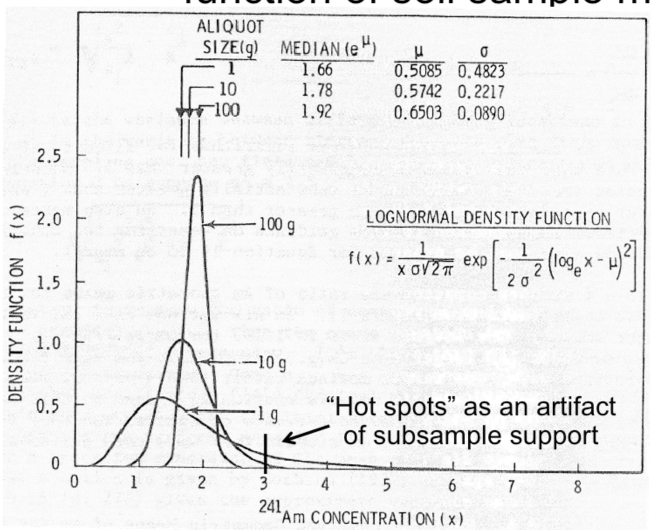
Goals Are Different From Those for Incremental-Averaging

- Not really interested in accurately knowing the average concentration
- Only interested in knowing whether concentrations are above or below some threshold
- The issue is how reliably can we identify situations when contamination is truly is above the relevant threshold given our sampling strategy

Decision Units and Searching for Contamination

- When looking for evidence of contamination (e.g. SI or RI phase), decision units are often not well-defined
- For contaminants that are not naturally occurring, contamination evidence is a positive detection or positive detection above some threshold
- For contaminants that are naturally occurring, comparison is often to a background threshold value (BTV) and/or to some screening level

Recall: variability & statistical distribution is a function of soil sample mass



Adapted from DOE (1978) americium-241 study

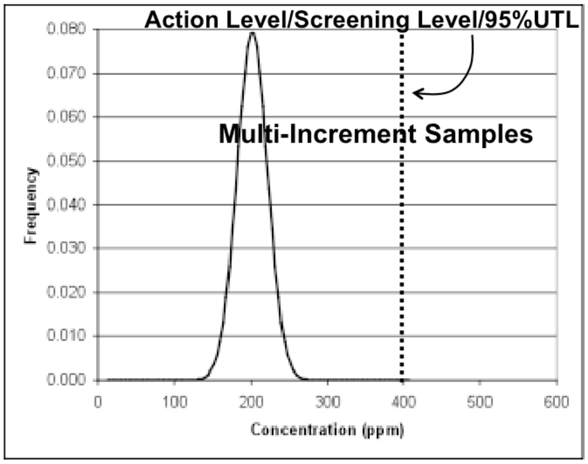
Each sample support has its own distribution of data values.

The larger the sample support, the closer to normal the distribution of data values becomes, and the fewer results appear in a distribution tail.

Explains why repeat sampling can fail to find the "hot spot"

Here's the problem:

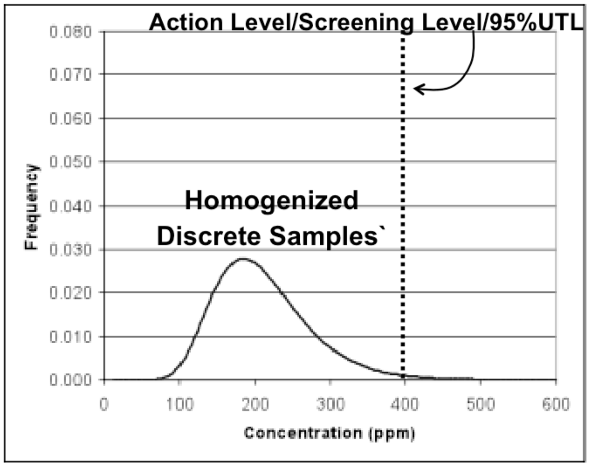
Underlying sample population distribution is tightly tied to the sample support: different sample supports have different statistical distributions



Large sample support (assume sample is properly processed for subsampling)

Here's the problem:

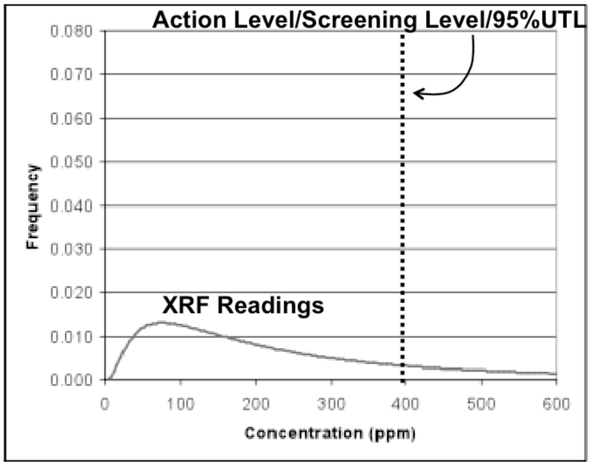
Underlying sample population distribution is tightly tied to the sample support: different sample supports have different statistical distributions

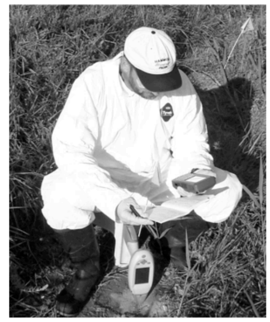


Medium sample support (assume sample is properly processed for subsampling)

Here's the problem:

Underlying sample population distribution is tightly tied to the sample support: different sample supports have different statistical distributions





Very small sample support

“Hot spot” Must be Defined

- Cannot scientifically defend defining a “hot spot” as any exceedance of an action level by a single sample result
- A scientifically defensible hot spot will be defined by concentration AND by the area/ volume over which that concentration is averaged

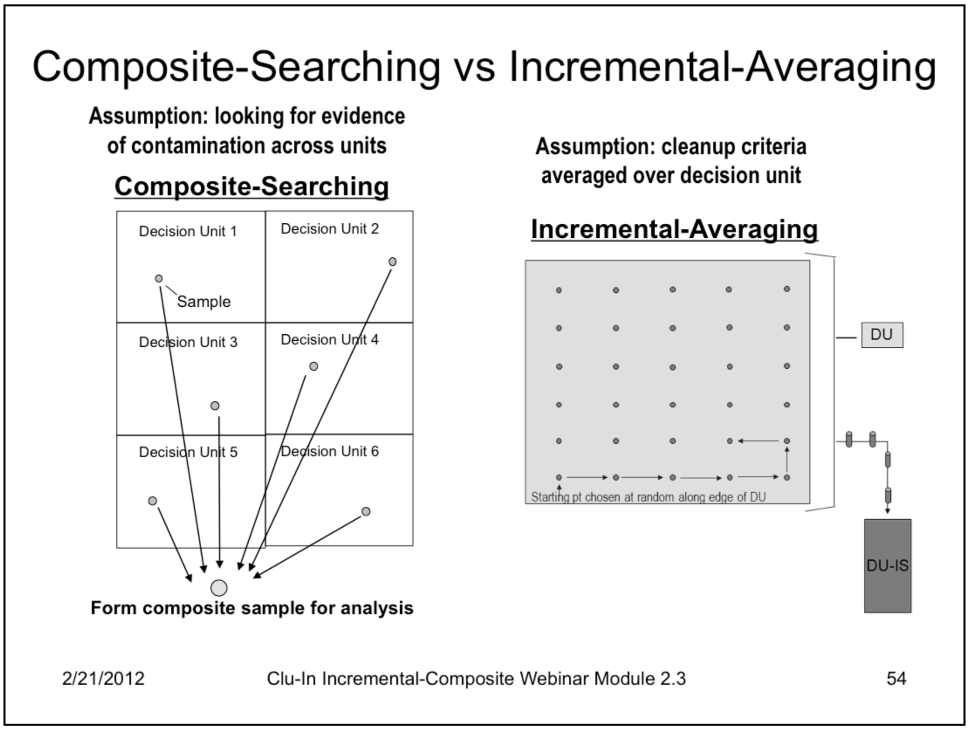
Be Careful!

- Incremental sampling may not be appropriate when evaluating the presence of contamination if the threshold is based on a discrete sample result
- When comparing a composite result to a BTV, it is essential that the background area used to develop the BTV be sampled with the same incremental-composite protocols
- Best when we are explicit about the derivation of the "decision unit" that is the basis for the decisions we are making, as well as the meaning of "hot spot"

UTL = upper tolerance level

A Consideration for Defining "Hot Spot"

- Hot spot properties (conc & area/volume) are set through negotiation between technical staff
- Balance what is ideal vs. what is practical
- Risk: a receptor is physically exposed to a mass, not to a concentration.
  - Conc alone not enough to assess exposure, must know how much media is involved
  - Example: common soil conc unit = mg analyte/*kg* soil
    - 500 ppm Pb = 500 mg Pb/kg soil
    - If ingest 1 g soil, exposure mass is 0.5 mg Pb
    - If ingest 100 g soil, exposure mass is 50 mg Pb



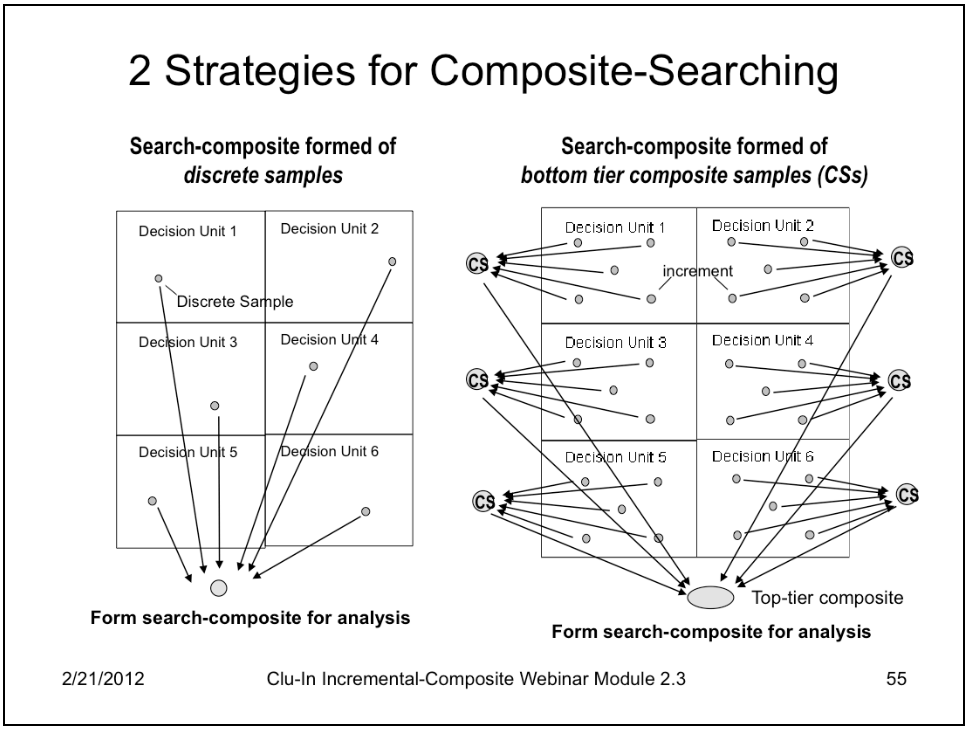
Here's the distinction between composite averaging and composite searching for the purposes of this presentation.

In the case of composite averaging (figure on the right), we collect multiple samples or soil increments from within a decision unit and then combine them into a composite sample for analysis. Our goal is to determine whether the average concentration within the decision unit is less than some cleanup criterion.

In the case of composite searching (figure on the left), we collect multiple samples or soil increments from either across decision units (as illustrated here) or within decision units and then combine them into a composite sample for analysis. Our goal is to determine whether any of those original increments might have had contamination levels above some specified threshold that would be indicative of the presence of contamination at levels of concern.

The balance of this discussion will focus on composite averaging. Composite searching will be discussed later (also referred to as adaptive compositing).

An important side point: “Dilution” is not a concern for composite averaging. “Dilution” is a concern for composite searching.



Here's the distinction between composite averaging and composite searching for the purposes of this presentation.

In the case of composite averaging (figure on the right), we collect multiple samples or soil increments from within a decision unit and then combine them into a composite sample for analysis. Our goal is to determine whether the average concentration within the decision unit is less than some cleanup criterion.

In the case of composite searching (figure on the left), we collect multiple samples or soil increments from either across decision units (as illustrated here) or within decision units and then combine them into a composite sample for analysis. Our goal is to determine whether any of those original increments might have had contamination levels above some specified threshold that would be indicative of the presence of contamination at levels of concern.

The balance of this discussion will focus on composite averaging. Composite searching will be discussed later (also referred to as adaptive compositing).

An important side point: “Dilution” is not a concern for composite averaging. “Dilution” is a concern for composite searching.

Adaptive Compositing Strategies for Searching

- Goal is to identify elevated areas
  - Looking for contamination > designated action level
- Assumptions:
  - Contamination believed to be spotty
  - Action level significantly > background levels
  - Sample acquisition/handling costs significantly < analytical costs
  - Appropriate methods exist for sample acquisition & aggregation

(continued)

56

Adaptive Composite-Searching (cont'd)

- Aggregate samples (discrete or IS) into composites for homogenization and analysis
  - Split each discrete (or IS, as the case may be)
  - Composite 1 set of splits
  - Archive the other set for re-analysis if necessary
- For the design:
  - Determine appropriate number of samples to composite (see next few slides)
  - Develop decision criteria for composites that indicate when analyses of archived splits are necessary (next)

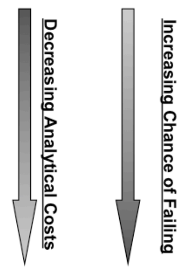
57

Recipe for Adaptive Composite-Searching

- Determine appropriate number of samples to composite & resulting decision criteria; use equal masses/volumes
- Decision criteria = [(action level - background) / (# of samples in composite)] + background
- Homogenize well & split samples – use one set of splits to form composites and archive other set
- If composite result < decision criteria
  - No more sampling required
- If composite result > decision criteria
  - Analyze archived splits contributing to composite

Example Decision Criterion

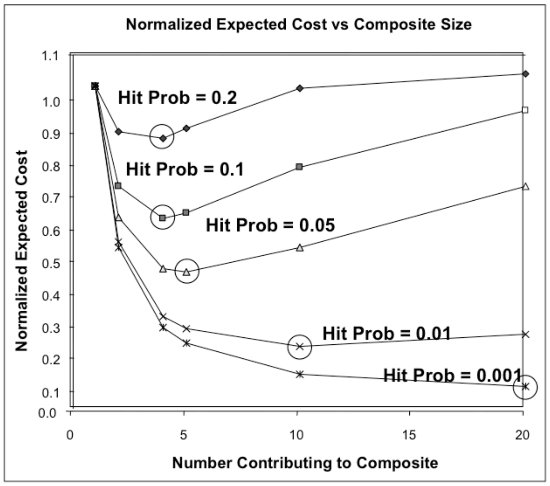
- Background: 10 ppm, Action Level: 100 ppm
- Determine decision criteria for 2-sample, 3-sample, 4-sample, 5-sample, and 6-sample composite:
  - 2-sample composite: 55 ppm
  - 3-sample composite: 40 ppm
  - 4-sample composite: 33 ppm
  - 5-sample composite: 28 ppm
  - 6-sample composite: 25 ppm
  - Don't want to get too close to background



When is Adaptive Compositing Cost-Effective?

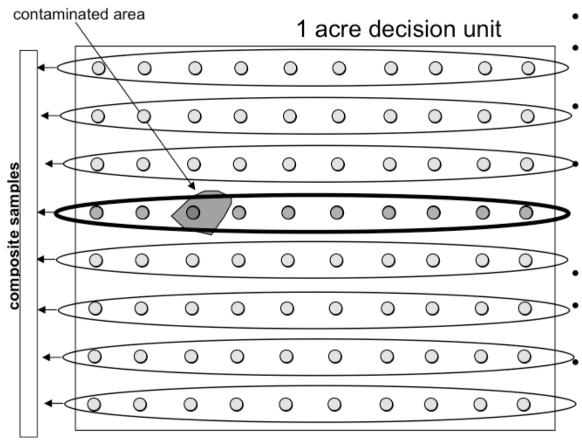
- The "spottier" contamination is, the better the performance (in contrast to discrete sampling)
- The greater the difference is between background and the action level, the better the performance
- The greater the difference between the action level and average contamination concentration, the better the performance
- Best case: no composite requires re-analysis
- Worst case: every composite requires re-analysis (will cost more than if the samples were just analyzed as discretes from the start)

Optimize Cost: How Many Samples to Composite?



- Is a function of the probability that contamination is present such that a composite will exceed the criteria
- Balance against the cost of going back & analyzing the archived increment splits when a hit occurs (which is why the lines go back up)
- The less likely it is that contamination is present, the more samples that can be composited
- The circled point identifies the ideal sample number to composite based likelihood of exceedence

Simplified Example…

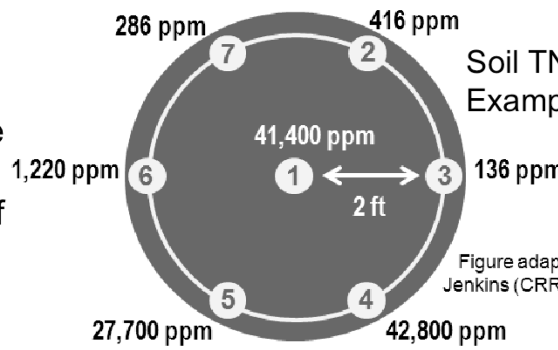


- Looking for PCBs > 50 ppm
- Could be anywhere in area of concern 1 acre in size
- Discrete sample/analyze all approach – 80 samples
- Alternative: take 80 samples and organize into 8 composites consisting of 10 samples each
- Decision criteria: **5 ppm**
- Would need to analyze 8 composites
- Plus analyze 10 archived samples from the guilty composite
- Find hotspot & calc average over all or portion of DU

More than 75% analytical cost reduction!!

A Flawed Assumption About "Hot Spots"…

– Average concentration across seven samples is 16,000 ppm
– Suppose "hot spot" level were 5,000 ppm
– Only a 43% chance of correctly identifying the presence of this hot spot if one relied on discrete samples



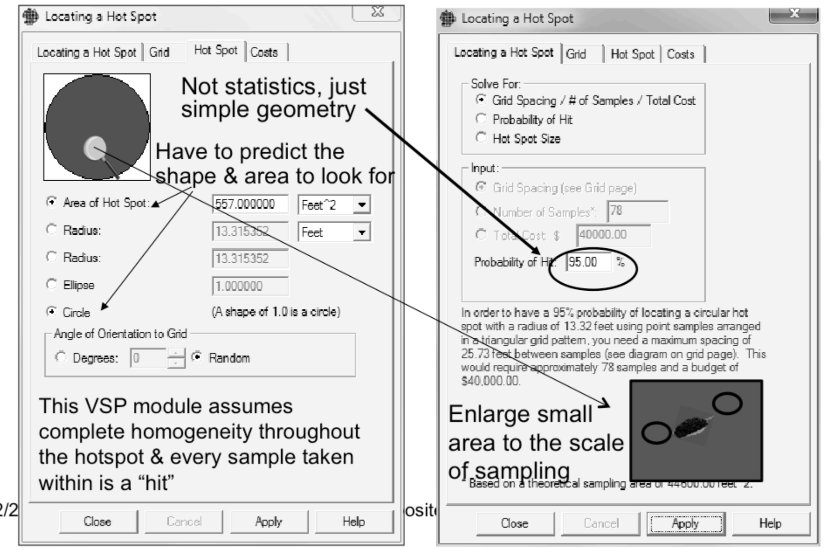
Soil TNT Example

Figure adapted from Jenkins (CRREL), 1996

2/20/2012 63

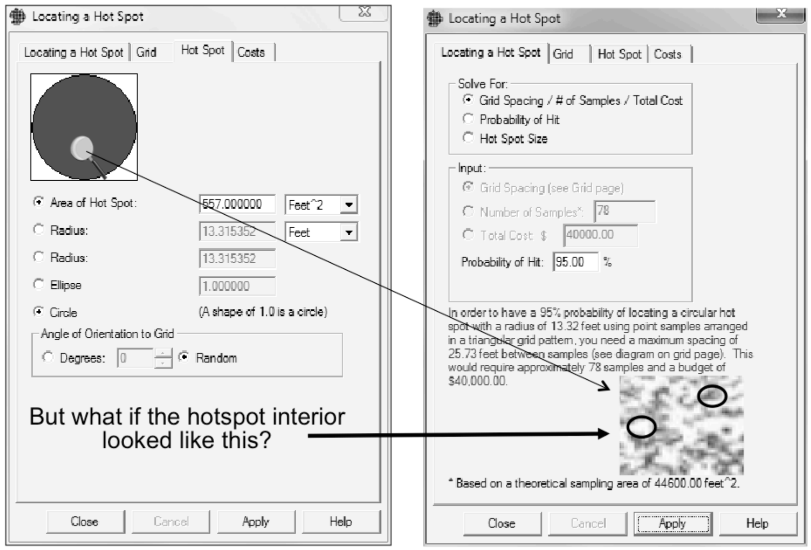
Why Use Tiered Compositing for Hotspot Detection?

Hot spot search algorithms can neglect short-scale heterogeneity

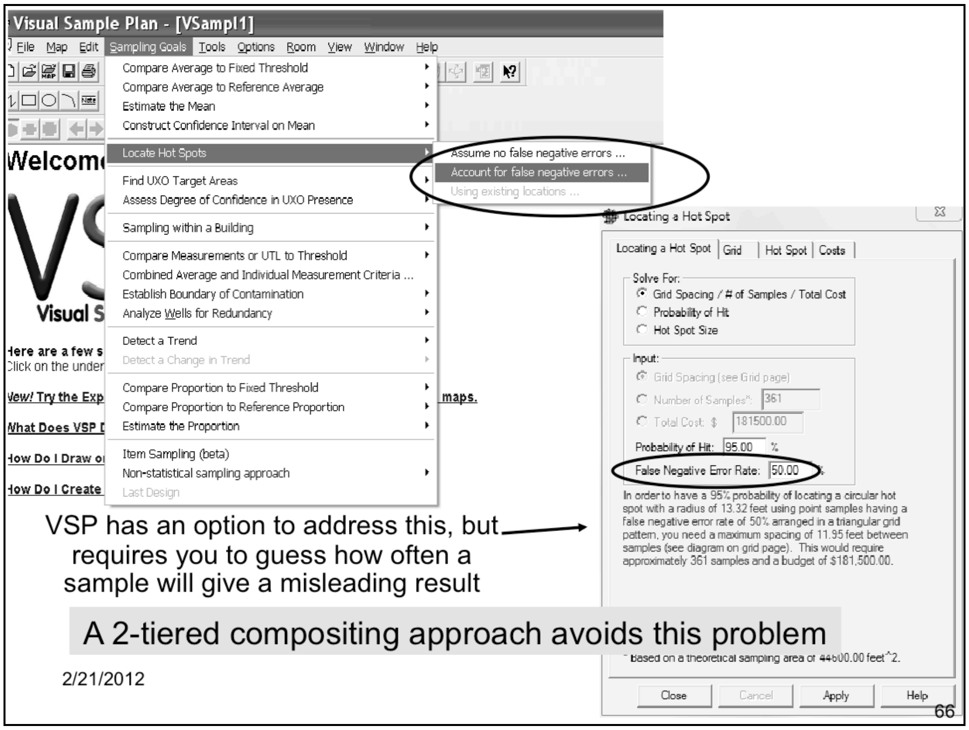


Why Use Tiered Compositing for Hotspot Detection?

Because the real world is not homogeneous



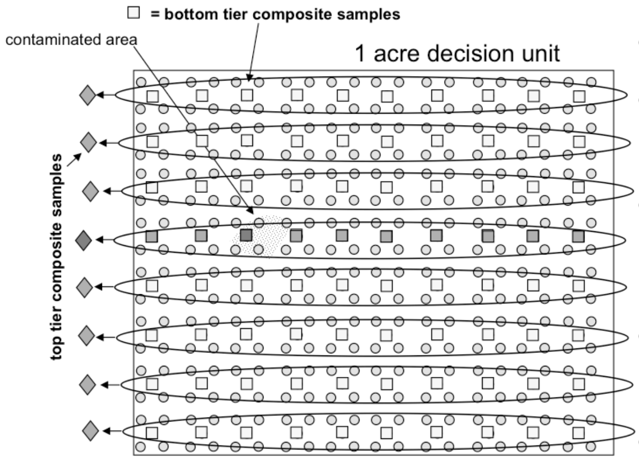
65


Visual Sample Plan - [VSampl1]
File Map Edit Sampling Goals Tools Options Room View Window Help
Compare Average to Fixed Threshold
Compare Average to Reference Average
Estimate the Mean
Construct Confidence Interval on Mean
Locate Hot Spots
Assume no false negative errors ...
Account for false negative errors ...
Using existing locations ...
Find UXO Target Areas
Assess Degree of Confidence in UXO Presence
Sampling within a Building
Compare Measurements or UTL to Threshold
Combined Average and Individual Measurement Criteria ...
Establish Boundary of Contamination
Analyze Wells for Redundancy
Detect a Trend
Detect a Change in Trend
Compare Proportion to Fixed Threshold
Compare Proportion to Reference Proportion
Estimate the Proportion
Item Sampling (beta)
Non-statistical sampling approach
Last Design
Locating a Hot Spot
Locating a Hot Spot | Grid | Hot Spot | Costs
Solve For:
Grid Spacing / # of Samples / Total Cost
Probability of Hit
Hot Spot Size
Input:
Grid Spacing (see Grid page)
Number of Samples*: 361
Total Cost: $ 181500.00
Probability of Hit: 95.00 %
False Negative Error Rate: 50.00 %
In order to have a 95% probability of locating a circular hot spot with a radius of 13.32 feet using point samples having a false negative error rate of 50% arranged in a triangular grid pattern, you need a maximum spacing of 11.95 feet between samples (see diagram on grid page). This would require approximately 361 samples and a budget of $181,500.00.
Close Cancel Apply Help
VSP has an option to address this, but requires you to guess how often a sample will give a misleading result
A 2-tiered compositing approach avoids this problem
2/21/2012
66

Incremental-Averaging & Composite-Searching Can Be Combined

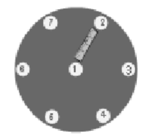
- Purpose
  - provide area average over larger decision units while effectively detecting hot spots at the same time
- Design
  - A “bottom-tier” of area-averaging is done on a small “local” (collocated) scale within a larger “top tier” strategy (minimizes error due to short-scale heterogeneity…prior example did not)
  - “Top-tier” strategy of incremental-averaging over the whole decision unit
  - Area-wide composite-searching for hot spot/pattern detection
- Provides overall mean estimate for the DU (or portions) while controlling analytical costs

Combined Design Example



- 400 soil increments ○ grouped by 5's
- Form 80 bottom-tier □ composite samples representative of 80 small areas
- 10 bottom-tier composites combined to form 8 top-tier composites, which are analyzed (◇ &◆)
- 10 archived bottom-tier splits required analysis due to the hit for a top-tier composite (◆)
- Rationale: improve likelihood of "seeing" contamination

Review Outputs from Combined Design



1. High density sampling coverage: controls short-scale heterogeneity & its problems
2. Finds hot spots: can go back & do finer delineation if desired
3. Can calculate DU average over whole DU or only selected parts
4. Provides vastly improved information for about same, or less, cost than much less informative sampling designs
5. Supports high confidence, high precision remedial designs

Any Questions?

Module 2.4

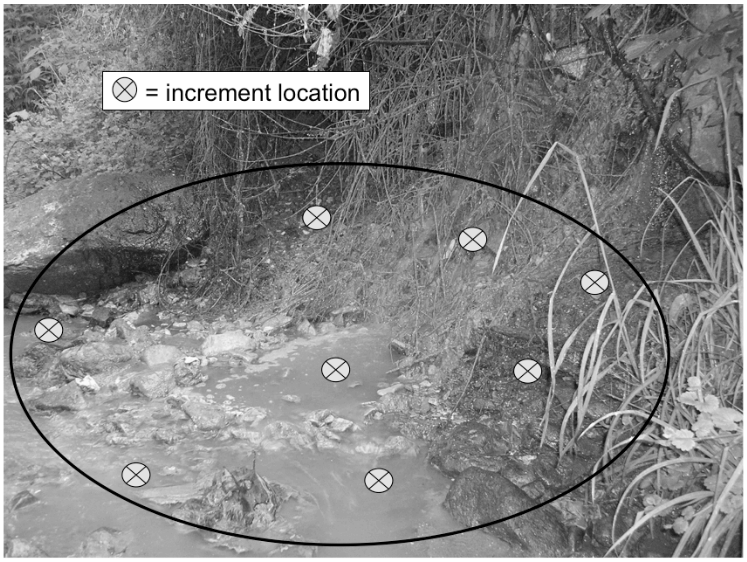
Composite-Incremental Sampling:

Limitations, Caveats and QC

71

Every soil sampling program can be improved by using some level of composite sampling in the field or in lab subsampling.

Do you trust a single small grab sample to determine a release?



Greater coverage by composite sampling makes it less likely to miss

Incremental-Composite Limitations

- Easier for surface than subsurface soil
  - Sample acquisition costs higher for subsurface
- Requires specialized handling for volatile contaminants (Hg, VOCs) (covered in ITRC ISM-1)
- Might not be as useful:
  - When *in situ* measurement techniques (*in situ* XRF) are available, depending on circumstances
  - Where cost of analysis is comparable to cost of sample acquisition

74

ICS Warnings

- Inadequate sample processing cancels field benefits
- Direct comparisons with other data sets (e.g., background) collected with a different protocol may not be appropriate
- Be very careful…
  - …about pooling incremental/composite results directly with discrete sample results (not advisable)
  - …when data goes into database, notes regarding DU size & increment number should be recorded
  - …when cleanup standards are based on different decision unit dimensions or sample support than an ICS strategy might reflect

The Need for Sample Processing

- Any form of composite sampling requires stringent sample processing to be effective
  - This is true for discrete samples too!
- Some commercial laboratories are investing in equipment and procedures that address soil processing
- If you are planning a project, discuss sample handling with labs or put requirements in RFPs
  - Does the lab have the equipment and know-how?
  - Find out BEFORE you get too far down the road!!

Measuring Adequacy of Sample Processing

- Routine lab QC provides information on sample-related precision
  - Field splits
  - Lab duplicates
  - Matrix spikes/matrix spike duplicates (MS/MSDs)

Practices that Reduce Within-Sample Heterogeneity

- IF the CSM identifies a particle size less than 10-mesh (2 mm) as the population of interest (e.g., an exposure pathway), sieving has the side effect of reducing particle effects.
- Other good practices:
  - Incremental subsampling using a “slabcake procedure” and taking increments to make the analytical sample
  - Increasing the analytical mass (to be digested or extracted)
  - Grinding/milling
  - See ITRC ISM-1 Sections 5 & 6 for complete discussions

Improve the Representativeness of Subsamples

- Take many increments to make up the analytical subsample ("incremental subsampling") →
- Equipment like rotary splitters

79

• Unlike routine discrete sampling programs, ISM specifically addresses sample support issues. A project team using ISM <u>must</u> consider the likelihood of nuggets, the analytical subsample's volume and particle size.

• Reducing the overall particle size by grinding prior to subsampling may sometimes be required.

• Increasing the mass of the subsample and incremental subsampling are common ways to reduce subsampling error.

• If a field sample needs to be split, there are specialized equipment and techniques, such as rotary splitters. Choice of technique is heavily dependent on soil properties.

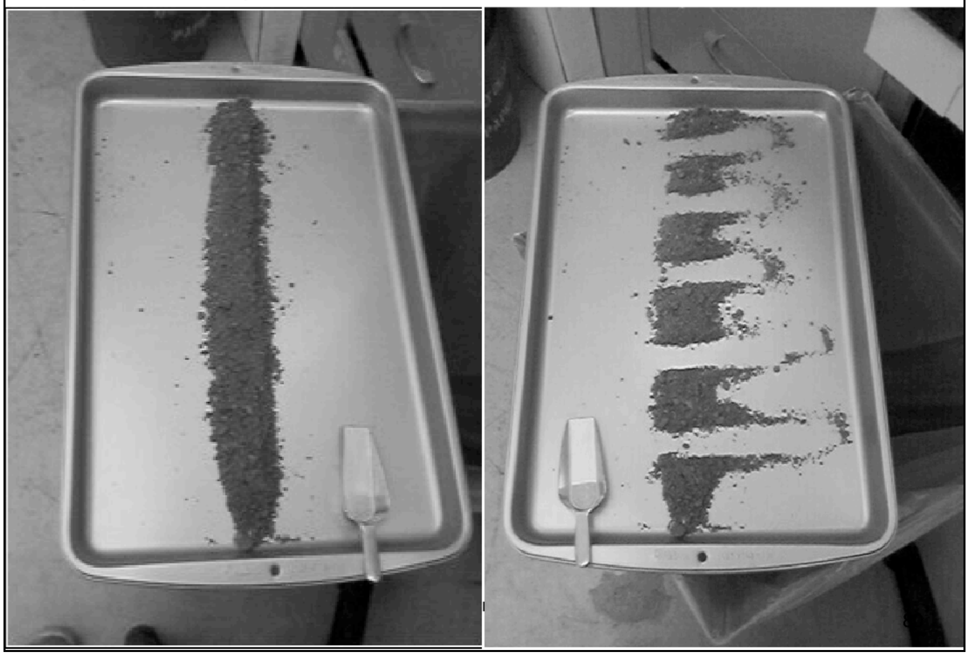
Supplemental Information

See ISM-1 Chapter 6

See also EPA guidance documents:

> • "Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples", EPA/600/R-03/027 (Nov 2003); and
>
> • "RCRA Waste Sampling Draft Technical Guidance", EPA 530-D-02-002 (August 2002), Chapter 6

Type of Gy-based Sample Volume Reduction



What about Grinding?

- Sample grinding is an open, unresolved issue
  - Important mechanism for processing samples
  - Allows smaller analytical subsamples to be representatively collected from large field samples.
  - Analytical methods usually designed to digest 0.5 - 2 g for metals
  - Increasing analytical sample mass generates more chemical waste and requires different lab equipment
- But there are concerns about grinding…

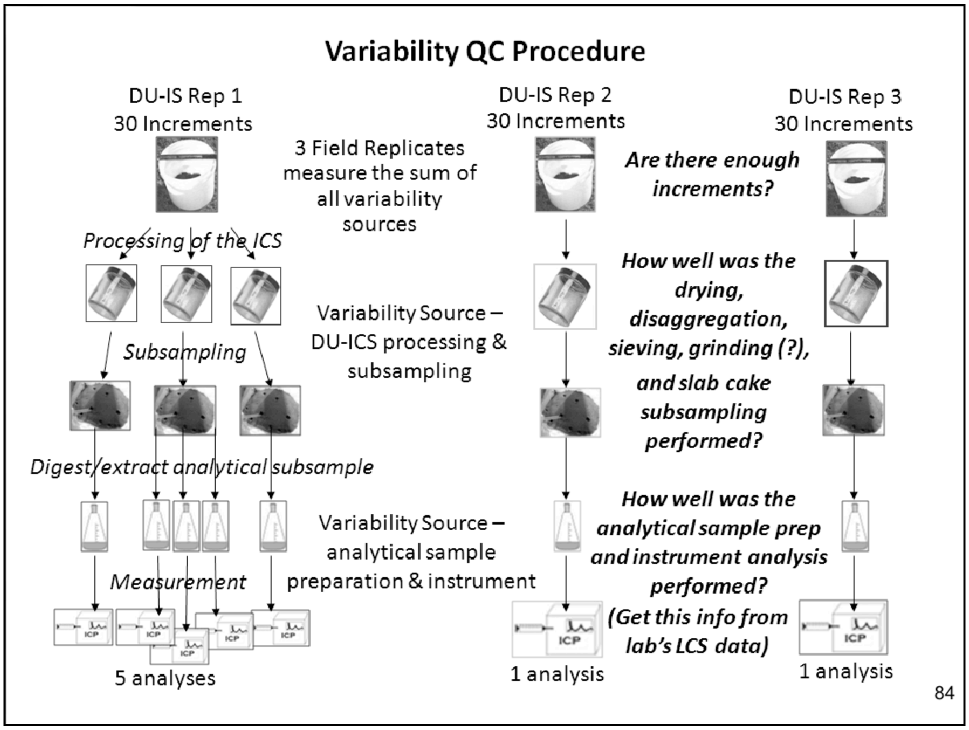
Grinding/Milling Concerns

- Does grinding increase solubilization of metals not released by typical acid digestion
  - May lead to higher sample concentrations that are less representative of bioavailable exposure
- This question is under investigation
  - Early studies suggest that increased solubilization can occur when grinding large size particle to much smaller sizes
  - Increased solubilization not noticeable when native particle size is already small (silts and clays)
  - May be a matrix, analyte, and site-specific phenomenon
- Need to make sure grinder does not add target metals (e.g., Cr from stainless steel)

Quality Assurance for Incremental-Compositing Designs

- IMPORTANT
- 3 replicates commonly used to calculate a UCL
  - Also provides a metric to assess overall quality
- Replication at the subsampling level provides a measure of sample processing adequacy and where to target corrective action
  - All for the cost of 4 additional analyses

QC procedures best if performed as part of a pilot study so that sampling and handling design can be perfected BEFORE the main sampling event.


Variability QC Procedure
DU-IS Rep 1
30 Increments
3 Field Replicates measure the sum of all variability sources
DU-IS Rep 2
30 Increments
Are there enough increments?
DU-IS Rep 3
30 Increments
Processing of the ICS
How well was the drying, disaggregation, sieving, grinding (?), and slab cake subsampling performed?
Subsampling
Variability Source – DU-ICS processing & subsampling
Digest/extract analytical subsample
Variability Source – analytical sample preparation & instrument
How well was the analytical sample prep and instrument analysis performed?
(Get this info from lab's LCS data)
Measurement
ICP
5 analyses
1 analysis
1 analysis
84

Variability Partitioning Equation

- Theoretical equation:

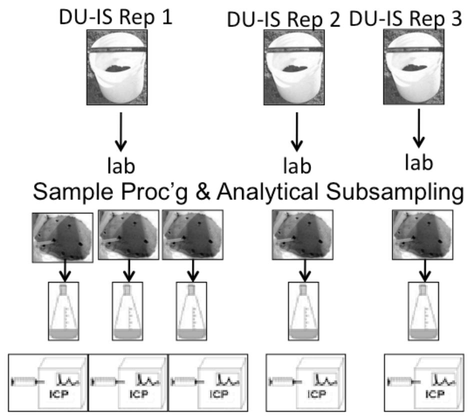
$$SD_{Total}^2 = SD_{LCS\text{-}instrument}^2 + SD_{analytical\ subsample}^2 + SD_{IS\ processing}^2 + SD_{bet\text{-}IS\ samples}^2$$

- In actual projects, likely not get all the info needed to partition variability to 4 sources
- Actual data example to illustrate follows

Total variability = sum of the variability for each component/step

Example from Actual Data

- 3 replicate IS samples collected (total error)
- Whole samples sent to lab for sample processing
- 3 replicate analytical subsamples
- For Sb & Pb, field heterogeneity was dominant error source
- For As, field & subsampling errors about equal



Analyte	Ave of Field Reps 1, 2 & 3	SD of Field Reps 1, 2 & 3	Field Rep 1 3 Subsmple Subsmple Ave	Field Rep 1 3 Subsmple Subsmple SD	Total measure-ment error (SD)	Total measurement error (RSD) (SD/mean*100)	Analytical + sample pro-cessing error (SD)	Analytical + sample pro-cessing error (RSD)	Field-scale (within-DU) error (SD)	Field-scale (within-DU) error (RSD)
ANTIMONY	4.0	1.5	5.1	0.8	1.5	38	0.8	16	1.3	33
ARSENIC	6.8	0.7	8.2	0.9	0.7	10.3	0.9	10.9	analytical > total	
LEAD	670	263	977	61	263	39	61	6	256	38

86

Most Helpful As Part of Pilot Study

- A pilot study can provide many benefits
  - Assess sources of data variability
  - If necessary, select corrective actions to reduce largest source
  - Use opportunity to fill CSM gaps or test critical assumptions underlying the sampling design
  - Determine optimal number of increments and/ or number of IS field replicates
  - Use as readiness review for field & lab staff

Potential Corrective Actions (1)

- For reduction of error (variability) from short- & long-scale heterogeneity
  - Increase mass of field increments
  - Increase number of field increments
- Improve sample handling/homogenization prior to splitting sample or subsampling
  - Break up clods better (coffee mill, mortar & pestle)
  - More careful sieving
  - Milling

Potential Corrective Actions (2)

- For reducing error in analytical subsampling
  - Increase number of increments in subsample
  - Increase mass of the analytical subsample
  - Improve rigor of analytical subsampling
    - Use more "correct" (per Gy theory) sampling tool
    - Exercise more care when preparing 2-D slabcake (need to avoid segregation of particles)
  - Perform replicate analytical subsampling & average them for a single result

Any Questions?

Resources & Feedback

- To view a complete list of resources for this seminar, please visit the **Additional Resources**
- Please complete the **Feedback Form** to help ensure events like this are offered in the future

Need confirmation of your participation today?

Fill out the feedback form and check box for confirmation email.

91