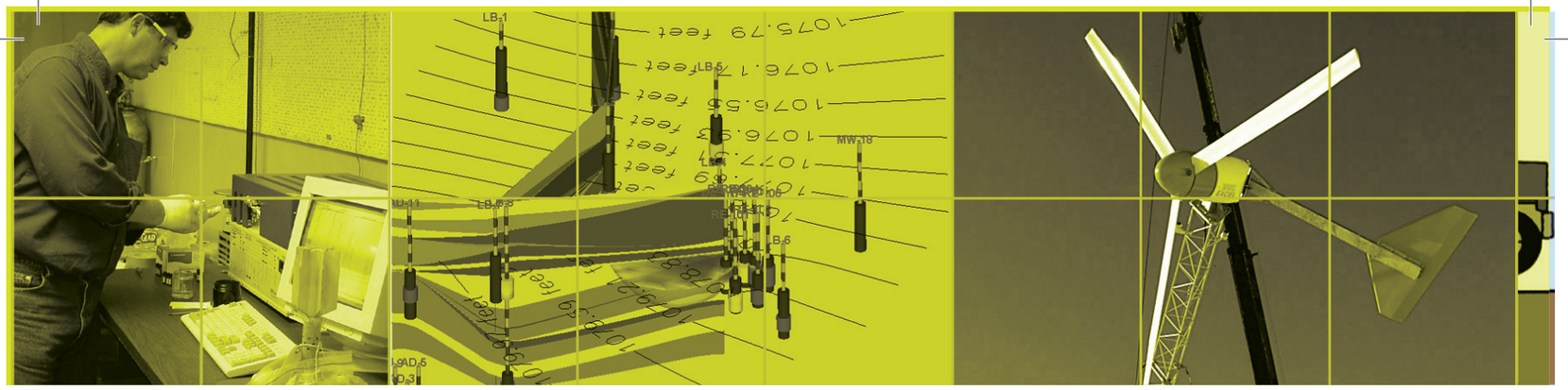
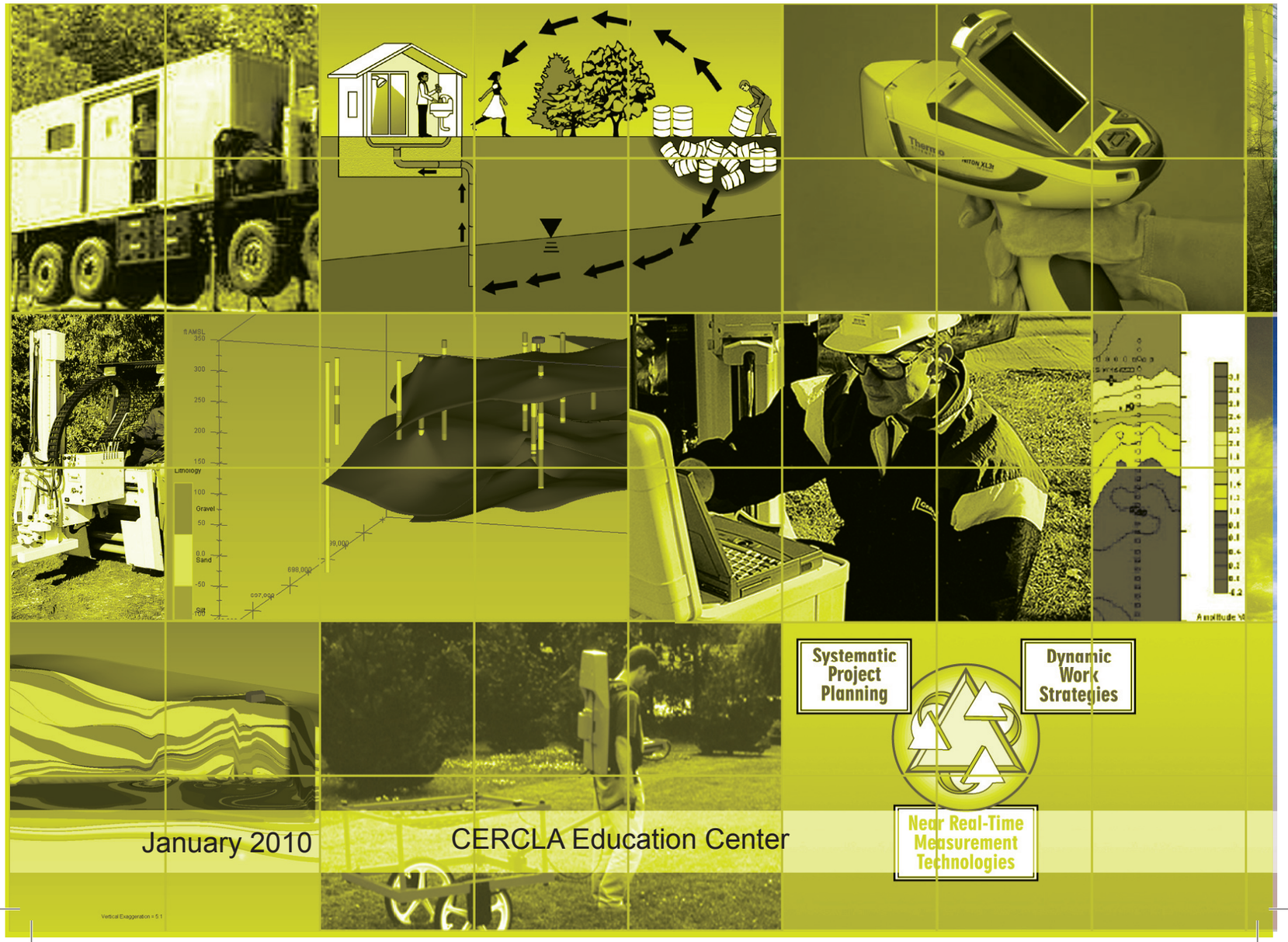


Advanced Design Application and Data Analysis for FP-XRF in Soil Matrices

2010 North American Environmental Field Conference and Exposition



January 2010

CERCLA Education Center

CLU-IN

EPA's Hazardous Waste Clean-Up Information (CLU-IN) Web site provides information about innovative treatment and site characterization technologies. It describes programs, organizations, publications, and other tools for federal and state personnel, consulting engineers, technology developers and vendors, remediation contractors, individual citizens, and all other waste remediation stakeholders. http://clu-in.org

Trainex Exchange

The Training Exchange Web site provides a wide range of training information to EPA, other federal agency, state, tribal, and local staff involved in hazardous waste management and remediation. The site provides training schedules for classroom and internet-based courses. Through Trainex, EPA works in partnership with other organizations and agencies to offer training relevant to hazardous waste remediation, site characterization, risk assessment, emergency response, site/incident management, counter-terrorism, and the community's role in site management and cleanup. www.trainex.org

Brownfields and Land Revitalization Technology Support Center

Coordinated through EPA's Technology Innovation Program, the Brownfields and Land Revitalization Technology Support Center ensures that Brownfields decision makers are aware of the full range of technologies available to make informed or "smart" technology decisions for their sites. The Brownfields Center provides a readily accessible resource for unbiased assessments and supporting information on options relevant to specific sites, technology-oriented reviews for investigation and clean-up plans, and information about other available support activities, such as those conducted by the Technical Assistance to Brownfields (TAB) Program located at the five regional Hazardous Substance Research Centers. Direct support is available to EPA regional staff, state staff, and local governments. www.brownfieldstsc.org

Technology Innovation Program Home Page on EPA's Web Site

The Technology Innovation Program's Web site provides information about characterization and treatment technologies for the hazardous waste remediation community. It offers technology selection tools and describes programs, organizations, and publications resources for federal and state personnel, consulting engineers, technology developers and vendors, remediation contractors, researchers, community groups, and individual citizens. Their goal is to create an information support network for technology decision makers who address contamination of soil or groundwater. www.epa.gov/tio

Triad Resource Center

This is the official web site for the Triad approach. EPA and a multiagency partnership developed the site to provide one-stop-shopping for Triad information, case studies, training opportunities, and news. www.triadcentral.org

Interstate Technology & Regulatory Council

The Interstate Technology and Regulatory Council (ITRC) is a state-led coalition working together with federal partners, industry, academia, and stakeholders to achieve regulatory acceptance of environmental technologies. In conjunction with EPA's Technology Innovation and Field Services Division, ITRC delivers training through the Internet to reach a geographically dispersed audience of regulators, consultants, and other members of the environmental community. www.itrcweb.org

Environmental Response Training Program

As part of EPA's comprehensive program for protecting the public and the environment from hazardous materials, EPA's Environmental Response Training Program's (ERTP) courses are designed for personnel who respond to spill events and investigate and clean up abandoned hazardous waste sites. Training is provided in health and safety and various technical operations needed to identify, evaluate, and control hazardous substances that have been released. ERTP quickly develops technically-focused courses to address emerging issues faced by responders (such as Anthrax, biohazards, and air monitoring for weapons of mass destruction). For information, visit www.trainex.org and select the ERTP.

U.S. EPA Technical Support Project

Provides technical assistance to Regional Remedial Project Managers, Corrective Action Staff, and On-Scene Coordinators. The Project consists of a network of Regional Forums and specialized Technical Support Centers located in ORD and the Office of Radiation Programs (ORP) laboratories, and OSWER's Environmental Response Team. www.epa.gov/tio/tsp

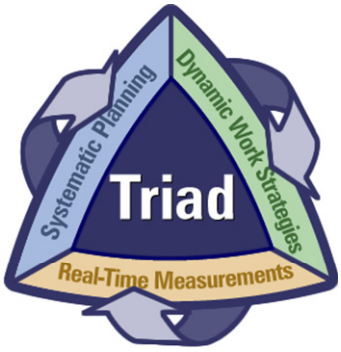
Federal Remediation Technologies Roundtable

The Federal Remediation Technologies Roundtable (FRTR) works to build a collaborative atmosphere among federal agencies involved in hazardous waste site cleanup. FRTR was established in 1990 to bring together top federal cleanup program managers and other remediation community representatives to share information and plan for the future. FRTR members-agencies include U.S. Department of Defense, U.S. Environmental Protection Agency, U.S. Department of Energy, U.S. Department of the Interior, and National Aeronautics and Space Administration. www.frtr.gov

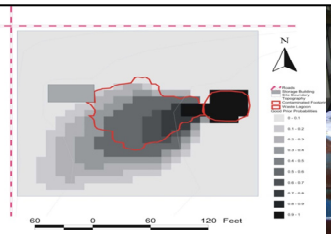
Green Remediation

Green Remediation Web helps stakeholders incorporate sustainable practices into hazardous site cleanup. The site provides access to information on technical issues such as treatment system optimization, renewable energy resources, and site management techniques. The site is part of EPA's efort to increase sustainability of site cleanup by sharing site-specific case studies, best management practices, and partnerships and opportunities such as government incentives and pilot demonstrations. http://clu-in.org/greenremediation

Advanced Design Application & Data Analysis for Field-Portable X-Ray Fluorescence



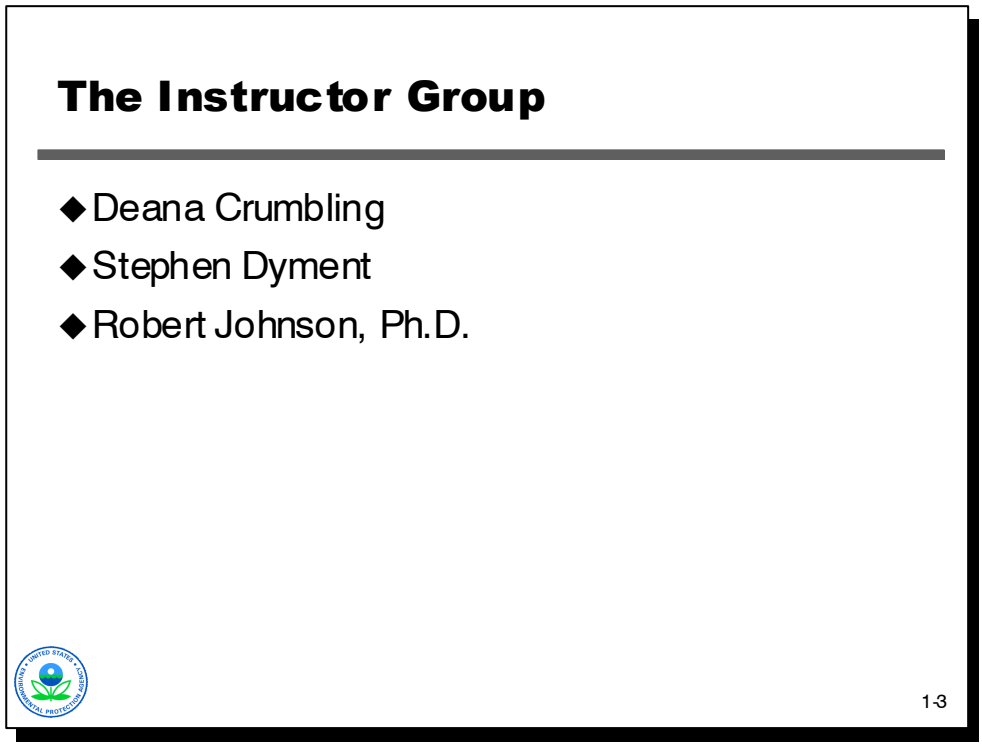
1-1



Module 1:

Introduction

1-2



Notes

The following professionals and subject matter experts constitute the instructor group for this course. Any three of the five individuals instruct each delivery of the course. All five individuals are excellent resources for questions about the use of the x-ray fluorescence (XRF) and the Triad approach to site characterization and remedial action sampling and monitoring.

- **Deana Crumbling**, U.S. Environmental Protection Agency (EPA) Office of Solid Waste and Emergency Response (OSWER); Office of Superfund Remediation and Technology Innovation (OSRTI)

  Deana Crumbling is a chemist with over 25 years experience, 12 of which have been in the waste cleanup industry. She worked for a state Superfund program and a consulting firm before joining EPA's Technology Innovation Program in 1997. Ms. Crumbling specializes in issues related to field analytical methods, sampling issues, statistics, and the Triad approach.

  Phone: (703) 603-0643
  Fax: (703) 603-9135
  E-mail: *crumbling.deana@epa.gov*

- **Stephen Dyment**, U.S. EPA OSWER, OSRTI

  Stephen Dyment is a chemist with more than 15 years experience including 4 years in a commercial analytical laboratory and 8 years in environmental consulting. He joined EPA in 2005 with a focus towards enhancing acceptance and use of emerging analytical technologies and sampling strategies. His perspective draws upon years of practical laboratory and field experience to

apply EPA's Triad approach at sites in Superfund, Brownfields, Resource Conservation and Recovery Act (RCRA), underground storage tanks (UST), and state programs. Mr. Dyment's efforts have resulted in the development of numerous EPA case studies, profiles, and training courses that outline successful strategies for the use and understanding of collaborative data sets, adaptive quality control (QC) programs, and real time analytics. He holds a B.S. in Environmental Science and Toxicology from the University of Massachusetts at Amherst.

Phone: (703) 603-9903
Fax: (703) 603-9135
E-mail: *dyment.stephen@epa.gov*

- **Robert Johnson**, Ph.D., Argonne National Laboratory, Environmental Assessment Division

  Dr. Johnson has worked at Argonne National Laboratory since 1991 as an environmental engineer, statistician, software developer, and project manager. He serves as principal investigator and technical lead for Argonne's adaptive sampling and analysis program (ASAP), and provides expert technical support to EPA on a number of Triad-related training courses and projects. Bob has significant experience in statistical sampling design and data analysis and interpretation using XRF through his U.S. Department of Energy (DOE), EPA, U.S. Department of Defense (DoD), and other technical support functions.

  Phone: (630) 252-7004
  Fax: (630) 252-3611
  E-mail: *rlj@anl.gov*

Who Will Benefit from this Course?

- Regulatory project managers and quality assurance reviewers who use XRF data
- Consultants and regulatory staff responsible for
  - » Designing and approving work plans that use XRF
  - » Interpreting XRF data

1-4

Notes

- **Regulatory project managers and quality assurance reviewers who use XRF data:** This course will benefit project managers from federal, state, and local regulatory programs and the quality assurance staff that review data collection plans and XRF data. The course will provide project managers and quality assurance (QA) staff a better understanding of how XRF data should be collected and how it can be used to reduce site characterization uncertainty.

- **Consultants and regulatory staff responsible for:** Consultants and regulatory staff who are responsible for designing and approving work plans that use XRF and who interpret XRF data also will benefit from this course. After taking this course, the participants should be able to design and implement XRF sampling strategies that provide more reliable and robust data sets. Participants also will be able to better interpret the XRF data generated from the improved sampling strategies.

Take Away Points...

- Spatial heterogeneity is a primary source of data uncertainty
- Traditional data strategies often are not cost-effective for addressing this data uncertainty
- More effective, efficient data designs involve
  - » Dynamic and adaptive field decision-making
  - » real-time data generation and management tools
- XRF is one of these tools
- Use of appropriate sampling designs, QA/QC, and collaborative data allow higher certainty and defensible decisions with XRF

1-5

Notes

- **Spatial heterogeneity is a primary source of data uncertainty:** Spatial heterogeneity, or the random distribution of contaminants in environmental media, is the primary contributor to data uncertainty. Data uncertainty associated with spatial heterogeneity is far greater than uncertainty associated with analytical errors. Reducing spatial heterogeneity is the way to significantly reduce data uncertainty.

- **Traditional data strategies often are not cost-effective for addressing this data uncertainty:** Traditional data strategies involve the collection of discrete samples that are analyzed in a fixed laboratory. Traditional strategies can only address this uncertainty by collecting a very large number of samples for laboratory analysis, which in most cases is cost prohibitive.

- **More effective, efficient data designs involve:** Data designs that are dynamic and can be adapted in the field combined with real-time data generation and management tools can overcome spatial heterogeneity effectively and efficiently.

- **XRF is one of these tools:** For many sites, the XRF can be used in a dynamic sampling strategy to generate real-time data and to more fully characterize contaminant levels and locations.

Day's Topics

- Welcome and Introduction
- XRF basics
- That mysterious, but critical data quality element: ***representativeness***
- XRF demonstration
- XRF and quality control
- XRF sampling designs
- Wrap up and resources

1-6

Instrument and Software Disclaimer

Referring to specific XRF instruments or software packages is for information purposes only and does NOT constitute endorsement.

- Manufacturers Niton and Innov-X
- Excel (Microsoft Office)
- Visual Sampling Plan (Pacific Northwest Lab: http://dqo.pnl.gov/)
- BAASS (Argonne National Lab: www.ead.anl.gov/project/dsp_topicdetail.cfm?topicid=23)
- Surfer/Grapher (Golden Software: www.goldensoftware.com)
- ArcView 3.x or 9.x (ESRI: www.esri.com)
- Freeware can be found at http://www.frtr.gov/decisionsupport/

1-7

Module 2

Basic XRF Concepts

Module 2:

Basic XRF Concepts

2-1

What Does An XRF Measure?

- X-ray source irradiates sample
- Elements emit characteristic x-rays in response
- Characteristic x-rays detected
- Spectrum produced (frequency and energy level of detect x-rays)
- Concentration present estimated based on sample assumptions

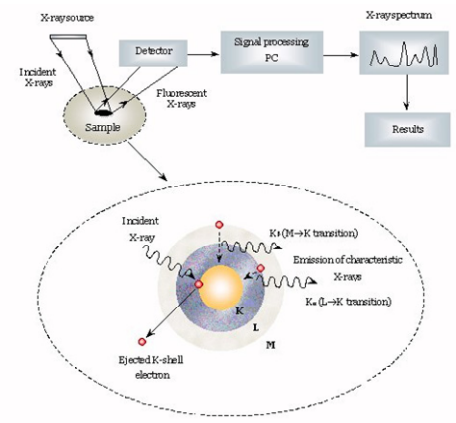


Figure 1: *The principle of XRF and the typical XRF detection arrangement*

2-2

Notes

- **X-ray source irradiates sample:** Modern XRF systems include basically three components:

  - » an x-ray source
  - » a detector
  - » a signal processing unit

  The x-ray source produces x-rays that irradiate the sample of interest. Traditionally x-ray sources were sealed radionuclide sources such as Fe-55, Cd-109, Am-241, or Cm-244. Each sealed source type emitted x-rays of a particular energy level. The selection of a sealed source depended on the elements of interest, since different elements respond best to different irradiating x-ray energy levels. Sealed sources, however, presented practical challenges:

  - » some had relatively short half-lives meaning that they had to be changed on a regular basis to maintain XRF performance
  - » they often required special licenses to be used
  - » each only addressed a relative small set of inorganic contaminants of concern

  Consequently manufacturers of XRF units have been moving to electronic x-ray tubes for producing the required x-rays.

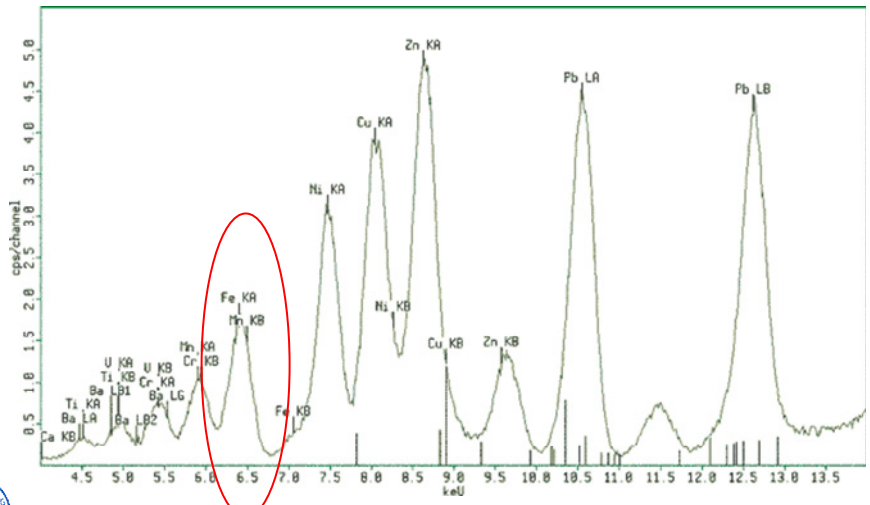
- **Elements emit characteristic x-rays in response:** When a sample is irradiated with x-rays, the x-rays interact with individual atoms, and these atoms respond by "fluorescing," or producing their own x-rays whose energy levels and abundance (number) are different for each element.

- **Characteristic x-rays detected:** The XRF detector captures these fluorescent x-rays, counting each and identifying their energy levels.

- **Spectrum produced (frequency and energy level of detect x-rays):** The signal processing unit takes the detector information and produces spectrum. Additional software processing converts the spectrum into element-specific estimates of the concentrations present based on element and sample media-specific calibrations.

- **Concentration present estimated based on sample assumptions:** Additional software processing converts the spectrum into element-specific estimates of the concentrations present based on sample assumptions.

An Example XRF Spectrum…



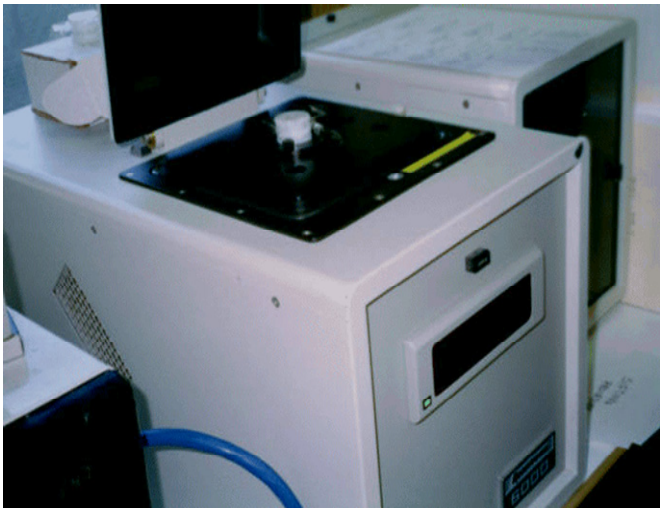
2-3

Notes

- This slide shows an example of x-ray spectrum produced by an XRF measurement. The x-axis is x-ray energy, and the y-axis shows the number of x-rays observed at each energy level. The peaks are indicative of the presence of unique elements. The heights of the peaks are proportional to the number of x-rays counted, which in turn is proportional to the mass of the element present in the sample. The width of the peaks, in general, is an indication of the detector's ability to "resolve" x-ray energies it observes, or in other words, to correctly identify the energy level of the x-ray it detected. The better the resolution, the tighter these peaks will be, the better the XRF will be in terms of performance (i.e., correctly identifying and quantifying the presence of a particular element).

  As this spectrum demonstrates, any particular element can have more than one peak associated with it, for example lead, or zinc, or iron in this spectrum. As this spectrum also demonstrates, peaks for individual elements may be so close that for all practical purposes they are indistinguishable. The Fe/Mn peak around 6.5 KeV is a good example, which is what causes what is known as interference. This topic will be discussed later in the course.

Bench-top XRF



2-4

Notes

- This slide shows a bench-top XRF unit. Samples from the field are brought to the unit which can be located in a trailer. XRF is a well-established analytical technique with a long history of use in a laboratory environment. In the last decade advances in electronics have allowed the development and refinement of field-deployable units. XRF analysis is different from most other inorganic techniques in that it is a non-destructive analysis. In other words, the original sample is not destroyed by the analytical process. There are no extraction or digestion steps. Consequently the same material can be analyzed repeatedly by an XRF unit, or analyzed by an XRF unit and then submitted for some other analysis.

How is an XRF Typically Used?

- Measurements on prepared samples

(continued) 2-5

How is an XRF Typically Used?

- Measurements through bagged samples (limited preparation)

(continued) 2-6

How is an XRF Typically Used?

- *In situ* measurements of exposed surfaces

2-7

Notes

Field deployable XRF units typically are used in one of three ways.

- **Measurements on prepared samples:** The first is a table-top system analyzing prepared samples. In this setting, samples are prepared (homogenized and dried), and then sub-sampled with sub-samples placed in the small cups shown here. Cups are covered with mylar, placed face down on the detector, and measured for a fixed period of time. The system shown is an Innov-X unit.

- **Measurements through bagged samples (limited preparation):** The second method is done by taking measurements through the walls of bagged samples. In this setting, the full mass of the soil sample is placed in a bag (smooth and thin-walled). The soil mass is kneaded to provide some homogenization, and then multiple relatively short measurements are taken systematically across each side of the bag. The results are averaged to provide an estimate of what the bag contains.

- ***In situ* measurements of exposed surfaces:** The third method is done by taking direct measurements of exposed soil surfaces. Soil surface preparation typically includes removing stones and organic material, and flattening the soil so that an even surface is presented to the detector. As with bagged samples, multiple relatively short measurements are usually taken across the soil surface, and the average used to estimate what is present. The system show is a Niton unit.

What Does an XRF Typically Report?

- Measurement date
- Measurement mode
- "Live time" for measurement acquisition
- Concentration estimates
- Analytical errors associated with estimates
- User defined fields

	A	C	D	E	F	G	H	O	P	Q	R	S	T
1	Date	Mode	LiveTime	Match1	MN1	Pass/Fail	Pass Fail Standard	Cr	Cr +/-	Mn	Mn +/-	Fe	Fe +/-
2	6-Dec-04	Standardization	53.46	0.0208	240	-0.0256	PASS						
13	6-Dec-04	Soil	76.53					<LOD	170.85	249.86	41.7	24698.13	184.68
14	6-Dec-04	Soil	80.24					196.44	60.31	263.28	43.37	25348.6	192.79
15	6-Dec-04	Soil	76					<LOD	175.72	373.11	43.1	21279.25	168.1
16	6-Dec-04	Soil	83.61					226.5	62.71	289.18	45.05	25161	197.15
17	6-Dec-04	Soil	83.07					<LOD	186.41	477.87	48.01	26620.96	205.4

2-8

Notes

- **What does an XRF typically report:** This is an example of an Innov-X output. The Mode represents the configuration of the system (i.e., what are the underlying assumptions behind the measurement…Soil? Filter? Standardization [a form of QC]. Live time refers to the fact that the detector is not actually "detecting" for the entire length of acquisition time. Different instruments handle this in different ways. For some instruments the acquisition time is set, and that is the time the instrument is actively acquiring data with a resulting "live time" that is somewhat less. For other instruments the acquisition time is set and the instrument continues acquiring data until the live time equals the required acquisition time.

  Note that the first record here ("Standardization") is an internal energy calibration check that is done periodically.

  In the case of the Innov-X, the error reported represents one standard deviation around the reported value. For the Niton, it is two standard deviations.

  In the case of the Innov-X as shown here, when a particular element's result is below the instrument-calculated detection limit, instead of the result "<LOD" is reported, and instead of the error, the detection limit is reported. The detection limit is 3 times the error associated with a result where the element of concern is at zero concentration, which can be seen here for chromium.

Which Elements Can an XRF Measure?

- Generally limited to elements with atomic number > 16
- Method 6200 lists 26 elements as potentially measurable
- XRF not effective for lithium, beryllium, sodium, magnesium, aluminum, silicon, or phosphorus
- In practice, interference effects among elements can make some elements "invisible" to the detector, or impossible to accurately quantify

2-9

Notes

- **Generally limited to elements with atomic number > 16:** The XRF is generally applicable to elements which have an atomic number greater than 16, or in other words, to elements starting with sulfur and moving up the periodic table. However, the XRF cannot necessarily measure all elements with an atomic number greater than 16 at concentrations that would be considered acceptable for environmental applications.

- **Method 6200 lists 26 elements as potentially measurable:** EPA Method 6200 for Field Portable X-Ray Fluorescence Spectrometry lists the following elements as being potentially measurable:

  - Antimony (Sb)
  - Arsenic (As)
  - Barium (Ba)
  - Cadmium (Cd)
  - Calcium (Ca)
  - Chromium (Cr)
  - Cobalt (Co)
  - Copper (Cu)
  - Iron (Fe)
  - Lead (Pb)
  - Manganese (Mn)
  - Mercury (Hg)
  - Molybdenum (Mo)
  - Nickel (Ni)
  - Potassium (K)
  - Rubidium (Rb)

- » Selenium (Se)
- » Silver (Ag)
- » Strontium (Sr)
- » Thallium (Tl)
- » Thorium (Th)
- » Tin (Sn)
- » Titanium (Ti)
- » Vanadium (V)
- » Zinc (Zn)
- » Zirconium (Zr)

- **XRF not effective for lithium, beryllium, sodium, magnesium, aluminum, silicon, or phosphorus:** Standard XRF systems are not effective for lithium, beryllium, sodium, magnesium, aluminum, silicon, or phosphorus.

- **In practice, interference effects among elements can make some elements "invisible" to the detector, or impossible to accurately quantify:** In practice, the performance of the XRF (as measured by detection limits and ability to accurately quantify an element) is highly variable from element to element. One of the factors contributing to variations in performance is the interference among elements whereby the elevated presence of one element may mask the elevated presence of another. A common example is arsenic masked by the presence of lead. Interference effects are real, element-specific, and at times significant.

How is an XRF Calibrated?

- **Fundamental Parameters Calibration** – calibration based on known detector response properties, "standardless" calibration, what is commonly done
- **Empirical Calibration** – calibration calculated using regression analysis and known standards, either site-specific media with known concentrations or prepared, spike standards
- **Compton Normalization** – calibration calculated using a combination of fundamental parameters and empirical calibration, most common for general environmental applications when concentrations are below percent range

In any case, the instrument will have a dynamic range over which a linear calibration is assumed to hold.

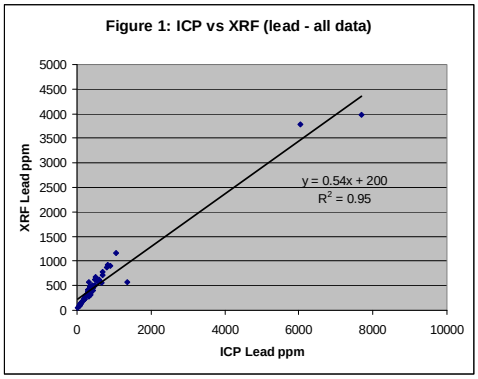
2-10

Notes

- Most, if not all, XRF vendors are willing to help users develop site-specific calibrations for their XRF applications. These can be particularly important where site-specific matrix effects are of particular concern, and/or when the element of interest is not one of the standard set used for factory standardless calibrations.

Dynamic Range a Potential Issue

- No analytical method is good over the entire range of concentrations potentially encountered with a single calibration
- XRF typically under-reports concentrations when calibration range has been exceeded
- Primarily an issue with risk assessments



2-11

Notes

- **No analytical method is good over the entire range of concentrations potentially encountered with a single calibration:** The scatter plot shows paired XRF/inductively coupled plasma (ICP) emission spectroscopy results for lead from sample splits. In this particular case the concentrations range over 3 orders of magnitude and come close to the percent range at the high end. A regression line developed from this data produces what appears to be an excellent fit as measured by the R^2, but visually inspecting the line as compared to the data suggests that linear calibration range of the instrument has been exceeded and that the instrument's response is not linear over this range of concentrations.

  Calibration ranges are issues for standard laboratories as well; however, in that instance when there is concern that a calibration range is being exceeded there is always the option of diluting the sample to bring the concentration down to an appropriate analytical level. For most deployments of an XRF, however, diluting the sample is not a practical option, so this is something to watch for.

Standard Innov-X Factory Calibration List

Antimony (Sb)	Iron (Fe)	Selenium (Se)
Arsenic (As)	Lead (Pb)	Silver (Ag)
Barium (Ba)	Manganese (Mn)	Strontium (Sr)
Cadmium (Cd)	Mercury (Hg)	Tin (Sn)
Chromium (Cr)	Molybdenum (Mo)	Titanium (Ti)
Cobalt (Co)	Nickel (Ni)	Zinc (Zn)
Copper (Cu)	Rubidium (Ru)	Zirconium (Zr)

2-12

Notes

- This slide shows the list of compounds available for the standard Innov-X factory calibrations.

How is XRF Performance Commonly Defined?

- Bias
- Precision
- Detection Limits
- Quantitation Limits
- Representativeness
- Comparability

2-13

Notes

- **How is XRF performance commonly defined:** The following factors are used to define how an XRF performs:
  - » **Bias** – does the instrument systematically under or over-estimate element concentrations?
  - » **Precision** – how much "scatter" solely attributable to analytics is present in repeated measurements of the same sample?
  - » **Detection Limits** – at what concentrations can the instrument reliably identify the presence of an element?
  - » **Quantitation Limits** – at what concentrations can the instrument reliably measure an element?
  - » **Representativeness** – how representative is the XRF result of information required to make a decision?
  - » **Comparability** – how do XRF results compare with results obtained using a standard laboratory technique?

  The following slides will discuss bias, precision, detection limits, and comparability in more detail.

What is Bias?

- Bias is a systematic difference between XRF data and what one observes in corresponding laboratory results
  - »Absolute bias reflected in a regression line y-intercept that is significantly different from zero
  - »Relative bias reflected in a regression line slope that is significantly different from one

2-14

Notes

- **Bias is a systematic difference between XRF data and what one observes in corresponding laboratory results:** Bias is a systematic disagreement between XRF results and fixed laboratory results for the same sample. Bias typically manifests itself in one of two ways (or both combined). The first is when an XRF over- or under-reports an element concentration at low concentrations. This is reflected in a comparability linear regression y-intercept that is significantly different than zero.

  The second is when an XRF systematically under- or over-reports an element's concentration by a certain fraction across a range of concentrations when compared to laboratory results. In comparability regressions this is reflected in a linear regression slope that is significantly different than one.

Where Does Bias Come From?

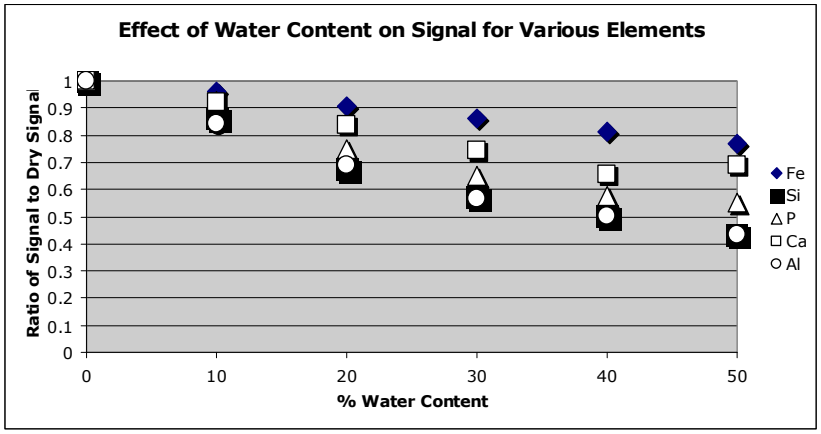
- Soil moisture effects on XRF
- Problems with XRF calibration
- Problems with XRF performance
- Concentrations outside calibration range of XRF
- Interference effects (lead impacting arsenic values)
- Matrix effects
- Extraction issues with laboratory procedure (antimony, barium)

2-15

Notes

- **Where does bias come from:** The first set of items on this slide shows bias "problems" potentially associated with the XRF. The last item is a problem that is associated with the laboratory and arises from the fact that ICP requires an extraction which usually does not involve complete digestion of the sub-sample being analyzed. Consequently for some combinations of elements and soil type the lab may actually be under-reporting the mass concentration present in the original sample. The presence of observable bias in XRF data when compared to laboratory data is not necessarily an indication of issues associated with the XRF.

Effect of Water Content For Various Elements



Slide courtesy of ThermoFisher Scientific

2-16

Notes

- This slide is courtesy of ThermoFisher Scientific, maker of the Niton line of XRF units. It shows XRF response to various metals as a function of soil sample moisture content. These data were obtained by starting with a dry sample, measuring its metal content by XRF, and then systematically re-hydrating the sample and re-measuring its metal content at each hydration step.

  As soil moisture content increases, XRF-estimated metal concentrations will decrease. The size of the effect or bias for any particular soil moisture level is element specific. The effect is approximately linear over the range of soil moistures typically encountered in the field.

Coffee Filter Dewatering Procedure from SDI

- This procedure is available at http://www.sdix.com/TechSupport/bulletins/t00078.pdf
- It was developed to dewater soil/sediment samples for immunoassay analyses.
- It can also be used to help dry samples for field XRF analysis.

2-17

DEWATERING OF MOIST SEDIMENTS AND SOILS FOR APPLICATION TO RaPID Prep™ AND RaPID ASSAYS® PROCEDURES

BACKGROUND INFORMATION

High moisture content (>30% by weight) can significantly lower concentrations of analytes detected in the RaPID Prep soil methods. In these methods, water associated with a sample will increase the volume of the extraction solution (liquid phase) and decrease the amount of soil (solid phase) in a weighed sample. In some cases, water may also decrease the extraction efficiency of the extracting solvents, lowering recovery even further. The incremental contribution of soil water to the final volume of soil extract can be corrected by adjusting the dilution factor used with the procedure. Estimation of and correction for loss of extraction efficency, however, is a more complex task. In situations where volatile chemicals are to be analyzed, air drying of moist samples is not feasible because analyte will be lost to vaporization. For these reasons, determination and correction of moisture content is not practical in many field testing situations.

The procedure described below is a quick and simple approach to reducing the moisture content of most soils and sediments to less than 30% for processing in the RaPID Prep Soil Collection and Sample Extraction kits prior to analysis by the RaPID Assays. For some analytes, it may be appropriate to also test the water phase of a moist sediment or soil sample. We recommend sampling the water phase prior to dewatering the sample for soil analysis.

MATRIX

Sediment or soil samples with greater than 30% moisture content

MATERIALS

SDI RaPID Prep Soil Collection kit and the appropriate RaPID Prep Sample Extraction kit.

Materials: Coffee filters (e.g. Mr. Coffee® #2 cone filters), paper towels and disposable plastic teaspoons, gloves, and disposable plastic cups (optional)

SAMPLE PREPARATION

Measure 4 teaspoons of moist soil or sediment sample into a coffee filter. Squeeze the sample within the coffee filter using gloved hands. A disposable plastic cup can be used to collect water from the sample as it drains. Paper towels can be wrapped around the outside of the coffee filter to assist in absorbing water. Remove as much moisture as possible from the sample.

Open up the coffee filter and remove a 10 gram sample of the treated sediment or soil. Process the soil according to the directions on the package inserts of the Soil Collection kit or the Sample Extraction kit. In most cases, no correction for moisture content is required when calculating the analyte concentration (refer to procedural notes).

PROCEDURAL NOTES

Sediments or soils high in organic matter content (>40%) may have greater than 30% moisture content after performing the above procedure. For organic samples, we recommend following the above procedure and adjusting the dilution factor for moisture content. Moisture content can be determined or estimated for site representative samples prior to analysis.

EXPECTED RESULTS

The RaPID Assay system was used to determine the PCB concentration of three samples which originally contained >40% moisture. Results were compared on split samples that were either dewatered according to the above procedure or air dried to remove moisture. Results were as follows:

% Moisture Content of Dewatered Sediment	PCB soil concentration (ppm) Dewatered Sediment	Air dried Sediment
18.4	1.3	1.5
21.4	30.0	34.3
15.4	100.0	110.6

TECHNICAL ASSISTANCE

STRATEGIC DIAGNOSTICS INC.
128 Sandy Drive
Newark DE 19713

(800) 544-8881
(302) 456-6789
FAX (302)456-6782

R060597

Analytical Precision Driven By …

- **Measurement time** – increasing measurement time reduces error
- **Element concentration present** – increasing concentrations increase error
- **Concentrations of other elements present** – as other element concentrations rise, general detection limits and errors rise as well

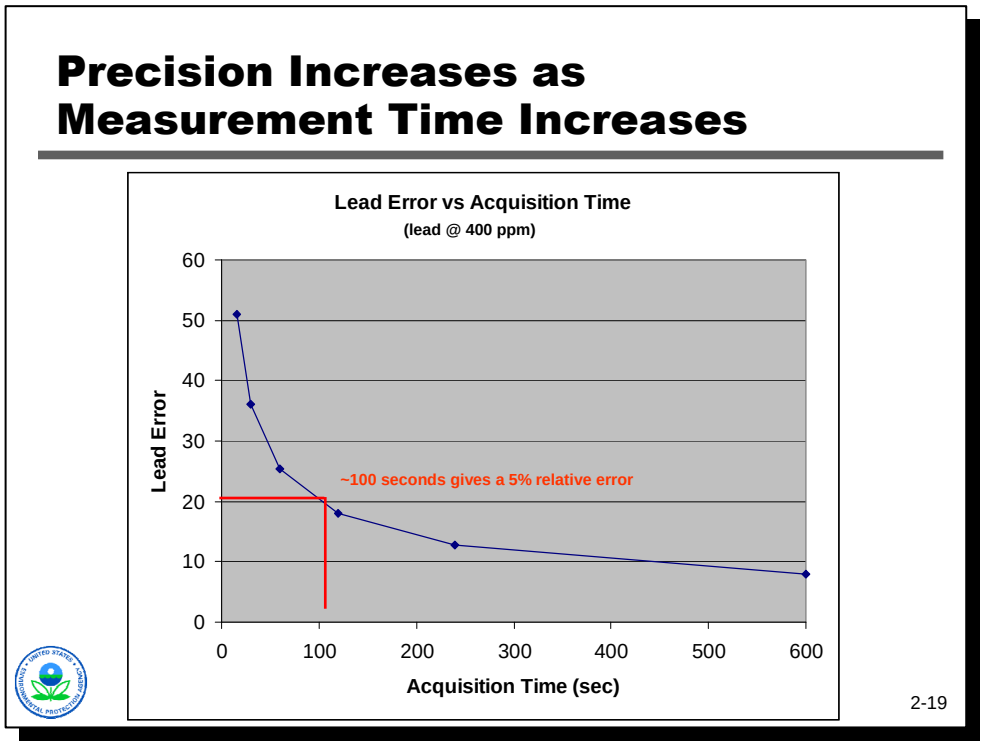
2-18

Notes

- **Measurement time:** Measurement time affects precision. Increasing the measurement time reduces error and increases precision.
- **Element concentration present:** The amount of the element of concern affects precision. Generally, increasing concentrations result in increased error and decreased precision.
- **Concentrations of other elements present:** The presence of other elements affects precision. As the concentration of other elements rise, general detection limits and errors rise, decreasing analytical precision.

  When the XRF reports a detectable concentration for a particular element, it also provides an error that is mathematically calculated based on counting statistics. The accuracy of the XRF error estimate can be checked to determine if it is accurate.

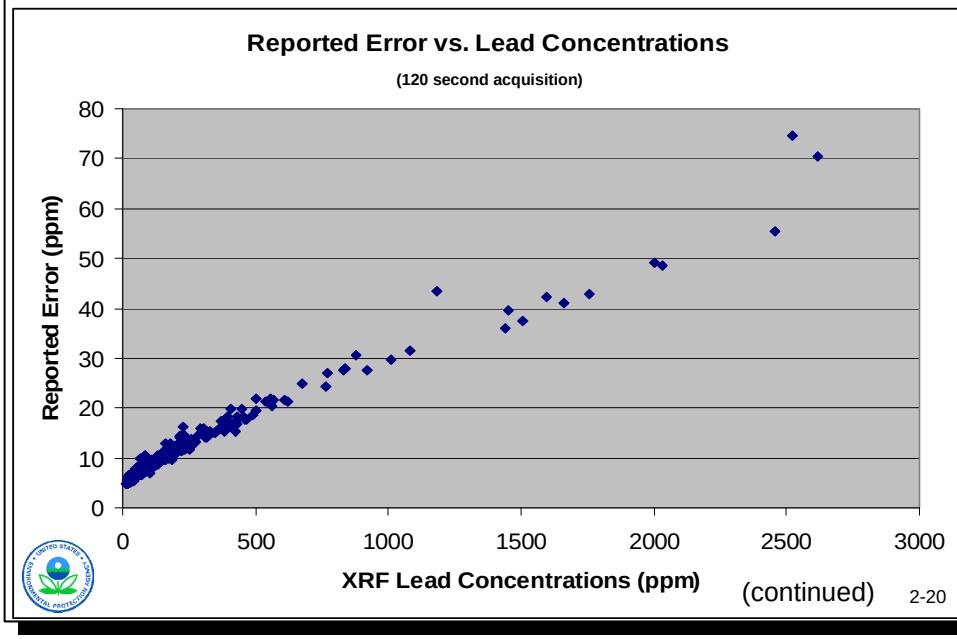
  To check the accuracy of XRF error estimates do repeated measurements of the same sample, note the average error reported, and then calculate the standard deviation of the reported results. The average reported error and the calculated standard deviation should be approximately the same.



Notes

- This graph shows an example relationship between XRF measurement error for lead at 400 parts per million (ppm) and acquisition time. As acquisition time increases, error decreases, but the relationship is not linear. Basically to cut error in half requires quadrupling acquisition time. For this particular XRF system, when lead is at its typical action level (400 ppm), a relative percent error of less than 10 percent is achieved with acquisition times of only 30 seconds.

Lead Example: Concentration Effect

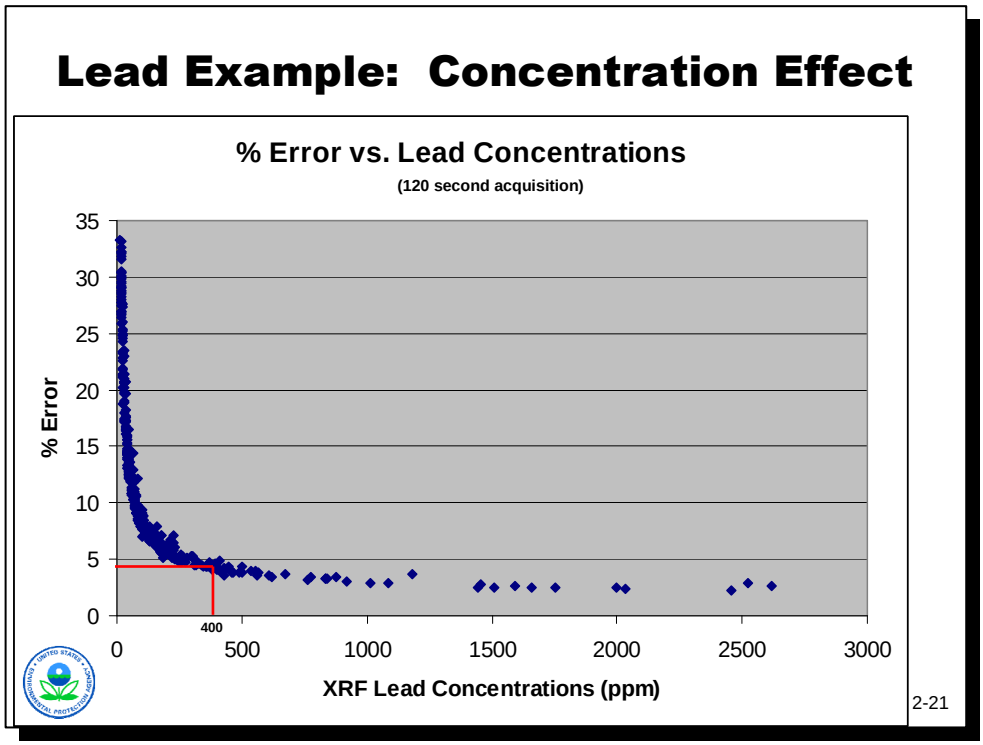


(continued) 2-20

Notes

- The next two slides show graphs that illustrate the effects of concentrations on reported measurement errors in the case of 434 lead measurements with an XRF. On this slide, the x-axis shows lead concentrations while the y-axis shows their associated reported errors. The expected relationship is indicated: error grows as the square root of concentration. In other words, to double the error it is necessary to quadruple the concentration.

  Note these relationships start to fall apart as XRF lead values become high, reflecting the contribution of other sources of error to measurement error (e.g., the presence of other elements that are very elevated).



Notes

- This graph also illustrates the effects of concentrations on reported measurement errors in the case of 434 lead measurements with an XRF. Percent error is plotted as a function of concentration. Notice that percent error is a maximum at the detection limits of the instrument, and is never more than approximately 30 percent.

  For lead values in the range of what is typically of interest (e.g., 400 ppm), percent error is less than 5 percent. This is an important fact to keep in mind. The expectation for standard laboratory analytical precision is less than 10 percent. In the case of this XRF example, the XRF meets that expectation for lead values greater than approximately 100 ppm. A general rule of thumb for any particular element is that for concentrations that are 10 times the XRF's detection limit, the analytical error of XRF measurements will be less than 10 percent.

XRF Detection Limit Calculations

- SW-846 Method 6200 defines DL as 3 X the SD attributable to the analytical variability (imprecision) at a low concentration
- XRF "measures" by counting X-ray pulses
- XRF instruments typically report DLs based on counting statistics using the 3 X SD definition
- SDs and associated DLs also can be calculated manually from repeated measurements of a sample (if concentrations are detectable to begin with)

2-22

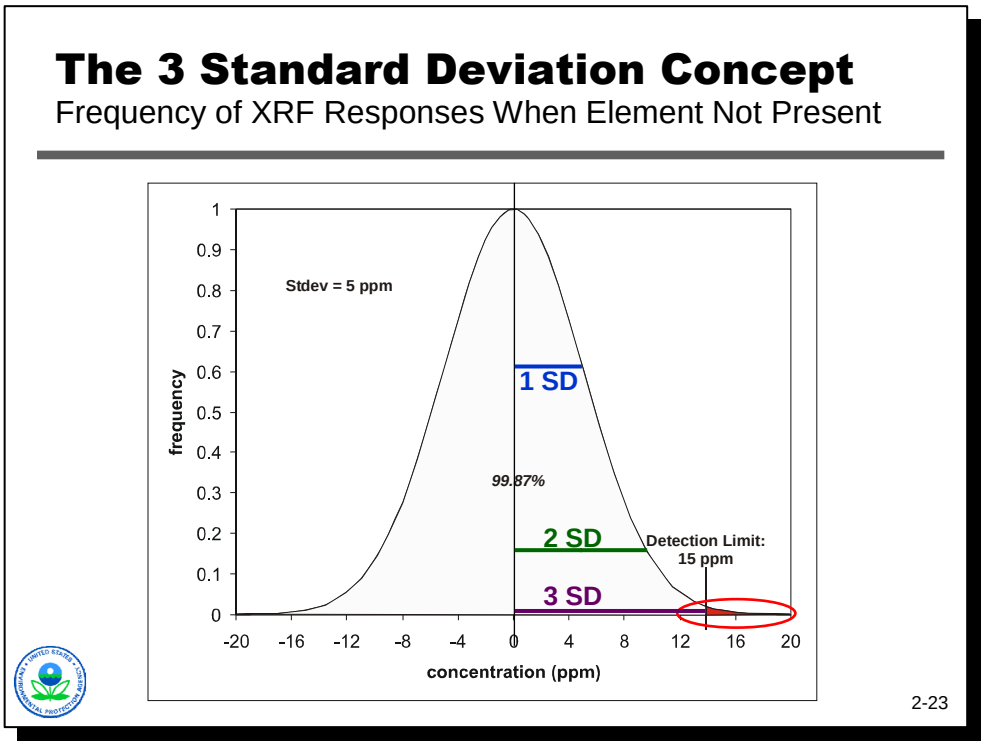
Notes

- **XRF detection limit (DL) calculations:** Remember that relative error or percent error (error divided by the concentration) falls as concentration increases. What this means is that using this definition of DLs, the percent error associated with an XRF measurement will never be more than approximately 30 percent, and usually will be significantly less.

  DLs estimated manually from repeated measurements of a sample will, in general, overestimate the actual DL of the instrument.

Internet

For information on SW-846 Method 6200 visit,
www.clu-in.org/conf/tio/xrf_082508/cd/sw-846-XRF-method-6200.pdf

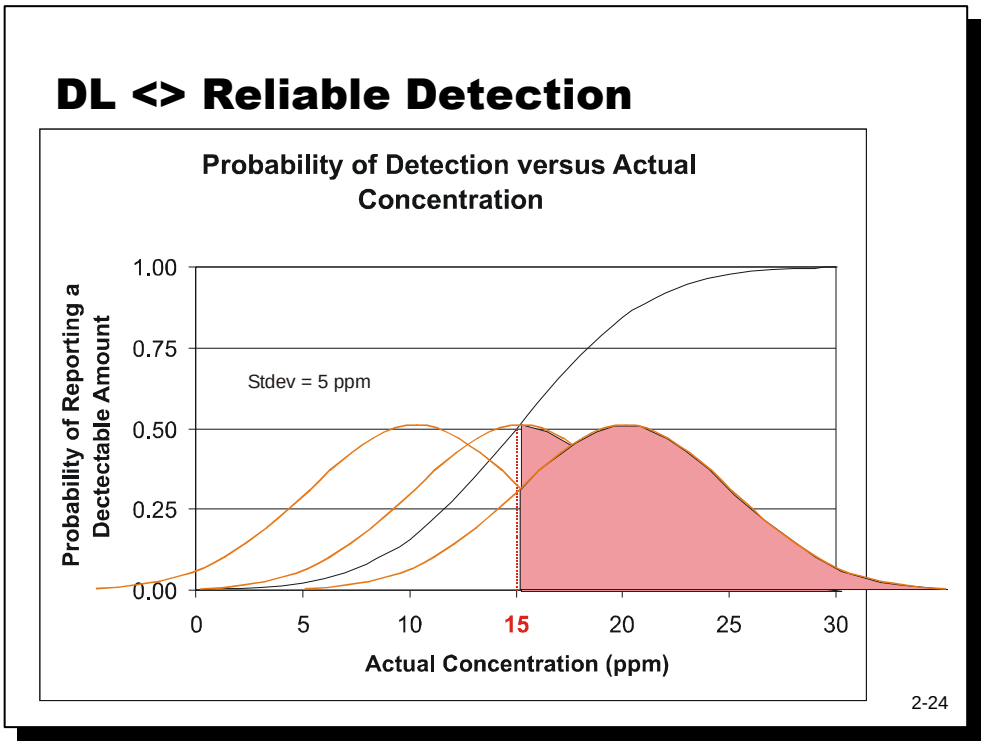


Notes

- The graphic above illustrates the frequency of XRF responses when the element is not present. Assume that a sample does not have an element present (or that it is present at trace levels). If a measurement of a sample is taken with an XRF, the XRF would record a concentration present for that element just because of the random nature of x-ray counting statistics. If a large number of repeat measurements were taken, a distribution or frequency plot of those "random" concentrations could be generated, as is shown here, with a measurable standard deviation (SD) and centered on zero (assuming there was no low end bias in the XRF calibration). By moving three standard deviations up from zero and calling that the DL (consistent with SW846 Method 6200), then almost 100 percent of the concentration values generated when the element is not present would be less than the DL. In other words, if the instrument records a result greater than this DL, then it is very likely that in fact the element is present.

  In practice, the SD (or error) associated with an XRF reading when an element is not present is estimated based on counting statistics, and this in turn is the basis for the detection limit reported by the instrument.

  Instruments frequently either report the detection limit when an element is not detected, or a result and associated error when an element is detected. A question to ask: what the detection limit was, say, for lead if the XRF reports lead at 900 ppm with an error of 100 ppm. Instruments typically do not provide the information necessary to answer that question. In this particular case, assuming the detection limit is 300 ppm (or 3 times 100 ppm) would not be correct, since reported error is a function of the actual amount of an element present.



Notes

- This definition of detection limits is somewhat arbitrary and not particularly useful in the case of metals since almost every environmental sample does have metals present at background levels.

  A more useful question to ask: what the concentration has to be for the XRF to reliably detect the presence of a particular element. It turns out that this level is twice the detection limit as reported by the instrument. The graphics on this slide illustrate this fact.

  Suppose for a particular element the detection limit of the instrument is 15 ppm (i.e., the measurement standard deviation associated with a very low concentration is 5 ppm). The black curve shows the probability of the XRF reporting a value above the detection limit as a function of the actual concentration present in a sample. Even if the actual concentration is less than the detection limit, there is a chance (although less than 50 percent) that the XRF will reported a detectable value. That probability grows to 50 percent at the detection limit itself, and eventually approaches one when one is twice the detection limit.

  If the actual concentration in a sample was 10 ppm and one did repeated measurements, one would generate a frequency plot that looks like the first bell-shaped curve. The portion shaded pink indicates the fraction of repeated measurements yielding a result greater than 15 ppm. As the second bell-shaped frequency plot demonstrates, when the actual concentration is at the detection limit, we have only a 50-50 chance of the instrument reporting a number greater than the detection limit. However, as the actual concentration increases, the probability of detection increases as well.

For Any Particular Instrument, Detection Limits Are Influenced By …

- Measurement time (quadrupling time cuts detection limits in half)
- Matrix effects
- Presence of interfering or highly elevated contamination levels

Consequently, the DL for any particular element will change, sometimes dramatically, from one sample to the next, depending on sample characteristics and operator choices

2-25

Notes

- **Measurement time:** The precision or reproducibility of a measurement will improve with increasing measurement time. Increasing the count time by a factor of 4 will provide 2 times better precision. Consequently increasing the count time by a factor of 4 will cut detection limits by a factor of two. Of course, increasing count time decreases sample throughput, so selecting the appropriate measurement time is a trade-off between the desired detection limits and per-sample measurement costs.

- **Matrix effects:** Physical matrix effects result from variations in the physical character of the sample. These variations may include such parameters as particle size, uniformity, homogeneity, and surface condition. One way to reduce error associated with variation in particle size is to grind and sieve all soil samples to a uniform particle size. Differences in matrix effects can result in differences in detection limits from one sample to the next.

- **Presence of interfering or highly elevated contamination levels:** Chemical matrix effects result from the differences in the concentrations of interfering elements. These effects occur as either spectral interferences (peak overlaps) or as x-ray absorption and enhancement phenomena. Both effects are common in soils contaminated with heavy metals. For example, iron tends to absorb copper x-rays, reducing the intensity of the copper measured by the detector, while chromium will be enhanced at the expense of iron because the absorption edge of chromium is slightly lower in energy than the fluorescent peak of iron. When present in a sample, certain x-ray lines from different elements can be very close in energy and, therefore, can cause interference by producing a severely overlapped spectrum. The presence of interference effects will raise detection limits.

Examples of DL …

(detection limits in black, "hits" in italicized **red**)

Analyte	Innov-X[1] 120 sec acquisition (**soil standard** – ppm)	Innov-X[1] 120 sec acquisition (**alluvial deposits** - ppm)	Innov-X[1] 120 sec acquisition (**elevated soil** - ppm)
Antimony (Sb)	61	55	232
Arsenic (As)	6	7	***29,200***
Barium (Ba)	NA	NA	NA
Cadmium (Cd)	34	30	**598**
Calcium (Ca)	NA	NA	NA
Chromium (Cr)	89	100	***188,000***
Cobalt (Co)	54	121	766
Copper (Cu)	21	17	661
Iron (Fe)	***2,950***	***22,300***	***33,300***
Lead (Pb)	***12***	8	***447,000***
Manganese (Mn)	56	***314***	1,960
Mercury (Hg)	10	8	481
Molybdenum (Mo)	11	9	***148***
Nickel (Ni)	***42***	31	451

2-26

Notes

- The table illustrates the fact that detection limits can change dramatically from sample to sample. Here we see three different sets of results from the same Innov-X unit, in each case collected with a 120-second acquisition time. The italicized red numbers in this table are actually quantified values (i.e., detects), while the black text numbers are detection limits. Results for three different samples are presented. The first is for a spiked matrix (the spiked element is not present in this table). The second is for a background soil sample taken from alluvial deposits. The third is for a highly contaminated sample taken beneath a leaking waste sewer line at a chemical facility.

  The effect on highly elevated lead and chromium on the detection limits for other elements is severe. The detection limit for mercury jumps from around 10 ppm to almost 500 ppm, a 50 times factor change.

  There is another thing to note about these data. The concentration levels reported for chromium and lead for the contaminated sample fall outside the calibrated range. These values would and should be taken with a large dose of skepticism … the levels of lead and chromium in this sample are undoubtedly extremely high, but the ability of the XRF to accurately quantify them at these levels would be very suspect.

To Report, or Not to Report: That is the Question!

- Not all instruments/software allow the reporting of XRF results below detection limits
- For those that do, manufacturer often recommends against doing it
- Can be valuable information if careful about its use … particularly true if one is trying to calculate average values over a set of measurements

2-27

Notes

- **Not all instruments/software allow the reporting of XRF results below detections limits:** The definition of DLs for the XRF is somewhat arbitrary. The fact is the instrument measures concentrations for all elements contained in its calibration library when a measurement is taken, although typically an XRF will then only report those measured values when they exceed its calculated detection limit.

- **For those that do, manufacturer often recommends against doing it:** Once non-detects are recorded and pushed forward, there is all conflicting guidance about how they should then be handled when constructing averages or calculating upper confidence limits on averages. Confusion can be avoided if one forces the XRF to report the original concentration estimate (above or below the DL) and then uses that data.

- **Can be valuable information if careful about its use . . . particularly true if one is trying to calculate average values over a set of measurements:** Values below DLs can be useful when calculating average values over a set of measurements. If the instrument's calibration is unbiased for low levels of the element of interest, using measured values below the instrument's detection limits can yield more accurate assessments of average concentrations that flagging readings as non-detects and substituting some arbitrary value such as the detection limit, or half the detection limit, in average value calculations. Great care and full disclosure are necessary when using values below detection limits.

XRF Data Comparability

- Comparability usually refers to comparing XRF results with standard laboratory data
- Assumption is one has samples analyzed by both XRF and laboratory
- Regression analysis is the ruler most commonly used to measure comparability
- SW-846 Method 6200: "If the r2 is 0.9 or greater…the data could potentially meet definitive level data criteria."

2-28

Notes

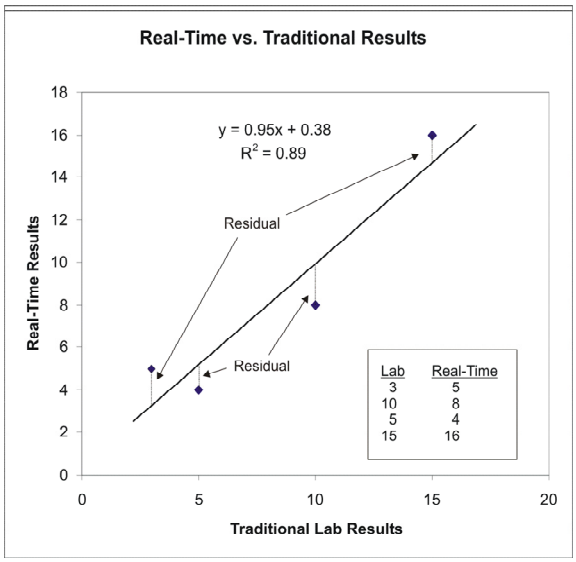
- **Comparability usually refers to comparing XRF results with standard laboratory data:** The comparability of the XRF analysis is determined by submitting XRF-analyzed samples for analysis at a laboratory. The XRF results are then compared with the laboratory results. The question of "comparability" or lack of "comparability" is what usually determines whether a technique's data sets are deemed useful or not.

- **Assumption is one has samples analyzed by both XRF and laboratory:** The confirmatory samples must be splits of well homogenized sample material. The confirmatory samples should be selected from the lower, middle, and upper range of concentrations measured by the XRF. They also should include samples with element concentrations at or near the site action levels.

- **Regression analysis is the ruler most commonly used to measure comparability:** The results of the confirmatory analysis and XRF analyses are usually evaluated with a least squares linear regression analysis.

- **SW-846 Method 6200: "If the r^2 is 0.9 or greater ... the data could potentially meet definitive level data criteria.:"** Method 6200 states that the method of confirmatory analysis must meet the project and XRF measurement data quality objectives. The method also suggests that the r^2 for the results should be 0.7 or greater for the XRF data to be considered screening level data. Finally, the method states that if the R^2 is 0.9 or greater and inferential statistics indicate the XRF data and the confirmatory data are statistically equivalent at a 99 percent confidence level, the data could potentially meet definitive level data criteria.

What is a Regression Line?



2-29

Notes

- The scatter-plot in this slide illustrates how a regression analysis works. The data in the lower-right table represents our collaborative data set: four samples, with each having both a traditional laboratory result and a real-time result (e.g. XRF). Plotting these data give us the scatter-plot shown. Assuming there's a linear relationship between results generated by the laboratory and results generated by the real-time technique, the question is finding that linear relationship.

  The line shown represents the results from a regression using these data. The regression line represents the "best fit" line. "Best fit" here is defined as the line that minimizes the sum of the squared residuals. A residual is the vertical distance separating a regression line and a data point.

Regression Terminology

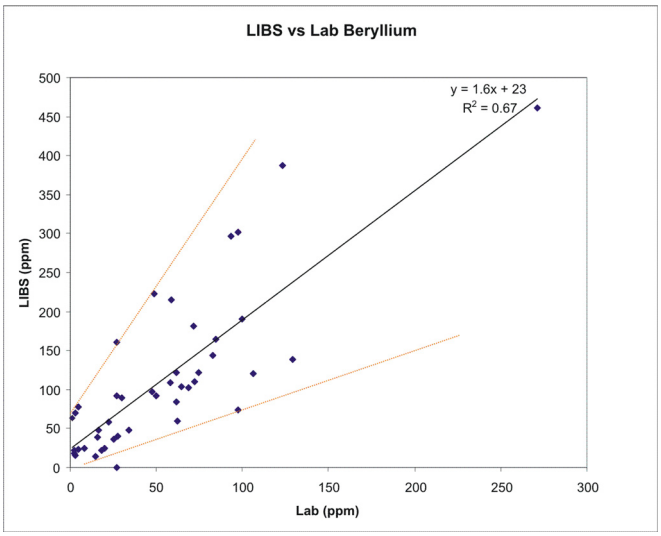
- Scatter Plot: graph showing paired sample results
- Independent Variable: x-axis values
- Dependent Variable: y-axis values
- Residuals: difference between dependent variable result predicted by regression line and observed dependent variable
- Adjusted R^2: a measure of goodness-of-fit of regression line
- Homoscedasticity/Heteroscedasticity: Refers to the size of observed residuals, and whether this size is constant over the range of the independent variable (homoscedastic) or changes (heteroscedastic)

2-30

Notes

- **Regression terminology:** The following are regression terms:
  - » Scatter Plot – graph showing paired sample results
  - » Independent Variable – x-axis values, usually the lab result
  - » Dependent Variable – y-axis values, usually the XRF result
  - » Residuals – difference between dependent variable result predicted by regression line and observed dependent variable
  - » Adjusted R^2 – a measure of goodness-of-fit of regression line
  - » Homoscedasticity/Heteroscedasticity – refers to the size of observed residuals, and whether this size is constant over the range of the independent variable (homoscedastic) or changes (heteroscedastic)

Heteroscedasticity is a Fact-of-Life for Environmental Data Sets



2-31

Notes

- Heteroscedasticity is unfortunately a fact-of-life for environmental collaborative data sets. The laser induced breakdown spectroscopy (LIBS)/laboratory scatter-plot illustrates the concept of heteroscedasticity. We can fit a regression line to these data, with the resulting line and its equation shown. The orange lines bracketing the regression line above and below give a sense for how the size of residuals change as concentrations increase. For low concentrations, the scatter-plot points are tightly clustered around the regression line, giving rise to relatively small residuals. As concentrations increase, the "scatter" of points around the line steadily increases. The result is that residuals for higher-concentration points are much larger than what they are for lower concentration values. This increasing residual size as concentrations increase is called heteroscedasticity.

  There is a simple physical explanation for heteroscedasticity in environmental collaborative data…analytical error tends to increase as concentrations increase.

  The concept is important because regression analyses often include upper confidence limit (UCL) lines or upper tolerance level (UTL) lines that bracket the regression line. The problem with this is that UCL and UTL calculations derived from a regression analysis are only valid if the underlying data are homoscedastic … which environmental data such as these never are.

Appropriate Regression Analysis

- Based on paired analytical results, ideally from same sub-sample
- Paired results focus on concentration ranges pertinent to decision-making
- Non-detects are removed from data set
- Best regression results obtained when pairs are balanced at opposite ends of range of interest

2-32

Notes

- **Based on paired analytical results, ideally from same sub-sample:** Such an analysis should be based on paired results, ideally with the analytical work done on the same sub-sample where possible to minimize the effects of sample preparation. Poor comparability results are often the result of poorly prepared samples and not analytical issues.

- **Paired results focus on concentration ranges pertinent to decision-making:** The paired results should focus on the concentration range pertinent to decision-making. Often times field analytical methods have a more limited dynamic range within which they provide accurate results. This means that it is unreasonable to expect a good, strong linear relationship for two methods over the complete range of concentrations (which may span several orders of magnitude) present at a site. What is important is to determine whether such a relationship exists over the range in which making decisions is important.

- **Non-detects are removed from data set:** Non-detects should be removed from a regression analysis because they will skew regression results.

- **Best regression results obtained when pairs are balanced at opposite ends of the range of interest:** The best regression results are obtained when the data used are balanced, for example, half are at the lower end of interest, and half are at the higher end of interest. WARNING: unbalanced data sets (i.e., data sets where most of the points are clustered at the low end with one or two high value) will yield unstable and likely misleading regressions.

Evaluating Regression Performance

- No evidence of inexplicable "outliers"
- Balanced data sets
- No signs of correlated residuals
- High R^2 values (close to 1)
- Constant residual variance (homoscedastic)

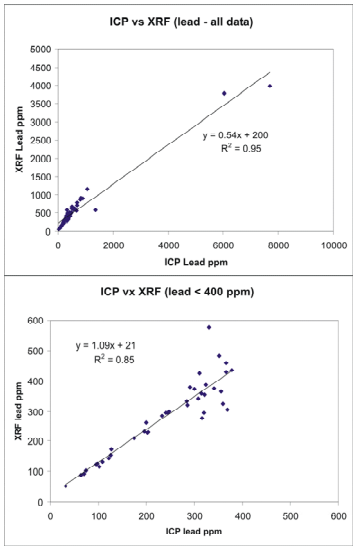
2-33

Notes

- **No evidence of inexplicable outliers:** There should not be any evidence of outliers. Outliers are points that clearly fall well away from the regression line and appear to be different than the rest.
- **Balanced data sets:** Data sets should be balanced.
- **No signs of correlated residuals:** There should not be any signs of correlated residuals. Correlated residuals refer to the situation where a group of points consistently fall above or below the regression line.
- **High R^2 values (close to 1):** A good regression should have a high R^2 value, preferably close to 1 (will range between 0 and 1).
- **Constant residual variance (homoscedastic):** A good regression also should have constant residual variance across the concentration range, or in other words the data should be homoscedastic. Unfortunately for environmental collaborative data sets, this is never the case.

Example: XRF and Lead

- Full data set:
  - » Wonderful R^2
  - » Unbalanced data
  - » Correlated residuals
  - » Apparently poor calibration
- Trimmed data set:
  - » Balanced data
  - » Correlation gone from residuals
  - » Excellent calibration
  - » R^2 drops significantly



2-34

Notes

- Here is an example based on XRF analyses of lead in soil samples. The top graphic shows a scatter plot based on the complete data set collected. The regression line has a wonderful R^2 value, but has several obvious visual deficiencies. These include unbalanced data (most of it clustered at the low end with only two points at the high end), correlated residuals, and what appears to be a poor calibration for the XRF based on the slope of the line.

  The second data set has had its data trimmed to include only those concentrations that fall within the range truly of interest from a decision-making perspective. These data are balanced across the concentration range of interest. The correlations are gone from the residuals. The slope corresponds to what would be expected from a calibrated XRF. Note that the R^2 value is actually less, though, then the first example, even though the second regression is clearly superior, underscoring the problems with simply using R^2 values as a measure of regression performance and hence field analytic data quality and usability.

  There is something else to note in the second scatter plot. The spread of the data around the line increases as concentrations increase. This is called heteroscedasticity - the variance of our data is not constant over the range of observed concentrations. The presence of heteroscedasticity is a given in environmental data, and complicates the interpretation of regression results. Care must be taken when interpreting UCLs and UTLs for regression lines when heteroscedasticity is present.

Converting XRF Data for Risk Assessment Use

- Purpose: making XRF data "comparable" to lab data for risk assessment purposes
- To consider:
  - » Need for "conversion" may be an indication of a bad regression
  - » XRF calibrations not linear over the range of concentrations potentially encountered
  - » Extra variability in XRF data not an issue (captured in UCL calculations when estimating EPC)
  - » Contaminant concentration distributions are typically skewed … lots of XRF data may provide a better UCL/EPC estimate than a few lab results even if the regression is not great

2-35

Notes

- There often is a desire to make XRF data "comparable" to laboratory data. Sometimes the desire is to pool XRF and laboratory data when calculating an average concentration for a given area, or when using the data to support a risk assessment. "Correcting" XRF data usually is done using a regression equation … plugging the XRF result in and popping out the equivalent laboratory result. The fact that "correction" is required at all suggests that there is some form of bias present in XRF data that need to be corrected to bring the XRF data set in line with laboratory data.

  While the desire to do this may be great, there are several points that need to be considered. The first and foremost is that the need for a conversion is an indication that something is not quite right with the XRF and/or the laboratory data (or with the regression that might be used to do the conversion). Before converting XRF data, develop an understanding of why the conversion is necessary. If the conversion appears to be required simply because the regression analysis was bad, there is the possibility that good XRF data will be converted into bad XRF data.

A Cautionary Example…

- Four lab lead results: 20, 24, 86, and 189 ppm
- ProUCL 95%UCL Calculations:
  - »Normal: 172 ppm
  - »Gamma: 434 ppm
  - »Lognormal: 246 – 33,835 ppm
  - »Non-parametric: 144 – 472 ppm
- *Four samples are not enough to either understand the variability present, or the underlying contamination distribution*

2-36

Notes

- Here is a cautionary example of how a few "definitive" laboratory data can cause trouble with interpretation.

  These are actual lead results from an exposure unit. These data were going to be used to support a risk calculation. The risk calculation required an Exposure Point Concentration (EPC). Typically EPCs are set to the 95 percent UCL on the mean.

  In this particular case, ProUCL, a free EPA software package, was used to calculate the 95 percent UCL. ProUCL allows calculation of the 95 percent UCL values under differing assumptions about the underlying distribution (i.e., normal, lognormal, gamma). As the numbers here demonstrate, depending on the underlying distribution selected the estimated EPC varies wildly.

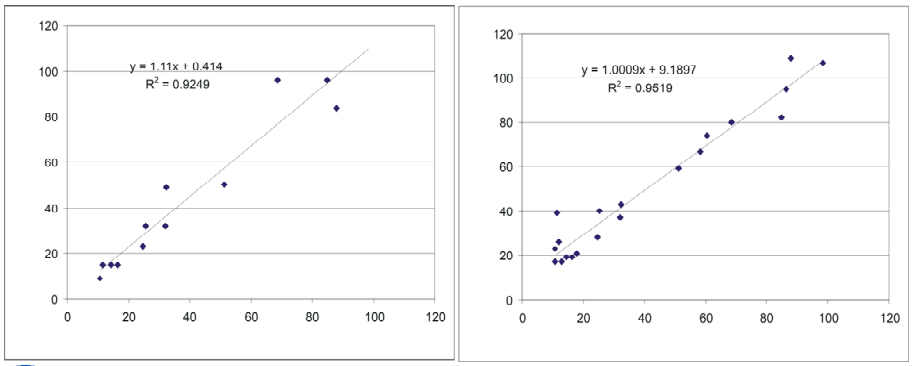
  The take away point is that a bunch of "poorer" analytical quality XRF data may actually provide a more sound basis for EPC estimation than a couple of "gold standard" laboratory sample results.

Internet

ProUCL can be found at *www.epa.gov/esd/tsc/software.htm*

Will the "Definitive" Data Please Stand Up?

One of these scatter plots shows the results of arsenic from two different ICP labs, and the other compares XRF and ICP arsenic results.
Which is which?

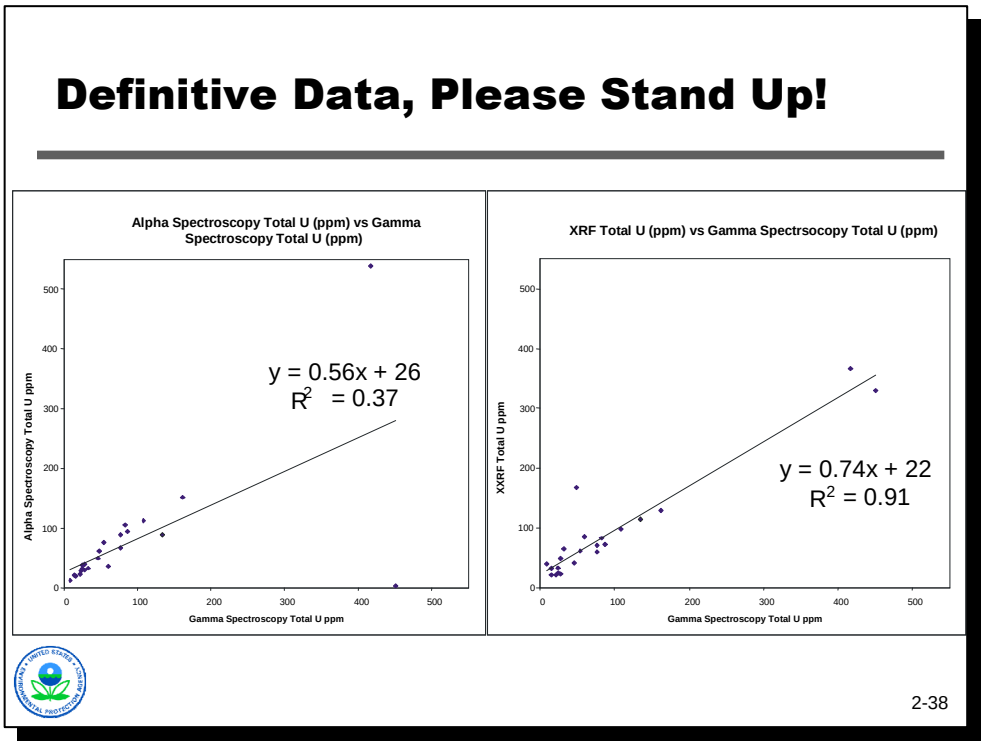


2-37

Notes

- These two scatter plots show paired data results for arsenic. In one case, samples first analyzed by XRF were then sent off for ICP analyses. In the other case, the same sample was split and sent for ICP analyses to two different laboratories. Which of these two correspond to the ICP/ICP comparison and which to the XRF/ICP comparison?

  The take home point is quite simple. Traditional analyses often are treated as though they are "definitive" and free from error. When the results of an alternative analysis such as an XRF are compared to those from a traditional laboratory, any differences observed are attributed to poor performance on the alternative analysis's part. The reality is not so simple. Traditional analyses also include "errors" that need to be recognized.



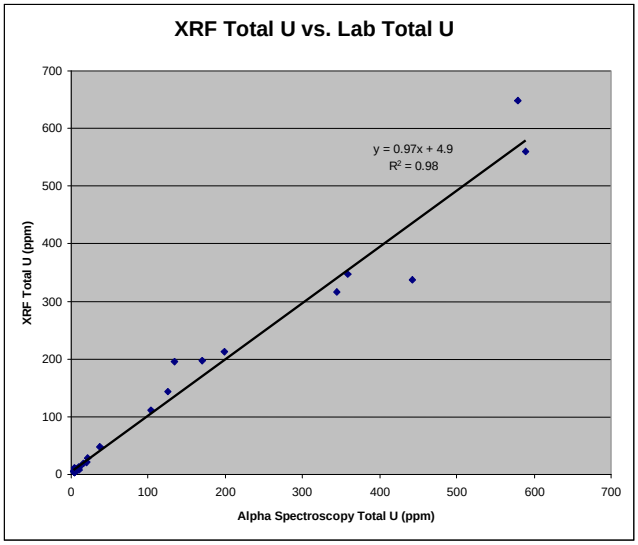
Notes

- At this site three different analytical techniques were employed for quantifying the presence of uranium. XRF was the field technique. Both gamma spectroscopy and alpha spectroscopy are traditional, widely accepted, "definitive" methods for measuring uranium. They differ in that gamma spectroscopy does not require an extraction, but has higher detection limits than alpha spectroscopy and also requires some assumptions be made about ratios of naturally occurring uranium isotopes. Alpha spectroscopy, on the other hand, has lower detection limits and can measure total uranium directly, but because it requires additional sample handling and preparation is more prone to errors introduced by the laboratory technician.

  The scatter plot on the right compares gamma spec and alpha spec results for split samples. The graph on the left compares XRF with gamma spec results.

  Note that there is excellent agreement between XRF and gamma spec, and poor agreement between gamma and alpha spec. A closer inspection of the gamma and alpha spec results identifies two samples as being particular problems. Which of these techniques deserves the "definitive" designation? Is the XRF any less "definitive" than either gamma or alpha spec?

How Good Can an XRF Be?



2-39

Notes

- This scatter plot shows exactly how good XRF can potentially be. This plot compared XRF uranium results for cup samples with alpha spectroscopy uranium results for the same cups that were originally analyzed by XRF. In this particular case, over this range of uranium concentrations, the XRF provided data that were highly comparable to alpha spectroscopy.

Take-Away Comparability Points

- Standard laboratory data can be "noisy" and are not necessarily an error-free representation of reality
- Regression R^2 values are a poor measure of comparability
- Focus should be on decision comparability, not laboratory result comparability
- Examine the laboratory duplicate paired results from traditional QC analysis - the split field vs. lab regression cannot be expected to be better than the lab's duplicate vs. duplicate regression

2-40

Notes

- **Standard laboratory data can be "noisy" and are not necessarily an error-free representation of reality:** It is a mistake to believe that standard laboratory data are free of errors. This can be seen when laboratory analyses from two different laboratories are compared to one another in the same way that XRF and laboratory data are compared.

- **Regression R^2 values are a poor measure of comparability:** Regression performance should be judged using a number of factors, not just the R^2 value.

- **Focus should be on decision comparability, not laboratory result comparability:** Decision comparability judges whether or not data is suitable for the decision at hand. XRF data may be suitable for decisions about whether an action level has been exceeded or for calculating UCL/EPC even when the regression is not perfect.

- **Examine the lab duplicate paired results from traditional QC analysis:** Frequently the regression from duplicate paired results is poor. It is unreasonable to expect the split field (XRF) versus laboratory regression to be better than the laboratory's duplicate versus duplicate regression.

What Affects XRF Performance?

- **Measurement time** – the longer the measurement, the better the precision
- **Contaminant concentrations** – potentially outside calibration ranges, absolute error increases, enhanced interference effects
- **Sample preparation** – the better the sample preparation, the more likely the XRF result will be representative

(continued)

2-41

Notes

- **Measurement time:** The longer the measurement time or count time, the better the precision will be.
- **Contaminant concentrations:** Contaminant concentrations may be outside of the calibration ranges. Other contaminants may cause interference effects.
- **Sample preparation:** The better the sample preparation, the more representative the XRF results will be of actual conditions.

What Affects XRF Performance?

- **Interference effects** – the spectral lines of elements may overlap
- **Matrix effects** – fine versus coarse grain materials may impact XRF performance, as well as the chemical characteristics of the matrix
- **Operator skills** – watching for problems, consistent and correct preparation and presentation of samples

2-42

Notes

- **Interference effects:** The spectral lines of elements may overlap distorting results for one or more elements.
- **Matrix effects:** Physical matrix effects, such as fine versus course grain materials, may impact XRF performance. In addition, chemical characteristics of the matrix also may impact XRF performance.
- **Operator skills:** The level of operator skill can affect XRF performance. The operator should watch for problems and should practice consistent and correct preparation and presentation of samples. Operator skills, cannot be over-emphasized. Today's hand-held units are relatively simple to operate. But unless care is taken by an experienced operator during data collection, the data generated may be of very questionable quality … or perhaps completely unknown quality.

  A later module will discuss appropriate QC in much more detail. The important take-away point here is that the XRF is technically capable of producing high quality analytical results *if* the appropriate QC is in place to identify and correct problems that potentially compromise data quality.

What Are Common XRF Environmental Applications?

- *In situ* and *ex situ* analysis of soil samples
- *Ex situ* analysis of sediment samples
- Swipe analysis for removable contamination on surfaces
- Filter analysis for filterable contamination in air and liquids
- Lead-in-paint applications

2-43

Recent XRF Technology Advancements…

- Miniaturization of electronics
- Improvements in detectors
- Improvements in battery life
- Improved electronic x-ray tubes
- Improved mathematical algorithms for interference corrections
- Bluetooth, coupled GPS, connectivity with PDAs and tablet computers

2-44

Notes

- **Recent XRF technology advancements:** The following advancements in XRF technology have improved the performance of the technology:
  - » Miniaturization of electronics – this has made the instruments more portable
  - » Improvements in detectors – with a corresponding lowering of detection limits
  - » Improvements in battery life – which increases sample throughput by reducing instrument downtime and improves general field application
  - » Improved electronic x-ray tubes – which improves performance of the units
  - » Improved mathematical algorithms for interference corrections – which expands the applicability of the technology
  - » Bluetooth, coupled global positioning system (GPS), connectivity with personal digital assistant (PDA) and tablet computers – which enhances data collection, management, and storage

… Contribute to Steadily Improving Performance

Analyte	DL in Quartz Sand by Method 6200 (600 sec – ppm)	TN 900 (60 to 100 sec) – ppm	Innov-X[1] 120 sec acquisition (soil standard – ppm)
Antimony (Sb)	40	55	61
Arsenic (As)	40	60	6
Barium (Ba)	20	60	NA
Cadmium (Cd)	100	NA	34
Chromium (Cr)	150	200	89
Cobalt (Co)	60	330	54
Copper (Cu)	50	85	21
Iron (Fe)	60	NA	***2,950***
Lead (Pb)	20	45	***12***
Manganese (Mn)	70	240	56
Mercury (Hg)	30	NA	10
Molybdenum (Mo)	10	25	11
Nickel (Ni)	50	100	***42***

2-45

Notes

- This last table shows the results of XRF technology improvements over the years. The first data column shows XRF detection limits as reported in Method 6200 in the best of conditions … quartz sand with a 600 second acquisition. The second column shows the performance of a TN 900 XRF in the mid- to late-1990s (the table containing these results is dated 1998) with a 60 to 100 second acquisition. One would expect these values to be less than half of what is reported if a 600 second acquisition time had been used. The last column shows data collected with an Innov-X unit in 2006 for a spiked soil standard (the spiking element is not present in this table). Results in bold red indicate actual measured data. Plain text results are reported detection limits. The detection limit differences are marked for a number of samples.

  For example, in the case of arsenic the Innov-X detection limit is one tenth that of the TN 900 back in the 1990s. This improvement is not vendor-specific … in fact all vendors of portable XRF technologies have made significant strides in improving instrument performance in the last decade. One would expect those improvements to continue and be reflected in falling detection limits and better handling of interference effects.

Q&A

2-46

Module 3

Representativeness

Module 3:

Representativeness

3-1

Data Collection Tied to Specific Decision

- Contaminant above background levels? (SI)
- Human health or ecological risks unacceptable? (RI)
- Contaminant concentrations above the cleanup criteria? If so, what should be done? (FS/RD)
- Should soil/sediment removal/treatment continue? be modified? or stop? (RA)
- Many kinds of data used to guide decisions.

3-2

Notes

There are many kinds of decisions made during the investigation and cleanup of contaminated sites. Data collection and interpretation MUST be designed specifically to answer a specific question in the context of a particular site and its characteristics. Site cleanup and its environmental context is simply too complex for a generic data set to be collected that can be used to answer any question that arises.

- **Contaminant above background levels? (SI):** The site inspection (SI) phase establishes whether contaminant levels are significantly above background so that they can be scored using the Hazard Ranking System model.

- **Human health or ecological risks unacceptable? (RI):** During the remediation investigation (RI), environmental data is collected to determine if the nature and extent of contamination present an unacceptable risk. Environmental data are also collected during the RI to evaluate remediation technologies.

- **Contaminant concentrations above the cleanup criteria? If so, what should be done? (FS/RD):** During the feasibility study (FS) and remedial design (RD), environmental data are used to determine if contaminant concentrations are above preliminary remediation goals in the FS report or cleanup levels in the Record of Decision (ROD). Environmental data collected during RD typically are used to better estimate the volume of contaminated media exceeding contaminant concentrations.

- **Should soil/sediment removal/treatment continue? be modified? or stop? (RA):** Environmental data collected during the remedial action (RA) are used to determine if the cleanup objectives have been met and to monitor and optimize the operations of ongoing treatment systems.

Nirvana

- Representative, fast, cheap method able to run lots of samples and provide "definitive data"
- Reality Bites:
  - »Expensive, time consuming analytics
  - »Few samples collected
  - »Measurement errors & interpretation issues exist

3-3

Notes

- **Representative, fast, cheap method able to run lots of samples and provide "definitive data:"** In a perfect world, the collection of representative data would be fast and inexpensive using methods that generate a large volume of data that is considered to be "definitive." What is the fundamental issue? If we had a less expensive way of providing high density data quickly that was definitive relative to whatever decision that needed to be made, we would not be having this workshop.

- **Reality bites:** In reality we are asked to make decisions about a site from a very limited number of samples, whose results, at times, are subject to interpretation and error themselves. Traditionally, data collection is very expensive and uses time consuming analytical techniques. Because of the time and expense, only a few samples are collected. Relying on just a few samples leads to interpretation issues because of predictable measurement errors.

"Representative Sample" is Defined in Regulation

RCRA: "Representative sample means a sample of a universe or whole (e.g., waste pile, lagoon, ground water) which can be expected to exhibit the average properties of the universe or the whole (40 CFR 260.10)."

Evades several questions, such as…
…how do you take such a sample?

3-4

Notes

- **RCRA definition:** Regulations under RCRA define the term representative sample as, "Representative sample means a sample of a universe or whole (e.g., waste pile, lagoon, ground water) which can be expected to exhibit the average properties of the universe or the whole (40 CFR 260.10)." The regulation makes it sound easy. However, the regulation completely overlooks the physical reality that complexities (heterogeneity) exist on several spatial scales within a "waste pile, a lagoon, ground water." The difficulty lies in taking such a sample that represents the average of the universe or whole.

  In common usage of the term "sample," the regulatory language refers to a physical sample … a portion of material taken from the bulk mass of a parent material. In the statistical sense, "sample" refers to the GROUP of physical samples taken from the parent material ("the whole").

"Representative Sample" per RCRA

- Language unclear whether statistical or single physical sample is intended
  - »Common usage of "sample" implies a single grab sample can represent the whole
  - »Can we expect a single tiny sample to represent the average for a "whole"?
  - »Reasonable ONLY if homogeneity throughout the whole can be assumed
  - »"Representative of…"

3-5

Notes

- **Language unclear whether statistical or single physical sample is intended:** The concept of "representativeness" is very vague and ill-defined for many working in the environmental field. This is because standard or regulatory definitions for "representativeness" also tend to be rather vague.

  » The RCRA solid waste regulations at 40 CFR §260.10 define a **representative sample** as: "a sample of a universe or whole (e.g., waste pile, lagoon, ground water) which can be expected to exhibit the average properties of the universe or whole."

Internet

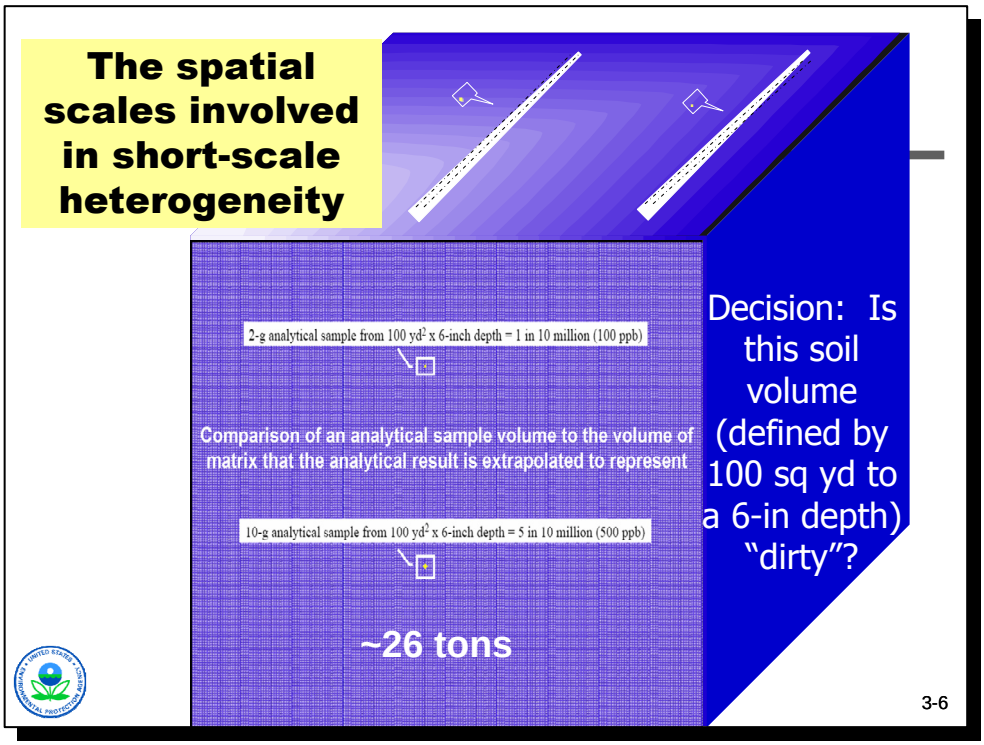
http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=817591009b20d11a9dd16ef3f173c6aa&rgn=div8&view=text&node=40:23.0.1.1.1.2.1.1&idno=40

» American Society for Testing and Materials (ASTM) (consensus standard D 6044-96) defines a representative sample as "a sample collected in such a manner that it reflects one or more characteristics of interest (as defined by the project objectives) of a population from which it was collected."

It is not clear from these definitions whether the term "sample" refers to a statistical sample (made up of a number of individual specimens) or to a single sample, or whether the authors intended to allow either interpretation. A critical issue with the RCRA regulatory definition is that representativeness is defined in terms of an "average." Operationalizing this definition for contaminated site cleanup poses problems. First, the extreme heterogeneity of environmental matrices and contaminants makes determination of a statistical "average" difficult and expensive. Second, how is it decided how to define the volume of the whole

over which a property is to be averaged? Third, some environmental and engineering decisions (notably, those decisions involved with selecting and designing remedial systems) should not be made based on an "average," if that average encompasses wide variation.

In order to be useful for managing projects in the environmental field, the concept of representativeness must be made more concrete and meaningful. This can be done by simply adding the word "of." This adjusts the terminology and people's thinking to make it clear that data or other information must be representative of the intended decision or specific property under investigation. In this way, "representativeness" becomes linked to a concrete decision and decision unit rather than just an abstract "average." The ASTM definition seems to reflect this same kind of approach ("…reflects one or more characteristics of interest (as defined by the project objectives)…"



Notes

- The cube represents the volume of soil encompassed by 100 square yards to a 6-inch depth (about 26 tons of soil). The higher dot represents the relative scale for a single 2-gram sample taken from that volume. The lower dot represents the relative scale for a single 10-gram sample.

  Assume a sample is designated as a "representative sample." The term is meaningless unless more information is supplied. What property is to be represented? Is it the highest concentration of one contaminant, of all contaminants at once? Is it the average contaminant concentration? What is the volume over which the result is to be represented?

  What is absolutely known is that the reported result is representative of the contaminant concentration of the analytical sample. Any extension of that concentration to other parts of matrix not analyzed must be supported by evidence that shows that extrapolation of the result to a larger volume is justified.

  Each dot represents a 2-gm soil sample within a 100 sq yd x 6-inch volume.

Calculations:

100 sq yd x 6 inch depth = 1.28 x 10exp(7) cu cm = 17 cu yd

Soil density from Internet = 120 lb per cu ft = 1080 lb per cu yd

120 lb/cu ft = 1.62 ton per cu yd = 5.41 x 10exp(4) gram/2.83 x 10exp4 cu cm = 1.9 g/cu cm (for in place, “consolidated” soil; not soil dug up & in bin)

1.28 x 10exp(7) cu cm x 1.9 g/cu cm = 2.4 x 10exp(7) gram total soil wt for the 100 sq yd x 6-inch depth soil matrix = 24 metric tons = 26.4 tons

Ratio of 2-g soil sample (for metals analysis) to total matrix wt = 0.83 x 10exp(-7) ~ 1 x 10exp(-7) = 100/10exp(9)

100/10exp(9) = 100 ppb

To get 10 million parts: (2.2 x 100) cubed = 10.6 x 10exp(6) = 10 million

Therefore, about 200 (length) x 200 (width) x 200 (ht) = 10 million parts in 3-D

No of cells in 2-D area = 220 x 220 = 48,400 grid cells

Then repeat in the other 2 dimensions.

"Representative Sample" per ASTM

- ASTM (consensus standard D 6044-96) defines a representative sample as:

  "a sample collected in such a manner that it reflects one or more characteristics of interest (as defined by the project objectives) of a population from which it was collected."

3-7

Notes

- The ASTM definition reflects an approach that grounds "representativeness" in the decisions, ("…reflects one or more characteristics of interest (as defined by the project objectives)…" It is not clear from the definition whether the term "sample" in a statistical sense (a group) or in a physical sense is intended. But if the purpose of sampling is tied to making a decision, the sampling design that is best able to support that decision will guide whether a single physical sample is sufficient, or whether a statistical sample is called for.

What is a "Representative Sample?"

- "A representative sample is one that answers a question about a population with a given confidence."
- "A sample that is representative for a specific question is most likely not representative for a different question."
- A sample whose result can be CONFIDENTLY extrapolated to a defined portion of the parent matrix from whence it came.

3-8

Notes

- **"A representative sample is one that answers a question about a population with a given confidence.":** Because of the deficiencies fostered by the ambiguity of the tem "representative sample," better definitions are needed. The definitions by Ramsey & Hewitt, like the ASTM definition, tie "representativeness" to project questions about a population that is relevant to the decision-making process.

- **"A sample that is representative for a specific question is most likely not representative for a different question.":** A very important observation is that once a sampling design is geared toward answering a particular question, that sampling design will probably not be representative for a different question. This fact makes it unlikely that generic sampling designs will properly serve the needs of all the different investigation, risk, remedial, legal, etc. questions that often arise for sites.

- **Reference:** From "A Methodology for Assessing Sample Representativeness" Charles Ramsey & Alan Hewitt in *Environmental Forensics*, 6:71 – 75, 2005

  ISSN: 1527 – 5922 print / 127 – 5930 online
  DOI: 10.1080/15275920590913877

Representativeness: Answering a Question…Making a Decision

- Intended decision: Does the lead (Pb) concentration in soil exceed a regulatory limit of 600 ppm for the top 4 ft of a former 1-acre lagoon area?
- The lagoon is sampled via a 4-ft core. A 400-g sample is taken from the center of a core and sent to the lab.
- In the lab, a 1-g subsample is taken from top of the sample jar and analyzed for Pb. The result is 75 ppm.
- Can the result of 75 ppm Pb be confidently extrapolated to represent the top 4 ft of the 1-acre lagoon area to decide whether the 600 ppm limit is exceeded?
- What do we absolutely know about what this analytical result represents?

3-9

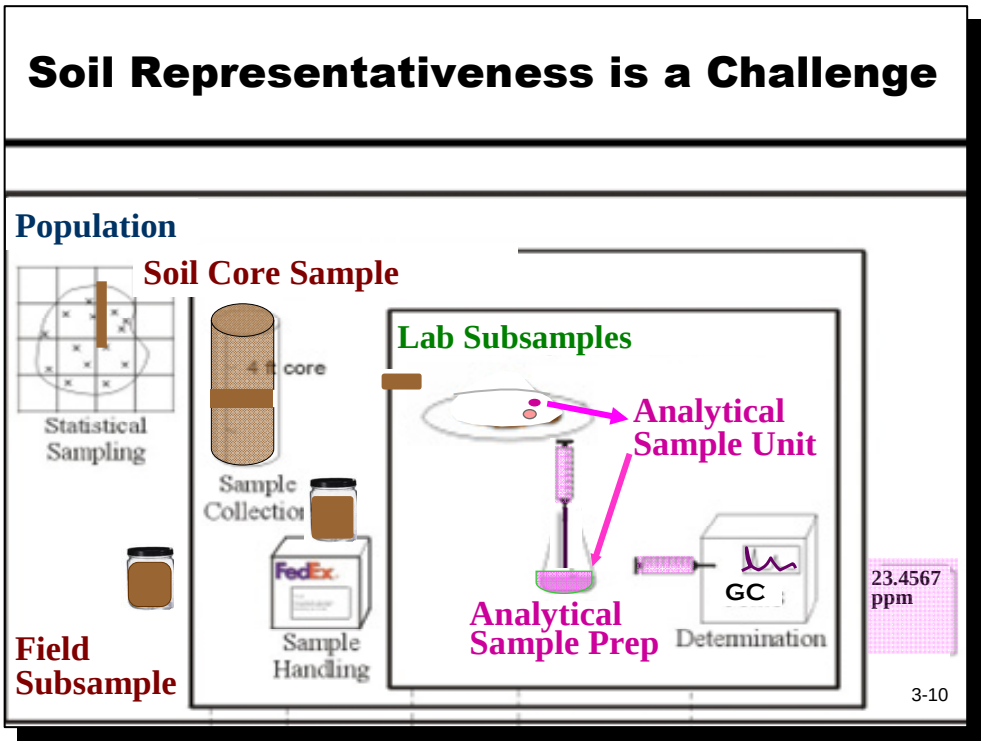
Notes

- The next several slides will explore the factors that complicate extrapolating the result of a 1-g subsample to a volume thousands of times larger.

  What do we know for sure that the 75 ppm result represents? Is it the soil in the former 1-acre lagoon? Is it the soil in the 4-ft core? Is it the soil in the 400-gram jar? Is it the soil in the 1-gram subsample which is digested and analyzed?

  The next question is … Can the result from the analytical sample be confidently used to answer our question for the 1-acre lagoon? In other words, do we have evidence that the concentration in the laboratory's 1-gram subsample is representative of the concentration for the whole 400 grams of soil in the jar? Then, are we sure that the concentration of the 400-gram soil sample is representative of the concentration for the entire 4-ft core? Finally, can we trust that the concentration of the 4-ft core represents the concentration across the whole 1-acre lagoon?

  What if we took another core in another place in the lagoon and repeated the procedure of core sampling, subsampling in the field for the jar, and then subsampling in the lab for the 1-gram analysis? How close can we expect the second result to be to the first?



Notes

- For soil and water analysis, the analytical sample is that volume of material that is ACTUALLY digested or extracted prior to analysis. No matter how evenly or unevenly the analyte is distributed within the analytical sample, they are homogenized within the stirred solution that frees them from the matrix. Any portion of liquid taken from the solution can be expected to be representative of the solution as a "whole."

  The big question is … "Does the analytical sample adequately represent the population from whence it came? The slide figure shows the long string of samples and subsamples that came before the analytical sample. If the analytical sample (and thus the analytical result) no longer represents the parent matrix, then the quality of the analysis itself is irrelevant. An analytical sample is "representative" if there is justification for extrapolating the concentration of the sample to the concentration of the parent matrix that sample is supposed to represent. If an analytical result is not representative, but the decision maker treats it as if it were, then decision errors are likely. All the while, the analytical method may be "perfect." Yet the data leads to an erroneous conclusion.

Within-Sample Heterogeneity: Micro-scale Sample Representativeness

Adapted from ITRC (2003)

Firing Range Soil Grain Size (Std Sieve Mesh Size)	Pb Concentration in fraction by AA (mg/kg)
Greater than 3/8" (0.375")	**10**
Between 3/8" and 4-mesh	**50**
Between 4- and 10-mesh	**108**
Between 10- and 50-mesh	**165**
Between 50- and 200-mesh	**836**
Less than 200-mesh	**1,970**
Bulk Total	**927 (wt-averaged)**

The decision determines representativeness

3-11

Internet

Data adapted from the Interstate Technology and Regulatory Counsel (ITRC). 2003. *Characterization and Remediation of Soils at Closed Small Arms Firing Ranges.* January. Available on-line at *http://www.itrcweb.org/SMART-1.pdf*

Notes

- The results of this study show how different particle sizes within the same jar of soil have different lead (Pb) concentrations. This is called "within-sample" or "micro-scale" soil heterogeneity because different concentrations of analyte occur on very small spatial levels within in a single jar of soil. Although the soil may look "homogenized," it really is not as long as different particles sizes exist in the sample jar. This would not matter IF the entire volume of soil in the jar was analyzed all at once. Analyzing the whole sample gives you the true concentration of the jar contents. However, jars usually contain 100 grams or more of soil. Common analytical methods for Pb (and other metals) use between 0.5 and 2 grams of soil for the analysis, depending on the laboratory's standard operating procedure (SOP). So the analytical sample is much, much smaller than the mass of soil in the jar.

  For this study, a large soil sample was taken from a firing range with Pb contamination. The soil sample was dried and clods were broken apart, but no grinding was performed. Visible fragments of Pb bullets were removed. The soil was then sieved into different-size fractions. The 6 particle size fractions that resulted are provided above. Particle size gets smaller as the mesh size increases. Each particle-size fraction was analyzed for Pb separately by atomic absorption (AA), a routine laboratory method for analyzing metals.

An obvious trend exists for this site's soil: the Pb concentration in a particle size fraction increases as the particle size decreases. Why should this be? There are a few reasons. The smaller the particle size, the more surface area is available to adsorb contaminants like Pb and the smallest fraction is more likely to have particles made of clay minerals. Clay minerals carry a negative charge that attracts and holds on to positively-charged metal ions. Over time, contaminants "partition" into the soil constituents that have properties that attract them. There may also be very, very tiny particles of Pb released by the gun's firing mechanism, from impacts of bullets into hard surfaces (like rocks), and by slow decay of bullet fragments.

Particle size effects on analytical results have ramifications for the sampling and analysis of soil. When soil is shipped to a laboratory, motions in transit cause a segregation of particle sizes within jars. When a sample jar arrives at the laboratory, larger particles are typically sitting on top, and smaller particles have moved toward the bottom. If a technician were to sample a jar by unscrewing the cap and simply scooping a subsample off the top, the Pb result would likely be a lot lower than the true Pb concentration for the whole jar of soil.

As mentioned above, metals analysis for soil typically involves digesting a very small mass, around 1 gram. So another variable that can affect the concentration of the analytical sample (and thus the reported result) is the size and shape of the utensil used to weigh out the nominal 1-gram. A variety of utensils of varying sizes and configurations can be used to scoop up small amounts of soil and ferry it from the jar to the weigh boat that sits on the balance. There is no standardization of what utensil should be used. Even within the same laboratory, different technicians may use different scoops. A larger, spoon-shaped utensil will retain the larger particles (which provide mass, but little Pb), but those particles could easily roll off a flat spatula or a much smaller scoop. Thus a larger bowl-shaped utensil will select FOR larger particles, whereas a flat or very small scoop surface will select AGAINST larger particle sizes.

Another variable is related to the motions the technician makes while weighing out the analytical sample. Say the target mass for an analytical sample was 1 gram. Weighing out samples takes time, and technicians are always under pressure to maintain high sample throughput. So the fewer scoops into the weigh boat needed to get close to 1 gram, the more samples a technician can process. So naturally, the technician will make the 1st scoop out of the jar larger to try to get close to 1 gram without going significantly over. If it does go overweight, the soil must be dumped and weighing started over. Although the analytical sample doesn't need to be exactly 1 gram, it should be close. If a larger sized scoop was used and the amount of sample in the scoop looks larger than 1 gram, the technician may give the scoop a little shake to dump some of the larger particles back into the jar. This action selects AGAINST larger particles.

Now, say the 1st scoop of soil brought the balance to 0.7 g. Then a smaller volume (with even fewer large particles that might "tip the weight over") may be scooped into the weigh boat. Say that now the balance says 0.9 g. To get the mass closer to 1.0, the technician will likely gently tap the side of the scoop while holding it over the weight boat in order to knock smaller particles in a little at a

time. This action selects AGAINST larger, low-Pb content particles and preferentially adds smaller, high Pb-content particles.

These very common techniques are fine when weighing out materials that are truly homogenous and have a uniform particle size. But for soils, variable selection for and against various particle sizes in the analyzed subsample changes the result. These various weighing techniques may all occur in the same sample weighing, or only one or none may occur. The fact that these variables are not controlled in routine laboratory practice is part of the reason why split sample results can be very different, and explains why laboratory duplicates from the same jar often have poor precision.

This raises a question: What does a "representative sample" mean in the face of this kind of matrix heterogeneity? Should a "representative subsample" be representative of the jar contents? In other words, should large and small particle sizes in the subsample have the same proportion as in the jar? Or, should the subsample be "representative" of the question that needs to be answered. In other words, should the analytical sample be representative of the decision the data will be used to make? Is "representative of the jar" and "representative of the decision" always the same thing? The scientific answer is the answer to use if we want reliable risk assessments or remedial designs that work. The scientific answer is that "sample representativeness" cannot be determined until the "characteristics of interest" (as worded in ASTM's definition, recall slide 3-7) are known. The decision that the data are expected to answer is what determines the "characteristics of interest," and the "population" that the data need to represent.

How would a "representative analytical sample" be prepared? Which particle size is representative? If the decision about this soil site is a risk decision where the anticipated exposure pathway is dust blowing off the site and into residential areas, the smallest (dust-sized) particle sizes are representative of dust exposure decisions. The concentration of Pb in the dust sticking on little kids hands could be very high even though the average concentration for the BULK soil (which contains a range of larger sized particles) may be much lower.

On the other hand, it is easy to imagine a decision where the concentration for the bulk soil is the characteristic of interest. The toxicity characteristic (TCLP) is a standard test to decide is material is safe to landfill. Since the bulk soil is what would be going into the landfill, the TCLP should be done on bulk soil.

There may be times when the larger particle sizes are more representative of the decision. Soil "washing" has been around a long time as a remedial technique to clean metal contaminated soils. Soil washing works by carrying away the smaller particle sizes in a stream of water. The larger particle sizes left behind have lower concentrations. Knowing the above information allows an engineering design that can tailor the force of the water so that the particle size combination that renders a soil below action levels can be targeted. The concentration of residual small soil particles requiring special disposal also can be predicted.

"Representative of a decision" may or may not be the same thing as "representative of the parent bulk material." Selecting what "representativeness" means for a particular sampling event MUST be driven by what question the data are supposed to answer. If "representative of the bulk" is desired, Gy theory can be used to design a correct sampling scheme that can accomplish that. Although Gy theory is very important, it will not be covered in this course due to time constraints. Commercial courses are available that teach Gy theory.

Size conversions:

- 3/8" = 0.375 in. = 9.525 mm
- ASTM (US std) nominal aperture mesh size (mm):
  - 4-mesh = 4.76 mm
  - 10-mesh = 2 mm
  - 50-mesh = 0.297 mm = 0.3 mm
  - 200-mesh = 0.074 mm

Micro-scale Heterogeneity Causes Highly Variable Data Results for Replicates

Analytical Subsample Support (dried, ball-milled, sieved to <10-mesh)	Range of Results [for 20 individual subsamples (ppb)]	Coeff of Var. (CV)	Number of subsamples required to estimate true sample concentration within a range of…	
			…± 25%* [ex: 1930 ± 25% = 1448 - 2412 ppb]	…± 10%* [ex: 1930 ± 10% = 1737 - 2123 ppb]
1 gram	**1010 - 8000**	**0.79**	**39**	**240**

The true concentration for Am-241 in this large sample is known to be 1930 ppb.

Adapted from DOE (1978)

* At 95% confidence

3-12

Notes

- Adapted from Source: Doctor, P.G. and R.O. Gilbert. (1978) "Two Studies in Variability for Soil Concentrations: with Aliquot Size and with Distance." in "Selected Environmental Plutonium Research Reports of the Nevada Applied Ecology Group" (NAEG), Las Vegas, NV. Pp. 405-423 and 439-442. 95 percent confidence for +/- percent range.

  This study, performed in the mid-1970s, was designed to look at how data variability was related to the subsample volume run through the analysis procedure.

  A very large soil sample (in the range of a couple of a kilogram [kg]) was dried, ground and sieved to less than 10-mesh. Although this is not the most thorough sample preparation possible, it is far more intensive than is routinely performed by environmental laboratories. Using radiological analysis that can measure large volumes of material, the true americium-241 (Am-241) concentration for the entire large sample was found to be 1930 ppb.

  The study was done by taking 20 replicate analytical samples of different volumes and analyzing each. This slide summarizes the results for the 20 1-g subsamples that were taken from the large prepared sample. Each subsample was analyzed, and the statistics for the 20-sample data set were calculated. The range of results for the 1-gram series is provided in the 2nd column. A statistical expression of the variability among those 20 results, the "coefficient of variation" is provided in the 3rd column. [The coefficient of variation (CV) also is called the relative standard deviation (RSD).] Columns 4 and 5 convey the effect of data variability on data confidence.

For these 1-g subsamples, the CV was 0.79. That level of data variability means that there is very little confidence that a single sample result is representative of the true concentration of the original sample. Instead, to achieve some reasonable level of data confidence requires that multiple subsamples be analyzed and averaged. For example, look at the 4th column. It says that in order to estimate the true concentration (which is 1930 ppb) with an accuracy of +/- 25 percent when using 1-g subsamples, ***39*** replicate subsample analyses must be done and the results averaged. In other words, we can be 95 percent statistically confident that averaging 39 replicate analyses should give us a value that is within a 25 percent range around the true concentration. So for this sample, we know that the true concentration of the original large sample is 1930 ppb. Knowing that, we can expect that the mean of 39 analysis on that large sample will fall within the range of 1448 to 2412 ppb.

If we wanted to be sure our subsamples are giving us an even better estimate of the true concentration, we would analyze more subsamples. If we wanted to get within 10 percent of the true concentration, we would have to analyze 240 subsamples.

What does that mean for the interpretation of data results? What if we used only a single 1-gram subsample to make a decision about the actual concentration? In this case, one of the subsamples had a result of 8000 ppb. If we only ran 1 analysis on the 2-kg sample, and got 8000 ppb as the result AND IF we believed that number and used it to make a decision, we could made a serious decision error. And unless we know what the subsampling variability is, we have no idea how much we can trust the result from any single analysis. Laboratory duplicates (or even better, triplicates or better) are the means by which we can estimate the variability in the matrix we are sampling. We do not need to do replicate analyses on every single sample. Once we know the variability for a given area that holds the same "population" of soil, AND we know how close to the true concentration we want to be, we can tailor an analytical design that meets those needs.

The Smaller the Analytical Sample, the More Likely that the Result is Non-Representative

Subsample Support (dried, ball-milled, sieved to <10-mesh)	Range of Results [for 20 individual subsamples (ppb)]	Coeff of Var. (CV)	Number of subsamples req'd to estimate true sample concentration within a range of…	
			…± 25%* [ex: 1930 ± 25% = 1448 - 2412 ppb]	…± 10%* [ex: 1930 ± 10% = 1737 - 2123 ppb]
1 g	1010 - 8000	0.79	39	240
10 g	1360 - 3430	0.27	5	28
50 g	1550 - 2460	0.12	1	6
100 g	1700 - 2300	0.09	1	4

How much confidence should be placed in any single result?

3-13

Notes

- Adapted from Source: Doctor, P.G. and R.O. Gilbert. (1978). "Two Studies in Variability for Soil Concentrations: with Aliquot Size and with Distance." in "Selected Environmental Plutonium Research Reports of the Nevada Applied Ecology Group" (NAEG), Las Vegas, NV. Pp. 405-423 and 439-442. 95 percent confidence for +/- percent range.

  In addition to the experimental trials using 1-g subsamples, additional rounds of sampling were done with 20 each of 10-g, 50-g, and 100-g subsample masses. The trend is obvious: the larger the subsample taken, the lower the variability in replicate results (i.e., the lower the CV), and the fewer replicate analyses would be needed to achieve a desired degree of data confidence.

  Advancing Technology Issue: Advancing analytical technologies and instrumentation are using smaller and smaller subsample aliquots. This phenomenon is reducing data quality by reducing the representativeness of individual sample results.

  What might an analytical design that controls for data uncertainty look like? To do the design, we have to know what decision we want to make. Say our action level is 2000 ppb for a decision unit that is an acre of land area, to a depth of 1 foot (ft). Say also we desire to have 95 percent statistical confidence we are making the right decision relative to the 2000 ppb action level. Further assume that this decision unit can be sampled using a multi-increment (MI) sampling strategy such that a large MI sample is representative of the average concentration over the decision unit volume defined by 1-acre x 1 ft. Our decision about the decision unit rests on the result from that one MI sample. The question becomes, how do we control for variability within that MI sample so that

our analytical subsample(s) give us a reliable estimate of the true concentration of the MI sample? Since we know that the MI sample is representative of the decision unit, we can simply extrapolate the MI result to the 1-acre decision unit?

Say that we did previous work on adjacent 1-acre decision units and they all appear to be similar. From them, we found out that the variability within the MI sample for 1-gram subsamples (after using our sample preparation procedures to achieve at least some homogenization of the MI sample) was a CV of 0.79. Because of our previous experience, we expect a similar CV of the current MI sample. How could the data uncertainty created by the variability within the MI sample be reduced, say to +/- 10 percent?

Using the chart above, we see that for a CV of 0.79 and 95 percent desired confidence that our MI sample result is within 10 percent of the true concentration, we see we would need to run 240 subsamples!!! There is no way that is going to happen. So what are the alternatives?

1) We can do better sample preparation of the MI sample to reduce its heterogeneity. If the CV is lowered, we can achieve the same decision confidence goals with fewer subsamples. For example, if better MI sample homogenization could lower the CV to 0.27, we'd only need to analyze 28 subsamples to achieve the same statistical decision confidence.

2) We could analyze larger subsamples and reduce the CV.

3) We can lower our expectation for statistical confidence (say to 80 percent), which will reduce the number of 1-gram subsamples needed.

4) Or, Instead of analyzing 240 individual subsample, we could get the same statistical power by taking those 240 subsamples (collected from throughout the whole MI sample collected from the decision unit), and pooling them into their own MI sample(s). Of course, then we would need to subsample that MI sample because pooling 240 1-gram subsamples creates a sample of 240 grams. The process of creating and mixing the new MI sample should produce better homogenization, which lowers its CV. If the CV could be reduced to 0.12, only 6 1-gram subsamples need be taken and analyzed. Having 6 subsamples analyzed allows for statistical evaluation of the data set, and determination of a confidence interval around the mean of those 6 data points.

A combination of several or all of these options can be wisely used to reduce the analytical and labor costs associated with following only 1 of the options.

Then that confidence interval is compared to the action level. Does the upper 95 percent confidence limit fall below 2000 ppb? Then you can decide that the analytical subsample was below 2000, and by extrapolation up the chain of representativeness, the true mean concentration across the entire 1-acre x 1 ft decision unit is below 2000 ppb (i.e., "clean").

The same principle applies to deciding if the site is "dirty." If the lower 95 percent confidence limit is greater than the action level, then there is 95 percent statistical confidence that the true mean for the decision unit is greater than the action level.

The problem comes when the 95 percent confidence interval straddles the action level. You cannot make a decision at 95 percent confidence. But perhaps the numbers work out that the respective confidence limit falls above or below the action level so that a clear decision can be made. So, if the expectation for decision confidence is lowered, you may be able to say that the site is "clean" or "dirty" at 80 percent confidence, which might be fine with the project stakeholders.

Data Uncertainty & Decision-Making

Result = 350

Mean ± 25% allowable uncertainty

If there is 25% allowable difference in the measurement system, you cannot claim that values between 263 and 438 are different

If action limit (AL) = 400, and you get 350 and have 25% variability in your system, can you confidently claim that the result is below the AL? **NO**

If AL = 500, and you get 350, can you claim result < AL? **YES**

3-14

Notes

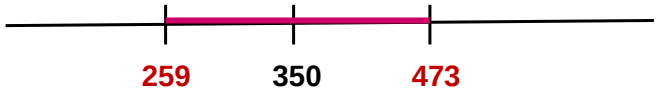
- When an acceptable uncertainty is set, such as +/- 25 percent, the meaning of that acceptable level of data uncertainty is that the measurement system (which includes sample collection and handling steps) is not expected to perform any better than +/- 25 percent. That means that each analytical result is bounded by an uncertainty interval of +25 percent on the upper side of the result, and an uncertainty interval of -25 percent on the lower side. By definition, within those bounds there is uncertainty as to whether the measurement system (which includes all steps from sample collection to instrumental analysis) can really "see" a difference in concentrations. In other words, although the result was reported as 350, the actual value may be anywhere between 263 & 438. If there is 25 percent uncertainty, 350 and 270 and 400 all look the same to the measurement system. Decision confidence is possible only if decision thresholds fall outside the data point's uncertainty interval.

  So what is data quality? Data quality is the suitability of the data for its intended use. The ability to use data to make decisions is a function of BOTH the uncertainty in the data AND the requirements of the decision it is being used for. Say your intended use of the data result of 350 is to decide if it is higher or lower than 400. If that data point has 25 percent uncertainty, you cannot use that data to confidently make that particular decision. (Of course, if decision confidence was not an issue, environmental decisions could be made by flipping a coin.) However, 25 percent uncertainty is ok for a result of 350 is the threshold it is compared to is 500.

QC checks serve to verify that the measurement system is performing no worse than the allowable level of data uncertainty. QC checks also may show that the measurement system is performing better than expected, which means that the widths of the uncertainty interval around a reported data point can be decreased. For example, if it is shown that the measurement system is precise enough so that the actual uncertainty is +/- 10 percent, then the uncertainty range around 350 is 315 – 385. That improvement in precision improves the usability of the data for making the decision at the 400 threshold. The reduction in data uncertainty is less important for decision making at the 500 action level.

The Relevance of QC to Decision-Making

Soil Result = 350 ppm
Lab Duplicate Result is allowed to be ± 30% RPD



This measurement system's acceptable "noise" means that values between 259 and 473 ppm cannot be distinguished as different from 350

If project decisions are to be made at an action level (AL) = 400 ppm and the permissible system noise is +/- 30% RPD, should you expect to decide that a result of 350 ppm is below the AL?

NO

3-15

Notes

- Note: RPD = relative percent difference. +/- X percent RPD is a different measure than +/- X percent. The equation used here for calculating RPD is the (1st value minus the 2nd value) divided by the average of the 2 values.

  An RPD of 25 percent produces a different uncertainty interval than just +/- a percentage.

  For example, the uncertainty interval around 100 with an RPD = 25 percent is 100 + 25 percent RPD = 129 at the upper end, and 100 - 25 percent RPD = 78 for the lower end. Thus the 25 percent RPD uncertainty interval around a result of 100 = 78 to 129. The asymmetry of the upper and lower intervals is due to the average function in the denominator of the RPD calculation.

  In contrast, the uncertainty interval designated as "+/- 25 percent" around 100 is 75 to 125.

  Whenever an acceptance limit for a QC check is set, such as +/- 25 percent RPD for lab duplicates, the meaning of that acceptance limit is that there is an acknowledgement that the measurement system (which includes sample handling steps) is not expected to be any more precise than +/- 25 percent RPD around some value.

Consider the measurement system referenced in this slide, where +/- 30percent RPD is the acceptance range for duplicates. Again, that means that the measurement system's performance is consistently expected to not see a difference between 259 and 350, or a difference between 350 and 473. For an action level of 400 ppm, and a result precision on soil samples of +/- 30 percent, values between 280 and 520 are considered equivalent to 400 ppm. Therefore, for this measurement system, a decision that a result is confidently less than 400 ppm requires that the reported result be less than 280 ppm. If it is desirable to show that a result of 350 ppm is less than an action level of 400 ppm, the measurement uncertainty must be reduced to at least 13 percent RPD. [((400-350)/375) x 100 = 13.3 percent RPD]

Something to Think About: Lab Data Are Considered to be "Definitive"

Are lab data meeting all the requirements for definitive data?

DQOs for Superfund guidance

"For the data to be definitive, either analytical or total measurement error must be determined." (p. 43)

3-16

Notes

- Is the measurement or analytical error of data being reported to project managers in a way that allows them to take data uncertainty into account when evaluating data against decision criteria?
- Data Quality Objectives Process for Superfund: Interim Final Guidance (September 1993)

Measurement Error, Data Variability & Sample Representativeness

- We need a concept that captures the fact that the volume of the matrix is a determinant of measured concentrations and data variability (measurement error)
- That concept is called "**support**"
- The term and its definition appear in the DQOs for *Superfund* guidance and other EPA guidance documents for the waste cleanup programs

3-17

Notes

- **We need a concept that captures the fact that the volume of the matrix is a determinant of measured concentrations and data variability (measurement error):** Fortunately, such a concept does exist!
- **That concept is called "support:"** Much of the remaining material in this module is dedicated to defining what "support" means.
- **The term and its definition appear in the Data Quality Objectives (DQO) for *Superfund* guidance and other EPA guidance for the waste cleanup programs:** The following guidance documents define "support:"
  - » Data Quality Objectives Process for Superfund: Interim Final Guidance (September 1993)
  - » Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples, EPA/600/R-03/027, November 2003
  - » Soil Sampling Quality Assurance User's Guide, 2nd Edition, EPA/600/8-89/046
  - » A Rationale for the Assessment of Errors in the Sampling of Soil, EPA/600/4-90/0-13

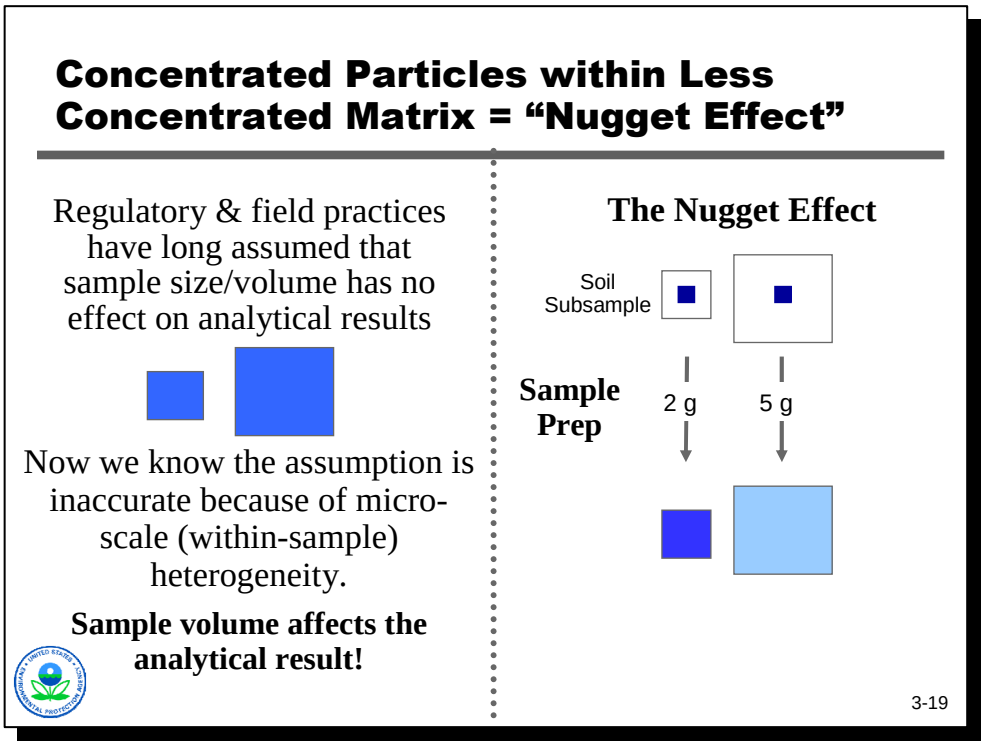
Sample Support

- **Sample Support** encompasses the physical properties of the sample that are relevant to the representativeness of the sample:
  - the **size** (mass or volume)
  - **shape** & **orientation** of a physical sample drawn from a matrix population (such as soil, sediment, or water)

3-18

Notes

- **Sample support:** Many people have not heard this term, but the concept is critical to sample representativeness and data quality. This is why it was mentioned in the "DQOs for Superfund" guidance. It was near the text that talked about analytical quality. The analytical quality piece caught on, the sample support piece did not. Yet it was recognized that sample support was an important factor that goes into "data quality."

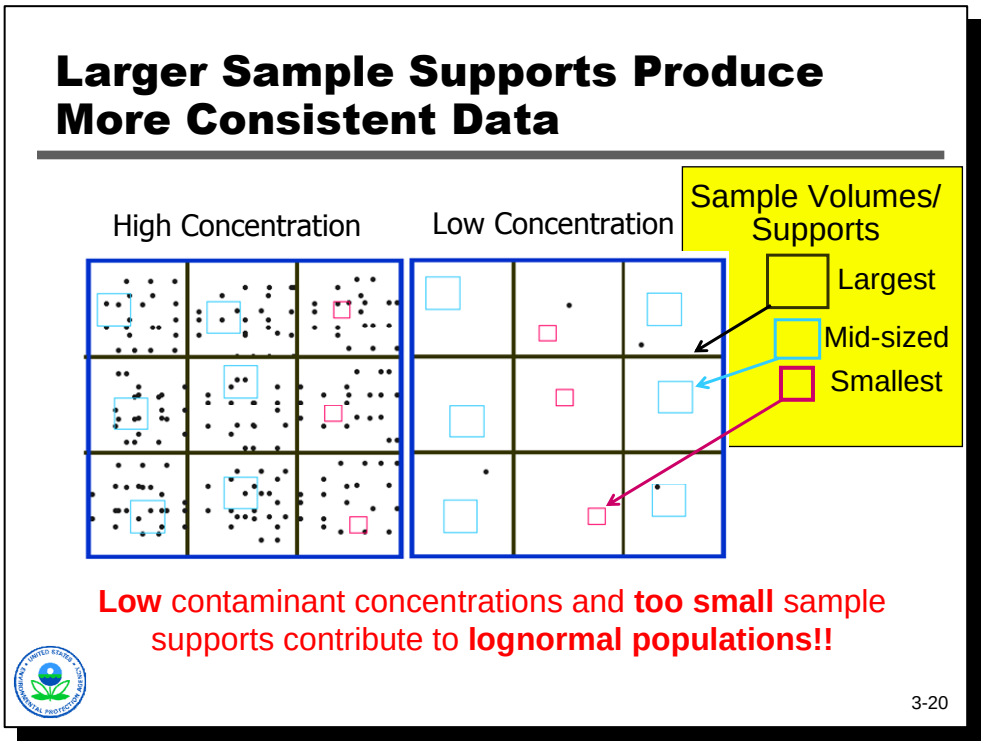


Notes

- The issue of "sample support" for heterogeneous environmental and waste matrices should make us reconsider the common (and usually unacknowledged) assumption that the reported concentration of an environmental sample should be the same no matter what volume of sample is collected.

- The volume of the sample is an important factor that influences the reported concentration for a sample, especially when contaminants are heterogeneously distributed throughout the parent matrix. All samples must be homogenized (through physical or chemical means) prior to analysis. For heterogeneous samples (that are affected by the nugget effect to a lesser or greater degree), the analytical result for a sample is determined by how much contaminant is captured in that sample, and how much cleaner matrix is contained in the sample (that serves to dilute the contaminant during homogenization). The nature of contaminant release to the environment (such as release to ground surface in the form of a powder or particulate) increases the probability of heterogeneity, as does contaminant solubility, mobility, and the age of the release. Obviously, environmental variables (such as precipitation, wind erosion, temperature, matrix composition) interact with the contaminants' properties to mitigate or aggravate heterogeneity. Contaminants that may at first have been more homogeneously released may become heterogeneously distributed throughout a matrix over time if their chemical properties cause them to preferentially partition onto mineral surfaces or into organic carbon that are themselves heterogeneously distributed, or inclusion of those matrix components into the analytical sample is variable or unpredictable.

- The issue of sample support is becoming an increasingly important determinant of analytical result as more sophisticated analytical technologies require smaller and smaller volumes of sample. At one extreme, sensors technologies currently under development will have miniscule sample supports, and data interpretation will be extremely difficult unless there is much greater awareness and management of sample support concepts.



Notes

- The nugget effect can occur when contamination occurs in particulate form (such as explosives residues deposited as a powder or lead fragments in a firing range), or when contaminants partition onto mineral surfaces or organic carbon which are themselves heterogeneously distributed. Gy theory relates the size of the matrix particles to the sample support mass needed to representative the larger matrix volume. The volume of the sample is an important factor that influences the reported concentration for the sample, especially when contaminants are heterogeneously distributed throughout the parent matrix.

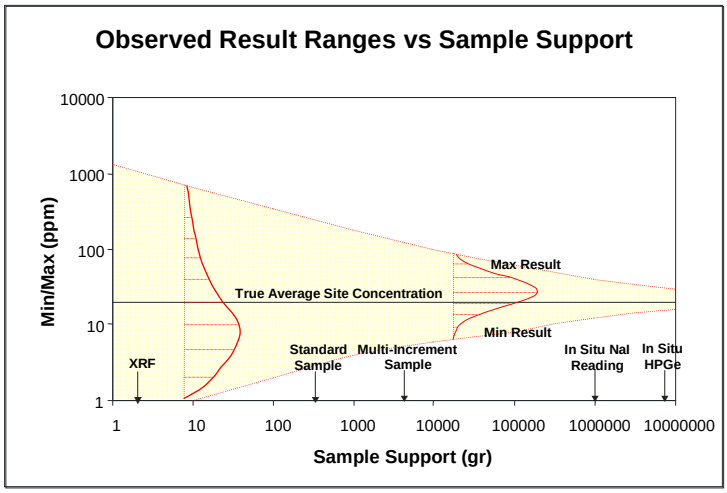
- Three different color-coded sample supports are illustrated in this figure. From largest to smallest sample support, the colors are black, light blue, and red. The dark dots ("particles") represent higher contaminated small particles in a matrix of "cleaner" particles (not shown in the figure). The figure depicts the variable capture rates of the "dirty" particles for higher and lower contaminant concentrations and for different sample supports (volumes).

- Since smaller sample supports have a lower capture rate of contaminated particles, there is a higher rate of non-detects. On the other hand, when a contaminated particle is captured, the low volume of cleaner matrix causes the concentration to be higher after sample prep. These two factors, a high number of non-detects/very low concentration samples, and very high concentrations in a few samples, produce the statistical lognormal frequency distribution common to environmental sampling.

- Specifying a regulatory threshold without specifying the sample support over which it applies (or at least recognizing that differences in sample support introduce variability into analytical data results) easily leads to widely different analytical results from one sample to the next. Since sample support is generally ignored in regulation, it is ignored in practice and the sample support is left to chance. This leads to uncontrolled (and usually undocumented) variations in sampling conditions and often widely varying results that are difficult to interpret.

  Unless the laboratory was in charge of field sampling and was involved in project planning and sampling analysis plan (SAP) and quality assurance project plan (QAPP) preparation, the laboratory cannot be held accountable for such variable results. The analytical result is probably correct from the standpoint of generating an accurate result on the analytical sample. Project planning was faulty for not ensuring that sample collection and handling procedures would produce samples representative of the decision.

Another Way to View Data Uncertainty as a Function of Sample Support



3-21

Notes

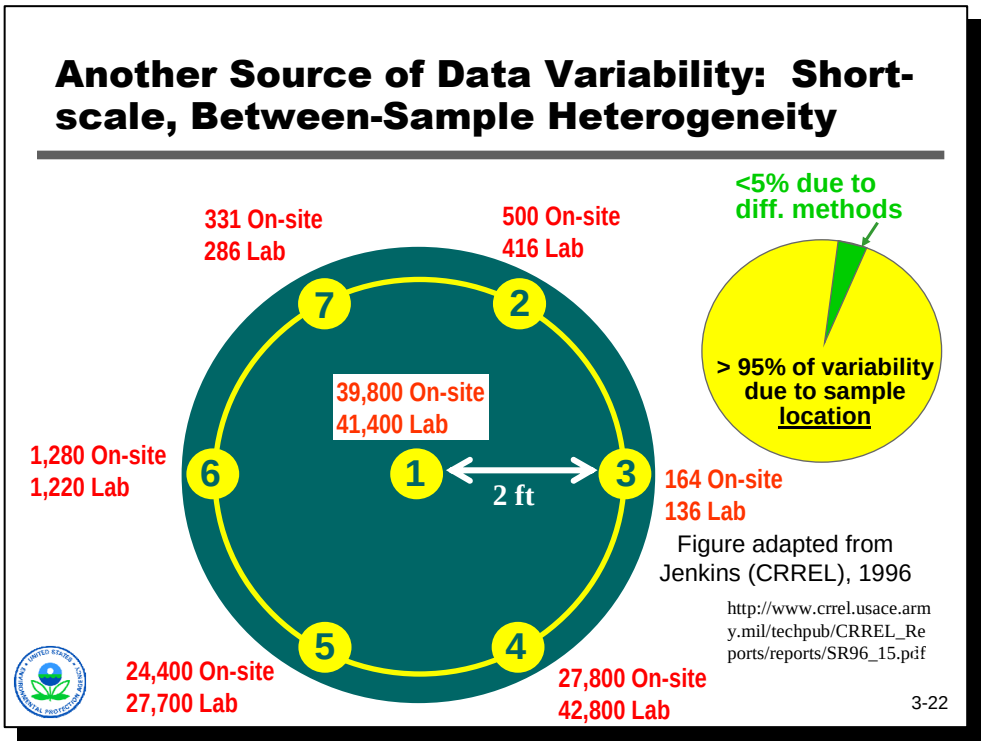
- This graph summarizes a study that looked at the range of results generated for different sample supports. This study is similar to the Am-241 study discussed earlier in this module. The analytical methods suitable for this analyte (uranium) have sample supports that range from 10,000 kg soil down to a couple of grams of soil. The larger the sample support, the more consistent multiple sample analyses are, and the closer each result is to the true mean. In contrast, very small sample supports create highly variable results. Many of the results are very far from the true mean.

- The superimposed frequency distributions illustrate how the lognormal distribution (the one on the left) is typical of small sample supports, whereas a more normal-shaped frequency distribution (the one on the right) is produced when larger sample supports are used. This directly connects to the previous two slides illustrating how smaller sample supports have a lower capture rate of contaminated particles, and a higher number of non-detect/very low values. The very high concentrations seen with small sample supports are due to the bias created when contaminated particles are captured within a small volume of cleaner matrix.

- It is important to know that the term “sample support” assumes that the sample volume is completely homogenized before any analytical subsampling is performed. If the sample volume is not appropriately homogenized prior to subsampling, then the volume of the analytical subsample is the actual sample support. On the graph, the sample support masses denoted as ”standard sample” and “multi-increment sample” require that the full sample mass be completely homogenized before analytical sampling.

- The two analytical methods on the far right are in situ, walk-over detection technologies. The sample support reported for each is the volume of soil the detector/sensor is able to "see" all at once during a single analytical measurement.

- Abbreviations on slide: XRF = X-ray fluorescence; NaI = sodium iodide detector; HPGe = high performance germanium detector.



Notes

- This figure shows how spatial heterogeneity can be measured on a short spatial scale of several feet. The heterogeneity causes sample concentrations to differ by at least two orders of magnitude.

  This study examined the relative variability introduced in sample results by two different analytical techniques (in this case a standard laboratory analysis and a field analytical method) and by the short-scale spatial variability actually present in contamination concentrations (in this case explosive residues). The protocol selected seven samples from a 4-foot diameter circle. Those samples were split and analyzed two different ways. The larger graphic displays the observed results. Based on an analysis of variance, the smaller graphic apportions the observed variability in the complete data set (field analytics and laboratory data combined) between that contributed by differences in results produced by the two different techniques, and that contributed by spatial heterogeneity in this small area. The conclusion: the variability associated with spatial heterogeneity was 20 times greater than that associated with the analytical methods.

Internet

Source: Example of characterizing sampling variability from Tom Jenkins, USACE Cold Regions Research and Engineering Laboratory's (CRREL) various reports at *http://www.crrel.usace.army.mil/techpub/CRREL_Reports/html_files/Cat_X.html*].

This example is from the Monite installation, which is contaminated with explosives residues (the facility reclaimed explosives from out-of-date munitions). The analyte in this figure is TNT.

How to Pick the Right Sample Support

- **Sample Support** – the **size** (mass or volume), **shape** and **orientation** of the physical sample taken to ***represent*** a specific population of interest.
- Need to know what the population of interest is
- Need another concept: Decision Unit

3-23

Notes

Sample support and other aspects of a sampling design must be developed on a site-specific basis.

- **Sample support:** The definition of sample support includes the population of interest.
- **Need to know what the population of interest is:** In order to select the appropriate sample support, the population of interest must be defined.
- **Need another concept:** The population of interest is defined by the decision to be made and includes the spatial area over which or the objects to which the decision will apply. This concept is called the decision unit (DU).

A Fundamental Concept for Ensuring Representativeness

- **Decision Unit:** An area, a volume, or a set of objects (e.g., ¼-acre area, bin of soil, set of drums) that is treated as a single unit when making decisions
  - »The decision unit may be a single item (such as a volume of soil)
  - »Or a decision unit may be a group of items united by a common property
  - »Examples: exposure unit, survey unit, remediation unit …

3-24

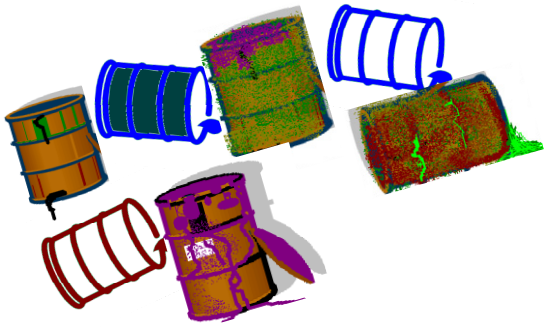
Notes

- **Decision unit:** A DU is an area or set of objects for which a decision needs to be made. The decision could be whether contamination is present in the unit or not, and if it is, whether it is at levels that exceed cleanup requirements.

  Examples of single decision units include a quarter acre, a city block, or a set of storage drums. Exposure units, survey units, or remediation units are all examples of types of decision units. Decision units can have a temporal component as well as a spatial dimension.

  When there is a decision of whether the concentration of an area is greater or less than some threshold, the DU is the unit over which the average concentration [as represented conservatively by the upper confidence interval (UCL)] for the area (or volume) is calculated and used to compare to the threshold.

Drum Example #1



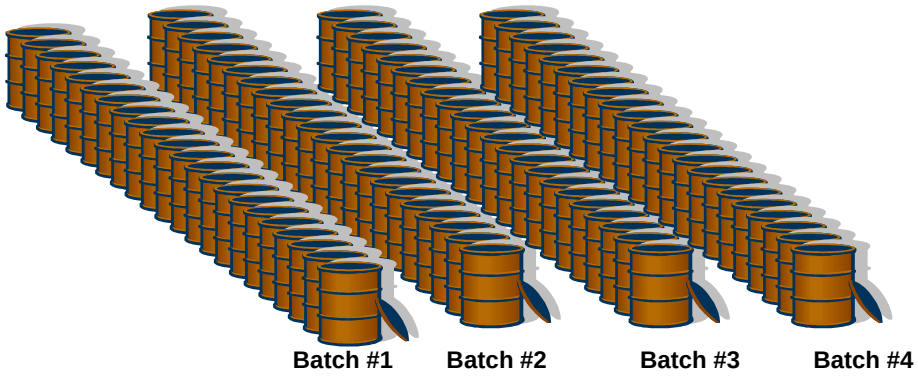
Valley of the Drums: These need to be characterized, transported, and disposed properly.
What is the decision unit? How do you sample it?

3-25

Notes

- The primary question is: "Can the results of a single drum be extrapolated to "represent" the contents and condition of the rest of the drums." The answer is "no." Each drum is likely different. Each may pose its own hazards, whether chemical (toxicity or burns) or physical (leakage, explosions or fire). Therefore each drum must be examined and sampled to decide how to handle each drum safely.

Drum Example #2



80 drums were cleaned in batches of 20. You need to ensure the cleaning process worked.
What is the decision unit and how would you sample it?

3-26

Notes

- Since a batch of similar drums are cleaned at the same time, it can be assumed that a batch is a single population because all members of the population have been subjected to the cleaning: same detergent strength, same sprayer force, same time periods for detergent exposure and rinsing, same heat temperature and time period. Since all members of a population can be considered as equivalent, random statistical sampling of each batch is appropriate.

  Statistical sampling means that a random selection and testing of a subset of the population is sufficient to draw conclusions about the whole population. The batch is the DU because the statistics are used to draw conclusions about the whole batch. The decision about the batch is made by extrapolating the results of the statistical sample decision to the entire batch. If the statistical sample fails inspection, the whole is considered to have failed. If it is valid to extrapolate the results of a subset (i.e., the statistical sample), then the statistical sample is said to "represent" or "be representative" of the whole population.

What is the Relationship Between Decision Unit and "Population?"

- **Population:** Set of objects or a volume of material sharing a common characteristic; can be synonymous with DU.
- Examples where they are not synonymous:
  - 2 "populations" exist within a single DU:
    - "clean" and "dirty" soil areas within a residential yard
      - A population is large enough so that more than 1 exposure unit is needed to cover it
    - 50 acres are suspected to be clean; but the DU (exposure unit) is 1 acre. So there are 50 DUs in that "suspected clean" population

3-27

Notes

- **Population:** A population is a set of objects or an area that shares some common characteristic. The way we will use population is often synonymous with decision units, however it need not be. A population might, in fact, be divided into several distinct decision units.

- A population distribution refers to the distribution (think histogram) of some population parameter that is of particular interest (e.g., contamination concentration). For example, the population might be everyone who lives in the Chicago metro region, the parameter of interest could be their height, and the population distribution for height would refer to the distribution of height present in Chicago's population.

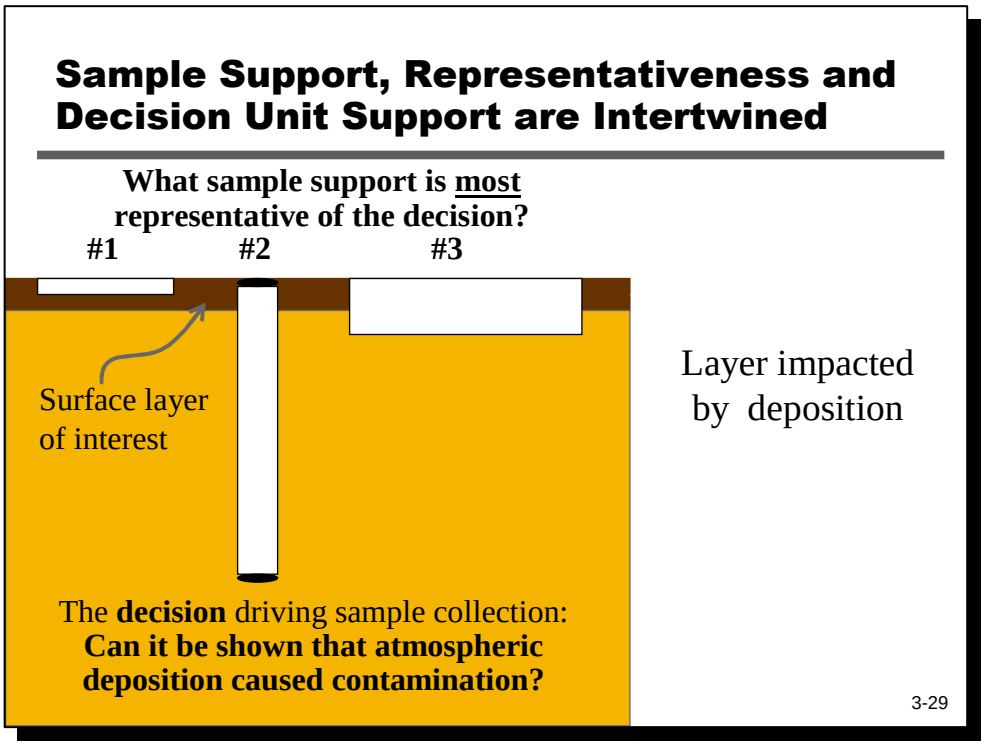
Sampling Design Requires a Progression of Supports

- Decision Unit Support – the spatial dimensions, mass, particle size or other physical properties that characterize the population of interest
  - »the decision DEFINES the population of interest, and
  - »the population DEFINES the properties of samples
- The sample support must mimic (on a small spatial scale) the decision unit support (on the larger spatial scale of the decision)

3-28

Notes

- **Sampling design requires a progression of supports:** Sampling program design starts "high" (at the intended decision) and works downward to representative sample support and analytics.



Notes

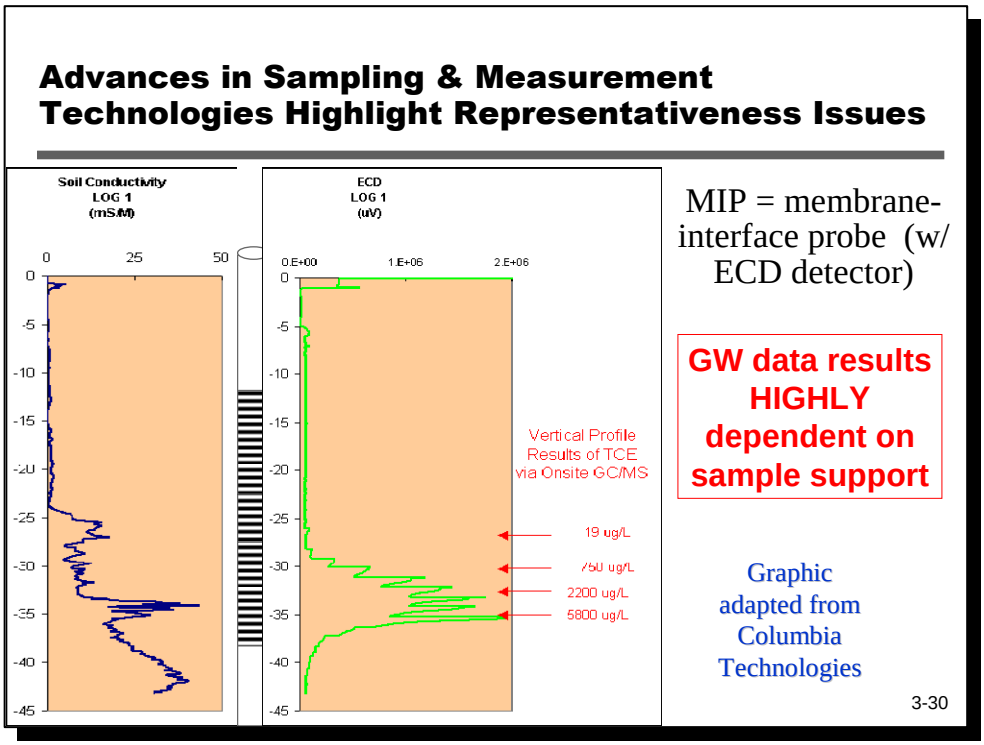
- This series of slides will illustrate concepts related to "sample support." These concepts are presented in a simplified form and do not attempt to portray the more exacting aspects of this topic.
- The panel on the left illustrates how sample volume and orientation must be selected to be representative of the decision to be made. Any of the 3 samples might be argued to represent true site conditions, but only one can be argued to be representative of site conditions in the context of the decision (atmospheric deposition).

  Color Key for left panel:

  - » Dark brown depicts surface soil impacted by surface deposition of lead from the atmosphere.
  - » Light brown depicts soil that would not be expected to be impacted by this atmospheric deposition.
  - » White areas depict the volume and orientation of material removed that becomes the "sample."

  Keep in mind that the entire sample is homogenized prior to subsampling for analysis.

The sample support (the physical dimensions of the sample) for Sample #1 would be representative of the matrix impacted by atmospheric deposition, but the sample supports of samples #2 and #3 would not be. Sample support #3 illustrates the importance of strict control over sample support in scenarios where careful stratification of populations is required to avoid biasing results by including non-representative sample. Even though the general orientation of sample collection in #3 is similar to #1, the concentration of lead in sample #3 would be expected to be "diluted" by the inclusion of "cleaner" soil from a non-representative layer into the sample.



Notes

- This slide shows the results from a direct-push membrane interface probe (MIP)-ECD taking readings every 2 inches going down to create a vertical profile of contamination in the subsurface.

  Soil conductivity results suggest transitions from sandy matrix to clay matrix (higher conductivity in clayey soil). The 7- to 8-ft wide band of contamination is associated with a clay layer in the subsurface. Small, discrete ground water samples (i.e., very small sample support) representative of point concentrations were collected using the direct-push (DP) probe and analyzed using gas chromatography-mass spectrometry (GC-MS).

  What analytical results (low, medium, or high laboratory results) would be expected if a monitoring well were screened over the various intervals depicted in the slide? (Keep in mind that clay layers may be rather non-permeable to water flow as compared to sandier layers.)

The Biggest Cause of Misleading Data

3-31

Notes

- We may try to pretend that we can sample soil and ground water as if it were homogeneous … as if the size of decision units and sample support did not matter. But Mother Nature is not required to go along with our self-deception. No matter how accurate an analytical method may be, getting the right result on a non-representative sample still will give the wrong answer. And we will waste resources cleaning up matrix that does not need cleaning up, calling something clean when it actually needs remediation, designing remedial systems that do not work, and making erroneous decisions about risk.

Sample Support is the Foundation of Sample Representativeness

- Sample support is critical to data quality
- Controlling sample support helps reduce the effects of micro-scale (within-sample) and short-scale (between-samples) matrix heterogeneity
- Sample support MUST ABSOLUTELY be controlled when splitting samples to assess analytical method comparability

3-32

Notes

- **Sample support is critical to data quality:** Sample support is the defined physical properties of the sample that are relevant to the representativeness of the sample, such as the size (mass or volume), shape, and orientation of a physical sample drawn from a matrix population (such as soil, sediment, or water).

- **Controlling sample support helps reduce the effects of micro-scale (within-sample) and short-scale (between-samples) matrix heterogeneity:** By controlling sample support, the heterogeneities associated with within-sample variability and between-sample variability can be reduced. Increasing sample volume and sample preparation (such as grinding to a homogenous particle size) reduces the effects of heterogeneity. Collocated samples provide a measure of the degree of short-scale matrix heterogeneity. The distance between collocated samples should always be stated in sampling plans and project reports. The results of collocated samples should be used to help determine the largest source of data uncertainty. If better control over uncertainty is required to make confident decisions, collocated information can be used to determine how and where additional sampling and analyses should be performed.

- **Sample support MUST ABSOLUTELY be controlled when splitting samples to assess analytical method comparability:** If micro-heterogeneity is not controlled when a single sample is split between two laboratories or between two analytical methods, the two splits may actually be different, and accurate analysis will show they are different. Poor matches between field duplicates, laboratory duplicates, and split samples are usually an indication that there has been poor control over matrix heterogeneity during sample handling.

Review of "Supports"

- Decision Unit Support
- Sample Support
- Measurement Support

3-33

Notes

- **Decision unit support:** Decision unit support identifies the spatial dimensions and other physical properties (such as particle size) that define the population of interest targeted by the decision.
- **Sample support:** Sample support encompasses the spatial dimensions and other physical properties (such as particle size) of a physical sample; it needs to be selected to mirror the decision unit support.
- **Measurement support:** Measurement support is the choice of analytical sample preparation (laboratory subsampling and digestion/extraction prior to instrumental measurement) that determines how much of the original sample content is actually "seen" and measured by the instrument.

Question:
Why do we need to understand all this?

Answer

Because XRF analytical programs perform much, **much, MUCH** better when XRF sampling and analysis designs apply this knowledge

Let's talk about the supports relevant to field-portable XRF instrumentation

3-34

XRF's "Supports"

- Analytical Sample Support
  - » The X-rays pass thru a window that is only a few cm2 in area
    - — They penetrate 1 - 2 mm depth into the soil surface
    - — The instrument "sees" ~ 2 g or less of a thin soil layer
  - » Soil particle size and its correlation to Me concentration determines how much Me is in the XRF's field of view
- Measurement support (same as analytical sample support)
  - » XRF measures the total element content of the soil volume for analytes it sees

3-35

Notes

- **Analytical sample support:** In contrast to the laboratory sample, the XRF measures about a 2 square centimeter area at a depth of a few millimeters, which represents about 2 grams or less of a thin soil layer. This is the XRF's analytical sample orientation. Soil particle size and its correlation to Me concentration determines how much Me is in the XRF's field of view.
- **Measurement support:** The XRF measures the total element content of the soil volume for analytes it sees.

Measurement Support: XRF vs. ICP

- ICP analysis requires digestion of soil mineral structure to free metals into solution
  - » The common nitric acid digestion does not free all metal present in the sample. How much is solubilized depends on soil mineralogy, the metal species released and how long ago, the redox and pH soil environment over the years…
  - » Hydrofluoric acid digests all minerals so total metal is measured by ICP analysis - but it is hard to find environmental labs with HF capability
- XRF directly measures total metals
  - » Results more comparable to HF digestion

3-36

Notes

- **ICP analysis requires digestion of soil mineral structure to free metals in solution:** Two digestion methods are available, the common nitric acid (HNO_3) method and the hydrofluoric acid (HF) method. Nitric acid does not free all metal present in the sample. The amount of metal solubilized depends on several factors including, soil mineralogy, the metal species released, the age of the release, the redox of the soil environment, and the pH of the soil environment. HF digests all minerals so total metal is measured by ICP analysis. However, it is difficult to find environmental laboratories with HF digestion capability.

- **XRF directly measures total metals:** The results of the XRF are more comparable to HF digestion than nitric acid digestion.

Controlling XRF Sample Support

- XRF's small sample support (1-2 grams, on the order of a lab subsample) makes it susceptible to non-representative readings (i.e., readings that are very different from the average concentration of the sample)
- The more uniform the distribution of soil particles, the more precise the XRF readings
- Uniformity depends on the adequacy of sample handling and preparation

3-37

Notes

- **XRF's small sample support makes it susceptible to non-representative readings:** Because the XRF "sees" only 1 – 2 grams of sample, it is susceptible to readings that are very different from the average concentration of the sample. Since ICP analysis also uses about 1 gram of soil for digestion and metals analysis, ICP data also are at risk of being non-representative.
- **The more uniform the distribution of soil particles, the more precise the XRF readings:** The precision of the XRF is controlled by how uniformly the element of interest and soil particles are distributed within the XRF's field of view. The more uniform the distribution, the more precise XRF measurements will be.
- **Uniformity depends on the adequacy of sample handling and preparation:** Because of inherent heterogeneity, uniformity is a function of sample preparation. The greater the sample preparation the greater the uniformity.

XRF Sample Handling/Preparation

- There are 3 main types of sample preparation for XRF—the effectiveness of all is influenced by operator effort and consistency
- Sample handling procedures should be defined in detail in the QAPP and followed meticulously
  - » Prepare soil for *in situ* "shots"
  - » Taking shots (readings) over a bagged sample
  - » Taking shots on a cup containing highly prepared soil

3-38

Notes

- **There are 3 types of sample preparation for XRF – the effectiveness of all is influenced by operator effort and consistency:** There are three types of sample preparation available for use with the XRF. Each can be performed in ways that degrade or enhance XRF performance. For example, the more care put into preparing the soil for an *in situ* shot, the more precise that shot will be.

- **Procedures defined in QAPP and followed meticulously:** The procedures for sample preparation and XRF use (including calibration procedures) should be described in detail in the QAPP and must be followed meticulously by the field team. There will be different procedures for the three types of XRF sample preparation, which are:

  - » Prepare soil for *in situ* "shots" – this procedure requires the least sample preparation

  - » Taking shots (readings) over a bagged sample – this procedure requires more sample preparation than the *in situ* method, but less than the cup method

  - » Taking shots on a cup containing highly prepared soil – this procedure requires the most sample preparation

Sample Preparation Option 1: *in situ*

- Grades of sample preparation for *in situ* "shots"
  - » Least preparation: "shoot" on bare ground with minimal debris removal and smoothing
  - » Most preparation: Loosen soil to selected depth of interest. Carefully pick out extraneous material. Crush and mix soil in situ until uniform appearance. Smooth/compress surface before placing XRF.
  - » Take several shots in the same prepared location. Re-position the XRF between shots to estimate matrix variability (evaluate the adequacy of sample prep)

3-39

Notes

- **Grades of sample preparation for *in situ* "shots:"** There are different levels of sample preparation for *in situ* use of the XRF:
  - » Least preparation – The least sample preparation involves "shooting" on bare ground with minimal debris removal and smoothing.
  - » Most preparation – The most sample preparation involves loosening the soil to a selected depth of interest. Extraneous material is then carefully picked out of the loosened soil. The loosened soil is then crushed and mixed *in situ* until it is uniform. The uniform soil is then smoothed and compressed before placing the XRF.
  - » Multiple shots – Several shots can also be taken for a single sample area, with repositioning of the XRF between the shots, to estimate sample variability and to evaluate the adequacy of sample preparation.

Sample Preparation Option 2: Bagged

- Bagged samples
  - » Increases sample support compared to in situ shots, especially when multiple shots are taken per bag
  - » Remove extraneous material. If necessary, crush soil before placing into bag (crushing a hard soil in a bag can damage the smoothness of the plastic)
  - » Mix bag by kneading (also breaks up aggregates) and/or turn bag end-over-end. Visually inspect to ensure uniform appearance. Do not just shake bag—will cause particle segregation and increase data variability
  - » Do not shoot through significant dimples or creases in the plastic—can cause increased reading variability

3-40

Notes

- **Bagged samples:** Before using a new batch of plastic bags, always check them to make sure the plastic itself will not cause interference. It is known to have happened and it causes the XRF results to be biased low. The plastic can be easily checked using the National Institute of Standards and Technology (NIST) reference samples (which are used as control samples). Take a few shots of the NIST standards you are using (all of them…the amount of bias differs with concentration) to get a good idea of the XRF average result for that cup. Then place the plastic bag over the cup. Shoot the XRF through the 2 layers of plastic of a closed bag. [Although you only shoot through a single layer during soil analysis, the interference effect, if it is there, will be easier to detect through 2 layers.] Take a few shots and get an average and the compare. Is there a significant difference? One way to quickly estimate whether there might be a significant difference is to see if ALL of the XRF readings through the bag are lower than the lowest reading you got without the bag. If the data are suggestive of a difference, but you are not quite sure, use statistics (a 2-sample t-test, which detects whether there is a difference between 2 means).

  Even a small bag-interference effect can cause a problem if it combines with another complication, such as soil moisture that is on the high side (between 10 and 20 percent moisture). The combination of the 2 interferences has been shown to cause XRF data to be significantly biased low when compared to ICP samples from the bag.

Notes

- The bags do not always need to be this large.

Notes

- The original XRF instrument design (the “brick”).

Shooting Bagged Samples

3-43

Notes

- This slide shows a newer XRF instrument design.

Benefits of Bagged Samples

- When sampled from 1 sampling location, bagged samples increase the sample support from that location (this strategy controls within-sample variability)
- MIS can be performed: Placing increments from across the decision unit into 1 bag further increases the sample support and is more representative of the decision unit average (this controls for short-scale matrix variability)
- Taking multiple shots over the bag estimates the degree of within-sample variability (particle size effects) and the statistical average of those readings provides a better estimate of the true concentration for the bag

3-44

Notes

- **Benefits of bagged samples:** Obviously, the larger the sample support, the greater the challenge of proper sample homogenization. The question is how much sample preparation is enough. For samples with low background levels of metals, one would expect those metals to be fairly uniformly distributed throughout the media. Consequently probably not much preparation is needed to get a prepared sample mass that would yield nearly identical subsamples for analysis. On the other hand, once one has heavily contaminated material, the story would be different, and significantly more sample preparation might be required to get the same repeatability in subsampling results.

Benefits of Bagged Samples

- Sample bagging and readings can be performed in real-time
  - » The number of shots over the bag can be adjusted in real-time (either up or down) to provide statistically valid results
  - » Within-bag variability that is too high can be addressed by additional kneading or other corrective actions, or by examination of bag (ex: paint chips?)
  - » Inexpensive
  - » Supports dynamic work strategy for field work by providing data of known and documented quality

3-45

Notes

- **Sample bagging and readings can be performed in real-time:** The XRF is almost unique (if not unique) in providing the ability to non-destructively and quickly evaluate the efficacy of sample preparation by multiple *in situ* readings across a sample's surface. The XRF also provides the possibility of substituting multiple readings across a soil's surface (either through a bagged sample's walls, or across a sample spread on a work area) for sample preparation when trying to obtain an accurate assessment of contaminant levels within the sample. Recall from earlier slides that you can get the same statistical confidence by increasing n (i.e., increasing the number of XRF shots going into calculating an average reading for the bag) or by decreasing the SD by improving sample homogeneity/uniformity.

  The XRF is unique in providing the ability to inexpensively (no consumables, just labor time), non-destructively and very rapidly evaluate the adequacy of sample preparation by taking multiple readings across a sample's surface (either *in situ* across a work area, or over a bag). Therefore the XRF provides the ability to use multiple readings, rather than more intensive sample preparation, to reduce data uncertainty and increase accuracy.

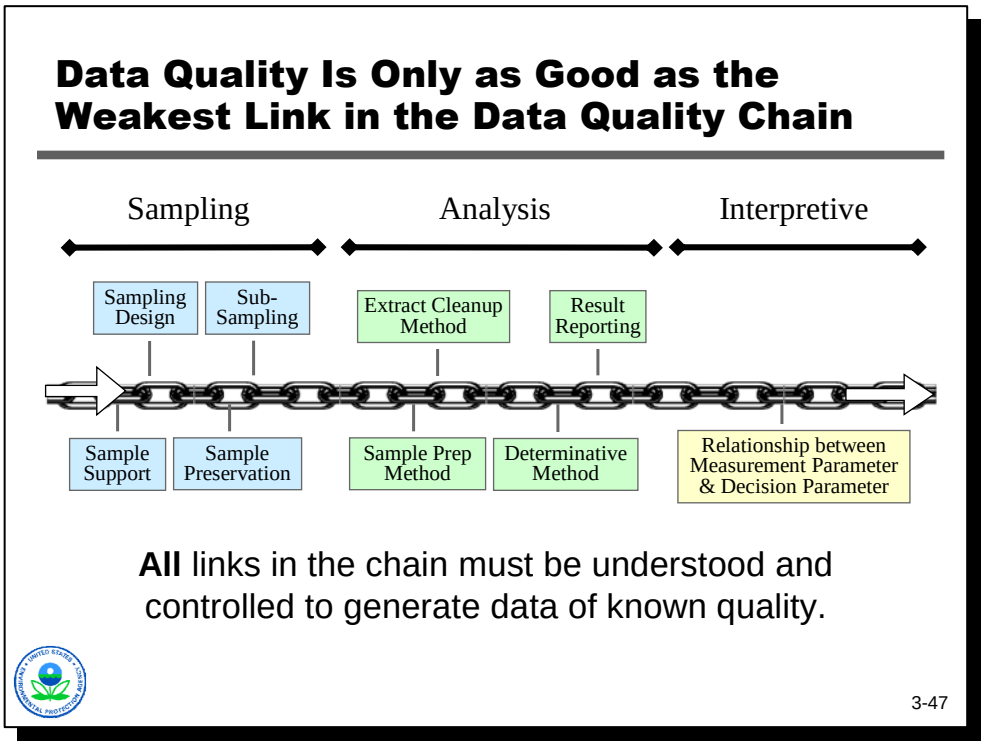
Sample Preparation Option 3: XRF Cups

- Cup samples
  - » Remove debris, dry, grind, and sieve to achieve uniform particle size
  - » Subsample properly and place into XRF cups
  - » Be CAREFUL tapping cups if particle size not completely uniform—will segregate fines
- Highest homogeneity = best precision
  - » Best for determining method comparability to ICP or AA
  - » Perform XRF and ICP/AA analysis on same cup

3-46

Notes

- **Cup samples:**
  - » Remove debris, dry, grind and sieve sample to achieve uniform particle size – Achieving uniform particle size is very important to increasing precision.
  - » Subsample properly and place subsamples into XRF cups – The subsample should be representative of the whole sample and each particle should have an equal chance of being included in the subsamples. EPA's 2003, "Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples," EPA/600/R-03/027, ranks several laboratory subsampling methods that could also be used in the field.
  - » Be CAREFUL tapping cups if particle size is not uniform – Too much tapping of subsamples that do not have a uniform particle size may cause partitioning of the various particle sizes and affect the ability of the XRF to "see" the total amount of the element of interest.
- **Highest homogeneity = best precision:** The cup method is the best method for determining comparability to ICP or AA. When conducting comparability analysis of confirmatory samples, it is very important to conduct the laboratory analysis on the same cup on which the XRF measurements were taken.

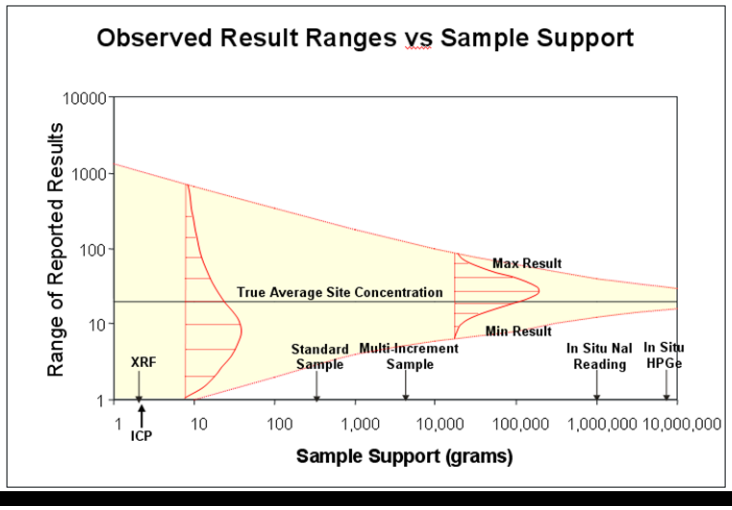


Notes

- Each link represents a variable contributing toward the quality of the analytical result. All links in the data quality chain must be intact for data to be of decision-making quality.

Data Variability is a Function of Sample Support

If every data point is compared individually to an action level, small sample supports will cause exceedances even if the true mean is well below the threshold.



3-48

Notes

- The graphic shows how the minimum and maximum values present in a set of samples changes as sample support changes. You have seen this graphic before, but this time we want to emphasize what happens to the decision making process when "not-to-exceed" thresholds are applied.

  As sample support shrinks, the amount of variability observed in sample results grows. This plays havoc with never-to-exceed standards. For example, if a never-to-exceed standard was around 80 ppm, data sets based on NaIi and HPGe measurements would conclude a site was in compliance, but data sets from an XRF, standard sampling, and multi-increment sampling programs would not. If the criteria were raised to a few hundred ppm, multi-increment sampling results would conclude there were no problems. If the criteria were raised to 900 ppm, the XRF would still be identifying problems, but results from standard samples would not. Depending on the measurement technologies deployed and their varying sample supports, one could potentially draw completely different conclusions about the site.

With these Issues, Can XRF Provide "Definitive Data?"

- Data Quality Objectives Process for Superfund: Interim Final Guidance (Sept. 1993)
  - » Page 43: *"For the data to be definitive, either analytical or total measurement error must be determined."*
- Measuring error requires taking multiple replicate analyses on a sample - this increases analytical costs, which is why error is seldom determined or reported
  - » However, XRF is unique in that replicate readings are very inexpensive
  - » Easy to determine error and meet SF's definition of definitive data

3-49

Notes

- Data Quality Objectives Process for Superfund: Interim Final Guidance (Sept. 1993), Page 43: *"For the data to be definitive, either analytical or total measurement error must be determined."*

  A precise, quantitative way to communicate uncertainty to decision makers could be phrased as "the analytical result = 398 ± 10 ppm Pb." This is much more meaningful to interpretation of data than relying on qualifiers.

  » Analytical error determination - Measures instrument precision by injecting replicates of the same sample extract into the instrument. Or determine precision and bias and an uncertainty interval by entering the method's most recent QC performance data into an uncertainty calculator spreadsheet (a Navy-developed uncertainty calculator is included on course CD). Contact Deana Crumbling (crumbling.deana@epa.gov) with questions for using it. But be aware that analytical error is very small compared to total error which includes sampling variability.

  » Total measurement error determination - Measures the overall precision of the entire measurement system (encompasses sample acquisition and processing on thru analysis). Contains the full expression of data uncertainty that affects confidence in interpreting the data and making decisions. Providing this ensures decision-makers are provided with full and transparent information. Replication of the sample acquisition and analytical process provides an estimate of total precision.

To do this, observe the soil volume which is to be represented by the sample. This is easily figured out if a systematic grid is used to lay out the sampling area and a sample is taken from each grid block. The concentration of the soil volume of that block is what the concentration result from the sample is supposed to represent. The sample result will be extrapolated to be the concentration for that block volume. So, how well do 2 samples collected from the same grid block (i.e., from the same sampling unit) provide the same results?

Summary: How to Reduce XRF Variability

- Analytical Variability
  - » Increase count time
  - » Use newer instrument
- Sample Handling Variability
  - » Homogenize samples
  - » Aggregate readings
  - » Handle samples consistently
- Sampling Variability
  - » More readings/samples per decision unit
  - » Increase sample support for each sample/reading

3-50

Summary: Controlling Heterogeneity Effects

- In a laboratory or field trailer
  - » Thoroughly prepare (dry, homogenize, grind) multi-increment bag contents and place sample in a XRF cup
- In the field
  - » XRF measurement aggregation
    - — Aggregate readings over the bag
  - » Multi-increment sampling over a decision unit into a large bag
  - » Multiple *in situ* shots (move XRF just slightly)

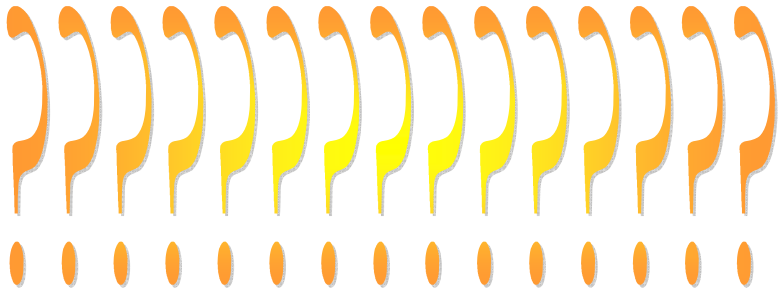
3-51

Notes

- Taking multiple in situ shots in basically the same location can help detect very high or very low results caused by a particle under the window that is not representative of the bulk soil.

How Many Increments?

Addressed in Module 6



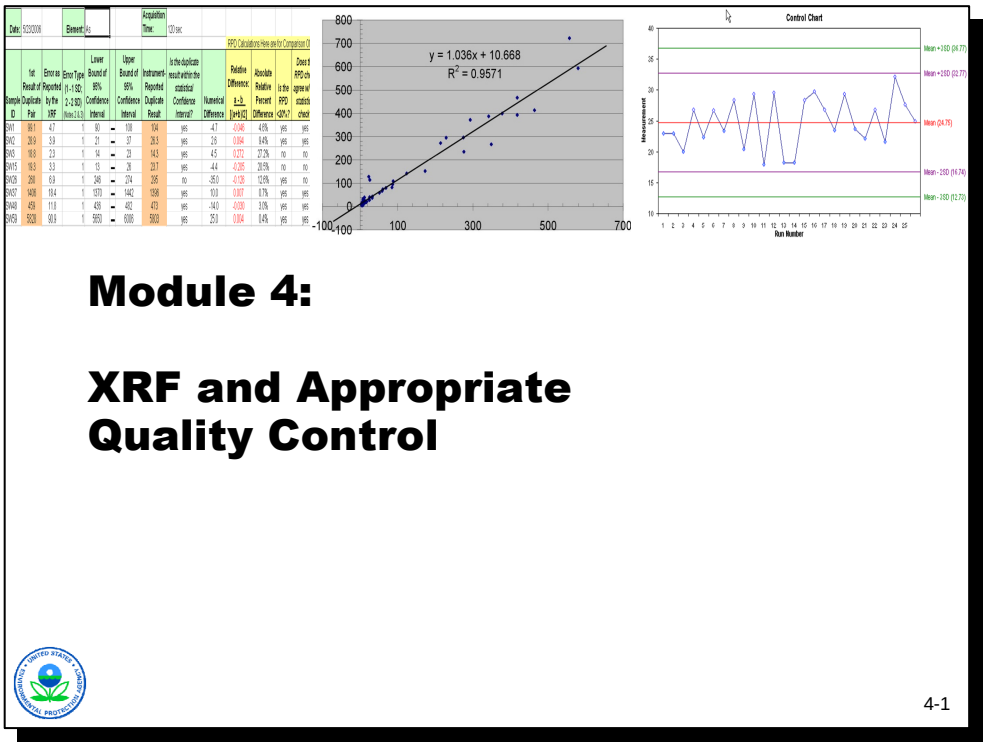
3-52

Q&A

3-53

Module 4

Quality Control



Module 4:

XRF and Appropriate Quality Control

4-1

Notes

- There are still issues surrounding the acceptance of XRF data for risk assessment and the collection of definitive versus screening data. A good QC program is critical to XRF data moving beyond the screening designation to more definitive uses such as delineation confirmation and risk assessment.
- The focus of this module is on the types of QC samples that are available, how they can be used, some examples of tools and strategies used successfully at other sites, and some potential pitfalls to look out for that highlight the critical need for good XRF QC.

How Did We Get Here? Why Do I Need Quality Control?

- Previous modules illustrate the need to manage:
  - » Spatial and sample variability
  - » Particle size and matrix effects
  - » CSM, instrument dynamic range
  - » Decision units and decision quality
- Desire to move beyond "screening" designation, ensure quality data
- Comparability with lab methods

4-2

Notes

- Most XRF projects perform an initial evaluation, often a comparison of split XRF and ICP samples using linear regression to obtain a correlation coefficient. For some projects, obtaining an initial good correlation coefficient becomes the sole effort of the QC program and once obtained little attention is paid to subsequent QC and XRF data quality.

  While method SW-846 6200 discusses use of energy calibrations, blanks, and precision measurements as important components of any XRF QC program, users often focus on the confirmatory analyses section. This section states "The correlation coefficient (r) for the results should be 0.7 or greater for the FP XRF data to be considered screening level data. If the r is 0.9 or greater and inferential statistics indicate the FP XRF data and the confirmatory data are statistically equivalent at a 99 percent confidence level, the data could potentially meet definitive level data criteria."

  Experience shows that while obtaining a good correlation with fixed laboratory data is extremely beneficial, maintaining a QC program to monitor instrument performance and understand potential issues for both XRF and lab methods is critical to ensuring that XRF data can be used for intended purposes. Rather than solely working to ensure that XRF data match ICP or other laboratory methods, a good QC program should ensure high quality XRF data that stands alone, can be used collaboratively with lab data or other information sources, and in some cases can highlight potential variability, matrix, or analytical issues associated with laboratory methods that otherwise can go un-noticed.

What Can Go Wrong with an XRF?

- Initial or continuing calibration problems
- Instrument drift
- Window contamination
- Interference effects
- Matrix effects
- Unacceptable detection limits
- Matrix heterogeneity effects
- Operator errors

4-3

Notes

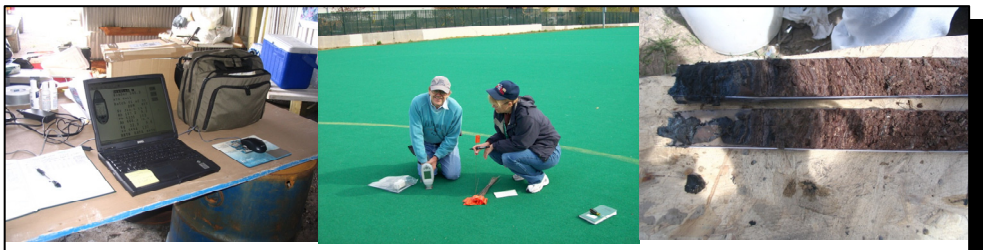
- This picture represents what some people may believe when it comes to QC - if we pretend it does not exist then we should not have an issue. In the past, we took a few splits, got a good R2 value, and knew the XRF was good. From there, people would operate the instrument in unconcerned bliss.

- However, there are a whole host of issues that can arise when using XRF at your site. To spot those issues before they have a significant impact on a project requires a QC program. This slide shows a list of the types of QC issues encountered when using FP XRF and these issues will be addressed in this module.

  Matrix heterogeneity and small scale variability are often critical sampling issues that need to be understood and managed. XRF is uniquely suited to help in this regard.

  Matrix effects example: moisture greater than 20 percent can impact performance. High levels of moisture can absorb or reflect x-rays resulting in bias.

  For *in-situ* operation good window contact and surface preparation are key.



Field Portable XRF Demonstration

4-4

Field Portable XRF
Daily Operation – Soil Mode

A. Power up: Stabilize 10 – 30 minutes
B. Instrument Detector Calibration
C. Application Verification
   1. SiO2 or Sand Blank (or both)
   2. SRMs 2709, 2710, 2711
   3. Other SRMs/certified standards as needed
   4. Precision Sample, e.g., LCS sample from ERA (~100-200 ppm most elements)

(continued)

4-5

Notes

◆ **Steps to startup, stabilize, and operate the portable XRF unit:**

A. Power up and stabilize
   - time may vary due to ambient conditions

B. Perform instrument detector calibration
   - inside of shutter or separate sample

C. Verify soil application
   - blank sample (results zero +/- Reportable Level)
   - NIST Standard Reference Materials (SRM) to check accuracy
   - precision sample for RSD calculation

Field Portable XRF
Daily Operation – Soil Mode

D. Run Soil Samples (in XRF sample cups)

1. Analyze 10 samples
2. Analyze MDL sample (e.g., SRM 2709)
3. Analyze Precision sample
4. Repeat steps 1-3 until all samples have been analyzed

(continued)

4-6

Notes

◆ **Steps to startup, stabilize, and operate the portable XRF unit:**

D. Analyze soil samples (contract laboratory program [CLP] - like protocol)
- prepared in XRF sample cups
- 10 samples
- Method Detection Limit (MDL) sample (concentrations near expected MDLs)
- Precision sample
- repeat until all samples analyzed

Field Portable XRF
Daily Operation – Soil Mode

E. Final QC
   1. Analyze MDL sample
   2. Analyze Precision sample
   3. Final detector calibration check (additional detector calib check may be needed every 2-4 hours)

F. Download results and backup data files

G. Process data and print preliminary results

4-7

Notes

- **Steps to startup, stabilize, and operate the portable XRF unit:**

  E. Final Daily QC
     - MDL sample
     - Precision sample
     - final detector calibration check

       NOTE: may need to check detector calibration every 2-4 hours depending on ambient conditions

  F. Download results to data files and backup files (e.g., USB drive)

  G. Process data, print daily results, update tables with all data

Field Portable XRF
Data Processing

A. Sort raw data (by sample ID) for elements of interest (Excel spreadsheet)

B. Copy QC results to appropriate sections of data processing spreadsheet

   1. MDL sample (e.g., SRM2709)
   2. Other SRM data
   3. SiO2 and/or Sand Blank

(continued)

4-8

Notes

◆ **Steps to process field portable XRF data:**

A. Sort raw data by sample ID for elements of interest
   - spreadsheet, database, or other software
     (will illustrate using an Excel spreadsheet)

B. QC data to the appropriate section of the spreadsheet
   - MDL sample data
   - data for other SRMs
   - SIO2 and/or Sand Blank data

Field Portable XRF
Data Processing

C. Sample Results
   1. Copy raw data to appropriate location in 'Raw_Data' sheet
   2. Final results in 'Final' sheet; 2-3 significant figures, RL qualified (if <RL, then report "U")

D. Calculations
   1. MDL = std dev * t-value (SRM 2709)
   2. QC samples: average, std dev, %RSD
   3. Precision sample: average, std dev, %RSD
   4. SiO2 / Sand blank: average

(continued)

4-9

Notes

◆ **Steps to process field portable XRF data:**

C. Sample Results
- copy raw data to appropriate section (e.g., 'Raw_Data' sheet)
- final results are in the 'Final' sheet
  2 or 3 significant figures (complicated equation)
  RL qualified (if < RL, report as "U")

D. Calculations
- MDL sample (typically SRM2709)
  average and std deviation (std dev) for all measurements
  MDL = std dev * t-value (Students t, 'TINV' function in Excel)
- QC samples
  average, std dev, percentRSD
  percentRSD = percent relative standard deviation
  = (std dev * 100) / average
- Precision sample
  average, std dev, percentRSD
- Blank sample (SIO2 or Sand)
  average

Field Portable XRF
Data Processing

E. Reporting Level (RL)

1. 1-5 times the calculated MDL (operator professional judgment)
2. Typical RL = 3 x MDL
3. Special case (e.g., Cr): RL ~ MDL

4-10

Notes

◆ **Steps to process field portable XRF data:**

E. Reporting Level (RL)
- 1-5 times the calculated MDL
  based on operator professional judgment
  experience with instrument, soil type, conc range, etc
- typical RL = 3 * MDL
- In special cases the RL may be approximately the MDL
  for example, often use MDL as the RL for chromium (Cr)
  because need best possible RL for Cr even though it
  is a difficult element by XRF

Typical MDLs (ppm)

Element	MDL	Element	MDL
Sb	63	Cu	26
As	9	Pb	12
Cd	26	Mn	150 *
Cr	43	Hg	6
Co	230 *	Ni	69 *

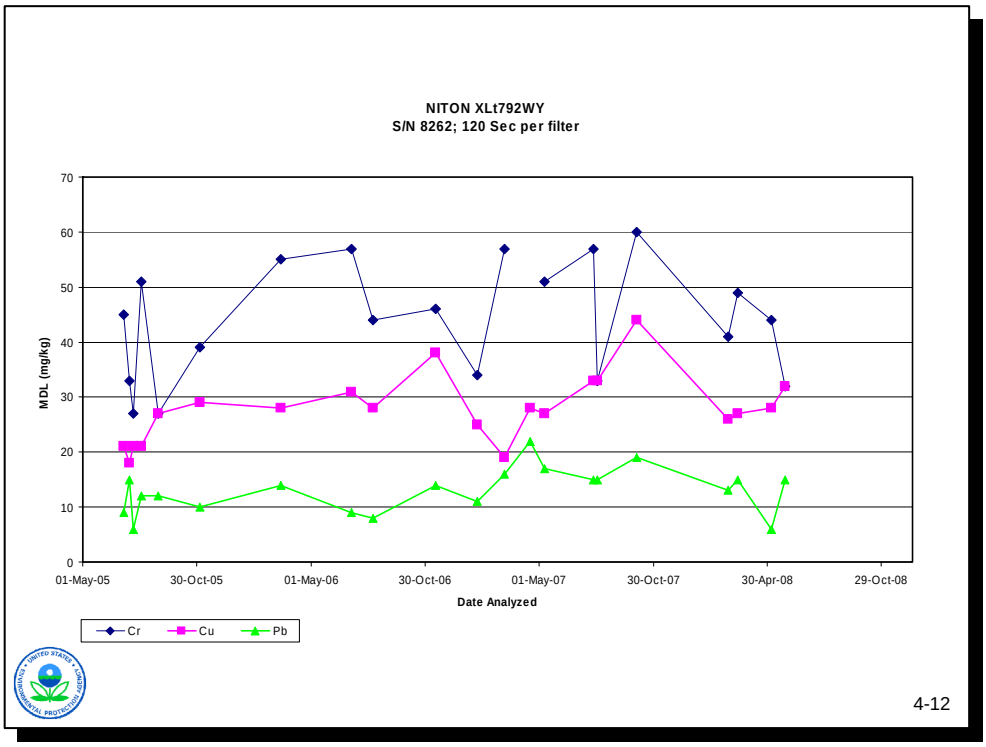
NITON XLt792YW FPXRF ANALYZER: 120 SEC PER FILTER; (240 SEC TOTAL), SRM 2709 SAMPLE

*** SIGNIFICANTLY BETTER MDL WITH SAND**

4-11

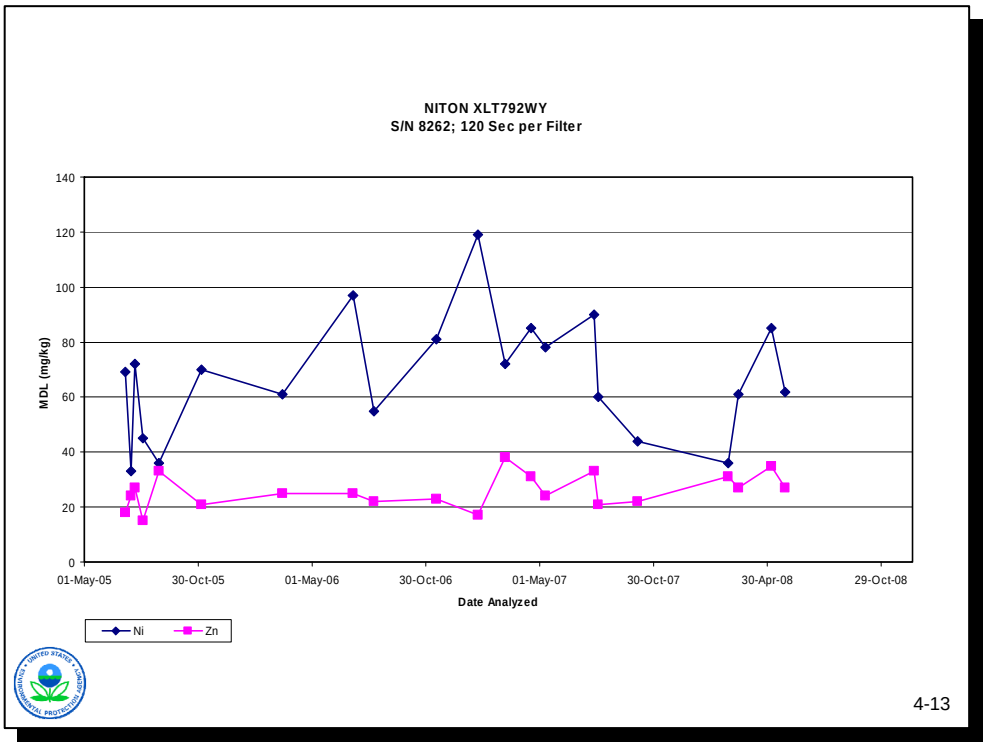
Notes

- This table shows TYPICAL MDLs for several soil contaminant elements commonly analyzed by field portable XRF. This is NOT a comprehensive list. Use SRM2709 as the MDL sample with 120 second measurement time for each measurement condition (2 for the NITON XLt792YW; 240 seconds TOTAL measurement time). NOTE: this is live measurement time, that is, the time that the analyzer is actually able to measure X-ray data. Clock time is greater than this because the counting electronics accounts for 'dead time' when the unit is busy processing data. NOTE that MDLs will be significantly better using sand because SRM2709 contains significant amounts of several elements, which leads to elevated MDL values.



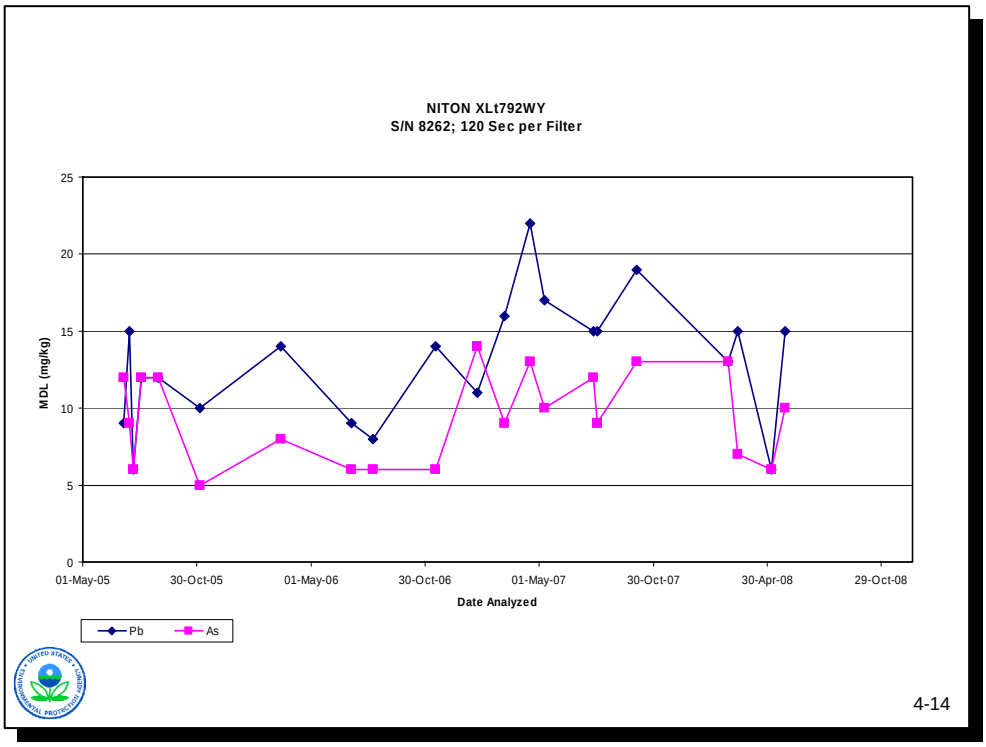
Notes

- This slide shows long-term stability of MDL values calculated for the NITON XLt792YW field portable XRF unit. More recent data (not shown) also falls in the same range. Good stability over a 3+ year time frame; this implies good stability on the X-ray tube output.



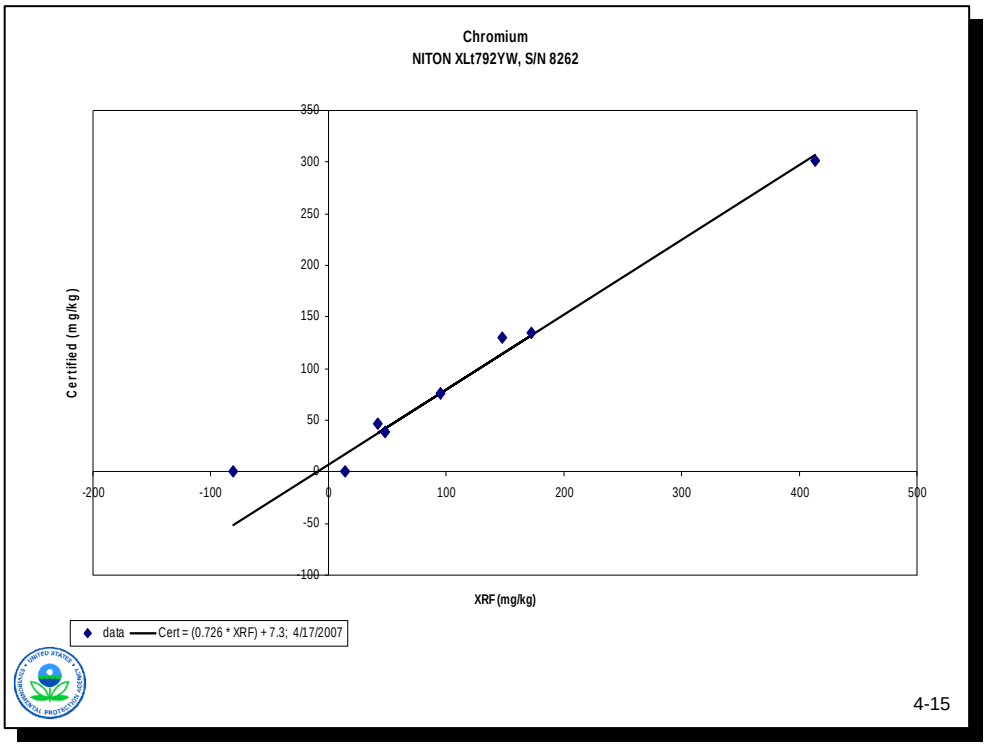
Notes

- This slide shows additional MDL stability data for Ni and Zn.



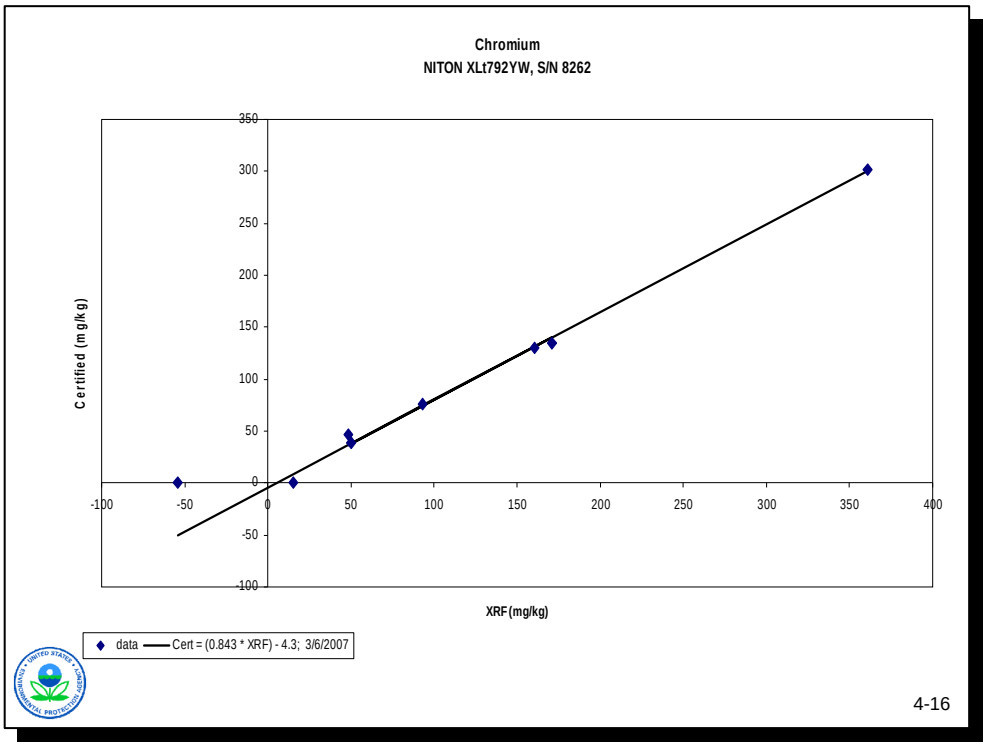
Notes

- This slide shows MDL stability data for Pb and As.



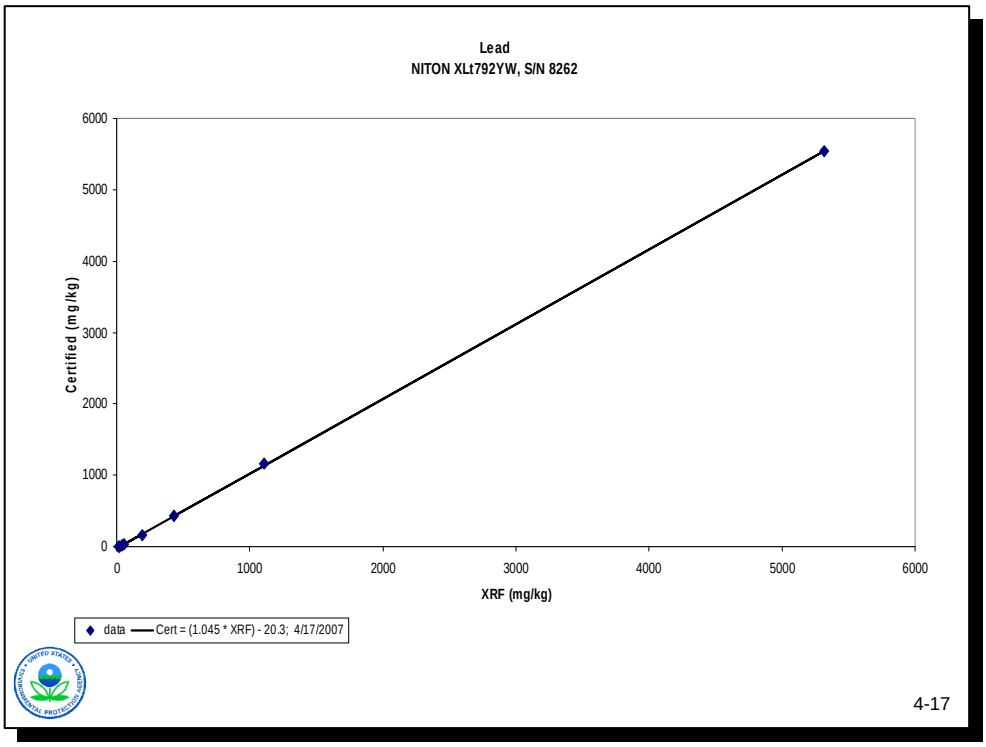
Notes

- This slide illustrates XRF performance for chromium with NIST certified SRMs using standard 120-second measurement times. There is a very good fit for Cr from the detection level to 300 ppm. Note that the two 'blanks' have quite different XRF readings; sand reads quite negative (approx -80 ppm) while SIO2 is close to zero and in better fit with the regression line. Note also, that XRF Cr biased high compared to certified values (calibration adjustment would take care of this).



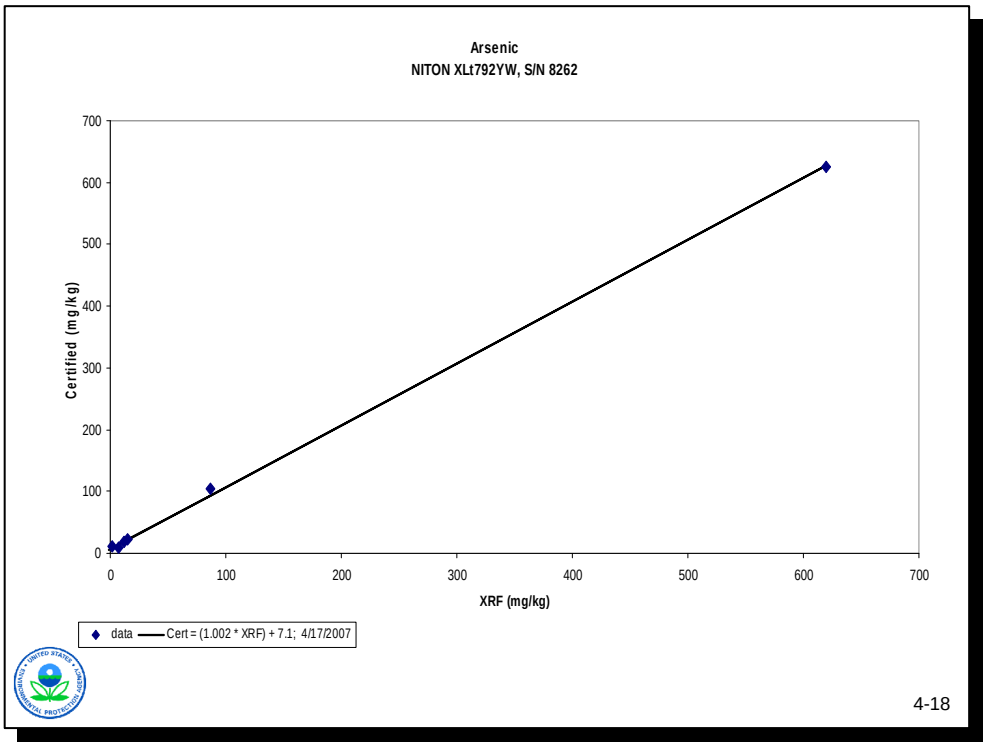
Notes

- This regression for Cr does not include the sand blank in the fit. There is a better regression without the sand blank.



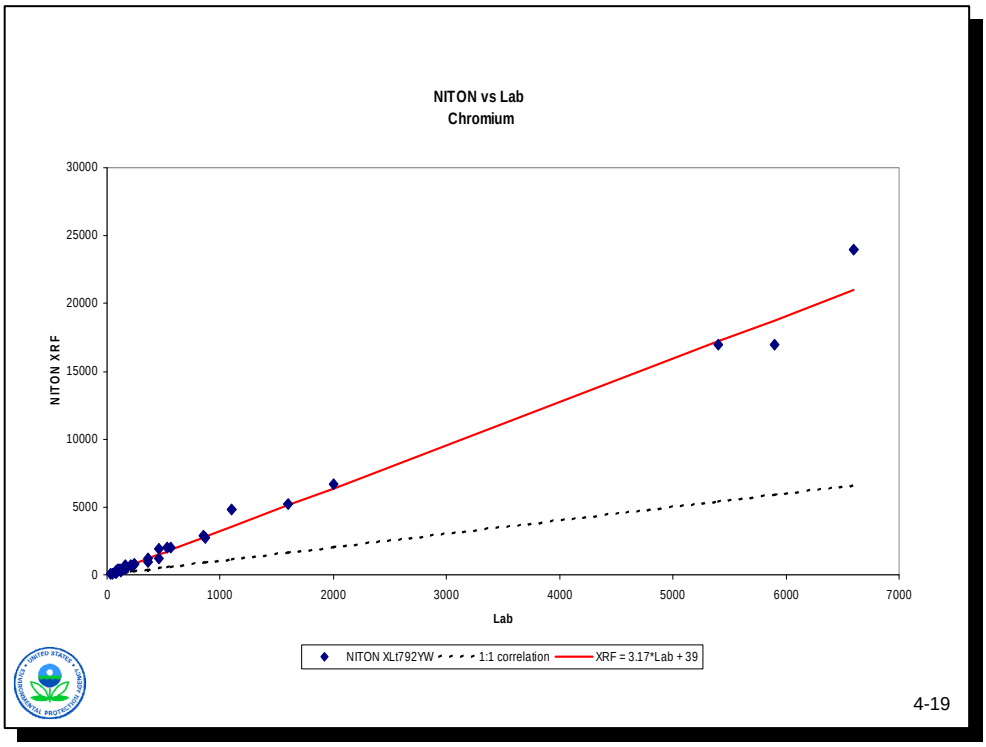
Notes

- This is another example of very good agreement between XRF and certified values for lead (Pb). Note that the slope is nearly 1.00 and the intercept is less than the typical Pb RL (approximately 50 ppm).



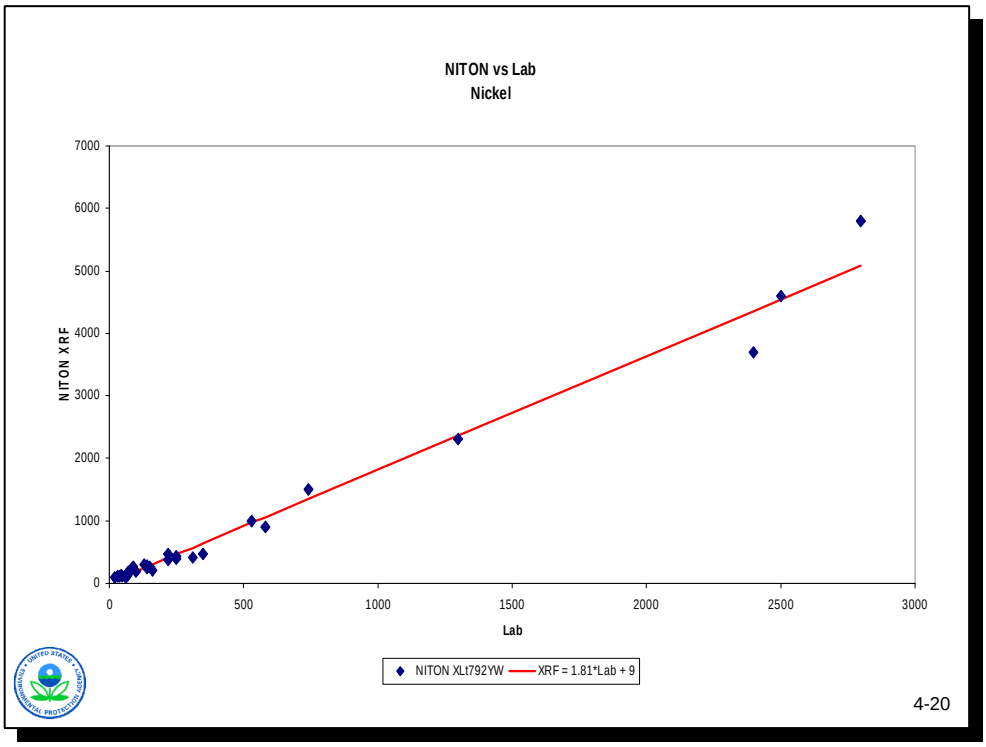
Notes

- This slide shows a very good regression for arsenic (As); the slope = 1.00 and there is a small intercept.



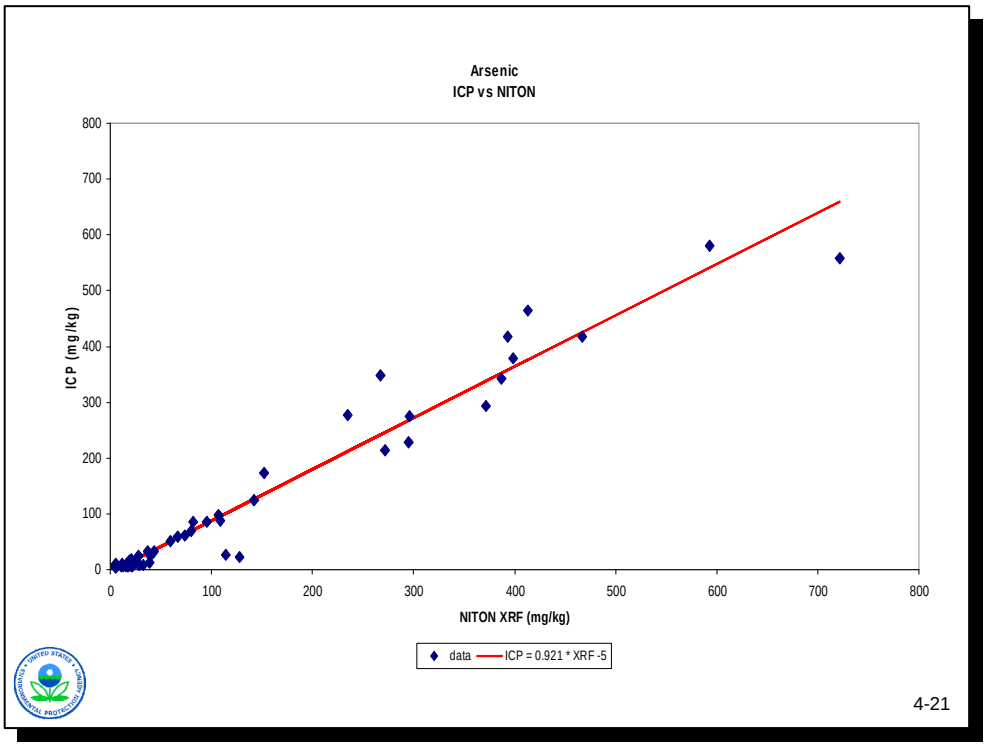
Notes

- This slide shows the regression for XRF versus laboratory analysis of Cr. Note that XRF values are approximately 3-times higher than the laboratory. This is to be expected since laboratory digestion methods are NOT 100 percent efficient and for Cr can range from 10 – 100 percent; typically about 30-50 percent. This depends on sample matrix. The XRF data are highly correlated with laboratory data and the regression may be used to "predict" laboratory values based XRF data PROVIDED the same soil matrix is considered.



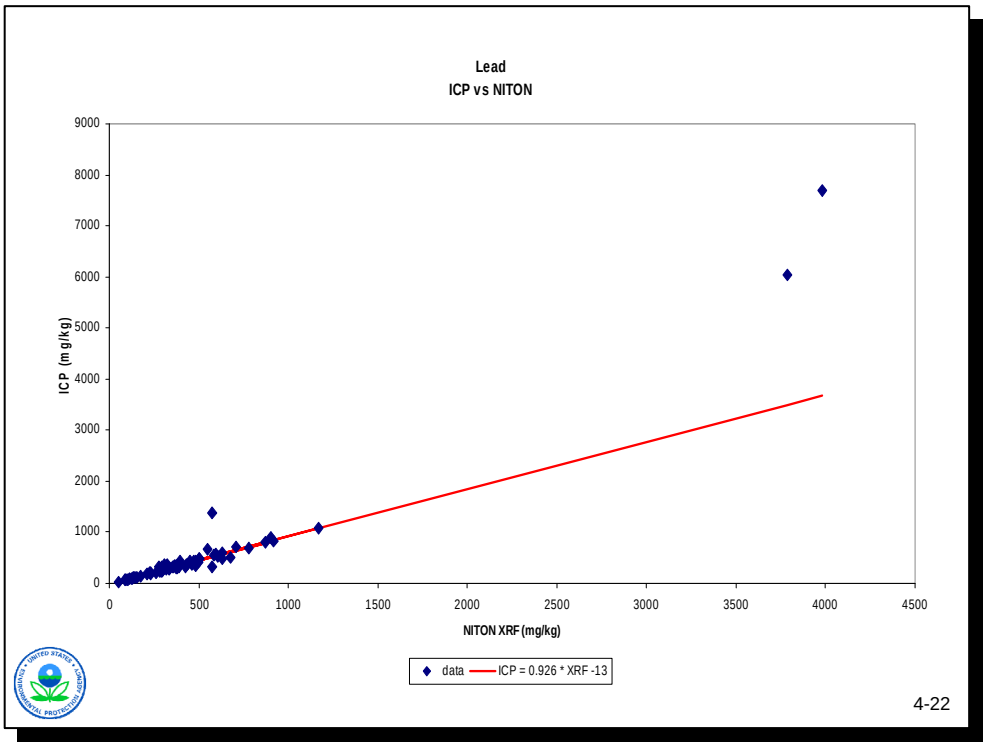
Notes

- Nickel by XRF is approximately 1.8-times higher than laboratory analysis; good correlation.



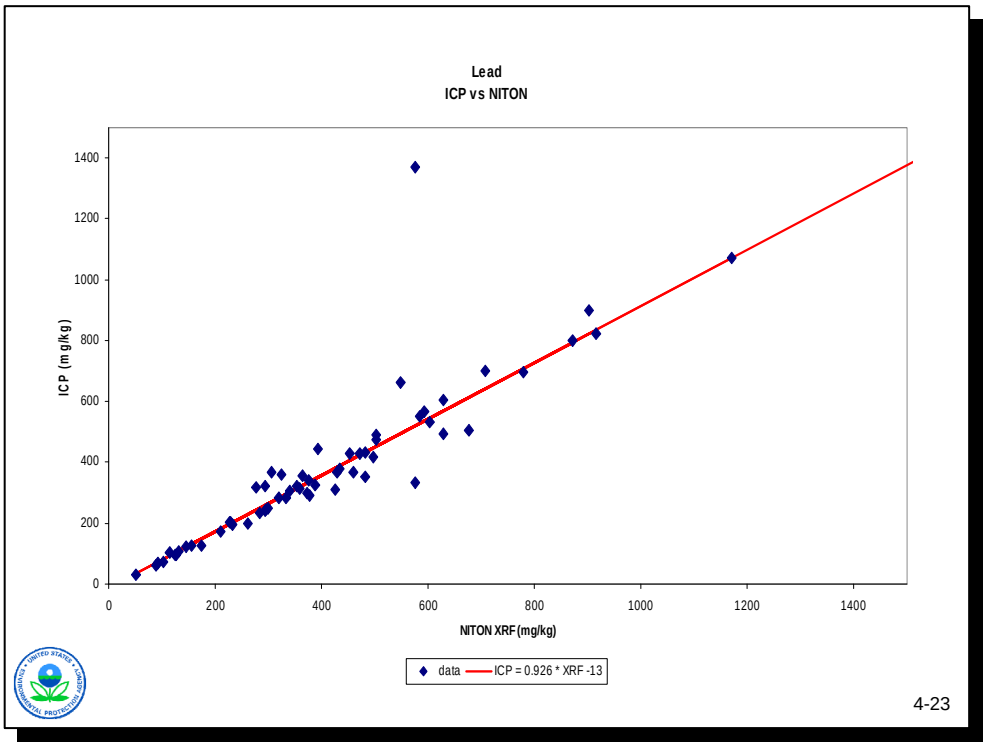
Notes

- Arsenic laboratory data is lower than XRF; laboratory values are approximately 92 percent of XRF value.



Notes

- This slide shows lead laboratory data also approximately 92 percent of XRF values. Note the 3 apparent outliers; the regression is without these outliers.



Notes

- This regression shows good agreement for Lead in the RL to 1200 ppm concentration range.

Field Portable XRF
High Throughput – Prepared Soil Samples

A. Dry, sieve, analyze in XRF cups
(Batch drying and sieving)

B. Two (2) people dedicated to XRF

1. First – receive and prepare samples
2. Second – XRF operation: QC/sample analysis (~ 5 min / sample), data processing, report preparation

(continued)

4-24

Notes

- This is an example illustrating High Throughput for XRF analysis of prepared soil samples.

  A. Dry, sieve, analyze in XRF cups (batch dry/sieve)

  B. Two (2) people dedicated to XRF analysis
     - first person to receive/prepare samples in XRF cups
     - second person to operate XRF unit (QC and site sample analysis, data processing, report preparation, report delivery)

Field Portable XRF
High Throughput – Prepared Soil Samples

C. Sample Throughput

1. Total of 214 samples (including prep dups) and delivered reports in approx 70 man-hours
2. Seven (7) elements analyzed by XRF
3. Prelim reports for all samples delivered to site manager
   - Two (2) significant figures, RL qualified
   - Sorted by location
   - Separate report sorted by Cr concentration

(continued)

4-25

Notes

- This is an example illustrating High Throughput for XRF analysis of prepared soil samples.

C. Sample Throughput
1. Total 214 samples (including preparation duplicates) and several delivered reports in approx 70 man-hours (~ 20 min/sample)
2. (daily reports and final reports delivered to site manager on final day of site activities)
3. Seven elements analyzed/reported by XRF
4. Prelim reports for all samples delivered to site manager on final day
   - two significant figures, RL qualified
   - sorted by sampling location
   - separate report sorted by Cr concentration (primary concern)

Field Portable XRF
High Throughput – Prepared Soil Samples

D. Time Breakdown

1. Sample Prep ~ 45%
2. Instrument operation/analysis ~ 40%
3. Data reduction/report prep ~ 15%

4-26

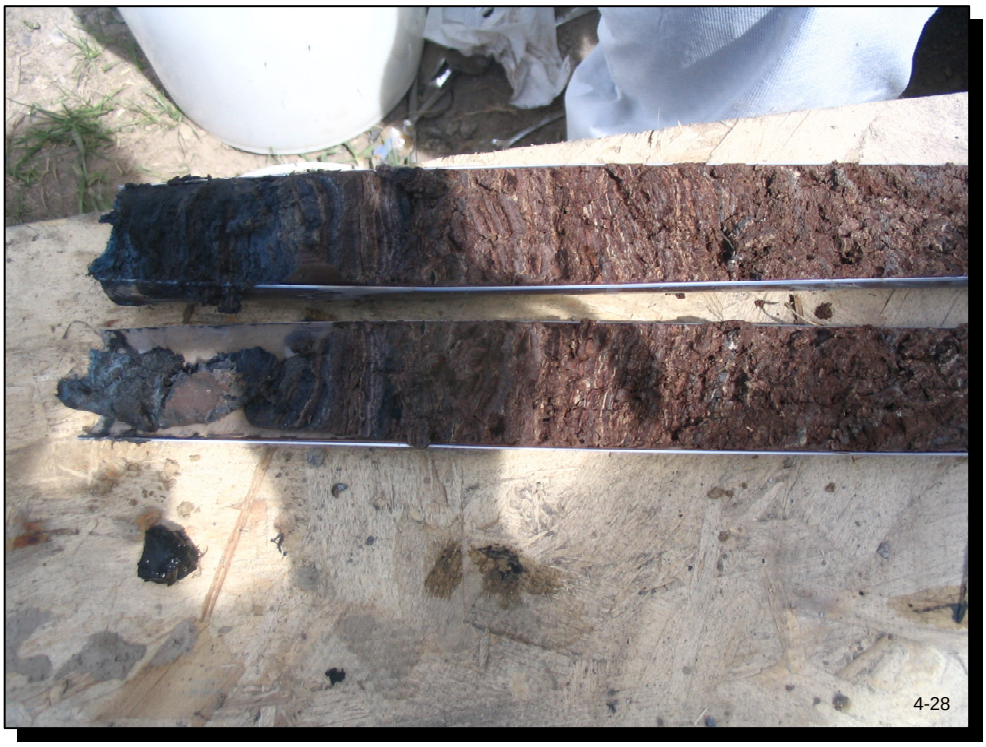
Notes

- This is an example illustrating High Throughput for XRF analysis of prepared soil samples.

  D. Time Breakdown (approx 20 minutes per sample)
  - Sample Prep approx 45 percent (~9 minutes)
  - Instrument operation/analysis approx 40 percent (8 minutes)
  - Data reduction/report preparation/report delivery approx 15 percent (3 minutes)

Notes

- This is the site where high throughput XRF analysis was done. The investigation team analyzed surface and core samples as well as sediment samples (creek running through the property).



Notes

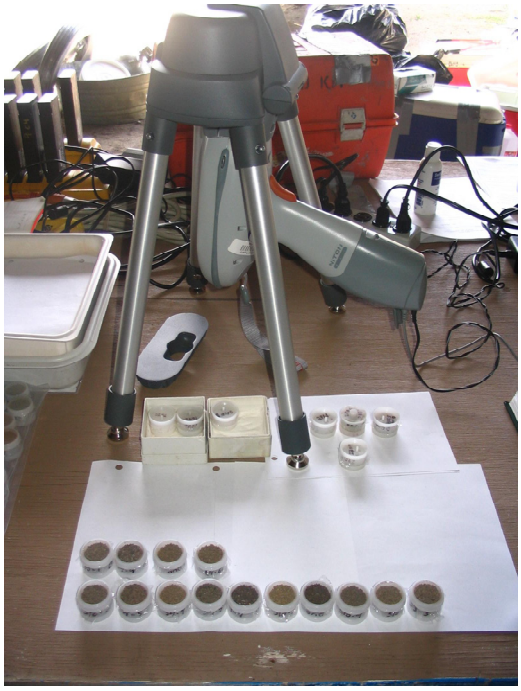
- Core samples were collected from various locations. A section corresponding to a depth range (e.g., 1-2ft) was submitted in a labeled plastic baggie for on-site XRF analysis. The sample was dried, sieved, placed in an XRF cup, and sealed with ¼-mil polypropylene X-ray window film prior to XRF analysis.

Notes

- Very “luxurious” accommodations by field standards:

  1. Inside a shed that had electricity and provided shade from the sun.
  2. Used an old sign as a table and barrels as the “legs.”

  Sample preparation was on the other side of the shed (left of this picture).



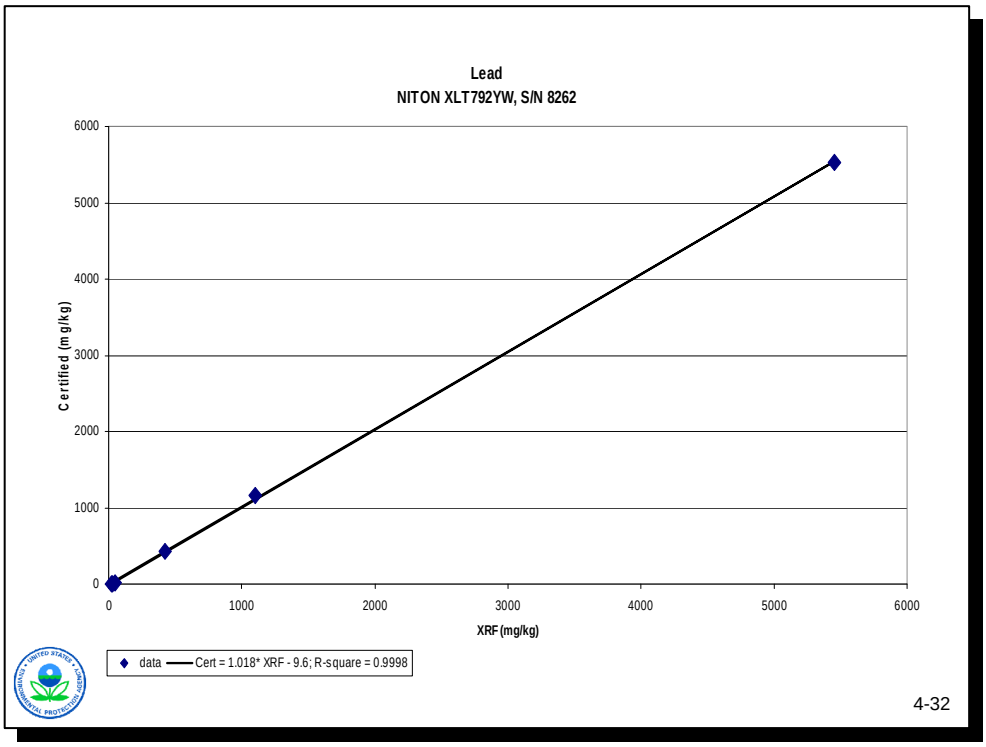
4-30

Notes

- This is a side view of the XRF unit in its Portable Test Stand with samples ready for analysis. There are QC samples next to the legs of the test stand.

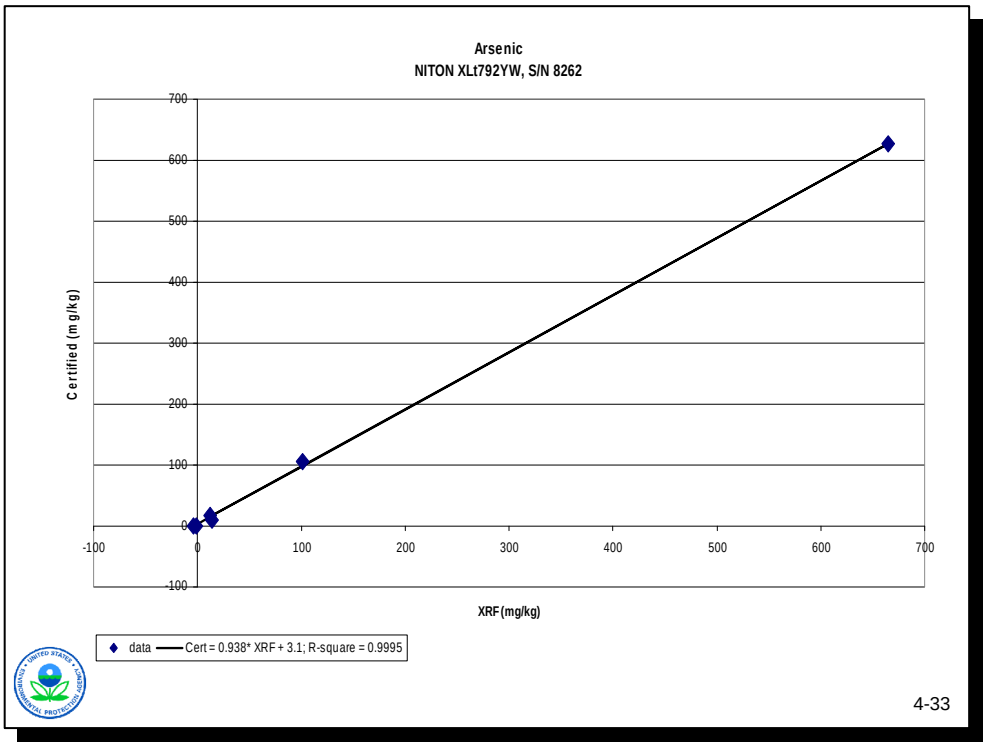
Notes

- This is a view of the results displayed on the portable computer. The computer is not required to run the analyzer and is used as a terminal to display results. XRF unit is self-contained and battery operated (or can use electricity as in this case). All data stored in the XRF unit and downloaded after analysis is completed for the day.



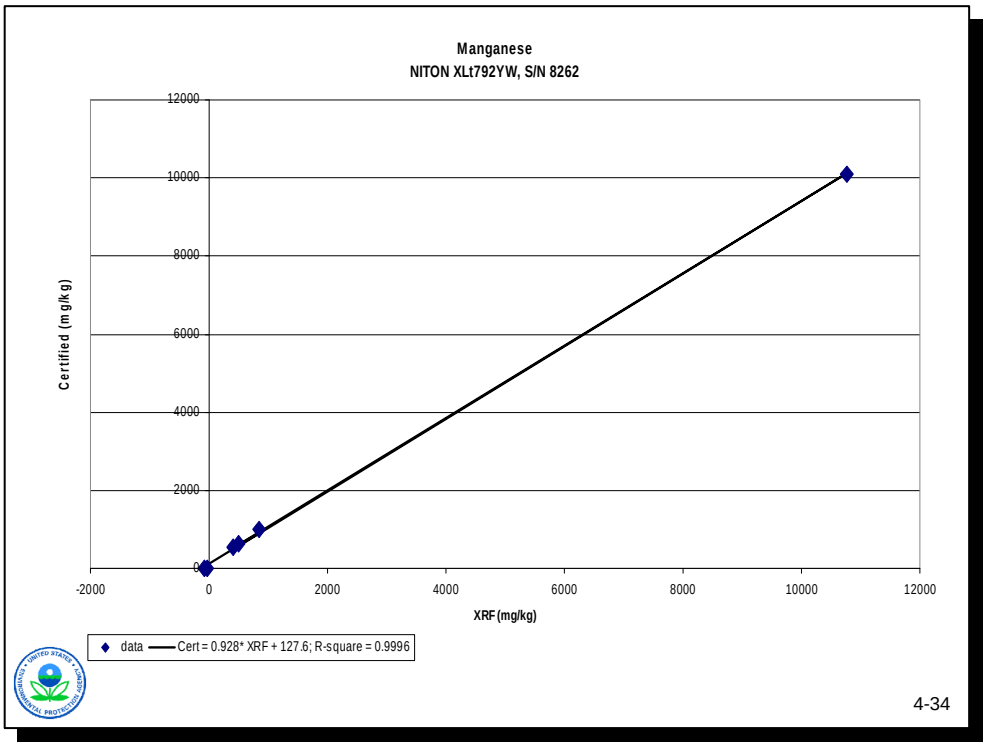
Notes

- This plot shows a very good performance for XRF versus certified values at this site; lead plot shown here (slope near 1.00 with very small intercept).



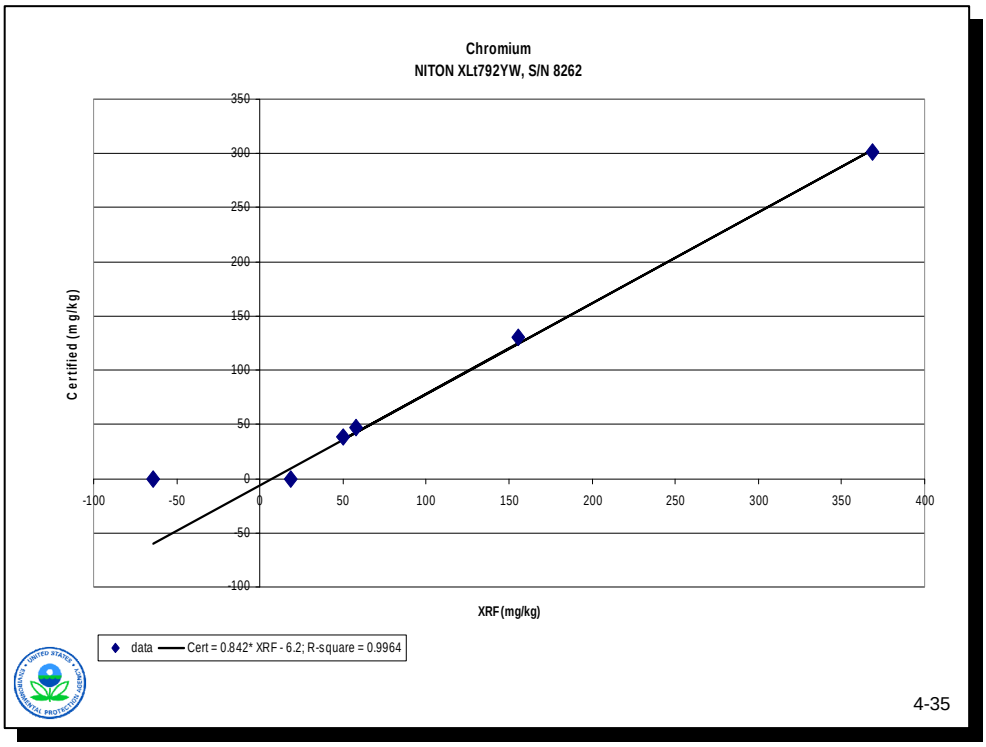
Notes

- Arsenic also is very good for SRM analysis; slope near 1.00 and very small intercept.



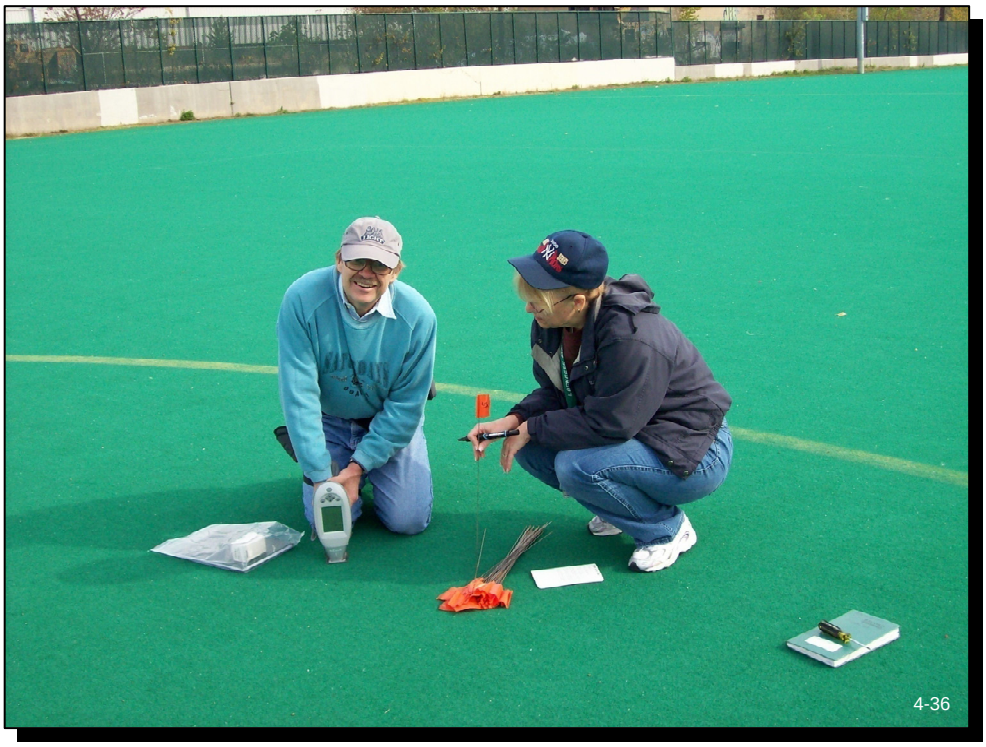
Notes

- This slide shows a similar performance for manganese.



Notes

- The plot for Chromium (primary concern) also is very good.

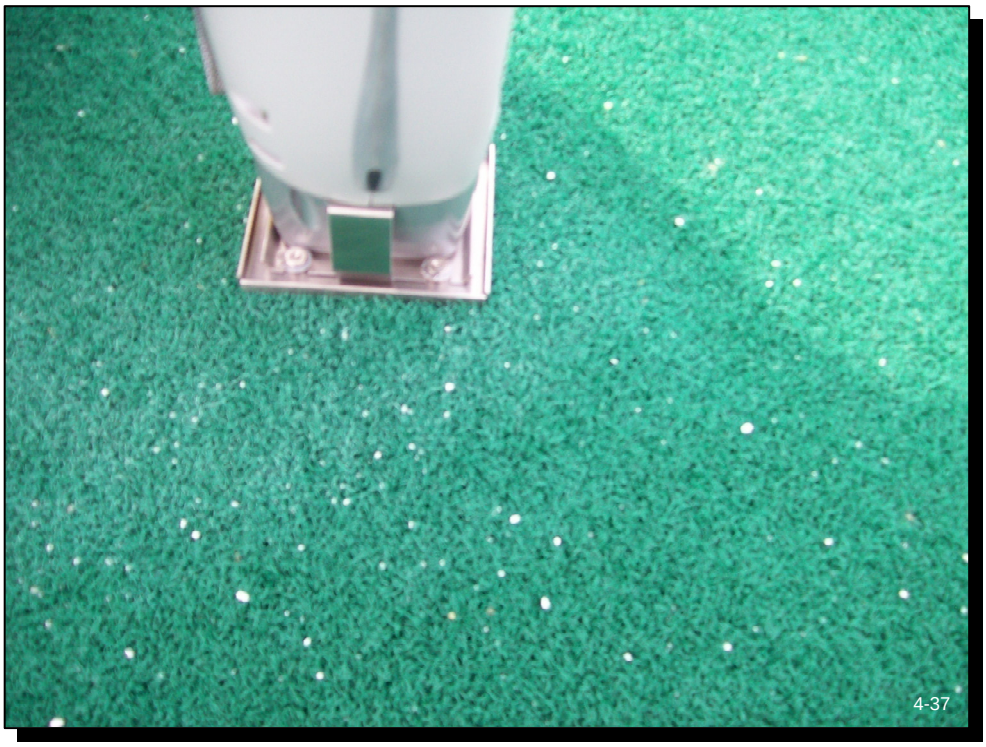


Notes

- We will switch gears and illustrate use of XRF as a screening tool. Here the XRF unit is being used to screen for contamination in turf on a playing field. This is not a soil matrix, therefore, screening data only – NOT QUANTITATIVE!

  It can, however, determine relative levels, i.e., can use ratios of readings.

  The measurement time was 60-seconds.

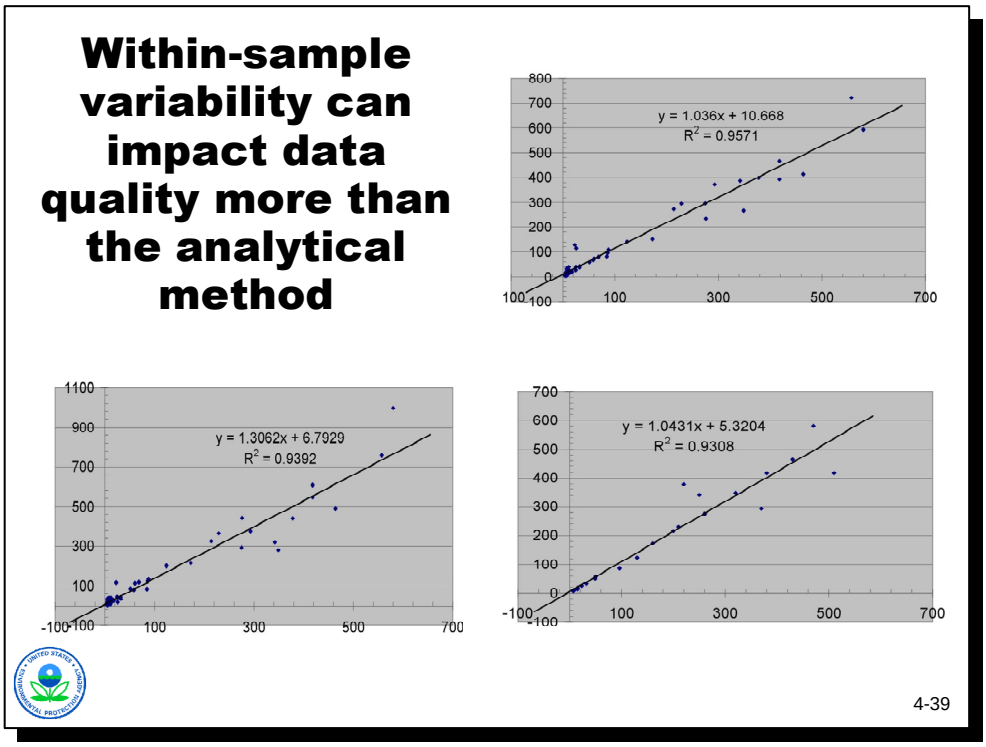


Notes

- This is a close-up of *in-situ* analysis adaptor for the XRF unit.

Notes

- The investigation team also measured the sandy perimeter of the field. The team screened more than 75 locations in a 3-hour period providing highly useful information about where to collect samples for laboratory analysis.



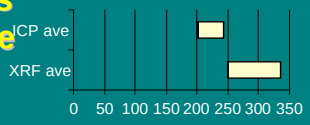
Notes

- As highlighted in previous modules, sample variability is the driver in terms of overall measurement uncertainty or variability. The next few slides illustrate the effects of within-sample variability. Care should be taken when generalizing in terms of which analytical methods provide better control on variability.
- Samples were archives (20-30 grams in a sandwich bag) analyzed for arsenic by ICP in 2005 by a Regional laboratory and again in 2006 by ERT. Correlation coefficients were better for both the Innov-X and Niton instrument cup samples and corresponding ICP analysis than they were for linear regressions that compared ICP from two different laboratories.

95% Confidence Interval (CI) Bar Graphs

2815 F1-SS1-3 jar As CI

Sometimes ICP is more precise…



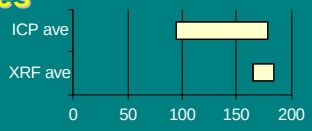
Arsenic CI Graph

As conc (ppm)

	XRF ave = 294	ICP ave = 203
width	86	40
LCL	250	202

2815 F1-SS1-3 jar Pb CI

…sometimes XRF is



Lead CI Graph

Pb conc (ppm)

	XRF ave = 174	ICP ave = 137
width	18	84
LCL	165	95

Triplicate subsamples of the same (<10 mesh) sample analyzed by both ICP & XRF

CI results for As & Pb shown here

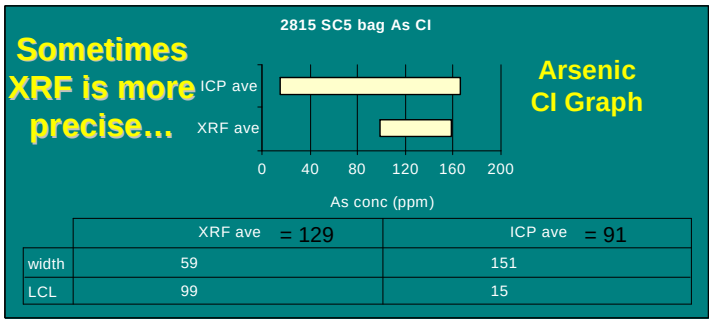
4-40

Notes

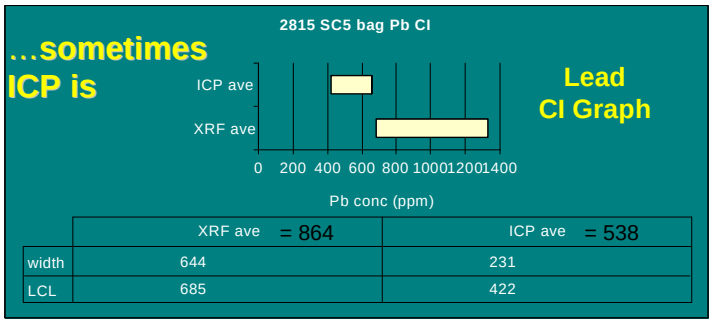
- It is difficult to generalize as to whether ICP or XRF is more precise, even within the same sample.

  The width of the confidence interval (CI) indicates the variability observed from triplicate analysis via ICP and XRF. In this sample, variability is higher for arsenic analyzed via XRF, while ICP has greater variability for lead.

95% CI Bar Graphs for Another Sample



	XRF ave = 129	ICP ave = 91
width	59	151
LCL	99	15



	XRF ave = 864	ICP ave = 538
width	644	231
LCL	685	422

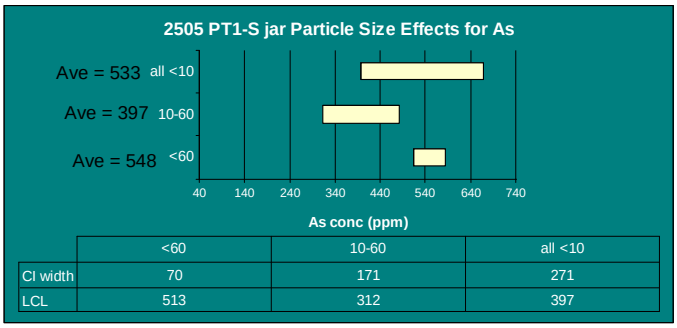
Here the situation is reversed for another sample from the same yard.

4-41

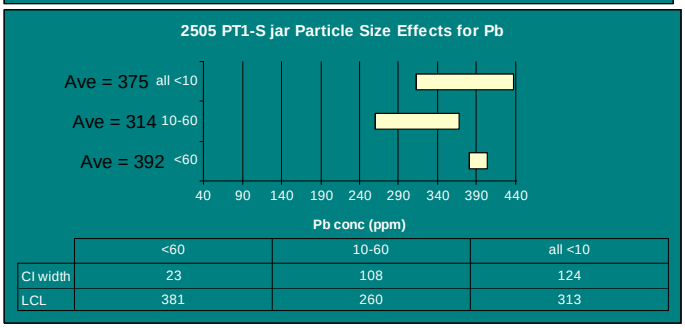
Notes

- In this series of images, the XRF average is always greater than the ICP. In most cases the XRF number can be considered conservative. If field based action levels are being used in addition to these conservative XRF concentrations, the project team can be very confident that the decision errors will be minimalized. However, for most applications the established laboratory method (in this case ICP) is expected to better represent the true mean.

95% CI Bar Graphs for Particle Size Effects



	<60	10-60	all <10
CI width	70	171	271
LCL	513	312	397



	<60	10-60	all <10
CI width	23	108	124
LCL	381	260	313

Particle Sizes

all <10 mesh = crushed soil put thru 10-mesh sieve & analyze all going thru

10-60 mesh = above then put thru 60-mesh sieve & what is retained on 60-mesh is analyzed

<60 mesh = that going thru 60-mesh sieve & is analyzed

4-42

Notes

- This graphic illustrates the huge impact particle size can have on both concentration and variability of results.

  Notice how variability generally decreases and concentrations generally increase with a reduction in particle size. Why?

  Greater surface area and negative charges of some finer grained clay materials for example, may be preferentially sorbing metals contamination.

  Understanding this trend at your site can have significant implications for defining appropriate decisions and decision units, XRF sample preparation, and method comparability with laboratory analyses.

What Can We Do?

4-43

Notes

- What can be done to stem the tide of uncertainty?

  Recognize that uncertainty exists within any method, then we can use QC programs to identify areas of variability or uncertainty and focus resources on those with the greatest potential impact to your project.

  Recognizing limitations and using the advantages of each method allows data sets to be evaluated collaboratively to control different areas of uncertainty (for example, laboratory methods to control analytical variability and XRF to control sample or spatial variability).

  Use of a Demonstration of Method Applicability (DMA) early in the project can help refine QA/QC goals and processes for field deployments, allow development of decision logic diagrams to drive dynamic work strategies, and provide a mechanism for adaptively managing uncertainty or variability that can have potentially significant impacts on your project.

NIST – Recognition of Variability

- Certified concentrations based on two or more independent methods requiring complete sample decomposition or nondestructive analysis
- Some of the most homogenous and well characterized material out there
- Yet. . . . . .

NIST 2709
Certified Values

Element	Mass Fraction (mg/kg)
Antimony	38.4 ± 3
Arsenic	626 ± 38
Barium	707 ± 51
Cadmium	21.8 ± 0.2
Copper	2950 ± 130
Lead	5532 ± 80
Mercury	32.6 ± 1.8
Nickel	14.3 ± 1.0
Silver	35.3 ± 1.5
Vanadium	76.6 ± 2.3
Zinc	6952 ± 91

4-44

Notes

- **Certified concentrations based on two or more independent methods requiring complete sample decomposition or nondestructive analysis:** The table on the right-hand side of this slide shows the NIST SRMs. The certified values are weighted means of results from two or more independent analytical methods, or the mean of results from a single definitive method, except for mercury. Mercury certification is based on cold vapor atomic absorption spectrometry used by two different laboratories employing different methods of sample preparation prior to measurement. The weights for the weighted means were computed according to the iterative procedure of Paule and Mandel [1]. **Note that there are uncertainties associated with the reference values.**

- **Some of the most homogenous and well characterized material out there:** NIST SRMs are some of the most homogenous and well characterized material available yet even these samples recognize and quantify uncertainty. The stated uncertainties include allowances for measurement imprecision, material variability, and differences among analytical methods. Each uncertainty is the sum of the half-width of a 95 percent prediction interval and includes an allowance for systematic error among the methods used. In the absence of systematic error, a 95 percent prediction interval predicts where the true concentrations of 95 percent of the samples of this SRM lie. XRF was actually used to assess heterogeneity of these materials as part of the certification process.

- The NIST certified values and associated uncertainty should highlight the fact that most laboratory methods do not quantify and report uncertainty or variability. XRF has the advantage of providing an estimate of variability based on the 1 or 2 standard deviations of the counting statistics reported by portable instruments. Higher variability often indicates the need to better control some portion of the process such as sample preparation or take additional XRF readings to calculate a sample mean and understand within-sample variability.

NIST – Recognition of Digestion Issues

See NIST 2709, 2710, 2711 Addendums

Element	Range			Median	N	% Leach Recovery†
	mg/kg					
Antimony	---		---	< 10	1	...
Arsenic	---		---	< 20	2	...
Barium	392	-	400	398	2	41
Cadmium	---		---	< 1	5	...
Chromium	60	-	115	79	5	61
Cobalt	10	-	15	12	5	90
Copper	26	-	40	32	7	92
Lead	12	-	18	13	5	69
Manganese	360	-	600	470	7	87
Molybdenum	---		---	< 2	2	...
Nickel	65	-	90	78	7	89
Selenium	nr	-	nr	0.014	1	< 1
Strontium	100	-	112	101	3	44
Vanadium	51	-	70	62	3	55
Zinc	87	-	120	100	7	94

† % Leach Recovery = $100 \times \left[\frac{\text{Median Value}}{\text{Certified/Information Value}}\right]$

--- at or below the detection limit
... no % Leach Recovery calculated
nr no range reported by the laboratory

4-45

Notes

- EPA has established a number of leach methods for the determination of labile or extractable elements. They include Methods 3015, 3050, and 3051. Of course the term "total metals" usually accompanies these methods.

  A number of cooperating laboratories using the variation to EPA Method 3050 for Flame Atomic Absorption Spectrometry (FAAS) and Inductively Coupled Plasma-Atomic Emission Spectrometry (ICP-AES) measurements, have reported data for SRMs 2709, 2710, and 2711. This variation of the method uses hydrochloric acid in its final step, which is different from Method 3050 for ICP-MS which I believe uses HNO3. Several laboratories provided replicate (3 to 6) analyses for each of the three soil SRMs. The number of results for a given element varied from only one to as many as nine, as indicated in the data presented in Tables 1 through 3. Because of the wide range of interlaboratory results for most elements, only the data range and median of the individual laboratory means are given. Ranges differ somewhat from those in reference [26], since this addendum is based on a larger data set than had been available previously.

  This slide shows a subset of the results. These are not considered "certified values" but they do illustrate the issues or complexities that are derived from using methods that do not completely digest all metals in the matrix. Chromium is a classic example of a metal that commonly analyzed for using ICP or XRF and has poor recoveries. The lead result shows only a 69 percent leach recovery. Incomplete digestion is an issue to be aware of particularly when assessing comparability of XRF and ICP or AA.

  For a number of sites, the DMA and XRF illustrated that in some cases the risk or regulatory drivers that were expected based on existing digestion/ICP data were not the major risk drivers. Instead, metals like antimony with its poor digestion efficiencies ended up driving the XRF delineation and excavation.

Demonstration of Method Applicability (DMA)

"I think that in the discussion of natural problems we ought to begin not with the Scriptures, but with experiments, and demonstrations."

Galileo Galilei

- Concept founded in SW-846, performance based measurement initiative
- Initial site-specific performance evaluation
  - » Analytical and direct sensing methods
  - » Sample design, sample collection techniques, sample preparation strategies
  - » Used to select information sources for field and off-site
- Goal is to establish that proposed technologies and strategies can provide information appropriate to meet project decision criteria

4-46

Notes

- **Concept founded in SW-846:** Even with well understood technologies like XRF, MIP, and laser-induced fluorescence (LIF) stakeholders will not rely on these tools and strategies without site-specific demonstrations to understand how tools or approaches can be used to effectively manage uncertainty and be used collaboratively with other information sources. The performance based measurement (PBMS) initiative fits nicely with Triad because PBMS conveys "what" needs to be accomplished, but not prescriptively "how" to do it. EPA defines PBMS as a set of processes wherein the data needs, mandates, or limitations of a program or project are specified, and serve as criteria for selecting appropriate methods to meet those needs in a cost-effective manner.

- **Initial site-specific performance evaluation:** Under a performance-based approach, EPA would specify:
  - » Questions to be answered by monitoring.
  - » Decisions to be supported by the data.
  - » **Level of uncertainty acceptable for making decisions.**
  - » Documentation to be generated to support this approach in the monitoring program.

- **Goal:** Data should be collected to meet project specificity, sensitivity, and reliability requirements.

Internet

Information on the PBMS initiative can be found at *http://www.epa.gov/sw-846/pbms.htm*

Why Do You Need a DMA?

- Triad usually involves real-time measurements to drive DWS
- Greatest sources of uncertainty usually sample heterogeneity and spatial variability
- Relationships with established laboratory methods often required to make defensible decisions
- Provides an initial look at CSM assumptions
- Not always appropriate
  - » Limited scope or sufficient resources
  - » Remember the Brownfields perception

4-47

Notes

- **Triad usually involves real-time measurements to drive dynamic work strategies (DWS):** Because XRF has been used for the past 15 years, performance of a DMA may seem, on the surface, to be unnecessary. However, Module 2 - the basics, illustrated why site-specific regression must be used carefully, and Module 3 showed the importance of understanding sample heterogeneity and developing strategies to deal with those uncertainties. Module 3 also introduced the concept of tailoring sample support to the decision at hand. A DMA can help do that.
- **Greatest sources of uncertainty usually sample heterogeneity and spatial variability:** The issues associated with heterogeneity and spatial variability cannot be managed if they are not recognized until after demobilization from the field. DMAs provide an early look at the significance of these issues and allow establishment of effective strategies (QA/QC) to deal with them. Many technologies still struggle with the "screening" stigma. XRF is a good example of a very well established technology for which many regulators still require fixed laboratory "confirmation" and stakeholder acceptance often requires this.
- **Relationships with established laboratory methods often required to make defensible decisions:** The relationship of XRF and laboratory methods allow the user to have confidence in the program. By establishing this relationship, the project team can develop field based action levels, ranges (clean, dirty, unsure), target collaborative samples, and monitor decision error rates. The relationship also can help highlight the fact that "gold plated fixed laboratory" results suffer from the same sampling and sub-sampling issues that innovative or field methods do. By establishing these relationships, the project team can demonstrate why stakeholders cannot expect field methods to compare any better than two laboratories or even the same laboratory.

- **Provides an initial look at CSM assumptions:** Most technologies and approaches requiring a DMA result in increased information density. The DMA should be used to determine if the preliminary conceptual site model (CSM) assumptions are true.

- **Not always appropriate:** Some components of investigation and cleanup such as the SI step within the Comprehensive Environmental Response Compensation and Liability Act (CERCLA) or Superfund, where guidance suggests collection of 20 or fewer samples, may not be appropriate for conducting a DMA. Similarly, some Brownfields sites, with grants that have very limited funding, may not be appropriate to accommodate a DMA. Projects with adequate resources to employ established mobile or fixed laboratory methods at sufficient density may be inappropriate, while those requiring method modifications or careful examination of sampling and spatial uncertainties may benefit significantly from DMAs. Even if only fixed laboratory methods are used, a DMA should be considered for a fixed laboratory method if there is any question about matrix interference effects. Just a few pilot samples could save millions of dollars of wasted analyses by detecting extraction or other problems at the start.

  It should be noted, however, that for most applications a DMA is beneficial precisely because a particular field analytical technique, direct sensing tool, or innovative strategy is identified as applicable to cost effectively increase data density, refine the CSM, or address small scale variability and matrix heterogeneity. In some cases, selection of sampling locations for an SI are obvious (for example: visual staining, product, lagoons, discharge points) while other cases are more complicated, making determination of appropriate sampling locations for those 20 samples problematic. Depending on the nature of suspected contamination some sample material can be archived and potentially used later as part of a DMA for an expanded SI or additional work.

  In the case of Brownfields applications, most assessment grants are in the range of $200,000 where it is reasonable to assume that resources allow for data densities greater than 20 samples. At sites with elevated expenditures associated with collection of subsurface samples, adding limited additional cost for analysis of field analytical methods, direct sensing tools, or other innovative technologies does not significantly raise project expenses. In these cases, inexpensive analytics and direct sensing tools can provide greater vertical density and help target locations for more expensive traditional laboratory samples.

  Finally, the very definition of a Brownfield - “a property, redevelopment, or reuse which may be complicated by the presence or potential presence, of a hazardous substance, pollutant, or contaminant” underscores the need for higher data density and collaborative data sets that often accompany DMAs. Regardless of whether significant contamination or the perception of contamination is present at such a property, DMAs and associated innovative tools allow for a higher data density that facilitates timely revitalization. These data sets are particularly helpful to address stakeholder concerns and provide a level of comfort that allows developers, insurance partners, risk partners, public stakeholders, state agencies, local agencies and others to be involved, invested, and reassured with a project outcome.

What's Involved?

- There is no template for DMAs!
  - » Format, timing, documentation depend heavily on site specifics, existing information, intended data use
- Performed early in program
- Go beyond simple technology evaluation to optimize full scale
  - » Sample design, decision and unit designations
  - » Sample prep, throughput, other logistics
  - » Data management issues
- Documentation

4-48

Notes

- **There is no template for DMAs:** DMAs can be performed easily and affordably before mobilization, or as an early component of a field program. It does not necessarily require a separate mobilization. The complexity of the DMA should be commensurate with the expected complexity and scope of the project and with the expected data use and decisions being made.

  Existing information and archived samples are often extremely valuable.

- **Performed early in program:** It is best to identify potential issues and design strategies early in the sampling program. The DMA process also allows planning for contingencies.

- **Go beyond simple technology evaluation to optimize full scale:** Effective DMAs go beyond simple "does it work at my site" questions to look at sampling, logistical, and data management issues. Communication, data sharing, visualization, collaborative data needs, staffing, project sequencing can all be optimized prior to full-scale implementation.

What to Look For....

- Effectiveness - Does it work as advertised?
- QA/QC issues
  - » Are DLs and RLs for site matrices sufficient?
  - » What is the expected variability? Precision?
  - » Bias, false positives/false negatives?
  - » How does sample support effect results?
  - » Develop initial relationships of collaborative data sets that provide framework of preliminary QC program
- Matrix issues?
- Do collaborative data sets lead to the same decision?
- Assessing alternative strategies as contingencies

4-49

Notes

- **Effectiveness:** A DMA can assess whether or not the technologies to be used will perform as advertised by the vendor. Sites can benefit from even simple DMAs. For example, on several sites, DMAs were not performed for technologies such as ground penetrating radar (GPR), EM, and resistivity. The project teams encountered depth issues with GPR, interferences with EM, and data processing/interpolation and surveying with resistivity. A DMA and performance-based contracting mechanism would have saved project resources because the data from these technologies was collected, but could not be used.

- **QA/QC issues:** A DMA can provide valuable information that can be used to optimize procedures to address variability. Variability effects statistically based sampling designs and tolerable uncertainty. For example, if a 95 percent UCL is being used to make decisions and there is significant variability, often times the 95 percent UCL is pushed above the action level. In such a case, some resources put toward understanding and controlling variability due to small scale or matrix heterogeneity will benefit a sampling design. With a high density real time tool it may be possible to isolate problem areas within a decision unit and address them separately, rather than taking action on the entire decision unit.

  Sample support is the size, shape and orientation of the sample. A DMA can assess if the level of effort required for advanced sample preparation is worth the higher precision, accuracy, or bias control achieved.

  QC samples are collected and analyzed to evaluate which uncertainties are the largest contributors to total measurement error. Project resources can then be allocated to control for those activities with the highest impact.

- **Matrix issues:** Some direct sensing tools have direct push limitations that can be evaluated and addressed during a DMA. Typical matrix issues include, XRF-lead arsenic peak overlap, moisture >20 percent, and high concentration of unexpected metals.

- **Do collaborative data sets lead to the same decision:** Examine whether collaborative data sets lead to the same decision as XRF alone.

- **Assessing alternative strategies as contingencies:** By conducting a DMA, alternative strategies can be assessed as contingencies that can be implemented should the performance of intended methods prove to be inadequate.

More DMA Benefits

- Augment planned data collection and CSM development activities
- Test drive communication and data management schemes
- Develop relationships between visual observations and direct sensing tools
- Flexibility to change tactics based on DMA rather than full implementation
- Establish initial decision logic for DWS
- Evaluate existing contract mechanisms
- Optimize sequencing, staffing, load balance, unitizing costs

4-50

Notes

- **Augment planned data collection and CSM development activities:** The DMA data has additional uses other than just to evaluate and optimize how the XRF will be used. The DMA data will augment the future data collection and CSM development activities.
- **Test drive communication and data management schemes:** The DMA allows the project team to test the communication and data management schemes that are available. These include sampling and statistical tools and visualization and data management tools.
- **Develop relationships between visual observations and direct sensing tools:** The DMA can be used to develop standard descriptions for visual observations. Although this can be extremely beneficial for technologies like MIP and LIF, it also applies to XRF. Things like tailings, high moisture, or matrix differences like ash layers, can be visually obvious and can warrant special consideration.
- **Flexibility to change tactics based on DMA rather than full implementation:** The DMA provides the flexibility and opportunity to change tactics based on the results. This is much easier to accomplish with the smaller-scale DMA rather than during full implementation of a sampling program.
- **Establish initial decision logic for DWS:** It is difficult to develop decision logic without some knowledge of how analytical tools or sampling strategies will work in the field. The DMA provides the information necessary to develop site-specific decision logic.

- **Evaluate existing contract mechanisms:** The results of the DMA are useful for evaluating existing contract mechanisms to determine if they are suitable for a dynamic work strategy.

- **Optimize sequencing, staffing, load balance, unitizing costs:** A DMA gives insight into logistical requirements. Understanding throughput and other logistics will allow the project team to balance personnel and optimize field efforts.

Choosing Samples for a DMA

- Ideally across the expected range of concentrations with a focus around action levels
- Limited by difficulty in obtaining material (depth, drilling)
- Multiple matrices?
- Problematic, interesting, or strange make a great choice
  - »Remember dynamic range issues

4-51

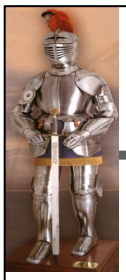
Notes

- **Ideally across the expected range of concentrations with a focus around action levels:** Determining which samples to split or analyze in the laboratory is a critical question. In addition, the method for splitting samples is just as important (for example, cups for XRF and then sent to a laboratory). If the XRF is mobilized to the site, it is best to collect a large number of samples that span the range of concentrations and focus around the action level.

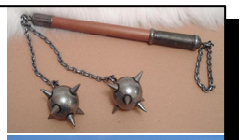
- **Limited by difficulty in obtaining material (depth, drilling):** XRF sampling is only limited by the difficulty in obtaining samples but the project team can conceivably evaluate 100 samples to choose the best 20, that span the expected range of concentrations, for analysis at the laboratory (for example, 5 high, 5 low, 10 around action levels). In most cases, some percentage of samples still will be required or collected (5-20 percent, to confirm excavation).

- **Multiple matrices:** In the case of applications with obviously different matrices (sand, soil, sediment, tailings), it is wise to evaluate each separately.

- **Problematic, interesting, or strange samples make a great choice:** The advantage of the real time information that XRF supplies is that as the sampling program progresses the project team can identify problematic, interesting, or strange samples for additional XRF measurements or targeted collaborative laboratory analysis. Remember the linearity issues associated with the instrument that were explained in Module 2; at percentage levels (10,000 ppm or greater) the XRF will have a low bias.



Your Quality Control Arsenal . . .*Weapons of Choice. . .*



- Energy calibration/standardization checks
- NIST-traceable SRM, preferably in media similar to what is expected at the site
- Blank silica/sand
- Well-characterized site samples
- Duplicates/replicates
- In-situ reference location
- Matrix spikes
- Examination of spectra

4-52

Notes

- **Your quality control arsenal:** Calibration checks serve several purposes:
  - » they identify whether the XRF unit is initially properly calibrated (provides unbiased measurements for the elements of concern in the range of concern),
  - » they are used over time to make sure the calibration holds,
  - » they can be used to identify and quantify potential interference effects. The latter is typically done with matrix spikes (e.g., sample spiked with two contaminants of concern at known concentrations that may interfere with each other from an XRF perspective) and/or using well characterized site-specific samples with known elevated concentrations of elements that are suspected to pose potential interference concerns for the XRF.

With all calibration QC, it is important that concentrations present in matrices used for calibration checks are in the range where decisions will be made. Concentrations that are too low may either be non-detectable or have so much measurement error associated with their results that their use as calibration checks are compromised. Concentrations that are too high may well fall outside of the linear calibration range of the instrument.

Blanks ensure clean window and minimize false positives.

Standardization or Energy Calibration

- Completed upon instrument start-up or when instrument identifies significant drift
- X-rays strike stainless steel plate or window shutter (known material)
- Instrument ensures that expected energies and responses are seen
- Follow manufacturer recommendations (typically several times a day)

4-53

Notes

- **Completed upon instrument start-up or when instrument identifies significant drift:** Instrument start up seems simple enough but most manufacturers and experienced users recommend the instrument should be allowed to "warm-up" for 20-30 minutes before beginning initial calibration checks. During this time x-ray generation and the detector components and instrument temperature stabilize. Significant temperature swings can sometimes impact instrument performance. Some Innov-X units will occasionally require re-standardization during operation if the software detects drift. The actual standardization only takes a minute and is followed by the running of a series of blanks, SRMs, and site-specific calibration standards (SSCS) to ensure everything is in control before continuing on with sample analysis.

- **X-rays strike stainless steel plate or window shutter (known material)**

- **Instrument ensures that expected energies and response are seen:** The purpose of this procedure is to perform an energy calibration so that the x-ray peaks will be located in the proper channels and the correct intensities (counts) will be recorded for each region of interest (ROI). Thus ADC channel number is calibrated in terms of energy or kiloelectron volts (keV) and the x-ray peaks show up where they are expected to be in the spectrum.

  The software looks for x-ray counts data for each analyte in a specific ROI. When properly calibrated, the centroid of the ROI will correspond to the desired x-ray line energy (usually expressed in keV). The energy calibration is accomplished by collecting a spectrum of a reference sample with distinct x-ray lines; one at low energy (keV) and one at high energy over the useful range of the detector. For example, a sample containing iron may be measured using a

silver anode x-ray tube; the resultant spectrum will contain a strong iron K-alpha line at 6.4 keV as well as a strong scattered tube line (silver K-alpha) at 22.1 keV. The energy calibration routine locates these lines and determines their centroids in terms of channel number. Next, the difference in energy is divided by the difference in centroid channel number resulting in keV/channel. The channel number (fraction) corresponding to "zero" energy is generally also determined (typically a number very close to zero). Thus ADC channel number is calibrated in terms of energy (keV) and the x-ray peaks show up where we expect them to be in the spectrum.

- **Follow manufacturer recommendations (typically several times a day):** The project team should follow the manufacturer recommendations for frequency of standardization or energy calibration.

Initial Calibration Checks

- Calibration SRMs and SSCS typically in cups
- Perform multiple (at least 10) repetitions of measuring a cup, removing the cup, and then placing it back for another measurement
- Compare observed standard deviation in results with average error reported by instrument
- Compare average result with standard's "known" concentration
- Use observed standard deviation for evaluating controls for on-going calibration checks (DMA)

4-54

Notes

This initial calibration check is a little more labor intensive than the continuing calibration checks, but it is important.

- **Calibration SRMs and SSCS typically in cups:** The initial calibration also can include the running of a series of blanks. SRMs and SSCs are typically run using cups.

- **Perform multiple (at least 10) repetitions of measuring a cup, removing the cup, and then placing it back for another measurement:** Multiple measurements are performed by measuring a cup, removing the cup, and then placing the cup back for another measurement. This should be repeated until at least 10 measurements have been made.

- **Compare observed standard deviation in results with average error reported by instrument:** For SRMs or SSCS, the expectation is that through a series of repetitions of at least 10 or more, a quick spreadsheet check of the standard deviation of those repetitions for each element of concern should be around the observed average error reported by the instrument. The instrument provides an error for each result (it is important to note if the instrument reports 1 or 2 standard deviations of the counting statistics). The average error reported by the instrument should closely match the SD of the 10+ measurements. The values do not have to match exactly but, for example, an average error from the instrument that is approaching ½, or in the other extreme twice, the SD of the repetitions, would be a flag that there may be calibration issues.

- **Compare average result with standard's "known" concentration:** The average result of the repetitions should also closely match the "known" concentration. So in the case of the NIST SRMs, the certified values, and in the case of SSCS, the average or expected value of the well characterized sample. Make sure XRF performance in relation to SRMs and SSCS is well understood initially. Even if one element reads slightly high or low for example, watch for trends as the program progresses.

- **Use observed standard deviation for evaluating controls for on-going calibration checks:** Assuming the results do closely mirror each other, then the observed SD can be used to set expectations for SRM and SSCS variability moving forward. The observations should mirror the DMA data set and the 2SD/3SD control limits used for control charts (which will be discussed later). In terms of the SSCS, evaluate:

  1) what detection limit performance can be expected

  2) what measurement times are required to get acceptable detection limits

  3) the presence of elements in background that may compromise system performance

Initial Calibration Check Example

Sample	# of Measurements	Known		Reported	
		U	Moly	U	Moly
SiO2 Blank	1	<LOD	<LOD	<LOD	<LOD
50 ppm U	3	**50**	NA	**<LOD**	14
150 ppm U	3	**150**	NA	**116**	23
50 ppm Moly	3	NA	50	55	42
150 ppm Moly	3	NA	150	<LOD	134
100 ppm U/Moly	6	**100**	100	**68**	112
Archived Site Sample	10	**100**	NA	**230**	21

4-55

Notes

- In many applications the unit may be rented, borrowed, or received from the manufacturer. When renting a unit, understand that the previous operator probably did not handle the unit with "kid gloves" or that those in charge of maintenance did not tune that instrument to point where it is running like new. This example shows why it is important to do an initial calibration check. Even if the unit is owned, it is still a good idea to perform this evaluation.

  The particular instrument in question here, had a "standardless" calibration done by the factory. The two primary elements of concern at the site in question were uranium and molybdenum. A blank was obtained, along with five spiked standards. There also was one well-characterized historical sample available.

  The initial check was not good…detection limits for uranium appeared to be significantly higher than expected, uranium results were different than the standards' reported values, and most importantly there was a huge discrepancy between the archived sample laboratory result for uranium and what the XRF was measuring. Ultimately it turned out to be a bad instrument or factory calibration that required replacement.

Continuing Calibration Checks

- At least twice a day (start and end), a higher frequency is recommended
- Frequency of checks is a balance between sample throughput and ease of sample collection or repeating analysis
- Use a series of blank, SRMs, and SSCS
- Based on initial calibration check, how is XRF performing?
- Watching for on-going calibration check results that might indicate problems or trends
  - » Typically controls set up based on DMA and initial calibration check work (a two SD rule)

4-56

Notes

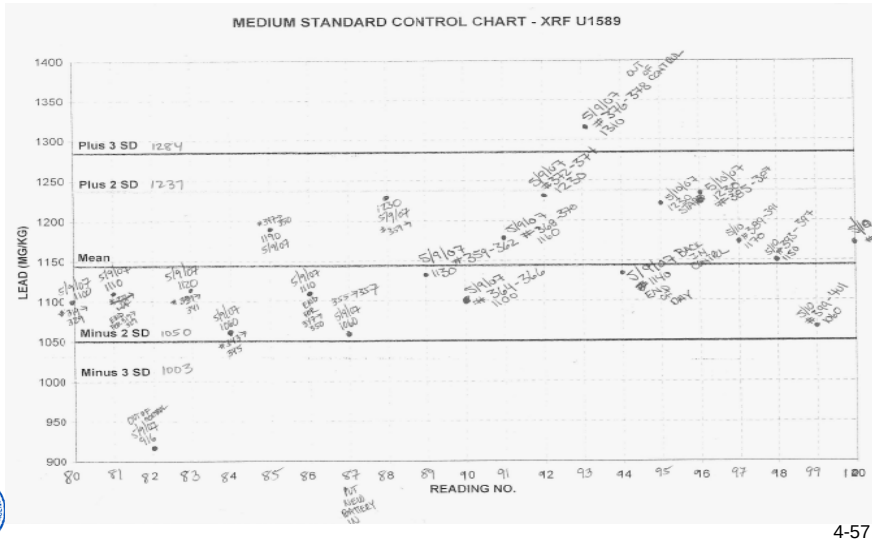
- **At least twice a day (start and end), a higher frequency is recommended:** Most projects use a higher calibration check frequency than one at the start of the day and one at the end of the day. Continuing calibration is recommended as often as every 10 or 20 samples. If an out of calibration situation is encountered then all data collected since previous check is "in question."

- **Frequency of checks is a balance between sample throughput and ease of sample collection or repeating analysis:** Balance between the time it takes to run QC checks and the possibility of losing large of amounts of data to QC problems. Weigh the time to run SRMs, SSCS, and blank against loosing or qualifying data. Determine how easy is it to reproduce the data or re-analyze samples. If samples are archived a lower frequency of checks may be acceptable. If collecting large quantities of in-situ shots per day, a higher frequency might be warranted particularly if remedial decisions are being made based on XRF results.

- **Use a series of blank, SRMs, and SSCS:** Use a series of a blank (usually quartz or sand), SRMs like NIST 2709 (low), 2710 (high), and 2711 (mid), and SSCS if available. Ideally, try to bracket the range of expected values.

- **Watching for on-going calibration check results that might indicate problems or trends:** Determine if the XRF is usually higher or lower than the known SRM value, and by what percent or PPM value. The next several slides illustrate why this understanding is important and how it can be used to spot "problems" with instrument performance over time.

Control Charting Your Continuing Calibration Checks



4-57

Notes

- There is an example of a spreadsheet developed as part of the XRF toolbox. This is an electronic version but simple handwritten and plotted results work just as well.

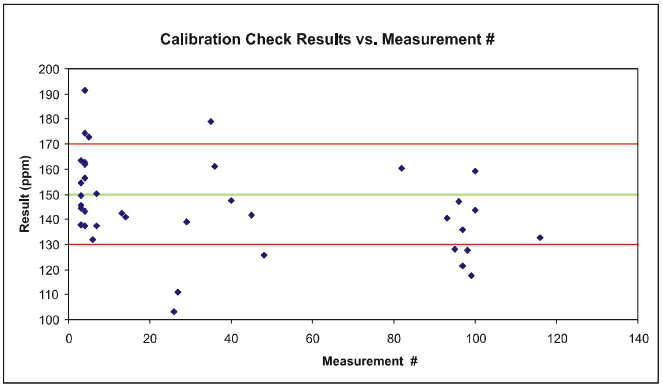
In general a control chart is used:

1. Use the DMA results and/or ICAL results to generate summary statistics for key COCs based on SRM and SSCS values. If Pb and As are of interest, 6 of these charts may be generated (As 2709, 2710, 2711 and Pb 2709, 2710, 2711). Another option is to choose only 1 SRM (for example 2709) that is in the range of interest and a SSCS.
2. Using the summary statistics provides an expected mean (red line) and SD, (2XSD (purple), 3XSD is green).
3. Plug in new values as continuing calibration checks are performed.
4. Take your initial reading, if the measurement exceeds 2 SD, take it again (using a 95 UCL you would expect 5 out of 100 readings to be outside the ±2SD value).
5. If the second reading is within the 2SD of the mean, continue and monitor for trends.
6. If the second reading is still outside the 2SD window, initiate corrective action (change battery, re-initialize, re-run SRMs and QC).
7. Likewise, any 1 reading that is outside the 3SD window would require corrective action.

Again, the value in doing this is that out of control situations can be identified as they happen and limit the number of samples that may need to be re-run. If continuing calibration is only done at the start and end of each day, and for example, one of your analytes is >3SD at the end of the day, all samples from that day will need to be re-run.

Continuing Calibration Checks
Example of What to Watch for…

- Two checks done each day, start and finish
- 150 ppm standard, w/ approx. +/- 9 ppm for 120 second measurement
- Observed standard deviation in calib check data: 18 ppm
- Average of initial check: 153 ppm
- Average of ending check: 138 ppm



4-58

Notes

- This data show what can happen to an XRF. This was a tube-based system with calibration checks performed at the start of the day and at the end. The initial measurement of the day was measurement "1," with the measurement # incrementing as the day went on and measurements were made. A 150 ppm standard was being used for the checks. The reported error for a 120 second measurement on the standard was approximately 9 ppm. The first indication that something was not right was when we were watching the SD associated with the check data as data accumulated…that SD stayed around 18 ppm, or twice what the reported measurement error was for each measurement.

  We then checked the average calibration result for start-of-day readings and end-of-day readings. The average start-of-day reading was 153 ppm, or almost spot on the calibration concentration. However, the average end of day reading was 138 ppm, about 10 percent below our calibration concentration. Graphing our calibration check results as a function of measurement number underscores the problem…the calibration check result falls as the number of measurements taken prior to the calibration check increases. Fitting a regression line through these data suggests that we were losing about 10 percent off of our calibration for every 100 measurements made, an effect most likely attributable to the battery degradation as the day wore on (the battery was recharged each night). The obvious fix … switching batteries after a set number of measurements.

Interference Effects

- Spectra too close for detector to accurately resolve
- Result: biased estimates for one or more quantified elements
- DMA, manufacturer recommendations, scatter plots used to identify conditions when interference effects would be a concern
- "Adaptive QC"…selectively send samples for confirmatory laboratory analysis when interference effects are a potential issue

4-59

Notes

- **Interference effects:** Another advantage to completing a DMA is to determine if other metals present at the site indicate the potential for interference issues. Interference issues can occur when detector resolution is not sufficient to determine if response at a particular energy level range (keV) is due to one element or another. All elements have multiple energy level responses depending on whether the electron that was struck by the x-ray was ejected from the inner K shell or the slightly higher state L and M shells. The primary photons generated by the x-ray tube or radioisotope sources generally have enough power to eject electrons in the 3 inner most shells. As a higher energy outer shell electron moves to fill the inner shell vacancy it also produces two characteristic x-rays of two spectral regions (alpha and beta lines).

Potential Interferences

Element	keV	Shell	Region (keV)		Possible Interference
			Min	Max	
P	2.02	Ka	1.85	2.2	**Y**, **Zr**, Nb, S
S	2.31	Ka	2.2	2.45	**Nb, Mo**, Tc, P,Cl
Cl	2.62	Ka	2.45	2.82	Tc, **Ru**, Rh
K	3.31	Ka	3.12	3.42	Ag, **Cd, In**
Ca	3.69	Ka	3.54	3.87	Sb, Te
Ti	4.51	Ka	4.3	4.7	Sc,Cs,**Ba**
Ba	4.83	Lb	4.7	5.1	**V**, Ti, La, Pr
Cr	5.41	Ka	5.2	5.6	**Pm,V**
Mn	5.9	Ka	5.7	6.1	Eu, Cr
Fe	6.4	Ka	6.2	6.6	Gd, Tb, Mn
Co	6.93	Ka	6.73	7.13	**Er,** Ho, Fe
Ni	7.48	Ka	7.28	7.68	Yb, Ho
Cu	8.05	Ka	7.85	8.25	Hf, Tm
Zn	8.64	Ka	8.44	8.84	Re, Lu
W	9.67	Lb	9.5	9.9	**Au**, Ge
Hg	9.99	La	9.8	10.2	Re, Ge, As
As	10.54	Ka	10.3	10.7	**Pb**

(continued)

4-60

Notes

- The table in this slide provides interference information. The periodic table x-ray energy references by Niton and Innov-X also are good sources of information. They include excitation energies for various spectra lines and source selection for isotope instruments.

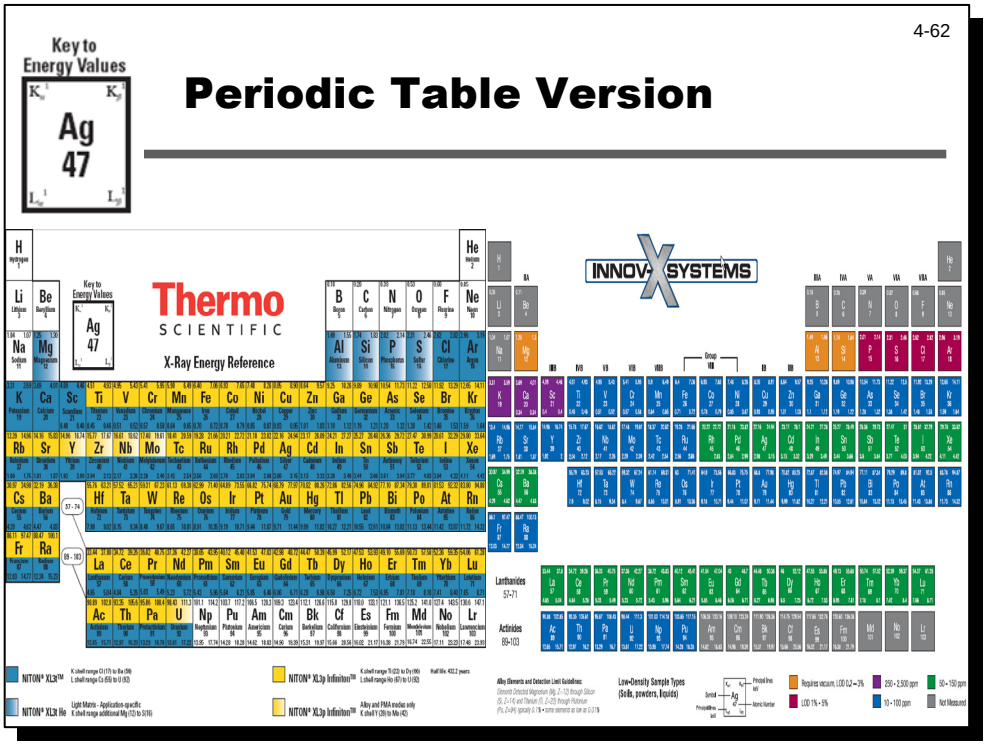
Potential Interferences

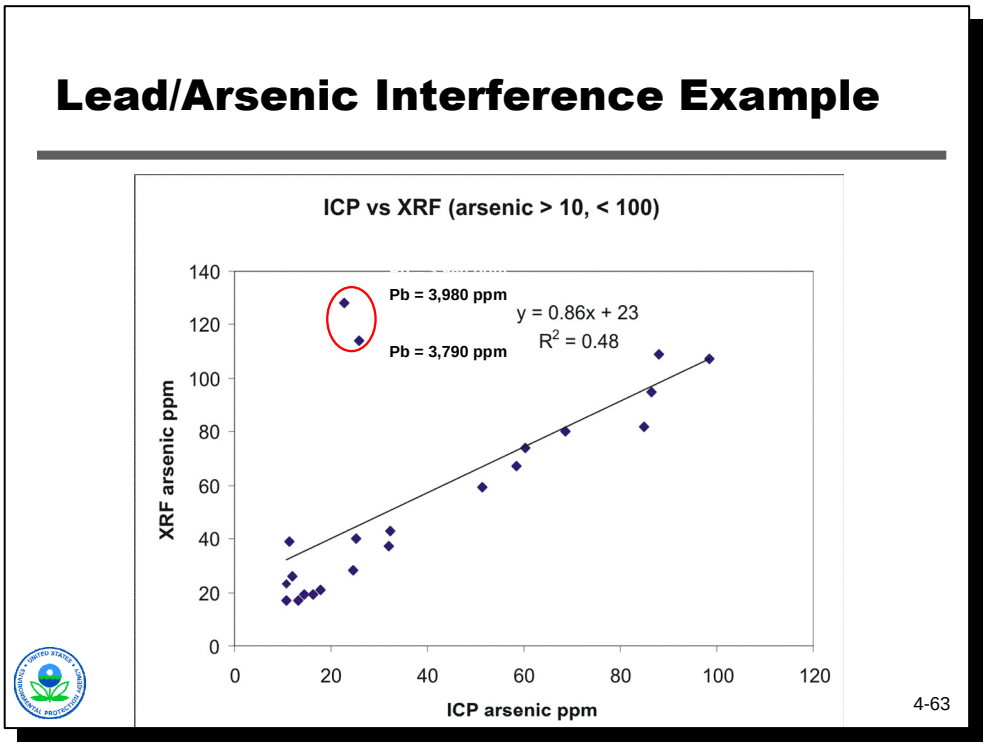
Element	keV	Shell	Region (keV)		Possible Interference
			Min	Max	
Se	11.22	Ka	11	11.4	Po
Au	11.44	Lb	11.3	11.675	**At**
Br	11.92	Ka	11.7	12.1	Hg, Fr
Pb	12.61	Lb	12.4	12.8	**Ac**, **Kr**, Se
Bi	13.02	Lb	12.85	13.15	
Rb	13.39	Ka	13.2	13.6	Pa, Po, Br
U	13.61	La	13.5	13.9	
Sr	14.16	Ka	13.9	14.3	Kr
Zr	15.77	Ka	15.4	16	Sr, Cf, **Ac**
Mo	17.48	Ka	17.1	17.7	
Ag	22.16	Ka	21.804	22.404	
Cd	23.17	Ka	22.81	23.14	
Sn	25.27	Ka	24.894	25.494	
Sb	26.36	Ka	25.974	26.574	

4-61

4-62

Periodic Table Version





Notes

- Many of the classic issues with linear regressions discussed in Module 2 are illustrated in this linear regression.

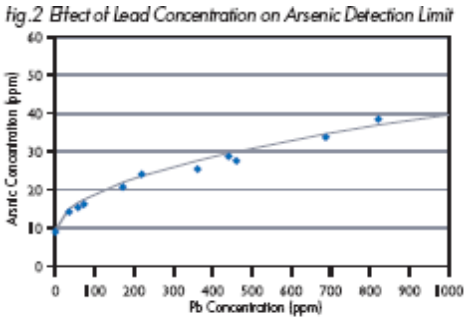
  The above linear regression is poor. Not only is the R^2 value poor at 0.48 but the regression suffers from many of the issues discussed in Module 2 including: correlated residuals, slope not close to 1 (0.86), and inexplicable outliers.

  An examination of the 2 points that seem to be driving the arsenic correlation reveals high concentrations of lead. The inability of the instrument to resolve peaks in the 10.5 keV range results in some portion of the lead contamination likely being attributed to arsenic.

Arsenic in the Presence of Lead One Vendor's Answer

Algorithm predicts lead Lα in 10.5 keV spectral region based on the "clean" lead Lβ signal. The lead contribution is subtracted leaving the arsenic Kα.

$$As\big|_{Pb} = As\big|_{noPb} + \sqrt{Pb(ppm)}$$



4-64

Notes

- The Innov-X software algorithm automatically corrects the arsenic result when lead is present. The algorithm predicts the contribution in the 10.5 keV spectral region from the lead La based on the interference-free measurement of the lead Lb. The lead La contribution is subtracted, yielding the peak intensity due solely to the arsenic Ka. However, the precision of the arsenic result (and the detection limit in the case of low arsenic concentrations) are affected because the statistical uncertainty of the lead La background subtraction yields a less precise result for the arsenic concentration. This effect does not occur if there is negligible lead present in the spectrum. The impact on both As detection limit and precision can be determined.

  The arsenic detection limit as a function of lead concentration is presented in Figure 2. Based on x-ray measurement statistics, the As detection limit increases as the square root of the increase in lead concentration, following the functional form in the equation.

  Example: If lead concentration = 500 ppm, square root = 22, arsenic DL = 10ppm around zero so arsenic DL with 500 ppm lead is around 32 ppm.

The Skeptical Chemist….

- Difficulty in resolving As concentrations when Pb was greater than 10X the As
- *"10 Times Rule" empirical rule of thumb*
  - » "J" any XRF detected values for arsenic below 1/10 of the lead value
- Example
  - » Pb detected at 350 ± 38 ppm
    - — As detected at 28 ± 6 ppm report as estimated "J"
    - — As detected at 48 ± 10 ppm would not require a "J"

4-65

Notes

- **The skeptical chemist:** Dr. Dennis Kalnicky is one of the contributors to EPA's XRF efforts and this course. Dr. Kalnicky is a chemist with Lockheed Martin and he works with EPA ERT in Edison, New Jersey. He has a wealth of XRF experiences (25+ years as a chemist and 15+ with these types of units).

  The premise is this slide is ……..

  When XRF measures As in the presence of Pb, there is a very serious overlap of the Pb L-alpha line with the As K-alpha line, which must be resolved by some type of overlap correction. The accuracy of that correction is extremely dependent on the reproducibility of the energy calibration for the unit. Even a small error in energy calibration can have large effects on the accuracy of As in the presence of much higher Pb. Now, it is entirely possible to statistically calculate uncertainty for the As concentration in this case, but almost impossible to factor in errors that may be due to slight energy calibration shifts. Innov-X in fact uses the statistical RL value for reporting As even in the presence of very much higher lead concentrations and as noted previously, this does not account for possible errors due to slight energy calibration variations.

  Discussions among course instructors about this issue indicated that some XRF users would typically not report XRF results detected for arsenic where the lead concentration in the same sample was 10X greater than that value because some felt very strongly that there was significant uncertainty around those DLs and reported values. In the case of NDs, some felt it was unwise to assume that the statistical uncertainty represents the detection level because it does not account for these slight energy calibration errors. In this case some would use the larger of the statistical RL provided by the instrument or 1/10 the Pb

concentration as the arsenic RL. In cases when both analytes are above the reporting level (RL, typically 3-5 times the MDL), if the Pb level does not exceed the As concentration by more than 10-times, the As data should be reliable.

Module 2 indicated that when determining mean concentration for decision units it may be best to have the instrument report a number even if below the calculated DL or RL rather than using ND, less than, or substituted value. In this case of As and Pb, this presents a problem. It is still possible to report values greater than the statistical RL and less than 1/10 the Pb concentration. Therefore, we now label values as estimated when the XRF detects As at less than 1/10 of the lead concentration and thus the 10X rule was born.

This is an empirical rule of thumb based on 15+ years of experience with field portable XRF analyzers. As with any rule of thumb, it may not apply in all cases and may change as technology improves. It is ***not*** a published rule of thumb, but it is a reasonable compromise until XRF technology can overcome these spectral interferences.

Monitoring Detection Limits

- Detection limits for XRF are not fixed for any particular element
- Measurement time, matrix effects, the presence of elevated contaminants…all have an impact on measurement DL
- Important to monitor detection limits for situations where they become unacceptable and alternative analyses are required

4-66

Notes

- **Monitoring detection limits:** Count times used are often 120 seconds but depending on site specifics, data needs, and operation or instrument mode many instruments can be run from 10-360 seconds or more. Performing a DMA can provide information to determine the best operating procedures for a specific site. Error readings associated with each analyte indicate how count times affect precision.

Monitoring Detection Limits

Analyte	Chemical Abstract Series Number	Innov-X[1] 120 sec acquisition (soil standard – ppm)	Innov-X[1] 120 sec acquisition (alluvial deposits - ppm)	Innov-X[1] 120 sec acquisition (elevated soil - ppm)
Antimony (Sb)	7440-36-0	61	55	232
Arsenic (As)	7440-38-0	6	7	**29,200**
Barium (Ba)	7440-39-3	NA	NA	NA
Cadmium (Cd)	7440-43-9	34	30	**598**
Calcium (Ca)	7440-70-2	NA	NA	NA
Chromium (Cr)	7440-47-3	89	100	**188,000**
Cobalt (Co)	7440-48-4	54	121	766
Copper (Cu)	7440-50-8	21	17	661
Iron (Fe)	7439-89-6	**2,950**	**22,300**	**33,300**
Lead (Pb)	7439-92-1	**12**	8	**447,000**
Manganese (Mn)	7439-96-5	56	**314**	1,960
Mercury (Hg)	7439-97-6	10	8	481
Molybdenum (Mo)	7439-93-7	11	9	**148**
Nickel (Ni)	7440-02-0	**42**	31	451
Potassium (K)	7440-09-7	NA	NA	NA

4-67

Notes

- This table was shown earlier, and is shown here again to illustrate the fact that detection limits for any particular instrument/element combination can vary widely.

  The last sample is an example of elevated detection limits for multiple elements based on high levels of nearby element response peaks. Also the likelihood that the detected values exceed the linear range of the instrument is high. Depending on the decision (highest concentration for risk) this sample would be a good choice for laboratory analysis.

  As a simple part of the QC program, watch for unacceptable detection limits. Although in this case, strictly from a decision perspective, although there may be an action level for antimony that is less than 232 ppm or for cadmium that is less than 598 ppm, it is likely not necessary to use these metals for a decision because As is 29,000 ppm and Pb is 45 percent. Note that the Pb levels are likely outside of the dynamic range for this instrument and even though the 45 percent value is uncertain, the sample is highly contaminated and would trigger a removal or remedial action.

Monitoring Dynamic Range

- Periodic, in response to XRF results exhibiting characteristics of concern (e.g., contaminants elevated above calibration range of instrument)…sample sent for confirmatory analysis
  - »Is there evidence that the linear calibration is not holding for high values?
  - »Should the characteristics used to identify samples of concern for dynamic range effects be revisited?

4-68

Notes

- **Monitoring dynamic range:** This is a good idea to monitor but it must be kept in context of the decision. If the action level is 400 ppm and results appear linear through an order of magnitude (4,000 ppm), then some loss of some linearity >4,000 ppm does not matter in the context of a clean versus dirty decision.

Matrix Effects

- In-field use of an XRF often precludes thorough sample preparation
- This can be overcome, to some degree, by multiple XRF measurements systematically covering "sample support" surface
- What level of heterogeneity is present, and how many measurements are required?
- "Reference point" for instrument performance and moisture check with *in-situ* applications

4-69

Notes

- **Matrix effects:** Matrix effects can significantly impact your project.

  Multiple XRF measurements can be individual readings to identify hotspots within the sample support area and develop a statistical mean for the sample support area. Sample preparation techniques to align particle size with decisions units and aggregate measurements can also be used to deal with matrix heterogeneity effects.

  The reference point is a single well defined and marked location that can be continuously accessed during the sampling event. Returning to this location daily or after rain events allows performance of the instrument to be monitored over time and after rain events for *in-situ* applications.

  In response to XRF results of concern (e.g., elevated lead when arsenic is the principal contaminant of concern)…send for confirmatory analysis. Determine if the XRF results are compromised by interference effects. Evaluate whether characteristics used to identify samples of concern for interference effects should be revisited. Always moisture check sample after rain events. Generally, it is wise to determine the characteristics of concern during the DMA and have field based action levels that trigger collection of collaborative data (i.e., ICP or AA analysis). Determine if results are too close to call clean or dirty.

Worried About Impacts From Bags?

- We've evaluated a variety of bags and found little impact
- Analyze a series of blank, SRM, and SSCS by analyzing replicates or repetitions through the bag
- Exceptions include bags with ribs and highly dimpled, damaged, creased bags
  - »Result in elevated DLs, reported errors

4-70

Notes

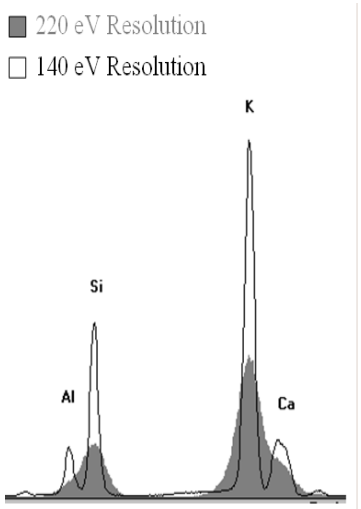
- **Worried about impacts from bags:** EPA has evaluated these impacts by shooting a series of analyses through the bag including different areas of the bag like clear areas and label areas. Based on the DMA and initial calibration, there should be a good understanding of the expected errors and variability associated with your QC samples that have been analyzed in cups through a Mylar film. If there is a potential impact to data quality from the bags then get different bags (thinner and clearer are better).

  The obvious explanation for instances where we have noticed some impact to instrument performance is the scattering of x-rays in non-flat and/or damaged surfaces.

  At one recent technical support site we did document a discernible impact, from plastic bags. By itself it might not have been so bad, but when combined with somewhat high moisture, it caused the XRF data to be significantly biased low. Plastic interference can be checked by using the NIST materials. Shoot a series of SRM analyses on the cup normally, then cover with the plastic bag and reshoot to see if there appears to be any bias as a result of the plastic bag.

Examination of Spectra

- Spectral response is actually a range in the 100's of electron volts
- Resolution of latest detectors <190-230 eV
- Older models ~280-300eV
- Can use spectra to evaluate high NDs or errors for target metals



4-71

Notes

- **Examination of spectra:** Recalling the tables in slides 20 and 21, spectral response is actually a range (example As K alpha response in the range of 10.3-10.7 keV or about 400 electron volts). Another element with a K, L, or M alpha or beta response in that keV range may show up in the spectra.

Controlling Sample Heterogeneity
Estimating Measurement Number Requirements

- Goal: reducing error due to heterogeneity to at least 30% and at most comparable to analytical error at action level (<10%)
- Ten measurements from a "sample support" at action level to estimate variability
- Aggregate error = SD/sqrt(n) where n is the number of samples contributing to aggregate
- Typically takes from four to sixteen aggregated measurements to achieve

4-72

Notes

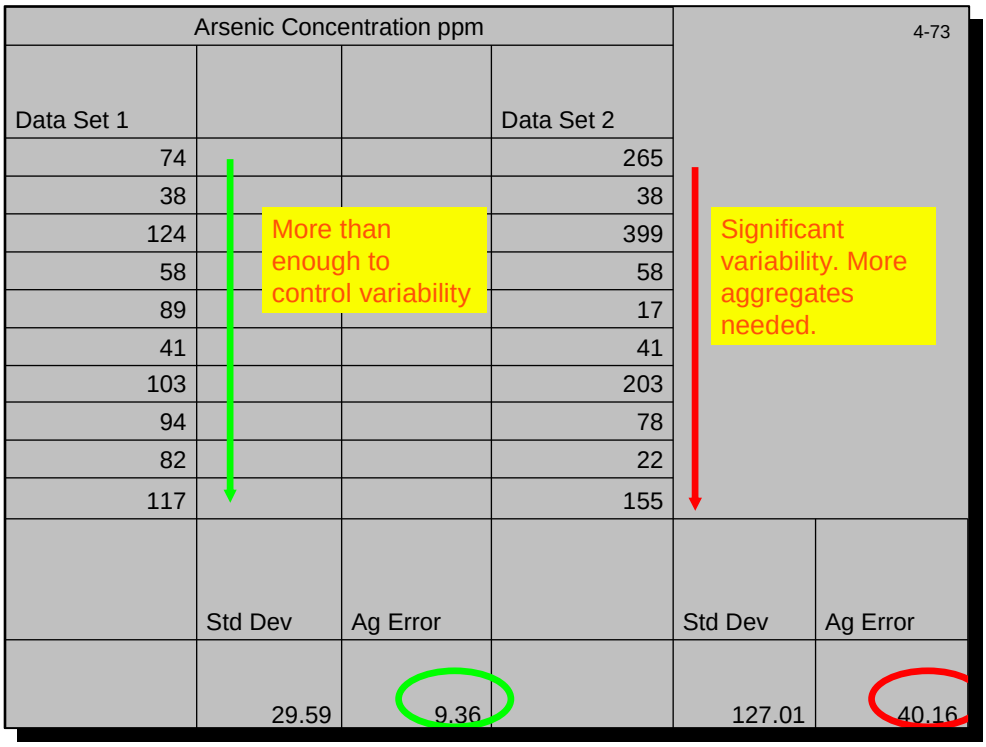
- **Controlling sample heterogeneity – estimating measurement number requirements:** Aggregate measurements are multiple short duration (30 second) readings across the sample support area to generate a mean concentration. Instrument returns a single "average" concentration for the aggregate measurements.

  To determine the appropriate number of "shots" in the aggregate mode, take 10 or more measurements from a sample support area to determine variability. The aggregate error is the SD / the square root of (n) where n is the number of shots contributing to the aggregate. Typically, it takes 4-16 shots to reduce the heterogeneity error to the 10-30 percent range.

  Initial estimates of heterogeneity at levels of concern and required for measurement aggregation (for *in situ* measurements or un-prepared sample measurements; repeated measurements systematically over exposed surface).

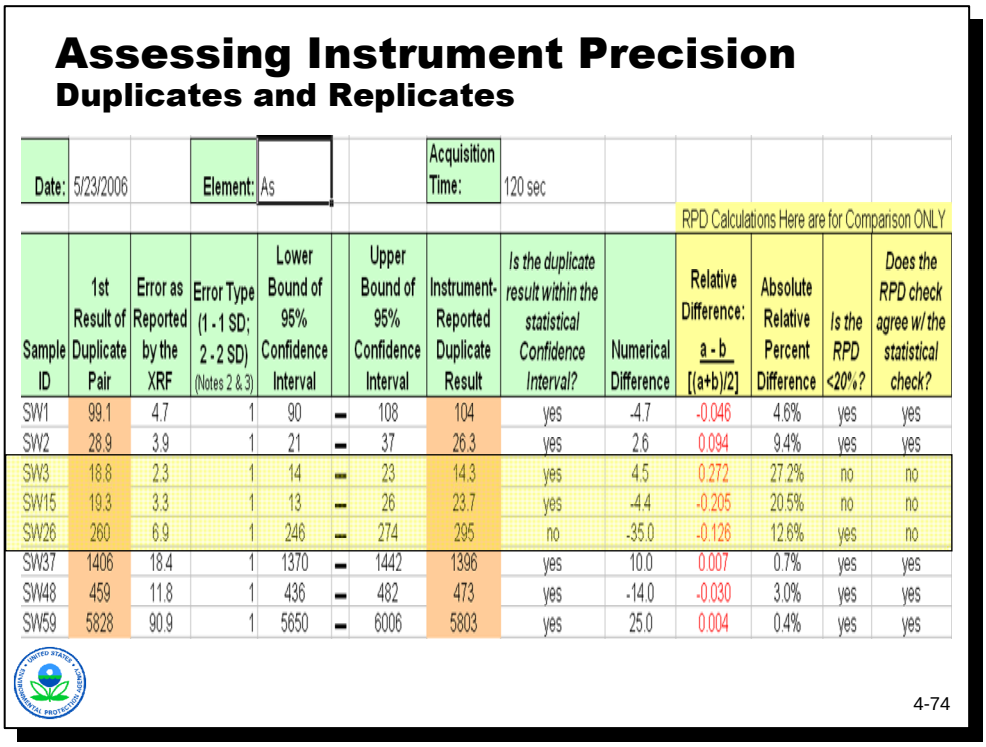
  1. How much measurement variability is presented attributable to within sample support heterogeneity?

  2. How many aggregated measurements are required to control heterogeneity effects?



Notes

- This is an example of aggregated measurements. Generally we do this with bagged samples.



Assessing Instrument Precision
Duplicates and Replicates

Date:	5/23/2006		Element:	As			Acquisition Time:	120 sec					
										RPD Calculations Here are for Comparison ONLY			
Sample ID	1st Result of Duplicate Pair	Error as Reported by the XRF	Error Type (1 - 1 SD; 2 - 2 SD) (Notes 2 & 3)	Lower Bound of 95% Confidence Interval		Upper Bound of 95% Confidence Interval	Instrument-Reported Duplicate Result	*Is the duplicate result within the statistical Confidence Interval?*	Numerical Difference	Relative Difference: a - b / [(a+b)/2]	Absolute Relative Percent Difference	*Is the RPD <20%?*	*Does the RPD check agree w/ the statistical check?*
SW1	99.1	4.7	1	90	–	108	104	yes	-4.7	-0.046	4.6%	yes	yes
SW2	28.9	3.9	1	21	–	37	26.3	yes	2.6	0.094	9.4%	yes	yes
SW3	18.8	2.3	1	14	–	23	14.3	yes	4.5	0.272	27.2%	no	no
SW15	19.3	3.3	1	13	–	26	23.7	yes	-4.4	-0.205	20.5%	no	no
SW26	260	6.9	1	246	–	274	295	no	-35.0	-0.126	12.6%	yes	no
SW37	1406	18.4	1	1370	–	1442	1396	yes	10.0	0.007	0.7%	yes	yes
SW48	459	11.8	1	436	–	482	473	yes	-14.0	-0.030	3.0%	yes	yes
SW59	5828	90.9	1	5650	–	6006	5803	yes	25.0	0.004	0.4%	yes	yes

4-74

Notes

- XRF instrument precision is measured through duplicates and replicates. This slide shows another spreadsheet from the XRF Toolbox. This was created as an alternative to using RPD as the sole means to assess duplicate agreement.

Spreadsheet for recording & assessing XRF instrument duplicate QC results

XRF instrument duplicates assess the reading-to-reading precision of the XRF instrument.

Instrumental precision can be affected by a low battery, extraneous material stuck on reading window, and operator mishandling of the instrument, such as slight shifting or tilting of the instrument during active counting periods.

BEFORE using this spreadsheet...Determine whether the instrument error value as reported by your XRF instrument represents either 1 or 2 SD for the instruments counting statistics (This information can be obtained from the instrument manufacturer).

Procedure

1. After taking the 1st shot of what will be the duplicate set, record the reported concentration value AND the error (reported by the XRF instrument for that reading) into the preprogrammed "Duplicate Calculator" sheet next to this Instruction sheet.

2. After taking the 1st reading, DO NOT MOVE the instrument from that spot!!

3. Enter the 1st instrument reading & its error into the spreadsheet calculator. Be sure to enter the error into the correct column for 1 or 2 SD. The spreadsheet will calculate the 95 percent Upper and Lower bounds for the CI around the 1st value. Note that the CI is based on the z-distribution, not the t-distribution. (The z-distribution is acceptable because the CI is based on instrument counting statistics, which have a normal distribution.)

4. Then take the next shot, but don't move the XRF instrument yet! Do not move the XRF until it is determined that the duplicates are within acceptable control limits.

5. Enter the 2nd reading into the appropriate column of the spreadsheet (under column heading of "Instrument Duplicate Result"). Determine whether the 2nd shot lies between the Upper and Lower Confidence Limits calculated from the 1st shot.

6. Answer the question of whether the duplicate QC result is acceptable.

If "yes," you may move the XRF and proceed to the next analysis.

If "no." DO NOT MOVE the XRF. Continue to Corrective Action below.

Corrective Action (to be taken if the 1st duplicate result is not within acceptable limits)

1) Take a 2nd duplicate reading, and determine If it is within the 95 percent confidence interval. If so, you can remove the XRF and move on.

2) If it is outside the confidence interval, run the NIST control(s), which is in an XRF cup.

3) If the NIST control(s) is(are) good, select another sample to perform the duplicate analysis (start from #1 above).

4) If the NIST control(s) is(are) out, troubleshoot the instrument by checking the battery, checking for cross-contamination or other possible problems. Once any problems are corrected, restandardize the instrument and rerun all controls to establish instrument performance.

5) If the NIST control(s) is(are) in and everything else about the XRF instrument is working ok, then the reason for poor instrument precision may be the operator.

Most likely the operator is not consistently holding the XRF steady in good contact with samples throughout the counting period. The corrective action is to counsel or retrain the operator, and verify that the operator is able to use the XRF properly to generate reproduceable results.

RPD calculations in the spreadsheet are for INFORMATIONAL purposes only.

The spreadsheet calculates the RPD between duplicates. The RPD calculation is for information only. Compare RPDs for various duplicate pairs at high and low concentrations. Observe that RPDs for low concentrations can be very high (and could exceed traditional RPD acceptance limits), even though the absolute difference between the 2 values is small.

Let's examine 3 samples to illustrate the differences, Samples SW3, SW15 and SW26. In the case of SW3 and SW15 the reported value for the duplicate is within the 95 percent confidence interval yet because the values are relatively low they would exceed <20 percent RPD criteria. The requirement of percent RPD < 20 is problematic for very low concentration samples because division by a low value causes the quotient to be high even when the numerical difference is minor.

In the case of sample SW26 we see that although the duplicate is outside the 95 percent confidence interval the values are sufficiently high and close enough together that the RPD is 12.6 percent. Remember though that a 95 percent CI means that 5 out of 100 (or 1 in 20) are expected to be outside this range; however if a measurement is repeated in triplicate the probability of both consecutive measurements being out of control without a problem existing is extremely low. In a case where we exceeded the LCI/UCI boundary we would shoot a 3 third analysis to determine the necessity of taking a corrective action.

This procedure is very similar to the control charting discussed on slide 5-22.

Duplicate Precision - Spreadsheet Procedure

- 1st shot of duplicate set- record concentration and error (1or 2 std dev)
  - » Spreadsheet calculates 95 UCL and LCL based on z-distribution (uses counting stats with a normal distribution)
- DO NOT MOVE INSTRUMENT- Take 2nd shot
  - » Is 2nd shot within 2 std dev (95 UCL/LCL)?
- Decision
  - » If not then take another replicate
    - — Within 95 UCL/LCL= under control, continue
    - — Outside 95 UCL/LCL=corrective action
- Corrective action- Run NIST standard
  - » If NIST is "in control" (2 std dev) then choose another duplicate sample and repeat
  - » If NIST is "out of control" (> 2 std dev) then troubleshoot (battery, cross contamination, operator issues?)
  - » After troubleshooting, re-standardize and run QC before proceeding

4-75

Field Based Action Levels
Part of Your QC Program?

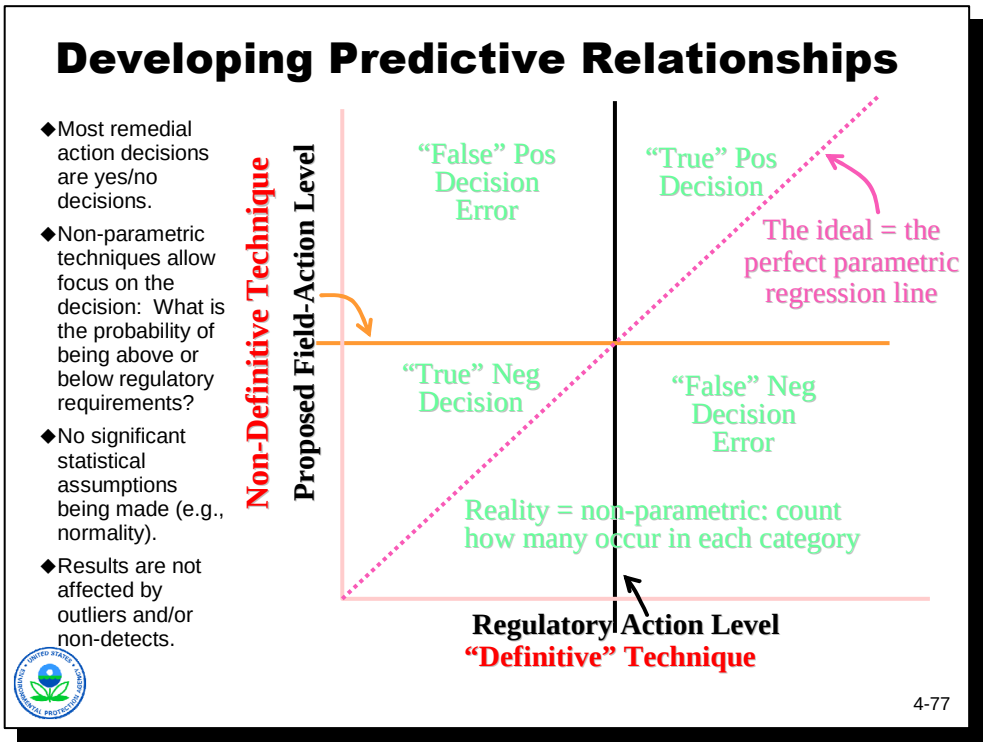
- As part of your QC program you will likely choose collaborative samples
  - » Helpful to ensure the quality of XRF decisions, monitor predictive relationships
- Some samples are obvious
  - » Spectral interferences, problematic matrices, samples outside instrument linear range, etc.
- Maximize the value of these samples
  - » Focus around action levels "too close to call"
  - » Some high, some low values
  - » Watch decision error rates

4-76

Notes

- **Field based action levels part of your QC program:** It is difficult to evaluate the comparability of ND pairs or 1 ND with detect and the information is of very little value. This course promotes moving away from a QC program that specifies that every 10th or 20th sample be sent for off-site laboratory confirmation. Instead, focus those samples where they will benefit the project most in terms of making good decisions.

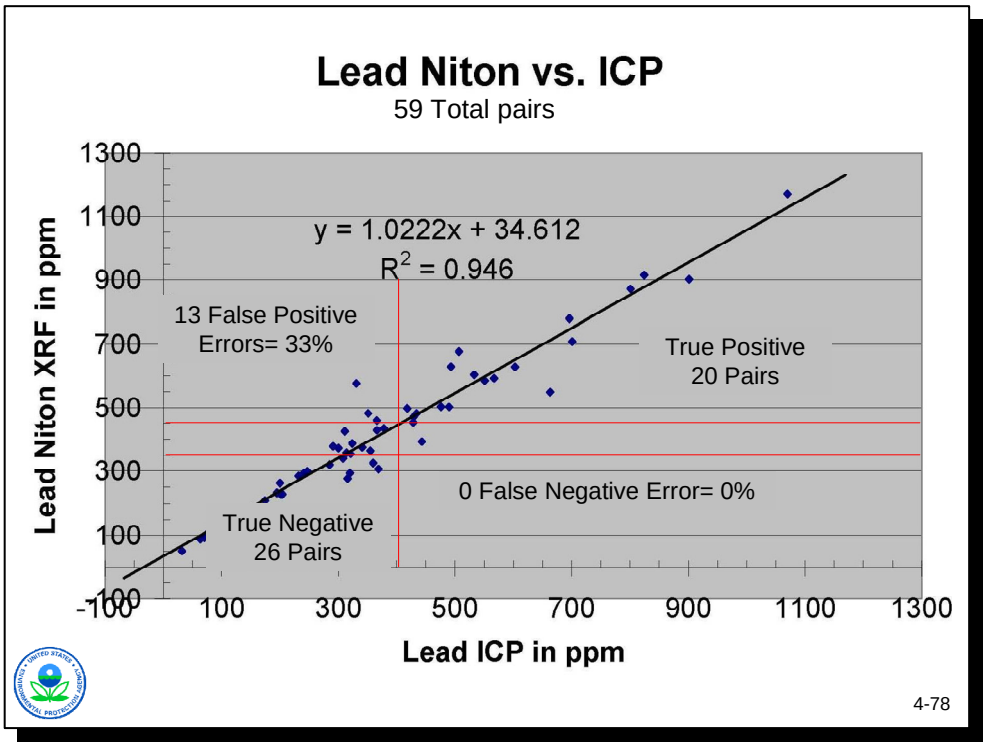
  It is still possible to maintain a laboratory split or collaborative sample frequency goal for a project, but the choices of samples for collaborative analyses is driven by the decision needs and uncertainties or variabilities as they unfold rather than solely on a frequency. That is one of greatest advantages of XRF, within seconds or minutes the project team has information that will determine their next move …… Maybe the sample is obviously clean, or obviously dirty, but maybe it is "too close to call," unusual, or different in some way that makes it a good candidate for ICP analysis. Under a more traditional approach, the project team likely would not have the information in a time frame that allows adjustments to the sampling scheme or QC frequencies.



Notes

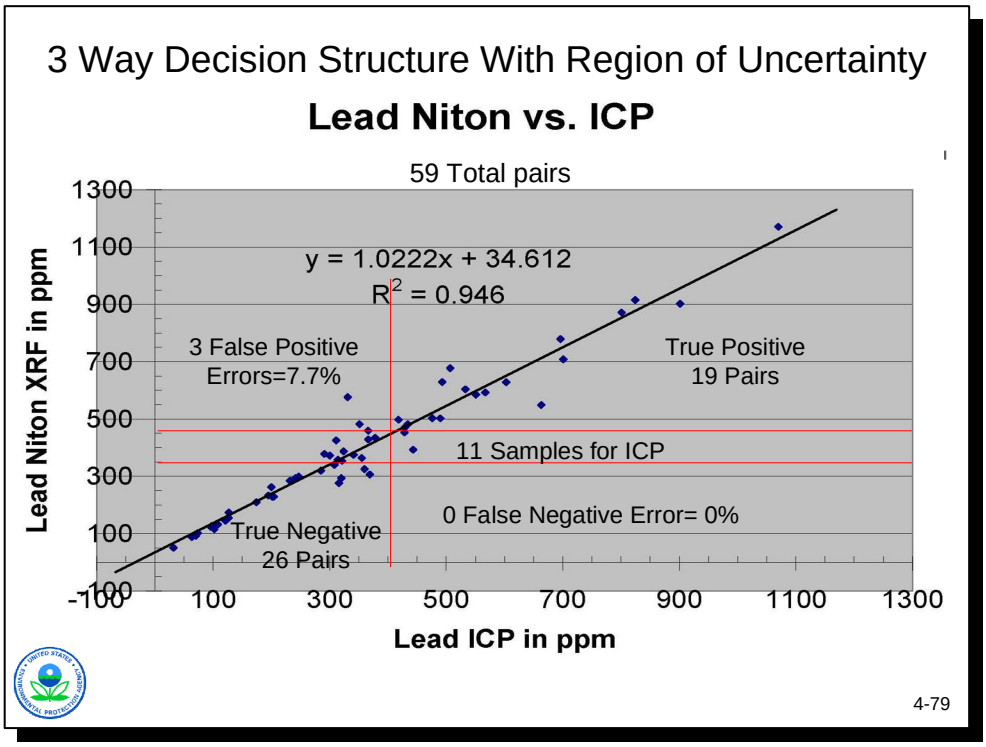
◆ **Developing predictive relationships:** Non-parametric methods don't make assumptions about underlying distribution of contaminants or use the estimation of parameters such as mean or standard deviation.

Non-parametric methods were developed to be used in cases when the researcher knows nothing about the parameters of the variable of interest in the population (hence the name *nonparametric*). In more technical terms, nonparametric methods do not rely on the estimation of parameters (such as the mean or the standard deviation) describing the distribution of the variable of interest in the population. Therefore, these methods are also sometimes (and more appropriately) called *parameter-free* methods or *distribution-free* methods.



Notes

- This is an example of a typical DMA product. It was used to develop field based action levels for XRF. This is a well correlated data set but the concept holds true for less "well behaved data sets." This uses a field based action level of 450 ppm.
- If there is greater concern about a Type I or false negative error, the action level could be reduced to 350 ppm, which decreases the false negative or "false clean" to 0. This would however result in a higher false positive or "false dirty" rate of 13 false positive errors or 33 percent. The project team would have to weigh the consequences of a false negative versus the costs associated with excavation or clean up at a rate of 33 percent false dirty. Many sites default to 5 percent error for false negative and 10 percent for false positive.



Notes

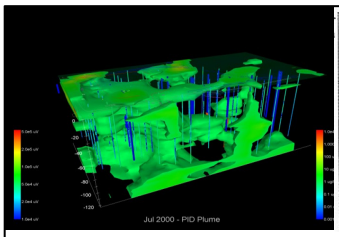
- This slide shows the structure of a 3 way decision. There are 19 true positives, 26 true negatives, 3 false positives, and 11 samples for ICP. The region of uncertainty is 350-450 ppm. Below 350 is definitely clean, above 450 is considered dirty while maintaining decision error rates less than 5 percent for false negatives and less than 10 percent for false positives.

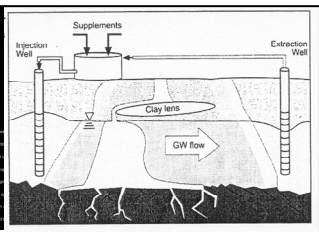
Q&A

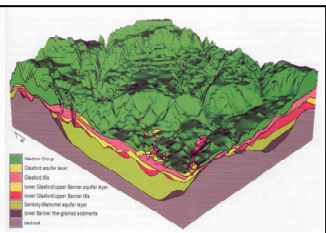
4-80

Module 5

Dynamic Work Strategies







Module 5:

Dynamic Work Strategies

5-1

Improving XRF Data Collection Performance Requires…

- **Planning systematically** (CSM)
- **Improving representativeness**
- **Increasing information** available for decision-making
- **Addressing the unknown** with dynamic work strategies

5-2

Notes

This module includes the following four broad areas:

- **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 denser, 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
- CSMs play a foundational role

5-3

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 also are 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.

- **CSMs 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

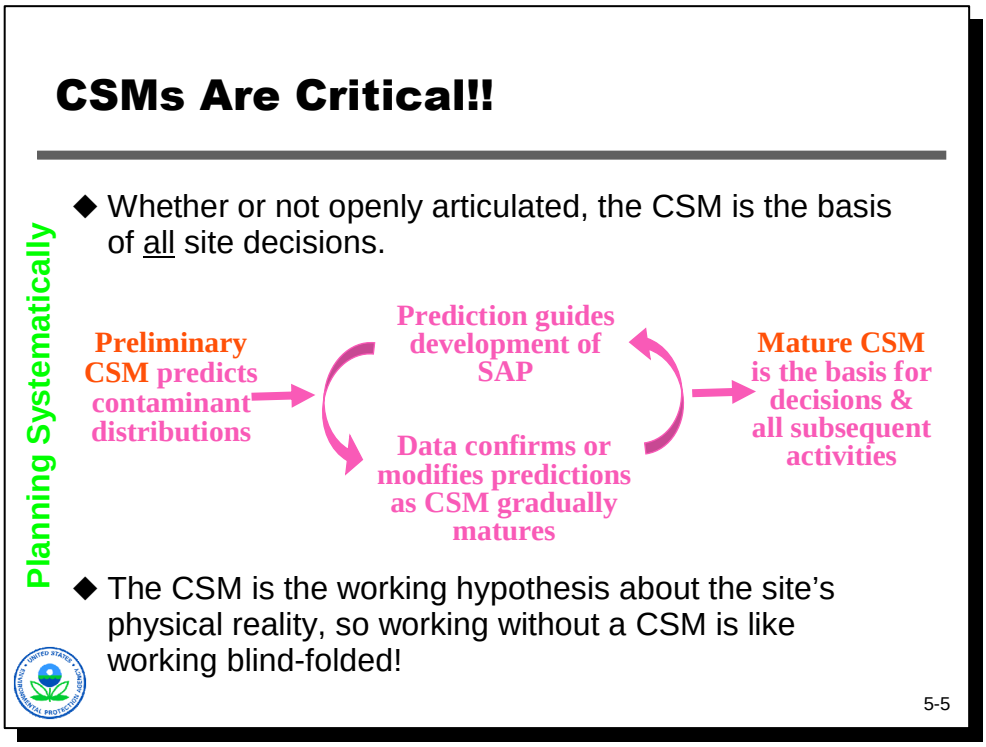
5-4

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



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.

CSM Uncertainty Drives Sampling Program Design

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

5-6

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.

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)
  - » Degree of short-scale (intra-sample) heterogeneity at action levels
  - » Degree of longer-scale (between sample) heterogeneity at action levels
  - » Vertical layering of contaminants

5-7

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) – 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

» Potential influence of soil moisture on XRF readings if in situ or bagged sample measurements are planned

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
  - »XRF applications to within-sample heterogeneity issues
- Controlling short-scale heterogeneity
  - »aggregating *in situ* measurements

5-8

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 by aggregating *in-situ* measurements. This is discussed in detail later in this module.

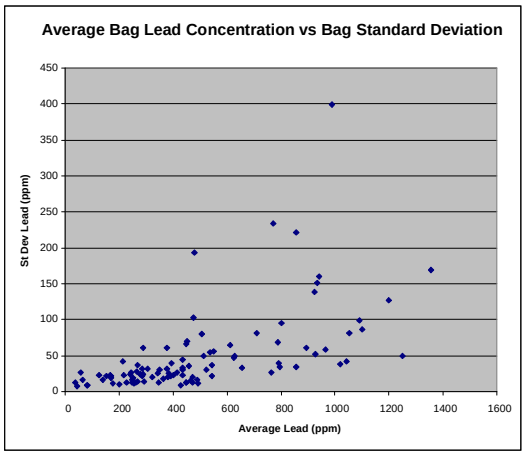
Internet

EPA's Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples, EPA/600/R-03/027, can be found at *http://clu-inor/download/char/epa_subsampling_guidance.pdf*

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, addressing within-sample variability becomes increasingly important



5-9

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. Most soil sampling techniques yield sample masses that range from 200 gram up to 1,000 grams. Metals laboratory techniques require only a small fraction of this mass for analysis (1 to 5 grams). Consequently soil samples are sub-sampled by the laboratory. Whether the sub-sample obtained is representative of the original soil mass is an open question. In general, within-sample heterogeneity is a function of contamination levels … typically for any particular site, the greater the concentration present, the greater the level of heterogeneity within samples as well.

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.
- **Analyzed multiple times for lead:** 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 preparation becomes increasingly important:** One way to interpret these data for any particular point is that the average lead value plus or minus twice its SD 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 SD 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 SD 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.

Verifying Sample Preparation by XRF

Improving Representativeness

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

5-10

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 contaminant of concern (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.
- **Target samples expected to have contamination around action levels:** It is best to target samples that have contamination concentrations close to the action level.
- **Perform multiple (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.
- **Compare resulting measurement variability:** Comparing the variability (i.e., SD) 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.

Bagged-Sample Measurements Can Substitute for Sample Preparation

Improving Representativeness

- Goal is to get an accurate estimate of the metal concentration within a sample as quickly and cheaply as possible
- Primary cost associated with an XRF is sample preparation
- Measuring through bag walls multiple times and averaging result substitutes for sample preparation
  - » How many shots through bag walls are required and what should the measurement times be?

5-11

Notes

- **Goal is to get an accurate estimate of the metal concentration within a sample as quickly and cheaply as possible:** Besides serving as a check on sample homogenization practices, an XRF can be used as a direct substitute for sample homogenization. The assumption is that the goal is to get an accurate estimate of metal concentrations within a sample as quickly and cheaply as possible.

- **Primary cost associated with an XRF is sample preparation:** The primary cost associated with XRF measurements is the labor associated with sample preparation. Multiple measurements through a bagged sample's wall is a substitute for expensive sample preparation.

- **Measuring through bag walls multiple times and averaging result substitutes for sample preparation:** One way of minimizing those costs while still obtaining an accurate estimate of metals concentrations within a bagged sample is by acquiring multiple XRF measurements systematically spaced across a bag's surface and using the average of those measurements. It then needs to be determined how many shots through bag walls are required and what should the measurement times be.

How Many Bagged Sample Measurements Are Required?

Improving Representativeness

- Select bagged sample with concentrations near action level
- Identify desired DL
- Estimate XRF measurement time required for DL and analytical error expected at action level
- Take ten shots on bag systematically (5 on a side) and observe variability (results' SD)
- Select measurement numbers

$$\frac{\text{Observed variability}}{\sqrt{\text{of measurement number}}} < \text{analytical action level}$$

- Individual measurement times equal time required for DL divided by number of measurements

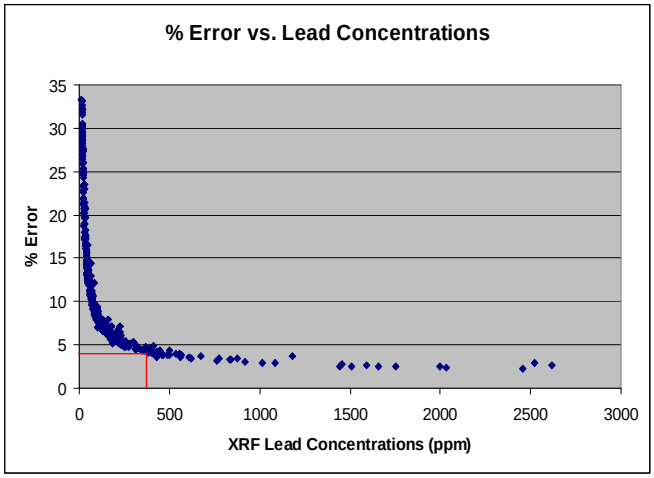
5-12

Notes

- **Select a bagged sample with concentrations near the action level:** The action level is of most interest.
- **Identify the desired DL:** Select the desired detection limit (DL) for the XRF instrument.
- **Estimate XRF measurement time required for DL and analytical error expected at action level:** Determine the XRF measurement time required to achieve the desired DL and estimate the analytical error that can be expected at the action level.
- **Take ten shots across the bag systematically (5 on a side) and observe variability (results' SD):** Take 10 XRF measurements across the bag systematically, measuring 5 times on each side of the bag. Observe the variability in the XRF measurements by calculating the SD of the results.
- **Select measurement numbers:** Select measurement numbers so that the observed variability divided by the square root of the measurement number is less than the expected analytical error at the action level.
- **Individual measurement times equal time required for DL divided by number of measurements:** The acquisition time for each shot would be the original acquisition time divided by the number of measurements to be taken.

Bagged Soil Pb Example: Determine Read Time & Instrument Error

- AL = 400 ppm
- Desired DL = 15 ppm
- DL of 15 ppm requires 120-sec read time
- @ 400 ppm, XRF instrument error < 5% for 120-sec readings



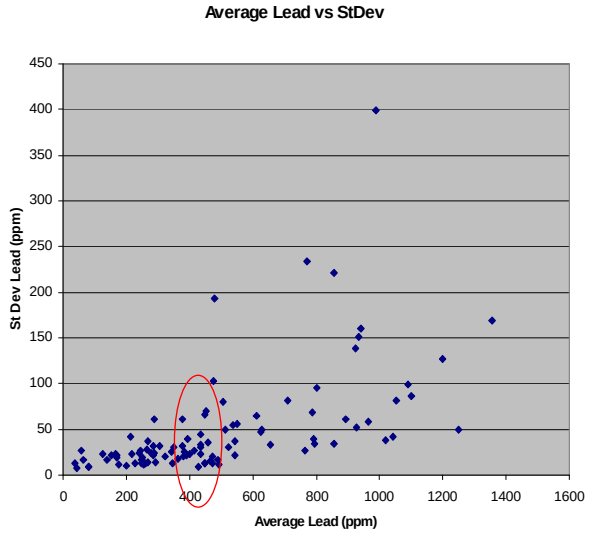
5-13

Notes

- In this example, the action level is 400 ppm, and a DL of 15 ppm is desired (which is around background levels for this site). For the instrument available, a 120-second reading will give a detection limit of 15 ppm. Around 400 ppm, the XRF relative error for a lead measurement with a 120-sec acquisition will be less than 5 percent (20 ppm). The observed error for these bags in that range (pertinent bags indicated by the red oval on the scatter plot) is about 8 percent, or 34 ppm. To cut this error in half, four measurements are needed (24 ppm divided by the square root of 4 is less than 20 ppm). Those four measurements need to total 120 seconds of acquisition time; that means each measurement should be 30 seconds long.

…Then Determine Within-Bag Variability to Find "n"

- @ 400 ppm, typical sample variability SD ~ 34 ppm (or ~ 8%)
- 4 30-sec shots per bag gives bag Pb conc with estimated error < 5% when reading is near the 400 ppm AL
- 4 30-sec shots



Same Concept Applies to XRF *In Situ* Measurements

Improving Representativeness

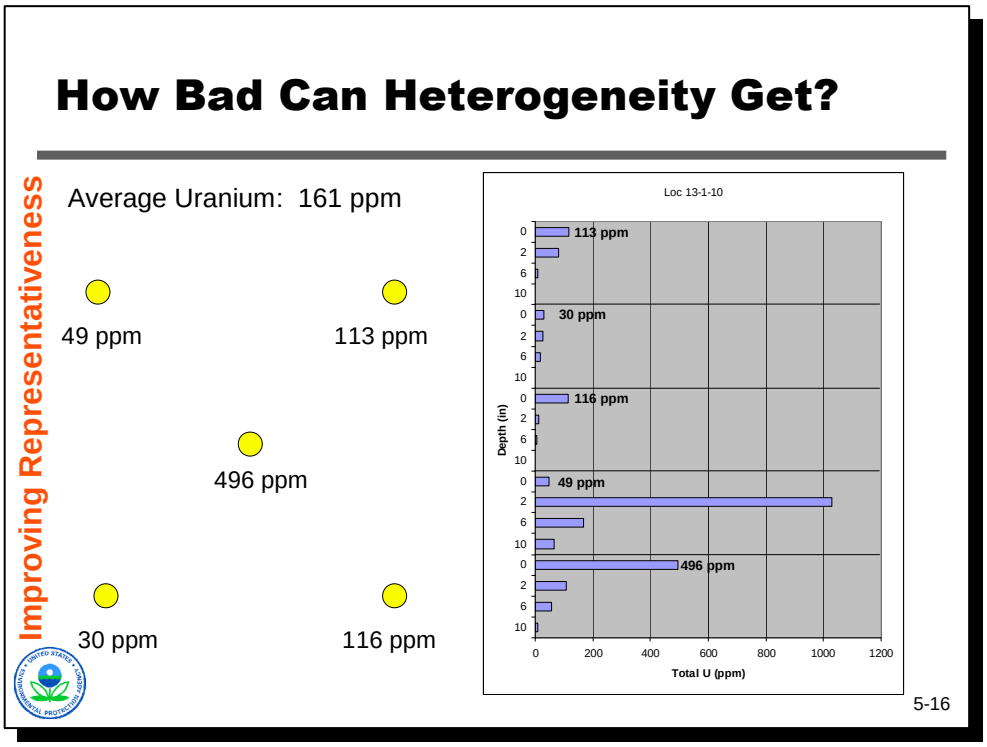
- Rapid means for quickly estimating concentration present in surface soils
- Short scale heterogeneity can be severe for impacted soils
- As with bagged samples, more shorter-acquisition-time measurements systematically across area of interest gives a more accurate assessment of the average concentration present

5-15

Notes

- The same concepts of using multiple XRF measurements to control for short-scale heterogeneity apply to *in situ* measurements as well. *In situ* measurements provide a means for quickly determining the concentration of metals present in exposed soil surfaces. The pictures show an XRF measurement from a ground surface, and an XRF measurement from soils placed in a pan…these soils were pulled from a soil core.

  The problem with this application is that short-scale heterogeneity can be severe in settings where a lot of contamination is present, and consequently a single point measurement may not accurately represent the true contamination conditions of immediately surrounding soils. As with bagged samples, one way to address this is to take multiple *in situ* XRF readings across the soil surface, and to calculate the average concentration observed.



Notes

- How bad can *in situ* short-scale heterogeneity get? Here are actual XRF data from a site contaminated with uranium. On the left are five *in situ* shots taken from an exposed soil surface over a one square foot area. The results range by an order of magnitude, from 50 to 500 ppm.

  A soil core down to 10 inches was retrieved from each of the five locations, and XRF readings taken at three different depths (2", 6", and 10"). The bar graph on the right shows the uranium results. In this approximately cubic foot of soil, XRF results ranges from background (~3 ppm) up to more than 1,000 ppm…in other words over almost 3 orders of magnitude.

One Additional XRF Not-So-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 for short acquisition times
- Particularly important for repeated *in situ* measurements or repeated measurements of bagged samples

5-17

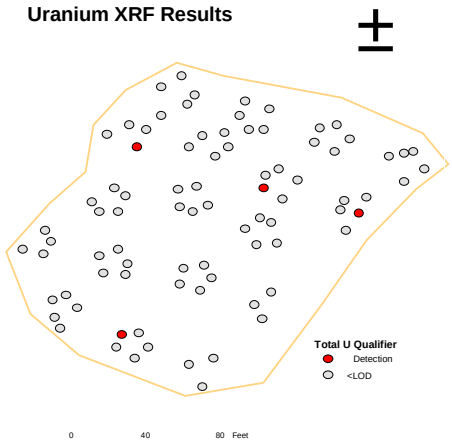
Notes

- **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:** This relationship holds true no matter whether one is talking about one long measurement, or whether one is averaging the results from many short measurements…the detection limit associated with an averaged result will reflect the aggregated time of all of the individual measurements.
- **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.

Working with <LOD Results: Exposure Units/Area Averages

Improving Representativeness

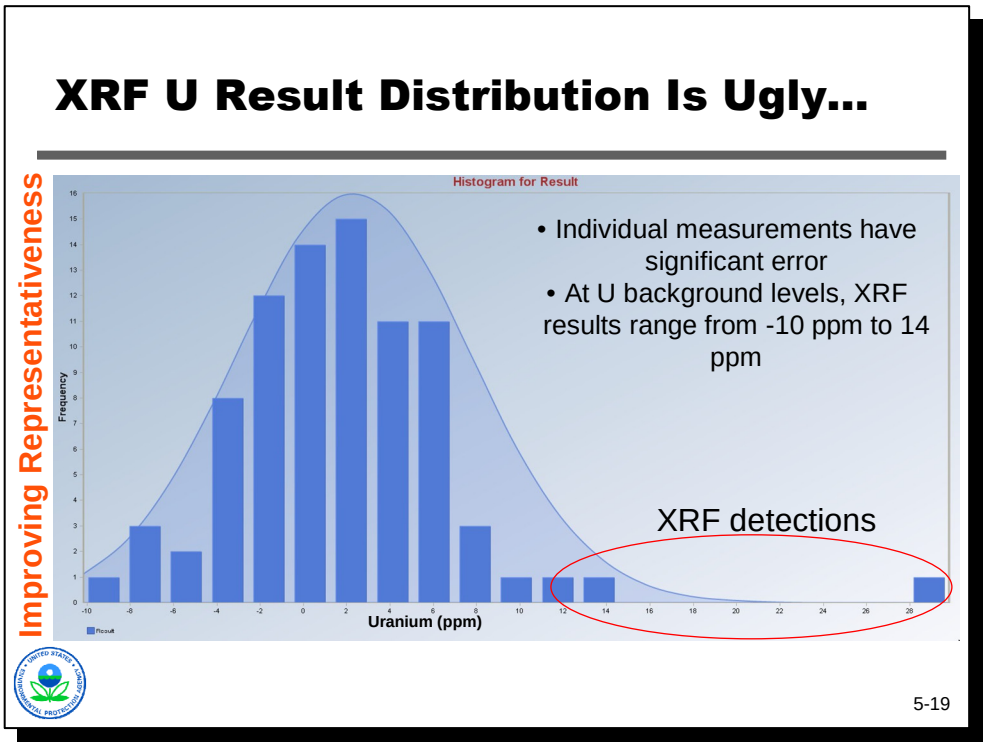
- 5 acre exposure unit
- Uranium issue
- 84 XRF samples
- All but 4 non-detects
- DL: ~15 ppm w/ 120 sec reading
- Background: ~ 3 ppm



5-18

Notes

- Here is an example of this principle in action. 84 XRF measurements (prepared cup samples) were collected from a 5 acre exposure unit (don't bother asking about the weird pattern…there's not a good explanation). Uranium was the contaminant of concern. 80 of these were "non-detects" (detection limit for each 120-sec reading was about 15 ppm). The actual background concentration for uranium for this site was around 3 ppm. In this particular case the individual XRF readings did not have detection limits sufficient to quantify uranium down to background levels.



Notes

- If one looks at the actual results for these 84 readings, one gets this "ugly" distribution, with the actual XRF data ranging from 10 ppm up to 14 ppm. The data underscore the fact that the individual data have significant measurement error associated with them, and as individual data points have little meaning for low levels of uranium.

…But EU Average U Calculation Provides Usable Results

Improving Representativeness

- With raw U XRF data:
  - » average: 2.3 ppm
  - » 95%LCL-UCL: 1.1 - 3.5 ppm
- Standard guidance would have discarded <LOD values when calculating mean
- End result would either mean rejecting XRF data, or using a much less accurate approach for estimating mean concentration (e.g., set <LOD results to LOD or ½ of LOD)

5-20

Notes

- However, if one averages this XRF data over the whole unit, the XRF data turn out to be valuable. The average is 2.3 ppm, right in line with what one would expect from background conditions, and the 95 percent LCL/UCL on the mean is relatively tight.

  Note that standard EPA guidance for handling the 80 non-detected uranium results would have meant either discarding the uranium XRF data, or using an obviously biased approach for estimating mean concentrations (e.g., setting non-detects to the detection limit, or half of the detection limit).

Working with <LOD Results: Bagged Sample Estimates

Improving Representativeness

Bag ID	Sample	Result	Error	Flag
BS-18	TOP-1	*10.9*	3.7	<LOD
	TOP-2	*5.4*	3.7	<LOD
	TOP-3	*0.0*	3.6	<LOD
	TOP-4	*2.0*	3.8	<LOD
	TOP-5	*7.1*	3.6	<LOD
	BOTTOM-1	*6.2*	3.8	<LOD
	BOTTOM-2	*1.3*	3.7	<LOD
	BOTTOM-3	*3.5*	3.8	<LOD
	BOTTOM-4	*8.8*	3.8	<LOD
	BOTTOM-5	*9.0*	3.5	<LOD

- Uranium concerns
- 30-sec readings
- DL: ~11 ppm/reading
- Averaging 10 readings for bag
- Equivalent to 300-sec reading
- Average = 5.4 ppm +/- 1.2 ppm
- DL for average: ~3.6 ppm

5-21

Notes

◆ Here's another example of working with non-detect XRF data … bagged samples with measurements through the bags. In this particular example (again uranium, this time with 30-sec readings and ~11 ppm detection limits per reading), for this particular bag, all of the results came back as non-detects.

If we use the raw data, however, the average uranium concentration for the bag was 5.4 ppm. The standard error for the estimated average was 1.2 ppm, which translates into a detection limit for the 10 sample aggregate of 3.6 ppm (i.e., DL is three times the standard error).

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*

5-22

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 Measurement Numbers Dynamically

Improving Representativeness

- Applicable to *in situ* and bagged sample readings
- XRF results quickly give a sense for what levels of contamination are present
- Measurement numbers 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 or within a soil core

5-23

Notes

We do not 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: 1st Foray Into Dynamic Work Strategies

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

5-24

Notes

- For this example, 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?).

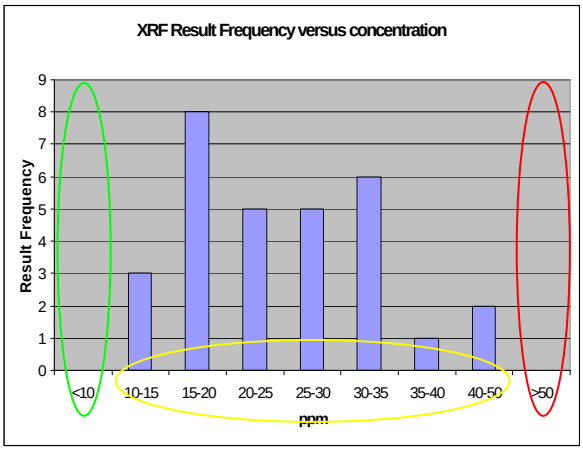
  For this particular example we have 3 bagged samples that we are quite confident came from an area with concentrations around our action level. We measured each bag systematically across their front and back ten times (5 on the front, 5 on the back). We observed that the average concentration reported 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

Simple Decision Rule:

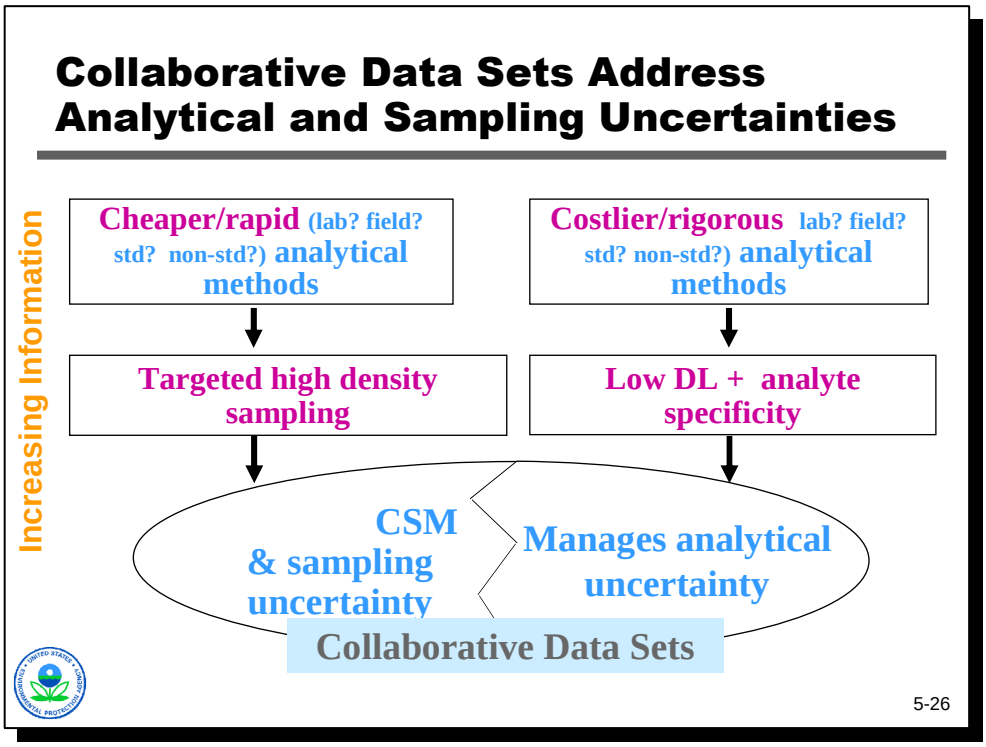
- if 1st measurement less than 10 ppm, stop, no action level problems
- if 1st measurement greater than 50 ppm, stop, action level problems
- if 1st measurement between 10 and 50 ppm, take another three measurements from bagged sample



5-25

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.



Notes

- This graphic shows that the less expensive and 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 when making decisions.
- Same budget allows a lot more XRF data points, improving average concentration estimates and hot spot identification
- Assumptions:
  - » XRF data is unbiased (or can be adjusted)
  - » Linear relationship exists w/ high correlation (SW-846 Method 6200 points to correlation coefficients >0.9 as producing "lab equivalent" data)
  - » Expensive traditional analyses still used for QC purposes
  - » Applicable to static or dynamic work plans
- Requirements: DMA study to establish relationship between XRF & lab method may be necessary. Perform on-going QC to verify comparability holds.

5-27

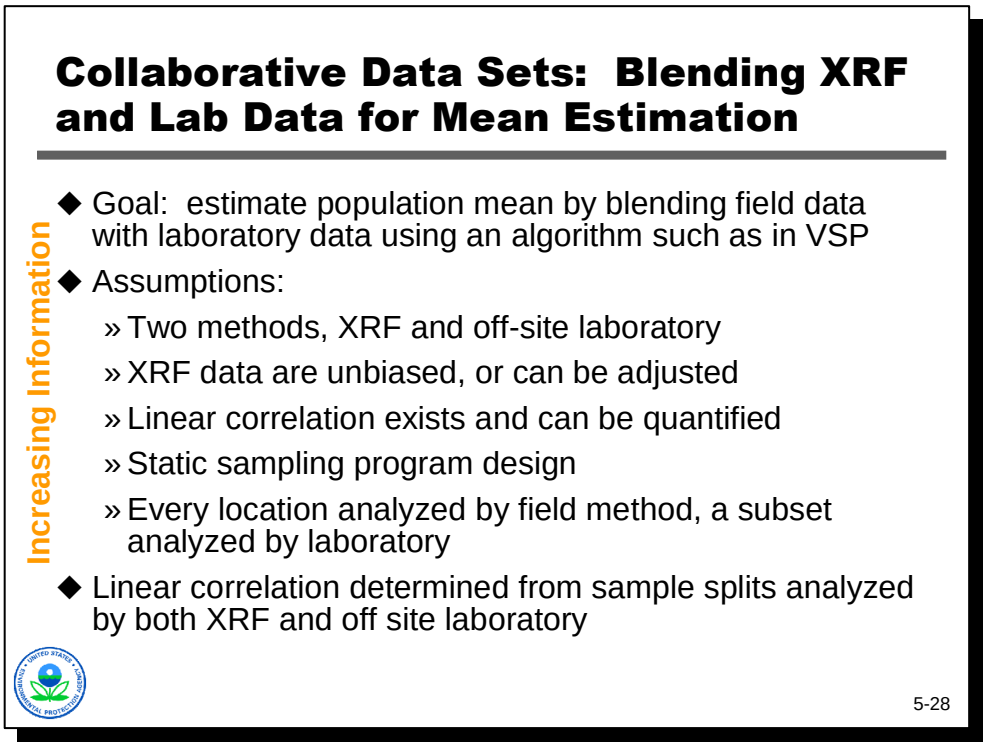
Notes

- **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 adjustable 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.



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 Sample Plan (VSP).

◆ **Assumptions:** The assumption is that our lower analytical quality data are unbiased (or if there is bias present, it can be adjusted), 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 analytical quality method, and a subset would also be analyzed by the more definitive analytical 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

- Potential 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 may not hold in actual site projects

5-29

Notes

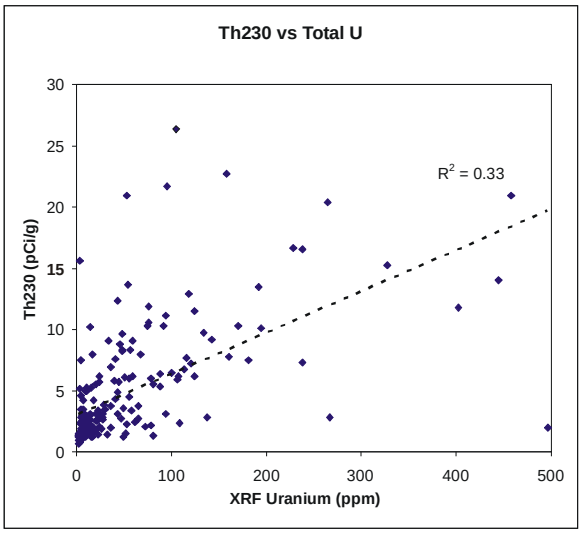
- **Potential issues with both previous approaches:** Both approaches assume that traditional laboratory data are "definitive," which is not always the case. Both approaches assume that the linear relationship between field and laboratory data holds over the whole range of data encountered, which is 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



5-30

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.

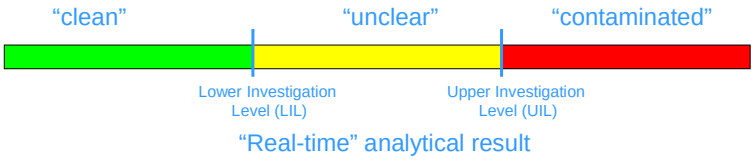
  The graphic shows an example from an XRF application. At this particular site, Th-230 was the risk driver. However, Th-230 doesn't lend itself to any convenient field analytical technique. At this site, however, the Th-230 was generally collocated with uranium, and uranium is measurable by XRF. This 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

Increasing Information

- 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
  - » LIL for real-time result below which we are confident contamination is not present
  - » UIL above which we are confident contamination is present



5-31

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 do not 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

5-32

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

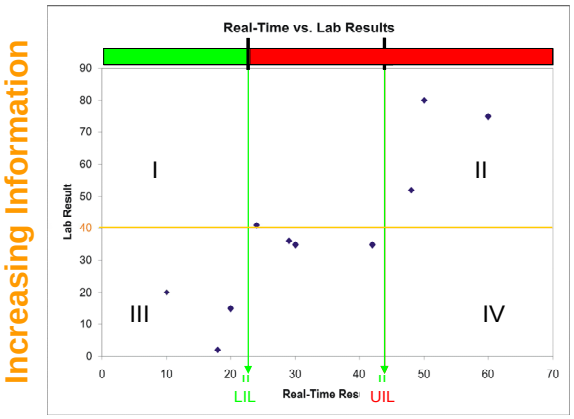
5-33

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



- 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%**

5-34

Notes

- This hypothetical example illustrates how one might 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 we have 10 samples plotted on the scatter plot. The x-axis is their real-time result (e.g., XRF). The y-axis is the corresponding lab 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 lab 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.

By properly selecting our LIL and UIL values, we can drive false clean and false contaminated rates towards zero. The price to be paid, for this example, is that we would need to send all samples with real-time results between the LIL and UIL to the laboratory for analysis.

Addressing the Unknown through Dynamic Work Strategies

Addressing the Unknown

- Adaptive analytics
  - »Hot spots
  - »Mean estimation
- Adaptive sampling
  - »Modifying sample/measurement numbers for mean estimation
  - »Simple decision rules for delineation

5-35

Notes

Dynamic work strategies can incorporate a variety of "if-then" scenarios to guide the progression of field work. In the context of XRF deployments, two types are applicable.

- Adaptive analytics is one example of a dynamic work strategy. Adaptive analytics makes use of collaborative data as described earlier. Field work is based on more than one analytical method, likely including at least one field deployable real-time technique along with standard off-site laboratory analyses. Typically all samples are analyzed by the real-time method, with off-site analyses reserved for those samples that meet specific criteria. Adaptive analytics can be used when looking for hot spots and when trying to estimate the mean concentrations for specific areas (e.g., an exposure unit).

- Adaptive sampling is another example of a dynamic work strategy. Adaptive sampling refers to modifying the number and/or locations of samples based on real-time results. Adaptive sampling has applications when estimating the mean concentration for a specific area (e.g., an exposure unit) and when trying to delineate contamination that has been encountered.

Adaptive Analytics

Addressing the Unknown

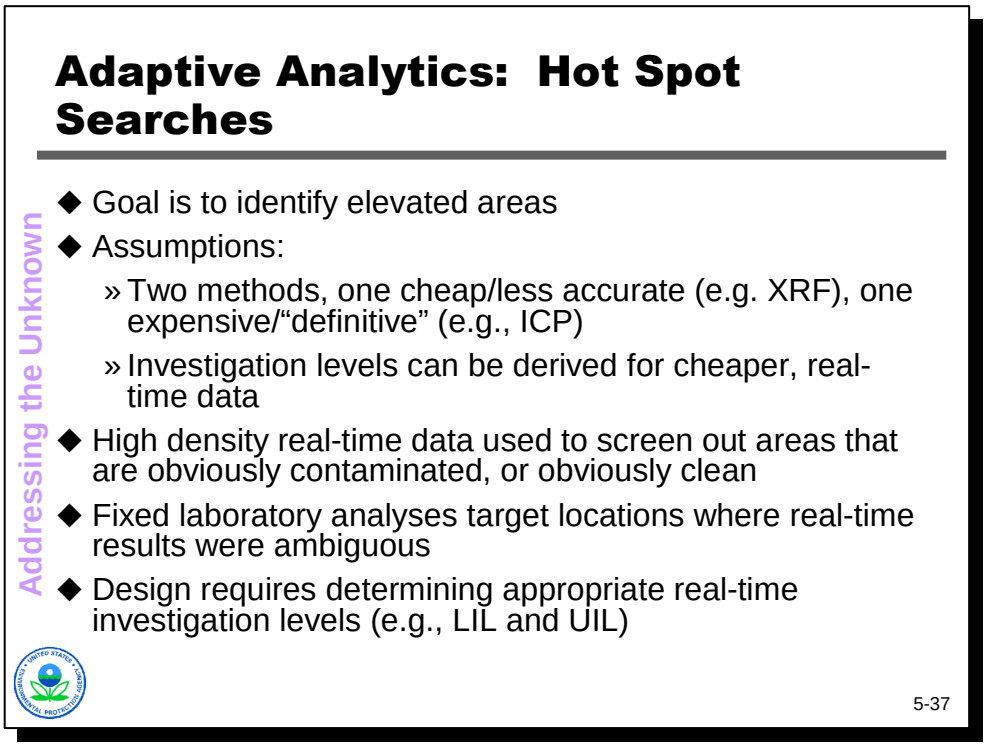
- Cheaper "real-time" method used to produce spatially dense data
- Based on "real-time" results, more expensive and definitive analyses done on selected sub-set of samples
- Decisions based on field investigation levels and/or other "triggers" (e.g., elevated presence of interfering element for XRF)

5-36

Notes

During this module we discussed the use of field investigation levels (lower investigation levels and upper investigation levels) to help with real-time data decision-making.

- **Cheaper "real-time" method used to produce spatially dense data:** Adaptive analytics uses less expensive "real-time" methods (such as the XRF) to produce data that provides dense coverage of the decision unit.
- **Based on "real-time" results, more expensive and definitive analyses are done on selected subset of samples:** The real-time data results are reviewed to guide the selection of a subset of samples for analysis by a laboratory method. The laboratory results are used to investigate real-time results that are of particular concern.
- **Decisions based on field investigation levels:** The decision as to whether to send a sample off for confirmatory laboratory analysis is driven by field investigation levels that are applied to the real-time results. These investigation levels guide decision-making.



Notes

Making use of adaptive analytics is one approach for combining collaborative data sets and dynamic data collection strategies. In this case the goal would be to identify elevated areas or delineate contamination. Sampling locations are fixed. The dynamic dimension of this type of program stems from the ability to select from different analytical techniques as work progresses.

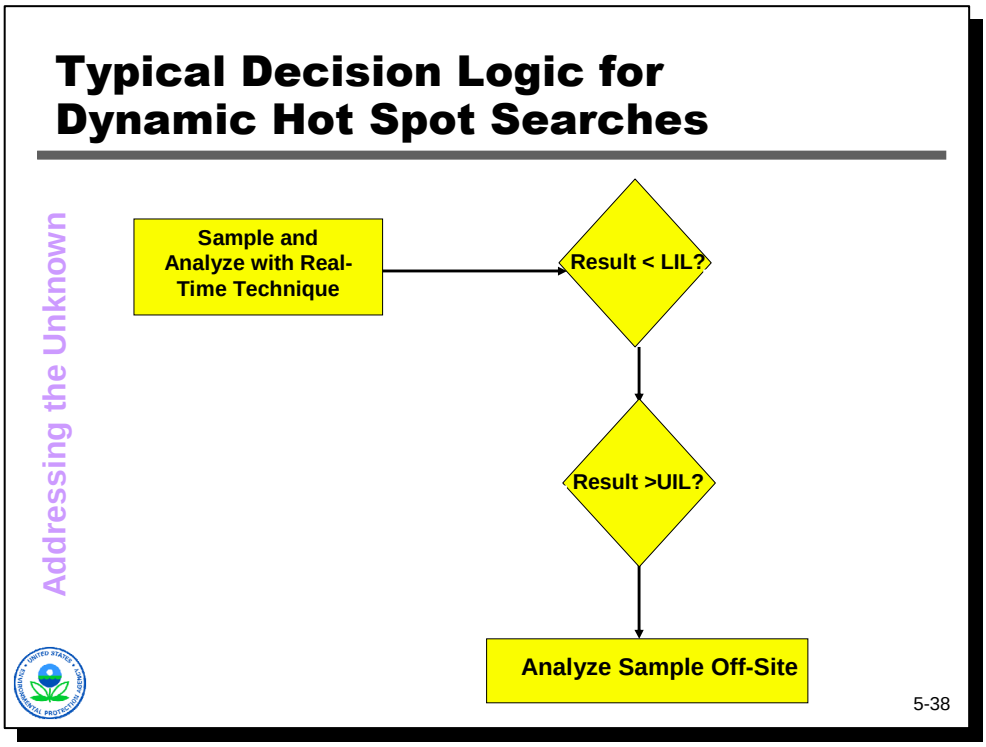
- **Goal is to identify elevated areas:** The goal of a hot spot search is to identify those areas of a site or decision unit that contain elevated levels of contamination that are significantly higher in concentration than other areas of the site. Areas with higher levels of contamination pose greater risks to human health and the environment and may need to be treated differently than other areas.

- **Assumptions:** The adaptive analytics approach assumes there are two methods available, one real-time method such as an XRF that provides data at a low cost but that is not highly accurate and another method such as ICP that is expensive but provides accurate data. This approach also assumes that investigation levels (lower investigation level and upper investigation level) can be derived for the less expensive real-time method.

- **High density real-time data used to screen out areas that are obviously contaminated, or obviously clean:** The real-time method is used to take many measurements (typically systematically) across the decision unit, creating a dense picture of contamination levels. The data is first used to screen out areas that are obviously contaminated, or obviously clean.

- **Fixed laboratory analyses target locations where real-time results were ambiguous:** The real-time data that is ambiguous, or that is between the established LIL and UIL, are targeted for fixed laboratory analyses.

- **Design requires determining appropriate real-time investigation levels (e.g., LIL and UIL):** This approach requires that the LIL and UIL for the real-time measurement levels be established. The LIL and the UIL define the obviously clean and obviously dirty areas and the ambiguous areas. Recall that the last module provided an example of how a LIL or UIL might be selected.



Notes

- This flow chart shows the decision logic for dynamic hot spot searches. Many samples are analyzed using a real-time method systematically across an area. Results that are less than the LIL indicate an area is clean of hot spot concerns. Results that are greater than the UIL indicate an area is contaminated at "hot spot" levels. Results in between the LIL and UIL are ambiguous and samples are sent off-site for laboratory analyses.

An Example…

Addressing the Unknown

- Looking for contaminated sediment layer
- Uranium used as a proxy for primary COC (which is not measurable by real-time technique)
- UIL and LIL derived for XRF uranium
- Every 6-in interval of 3-ft cores screened by XRF
  - » If all intervals < LIL, core considered clean
  - » If one or more intervals > UIL, core considered impacted
  - » If one or more intervals between LIL and UIL (but none above UIL), highest interval sent for lab analysis

5-39

Notes

- Here is an example of this logic at work. This site has contaminated sediment concerns. The contaminated sediment layer, when it is present, can exist at varying depths (i.e., close to the surface or at the surface in some areas, but buried deeper in other areas). The purpose of the GeoProbe work was to identify areas where contaminated sediments were a concern.
- The primary COC was a contaminant that was not amenable to real-time techniques. Fortunately, however, historical data indicated it was collocated with elevated uranium. Uranium is something that can be easily measured by XRF.
- A UIL and LIL were derived for the XRF and uranium based on a review of historical data. The LIL was selected so that if XRF uranium results were below that value, there was little chance the primary COC was present at levels of concern. The UIL was selected so that if XRF uranium results were above that value, there was a high probability that the primary COC was present at levels of concern.
- GeoProbe cores were systematically placed across the area of interest, with coring done down to a depth of 3 feet. Each six inch interval of the each core was screened by XRF. If all of the XRF uranium results were below the LIL, the conclusion was that there were no risk concerns at that location. If at least one XRF uranium result was above the UIL, then the assumption was that the contaminated sediment layer was present. If one or more XRF uranium results were above the LIL, but none were above the UIL, the core interval with the highest XRF uranium reading was selected and sent off-site for laboratory analysis.

When are Adaptive Analytics Cost-Effective for Searches?

Addressing the Unknown

- The "smaller" the "unclear" zone, the better the performance
- 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 follow-up fixed laboratory sampling required beyond QA/QC needs
- Worst case: every sample requires follow-up laboratory analysis

5-40

Notes

- **The "smaller" the "unclear" zone, the better the performance:** Adaptive analytics will be more cost effective when the difference between the LIL and the UIL is small. Fewer samples results will be between the LIL and the UIL that require subsequent laboratory analysis.
- **The greater the difference is between background and the action level, the better the performance:** Adaptive analytics will be more cost effective when the action level is much greater than the background level because more sample results will be below the LIL and clearly defined as "clean."
- **The greater the difference between the action level and average contamination concentration, the better the performance:** Adaptive analytics will be more cost effective if there is a large difference between the action level and the average concentration present because it will be less likely that the real-time method will yield a result between the LIL and UIL.
- **Best case:** In the best case, the real-time technique can be relied upon without additional follow-up using fixed laboratory sampling except for that required for quality assurance/quality control.
- **Worst case:** In the worst case, the real-time technique yields useless data and every sample requires follow-up laboratory analysis.

Adaptive Analytics: Mean Estimation

Addressing the Unknown

- Goal: Identify areas of concern and estimate their mean concentration
- Assumptions:
  - » Two methods, one cheap/less accurate (e.g., XRF), one expensive/"definitive" (e.g., ICP)
- XRF data identifies areas of concern
- XRF data used to estimate number of more expensive analyses required to estimate mean accurately
- More expensive, higher analytical quality data used to estimate average concentrations

5-41

Notes

- **Goal:** The idea is that lots of cheap, lower analytical quality data can be used to identify areas of concern, and then limited sampling with more expensive, higher analytical quality data can provide definitive information about those areas (e.g., estimate average contamination concentrations for an area).
- **Assumptions:** This approach assumes there are two methods, one of which is inexpensive and not highly accurate (e.g. XRF) and another which is expensive but accurate (e.g., ICP). The only requirement for the cheaper technique is that it has sufficient detection capabilities to confidently identify areas or situations that would be of concern.
- **Cheaper, lower quality analytical data identifies areas of concern . . . data used to estimate number of more expensive analyses required:** From a dynamic work plan perspective, the results from cheaper, "real-time" methods can be used to determine which areas requires more definitive sampling, and how many samples should be used.
- **More expensive, higher analytical quality data used to estimate average concentrations:** The off-site laboratory analyses are used to estimate the average concentrations within the decision unit.

An Example….

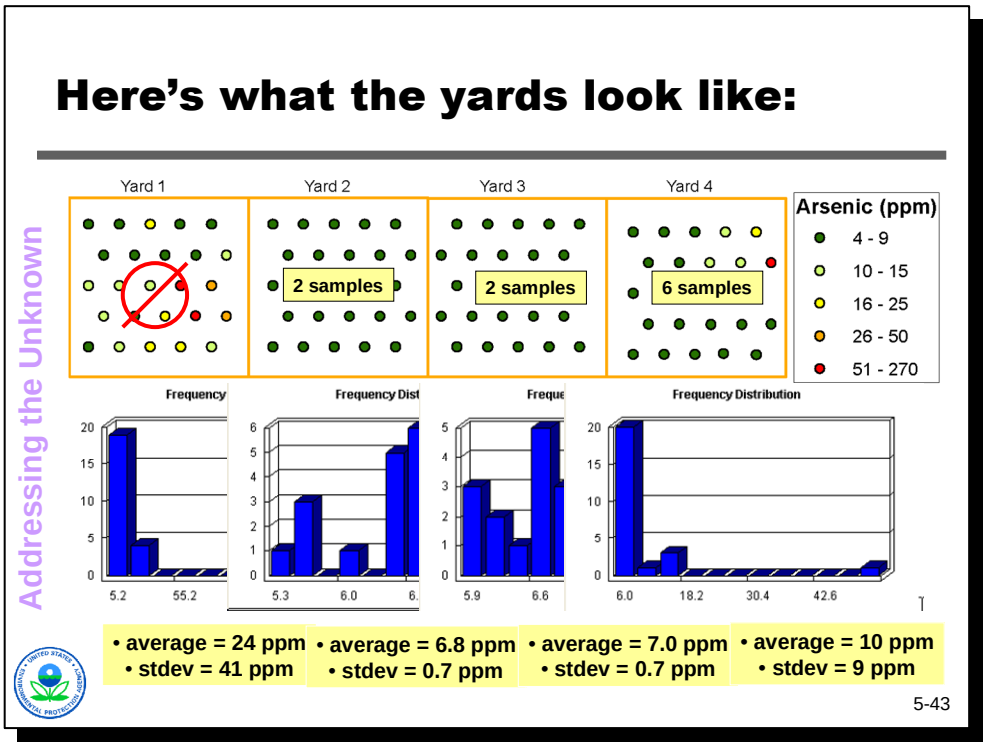
Addressing the Unknown

- 4 residential backyards screened by XRF for arsenic with action level of 25 ppm averaged over yard
- Regulator insists final release decision be based on lab data
- Use XRF to determine:
  - »whether each yard is likely above or below action level, and
  - »if below, how many lab samples are required to statistically show it?

5-42

Notes

- Here is an example of mean estimation. In the case of this project, XRF data were deemed by the regulator involved as not sufficient to establish that individual yards met arsenic action levels for release purposes. However, *in situ* XRF data could be used to quickly get a sense for whether a yard was a candidate for closure (or conversely was going to require remediation), and if it was ready for closure to identify how many "definitive" laboratory samples would be required to statistically establish that the 25 ppm requirement had been met.



Notes

- The map shows the four yards and the results of 25 *in situ* XRF readings systematically spread across each yard, color coded by arsenic values (4 – 9 is definitely background). Below each yard is a histogram of those 25 results along with the observed average and standard deviation (note that none of these, including the background yards, look particularly "normally" distributed).

- As should be clear from these XRF data, two yards are not impacted, while the other two appear to be impacted at varying levels. In the case of the yard at the far left, the average is very close to the 25 ppm standard, indicating it probably is not a candidate for release. In the case of the yard at the far right, the average is still well below 25 ppm despite the impacts, suggesting that with enough laboratory samples it may be able to be released.

- The XRF data allow estimation of the average arsenic concentration in each yard and the variability in arsenic concentrations that is present. That information, in turn, allows for customization of the number of discrete samples sent off to the laboratory for each yard to demonstrate compliance with the cleanup criteria, assuming a Student t test would be used to make that determination. Note that the yard to the left is not a candidate for closure…the XRF data suggests it would be futile to try release this yard. Note too that the number of samples required varies significantly from yard to yard.

Adaptive Analytics Also Pertinent to QC

Addressing the Unknown

- Heavy lifting for decision-making done using real-time techniques (i.e., XRF)
- Adaptive analytics used to target real-time technique QC
- Criteria developed that flag real-time samples as candidates for off-site lab analysis in support of QC

5-44

Notes

- These same concepts also apply to the use of adaptive analytics for QC purposes. Typically when collaborative data sets are used, a fixed percentage (e.g., 5 percent or 10 percent) of samples are sent off-site for more definitive laboratory analyses. Often these samples are either randomly identified, or are sent at specific intervals (e.g., after every 10 or 20 samples collected). The problem with this is that catching analytical problems this way is a real hit-or-miss affair.

  By understanding the ways that a real-time method such as an XRF might go "bad" (e.g., particular soil matrices, presence of other contaminants that interfere, extremely high or low results, etc.), then the project team can be much smarter in designing criteria that flag real-time samples as candidates for off-site laboratory analysis, and stand a much better chance of catching and correcting "problems" before they jeopardize the outcome of the field effort.

Example: Arsenic and Lead

Addressing the Unknown

- Site has known arsenic problems with arsenic decision-making driver
- Data collection primarily based on field-deployed XRF targeting arsenic
- Elevated lead also present, potentially at levels interfering with XRF arsenic readings
- Decision rule: if XRF lead > 10x arsenic result, require ICP analysis on sample

5-45

Notes

- Applying the XRF to an arsenic problem when elevated lead is present illustrates this. It would be most efficient to make decisions based on the XRF, but the project team needs to be wary of potential problems that the lead might introduce. The simple decision rule in the case for weeding out samples where our XRF arsenic data might be "bad" is to send off every sample for ICP analysis when the lead concentration is greater than ten times the arsenic concentration reported by the XRF.

Adaptive Strategies for Mean Estimation with XRF

Addressing the Unknown

- Use XRF as the primary data source
- Goal is to determine whether average concentration is above or below standard
- Will be estimating mean and 95% LCL/UCL based on bagged sample XRF data:
  - »How many sampling locations?
  - »How many shots per bag?

5-46

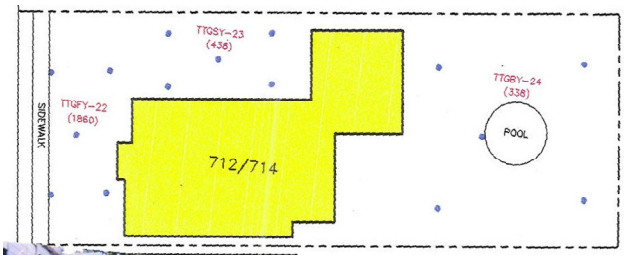
Notes

- As a last example of XRF dynamic work strategies, the remaining slides describe using an XRF combined with adaptive strategies to estimate average concentrations across decision units. The assumptions here are that XRF data quality is sufficient so that it can be used as the primary data source for decision making, and that our sampling goal is to determine whether the average concentrations within individual decision units are above or below some standard or action level.

Example: Lead in Residential Properties

Addressing the Unknown

Problem: historical sampling yielded ambiguous results re: lead contamination status in yards



XRF-Based Approach:

- 500 ppm goal, averaged over a property
- Stratify each property by yards (front, side, and back)
- Collect minimum data per yard: 5 locations per yard with each resulting bag shot 4 times
- Sample, measure, and calculate 95% LCL and 95% UCL for mean for each yard

5-47

Notes

- The site is a residential area adjacent to a facility with known lead problems. Historical sampling from these properties using standard techniques (i.e., five point composite, one sample formed from the composite and submitted for analysis) yielded ambiguous results. The goal is to use the XRF to come to a definitive conclusion about each property's contamination status.

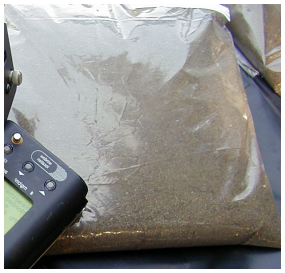
  The cleanup goal for this project was 500 ppm, averaged over a property. Individual properties were typically comprised from front, side, and back yards. Back yards were typically by far the largest area, but also were believed to be the least impacted by lead (i.e., higher concentrations were expected in front and side yards). The minimum sampling proposed for each property was five bagged samples from each yard (total of 15 for the property), with each bag analyzed four times, twice on side of the bag, twice on the other. These data were then used to estimate the average lead concentrations in each yard along with their corresponding 95 percent LCL and 95 percent UCL for the estimated means.

  The "real-time" question for each yard was whether the collected data were sufficient to unambiguously make decisions for the property, or whether either additional samples were necessary and/or additional readings on existing bagged samples were required to be definitive.

Example

Addressing the Unknown

- If 95% UCL < 500, stop! If 95% LCL > 500, stop!
- If neither, uncertainty sources evaluated and either:
  - » Bagged samples re-measured more times (e.g., 6 more times), or
  - » Additional 5 sampling locations added to the yard, or
  - » Both done
  - » LCL and UCL then re-calculated
- 95% UCL for property calculated from yard results (area-weighted stratified sampling design) and compared to 500 ppm



5-48

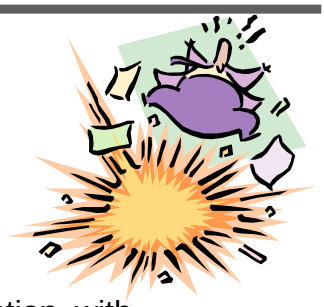
Notes

- The decision rule for each yard was straightforward. If the 95 percent UCL was less than 500 ppm, or the 95 percent LCL was greater than 500 ppm, data collection for that yard could stop. Enough data had been collected.

  If neither were true, then the data sets were reviewed to determine where data uncertainty was coming from…variability within bagged samples (which would indicate bags should be re-analyzed more times to further pin down the lead concentrations in the bags), or whether the variability was from differences in lead concentrations across a yard, in which case more samples were required (or perhaps both). If necessary, more data was collected and the mean, 95 percent LCL and 95 percent UCL were recalculated. If the decision rule still could not be met, then the yard was determined to be so close to the 500 ppm action level that it would be presumed contaminated.

The Biggest XRF Bang Comes from Combining…

- CSM knowledge, with…
- Collaborative data sets, with…
- Adaptive analytics, with…
- Adaptive QC & data uncertainty reduction, with…
- Adaptive sampling.



5-49

Q&A

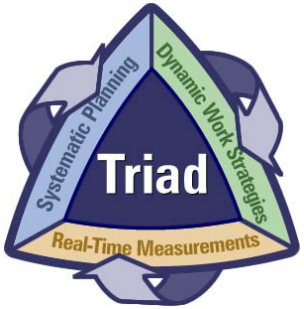
5-50

Module 6

Case Study

Module 6

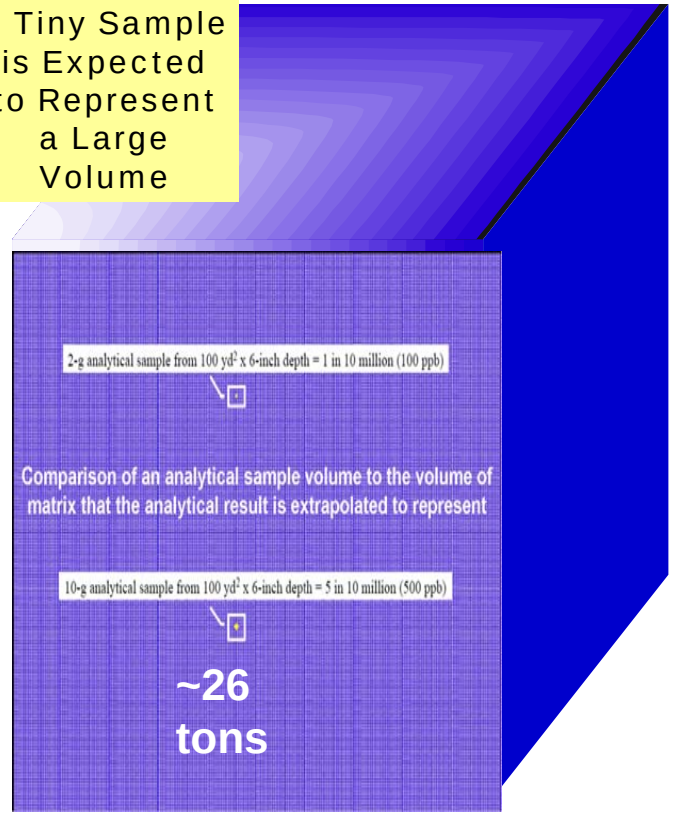
Case Study
Field XRF & Stratified Sampling Design for Residential Soil Lead



First, a Review of Representativeness Issues

- This case study illustrates using real-time data as feedback as to whether the **"representativeness"** aspect of data quality is sufficient to allow confident decisions about… **("representative of these decisions")**
  - Pb concentration around an action level
  - Whether the RCRA facility is responsible for any cleanup needed
  - If any cleanup is performed or recommended, what areas should be remediated to get the most contaminant mass removed per dollar.

1 Tiny Sample is Expected to Represent a Large Volume

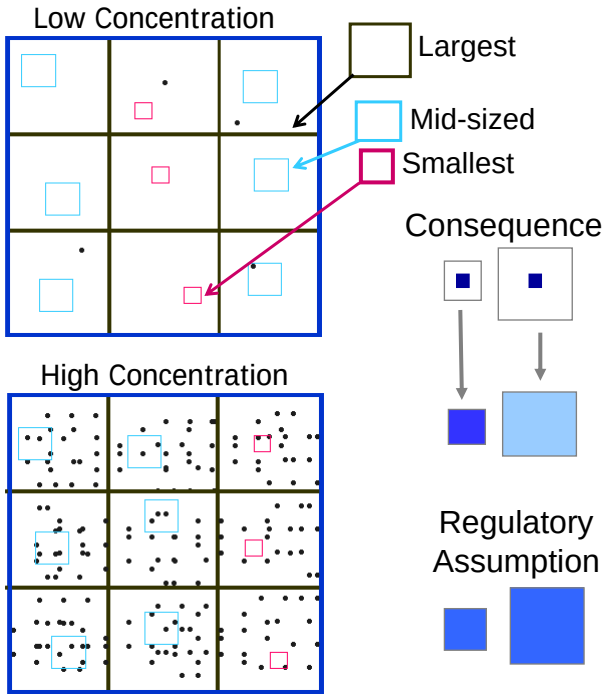


Which Tiny Sample?

Adapted from Interstate Technology and Regulatory Counsel (ITRC), 2003 http://www.itrcweb.org/SMART-1.pdf

Firing Range Soil Grain Size (Std Sieve Mesh Size)	Pb Concentration in fraction by AA (mg/kg)
Greater than 3/8" (0.375")	**10**
Between 3/8" and 4-mesh	**50**
Between 4- and 10-mesh	**108**
Between 10- and 50-mesh	**165**
Between 50- and 200-mesh	**836**
Less than 200-mesh	**1,970**
Bulk Total	**927 (wt-averaged)**

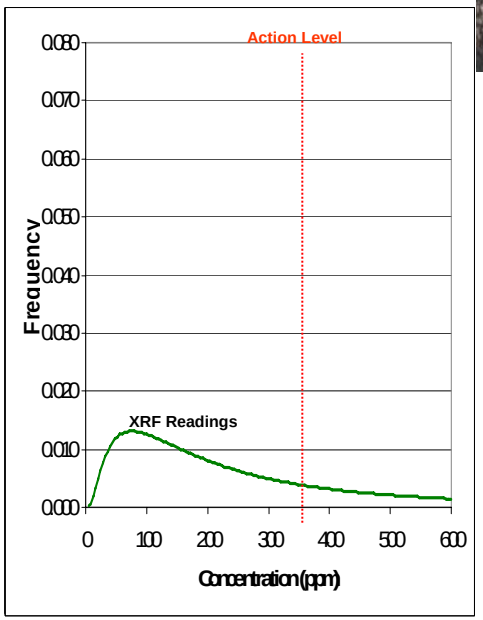
Larger Sample Supports Produce More Consistent Data



Experimental Data on Sample Support

Subsample Support	Coeff of Var. (CV)	Number of subsamples req'd to estimate true sample concentration within a range of…	
		…± 25% ex: 1930 ± 25% = 1448 – 2412 Less accurate	…± 10% ex: 1930 ± 10% = 1737 – 2123 More accurate
1 g	0.79	39	240
10 g	0.27	5	28
50 g	0.12	1	6
100 g	0.09	1	4

Extra Care Required for XRF



Can XRF Provide "Definitive Data?"

- Data Quality Objectives Process for Superfund: Interim Final Guidance (Sept. 1993)
  - Page 43: *"For the data to be definitive, either analytical or total measurement error must be determined."*
- Measuring error requires taking multiple replicate analyses on a sample - this increases analytical costs, which is why error is seldom determined or reported
  - However, XRF is unique in that replicate readings are very inexpensive
  - Easy to determine error and meet SF's definition of definitive data

Real-time Data & Managing Decision Confidence

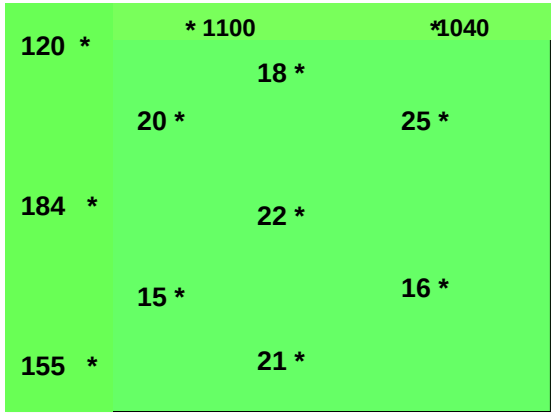
Real-time availability of results

- Instantly recognize data uncertainty that interferes with confident decision-making
- Increase replicates and standard reference materials as needed to calculate data "error" (imprecision & bias)
- Adapt sample processing & analysis to reduce data error to acceptable (to decision-making) levels
- Document statistical decision confidence
- Real-time maturation of decision-focused CSM is THE most powerful QA mechanism available

(Statistical) Stratified Sampling Design

- **Purpose:** determine the overall mean & UCL for a decision unit (DU) when different sections of the DU have different means & standard deviations (SDs).
- *Methods for Evaluating the Attainment of Cleanup Standards Volume 1: Soils and Solid Media",* 1989, section 6.4 http://www.cluin.org/download/stats/vol1soils.pdf
- *Guidance on Choosing a Sampling Design for Environmental Data Collection* (EPA QA/G-5S), 2002, Chap 6. http://www.epa.gov/quality/qs-docs/g5s-final.pdf
- *Data Quality Assessment: Statistical Methods for Practitioners* (EPA QA/G-9S), 2006, section 3.2.1.3 http://www.epa.gov/quality/qs-docs/g9s-final.pdf

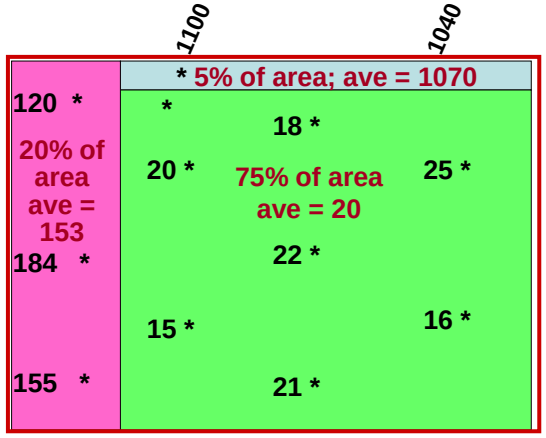
What Makes a Stratified Design Different?



To calculate average over the entire area, **routine practice** is that data go straight into a database, and then…

Sum(all) = 2736; then 2736 ÷ 12 = 228 ppm

"Dividing by 12" assumes equal weight is given to each sample (1/12th of total area)



But the CSM supports partitioning the site into 3 distinct portions based on similar populations

20(0.75) + 153(0.20) + 1070(0.05) = 99 ppm

A spatially weighted mean makes a difference!

Area	High	Mid	Low	Routine	Stratified
Mean	1070	153	20	228	**99**
SD	42	32	4	398	**80**
	95% (t-dist) UCL			434	**143**

Basic Principles of a Stratified Sampling Design

The CSM is the basis for defining both the DU & its strata

- Decision Unit (DU) = a unit for which a decision is made: a single drum, a batch of drums, risk exposure unit, remediation unit, etc.
- The DU is the volume & dimensions over which an average conc is desired
- Strata are created by **different release or transport** mechanisms – cause different contaminant patterns in within the DU
  - Target properties, like conc level & variability, differ from strata to strata w/in the DU
  - Strata must have a physical basis!!

Basic Principles (cont'd)

- DU is delineated (stratified) into non-overlapping subsections according to the CSM
- Each stratum's area/volume is recorded as a fraction of the DU's area/volume
- Each stratum's conc mean & SD determined
- The means & SDs are weighted & mathematically combined → overall mean & UCL for the DU
- Can apply stratification to data analysis even if not planned into sampling, but must have spatial info & final CSM available

Benefits of a Stratified Sampling Design

- ❖ Small areas of very high or low conc **do not bias** the overall mean of the DU.
- ❖ Reduces variability (SD) in the DU data set
  - ➢ **Reduces statistical uncertainty** (as measured by the distance between the mean & **UCL**)
- ➢ Preserves spatial information to identify **source/transport** mechanisms & support **remedial design**.

Case Study Background

- Aerial deposition of Pb from a smelter over a town.
- 10 yr ago - most properties cleaned
- These 6 gave confusing data results & thought to be outside deposition area
- Data hinted that highest Pb was in the front yards, along the street
- The street was the main road thru town & was heavily traveled by facility trucks
- Residents suspicious that cast-off from trucks was cause of high Pb & wanted facility to remediate
- Characterization project performed by EPA.
- Any potential remediation – under RCRA

Project Decision Goals

- **Resolve confusion** over past conflicting data
- Estimate **mean** (95% UCL) for the exposure unit (entire yard)
  - Compare to 500 ppm risk-based AL
  - If over, cleanup high concentration areas
- Pb source? Suggested by spatial contaminant pattern
  - Is there evidence the facility is the source & so would be responsible for any cleanup
- Summary: want to compare yard average to AL, but have spatial info to suggest attribution & guide any cleanup

Stratified Data Collection Design

- Yard divided into 3 physical sections (stratum 1, 2, and 3)
  - **S1: Front yard** (very small area)
  - **S2: Side yard** (medium, if present)
  - **S3: Back yard** (large area)
- Each strata divided into 5 ~equal subsections
- Measure area of each yard stratum & subsections
- 1 grab soil sample (~300 g) per subsection into a plastic bag (i.e., 5 samples per yard section)

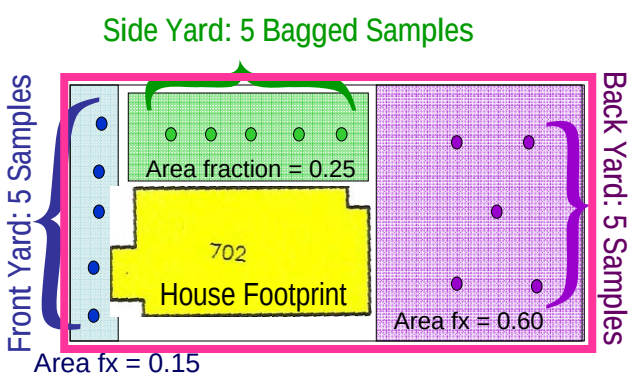
XRF work done in back of truck

Samples brought to the XRF

XRF Pb data from bagged soil samples (~300 gram)

Example Property & Preliminary CSM

Action Level (entire yard) = 500 ppm



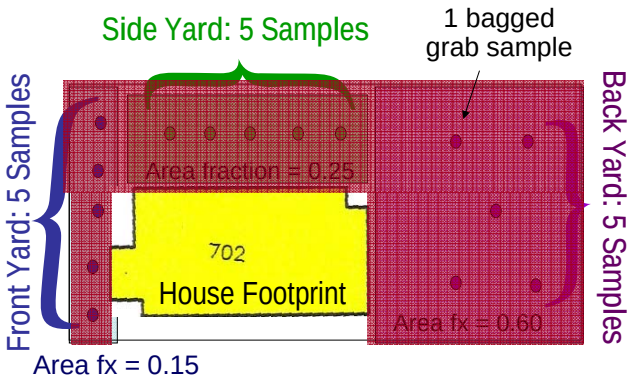
Potential release: facility trucks, Pb'd gasoline, atm deposition from facility, Pb paint, or combination of some or all. Pb conc expected to be high.

Potential release: Pb paint and/or atmos deposition. Expectation for Pb conc uncertain.

Potential release: Pb paint (close to structures) or atmos deposition alone (further from structures). Lower Pb conc expected.

Illustrative Sampling Design & Results

Action Level (entire yard) = 500 ppm



Front yard individual average (at 95% statistical confidence) = 700 +/-150 (**550 – 850** ppm Pb)

Side yard average = 500 +/-100 (**400 – 600** ppm)

Back yard average = 300 +/-50 (**250 – 350** ppm)

Area-weighted total yard average determined statistically as 410 +/- 25 (385 – 435 ppm Pb)

XRF Bag Analysis

- 4 30-sec readings on bag
  - (2 on front/2 on back)
- Results entered real-time into pre-programmed spreadsheet
- Spreadsheet immediately calculates:
  1. ave & SD for each bag
  2. ave & SD within each stratum
  3. ave & UCL for the decision unit (entire property).
  4. within-bag vs. between-bag variability & which is more significant
- **IF** statistical uncertainty interferes w/ desired decision confidence for DU:
  - Use #4 & a series of decision trees to reduce statistical uncertain until a confident decision is possible

Decision Tree #1

Evaluate statistical results for the ***yard*** & compare to the 500 ppm Action Level (AL)

Is there statistical confidence that mean is below AL?

yes

Decide Pb conc for the yard is below AL

200 +/- 50 (150 – 250)

Confident that no action needed

If neither condition is true

Decision too uncertain: more information needed

300 +/- 100 (150 – 520)

Is there statistical confidence that mean is above AL?

yes

Decide Pb conc for the yard is above AL

700 +/- 150 (550 – 850)

Confident that action is required

Go to Decision Tree #2

Decision Tree #2

Determine the greater source of data variability (decision uncertainty)

Is **within-bag** variability GREATER than **between-bag** variability?

yes

Go to Decision Tree # 3

no

Is **within-bag** variability LESS than **between-bag** variability?

no, they are ~equal

Go to Decision Tree #5

yes

Go to Decision Tree #4

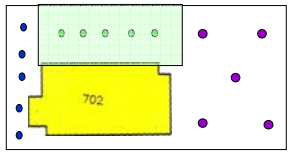
Getting the Data to Feed into Decision Tree #2

Identifying the Most Important Source of Data Variability (aka, statistical "error")

- Look at the ***average within-bag*** "error" (std dev, SD) for each of the 5 bags from a yard section
- Look at the ***between-bag*** "error" SD for all bags from a yard section.
- **Compare** the two: which is larger?
- See example data set

"Error" Measurement in Data Results for a Single **Yard Stratum**

Within-Bag Data Variability



Sect. Bag #1		Bag #2		Bag #3		Bag #4		Bag #5	
Shot #1	700	#1	550	#1	534	#1	769	#1	450
#2	670	#2	534	#2	440	#2	710	#2	400
#3	740	#3	654	#3	430	#3	960	#3	500
#4	650	#4	590	#4	420	#4	800	#4	550
Bag Mean	690		582		456		810		475
W/in-Bag SD	39		54		53		107		65

Mean of within-bag SDs = **(39+54+53+107+65)/5 = *63***

To get between-bag variability

690 582 456 810 475

Between-bag "error" (SD) for 5 bag means = *150*

Decision Tree #2

Determine the greater source of data variability (decision uncertainty)

Is **within-bag** variability GREATER than **between-bag** variability?

yes → **Go to Decision Tree # 3**

no → Is **within-bag** variability LESS than **between-bag** variability?

no, they are ~equal → **Go to Decision Tree #5**

yes → **Go to Decision Tree #4**

Decision Tree #3

Within-bag variability (SD) of Pb replicate results is GREATER than 1.5 x between-bag variability (SD). [Major source of data error is from heterogeneity within the sample (subsampling error)]

To control this source of variability:

Re-shoot each bag another 4 times & add results to spreadsheet & recalculate statistics for bags, for yard sections & for whole yard. Examine results.

Did the within-bag SD decrease so that it is more similar to the between-bag SD?

no → **Re-crush & mix sample bags. Evaluate for paint chips, etc. Shoot another 4 times & re-calculate within-bag SDs.** → (back to the question)

yes → **Is decision uncertainty now resolved?**

no → Go to Decision Tree #6

yes → **Done**

Decision Tree #6

Real-time efforts to reduce data variability have been insufficient to reduce statistical decision uncertainty at the degree of confidence desired.

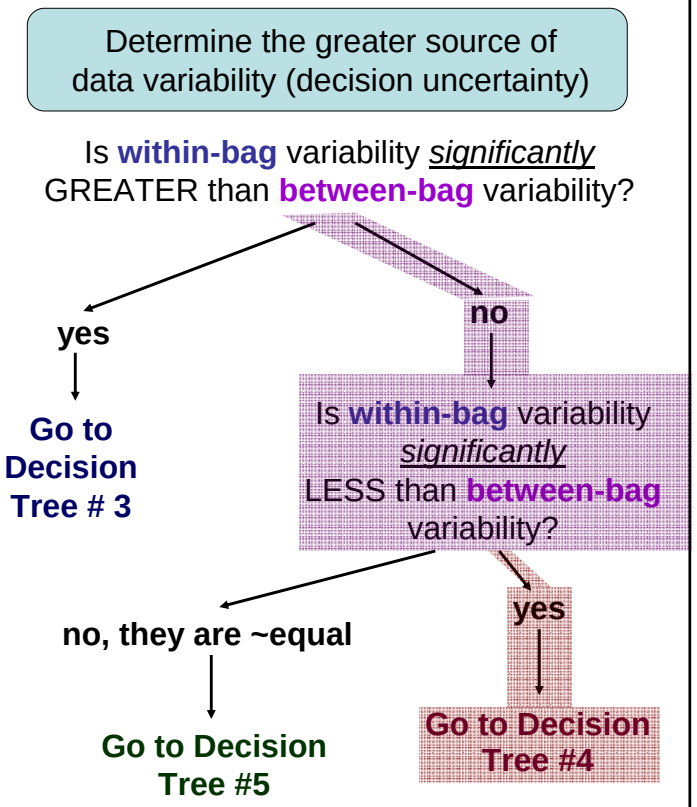
Options for path forward

1) If consequences of "assuming the worst" < cost of add'l sampling & analysis, default to the most protective decision without additional investigation.

2) If add'l investigation preferable to "assuming the worst" & statistical confidence is desired, design a follow-on sampling & analytical program. Perhaps do soil composition analysis for Pb-bearing particles (degraded paint chips, smelting slag, or Pb-battery fragments)

3) Negotiate for accepting a lower statistical confidence

Decision Tree #2



Decision Tree #4

Within-bag variability (SD) of Pb replicate results is LESS than 0.5 x between-bag variability (SD).
(Major source of data error is from concentration variations across the yard section area.)

To control this source of variability:

Collect another 5 bag samples from section area. Analyze 4 times/bag. Add results to spreadsheet & recalculate statistics for yard section & for whole yard.

Did the between-bag SD decrease so that it is more similar to the within-bag SD?

no

Increase sampling density via add'l grab samples, *in situ* XRF, or multi-increment design. Check for paint chips, etc. Calc new between-bag/sample SD.

Go to Decision Tree #6

yes

Is decision uncertainty now resolved?

no

Go to Decision Tree #6

yes

Done

Decision Tree #6

Real-time efforts to reduce data variability have been insufficient to reduce statistical decision uncertainty at the degree of confidence desired.

Options for path forward

1) If consequences of "assuming the worst" < cost of add'l sampling & analysis, default to the most protective decision without additional investigation.

2) If add'l investigation preferable to "assuming the worst" & statistical confidence is desired, design a follow-on sampling & analytical program. Perhaps do soil composition analysis for Pb-bearing particles (degraded paint chips, smelting slag, or Pb-battery fragments)

3) Negotiate for accepting a lower statistical confidence

Decision Tree #2

Determine the greater source of data variability (decision uncertainty)

Is **within-bag** variability *significantly* GREATER than **between-bag** variability?

yes → **Go to Decision Tree # 3**

no → Is **within-bag** variability *significantly* LESS than **between-bag** variability?

no, they are ~equal → **Go to Decision Tree #5**

yes → **Go to Decision Tree #4**

Decision Tree #5

1st Round sampling shows within-bag SD not significantly from between-bag SD. (Concentration variability across the yard section & within sample bags about the same.)

To control these sources simultaneously:

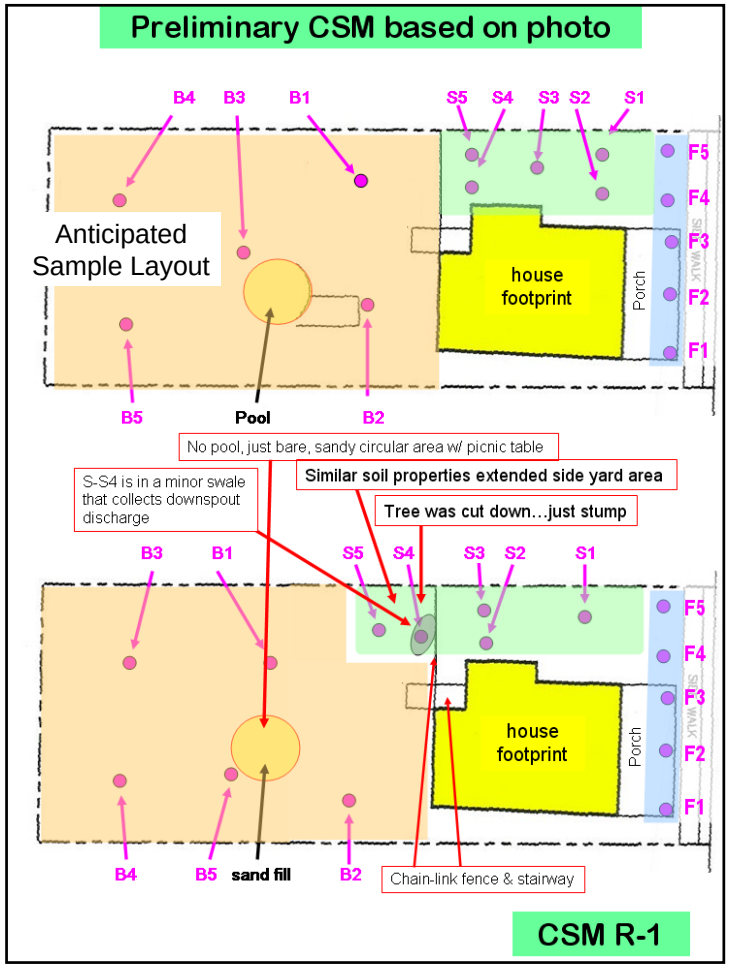
Analyze original bags an add'l 4 times each. Collect another 5 bag samples from the section & analyze 8 times each. Add results to spreadsheet & recalculate statistics for yard section & for whole yard.

Is decision uncertainty now resolved?

no → **Go to Decision Tree #6**

yes → **Done**

Property X CSM Development



After 1st Round of Sampling

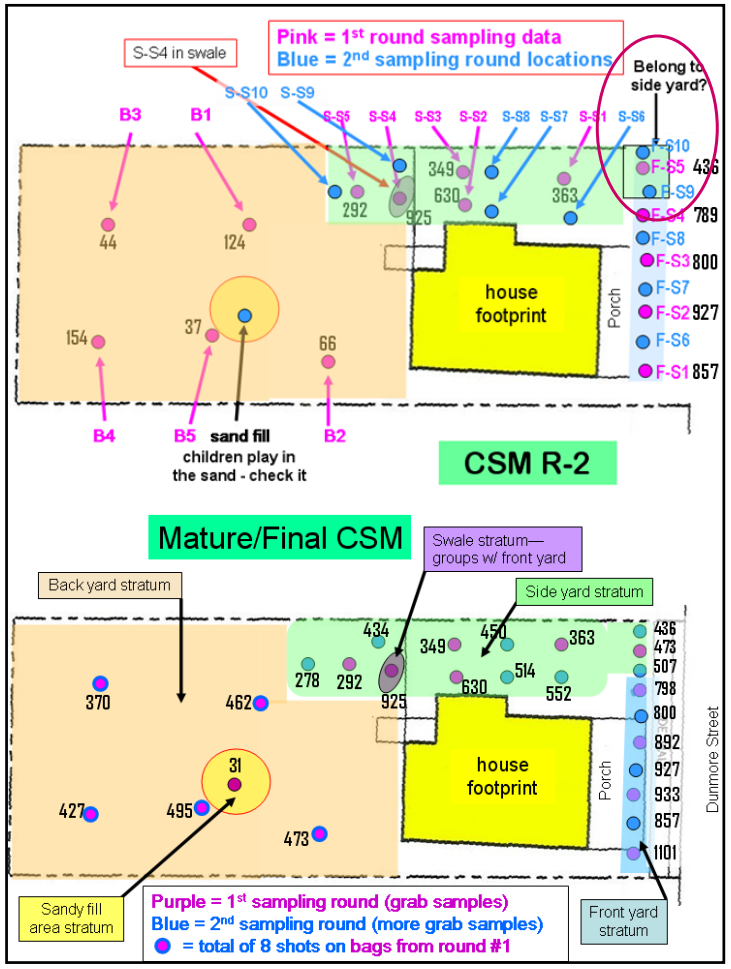
F Section		
	Section Mean =	762
	Section SD =	190
	Section 95%LCL =	526
	Section 95%UCL =	998
	Average within-bag SD =	92
	Between-bag SD =	190
	Within-bag SD > 1.5 x between-bag SD?	no
	Within-bag SD < 0.5 x between-bag SD?	yes

n = 5 bags

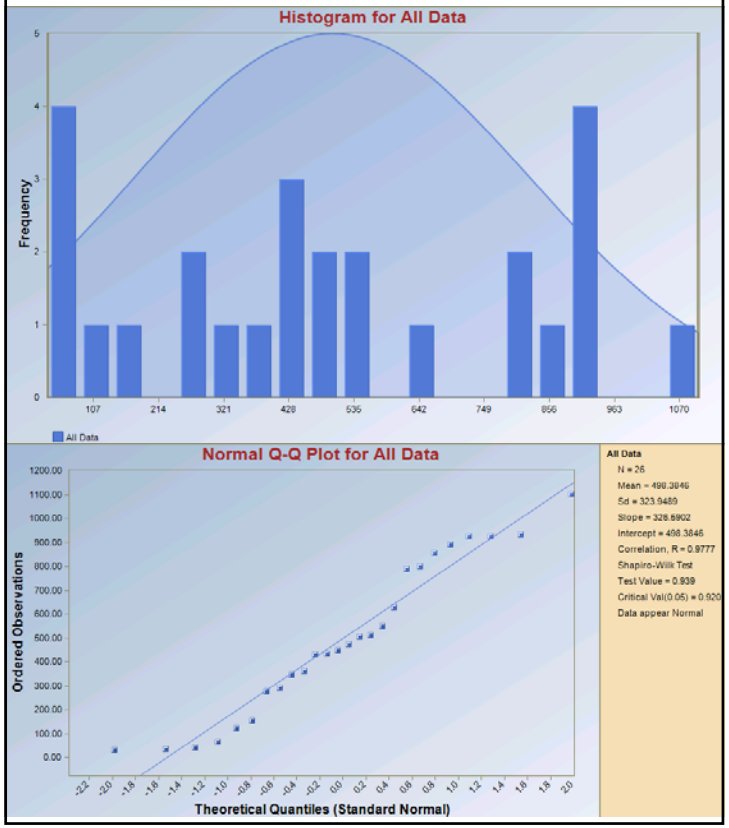
S Section		
	Section Mean =	512
	Section SD =	265
	Section 95%LCL =	183
	Section 95%UCL =	841
	Average within-bag SD =	50
	Between-bag SD =	265
	Within-bag SD > 1.5 x between-bag SD?	no
	Within-bag SD < 0.5 x between-bag SD?	yes

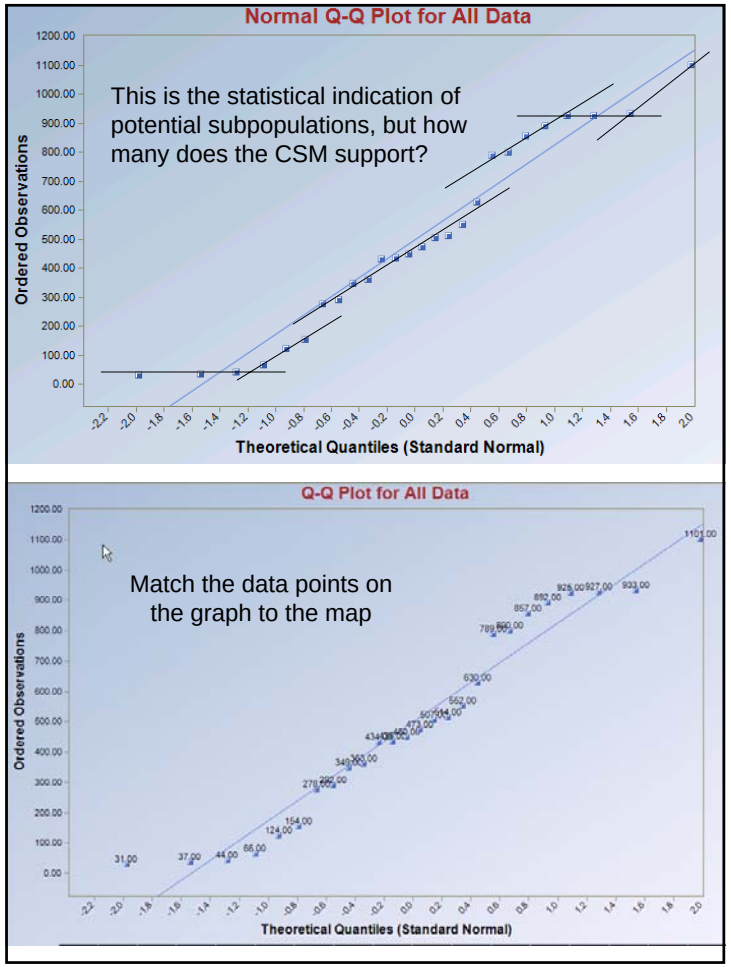
n = 5 bags

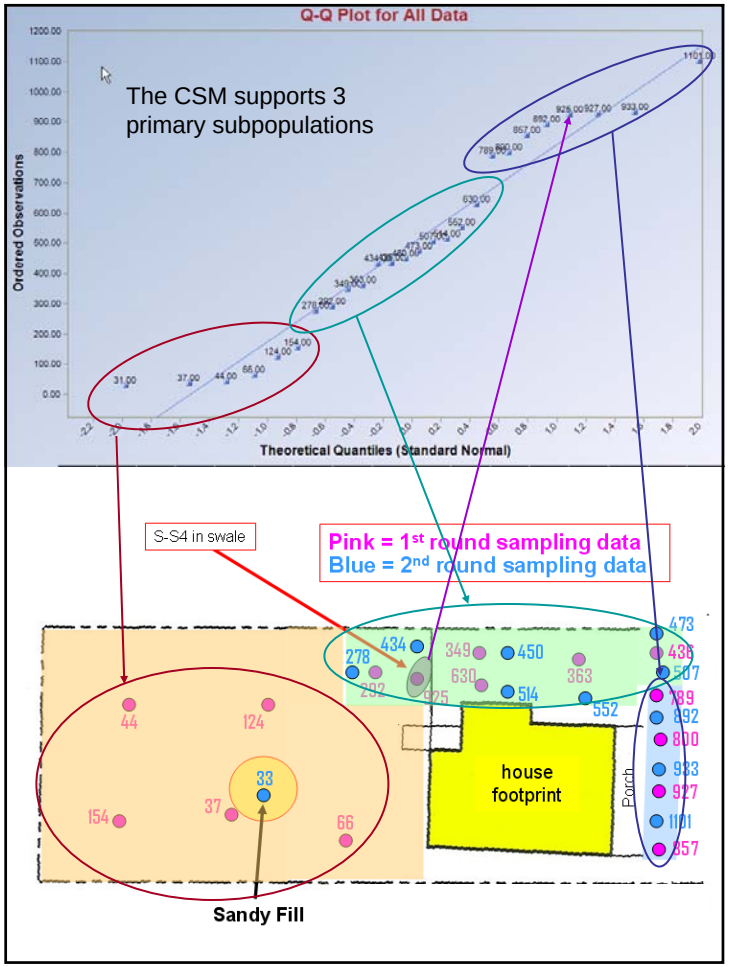
B Section		
	Section Mean =	85
	Section SD =	51
	Section 95%LCL =	21
	Section 95%UCL =	149
	Average within-bag SD =	16.0
	Between-bag SD =	51.4
	Within-bag SD > 1.5 x between-bag SD?	no
	Within-bag SD < 0.5 x between-bag SD?	yes

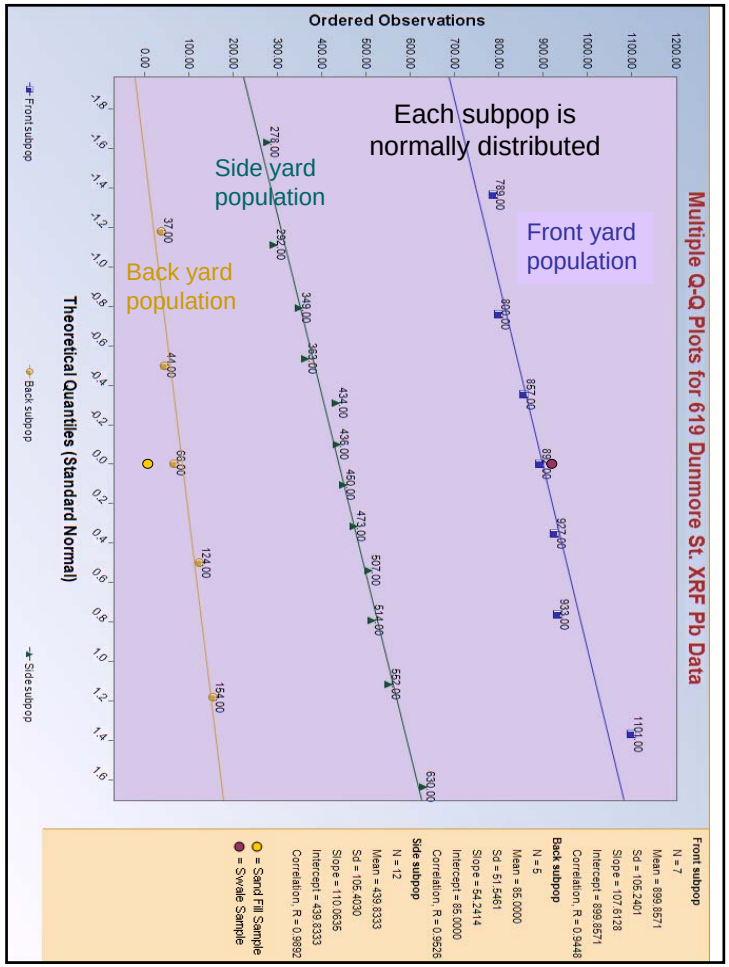


ProUCL Exploration of Entire Data Set






Q-Q Plot for All Data
Ordered Observations
Theoretical Quantiles (Standard Normal)
The CSM supports 3 primary subpopulations
S-S4 in swale
Pink = 1st round sampling data
Blue = 2nd round sampling data
house footprint
Porch
Sandy Fill


Ordered Observations
Theoretical Quantiles (Standard Normal)
Each subpop is normally distributed
Side yard population
Front yard population
Back yard population
Multiple Q-Q Plots for 619 Dunmore St. XRF Pb Data
Front subpop
Back subpop
Side subpop
= Sand Fill Sample
= Swale Sample

Calculation of property Mean & UCL using stratified statistics & Preliminary CSM

Area weighted for the entire property

	mean	stdev	weight	# samples	total area = 5052 sq ft
Front (area = 400)	761	202	0.079	5	
Side (area = 675)	512	265	0.134	5	
Back (area = 3977)	85	51	0.787	5	
	check sum =		1.000	15	= sum
				(no extra sand fill sample)	

Property Mean, Standard Deviation, and LCL/UCLs

Wt'd section	mean	stderror	LCL	UCL	
	196	25	147	245	95% (as 2-sided)
				237	95% (as 1-sided)
				228	90% (as 1-sided)

Calculation of property Mean & UCL for Preliminary CSM & 2nd Data Round

Area weighted for the entire property

	mean	stdev	weight	# samples	total area = 5052 sq ft
Front (area = 400)	771	234	0.079	10	
Side (area = 675)	479	228	0.134	10	
Back (area = 3977)	85	51	0.787	5	
	check sum =		1.000	25	= sum
				(no extra sand fill sample)	

Property Mean, Standard Deviation, and LCL/UCLs

Wt'd section	mean	stderror	LCL	UCL	
	192	21	150	234	95% (as 2-sided)
				227	95% (as 1-sided)
				219	90% (as 1-sided)

Area weighted for the entire yard values & Final CSM

	mean	stdev	weight	# samples	total area = 5029 sq ft
Front (area = 280)	900	105	0.056	7	
Side (area = 705)	440	103	0.140	12	
Back (area = 3700)	97	51	0.736	4	
Swale (area = 90)	925	0	0.0179	1	
Sand Fill (area=254)	34	5	0.051	2	
	check sum =		1.000	26	= sum

Section Mean, Standard Deviation, and LCL/UCLs

Wt'd section	mean	stderror	LCL	UCL	
	201	19	164	239	95% (as 2-sided)
				233	95% (as 1-sided)
				226	90% (as 1-sided)

Progressive measurement of error ("definitive data") & reduction of error

519 Dunmore 1st 15

General Statistics

Number of Valid Observations	15	Number of Distinct Observations	15

Raw Statistics		**Log-transformed Statistics**	
Minimum	37	Minimum of Log Data	3.611
Maximum	927	Maximum of Log Data	6.832
Mean	452.9	Mean of log Data	5.677
Median	363	SD of log Data	1.13
SD	338.9		
Coefficient of Variation	0.748		
Skewness	0.205		

Relevant UCL Statistics

Normal Distribution Test		**Lognormal Distribution Test**	
Shapiro Wilk Test Statistic	0.886	Shapiro Wilk Test Statistic	0.875
Shapiro Wilk Critical Value	0.881	Shapiro Wilk Critical Value	0.881
Data appear Normal at 5% Significance Level		Data not Lognormal at 5% Significance Level	
Gamma Distribution Test		**Data Distribution**	
k star (bias corrected)	1.07	Data appear Normal at 5% Significance Level	
Theta Star	423.1		
MLE of Mean	452.9		
MLE of Standard Deviation	437.7		
nu star	32.11		
Approximate Chi Square Value (.05)	20.16	**Nonparametric Statistics**	
Adjusted Level of Significance	0.0324	95% CLT UCL	596.8
Adjusted Chi Square Value	19	95% Jackknife UCL	607
		95% Standard Bootstrap UCL	590.8
Anderson-Darling Test Statistic	0.55	95% Bootstrap-t UCL	634.1
Anderson-Darling 5% Critical Value	0.758	95% Hall's Bootstrap UCL	590.8
Kolmogorov-Smirnov Test Statistic	0.169	95% Percentile Bootstrap UCL	602.7
Kolmogorov-Smirnov 5% Critical Value	0.227	95% BCA Bootstrap UCL	596.8
Data appear Gamma Distributed at 5% Significance Level		95% Chebyshev(Mean, Sd) UCL	834.3
		97.5% Chebyshev(Mean, Sd) UCL	999.4
Assuming Gamma Distribution		99% Chebyshev(Mean, Sd) UCL	1324
95% Approximate Gamma UCL	721.3		
95% Adjusted Gamma UCL	765.4		
Potential UCL to Use		Use 95% Student's-t UCL	607

If just plug initial 15 data points into ProUCL; compare to stratified at 237

619 Dunmore all 26

General Statistics			
Number of Valid Observations	26	Number of Distinct Observations	26

Raw Statistics		Log-transformed Statistics	
Minimum	32	Minimum of Log Data	3.466
Maximum	1101	Maximum of Log Data	7.004
Mean	464.2	Mean of log Data	5.72
Median	443	SD of log Data	1.131
SD	324.1		
Coefficient of Variation	0.698		
Skewness	0.278		

Relevant UCL Statistics			
Normal Distribution Test		**Lognormal Distribution Test**	
Shapiro Wilk Test Statistic	0.935	Shapiro Wilk Test Statistic	0.845
Shapiro Wilk Critical Value	0.92	Shapiro Wilk Critical Value	0.92
Data appear Normal at 5% Significance Level		Data not Lognormal at 5% Significance Level	
Gamma Distribution Test		**Data Distribution**	
k star (bias corrected)	1.203	Data appear Normal at 5% Significance Level	
Theta Star	385.8		
MLE of Mean	464.2		
MLE of Standard Deviation	423.1		
nu star	62.57		
Approximate Chi Square Value (.05)	45.37	**Nonparametric Statistics**	
Adjusted Level of Significance	0.0398	95% CLT UCL	568.7
Adjusted Chi Square Value	44.41	95% Jackknife UCL	572.7
		95% Standard Bootstrap UCL	565.1
Anderson-Darling Test Statistic	0.898	95% Bootstrap-t UCL	576.2
Anderson-Darling 5% Critical Value	0.766	95% Hall's Bootstrap UCL	566.8
Kolmogorov-Smirnov Test Statistic	0.163	95% Percentile Bootstrap UCL	568.5
Kolmogorov-Smirnov 5% Critical Value	0.175	95% BCA Bootstrap UCL	572.6
Data follow Appr. Gamma Distribution at 5% Significance Level		95% Chebyshev(Mean, Sd) UCL	741.2
		97.5% Chebyshev(Mean, Sd) UCL	861.1
Assuming Gamma Distribution		99% Chebyshev(Mean, Sd) UCL	1097
95% Approximate Gamma UCL	640.1		
95% Adjusted Gamma UCL	654		
Potential UCL to Use		Use 95% Student's-t UCL	572.7

If just plug full 26 data points into ProUCL; compare to stratified at 233

Control of Variability Produces Statistical Confidence for EU

NOTE: "Routine" calculation applies same weighting to data points & databases lose their spatial representativeness
(all rows apply to whole yard)

Strategy & Results for Example Yard	**Mean** (XRF)	**95UCL** (1/2 CI width)
uncontrolled micro-scale (within-bag) variability (single analysis) & routine calc	**476**	**647 (171)**
control within-bag variability (replicates); still use routine EPC calculation (15 samples)	453	607 (154)
stratified sampling & data analysis on **preliminary** CSM (15 samples)	196	237 (41)
stratified sampling & data analysis on **mature** CSM	**201**	**233 (32)**

Note: ½ CI width = mean-to-UCL width

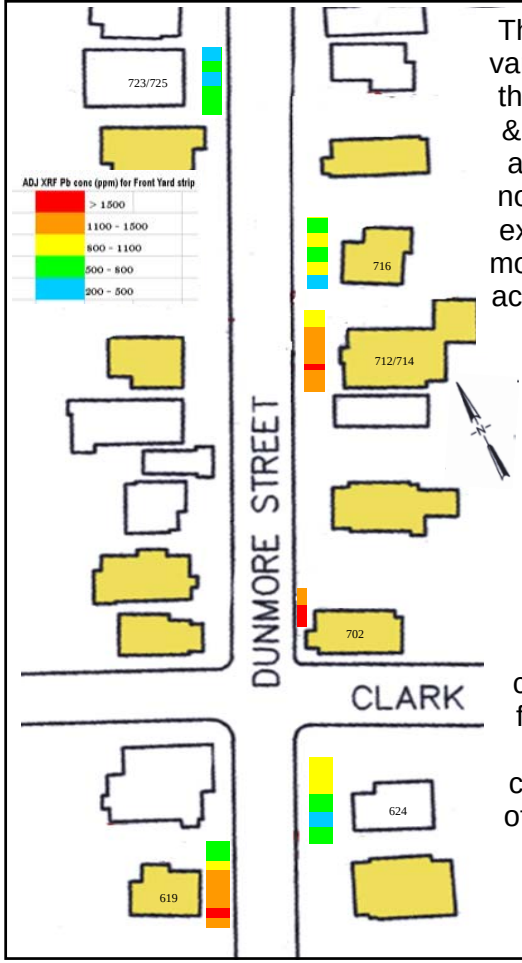
Additive Data Uncertainty (Variability) Components
(1st 4 rows apply to Front Yard only)

Component	Coefficient of Variation (CV)
1 XRF reading on 1 Front yard (FY) bag (instrument-reported error)	0.041 XRF instrument only
1 Bag (4 XRF readings on same bag)	0.087 + micro-scale (within-bag heterogeneity)
Immature CSM, FY section only (10 bag samples)	0.30 + short-scale (between-bag heterogeneity)
Mature CSM, revised FY section only (7 bags)	0.12 + CSM (correctly delineates statistical populations)
Combine w/ Side & Back sections → mature CSM, entire yard (area-wt'd)	0.49 + long-scale variability over property when combining to get property average

Outcome & Decisions

- After waiting 10 yrs, residents had their results that day
- High Pb nearest painted items
- In 2 yards, paint chips present from recent stripping of old paint
  - Toddlers present in worst yard
  - PM provided immediate advice to parents
  - Paint chips tested by XRF
    - 1 multi-layer chip = 18% Pb
    - SCREENING result: XRF calibrated for soil is not accurate for paint—WAY outside linear range
    - Still, the culprit was obvious

So far, no proof that trucks made some contribution



This degree of variability within the front yards & along street as a whole is not the pattern expected from moving sources acting over 25+ yrs

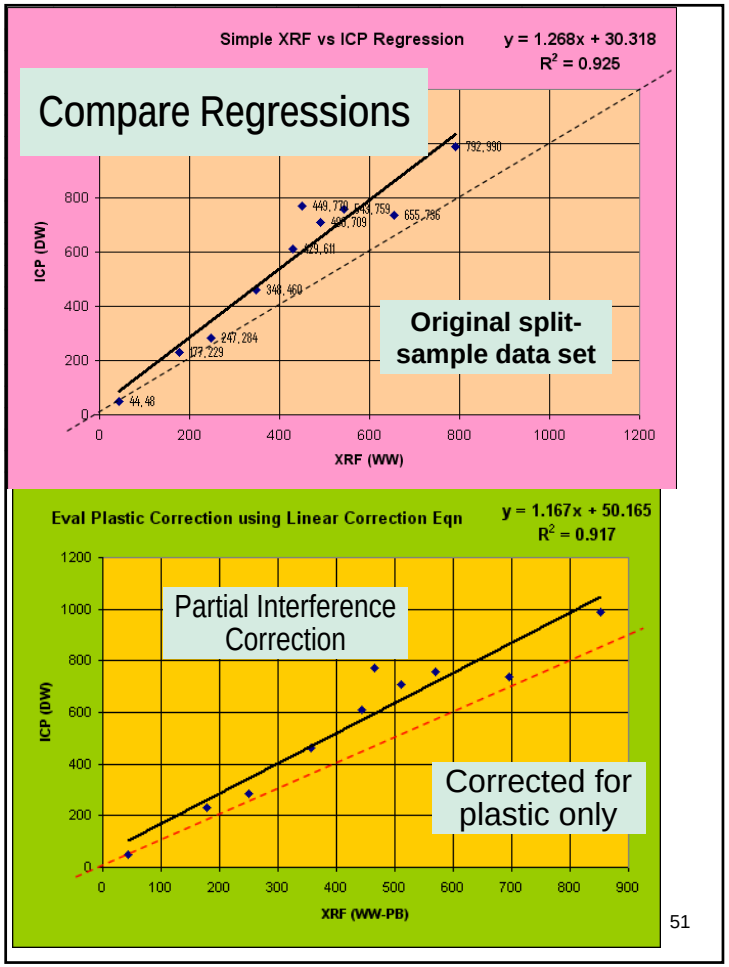
Weight of evidence:

There is no reason to conclude that facility trucks caused Pb contamination of these yards.

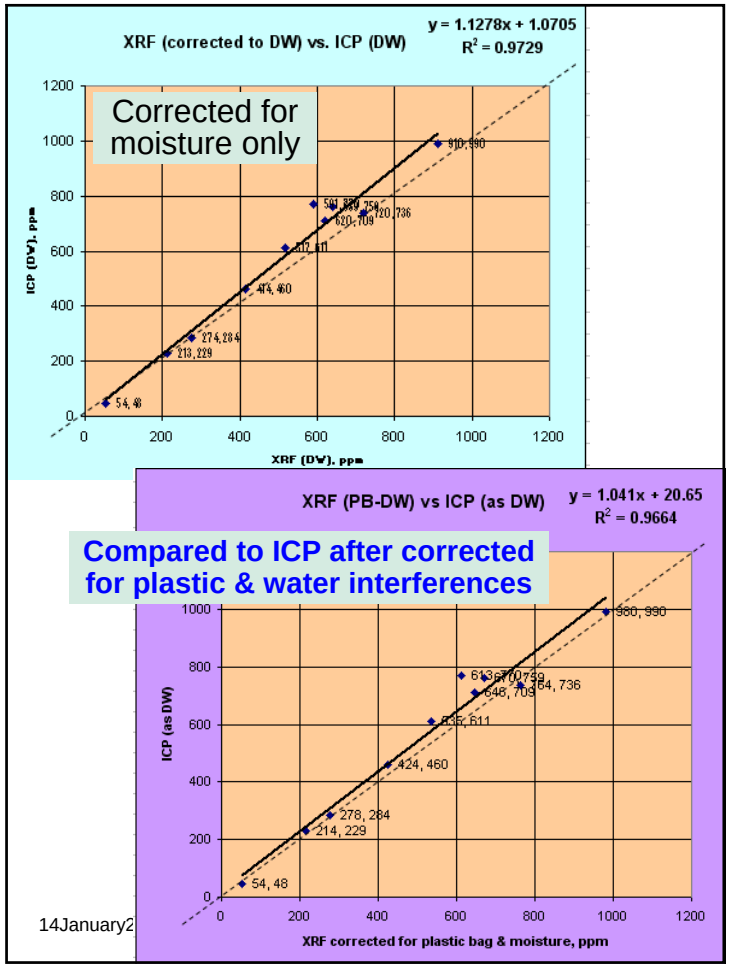
49

What about XRF-ICP Data Comparability?

- There was a problem w/ XRF-ICP comparability
- **XRF was significantly biased *LOW* compared to ICP**
- Investigation found the plastic bags that were used decreased XRF signal.
- When **plastic** interference ***combined with* moisture** >10%, the XRF data needed adjustment to be more comparable to the ICP data.
- Adjusted XRF data were usable for decisions.



51



14January2

Effect of interference corrections on this project's method comparability for bagged samples

(as measured by progressive improvement in the regression eqn)

Original regression (plastic & moisture interferences present)

ICP(DW) = 1.3[XRF(WW)] + 30, R^2 = 0.93

Corrected for plastic bag interf. only

ICP(DW) = 1.2[XRF(PB-WW)] + 50, R^2 = 0.92

Corrected for moisture interf. only

ICP(DW) = 1.1[XRF(DW)] + 1, R^2 = 0.97

Corrected for both interferences

ICP(DW) = 1.0[XRF(PB-DW)] + 21, R^2 = 0.97

Module 7

Wrap Up

Module 7:

Wrap Up

7-1

General and Specific Technical Resources

RESOURCES

7-2

General On-Line Resources

- Clu-In Web site
  http://www.cluin.org
- Brownfields Technology Support Center
  http://www.brownfieldstsc.org
- Field Analytics Encyclopedia Web site
  http://clu-in.org/char/technologies
- Archived Internet seminars
  http://cluin.org/studio/seminar.cfm
- ITRC Web site
  http://www.itrcweb.org
- Argonne National Laboratory ASAP Web site
  http://www.ead.anl.gov/project/dsp_topicdetail.cfm?topicid=23

7-3

More General On-Line Resources

- Free geostatistical-based decision assistance software (SADA)
  http://www.tiem.utk.edu/~sada/
- DOE DQO/statistics training materials Web site & VSP links
  http://www.hanford.gov/dqo/training/contents1.html
- USACE Engineering Manuals (EMs) [Especially see manuals for CSM (EM 1110-1-1200) & systematic planning (TPP) (EM 200-1-2)]
  http://www.usace.army.mil/inet/usace-docs/eng-manuals/em.htm

7-4

Sampling Design Assistance

- Collected items on the Clu-In Web site (www.cluin.org)
  - » Sample Collection and Handling
    http://cluin.org/char1_edu.cfm#samp_coll
  - » Statistics/Sampling Design
    http://cluin.org/char1_edu.cfm#stat_samp
- RCRA Waste Sampling Draft Technical Guidance
  http://www.epa.gov/epaoswer/hazwaste/test/samp_guid.htm
- EPA statistical sampling guidance (USEPA QA/G-5S)
  http://www.epa.gov/quality/qs-docs/g5s-final.pdf
- FRTR long-term monitoring optimization
  http://www.frtr.gov/optimization/index.htm

7-5

Sample Collection Assistance

- Sampling procedures (USEPA Region 4) http://www.epa.gov/Region4/sesd/eisopqam/eisopqam.html and http://www.epa.gov/athens/learn2model/part-one/field/index.html
- USEPA ERT Web page http://www.ert.org/mainContent.asp?section=Products&subsection=List
- EPA ORD Soil Sampling Quality Assurance User's Guide http://www.triadcentral.org/ref/ref/documents/soilsamp.pdf
- EPA ORD Subsampling Guidance http://www.cluin.org/download/char/epa_subsampling_guidance.pdf

(continued)

7-6

Sample Collection Assistance

- ASTM D6232: Selecting Sampling Equipment
- USACE CRREL Reports http://www.crrel.usace.army.mil/products/products.html
- USACE Waterways Experimental Station reports http://itl.erdc.usace.army.mil/library/
- VOCs in solid samples
  - » EPA OSW developing sampling guidance; USACE also has guide available - see: http://cluin.org/char1_edu.cfm
- Explosive residues in soil sampling design & handling guidance in SW-846 Method 8330 (see App. A) http://www.epa.gov/epaoswer/hazwaste/test/pdfs/8330b.pdf

7-7

Innovative Sampling Technologies
Direct Push

- In situ measurement of subsurface properties (stratigraphic logging) with CPT
  - » DOE Innovative Technology report on CPT http://web.em.doe.gov/plumesfa/intech/conepen/index.html
  - » EPA information: Direct Push technologies
    - — http://www.epa.gov/athens/learn2model/part-one/field/b-probing_field.htm
    - — http://www.epa.gov/swerust1/pubs/esa-ch5.pdf
    - — http://clu-in.org/char/technologies/

7-8

Selected Articles Describing the Triad Approach

See the Technical Components & References sections in the Triad Resource Center:

http://www.triadcentral.org/ref/index.cfm

- Interstate Technology & Regulatory Council TechReg Guideline for Triad: http://www.triadcentral.org/ref/ref/documents/SCM-1.pdf
- 2001 ES&T "Managing Uncertainty in Environmental Decisions" article: http://www.triadcentral.org/tech/documents/oct01est.pdf
- 2001 Quality Assurance journal "Representativeness" article: http://www.triadcentral.org/tech/documents/dcrumbling.pdf

(continued) 7-10

Selected Articles Describing the Triad Approach

- 2003 Remediation journal "Next Generation Practices" article: http://www.triadcentral.org/tech/documents/spring2003v13n2p91.pdf
- 2003 Remediation journal "Insurance" article: http://www.triadcentral.org/ref/doc/Remediation_preprint_Triad-Insurance.pdf
- Fall 2004 Remediation journal "Triad Myths" article: http://www.triadcentral.org/ref/doc/Fall04RemediationArticlePostprint.pdf

(continued)

7-11

Selected Articles Describing the Triad Approach

- Winter 2004 Remediation journal articles:
  - "Triad as Catalyst" article: http://www.triadcentral.org/ref/doc/RemediationCatalystPostprint.pdf
  - Triad Case Study: Rattlesnake Creek: http://www.triadcentral.org/ref/doc/TriadCaseStudy_RattlesnakeCreek.pdf
  - Triad Case Study: Marine Corps Base Camp Pendleton http://www.triadcentral.org/ref/doc/Winter_04_Remediation_Preprint_Navy_Case_Study.pdf
  - Triad Case Study: Former Small Arms Training Range http://www.triadcentral.org/ref/doc/ShawTriadCaseStudypreprint.pdf

7-12

Final Questions?

7-13

Final Words

- Please fill out the evaluation form
  - »They help us to continually improve this course
- We hope this course helps you implement effective and efficient sampling programs using XRF!
- GOOD LUCK in your future endeavors!

7-14