Conducting a Demonstration of Method Applicability and Designing Quality Control Programs for X-Ray Fluorescence in Soil

ConSoil 2008 – Milan

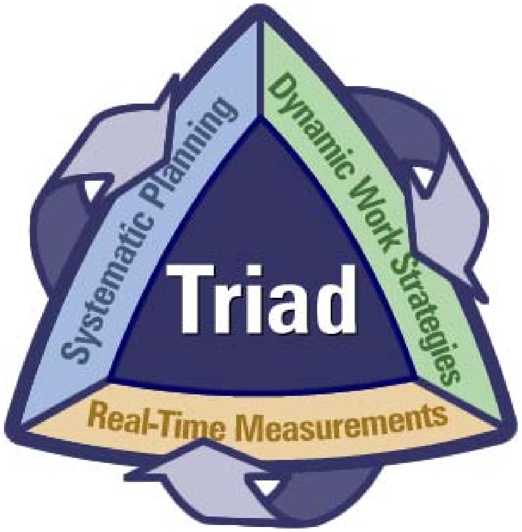
Stephen Dyment

U.S. EPA

Technology Innovation Field Services Division
dyment.stephen@epa.gov

Technical Session Objectives

- Provide an overview of the demonstration of method applicability (DMA) process used in a Triad Approach
  - Highlight activities often conducted during evaluations of field portable x-ray fluorescence (XRF) instruments
- Translate common DMA findings into a comprehensive quality control (QC) program for field activities involving XRF analysis of soil and sediment matrices
  - Indicate QC sample types, function, strategies for analysis, and effective use of results in real time
- Showcase project benefits of real time analysis and collaborative data sets



DMA History

- Concept founded in SW-846, performance based measurement (PBMS) initiative
  - http://www.epa.gov/sw-846/pbms.htm
- 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

Why Do I Need a DMA?

- ◆ Triad usually involves real-time measurements to drive dynamic work strategies
- ◆ Greatest sources of uncertainty are usually sample heterogeneity and spatial variability
- ◆ Relationships with established laboratory methods often required – educate stakeholders
- ◆ Early identification of potential issues
  - » Develop strategies to manage uncertainties
- ◆ Provides an initial look at CSM assumptions

What’s Involved?

- There is no template for DMAs!
  - Format, timing, documentation, etc. depend heavily on site specifics, existing information, and intended data use
- Perform early in program
- Go beyond simple technology evaluation to optimize full scale implementation
  - Method comparison, statistical analysis
  - Sample design, field based action levels
  - Sample prep, throughput, other logistics
  - Data management issues

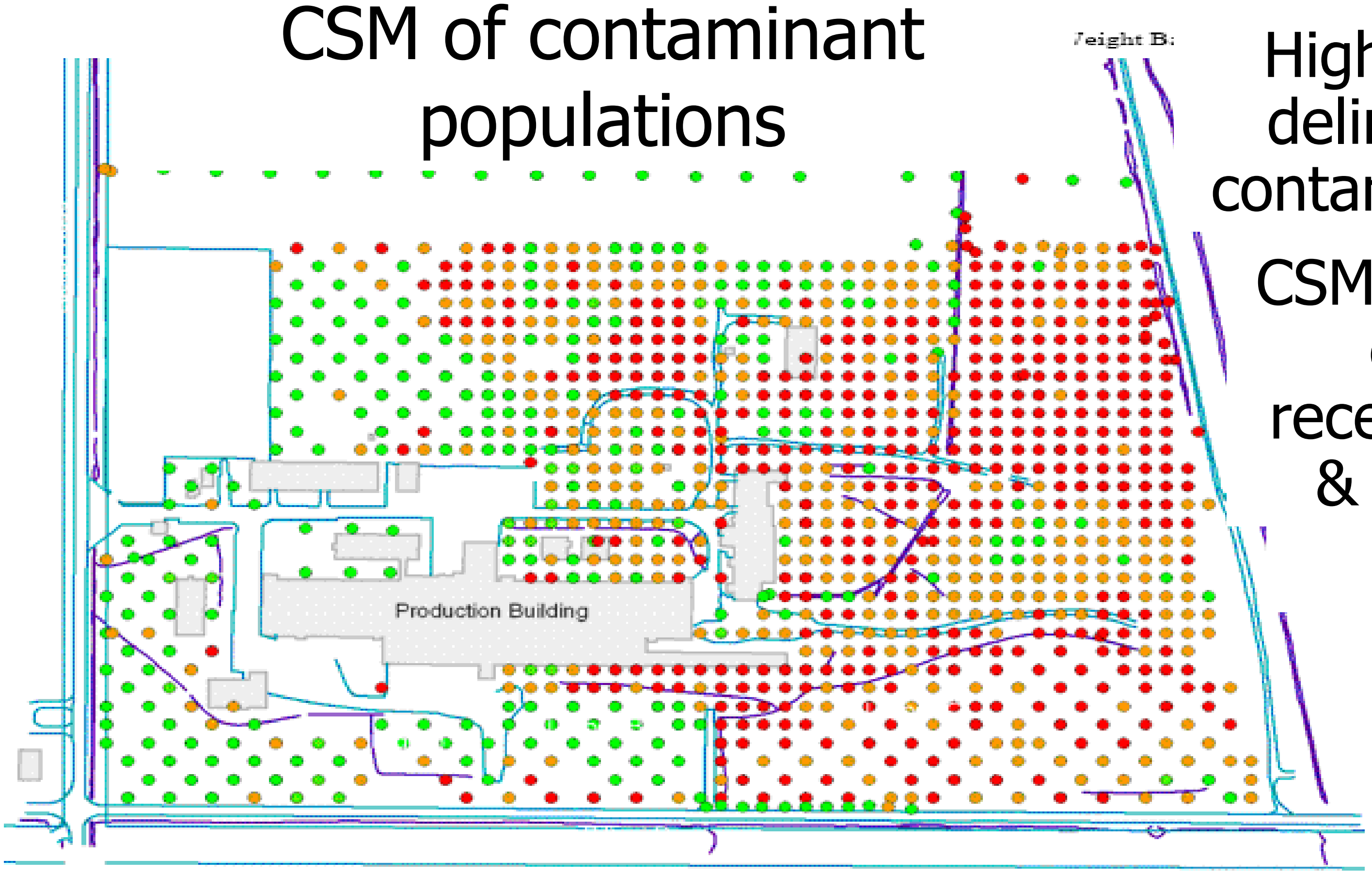
What to Look For

- 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

More Benefits

- Augment planned data collection and CSM development activities
- Test drive decision support tools
  - Sampling and statistical tools
  - Visualization tools
- Develop relationships between visual observations and direct sensing tools
- Flexibility to change tactics based on DMA rather than full implementation
- Establish decision logic for dynamic work strategies
- Evaluate existing contract mechanisms
- Optimize sequencing, load balance, unitizing costs

Typical DMA Products – Summary Statistics



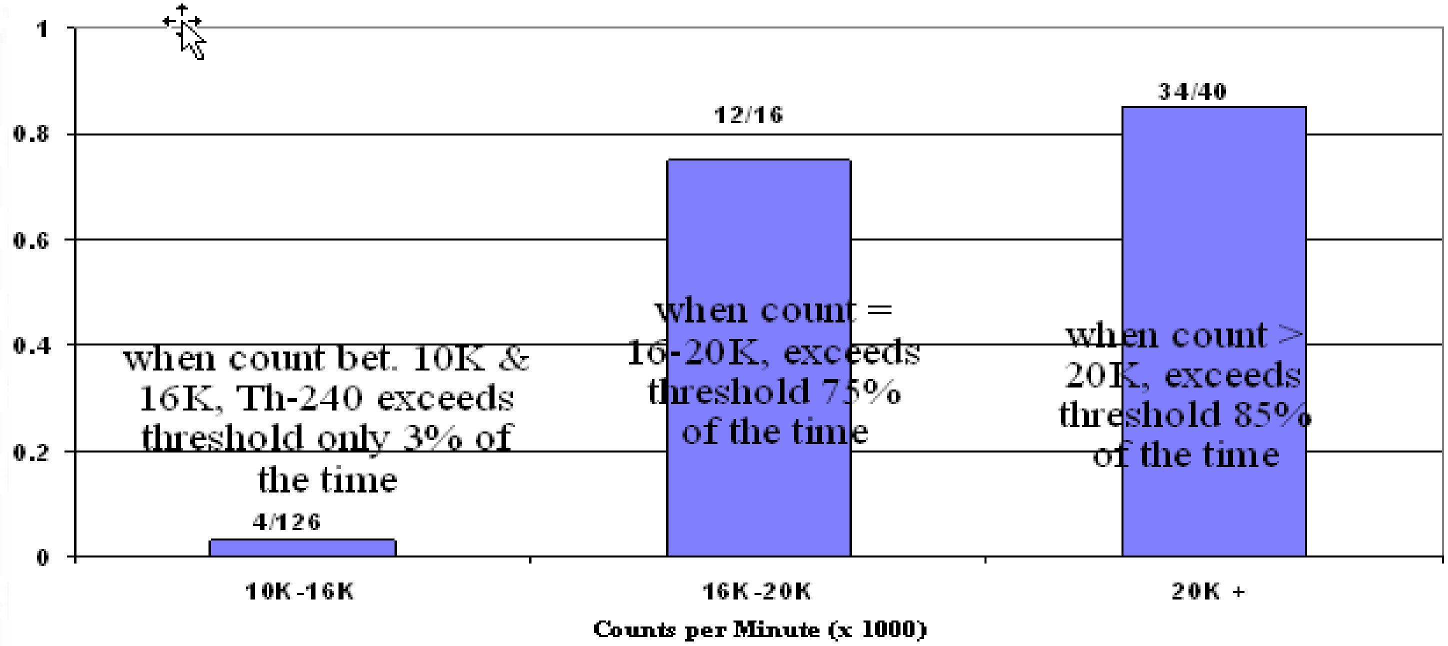
CSM of contaminant populations

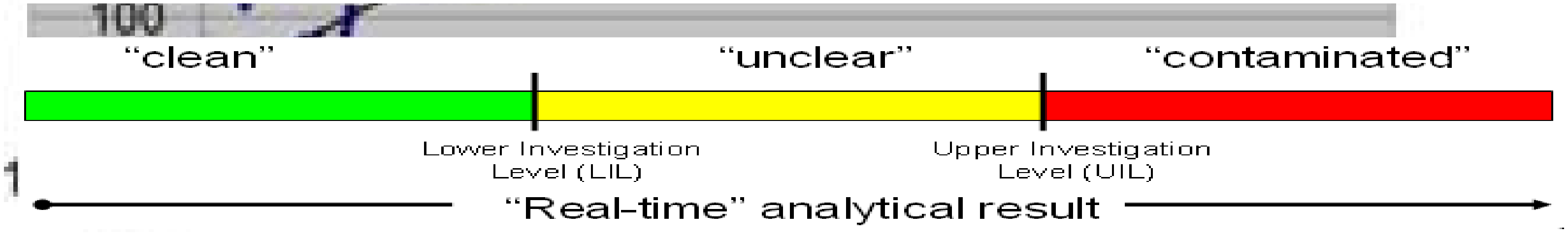
High density data delineates spatial contaminant patterns

CSM guides informed evaluation of receptors, pathways & exposure units

Typical DMA Products – Statistical Evaluations/Method Comparisons

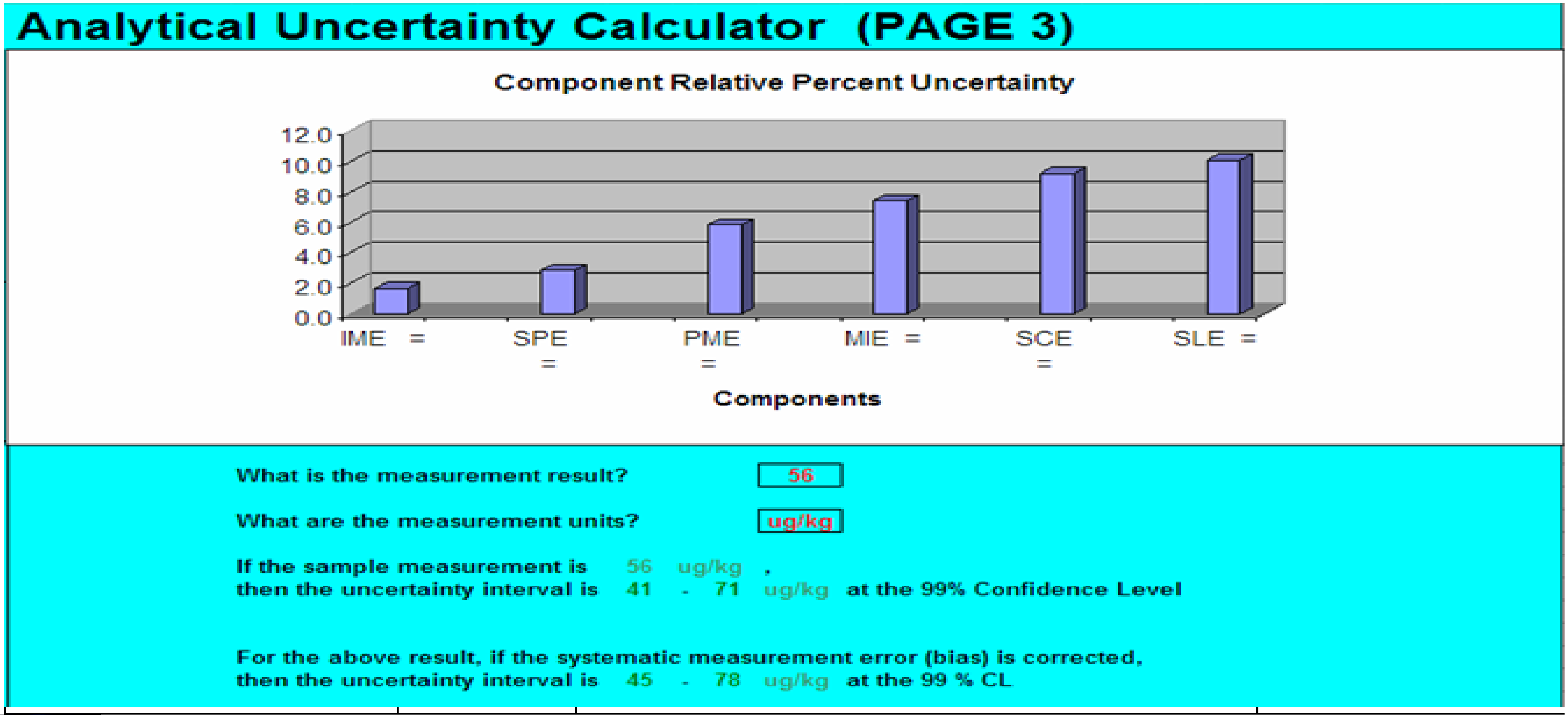
- ◆ Parametric - linear regressions
- ◆ Non-parametric - ranges or bins





Typical DMA Products – Uncertainty Evaluations

- Example: Ingersoll Uncertainty Calculator



Analytical Uncertainty Calculator (PAGE 3)

What is the measurement result? 56

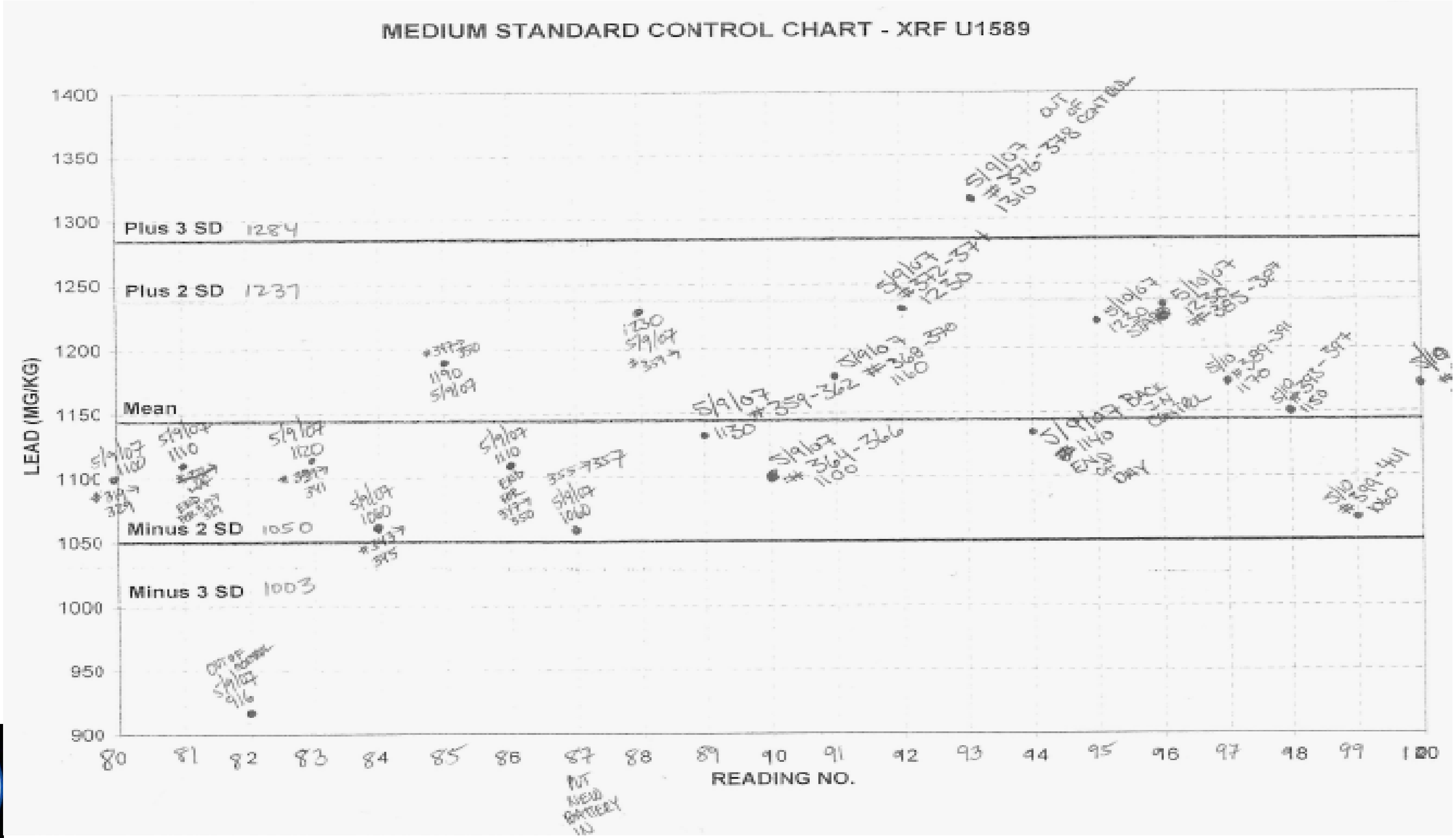
What are the measurement units? ug/kg

If the sample measurement is 56 ug/kg ,
then the uncertainty interval is 41 - 71 ug/kg at the 99% Confidence Level

For the above result, if the systematic measurement error (bias) is corrected,
then the uncertainty interval is 45 - 78 ug/kg at the 99 % CL

Typical DMA Products

- QC program worksheets



The Specifics of X-ray Fluorescence

- XRF-basics and principles of operation
- Translating DMA results
- Developing a QC program
    - QC sample types
    - QC sample function, corrective action
- Developing a dynamic sampling protocol
    - Choosing collaborative samples

Principle of XRF Operation

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

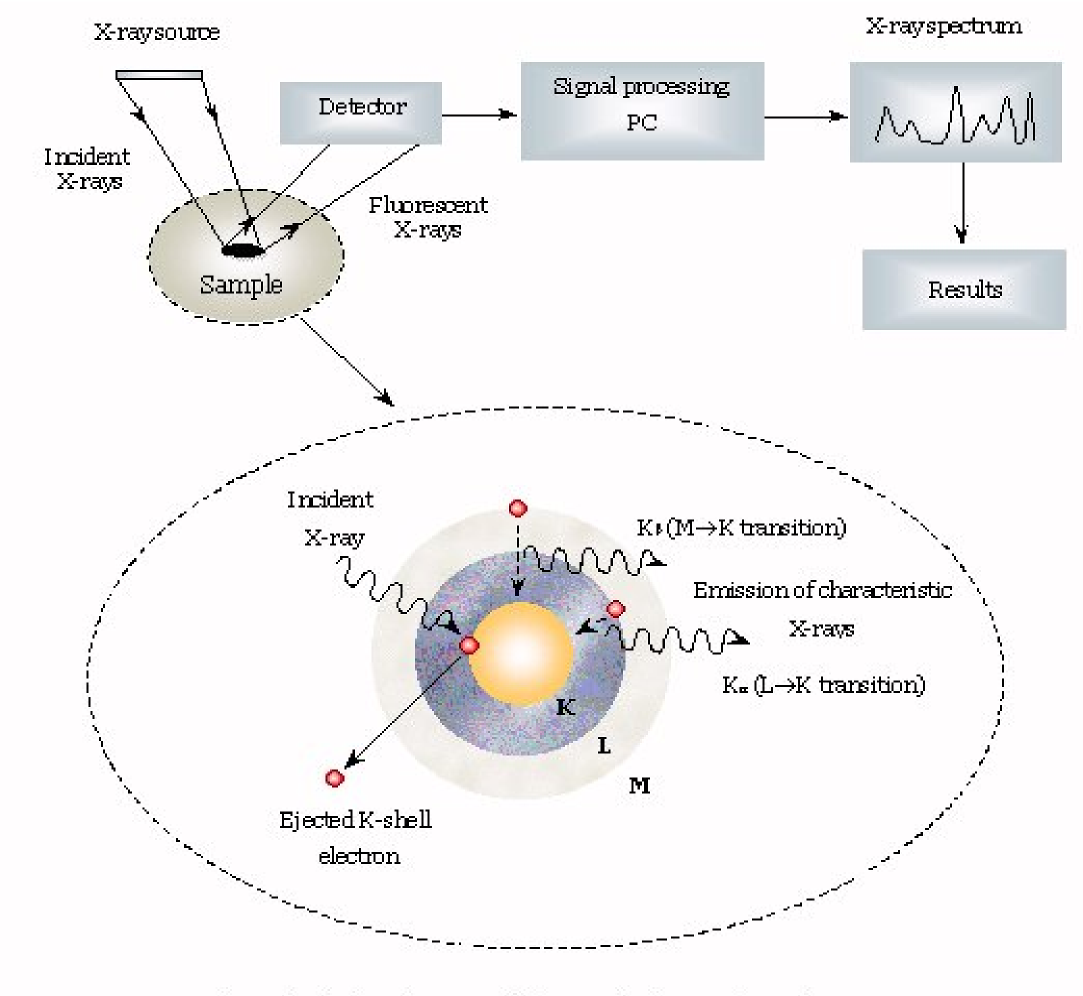
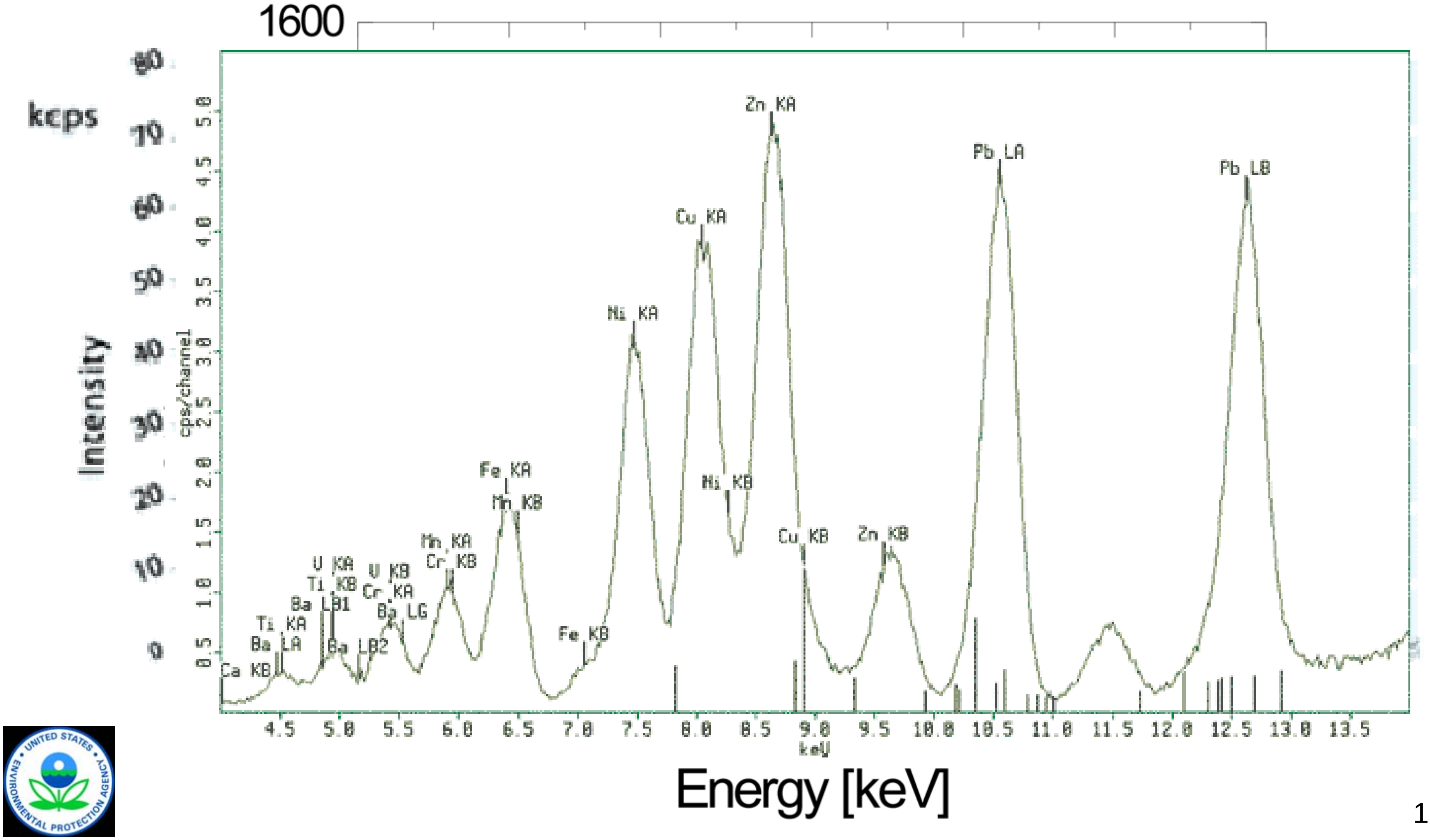


Figure 1: *The principle of XRF and the typical XRF detection arrangement*

Source: http://omega.physics.uoi.gr/xrf/english/images/PRINCIP.jpg

Some Example XRF Spectra…

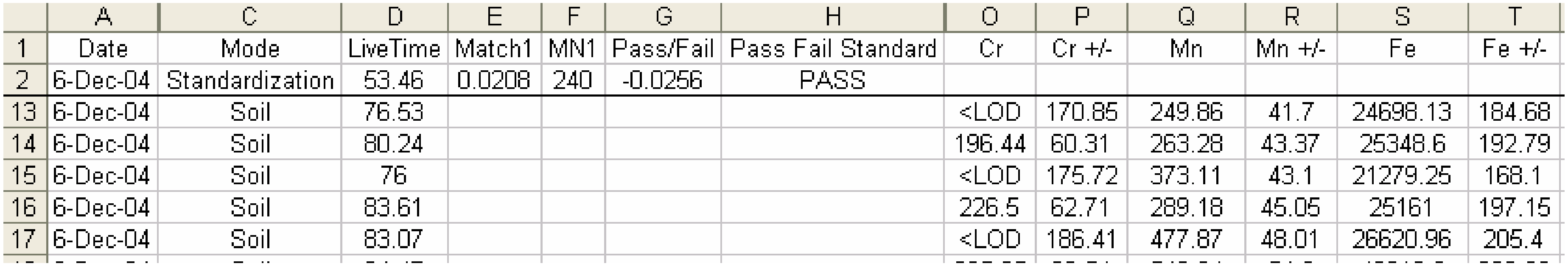


How is a Field Portable XRF Typically Used?

- Measurements on prepared samples
- Measurements through bagged samples (limited preparation)
- In situ measurements of exposed surfaces

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

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

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)**

How Is An XRF Calibrated?

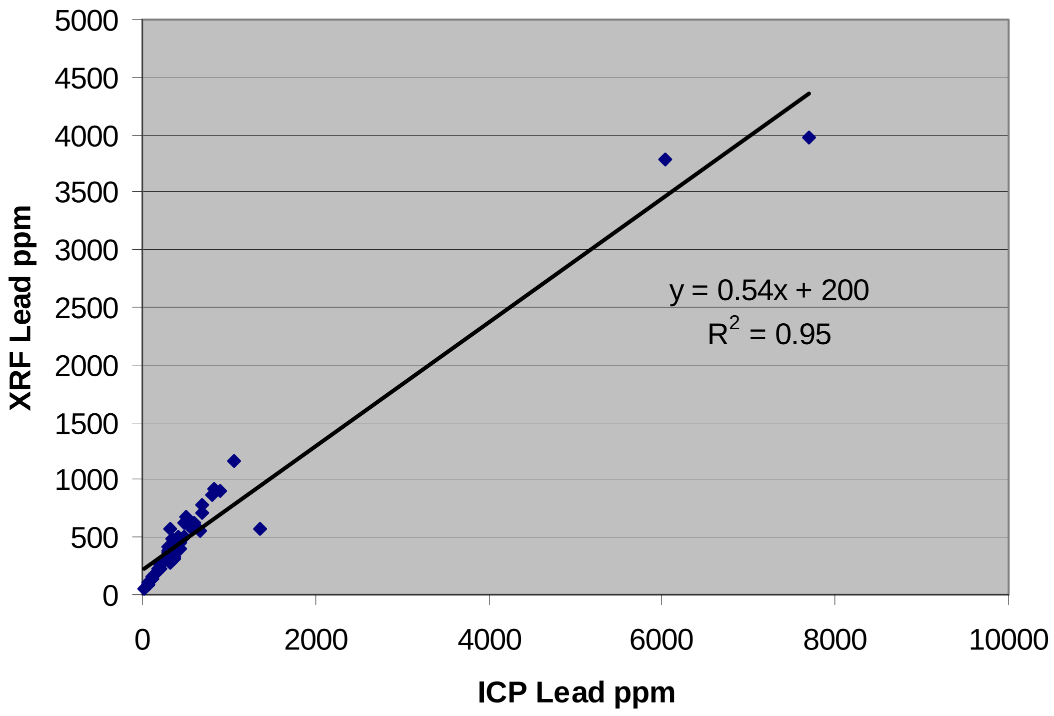
- ***Fundamental Parameters Calibration*** – calibration based on known detector response properties, “standardless” calibration
- ***Empirical Calibration*** – calibration calculated using regression analysis and known standards, either site-specific media with known concentrations or prepared, spike standards
- ***Compton Normalization*** – incorporates elements of both empirical and FP calibration. A single, well-characterized standard, such as an SRM or a SSCS, is analyzed, and the data are normalized for the Compton peak

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

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

Figure 1: ICP vs XRF (lead - all data)



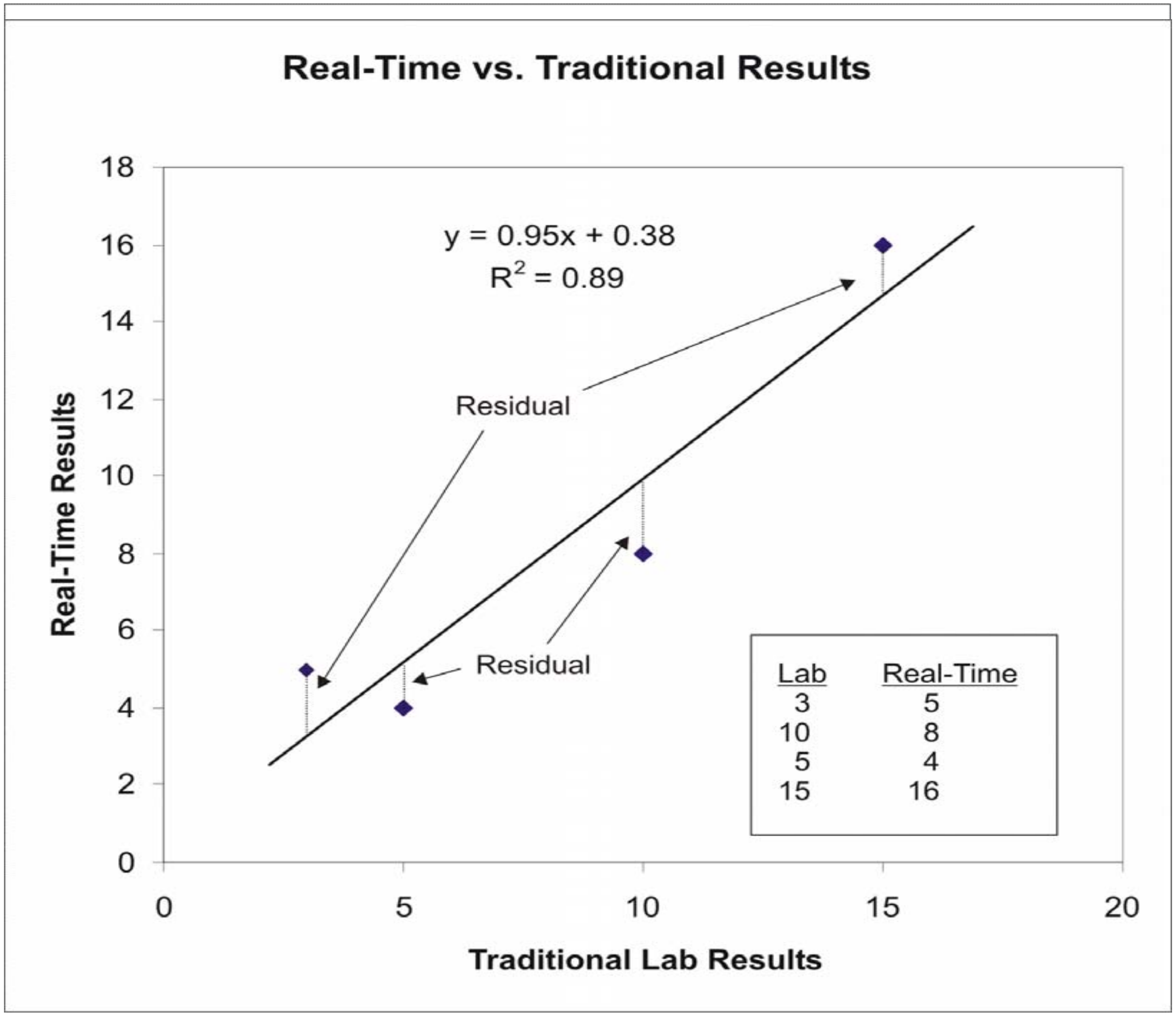
How is XRF Performance Commonly Defined?

- **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 concentration can the instrument reliably identify the presence of an element?
  - » 3 standard deviation rule
  - » Rule of thumb: 4X increase in count time = 1/2 reduction in DL
- **Quantitation Limits** – at what concentration 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?

Translating XRF DMA Results

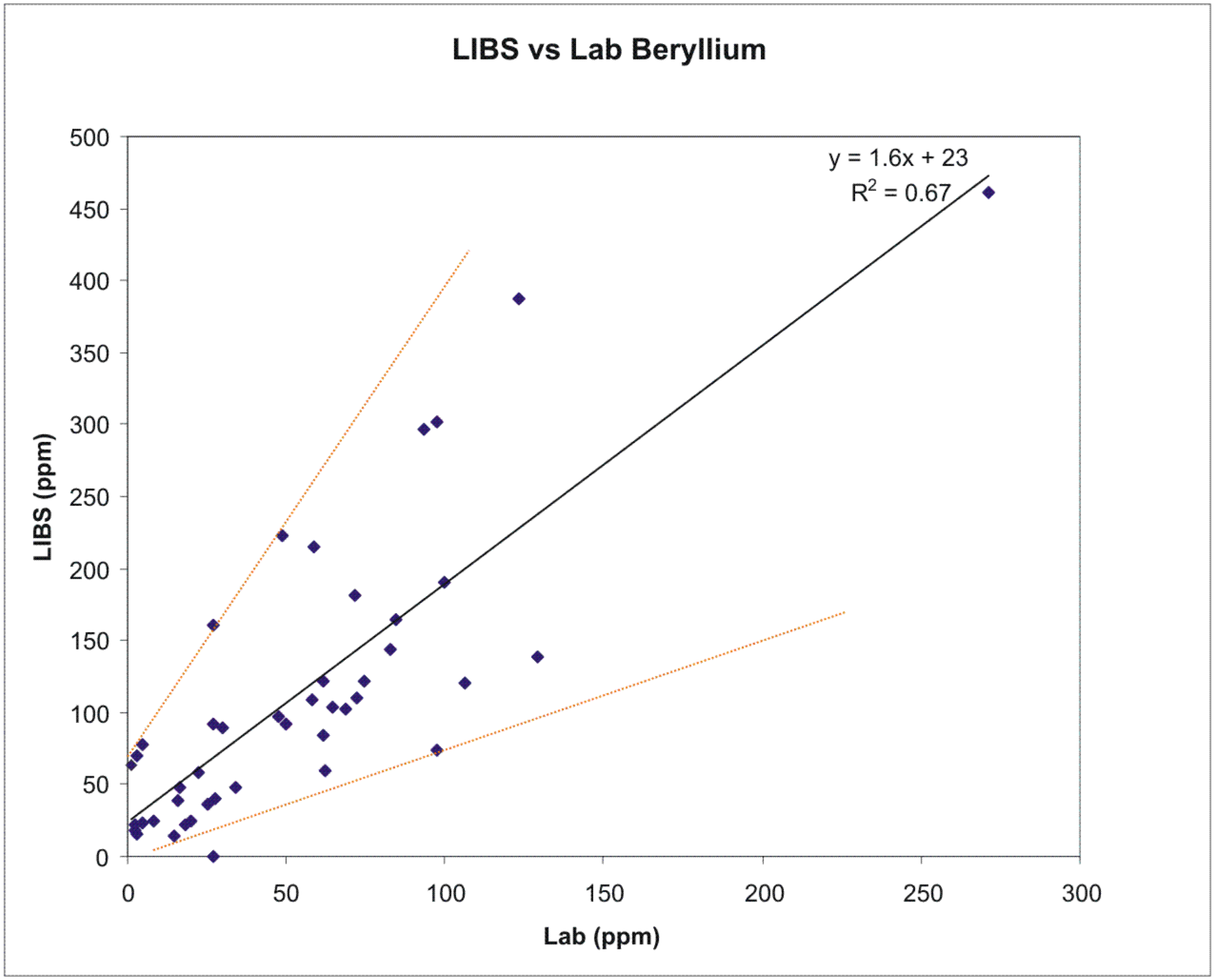
- Comparability - usually with ICP or AA methods
  - Regression analysis is the ruler most commonly used to measure comparability
  - Standard laboratory data can be “noisy” and are not necessarily an error-free representation of reality
  - SW-846 Method 6200: “If the r2 is 0.9 or greater…the data could potentially meet definitive level data criteria.”
- Focus should be on decision comparability, not laboratory result comparability
  - Parametric and non-parametric techniques available

What is a Regression Line?



Lab	Real-Time
3	5
10	8
5	4
15	16

Heteroscedasticity is a Fact of Life for Environmental Data Sets

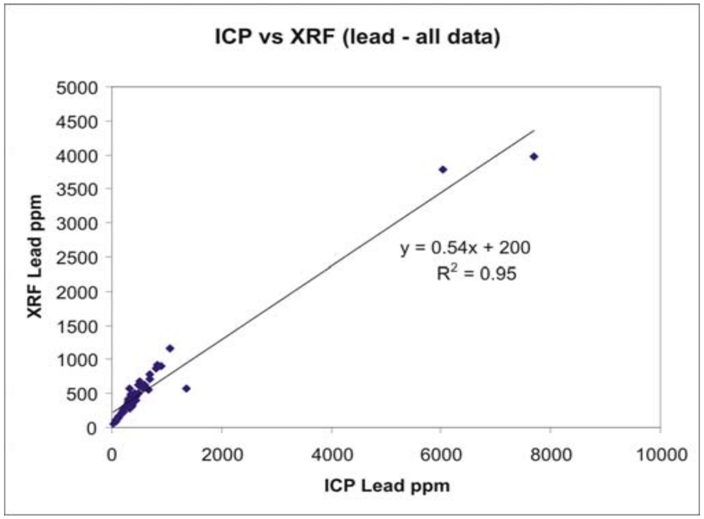


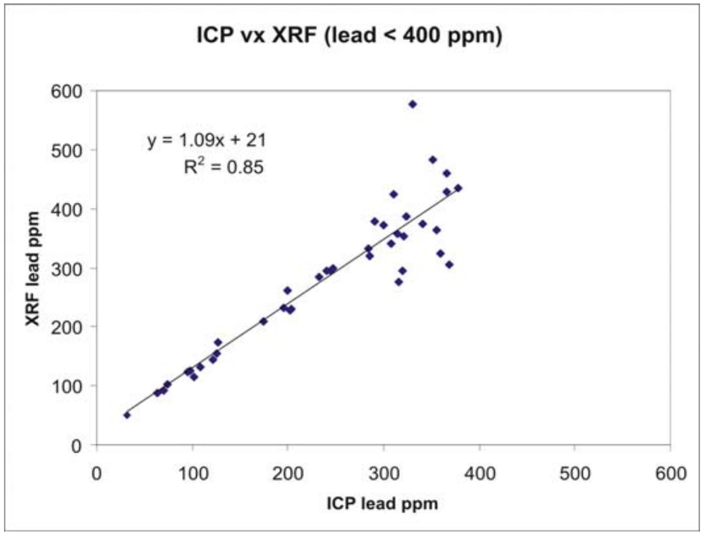
Appropriate Regression Analysis

- 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
- No evidence of inexplicable “outliers”
- No signs of correlated residuals
- High R2 values (close to 1)
- Constant residual variance (homoscedastic) is nice but unrealistic

Example: XRF and Lead

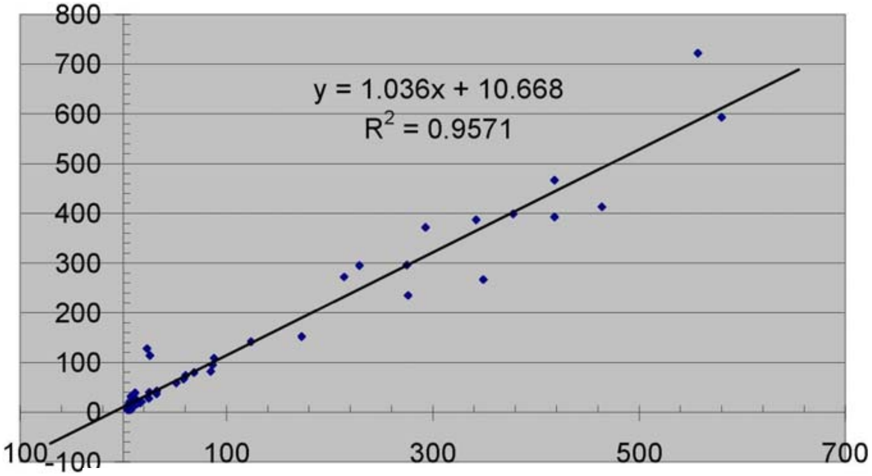
- Full data set:
  - » Wonderful R^2
  - » Unbalanced data
  - » Correlated residuals
  - » Apparently poor calibration
- Trimmed data set:
  - » Balanced data
  - » Correlation gone from residuals
  - » Excellent calibration
  - » R^2 drops significantly

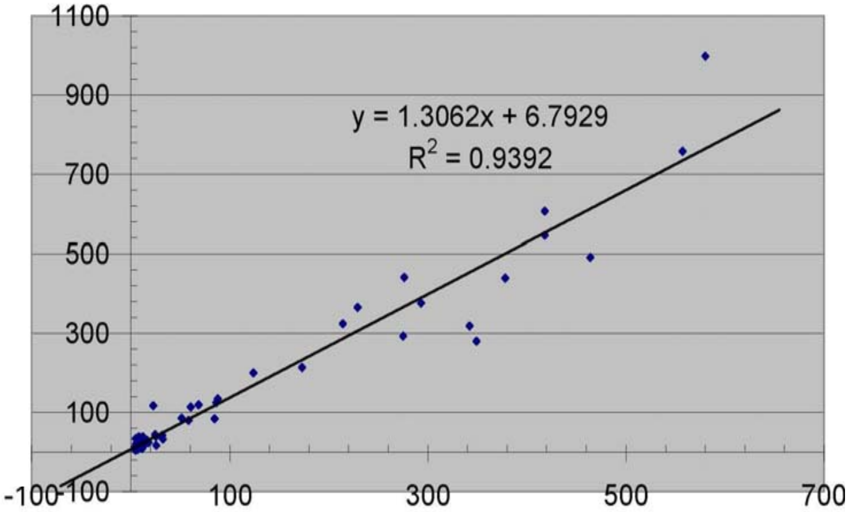


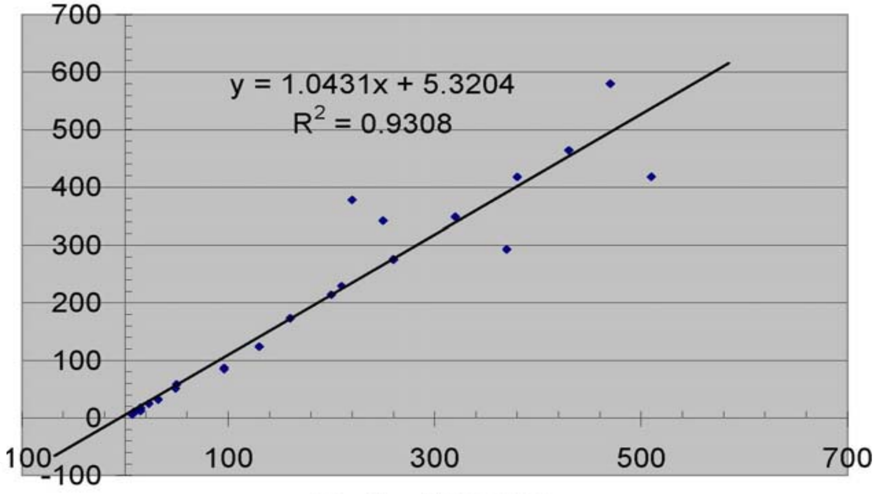


Cautionary Tale

Small scale variability can impact data quality more than the analytical method







A Properly Designed QC Program Will Help You Manage…

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

XRF Quality Control Procedures

- Initial warm-up (30 minutes)
  - Energy calibration/standardization checks
- Blanks - silica or sand
- Calibration checks - initial and continuing
- Detection limit evaluation and monitoring
- Duplicates - instrument, sample replicates
- Monitor for inference effects, trends
- Matrix effects - variability, moisture
- Watch sample or decision unit variability
- Watch dynamic range
- Decision error rates

Basic XRF QC Requirements: Initial Calibration Check

- Energy calibration/standardization checks
- Calibration checks using NIST-traceable standard reference material (SRM), preferably in media similar to what is expected at the site
- Calibration checks using blank silica/sand
- Calibration checks using matrix spikes
- Calibration checks using well-characterized site samples

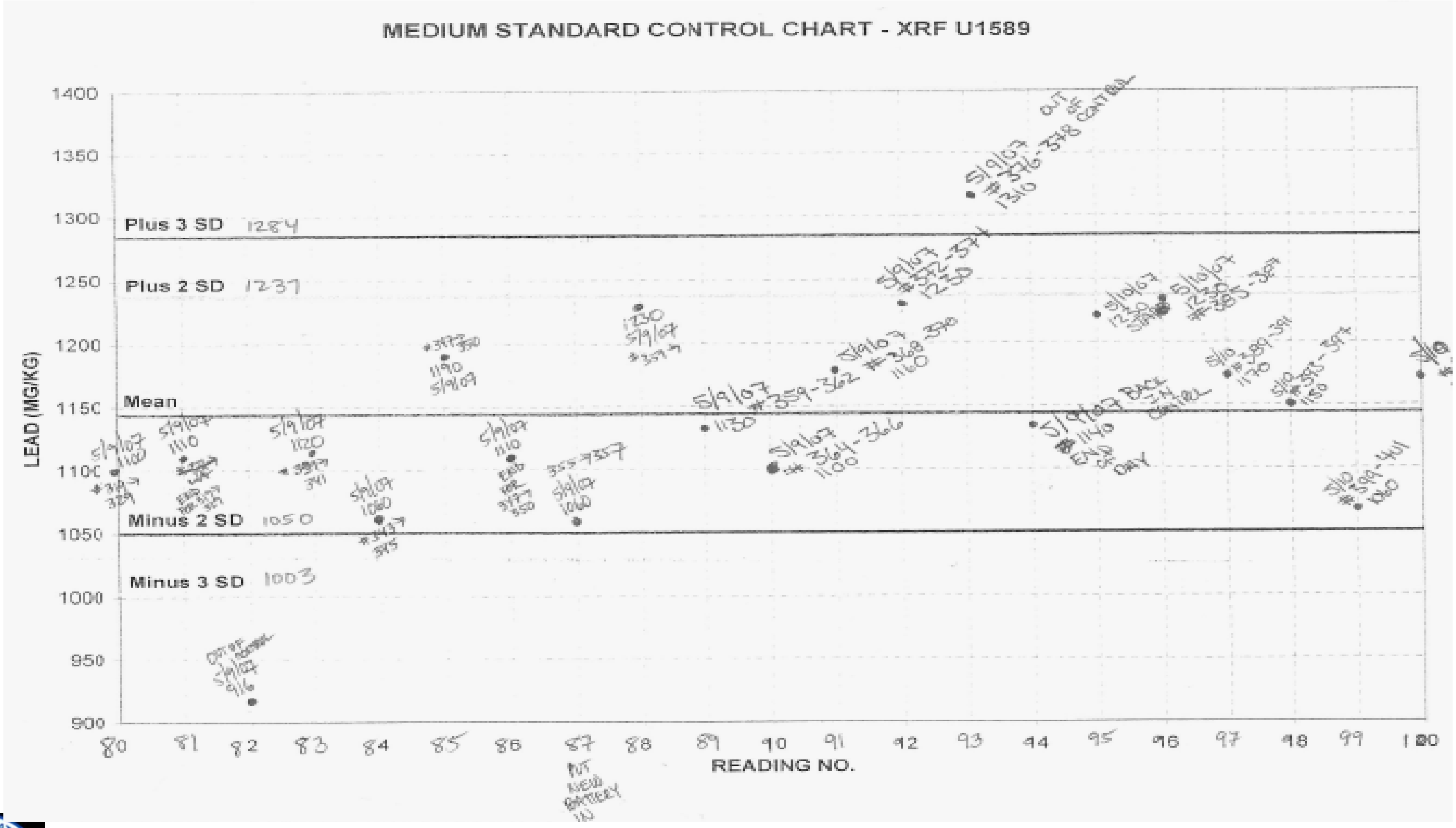
Initial Calibration Check Example

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

Basic XRF QC Requirements: Continuing Calibration

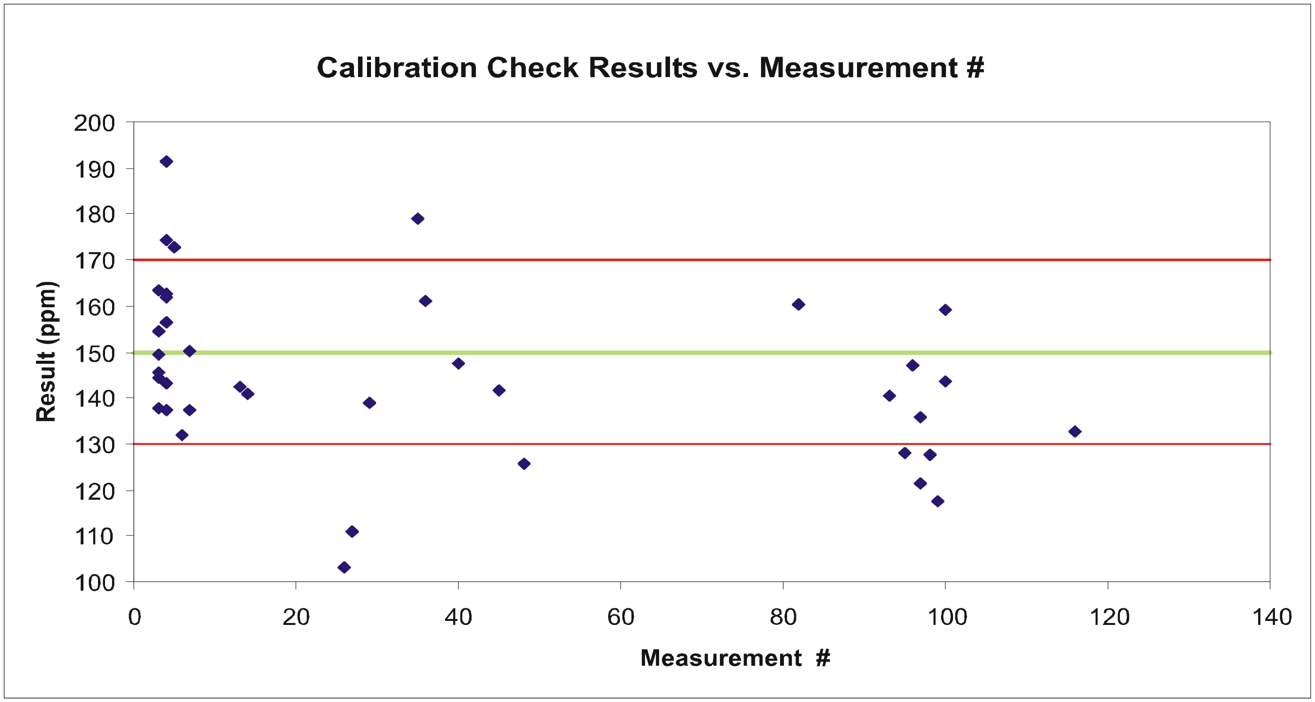
- Standardization checks: follow manufacturer recommendations (typically several times a day)
- On-going calibration checks: at least twice a day (start and end), a higher frequency is recommended
- Make sure XRF performance in relation to SRMs is well understood initially - watch for trends that indicate problems
- Typically controls set up based on initial calibration check work (i.e., a two standard deviation rule)
- Frequency of checks is a balance between sample throughput and ease of sample collection or repeating analysis

Control Charts – A Simple Continuing Calibration Check



Example of What to Watch for…

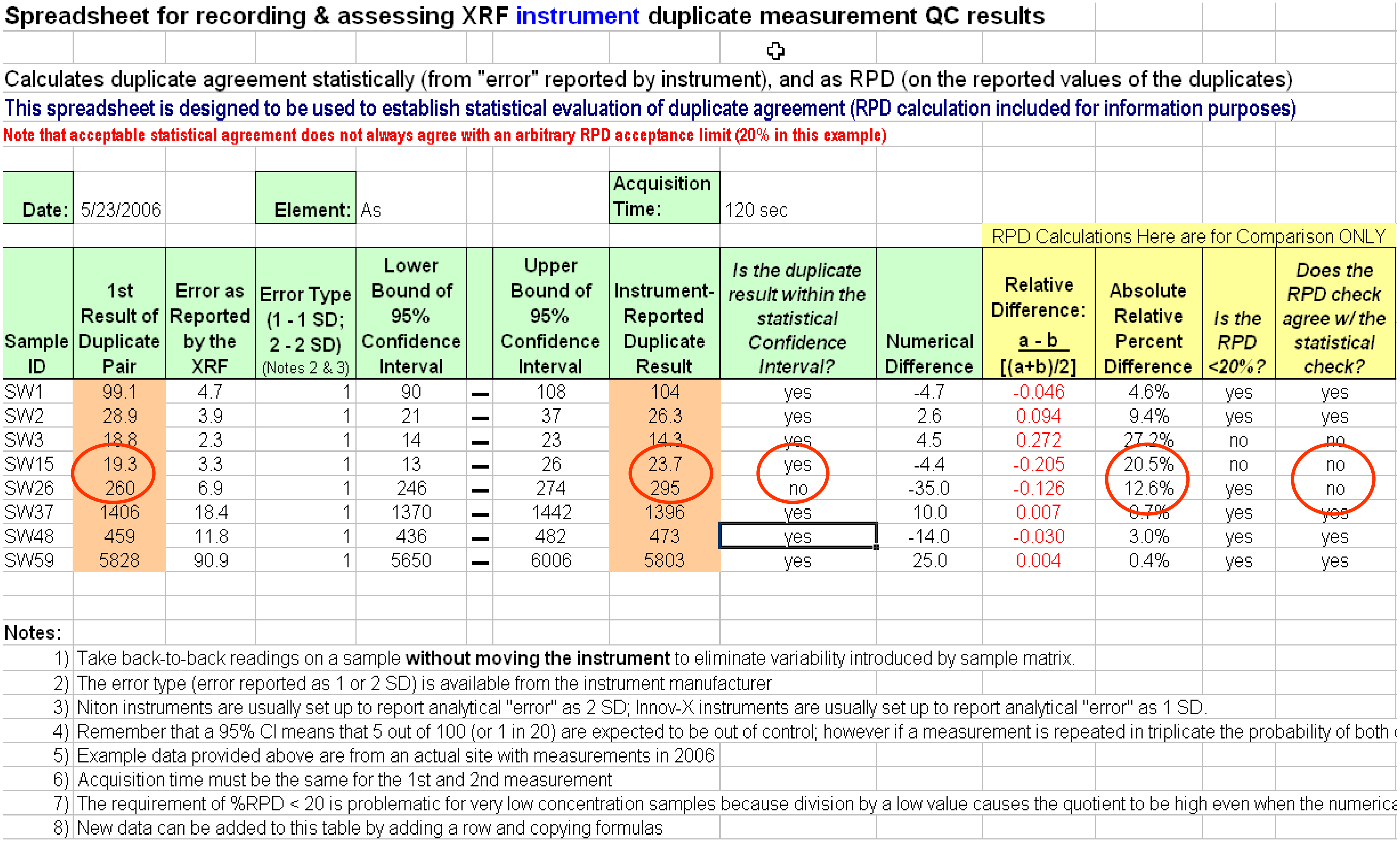
- Two checks done each day, start and finish
- 150 ppm standard with approximately +/- 9 ppm for 120 second measurement
- Observed standard deviation in calibration check data: 18 ppm
- Average of initial check: 153 ppm
- Average of ending check: 138 ppm



Monitoring Detection Limits One Example

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

Duplicates and Replicates



Spreadsheet for recording & assessing XRF instrument duplicate measurement QC results

Calculates duplicate agreement statistically (from "error" reported by instrument), and as RPD (on the reported values of the duplicates)

This spreadsheet is designed to be used to establish statistical evaluation of duplicate agreement (RPD calculation included for information purposes)

Note that acceptable statistical agreement does not always agree with an arbitrary RPD acceptance limit (20% in this example)

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

Notes:

1) Take back-to-back readings on a sample **without moving the instrument** to eliminate variability introduced by sample matrix.
2) The error type (error reported as 1 or 2 SD) is available from the instrument manufacturer
3) Niton instruments are usually set up to report analytical "error" as 2 SD; Innov-X instruments are usually set up to report analytical "error" as 1 SD.
4) Remember that a 95% CI means that 5 out of 100 (or 1 in 20) are expected to be out of control; however if a measurement is repeated in triplicate the probability of both
5) Example data provided above are from an actual site with measurements in 2006
6) Acquisition time must be the same for the 1st and 2nd measurement
7) The requirement of %RPD < 20 is problematic for very low concentration samples because division by a low value causes the quotient to be high even when the numerica
8) New data can be added to this table by adding a row and copying formulas

Interference Effects

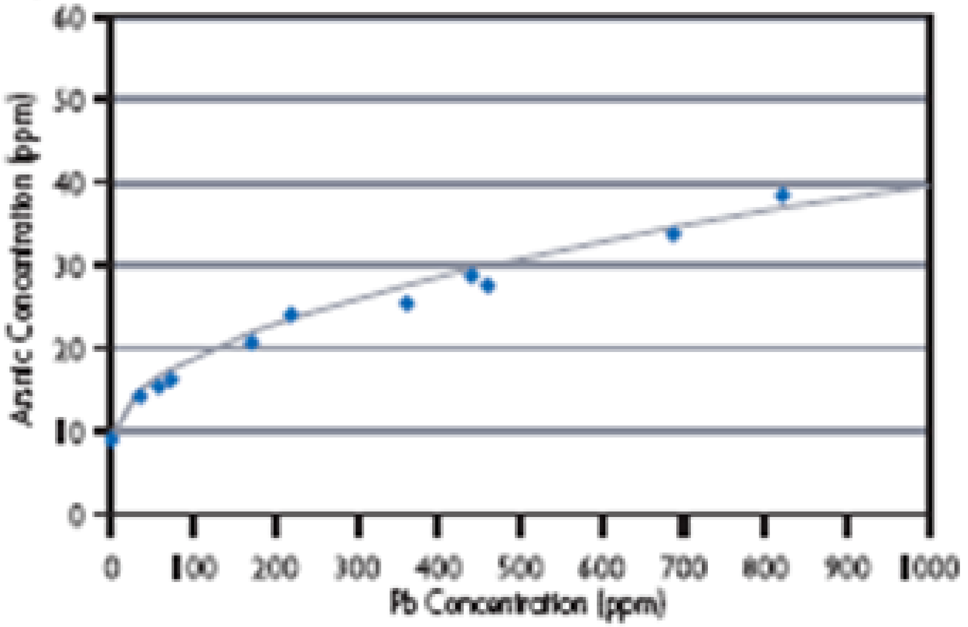
- Spectra too close for detector to accurately resolve
  - As Kα=10.55 KeV
  - Pb Lα=10.54 KeV
- Result: biased estimates for one or more quantified elements
- DMA, manufacturer recommendations, scatter plots used to identify conditions when interference effects would be a concern
- “Adaptive QC”…selectively send samples for laboratory analysis when interference effects are a potential issue

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.

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

fig.2 Effect of Lead Concentration on Arsenic Detection Limit



Matrix Heterogeneity Small Scale Variability 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

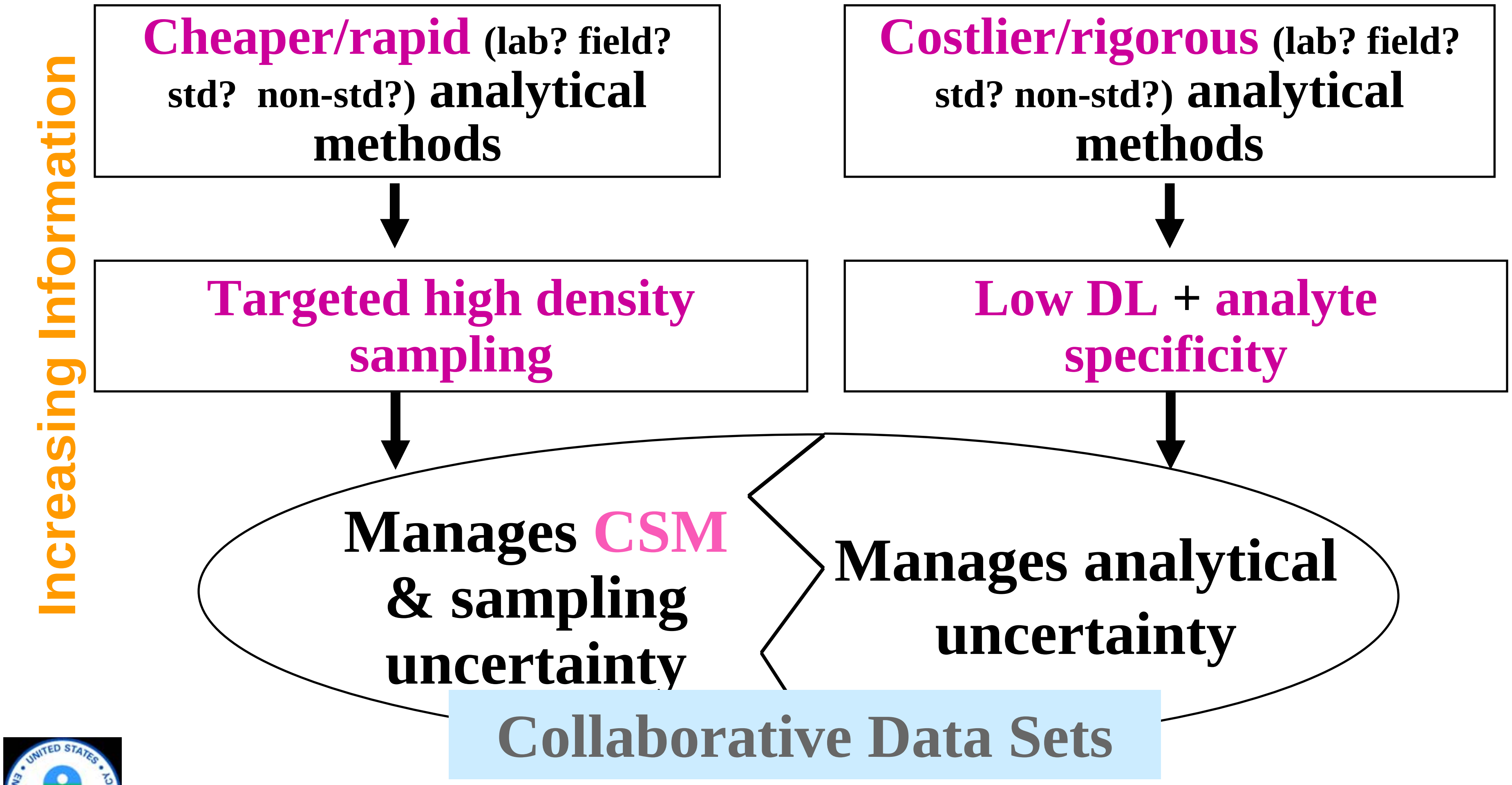
Micro-scale Contaminant – Matrix Relationships Cause Within – Sample Heterogeneity

Small Arms Firing Range Soil Grain Size (Std Sieve Mesh Size)	Pb Conc. 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
Totals	927 (wt-averaged)

Adapted from ITRC (2003)

What particle fraction is “representative”?

Collaborative Data Sets Address Analytical and Sampling Uncertainties



Increasing Information

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

Targeted high density sampling

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

Low DL + analyte specificity

Manages CSM & sampling uncertainty

Manages analytical uncertainty

Collaborative Data Sets

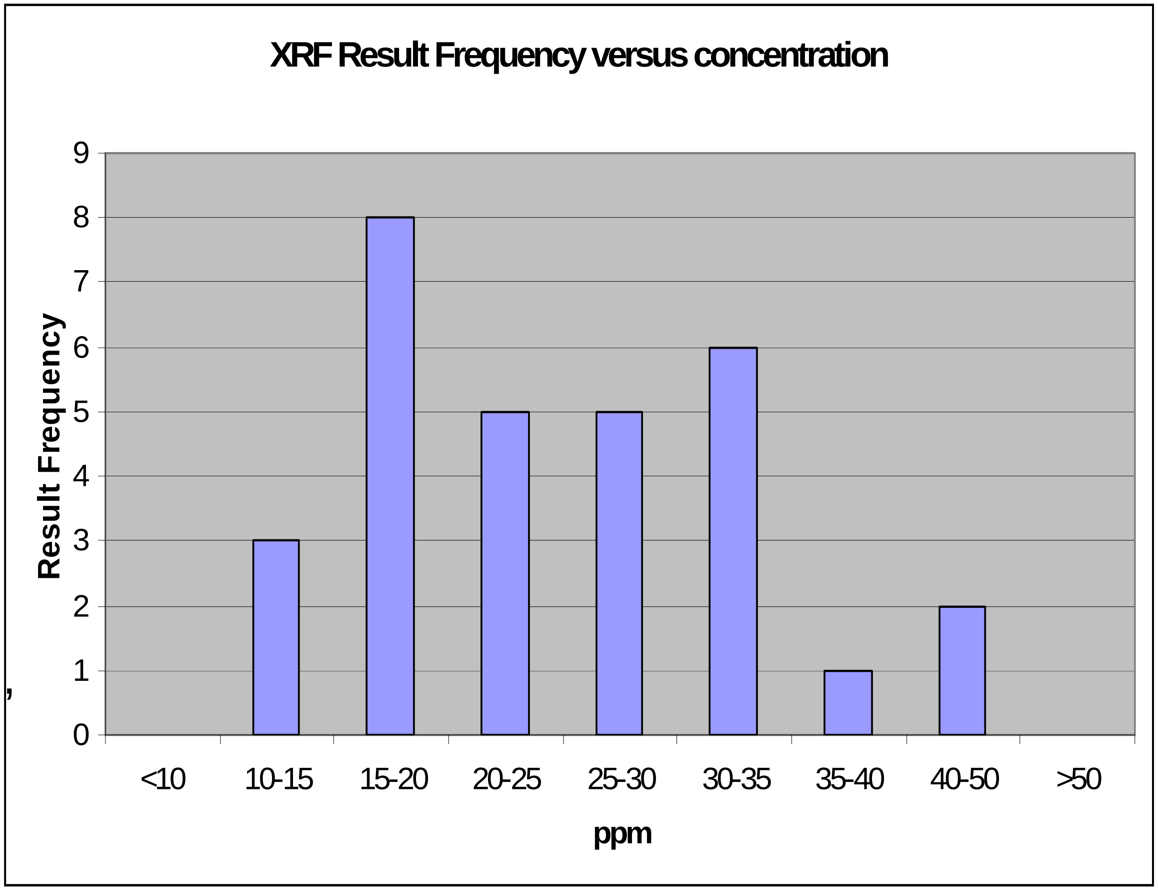
Dynamic Measurement Example

- 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

Example (cont.)

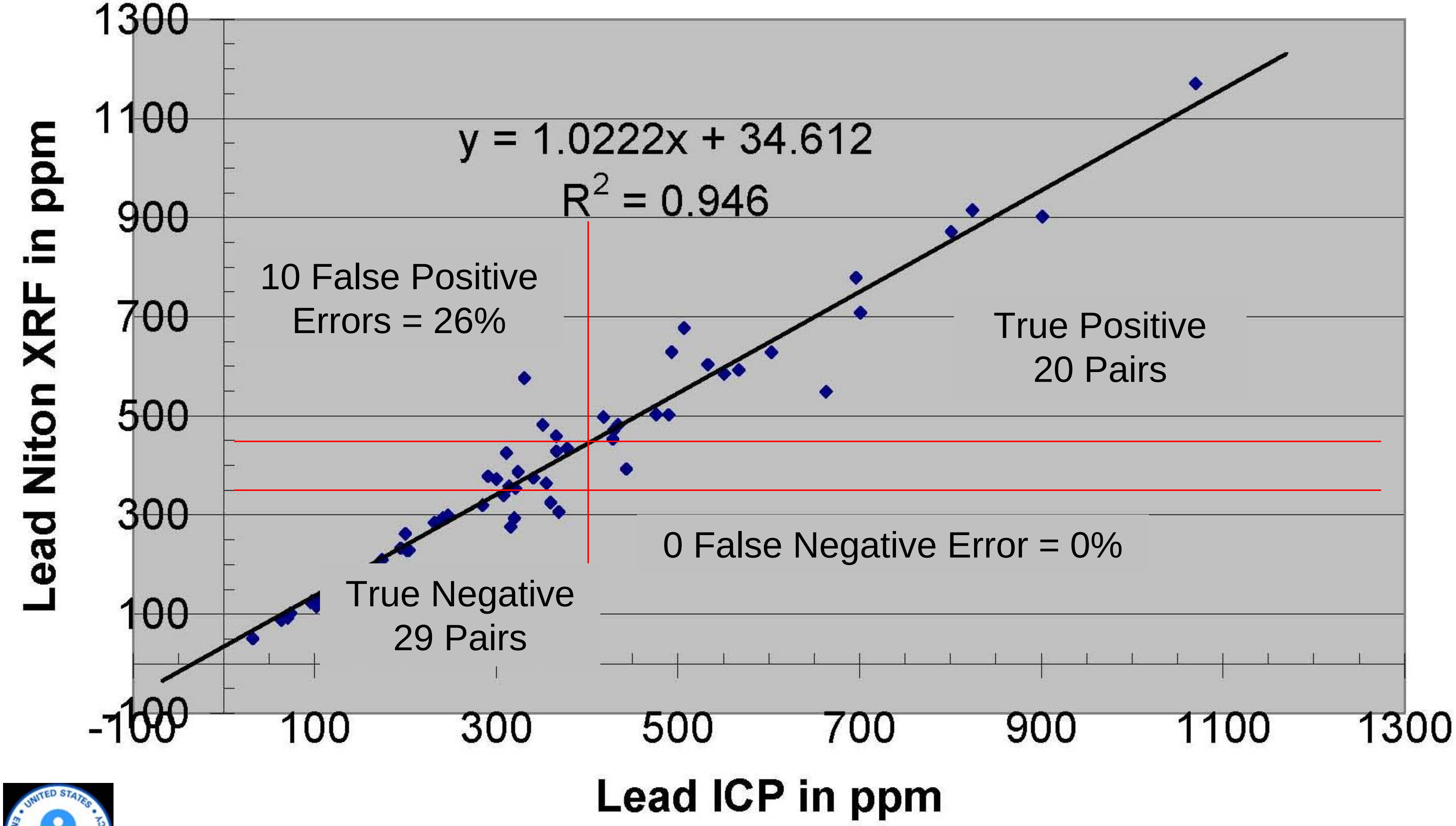
Simple Decision Rule:

- 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



Lead Niton vs. ICP

59 Total pairs



10 False Positive Errors = 26%

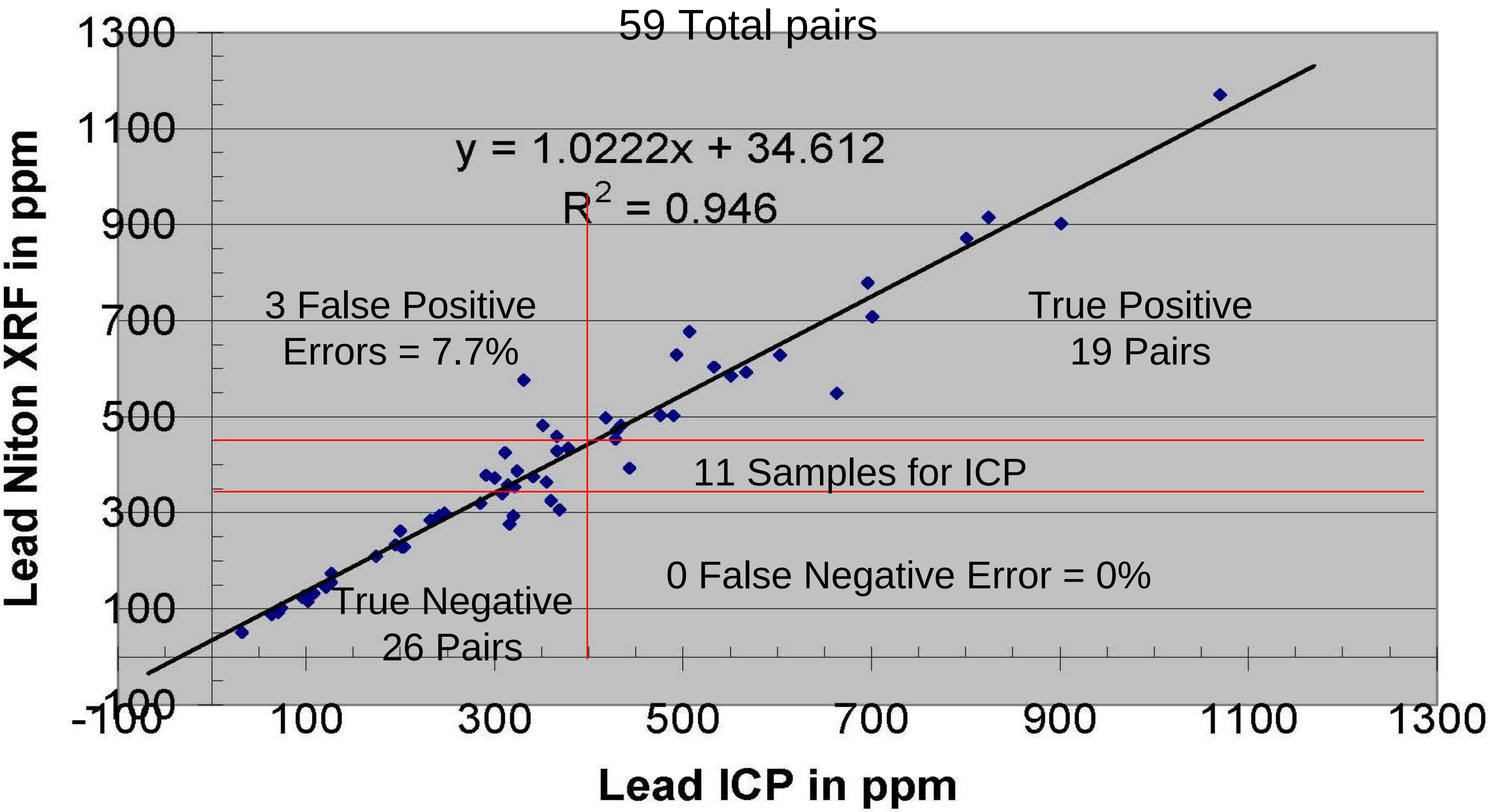
True Positive 20 Pairs

0 False Negative Error = 0%

True Negative 29 Pairs

3 Way Decision Structure with Region of Uncertainty

Lead Niton vs. ICP



59 Total pairs

$y = 1.0222x + 34.612$

$R^2 = 0.946$

3 False Positive Errors = 7.7%

True Positive 19 Pairs

11 Samples for ICP

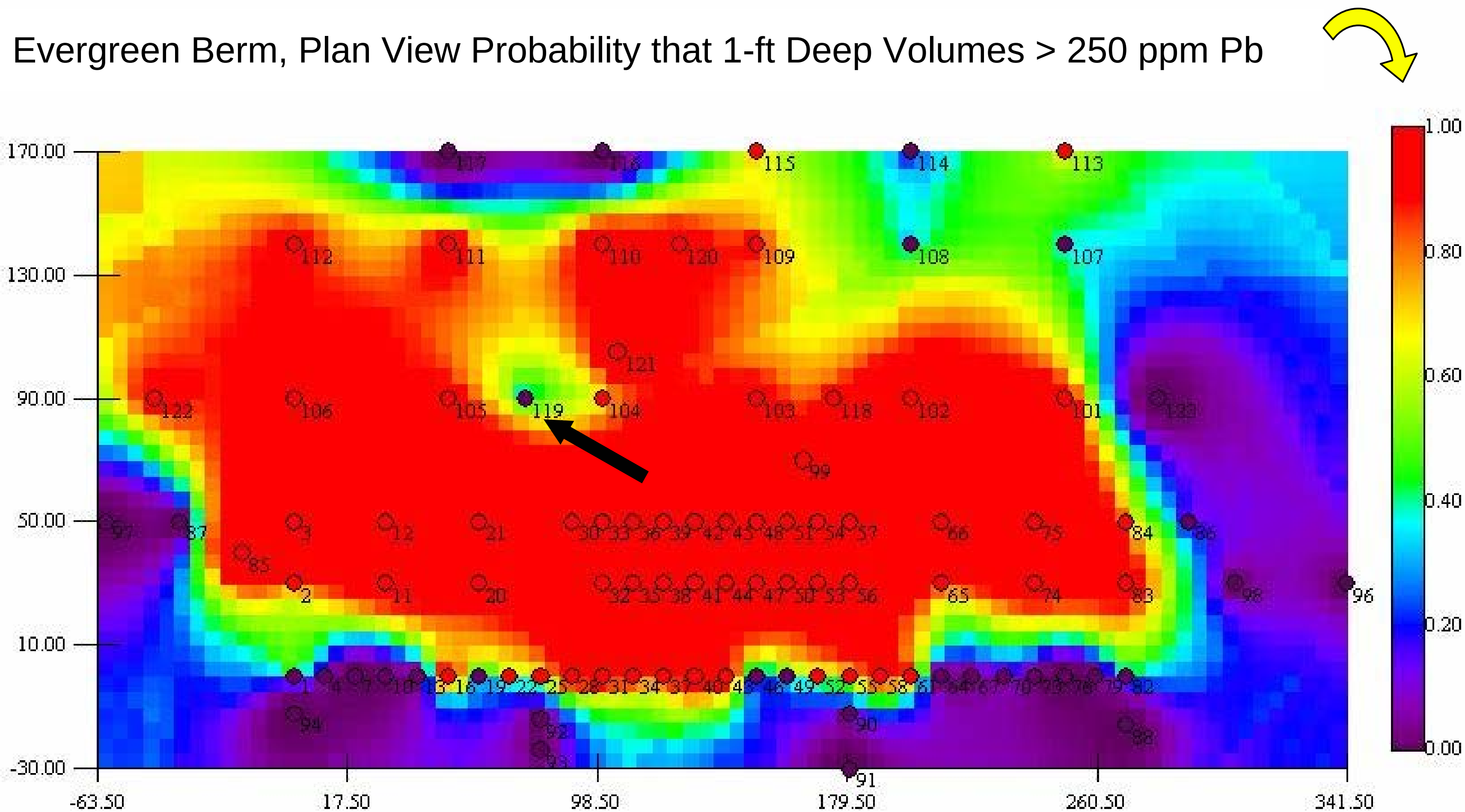
0 False Negative Error = 0%

True Negative 26 Pairs

Communicating Uncertainty in a XRF CSM

Addressing the Unknown

Evergreen Berm, Plan View Probability that 1-ft Deep Volumes > 250 ppm Pb



Note: Sample locations are numbered sequentially in time. See #119 (arrow) as example of adaptive fill-in of uncertain areas to firm up contaminant boundaries

Easting

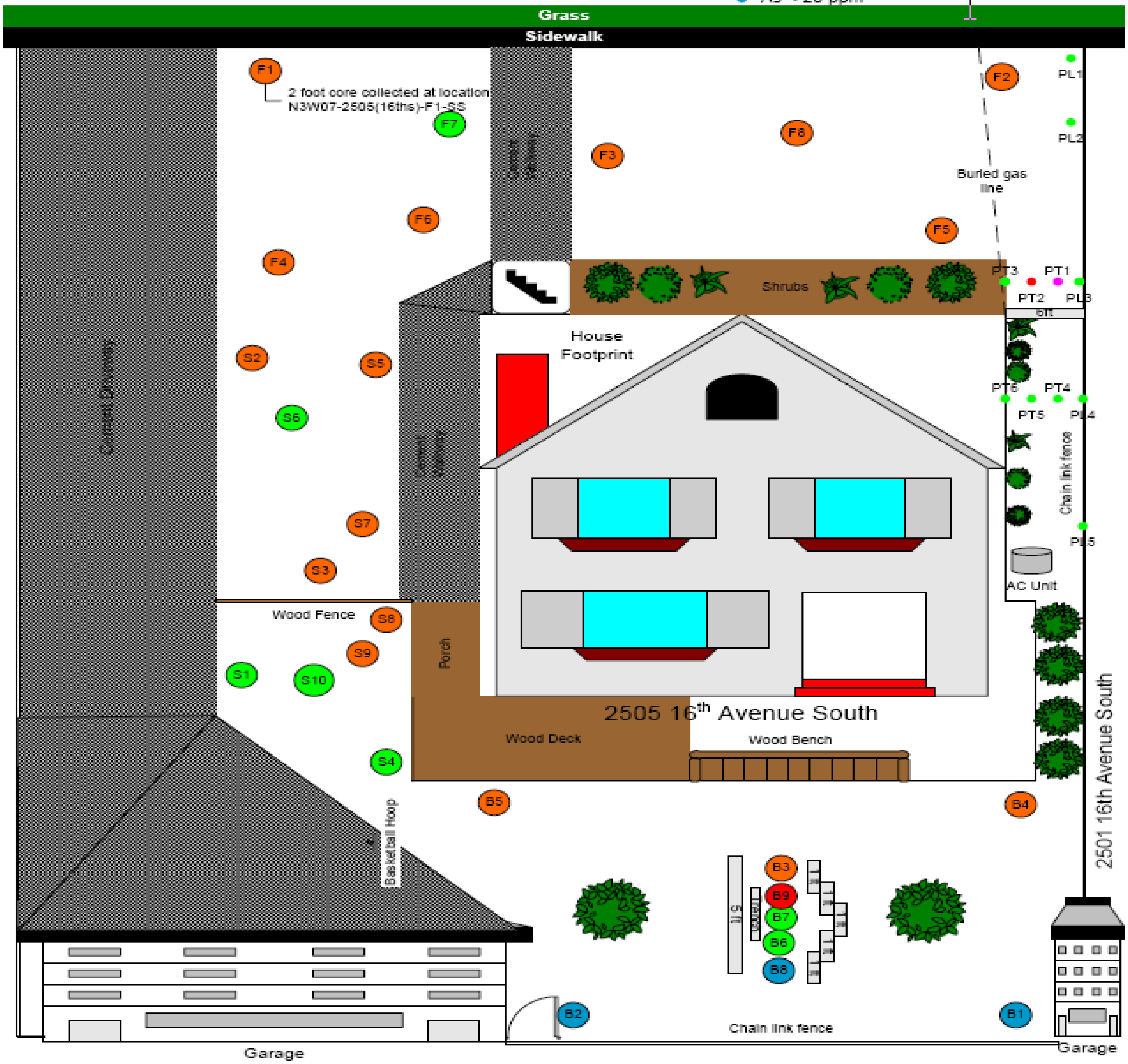
16th Avenue South

Drawing not to scale

- ● As > 300 ppm
- ● As 200-300 ppm
- ● As 95-200 ppm
- ● As 20-95 ppm
- ● As < 20 ppm

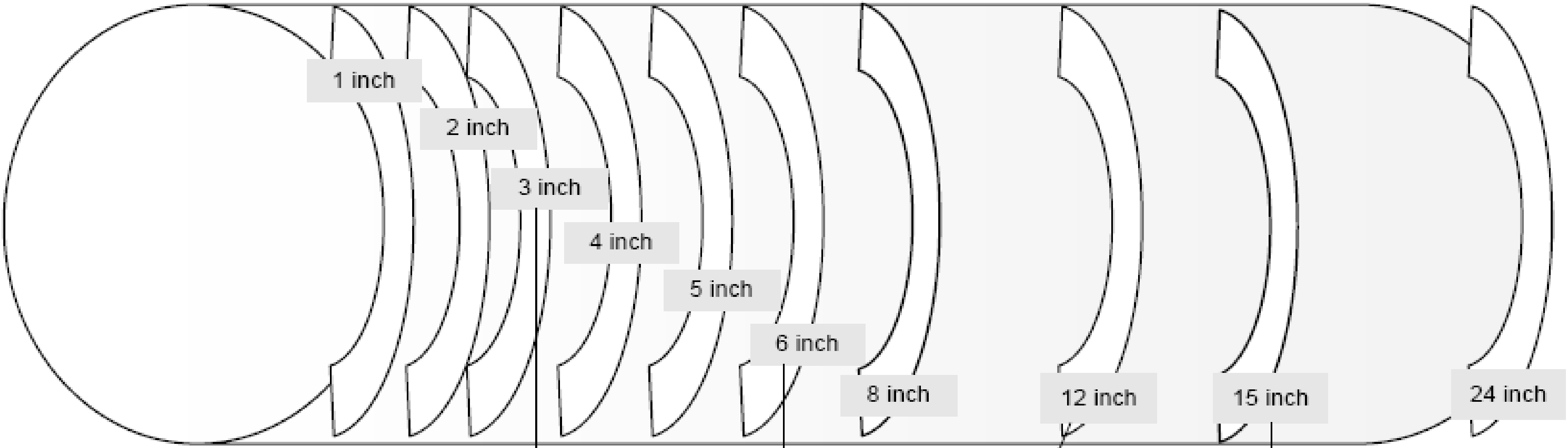
Chain link fence

Buried gas line



Samples Collected from Trench, North Sidewall Extending Between Locations B7 and B9 at 2505 16th Avenue South

7/19/2008



Sample ID	As ppm
N3W07-2505(16ths)-TR-SS3-PCR1	177
N3W07-2505(16ths)-TR-SS3-PCR2	278
N3W07-2505(16ths)-TR-SS3-PCR3	491
Average Concentration	315

Sample ID	As ppm
N3W07-2505(16ths)-TR-SS6-PCR1	290
N3W07-2505(16ths)-TR-SS6-PCR2	212
Average Concentration	251

Sample ID	As ppm
N3W07-2505(16ths)-TR-SS12-PCR1	65
N3W07-2505(16ths)-TR-SS12-PCR2	100
N3W07-2505(16ths)-TR-SS12-PCR3	140
Average Concentration	102

Sample ID	As ppm
N3W07-2505(16ths)-TR-SS15-PCR1	ND 5
N3W07-2505(16ths)-TR-SS15-PCR2	ND 6
Average Concentration	ND 6

Notes:
1) PCR IDs in trench samples indicate in situ replicates of unprepared samples shot in plastic cups using a different orientation of the XRF window for samples collected from the same depth interval.

Resources

- Case studies on resources CD provided
- Case studies and profiles on http://www.triadcentral.org/
- U.S. EPA Technical Bulletin - "Performing Demonstrations of Method Applicability Under a Triad Approach"
  - Due out this year: http://www.clu-in.org/
- Discussions with European and US Triad practitioners