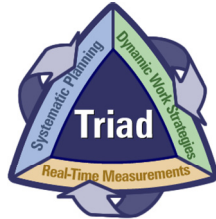# Advanced Design Application & Data Analysis for Field-Portable XRF

A Series of Web-based Seminars Sponsored by
Superfund's Technology & Field Services Division

Session 6

Q&A for Session 5
Module 6.1 – Dynamic Work Strategies Part 1

6.1-1

How To . . .

◆ Ask questions

  » "?" button on CLU-IN page

◆ Control slides as presentation proceeds

  » manually advance slides

◆ Review archived sessions

  » http://www.clu-in.org/live/archive.cfm

◆ Contact instructors

6.1-2

Notes

◆ When you registered, you were directed to this seminar's specific URL, which is the front page of today's seminar. The Front Page of the web cast contains a short abstract of today's session. We have also included pictures and short biosketches of the presenters. Please note the presenters' email addresses are hotlinked on that page in case you have any questions for one of them after today's presentation.

◆ For those of you joining us via the phone lines, we request that you put your phone on mute for the seminar. We will have Q&A sessions at which point you are welcome to take your phone off mute and ask the question. If you do not have a mute button on your phone, we ask that you take a moment RIGHT NOW to hit *6 to place your phone on MUTE. When we get to the question and answer periods you can hit #6 to unmute the phone. This will greatly reduce the background noises that can disrupt the quality of the audio transmission.

◆ Also, please do not put us on HOLD. Many organizations have hold music or advertisements that can be very disruptive to the call. Again, keep us on MUTE. DO NOT put us on HOLD.

◆ Also, if you experience technical difficulties with the audio stream, you may use the ? icon to alert us to the technical difficulties you are encountering. Please include a telephone number where you can be reached and we will try to help you troubleshoot your problem.

◆ Instructor contact information:

Deana Crumbling, U.S. EPA
Phone: (703) 603-0643
Fax: (703) 603-9135
E-mail: crumbling.deana@epa.gov

Robert Johnson, Argonne National Laboratory
Phone: (630) 252-7004
Fax: (630) 252-3611
E-mail: rlj@anl.gov

Stephen Dyment, U.S. EPA
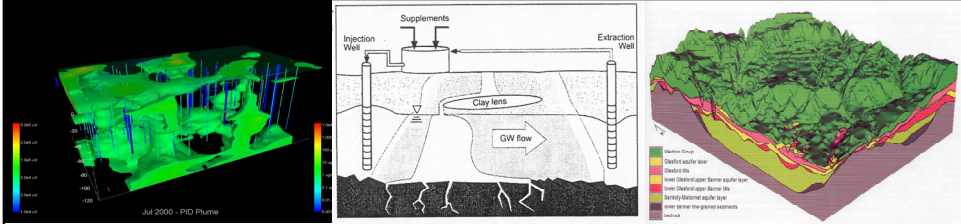Phone: (703) 603-9903
Fax: (703) 603-9135
E-mail: dyment.stephen@epa.gov

Q&A For Session 5 – Quality Control

6.1-3

Module 6.1:

Dynamic Work Strategies – Part 1

6.1-4

Improving XRF Data Collection Performance Requires...

◆ **Planning systematically** (CSM)

◆ **Improving representativeness**

◆ **Increasing information** available for decision-making

◆ **Addressing the unknown** with dynamic work strategies

6.1-5

Notes

The next two modules include four broad areas. Today's module will cover the first three of these. The last, different forms of dynamic work strategies, will be the focus of the second module.

◆ **Planning systematically (CSM):** The systematic planning process involves preparation of a conceptual site model, which is then used as the foundation for further work and is updated as the site becomes better characterized. Systematic planning also includes other important steps that will be discussed later.

◆ **Improving representativeness:** The XRF data that is collected should be representative of the actual site conditions in the decision units being investigated. There are many ways to improve the representativeness of the data.

◆ **Increasing information available for decision-making:** XRF data can increase information available for decision-making by providing a more dense, and therefore, more reliable picture of site conditions.

◆ **Addressing the unknown with dynamic work strategies:** Dynamic work strategies are adjusted to site conditions as they are learned, which makes subsequent data more and more useful for decision-making.

Systematic Planning and Data Collection Design

Planning Systematically

◆ Systematic planning defines decisions, decision units, and sample support requirements

◆ Systematic planning identifies sources of decision uncertainty and strategies for uncertainty management

◆ Clearly defined cleanup standards are critical to the systematic planning process

◆ Conceptual Site Models (CSMs) play a foundational role

6.1-6

Notes

◆ **Systematic planning defines decisions, decision units, and sample support requirements:** During the systematic planning process, the decisions to be made are clearly articulated and the spatial boundaries of the decision units are defined. The sample support requirements for data collection are also determined during the planning process.

◆ **Systematic planning identifies sources of decision uncertainty and strategies for uncertainty management:** Sources of decision uncertainty include any factor that may hamper the ability to make a decision, such as changing regulatory requirements, reuse issues, and site characterization issues. The planning process seeks to identify all sources of decision uncertainty and lay out a strategy for addressing and managing the uncertainty.

◆ **Clearly defined cleanup standards are critical to the systematic process:** The systematic planning process depends on the identification of clearly defined cleanup standards. A complete definition of cleanup criteria includes the area over which the standard, on average, is to be applied. Beware of "never to exceed" standards! These give the semblance of "conservative" cleanups but in fact are impossible to verify with technically defensible cleanup programs, and are susceptible to sample support complications.

◆ **Conceptual Site Models (CSM) play a foundational role:** The CSM becomes the foundation for all investigative and cleanup work to be conducted at the site. It represents the best understanding of the conditions at the site and is the tool for incorporating new information and planning future work.

The Conceptual Site Model (CSM) is Key to Successful Projects

Planning Systematically

Not to be confused with a fate/transport or exposure scenario model (although these may be components).

THE basis for cost-effective, confident decisions

◆ Decision-maker's mental picture of site characteristics pertinent to risk and cleanup

◆ A **CSM** can include any component that represents contaminant populations to make predictions about

» Nature, extent, and fate of contamination,

» Exposure to contamination, and

» Strategies to reduce risks from contamination

6.1-7

Notes

◆ **Decision-maker's mental picture of site characteristics pertinent to risk and cleanup:** The CSM is the decision maker's concept or mental picture of the site characteristics as they pertain to human health and environmental risk and cleanup. The CSM that results from systematic planning is not the same as the fate/transport or exposure scenario model that is developed for risk assessments, although an exposure scenario model may be a component of the CSM.

◆ **A CSM can include any component that represents contaminant populations to make predictions about:** The CSM includes any component that represents site conditions and makes predictions about the following:

» Nature, extent, and fate of contamination

» Exposure to contamination, and

» Strategies to reduce risks from contamination

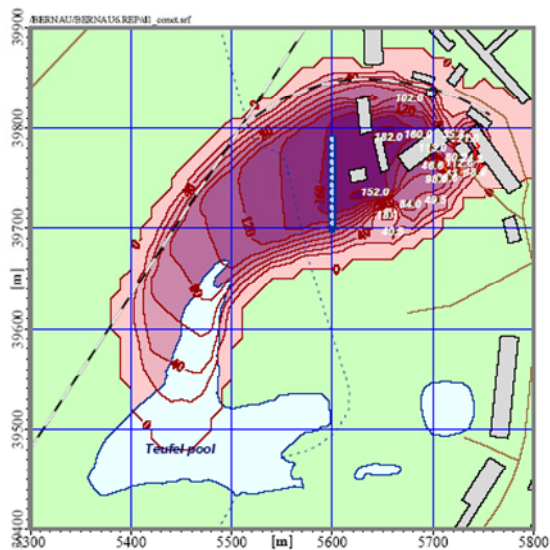# How well does the idealized mental model match reality?

Planning Systematically

6.1-8

Notes ◆ This type of graphic is often used to convey a sense of the spatial distribution of contaminants. Unfortunately, it gives the sense that location and level of contamination is known with a high degree of precision, although this kind of graphic might be based on only a few data points. It also suggests that contaminant levels are smoothly varying spatially, and that within a particular isopleth, two samples close together would give approximately the same answer. Finally, it gives no sense at all of whether there is uncertainty present in what it depicts, or where that uncertainty is greatest.

The World is Usually Messier Than Models Portray

Planning Systematically
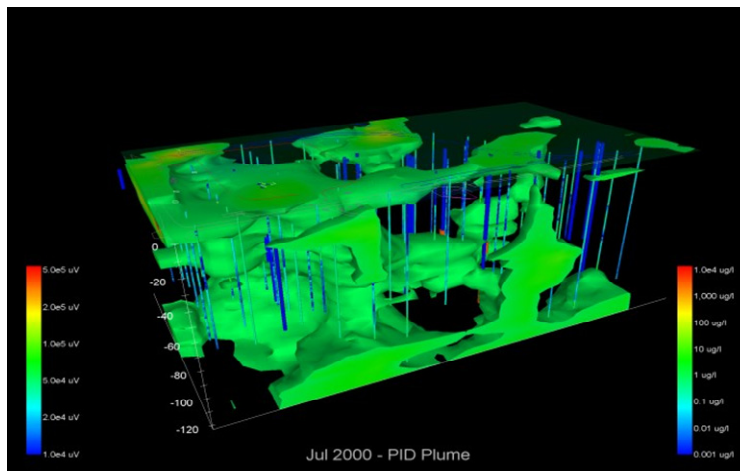
Slide adapted from Columbia Technologies, Inc., 2003

Jul 2000 - PID Plume

(Subsurface CSM from high density data using DP-MIP sensing)

6.1-9

Notes

◆ The reality of contaminant distributions in soils is really much more complicated. This particular graphic was based on relatively dense MIP results, and shows how complex the subsurface distribution of contaminants can be. Even this is likely a simplification of reality, since the "blobs" one sees are being interpolated between soil cores.

CSMs Are Critical!!

◆ Whether or not openly articulated, the CSM is the basis of <u>all</u> site decisions.

Planning Systematically

Preliminary CSM predicts contaminant distributions

Prediction guides development of SAP

Data confirms or modifies predictions as CSM gradually matures

Mature CSM is the basis for decisions & all subsequent activities

◆ The CSM is the working hypothesis about the site's physical reality, so working without a CSM is like working blind-folded!

6.1-10

Notes

◆ **Whether or not openly articulated, the CSM is the basis of all site decisions:** The CSM is the basis for all decisions about risk, remediation, and reuse. Unarticulated CSMs create conflict, are often based on untested assumptions, and lead to faulty project designs. The preliminary CSM predicts contaminant distributions and makes basic assumptions about cleanup levels and reuse. These predictions guide the development of the sampling program and the data confirm or modify the predictions as the CSM matures. The mature CSM is the basis for decisions and subsequent activities.

◆ **The CSM is the working hypothesis about the site's physical reality, so working without a CSM is like working blind-folded!:** The working hypothesis helps the investigative team make sense of the data collected at the site. Throughout the investigative process, the site team should be striving to learn the true physical reality about the contamination at the site and challenging each other when conception does not match reality.

CSMs Articulate Uncertainty

Planning Systematically

◆ CSM captures understanding about site conditions

◆ CSM identifies uncertainty that prevents confident decision-making

◆ A well-articulated CSM serves as the point of consensus about uncertainty sources

◆ Data collection needs and design flow from the CSM:

  » Data collection to reduce CSM uncertainties

  » Data collection to test CSM assumptions

◆ The CSM is living…as new data become available, the CSM is revisited, updated, and matures

6.1-11

Notes

◆ **CSM captures understanding about site conditions:** The CSM uses all existing information to provide an initial understanding about site conditions. The CSM explains what contamination is present, where the contamination is located, where the contamination may be migrating, and what types of actions may be available to address the contamination problems. In the early stages, some of these elements may be educated guesses rather than well-supported facts.

◆ **CSM identifies uncertainty that prevents confident decision-making:** The CSM identities those elements that are uncertain and for which additional information is necessary. The additional information should increase the certainty associated with the particular element so that decisions can be made with confidence.

◆ **A well-articulated CSM serves as the point of consensus about uncertainty sources:** The CSM that results from careful systematic planning represents a consensus about the sources of uncertainty and points the way forward for addressing the uncertainty.

◆ **Data collection needs and design flow from the CSM:** The CSM guides the data collection effort because it shows what data are needed to reduce CSM uncertainties and what data are needed to test CSM assumptions.

◆ **The CSM is living . . . as new data become available, the CSM is revisited, updated, and matures:** The CSM is not a static model. It is a living tool that must incorporate new data and change to reflect the new concept of reality. The CSM is mature when it reflects reality.
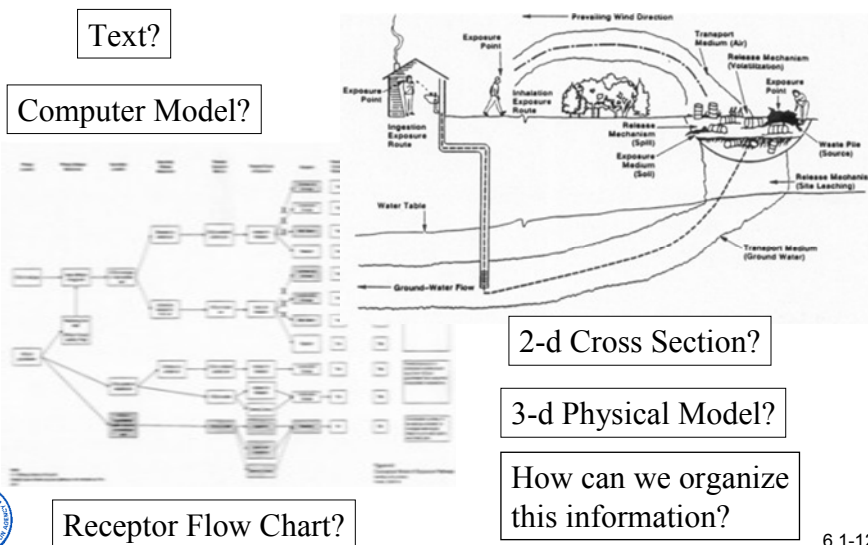
Notes

◆ But what does a CSM actually look like? The two figures depict a geologic cross-section and a receptor flow chart that are familiar to most of us. Environmental professionals often conjure up images such as these when they hear the term "CSM." Both are correct but directed towards different ways of understanding the site. A CSM will always include a textual description. A CSM might also be described with results from a computer model, or displayed with an actual physical model.
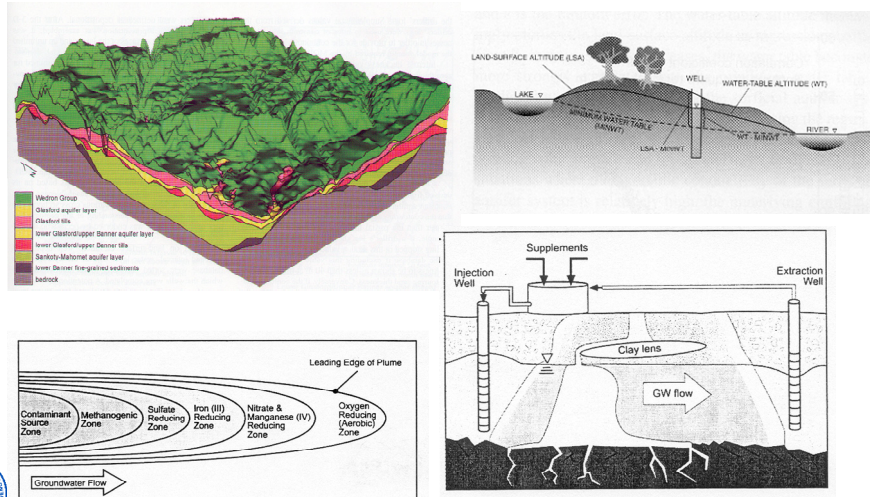
Notes

◆ There are many other visual images that can also be imagined as a component of a CSM. These figures, beginning on the upper left, depict a 3-dimensional geologic model, a schematic hydrogeologic cross-section, a plan view of changes in groundwater geochemistry, and a schematic cross-section depicting a groundwater treatment process. The organization of the CSM must therefore be flexible enough to consider varied information and requirements.

The CSM and XRF

Planning Systematically

◆ The following CSM elements are critical to consider when conducting systematic planning that involves use of the XRF
  » Decisions driving the data collection
  » Spatial definition of decisions or action levels
  » Contaminants of concern and their action levels
  » Matrix characteristics/co-contaminants that might affect XRF
  » Spatial contamination patterns (shotgun, air deposition, etc.)
  » Degree of short-scale (intra-sample) heterogeneity at action levels
  » Degree of longer-scale (between sample) heterogeneity at action levels
  » Vertical layering of contaminants

6.1-14

Notes

◆ **The following CSM elements are critical to consider when conducting systematic planning that involves use of the XRF:** The CSM that supports and guides an XRF investigation must address the following elements in order to be successful:

» Decisions driving data collection – determine what exactly is being decided, which may driven by what phase the project is entering. Is one interested in average concentrations across a yard? Systematically looking for hot spots in a larger area? Determining the depth of contamination via soil cores? Defining the boundaries of a contaminated area?

» Spatial definition of decision/action levels – define decision units and/or the action levels that will apply.

» Contaminants of concern and their action levels – assess ability of XRF to detect the contaminants

» Matrix characteristics/co-contaminants that might affect XRF – assess the potential for interference affects

» Spatial contamination patterns (shotgun, air deposition, etc.) – define sample supports

» Degree of short-scale (intra-sample) heterogeneity at action levels – define sample supports

» Degree of longer-scale (between sample) heterogeneity at action levels – define sample supports and sample design

» Vertical layering of contaminants – define sample supports and sample design, determine whether surface in situ readings are appropriate

Improving Data Representativeness

Improving Representativeness

◆ Sample support
  » matching sample support with decision needs
  » field of view for in situ analyses
◆ Controlling within-sample heterogeneity
  » Appropriate sample preparation important (see EPA EPA/600/R-03/027 for additional detail)
  » Uncertainty effects quantified by appropriate sub-sample replicate analyses
◆ *Controlling short-scale heterogeneity*
  » multi-increment sampling
  » aggregating in situ measurements

6.1-15

Notes

◆ **Sample support:** Data representativeness can be improved by developing appropriate sample supports that match sample support with decision needs and that improve the field of view of the XRF for in-situ analyses. This topic was covered in depth by the 2nd and 3rd module and so will not be discussed further as part of this module.

◆ **Controlling within-sample heterogeneity:** The heterogeneity inherent within a single sample can be reduced by careful sample preparation and homogenization, a topic thoroughly discussed in previous modules. Uncertainty effects can be quantified by appropriate sub-sample replicate analysis using laboratory methods. An XRF has some very interesting applications as a way of checking the effectiveness of sample preparation that we will discuss a bit further.

◆ **Controlling short-scale heterogeneity:** Short-scale heterogeneity can be controlled using multi-increment sampling and by aggregating in-situ measurements. We will be talking at length about this.

Verifying Sample Preparation by XRF

Improving Representativeness

◆ XRF can play a unique role in verifying sample preparation
  » XRF measurements are non-destructive
  » XRF measurements are fast
◆ Works when XRF-detectable metals are either primary COCs or are correlated with primary COCs
◆ Perform multiple (e.g., 5 to 10) direct measurements on sample (bagged or exposed) pre- and post-preparation
◆ Target samples expected to have contamination around action levels
◆ Review resulting measurement variability
◆ Can be part of a DMA and/or part of on-going QC

6.1-16

Notes

◆ **The XRF can play a unique role in evaluating the effectiveness of sample preparation:** An XRF measurement is fast and non-destructive.

◆ **Works when XRF-detectable metals are either primary COCs or are correlated with primary COCs:** The XRF can be used to check sample preparation when either the primary COC is an element measurable by the XRF, or when a metal measurable by XRF is collocated with the primary COC and strongly correlated from a concentration perspective.

◆ **Perform multiple (e.g., 5 to 10) direct measurements on sample (bagged or exposed) pre- and post-preparation:** To verify sample preparation, bag a sample and measure through the bag multiple times prior to sample preparation, then prepare the sample, re-bag, and re-measure the bag by XRF.

◆ **Target samples expected to have contamination around action levels:** It is best to target samples that have contamination concentrations close to the action level.

◆ **Review resulting measurement variability:** Comparing the variability (i.e., standard deviation) observed in pre-preparation XRF data with that observed in post-preparation XRF data will indicate how effective the preparation process was in reducing within-sample heterogeneity.

◆ **Can be part of a DMA and/or part of on-going QC:** This type of evaluation can be done as part of a pre-field work DMA or built into an on-going QC process.

Within-Sample Variability is a Function of Concentration

Improving Representativeness

◆ 100 bagged samples

◆ Analyzed multiple times for lead

◆ Variability observed a function of lead present

◆ As concentrations rise, sample prep becomes increasingly important

◆ Important point to remember as discussion turns to MI sampling

Average Lead vs StDev
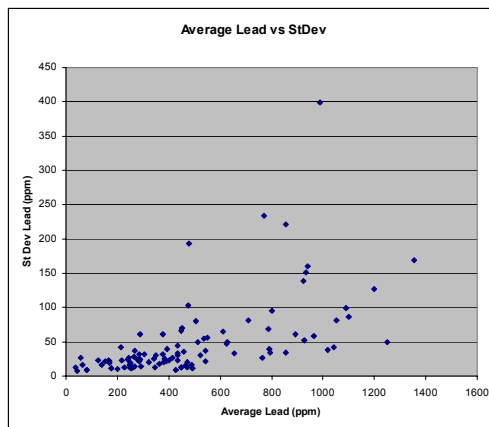
6.1-17

Notes

The contaminant heterogeneity that is present within a soil sample is, in general, a function of concentration. As concentrations increase, the variability present increases too. The data shown in this slide illustrate that fact.

◆ **100 bagged samples:** This is a data set where 100 bagged samples were analyzed as part of a lead-in-soil characterization effort. In each case the bagged sample was measured multiple times across the bag's surface by XRF, allowing calculation of both an average concentration for a bagged sample, and the standard deviation (a measure of variability) for that bagged sample's results.

◆ **Variability observed a function of lead present:** Each point in the scatter plot represents a bag. The x-axis shows the average concentration for lead. The y-axis shows the observed standard deviation. As lead concentrations increased, so, in general, did the variability as measured by standard deviation.

◆ **As concentrations rise, sample prep becomes increasingly important:** One way to interpret these data for any particular point is that the average lead value plus or minus twice its standard deviation would provide bounds on the concentration expected from a cup sample analyzed from that bag by XRF or ICP. For example, there is one bag with a lead concentration close to 500 ppm that yielded a standard deviation of 200 ppm. If a subsample from that bag were analyzed by ICP without any further preparation of the subsample, the expected result would range anywhere from 100 to 900 ppm. That bag happened to be particularly "bad" from a heterogeneity perspective, but it illustrates the point.

In contrast, the bags with concentrations less than 200 ppm had a typical standard deviation of only 16 ppm. At a bagged sample with an average concentration of 100 ppm, this would correspond to potential ICP/XRF cup readings ranging between 68 and 132 ppm.

◆ **Important to remember as discussion turns to MI sampling:** The relationship between within-bag heterogeneity is an important one to remember as this discussion moves to multi-increment sampling.

Notes

◆ Contrary to the image conjured by the cartoon, multi-increment sampling is more than just a new statistical buzzword. The notion of "multi-increment sampling" has gained some momentum in recent years. Most people would probably equate multi-increment sampling with compositing, although multi-increment sampling so far has avoided the knee-jerk, negative, "dilution" perception associated with compositing.

The goal of this module is not to unnecessarily complicate life by introducing new terms, but for the sake of clarity the remainder of this presentation will distinguish between compositing and multi-increment sampling as two distinctively different activities that share similar mechanics (i.e., bringing sample material together from different locations and homogenizing into one sample) but for completely different purposes.

Guidance on Multi-Increment Sampling/Compositing is Conflicting

Improving Representativeness

◆ *Verification of PCB Spill Cleanup by Sampling and Analysis* (EPA-560/5-85-026, August, 1985)
  » up to 10 adjacent samples allowed

◆ *Cleanup Standards for Ground Water and Soil, Interim Final Guidance* (State of Maryland, 2001)
  » no more than 3 adjacent samples allowed

◆ *SW-846 Method 8330b* (EPA Rev 2, October, 2006)
  » 30 adjacent samples recommended

◆ *Draft Guidance on Multi-Increment Soil Sampling* (State of Alaska, 2007)
  » 30 – 50 samples for "compositing"

6.1-19

Notes

◆ **Guidance on multi-increment sampling/compositing is conflicting:** It's not surprising there are different perceptions about whether compositing samples is a good or bad thing. The fact is the guidance (state and federal) currently is conflicting and confusing.

In 1985, TSCA sampling guidance recognized that compositing might be a cost-effective way of looking for PCBs, and went so far as to recommend compositing up to 10 adjacent samples.

In contrast, in 1991, the State of Maryland put our interim final guidance for cleanup standards and said that no more than 3 samples should ever be combined to form a composite.

In 2006, a revision to Method 8330b recommended compositing as many as 30 adjacent samples before analysis.

More recently, draft Alaskan guidance talks about compositing 30 to 50 samples during characterization work.
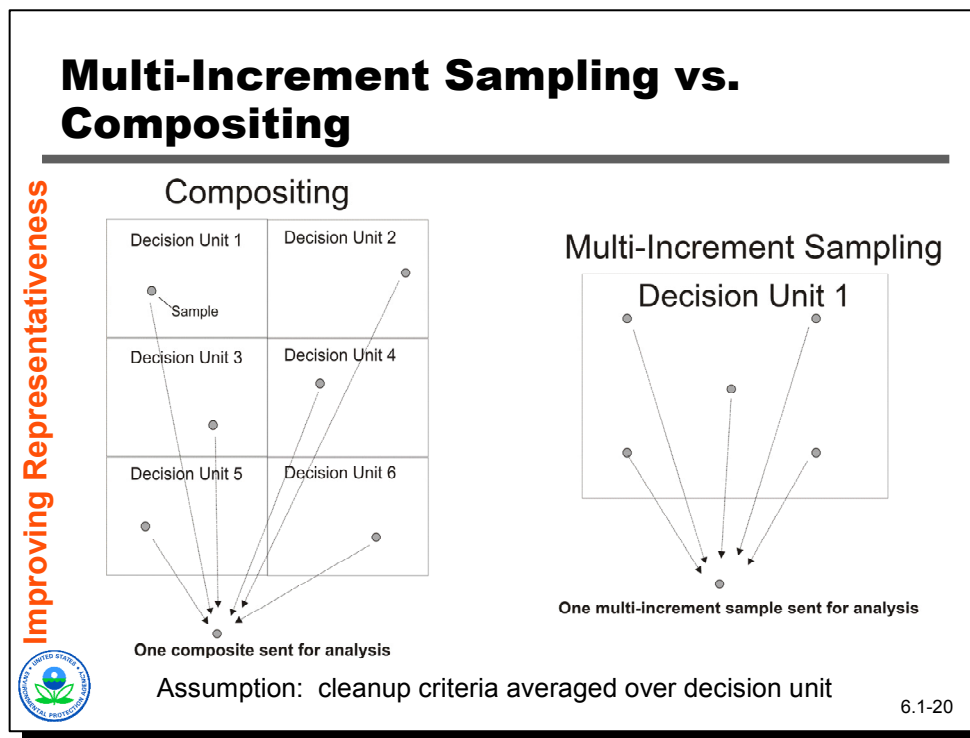
Notes

◆ This graphic shows the difference between compositing and multi-increment sampling. In compositing, samples are combined from multiple decision units into a single sample that is analyzed and used to make decisions about the presence or absence of contamination with each decision unit. In multi-increment sampling, multiple samples from a single decision unit are combined into a single sample that is homogenized and analyzed and used to make decisions about the contamination status of that single decision unit. More than one MI sample can be collected and formed for any particular decision unit.

Multi-Increment Sampling vs. Compositing

Improving Representativeness

◆ Multi-increment sampling: a strategy to control the effects of heterogeneity cost-effectively – *multi-increment averaging*

◆ Compositing: a strategy to reduce overall analytical costs when conditions are favorable – *composite searching* – topic in next module

6.1-21

Notes

◆ **Multi-increment sampling:** Multi-increment sampling is a strategy to control the effects of heterogeneity cost-effectively. The technique is also called multi-increment averaging. MIS is used to obtain better estimates of the average level of contamination present within a decision unit. Dilution is not a concern with MIS, just as one is not concerned that the mathematical average of several samples yields a result that will be less than some of the sample values contributing to the average. The term has been trademarked and EPA has approved the method for use in analyzing energetics in soil.

◆ **Compositing:** Compositing is a strategy to reduce overall analytical costs when conditions are favorable. The technique is also called composite searching. Adaptive compositing is used to systematically look for contamination issues across areas. Dilution is a potential issue with adaptive compositing, but is addressed by the use of appropriate field investigation levels. Adaptive compositing will be discussed in greater detail in the next module.

Multi-Increment Averaging

Improving Representativeness

◆ Applicable when goal is to get a better estimate of average concentration over some specified area or volume of soil

◆ Used to cost-effectively suppress short-scale heterogeneity

◆ Multiple sub-samples contribute to sample that is analyzed

◆ Sub-samples systematically distributed over an area equivalent to or less than decision requirements

◆ Effective when the cost of analysis is significantly greater than the cost of sample acquisition

6.1-22

Notes

◆ **Applicable when the goal is to obtain an estimate of the average concentration level across a decision unit:** Multi-increment averaging is applicable when the goal of the sampling is to obtain an estimate of the average concentration level across a decision unit.

◆ **Used to cost-effectively suppress short-scale heterogeneity:** The effects of short-scale heterogeneity on average concentration estimates are controlled by collecting an MI sample that is more representative of the decision unit characteristics.

◆ **Multiple sub-samples contribute to sample that is analyzed:** Multiple sub-samples are collected and mixed together in a larger sample that is prepared and analyzed. When sample preparation is conducted in a laboratory, the sample must be ground to a consistent particle size and then subsampled according to EPA's subsampling guidance.

◆ **Sub-samples systematically distributed over an area equivalent to or less than decision requirements:** The sub-samples are distributed over the area of the decision unit. The multi-increment sampling is representative of the average contamination concentration conditions within the decision unit.

◆ **Effective when the cost of analysis is significantly greater than the cost of sample acquisition:** This method is effective when is it more expensive to analyze a single sample than it is to collect the sample. This is an important assumption pertinent to the use of XRF technologies that will be discussed in greater detail in the next slide.

Concept Applies to XRF In Situ, Bag, and Cup Measurements

Improving Representativeness

◆ XRF in situ measurements - more measurements with shorter acquisition times is equivalent to multi-increment sampling (e.g., across a surface area or down a soil core)

◆ XRF bag measurements - multi-increment sampling addresses sampling error while multiple measurements on bag substitutes for sample homogenization

◆ XRF cup measurements - multi-increment sampling addresses sampling error

◆ *In general, MIS is not useful if an XRF can address the COCs of concern, although the concepts still apply*

6.1-23

Notes

◆ **XRF in situ measurement:** Taking more measurements with shorter acquisition times is equivalent to multi-increment sampling.

◆ **XRF bag measurements:** The sample in the bag is a multi-increment sample which addresses sampling error and taking multiple measurements of the bag substitutes for sample homogenization.

◆ **XRF cup measurements:** The sample used for cup sampling can be a multi-increment sampling which reduces sampling error. Cup sampling involves sample homogenization to create a uniform particle size.

◆ **In general, MIS is not useful if an XRF can adequately address the COCs of concern, although the concepts still apply:** Multi-increment sampling is not useful if an XRF can adequately address the COCs, although the concepts still apply for aggregating XRF measurements.

How Many MI Sample Increments?

Improving Representativeness

◆ Assume goal is to estimate average concentration over decision unit (e.g., a yard)

◆ VSP can be used to determine how many samples would be required if all were analyzed

◆ VSP calculation requires knowledge of expected contamination levels and the variability present

◆ Information can potentially be obtained by XRF

◆ The number of increments should be at least as great as identified by VSP

» Lumped into one MI sample for analysis?

» Apportioned into several MI samples for analysis?

6.1-24

Notes

Suppose one wants to use MI sampling. The natural question is how many soil sample increments should be used, and how many MI samples should be formed from those increments for analysis.

◆ **Assume goal is to estimate average concentration over decision unit (e.g., a yard):** The fundamental assumption, is that the project team wants to estimate the average concentration across a decision unit (e.g., a yard), and wants to use MI sampling to do that cost-effectively.

◆ **VSP can be used to determine how many samples would be required if all were analyzed:** A traditional sampling program design analysis might use VSP to calculate the required number of samples. VSP has routines to do just that.

◆ **VSP calculation requires knowledge of expected contamination levels and the variability present:** The VSP routines all require key pieces of information to do the calculations. Among those pieces of information are the expected average concentration that would be encountered in the decision unit, and the amount of variability expected in individual sample results from the decision unit.

◆ **Information can potentially be obtained by XRF:** In a traditional sampling program design effort, one would guess at the information VSP required. The XRF can be used to determine that information more exactly.

◆ **The number of increments should be at least as great as identified by VSP:** If the VSP calculations have been done, then the number of increments one would want to collect would be at least as many as the sample numbers VSP identified.

The final question would then be how many MI samples should be formed from those increments for analysis for the decision unit in question. The advice varies. Some practitioners advocate forming just one sample per decision unit and comparing the MI sample result with the action level. This clearly keeps analytical costs to the bare minimum, but provides little information about how much confidence there is that the MI sample result is an accurate representation of reality. Note that this is a strategy that's very similar to standard protocols for lead sampling in residential yards (i.e., a five-increment composite with one analysis done). An alternative would be to form several (e.g., four or five) samples from the increments and submit those for analysis. The statistics with those analytical results could then be used to determine how much confidence there is that the true average was above or below the action level.

One Additional XRF Basic Concept...

Improving Representativeness

◆ Recall that XRF relative measurement error and DL decrease with increasing count time

◆ Suppose one has established a DL goal and determined a necessary count time to achieve it

◆ *It doesn't matter whether one long shot is taken, or repeated shorter measurements with an average concentration determined from the shorter measurements!*

◆ This is why reporting <DL XRF results can be very useful…we need those results to calculate meaningful averages

◆ Particularly important for repeated in situ measurements or repeated measurements of bagged samples

6.1-25

Notes

◆ **Recall that XRF detection limits and relative analytical error drop as measurement time increases:** XRF detection limits and relative analytical error drop as instrument measurement time increases.

◆ **Suppose one has established a DL goal and determined a necessary count time to achieve it:** The project team can established a DL goal for a particular metal of interest, and then use the goal to determine the necessary associated XRF measurement time.

◆ **It doesn't matter whether one long shot is taken:** The basic XRF concept that applies here is that it doesn't matter whether that acquisition time is represented by one single acquisition, or by several sequential, shorter acquisitions whose results are then averaged…the resulting measurement error and detection limit are the same.

◆ **This is why reporting <DL XRF results can be very useful . . . we need those results to calculate meaningful averages:** This is why reporting XRF results that are less than the DL is so useful. As an example, a single 120-second acquisition is equivalent to (from the perspective of detection limits and measurement error) two 60-sec acquisitions, or three 40-sec acquisitions, or four 30-sec acquisitions, or five 24 sec acquisitions, or six 20-sec acquisitions, or twelve 10-sec acquisitions. What will happen as the acquisition times are shortened is that more and more of the individual acquisitions will potentially yield <LOD results (caused by rising DLs for individual readings). . . those results are necessary to obtain a meaningful average.

◆ **Particularly important for repeated in situ measurements or repeated measurements of bagged samples:** This concept is particularly important for bagged samples or in situ surface measurements.

How Many XRF Measurements for Bag or In Situ Shots at a Particular Location?

Improving Representativeness

◆ Assume goal is to get an accurate estimate of average bag concentration, or the concentration at a particular location

◆ Majority of cost of XRF deployment is sample preparation – bagged sample XRF readings potentially circumvent costly sample prep

◆ Select a bag or location with concentrations thought to be near action level

◆ Identify required DL and estimate XRF measurement time required for DL along with expected analytical error at action level

◆ Take ten shots and observe variability present

◆ Select measurement numbers so that observed variability divided by square root of measurement number is less than expected analytical error at the action level

6.1-26

Notes

Suppose now we are using the XRF for bagged sample analyses, or for in situ shots to determine the concentration in surface soils for a particular location.

◆ Assume goal is to accurately determine concentration in a bag, or at a particular location (so, again, an average but just at a much smaller scale)

◆ Why do this for a bagged sample? Because the majority of the cost of XRF deployment is sample preparation. Multiple measurements through a bagged sample's wall is a substitute for expensive sample preparation.

◆ Select a bagged sample or location thought to have concentrations around the action level.

◆ Identify the desired DL, associated required acquisition time, and expected XRF analytical error at the action level.

◆ Take ten shots across the bag/location and observe variability present in results.

◆ Select measurement number so that the variability divided by the square root of the measurement number is less than the expected XRF analytical error identified earlier. The acquisition time for each shot would be the original acquisition time divided by the number of measurements to be taken.

Revisiting Bagged Soil Lead Example...

Improving Representativeness

◆ Action level is 400 ppm

◆ Around 400 ppm, XRF measurement error < 5% for 120-sec readings

◆ Around 400 ppm, typical standard deviation ~ 34 ppm (or ~ 8%)

◆ 4 30-sec shots per bag would reduce error for bag lead estimate to less than 5%

Average Lead vs StDev
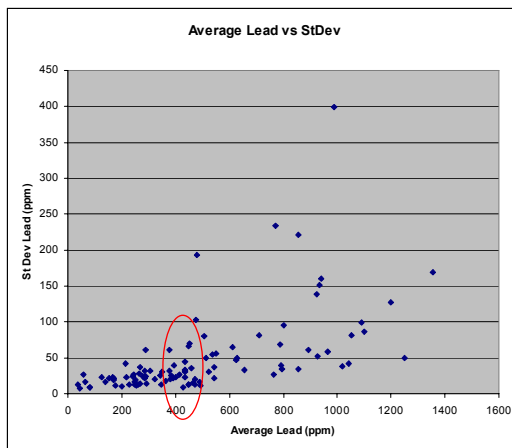
6.1-27

Notes

◆ The previous lead-in-soil bagged sample example illustrates how to calculate the number of measurements to take over a bagged sample. In this example, the action level is 400 ppm. Around 400 ppm, the XRF error for a lead measurement with a 120-sec acquisition will be less than 5% (relative error). The observed error for these bags in that range (pertinent bags indicated by the red oval on the scatter plot) is about 8%, or 34 ppm. To cut this error in half, four measurements are needed. Four measurements would translate into four 30-sec acquisitions.

Aggregating XRF Measurements

Improving Representativeness

◆ Can be done either automatically by the XRF unit (if set up to do so) or manually by recording multiple measurements, downloading, and calculating averages for sets of measurements in a spreadsheet

◆ If automatically, *be aware that the XRF-reported error and DL will be incorrect for the measurement aggregate*

6.1-28

Notes

◆ **Can be done either automatically by the XRF unit or manually:** Aggregating XRF readings can be done either automatically or manually. Both Niton and Innov-X allow their instruments to be set-up such that the instrument will report the average concentration from a sequence of measurements (e.g., after four measurements, or after eight measurements). The other alternative is to collect measurements in a standard sort of way, and then download the data to a spreadsheet and manually calculate the average concentration using spreadsheet functions.

◆ **If automatically be aware:** The former is easier, but be aware that the error and DL reported by the instrument will be wrong for automatically-calculated average values. The latter is more work, but by doing the math correctly one can calculate the correct analytical error and detection limit (if the average result is low enough to qualify as a non-detect).

XRF Results Can Drive Number of Measurements Dynamically

Improving Representativeness

◆ Applicable to in situ and bagged sample readings
◆ XRF results quickly give a sense for what levels of contamination are present
◆ Number of measurements can be adjusted accordingly:
  » At background levels or very high levels, fewer
  » Maximum number when results are in range of action level
◆ Particularly effective when looking for the presence or absence of contamination above/below an action level within a sample or within a decision unit

6.1-29

Notes

We don't necessarily need to have a fixed number of measurements per bag or per in situ location. We could let the number vary depending on what we encounter with the XRF. This is our first foray into dynamic work strategies.

◆ **Applicable to in situ and bagged sample readings:** For in situ and bagged sample readings, the measurement results should influence the number of readings that are taken. For highly variable results at or near the action level, more readings should be taken.

◆ **XRF results quickly give a sense for what levels of contamination are present:** The XRF is an excellent tool for giving a quick snapshot of what levels of contamination are present within a bagged sample or at a location.

◆ **Number of measurements can be adjusted accordingly:** The number of measurements should be adjusted based on the data generated:

  » At background levels or very high levels, fewer measurements are needed

  » When results are in the range of the action level, the maximum number of measurements should be taken

◆ **Particularly effective when looking for the presence or absence of contamination above/below an action level within a sample or within a decision unit:** The XRF is very effective when looking for the presence or absence of contamination above or below an action level within a sample or at a particular location. It has been used effectively to identify areas of a decision unit that require action and areas of a decision unit that do not require action.

Example

Improving Representativeness

◆ Bagged samples, measurements through bag

◆ Need decision rule for measurement numbers for each bag

◆ Action level: 25 ppm

◆ 3 bagged samples measured systematically across bag 10 times each

◆ Average concentrations: 19, 22, and 32 ppm

  » 30 measurements total

(continued)

6.1-30

Notes

◆ For this example, let's assume we are using bagged samples to try and quickly determine whether specific locations have concentrations that are of regulatory concern (e.g., perhaps we are bounding the footprint of contamination, or perhaps we are looking for "hot spots" over large areas that are otherwise expected to be at background conditions). We will be collecting samples from individual locations, bagging them, and then measuring them through the bag walls. We would like a decision rule that will expedite our bag screening process (i.e., how many measurements do we need to do on each of the bags?).

For this particular example we have 3 bagged samples that we are quite confident came from an area with concentrations around our action level. We measured each bag systematically across their front and back ten times (5 on the front, 5 on the back). We observed that the average concentration reported by the XRF for each of the bags was 19, 22, and 32 ppm, indicating that we are in fact around our action level. We now have 30 individual measurements to work with.

Example

Improving Representativeness

<u>Simple Decision Rule</u>:

• if 1st measurement less than 10 ppm, stop, no action level problems

• if 1st measurement greater than 50 ppm, stop, action level problems

• if 1st measurement between 10 and 50 ppm, take another three measurements from bagged sample

XRF Result Frequency versus concentration
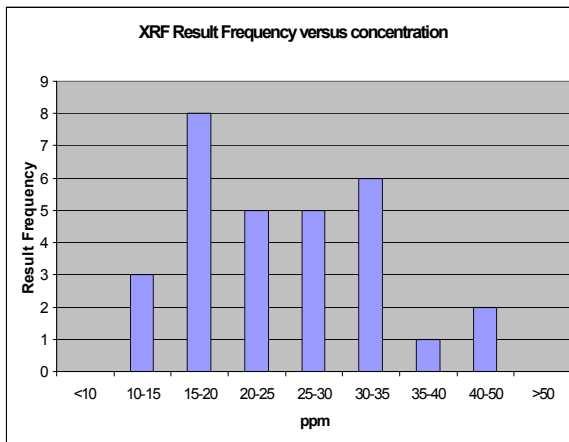
6.1-31

Notes

Using these 30 data points we can construct a simple decision rule, illustrated by the histogram shown on this slide. This histogram shows how many times particular ranges of concentrations were observed in this set of 30 measurements. We notice that for these samples, none of the individual XRF measurements were less than 10 ppm, and none were greater than 50 ppm. The decision rule that falls out:

◆ **If the 1st measurement is less than 10 ppm, stop:** That bag is unlikely to contain an average concentration at a level that would be of concern.

◆ **If the 1st measurement is greater than 50 ppm, stop:** The bag is very likely to contain an average concentration at levels that would be concern.

◆ **If the 1st measurement is between 10 and 50 ppm:** Collect another 3 measurements to better determine exactly what is in the bag.

MI Warning!!

Improving Representativeness

For sampling programs that use multi-increment (MI) sampling, one would expect MI sampling to significantly increase within sample heterogeneity. This would exacerbate the effects of poor sample preparation on either XRF cup analyses or off-site laboratory analyses (e.g., ICP).

6.1-32

Notes

◆ There is a warning here for MIS users. Within-sample heterogeneity increases as concentrations increase. This effect is exacerbated by MIS where background increments may be combined with highly impacted increments. If careful attention is not paid to sample preparation prior to XRF cup analyses or off-site laboratory analyses, the value of MIS will be lost. The flip side, though, is that if metals are the issue, an XRF can be used to determine whether the sample preparation on an MI sample was adequate through bagged sample analyses.

Collaborative Data Sets Address Analytical and Sampling Uncertainties

Increasing Information

Cheaper/rapid (lab? field? std? non-std?) **analytical methods**	**Costlier/rigorous** (lab? field? std? non-std?) **analytical methods**

Targeted high density sampling	**Low DL** + **analyte specificity**

Manages CSM & sampling uncertainty

Manages analytical uncertainty

Collaborative Data Sets

6.1-33
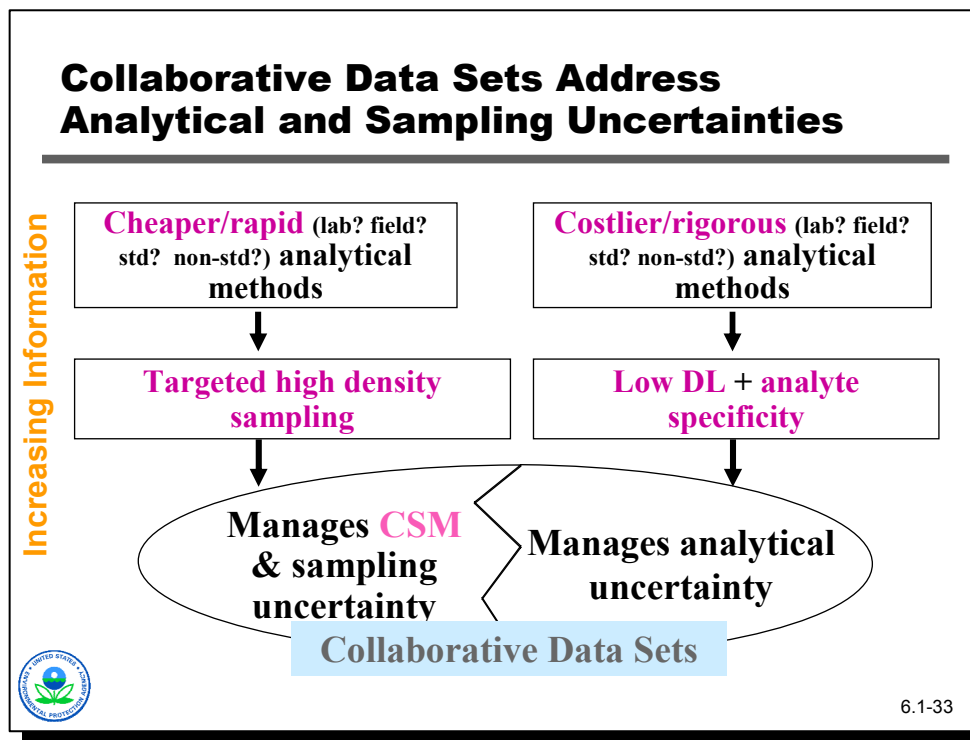
Notes

◆ This graphic shows that the cheaper/rapid analytical methods generate targeted high density sampling which manages CSM and sampling uncertainty while the costlier/rigorous analytical methods can achieve low detections limits for specific analytes which manages analytical uncertainty. The two data sets used collaboratively together can address most uncertainty associated with site characterization. In our XRF world, the XRF is our cheaper, rapid method while ICP is the costlier and more rigorous method. The primary point here is that each data source typically serves a different function. Off-site laboratory analyses should never be completely eliminated when using an XRF.

Collaborative Data Sets: Replacing Lab Data with XRF

Increasing Information

◆ Goal: replace more expensive traditional analytical results with cheaper field-analytics.

◆ Same budget allows a lot more XRF data points, improving average concentration estimates

◆ Assumptions:

» Cheaper method unbiased (or can be corrected)

» Linear relationship exists w/ high correlation (SW-846 Method 6200 points to correlation coefficients >0.9 as producing "lab equivalent" data)

» Expensive traditional analyses used for QC purposes

» Applicable to static or dynamic work plans

◆ Requirements: Method applicability study (DMA) to establish relationship between cheaper & more expensive method may be necessary. Perform on-going QC to verify relationship holds.

6.1-34

Notes

◆ **Goal:** This is the best case scenario for collaborative data. XRF analytical data can be used as a replacement for more expensive traditional laboratory data. SW-846 points to correlation coefficients > 0.9 as potentially indicating "laboratory equivalent" data.

◆ **Assumptions:** The assumptions are that the XRF exists that produces unbiased or correctable data and that a strong linear relationship exists between the cheaper and more expensive technique over the range of concentrations expected to be encountered. Regression analysis is usually used to demonstrate the existence and strength of such relationships. Laboratory analyses are not eliminated; they are reserved for a QC role.

◆ **Requirements:** In this context, more expensive data are typically used for two purposes: to establish that the relationship exists (perhaps through a demonstration of methods applicability study), and to watch for conditions or situations where the cheaper data might be suspect (e.g., interference from other contaminants or matrix effects).

The requirements for this typically are a method applicability study and a formal QA/QC process that watches for indications that the relationship is no longer valid, or is invalid under certain conditions.

Collaborative Data Sets: Blending XRF and Lab Data for Mean Estimation

Increasing Information

◆ Goal: estimate population mean by blending field data with laboratory data using an algorithm such as in Visual Sampling Plan (VSP)

◆ Assumptions:

  » Two methods, XRF and off-site laboratory

  » XRF data are unbiased, or can be corrected

  » Linear correlation exists and can be quantified

  » Static sampling program

  » Every location analyzed by field method, a subset analyzed by lab

◆ Linear correlation determined from sample splits analyzed by both XRF and off site laboratory

6.1-35

Notes

◆ **Goal:** This is the second best case scenario for collaborative data sets, to estimate a population mean by blending XRF data with laboratory data using an algorithm such as found in Visual Sampling Plan (VSP).

◆ **Assumptions:** The assumption is that our lower analytical quality data are unbiased (or if there is bias present, it can be corrected), and that there is a reasonably high linear correlation between our cheaper method and more definitive techniques. If these assumptions are true, the two data sets can be blended together in a statistical fashion to support things like estimating average contamination levels. In this case, every sample would be analyzed by the cheaper, lower quality method, and a subset would also be analyzed by the more definitive method.

In implementing this type of approach, one needs to determine how many locations should be analyzed with the cheaper, less reliable technique, and how many of those should also be analyzed by the more expensive approach.

The use of XRF for certain metals (e.g., lead) is probably the best example of a setting where this type of approach would be appropriate.

◆ **Linear correlation determined from sample splits analyzed by both XRF and off site laboratory:** The linear correlation is determined from sample splits analyzed by both XRF and a laboratory.

These Two Approaches Are Not Always Applicable

Increasing Information

◆ Issues with both previous approaches

» Assume that traditional lab data are "definitive"

» Assume that the linear relationship holds over the whole range of data encountered

» Assume an "excellent" correlation

» Assume the underlying contaminant distribution is normally distributed (in the 2nd approach)

◆ These assumptions frequently do not hold in actual site projects.

6.1-36

Notes

◆ **Issues with both previous approaches:** Both approaches assume that traditional laboratory data are "definitive," which is not always the case. Both approaches assume that the linear relationship between field and laboratory data holds over the whole range of data encountered, which is not always the case. The second approach assumes the underlying contaminant distribution is normally distributed, which is not always the case.

◆ **These assumptions frequently do not hold in actual site projects:** Data from existing investigations shows that these assumptions may not apply to many projects. There are situations where one would like to use something like the XRF, but comparability with laboratory results is not terrific. What to do?

Often Linear Regression Analyses Are Not Possible with Collaborative Data

Increasing Information

◆ Outlier problems

◆ Non-linear relationships

◆ Non-detects

◆ Result: data sets cannot be substituted or merged quantitatively

Th230 vs Total U

$R^2 = 0.33$

Th230 (pCi/g) — vertical axis: 0, 5, 10, 15, 20, 25, 30

XRF Uranium (ppm) — horizontal axis: 0, 100, 200, 300, 400, 500
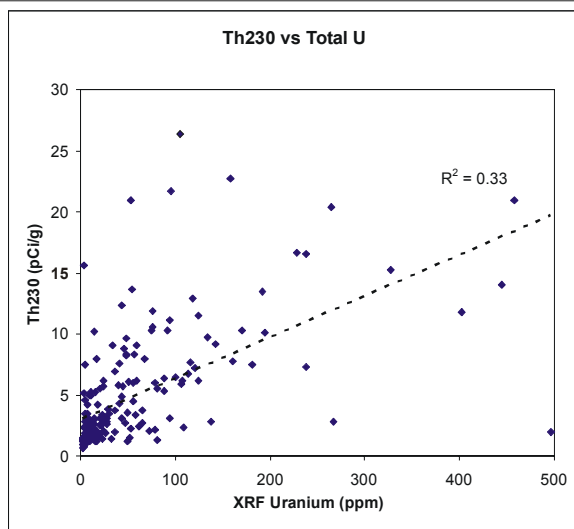
6.1-37

Notes

◆ Unfortunately, often times two different analytical techniques do not lend themselves to a simple linear regression, and so they cannot be combined directly in a quantitative way. Examples that can cause this include measuring two different parameters (e.g., PCB test kits versus GC results), outlier problems in data sets, non-linear relationships between two methods, issues with non-detects for one of the methods, etc.

The graphic shows an example from an XRF application. At this particular site, Th-230 was the risk driver. However, Th-230 doesn't lend itself to any convenient field analytical technique. At this site, however, the Th-230 was generally collocated with uranium, and uranium is measurable by XRF. This is a scatter plot of samples analyzed for Th-230 via alpha spectroscopy (a more definitive laboratory method) versus uranium results obtained by XRF, something that could be done in the field. The resulting linear regression and associated R2 value are not good.

However, if a relationship (from a decision-making perspective) can be established between the results obtained from XRF data, and those from more definitive analyses, then cheaper data can be used to directly support decision-making.

Non-Parametric Analysis Can Be a Useful Alternative

◆ Decision focus is yes/no

  » Is contamination present at levels of concern?

  » Should a sample be sent off-site for more definitive analysis?

◆ Goal is to identify investigation levels for real-time method that will guide decision making

  » Lower investigation level (LIL) for real-time result below which we are confident contamination is not present

  » Upper investigation level (UIL) above which we are confident contamination is present

Increasing Information

"clean"   "unclear"   "contaminated"

Lower Investigation Level (LIL)   Upper Investigation Level (UIL)

"Real-time" analytical result
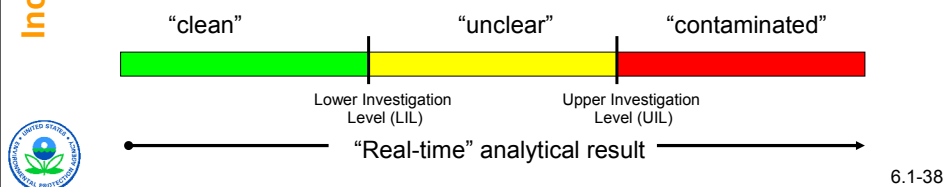
6.1-38

Notes

◆ A good example is the use of non-parametric statistical techniques that focus on the decision that needs to be made. Often times the decision is binary (e.g., is the contamination above or below requirements?). The idea is to determine investigation levels for the cheaper technique that are directly connected to decision requirements…e.g., if the result is below this investigation level, then it can be concluded that there is nothing of concern, but if it is above that investigation level, then there is certainly a problem. If there is a result between the two, then the cheaper technique (e.g., XRF) is not providing enough information to support the decision. In this case, the role of more definitive analyses is limited to establishing the investigation levels and clarifying results from the cheaper techniques.

The graphic illustrates the concept of a Lower Investigation Level (LIL) and Upper Investigation Level (UIL).

This is our second foray into dynamic work strategies. Note that this approach lends itself to dynamic work plans…cheaper, "real-time" results such as the XRF can be used both to drive sample location selection and determine whether more definitive sample analyses are required for specific areas. Up front we don't know (although we might guess) how many samples will be required for off-site laboratory analysis, because we don't know how many real-time results will yield a result that falls into the "unclear" category.

Selection of LIL and UIL Driven by Acceptable Error Rates...

Increasing Information

◆ Fraction of "contaminated" locations missed using a real-time investigation level: false clean error rate

◆ Fraction of "clean" locations identified as contaminated by a real-time investigation level: false contaminated error rate

◆ The lower the LIL, the lower the false clean error rate

◆ The higher the UIL, the lower the false contaminated error rate

6.1-39

Notes

◆ **Fraction of "contaminated" locations missed using a real-time investigation level: false clean error rate:** The false clean error rate is usually set quite low. This is the type of error EPA tries to avoid because it has potential human health impacts. The false clean error rate is the fraction of contaminated locations that might be missed using a real-time technique.

◆ **Fraction of "clean" locations identified as contaminated by a real-time investigation level: false contaminated error rate:** This error may cause remediation of areas that are not actually contaminated, which is costly if the area is large. This is the fraction of clean locations that a real-time technique mistakenly identifies as contaminated.

◆ **The lower the LIL, the lower the false clean error rate:** To achieve a low false clean error rate, the LIL should be set at a low level. The lower the LIL, the lower the false clean error rate.

◆ **The higher the UIL, the lower the false contaminated error rate:** To achieve a low false contaminated error rate, the UIL should be established at a high level. The higher the UIL, the lower the false contaminated error rate.

...and Costs

Increasing Information

◆ The greater the separation between the LIL and UIL, the greater the number of samples that may require confirmatory analysis

◆ The break-even cost analysis for collaborative data collection:

$$C_{rt}/C_f \;<\; (N_{rt} - N_f)/N_{rt}$$

where

» C_{rt} = cost of real-time,

» C_f = cost of lab analysis,

» N_{rt} is the # of real-time analyses, and

» N_f is the expected number of confirmatory lab analyses

6.1-40

Notes

◆ **The greater the separation between the LIL and UIL, the greater the number of samples that may require confirmatory analysis:** A larger number of samples may need confirmatory analyses in a laboratory if there is a large difference between the LIL and the UIL for the real-time measurements and a pre-ponderance of real-time results fall into the range of concentrations between the LIL and the UIL.

◆ **The break-even cost analysis for collaborative data collection:** The break even costs for collaborative data collection can be calculated using the equation above.

Hypothetical Example
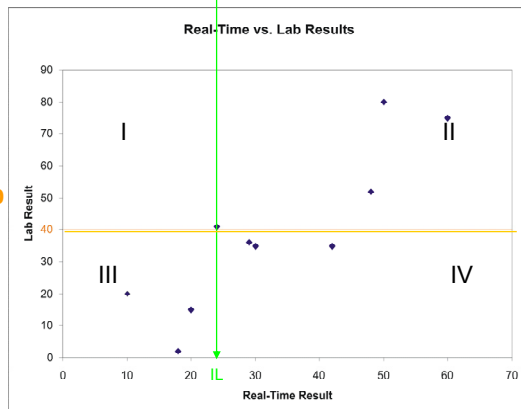
Real-Time vs. Lab Results

- I: False Clean
- II: Correctly Identified Contaminated
- III: Correctly Identified Clean
- IV: False Contaminated

- I/(I+II)*100: % of contaminated samples missed by LIL (false clean rate)
- I/(I+III)*100: % of "clean" samples that are contaminated
- IV/(II+IV)*100: % of "contaminated" samples that are clean
- IV/(III+IV)*100: % of clean samples above the LIL (false contaminated rate)

False Clean Rate: 0%   False Contaminated Rate: 50%   (continued)
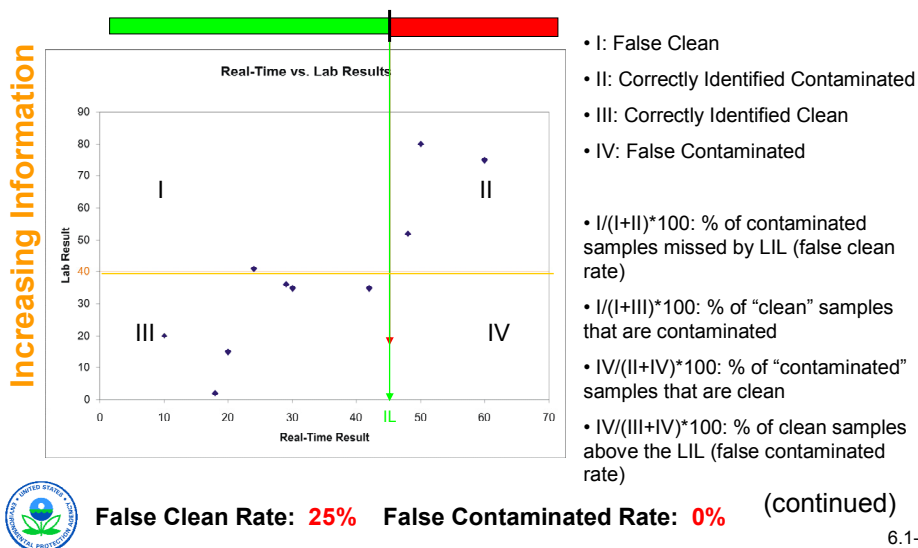
6.1-41

Notes

◆ This hypothetical example illustrates how to estimate the appropriate LIL or UIL given a set of paired data (i.e., samples that were analyzed both by XRF and off-site laboratory). The paired data could have been the product of a DMA, or they could have been generated during the course of historical characterization work at a site.

In this particular example, there are 10 samples plotted on the scatter plot. The x-axis is their real-time result (e.g., XRF). The y-axis is the corresponding laboratory result. The action level for this example is 40 ppm, denoted by the orange horizontal line that passes through the y-axis. Rather than trying to fit a regression line, we notice that if we identify one investigation level for the real-time technique and use that investigation level to classify samples as either "clean" or "dirty", that investigation level, combined with the action level, divide our graph into four regions: I, II, III, and IV. Sample points that fall into region I are "false clean" points…they are samples that the real-time technique would have labeled clean but the laboratory contaminated. Sample points that fall into region II are correctly identified by the real-time technique as contaminated. The number of points falling in region I divided by the sum of the points in region I and II is the false clean rate…the fraction of contaminated points that the real-time technique misses using that particular investigation level.

In a similar fashion, regions III and IV define the number of samples identified as correctly clean and as "false contaminated", respectively. Dividing the number of samples in region III by the sum of the sample numbers in regions III and IV gives our false contaminated rate.

Notice for that for this investigation level (~23 ppm), the result is a great false clean rate, but a terrible false contaminated rate.
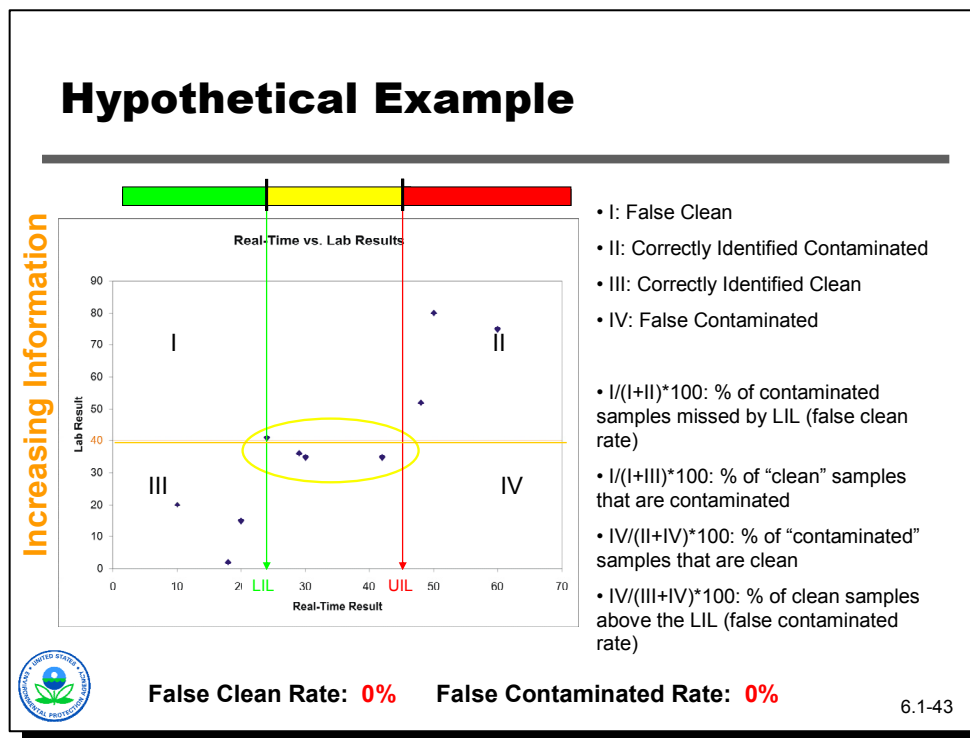
Hypothetical Example

Increasing Information

Real-Time vs. Lab Results

(y-axis: Lab Result, values 0 to 90; x-axis: Real-Time Result, values 0 to 70)

Quadrants: I, II, III, IV

- I: False Clean
- II: Correctly Identified Contaminated
- III: Correctly Identified Clean
- IV: False Contaminated

- I/(I+II)*100: % of contaminated samples missed by LIL (false clean rate)
- I/(I+III)*100: % of "clean" samples that are contaminated
- IV/(II+IV)*100: % of "contaminated" samples that are clean
- IV/(III+IV)*100: % of clean samples above the LIL (false contaminated rate)

False Clean Rate: 25%   False Contaminated Rate: 0%

(continued)

6.1-42

Notes

◆ If we are not happy with this false contaminated rate we can reduce it by raising the investigation level…in this case we increase it to 45 ppm. This has the desired effect on our false contaminated rate…it now drops to zero. But our false clean rate now jumps to 25%...a rate that will probably be unacceptable. This illustrates the problem of using one investigation level for real-time techniques. With one investigation level one can control the false clean error rate, or the false contaminated error rate, but not both at the same time. Decreasing the false contaminated error rate by raising the investigation level will raise the false clean rate, and vice versa.

Hypothetical Example

Real-Time vs. Lab Results

- I: False Clean
- II: Correctly Identified Contaminated
- III: Correctly Identified Clean
- IV: False Contaminated

- I/(I+II)*100: % of contaminated samples missed by LIL (false clean rate)
- I/(I+III)*100: % of "clean" samples that are contaminated
- IV/(II+IV)*100: % of "contaminated" samples that are clean
- IV/(III+IV)*100: % of clean samples above the LIL (false contaminated rate)

False Clean Rate: 0%   False Contaminated Rate: 0%

6.1-43

Notes

◆ The solution is to use two investigation levels, a lower investigation level and an upper investigation level. The lower investigation level guarantees we get the false clean rate we want, while the upper investigation level guarantees we get the false contaminated rate we desire.

The price we pay is that a certain set of our points (in this case 4, or 40% of our points) fall between the two investigation levels. These would be candidates for off-site laboratory analysis to determine exactly what is going on.

Notice that with this rate of off-site laboratory use (i.e., 40% of our samples sent off-site after real-time analysis), our real-time technique would have to cost less than 60% of our off-site laboratory for this to make sense from a cost perspective. If not, we would be better off sending all the samples to the laboratory and foregoing the real-time technique. Note, however, that even if the cost numbers are in favor of the real-time technique, there may be still other reasons for deploying it (e.g., rapidly identifying hot spots that can be delineated).

Next Session

◆ Module 6.2

  » Addressing the Unknown

6.1-44

Q&A – If Time Allows

6.1-45