Incremental-Composite Sampling (ICS) and XRF: Tools for Improved Soil Data

Deana Crumbling
USEPA Office of Superfund Remediation and Technology Innovation
Technology and Field Services Division
crumbling.deana@epa.gov 703-603-0643

1

Take-Away Points

- **The Problem:** Soil data can mislead decisions about risk & cleanup!
- **Why?** Common practice generates a concentration result from a *few grams* of soil and then *assumes that tons* of soil in the field have that *same* concentration.
- This presentation will show:
  - "Representativeness" for soil samples is only meaningful in terms of a sample, or a set of samples, that provide an **average** over some defined soil mass.
  - **"Sample representativeness" does not exist** until the RPM defines for a *specific* sampling event what field soil volume and particle size a soil sample is supposed to represent.
- A defined field soil volume/mass is called a **Decision Unit** (DU); DUs must be described in the QAPP/FSP.

2

Using incremental and/or composite sampling vastly improves the representativeness of soil or sediment data. Why can I make that claim?

Soil Sampling Is NOT Simple

- Effect of short-scale, *between*-sample heterogeneity
  - A grab field sample does not represent the field concentration
  - Misleading data possible if decision based on 1 grab sample
  - Remedy: In the **field**, use large discrete data sets or many-increment composites, use QC checks on sampling design
- Effect of micro-scale, *within*-sample heterogeneity
  - A grab analytical subsample does not represent the sample
  - Misleading data possible if decision is based on 1 grab subsample
  - Remedy: In the **laboratory**, isolate target soil particle size, avoid sample segregation errors, match subsample mass to sample particle size, form subsample from many increments

3

Using incremental and/or composite sampling vastly improves the representativeness of soil or sediment data. Why can I make that claim?

Tools for Reliable Soil Data Are Available

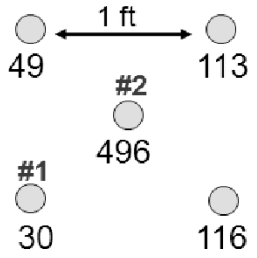
- Incremental-composite sampling (ICS) addresses:
  - Short-scale heterogeneity by collecting many field increments
  - Micro-scale heterogeneity by specialized sample processing and subsampling procedures
- X-ray fluorescence (XRF) instruments
  - ICS + real-time XRF data = powerful, efficient sampling designs
  - XRF can guide real-time, in-field choice of increment number, set DU boundaries & evaluate sample processing
  - Proper XRF application requires sufficient QC and documentation
    - XRF & ICP comparisons usually done incorrectly

4

Using incremental and/or composite sampling vastly improves the representativeness of soil or sediment data. Why can I make that claim?

Short-Scale Heterogeneity

- Differences in concentration at the scale of collocated field QC samples (inches to a few feet)
- Collocated samples are considered equivalent, but very different results are common
- If decision is based on a single grab sample, **chance** ("the luck of the grab") may determine outcome
- Decisions based on single samples:
  - "Hot spot" presence/absence
  - Drawing concentration contour lines



Set of collocated samples for uranium (mg/kg)

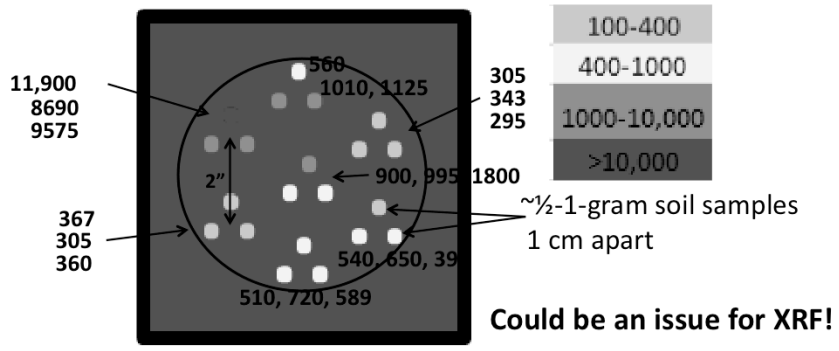
5

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

Very Short Short-Scale Heterogeneity

Figure: 21 separate ~½-1-gram samples (~16 g total) within a 4-inch diameter circle with ½-in depth (analyzed by ICP)

Assumed mean for the 160 g in the 6.5-cu.in. volume = 1994 ppm



Could be an issue for XRF!

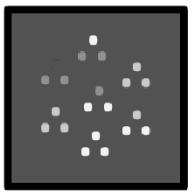
6

Joanna Becker, Perdue Univ. PhD thesis, 2005, Centimeter scale analysis of soil heterogeneities within a long-term, heavy metal contaminated site.
Becker, Joanna M., T. Parkin, C.H. Nakatsu, J.D. Wilbur and A. Konopka (2006) Bacterial Activity, Community Structure, and Centimeter-Scale Spatial Heterogeneity in Contaminated Soil. Microbial Ecology Vol. 51, 220-231.

Mass of soil in 4-inch circle (to ½-inch depth) = 160 g (assuming soil density of 1.5 g/ cubic cm)

A Grab Sample is "Representative" of …?

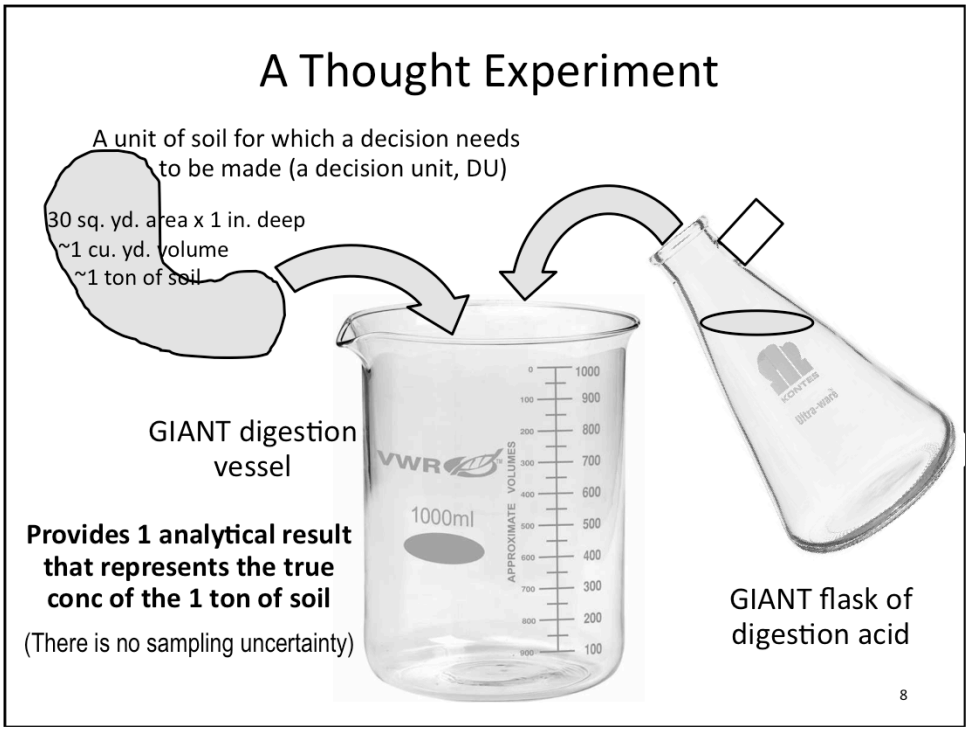
- …its own mass.
  - Do you make decisions at the scale of 100 grams?
- Is there evidence that a jar represents a larger field volume?
- "Sampling uncertainty": Unmanaged heterogeneity raises the question of whether the sample's concentration is the same as (i.e., represents) the concentration of a larger mass.



Think about the typical dimensions for the soil you make decisions about…

…the concentration for <u>that mass</u> is what you need to know.

7



If the entire DU could be analyzed in a single giant analysis, there would be no uncertainty about the true Pb concentration. Note that this process would produce a result that represents a giant composite of all soil particles in the DU.

Alternative:
Divvy the Whole Mass into Analytical Samples

Analyze entire 1-ton mass as 1-gram analytical samples (·)

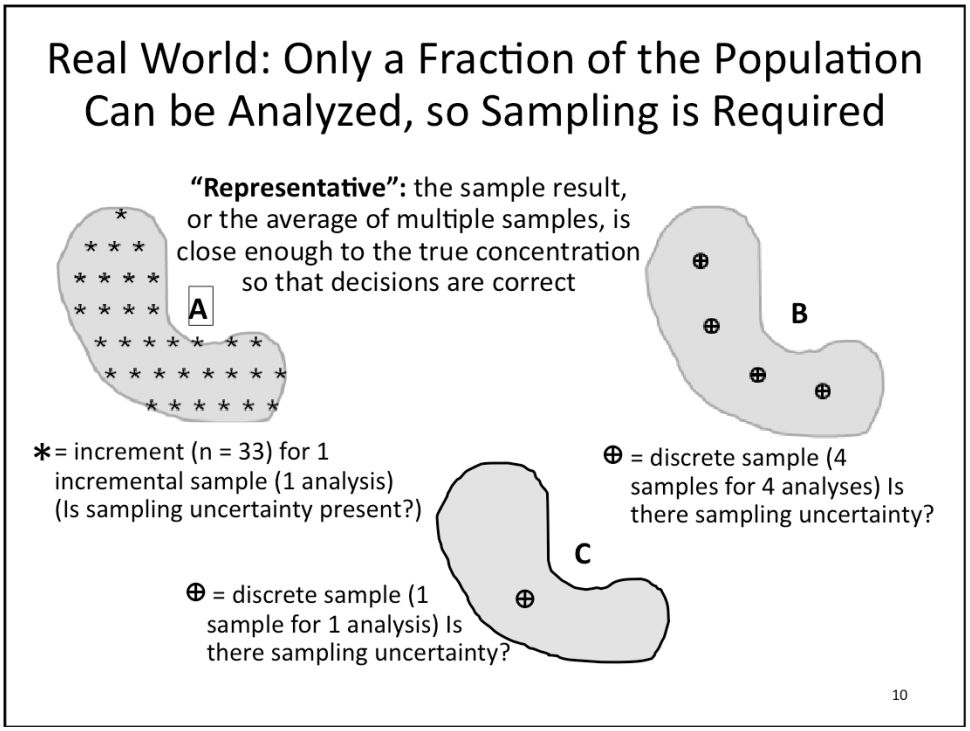
n = 1.4 x 10^6 samples & analyses = the statistical "population"

Take the 1.4 million data results & calculate their average = true conc for the 1-ton soil mass.

(Since the entire population of 1-gram samples is analyzed, there is no sampling uncertainty)

9

Assume no analytical error.



For this thought experiment, again assume that there is no analytical error.

Since a single DU cannot be analyzed in a single analysis, we must take samples, analyze them, and then draw conclusions about the DU concentration from the concentration of the samples.

In scenario A, we take more samples (n = 33), but it is expensive to analyze them all. So we perform a physical averaging by combining all the samples (now called increments) together to form a single composite called an incremental sample, which is analyzed. This is equivalent to taking 33 samples and analyzing all individually, then mathematically averaging all 33 results. Because this is not an analysis of the entire volume, there is uncertainty about how close the sample average is to the true concentration. With only 1 analysis, it is not possible to determine how much uncertainty is present in the result. However, if we take multiple independent incremental samples, we can determine uncertainty.

In scenario B, we take 4 discrete (grab) samples. Because we want to use those samples to determine the actual concentration for the entire DU, we take the average of the 4 data points. Since we are using 4 small samples taken from a heterogeneous medium (soil), there is uncertainty in whether the average of the 4 data points accurately represents the concentration for the DU. We can calculate an estimate of the uncertainty from the variability between the 4 results.

In scenario C, we take 1 discrete sample. There is sampling uncertainty present, but we have no way to estimate how large that uncertainty is.

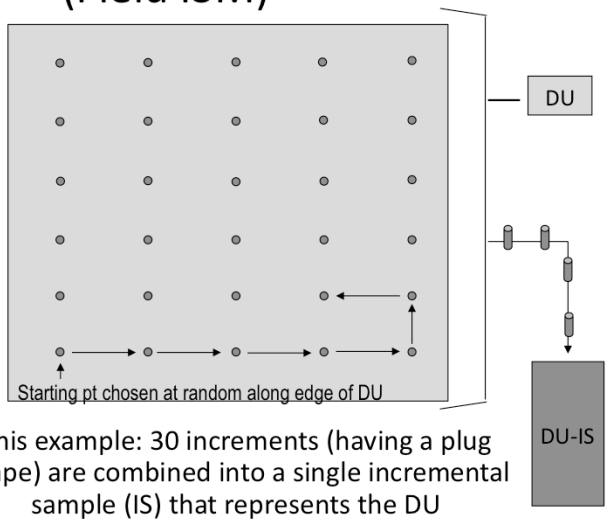
Which design looks like it would be more representative of the true concentration of the DU?

Incremental Sampling Methodology (Field ISM)

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

ISM definitive guidance is the ITRC ISM Tech Reg web document

www.itrcweb.org/ISM-1



This example: 30 increments (having a plug shape) are combined into a single incremental sample (IS) that represents the DU

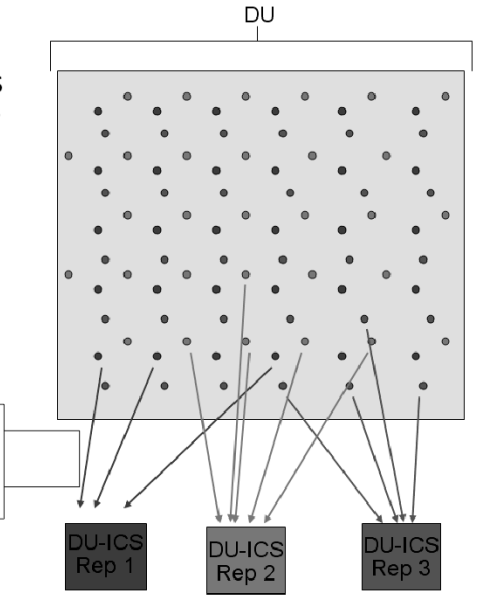
11

Replicate ICSs per DU

Need at least 3 independent replicate ICSs if want to calculate UCL or measure data uncertainty

Each replicate ICS result represents an estimate of the DU mean.

Example: 3 replicate ICSs of 30 increments each = 90 increments total in DU



12

Sample Processing & Correct Subsampling Critical for Reliable Data

- Micro-scale, within-sample heterogene caused by differences in particle size & composition
- Tiny particles are often composed of minerals that readily adsorb contaminants
  - Iron oxides
  - Clay minerals
  - "Contamination is in the fines"

See "Reference version" for this PPT presentation for more details.

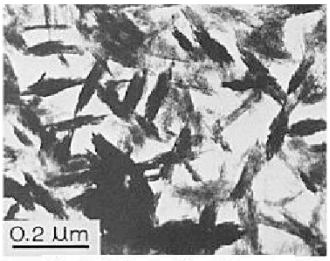
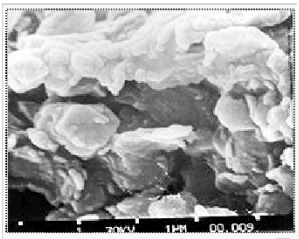


Fig. 2.19 Transmission electron micrograph showing clusters of many small acicular goethite crystals (courtesy of A. Suddhiprakarn and R. J. Gilkes).



electron microscope photograph of smectite clay - magnification 23,500

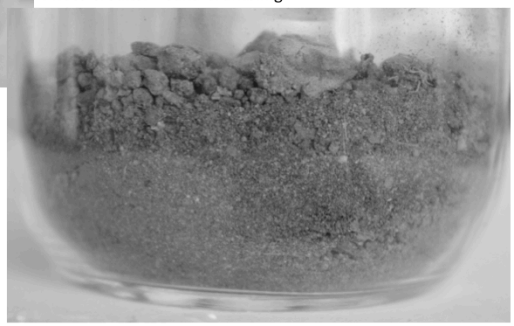
13

Particle Size vs. Routine Lab Subsampling

Freshly collected soil sample – Particles of many sizes & composition

Photo credits: Deana Crumbling

Same sample jar after jostling to mimic transport to lab: particles segregate.



What if just scoop subsample off the top?

14

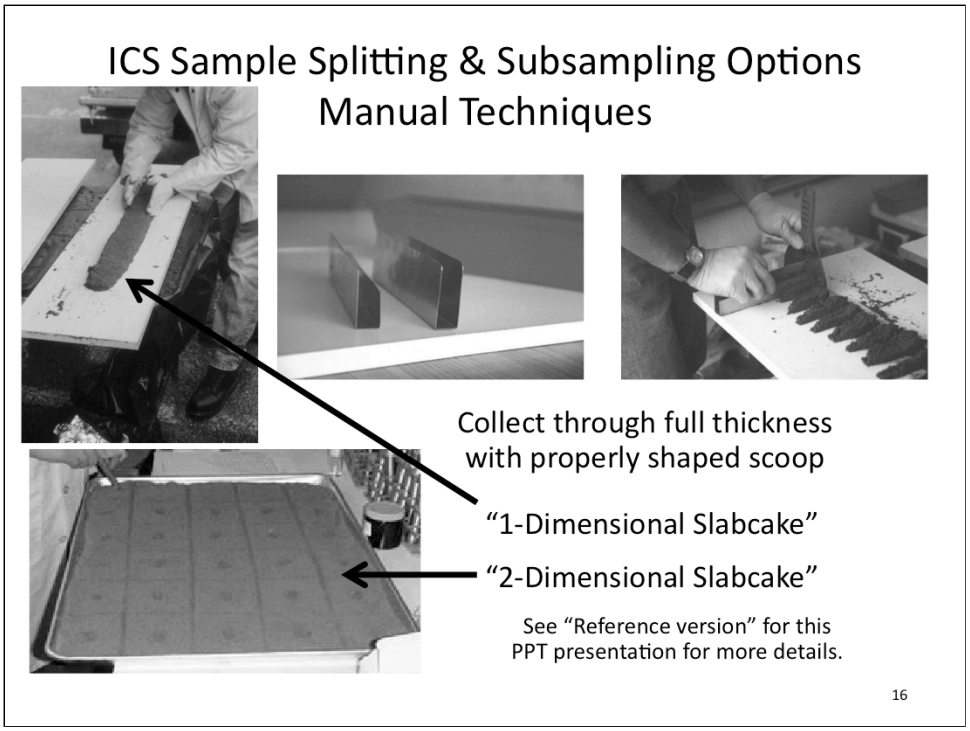
Here's what we mean by "particle segregation."

- These photos contrast non-segregated soil with segregated soil
- With shaking or jiggling, larger particles migrate to the top while smaller particles settle downward
- Stirring to "mix" is ineffectual to redistribute particles; often makes segregation worse
- If subsampling involves scooping off the top, could predominately get larger particles; but this depends on another factor (see next slide)

Micro-Scale Heterogeneity & Sample Handling

- Labs assume the sample they get is ready for analysis "as is"
- May stir to "mix"— makes particle segregation worse
- Lab duplicates often don't match
  - Reveals need for better sample processing & subsampling
- Good sample processing may include drying, disaggregation, sieving, and perhaps grinding
  - Match subsample mass to soil particle size (see equation in EPA530-D-02-002, Aug 2002, App. D)
- Subsampling performed using incremental technique or mechanical splitting
- QC includes replicates to calculate subsampling precision
- See "Reference version" for this PPT presentation for details

15



Speaker bullets

- 2 D slabcake
- Lower cost than sectorial splitter
- Pretty good representativeness
- Wet or dry sample
- Systematic Random design
- All increments combined = analytical subsample

Narrative

The 2 dimensional Japanese slabcake frequently provides acceptable subsample representativeness at a lower cost than the sectorial splitter. This process is a miniaturized version of what takes place in the systematic random field sample collection process. The wet or dry processed sample is spread evenly in rectangular slabcake and divided into grids as determined in project planning. The default is 30. The analyst removes a small increment from a random location in the first grid. Subsequent increments are collected from the same location in the other grids. All increments are combined to form the subsample for digestion or extraction so the size of the increment must be appropriate for the number of increments and the target subsample size.

2D slabcake subsampling can minimize bias and improve precision.

Supplemental information

See Section 6.2.2.7

Advantages and Limitations of Incremental Sampling

Advantages	Effect
Improved spatial coverage (increments x replicates)	• Sample includes high and low concentrations in same proportions as present within decision unit (DU)
Higher field sample mass	• Sample is more representative of field conditions; statistical distribution of replicate results is normalized
Optimized processing	• Reduces subsampling errors so analytical sample is more representative of field sample
Fewer non-detects	• Simplifies statistical analysis
More consistent data	• More confident decisions; more regulator & RP agreement on data interpretation

Limitations	Effect
Small number of replicates	• Limits UCL calculation methods (t-UCL & Chebyshev-UCL)
No spatial resolution within Decision Unit	• Limits remediation options within a DU unless a more complex ICS design is used or have 2nd remobilization

17

ISM has both advantages and disadvantages from a sampling design perspective.
Can't directly compare discrete and ISM samples because each measure different properties of the population.

Under disadvantages, discrete sampling allows for calculations of ratios of two variables – allows for correlations among constituents, or estimates of bioaccumulation factors (update from abiotic media to organisms) that you cannot get from ISM.
When assessing acute toxicity issues, the decision unit would have to be very small for incremental sampling. ISM may not be practical.

XRF: Great Partner with Incremental Sampling for Metals Analysis in Soil

18

Managing XRF's Micro-Scale Heterogeneity

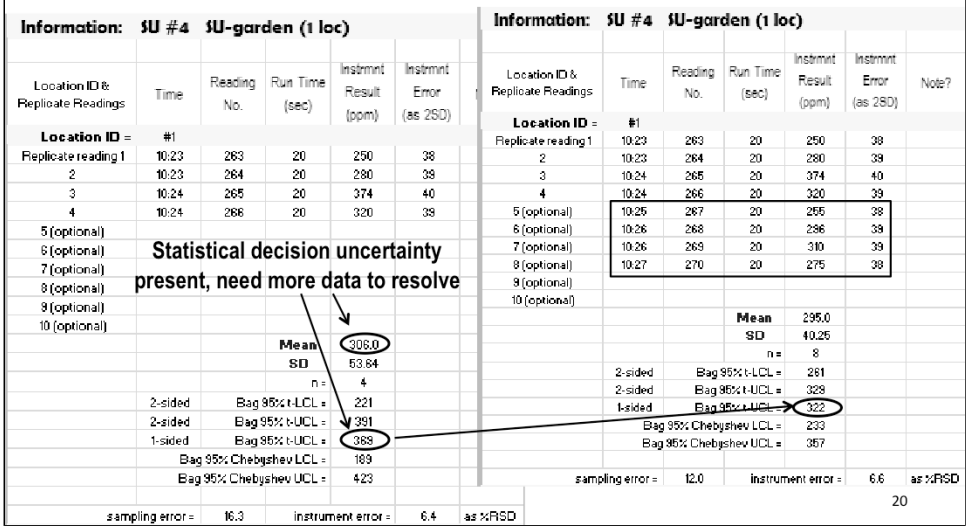
- Use replicate readings to understand degree of short-scale (for *in situ* readings) and micro-scale heterogeneities
- Replicate readings can substitute for, or complement, sample processing
  - Use reps' arithmetic average as the "result"
  - How many XRF replicate shots? Depends on data variability & closeness to decision threshold; decide in real-time.
  - How many seconds of read time? Depends on desired quant limit
  - Program the calculations into spreadsheet for fast decision-making

- Replicate readings do not add any consumables cost (only labor)

19

Programmed Spreadsheet

After the initial 4 readings per bag, can take additional readings until decision (Is the mean conc < 350 ppm?) is without statistical uncertainty, i.e., the 95% upper confidence limit (UCL) is < 350 ppm.



Information: SU #4 SU-garden (1 loc)

Location ID & Replicate Readings	Time	Reading No.	Run Time (sec)	Instrmnt Result (ppm)	Instrmnt Error (as 2SD)
Location ID =	#1				
Replicate reading 1	10:23	263	20	250	38
2	10:23	264	20	280	39
3	10:24	265	20	374	40
4	10:24	266	20	320	39
5 (optional)					
6 (optional)					
7 (optional)					
8 (optional)					
9 (optional)					
10 (optional)					
			Mean	306.0	
			SD	53.84	
			n =	4	
	2-sided	Bag 95% t-UCL =		221	
	2-sided	Bag 95% t-UCL =		391	
	1-sided	Bag 95% t-UCL =		369	
		Bag 95% Chebyshev LCL =		189	
		Bag 95% Chebyshev UCL =		423	
sampling error =		16.3	instrument error =	6.4	as %RSD

Statistical decision uncertainty present, need more data to resolve

Information: SU #4 SU-garden (1 loc)

Location ID & Replicate Readings	Time	Reading No.	Run Time (sec)	Instrmnt Result (ppm)	Instrmnt Error (as 2SD)	Note?
Location ID =	#1					
Replicate reading 1	10:23	263	20	250	38	
2	10:23	264	20	280	39	
3	10:24	265	20	374	40	
4	10:24	266	20	320	39	
5 (optional)	10:25	267	20	255	38	
6 (optional)	10:26	268	20	286	39	
7 (optional)	10:26	269	20	310	39	
8 (optional)	10:27	270	20	275	39	
9 (optional)						
10 (optional)						
			Mean	295.0		
			SD	40.25		
			n =	8		
	2-sided	Bag 95% t-UCL =		281		
	2-sided	Bag 95% t-UCL =		328		
	1-sided	Bag 95% t-UCL =		322		
		Bag 95% Chebyshev LCL =		233		
		Bag 95% Chebyshev UCL =		357		
sampling error =		12.0	instrument error =	6.6	as %RSD	

20

For more information, or to obtain a copy of the spreadsheet, contact Deana Crumbling, USEPA, crumbling.deana@epa.gov

See “Reference version” for this PPT presentation for details.

Warnings about XRF-ICP Data Comparability

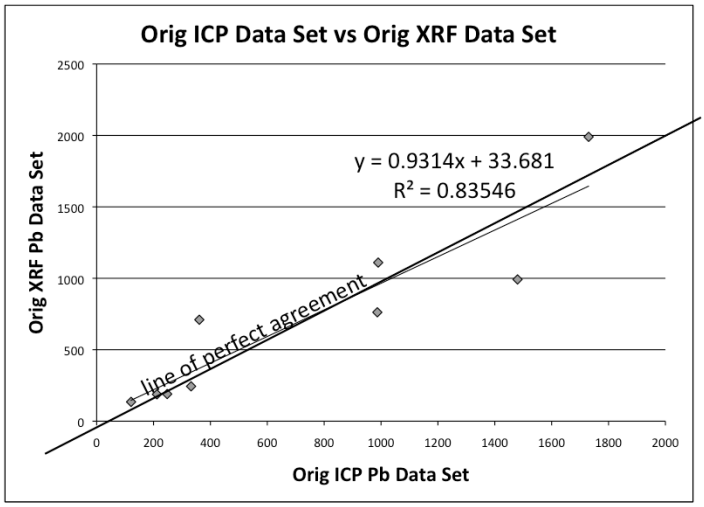
- "Comparability" refers to comparing XRF results to lab data
- SAME samples must be analyzed by XRF and lab
- Regression analysis commonly used to measure comparability; generates a line: $y = mx + b$
- R^2 is the commonly used "goodness" metric…

BUT IT SHOULD NOT BE!!

  - R^2 greatly influenced by sampling error: XRF data cannot match ICP data any better than ICP data can match itself!
  - m (slope) & b (intercept) are more important than R^2:
    - Intercept measures "bias", the difference between total metal (via XRF) & dissolvable/"available" metal (via 3050B digestion & ICP)
    - Slope should be close to 1.0
    - Regression line should be close to "line of perfect agreement"

21

Common ICP vs XRF Regression Techniques Ignore the Effects of Sampling Variability

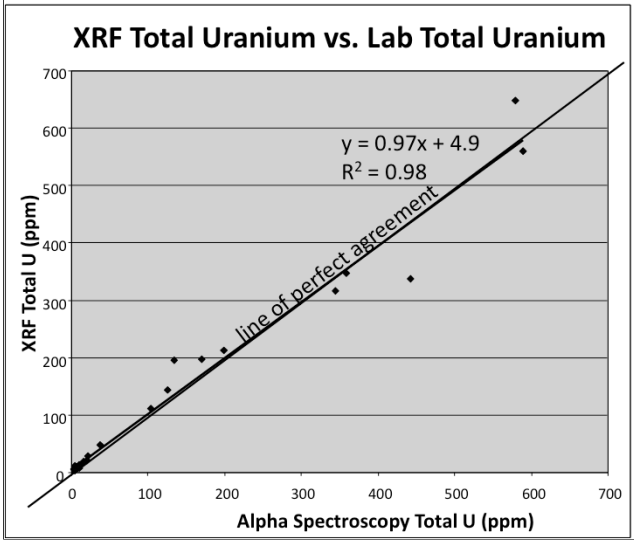


Falsely assumes the ICP data are without error; any differences "blamed" on XRF

Orig ICP	Orig XRF
332	244
121	135
212	188
1730	1990
1480	992
990	1110
248	190
361	710
987	762

22

When Sampling Variability is Controlled, XRF-Lab Comparability Can be Excellent



Other factors that can degrade comparability:

- Differences in moisture content
- Plastic bags holding XRF samples not free of interferences (this is easily checked before the start of the project).
- Interfering minerals and elements

23

Comparability Done the Right Way

Each comparability sample is analyzed twice by both methods

1 XRF bag or cup, 2 XRF readings (orig & dup) → Send bag/ cup to the Lab → 2 separate ICP analyses (orig & dup)

Analyte is Pb (ppm)

Sample ID	orig XRF	dup XRF	orig ICP	dup ICP	
NW5-B-0-2/D	244	371	332	412	
SWMW5-D-0-2.D	2150	1600			ICP results in the 3000 range--delete from plot
NW3-D-0-2/D	135	152	121	133	(because the single high value biases the regression)
WS4-B-0-2/D	188	156	212	240	
NW17-B-0-2/D	1990	2500	1730	2580	
W5-A-0-2/D	992	954	1480	1420	
NW12-A-0-2/D	1110	1010	990	1000	
WS7-D-0-2/D	190	141	248	223	
SWNW1-B-2-10/D	710	555	361	595	For more info, contact Deana Crumbling,
SWNW5-B-2-10/D	762	460	987	1030	crumbling.deana@epa.gov

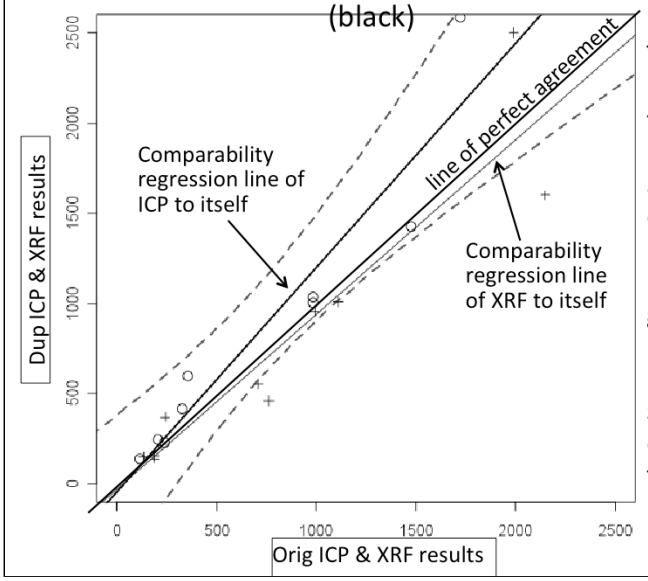
Measures: 1) how well XRF dups agree;
2) how well ICP dups agree; and
3) how well XRF & ICP agree

24

For more information, contact Deana Crumbling, USEPA, crumbling.deana@epa.gov

An Unbiased Regression Technique for Comparability

95% confidence interval (dashed red lines) bound the ICP vs ICP-dup regression line (black)



The XRF-XRF dup regression line (blue) falls within ICP's CI (red), meaning the XRF data is as comparable to the ICP data as the ICP is to itself. Near the action level (400), there is good agreement.

Note that the XRF line stays closer to the line of perfect agreement than the ICP line.

25

For more information, contact Deana Crumbling, USEPA, crumbling.deana@epa.gov

Using XRF to Guide Aspects of Incremental Sampling for Metals

26

XRF & ICS: Perfect Together

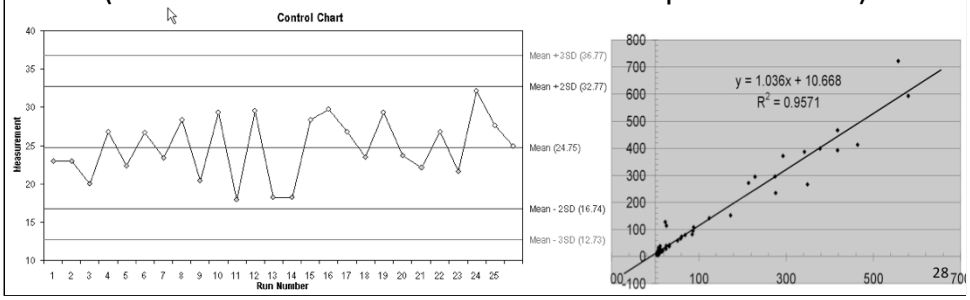
- XRF aids developing and verifying ICS sample processing procedures prior to lab metals analysis.
- Set DU boundaries to avoid mixing large "clean" and "dirty" areas into same DU for purposes of remediation & source delineation.
- Use XRF to approximate mean and SD across a DU.
  - How many increments per incremental sample?
  - 😐 Enlarge the XRF sample support to ~same mass as the increment sample support, or will over-estimate between-increment variability!
- Use XRF to evaluate IS samples before leave the DU:
  - Do you have enough replicate ISs to meet statistical decision goals?
  - See "Reference version" for this PPT presentation for details

27

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

Ensure Sufficient & Appropriate XRF Quality Control

(see "Reference version" for this PPT presentation)



28