

Module 4

# Where Does Uncertainty Come from When Making Restoration Decisions?



*“As we know, there are known knowns. There are things we know we know. We also know there are known unknowns. That is to say we know there are some things we do not know. But there are also unknown unknowns, the ones we don't know we don't know.”*

Donald Rumsfeld, Feb. 12, 2002,  
Department of Defense news briefing

## Decision Uncertainty Comes from a Variety of Sources

- Political, economic, organizational, and social uncertainty (outside scope of discussion)
- Model uncertainty (also outside discussion scope, although approaches to be discussed may provide mechanisms for addressing this)
- Data uncertainty. Data uncertainty refers to the uncertainty introduced into decision-making by uncertainty associated with data sets used to support decisions. ***Primary focus of the training...where statistics play a role.***

## Experience Has Demonstrated That Cleanup Work Is Filled with Uncertainty

- Removed soils volumes always greater than those estimated during the design phase
- DOE Ohio experience:
  - Fernald, 817,500 yd<sup>3</sup> more soil than expected requiring off-site disposal
  - West Jefferson, three times as much soil as expected
  - Mound, twice as much soil as expected
- FUSRAP experience:
  - Wayne: estimated – 58,000 yd<sup>3</sup>; revised – 110,000 yd<sup>3</sup>; actual 96,000 yd<sup>3</sup>
  - Middlesex: estimated – 24,000 yd<sup>3</sup>; actual – 41,000 yd<sup>3</sup>
- Complicates:
  - Program planning
  - Cost estimation
  - Remedial design and implementation



September 12, 2012

Portsmouth Training

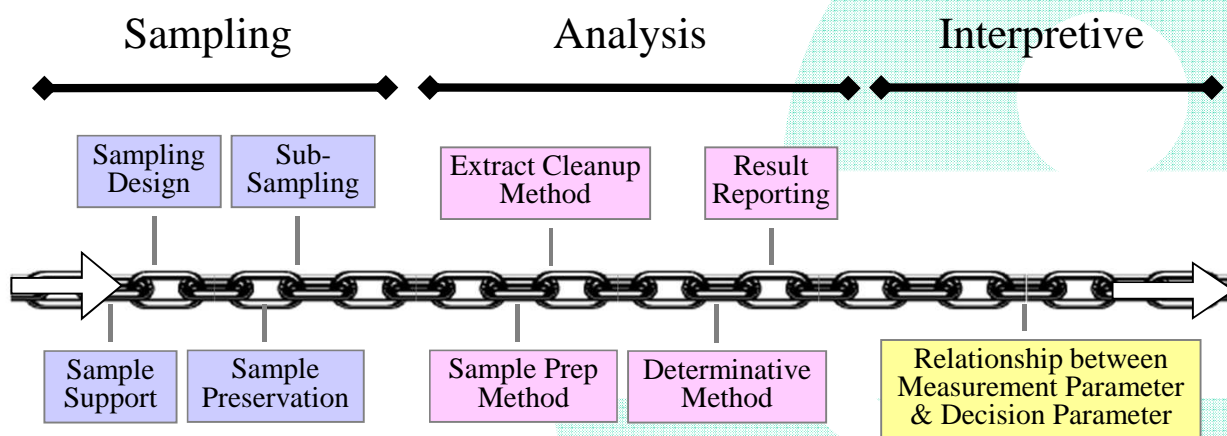
## Uncertainty Arises from Complexity and Heterogeneity of Natural Systems Combined with the Sparseness of Characterizing Data

- Spatial heterogeneity is primary source of variability observed in environmental sample results
- Sample results can vary by orders of magnitude for proximal samples
- Historically the cost of collecting/analyzing samples has been significant, limiting data that are available
- The result: decision-making taking place in a fog of uncertainty

September 12, 2012

Portsmouth Training

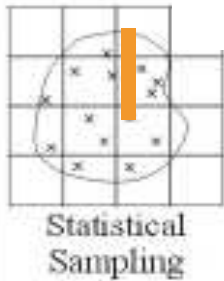
## Decision Quality Only as Good as the Weakest Link in the Data Quality Chain



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!

# Taking a Sample for Analysis

## Population



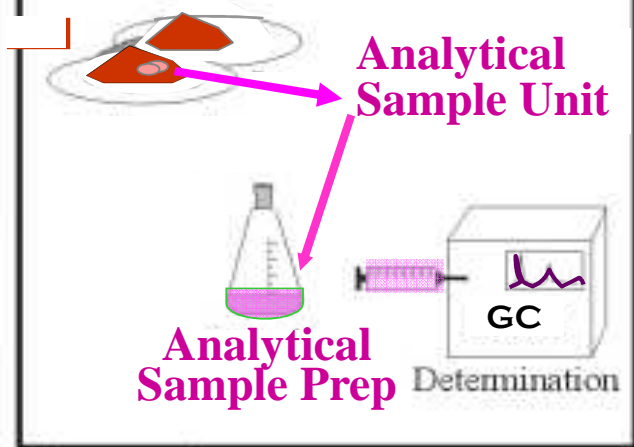
## Soil Core Sample



## Field Subsample



## Lab Subsamples (Duplicates)



23.4567 ppm

## What Contributes to Observed Variability in Sample/Measurement Results?

- Measurement error (RSD < 10%)
- Sample preparation/homogenization OR geometry/environmental variations for in situ measurements (RSD < 20%)
- Natural heterogeneity (RSD > 20%)
- All three of these are a function of concentration (i.e., all three are at a minimum when concentrations are at background levels, but grow as concentration levels in an area increase)

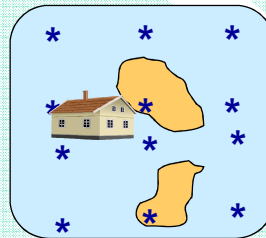
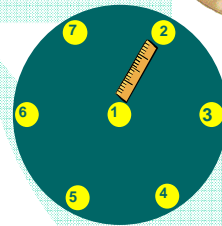
September 12, 2012

Portsmouth Training



# Heterogeneity Causes Data Variability

- **Heterogeneity:** Variations in matrix properties
  - **Within-sample** micro-scale heterogeneity
  - **Short-scale, between-sample** heterogeneity (affects agreement between collocated samples)
  - **Large-scale, between-sample** heterogeneity (on scale of conventional distances between samples)



September 12, 2012

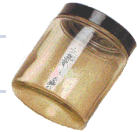
Portsmouth Training

9

# Within Sample Variability

Micro-scale, within sample jar: measured by lab duplicates  
(2 analytical subsamples taken from same sample jar)

## Pb: ICP lab duplicates



Lab dup examples from  
actual site data. Population  
characteristic being  
measured = Pb conc

	original	dup
	1038	688
	332	412
	874	2187
	34	49
	248	223
	69	45

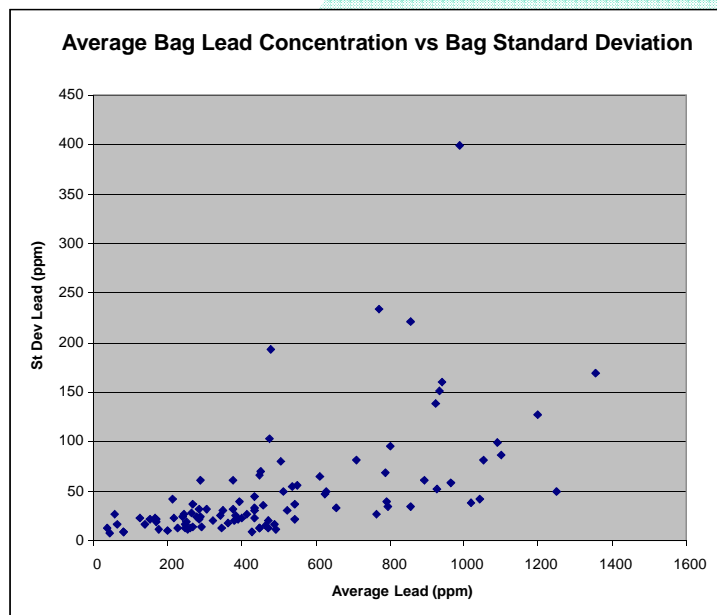
September 12, 2012

Portsmouth Training

10

# Within-Sample Variability

- 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



September 12, 2012

Portsmouth Training

11

## Within-Sample Variability: Interaction between Contaminant & Matrix Materials

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")	<b>10</b>
Between 3/8 and 4-mesh"	<b>50</b>
Between 4- and 10-mesh	<b>108</b>
Between 10- and 50-mesh	<b>165</b>
Between 50- and 200-mesh	<b>836</b>
Less than 200-mesh	<b>1,970</b>
<b>Bulk Total</b>	<b>927 (wt-averaged)</b>

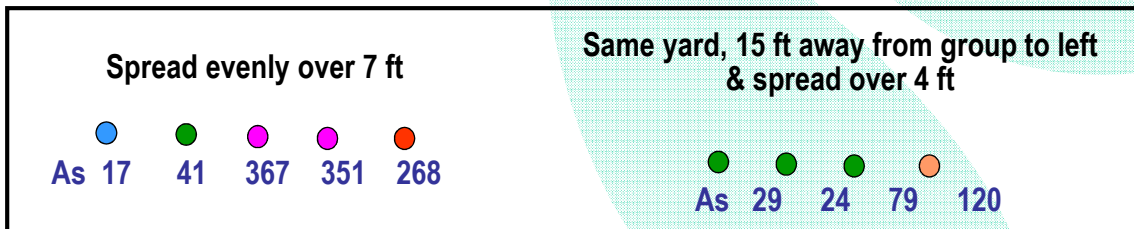
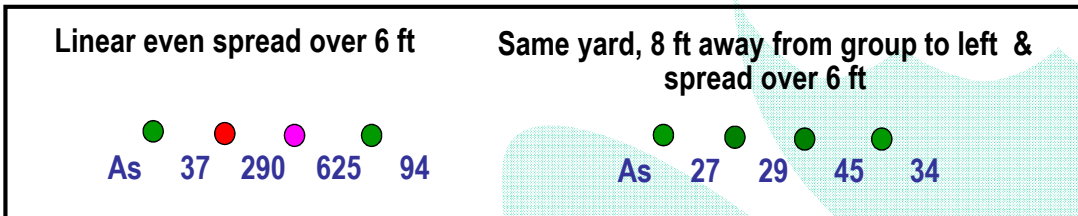
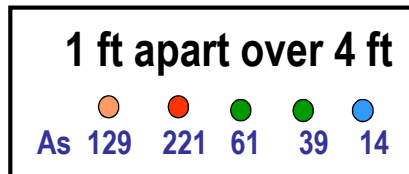
**The decision determines representativeness**

September 12, 2012

Portsmouth Training

12

# Short-Scale Heterogeneity Can Be Significant: Arsenic in Samples from 3 Residential Yards

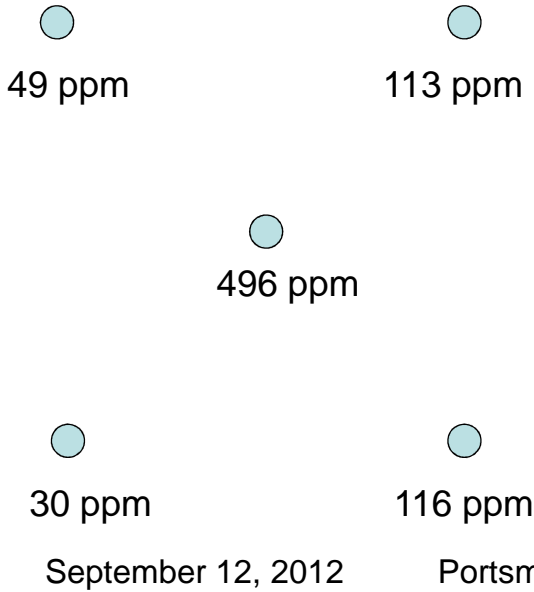


September 12, 2012

Portsmouth Training

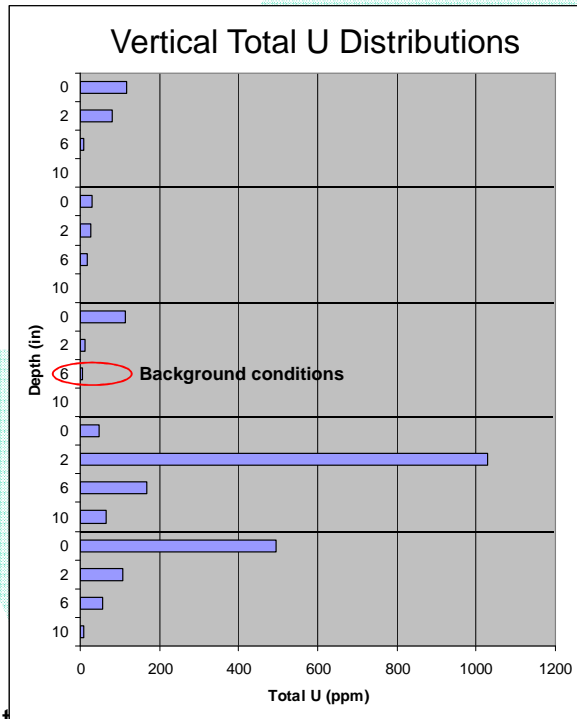
# Short-Scale Variability Can Be Significant: Uranium in Soils

Uranium over 1-ft<sup>2</sup> surface area



September 12, 2012

Portsmouth Training



# Short-Scale Variability Can Be Significant: Explosives

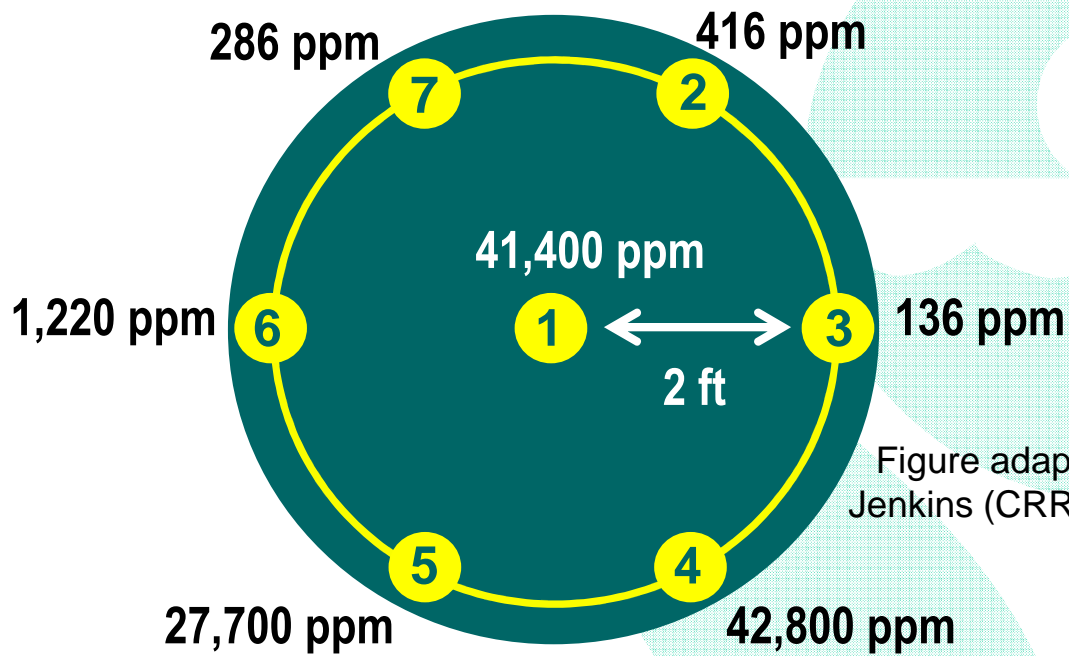


Figure adapted from Jenkins (CRREL), 1996

September 12, 2012

Portsmouth Training

15

# Heterogeneity Overwhelms Variability from Different Analytical Techniques

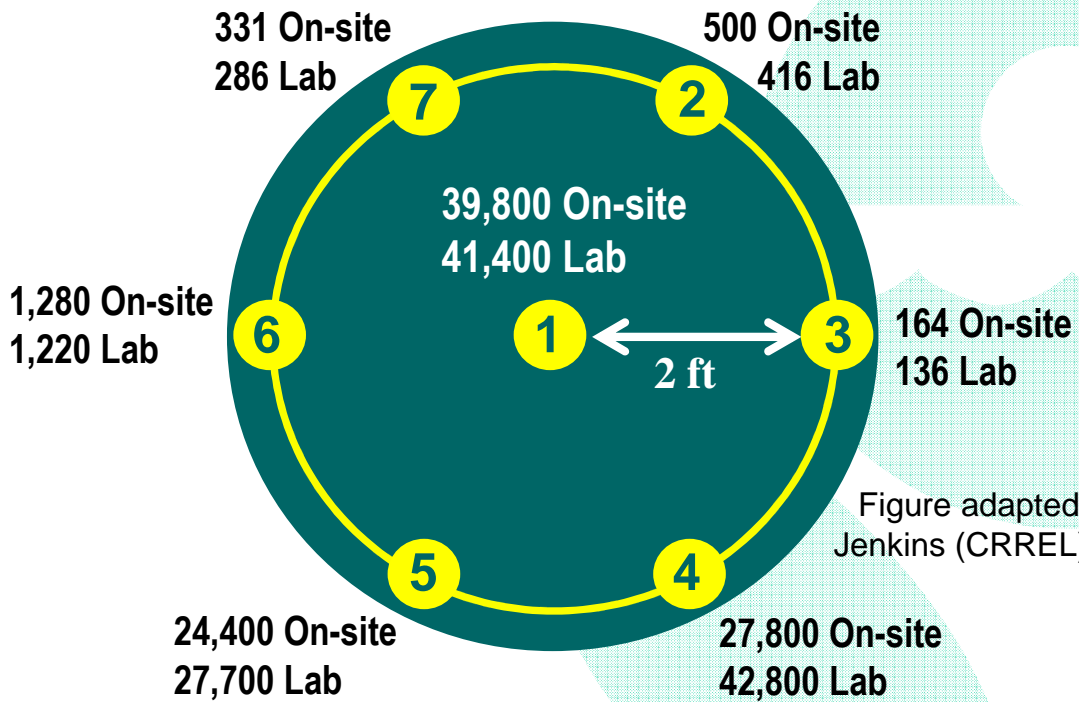


Figure adapted from Jenkins (CRREL), 1996

September 12, 2012

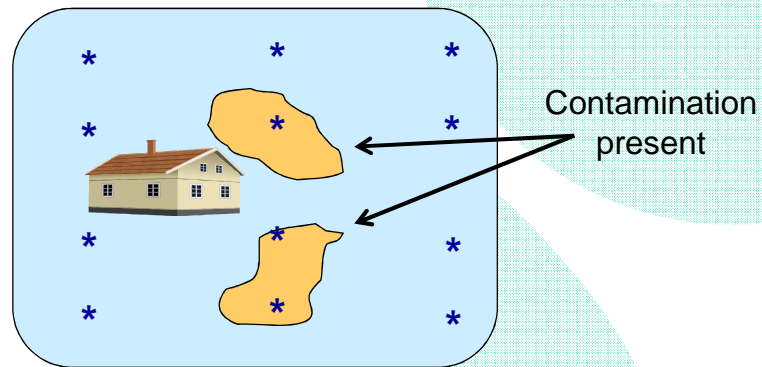
Portsmouth Training

16



# Variability at Highest Spatial Scale

Large-scale variability: differences in concentration at the scale of typical sampling design spacing; the kind of conc. variability traditional sampling designs are trying to find



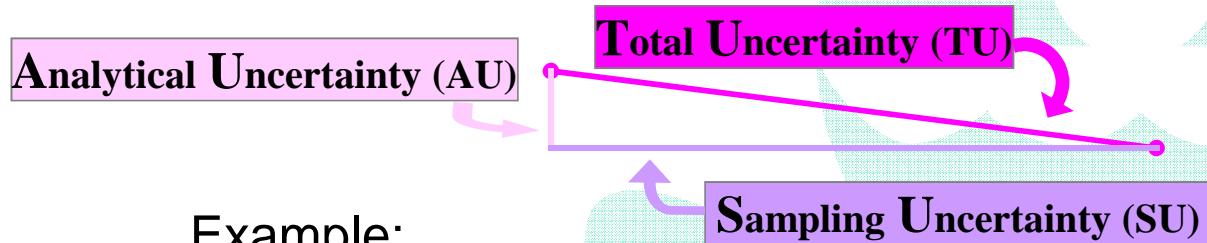
September 12, 2012

Portsmouth Training

17

## Uncertainty Math Magnifies Weakest Link's Effects in Data Quality Chain

Uncertainties add according to  $a^2 + b^2 = c^2$



### Example:

- AU = 10 ppm, SU = 80 ppm: **TU = 81 ppm**
- AU = 5 ppm, SU = 80 ppm: **TU = 80 ppm**
- AU = 10 ppm, SU = 40 ppm: **TU = 41 ppm**
- AU = 20 ppm, SU = 40 ppm: **TU = 45 ppm**

## Historically Focus Has Been Analytical Quality, But...

- Emphasis on fixed laboratory analyses following well-defined protocols
- Analytical costs driven to a large degree by QC/QC requirements
- Result:
  - analytical error typically on order of 30% or less for replicate analyses
  - traditional laboratory data treated as “definitive”...but definitive about what?

The Biggest Cause of Misleading Data

Heterogeneity Rules!



*You Can't Fool Mother Nature*

September 12, 2012

Portsmouth Training

20

## How Do We Reduce Data Uncertainty?

- For analytical errors:
  - Switch to a better analytical technique
  - Improve QC on existing techniques
- For sample prep and handling errors:
  - Improve sample preparation
- For sampling errors:
  - Collect samples from more locations

September 12, 2012

Portsmouth Training

21

***We can't control the effects of uncertainty on our decisions if we don't know where it is coming from.***

***Historically the cleanup business has focused resources on the wrong sources of data uncertainty.***

# Any Questions?



September 12, 2012

Portsmouth Training

23