Data Quality Performance for Different Sampling Strategies for Residential Soil Pb

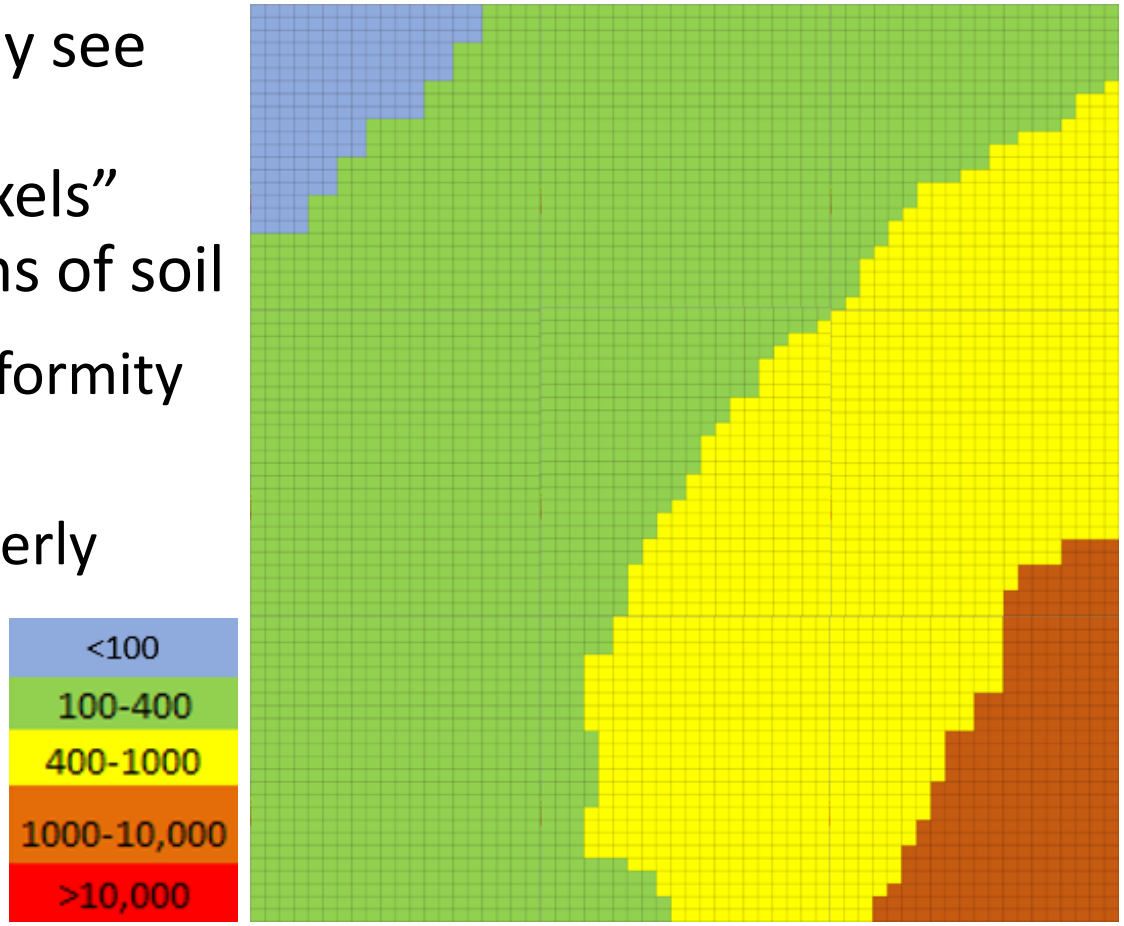
Summary of 3 Studies of Residential Properties in Mining-Contaminated Superfund Sites

Presenter: Deana Crumbling, USEPA

NEMC 2017, Washington, DC

Common Conception of Soil Contamination

- Graphic representations similar to below
- But if we could visually see actual contaminant concentrations as “pixels” representing 0.5 grams of soil
  - ➢Would there be uniformity within a zone?
  - ➢Would there be orderly transitions from one zone to another?



A Grad Student Research Study

Becker Thesis Study (Purdue Univ.)

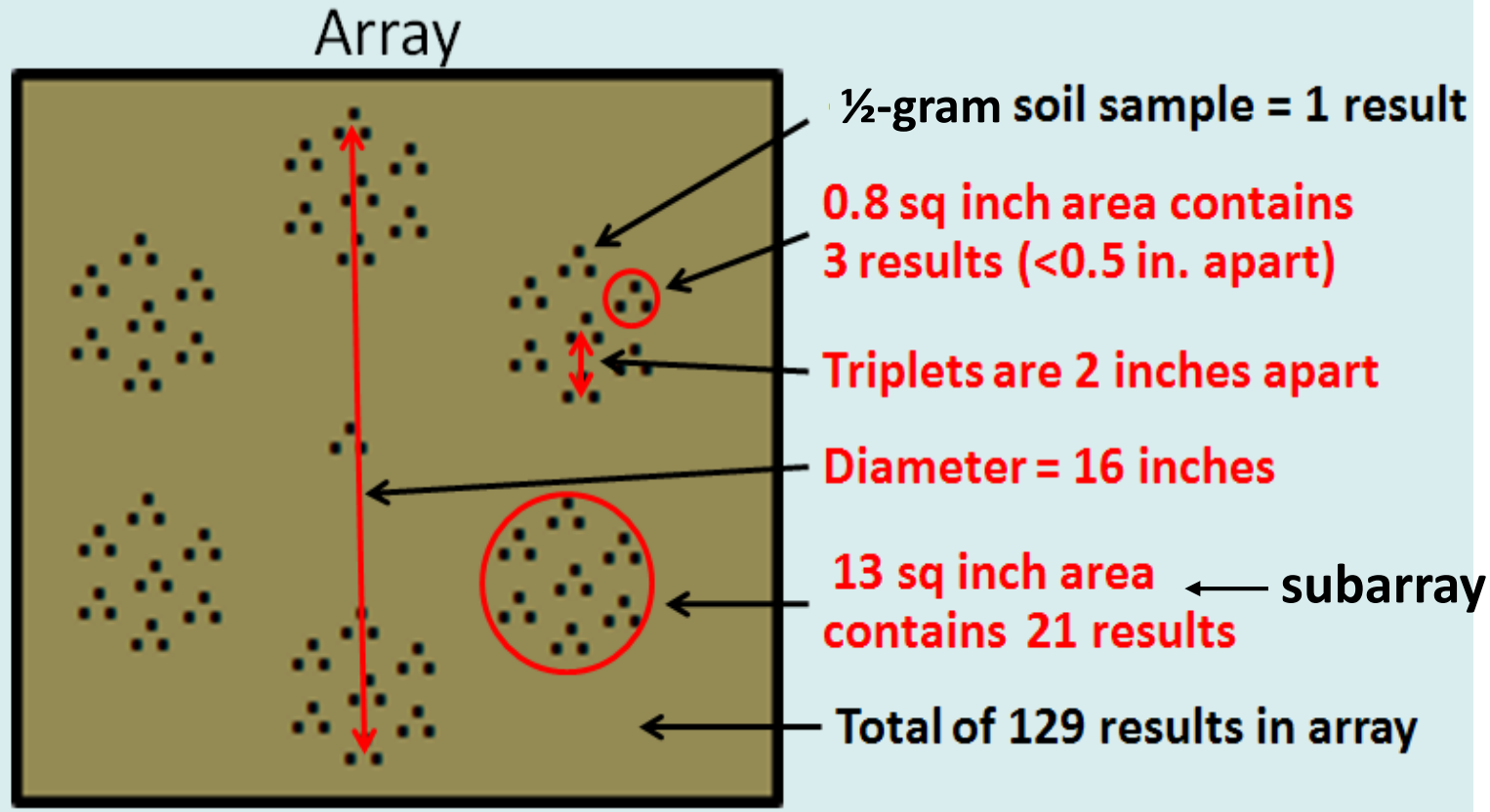
- **0.5-gram** samples collected from soil surface in a soil area contaminated by DOT **road paint** (white & yellow line paint)
- White paint contained **Pb**; yellow contained Cr
- Samples on the ground were arranged in 4 arrays
- Each array holds 129 0.5-g samples over a 16-inch circular diameter (1.4 ft^2), arranged in 6 subarrays
- Each subarray contains 21 samples over a 4-inch diameter, arranged in 7 groups of 3 (triplets)
- Triplet groups are about 2 inches apart
- Samples in a triplet are 1 cm (<0.5 inch) apart
- Total of **516 individual samples over about 6 ft^2**

Becker Thesis Study

Each array holds 129 0.5-g samples over 1.4 sq.ft.

Each subarray contains 21 samples over 13 sq.in.

Samples in triplets are 1 cm apart.

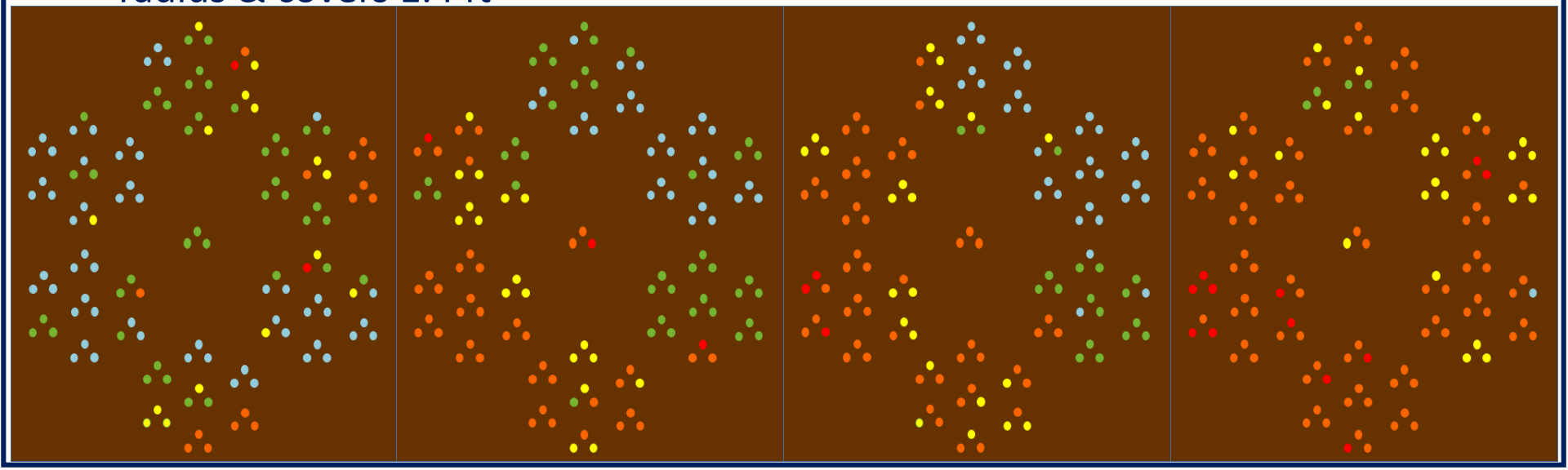


Pb Data Variability Due to Soil Heterogeneity

516 individual data points over ~6 sq.ft. of paint-contaminated soil

Each circular array has 8-inch radius & covers 1.4 ft^2

Average over all 4 arrays = 1907 ppm Pb



Concentration of the 516 individual samples in the contaminated area ranged from 3 to 29,000 ppm

Although overall average ~2000 ppm,

43% were <400 ppm & 23% were <100 ppm

Non-Representative Data

- "Non-representative" when data does not represent the concentration of the soil around it with respect to the decision
- Scale Mismatch: Scale of data generation (sample collection & analysis for contaminants) is 10-100 *million* times smaller than scales of cleanup-related decisions
- Heterogeneity: Contaminants selectively bind to soil particles so contaminants segregate by particle size, creating strong biases during laboratory subsampling
- Interaction between scale mismatch & heterogeneity causes the "representativeness" problem
  - Soil data are highly variable & often non-reproducible at scales of both field sampling and lab subsampling
- Routine practices & procedures not currently structured to control soil heterogeneity

Meta-analysis to Compare Routine vs. Newer Soil Sampling Techniques for Pb

- Data sets from 3 separate field studies analyzed & summarized with respect to 2 aspects of performance relating to data representativeness
  - How much data variability is present in the data and what are the implications for basing decisions on single sample results?
  - What are the false positive (FP) & false negative (FN) decision error rates around thresholds of 200 and 400 ppm Pb?
    - If a single concentration result is above a decision threshold (a "positive" result), what is the likelihood it represents a FP decision error?
    - If a concentration result is below a decision threshold (a "negative" result), what is the likelihood it represents a FN decision error?

The 3 Field Studies

- CO Pueblo: CO State study of 81 residential yard areas (i.e, decision units, DUs) within town of Pueblo, CO (Pb & As from smelter operations); surficial **grab** samples only; routine EPA procedures used, meaning that both **within-DU & within-sample variabilities were uncontrolled**
- EPA Pueblo: EPA R8 study of 155 residential yard DUs within same area of town; 4 depth intervals to 1.5 ft bgs; 18 DUs had triplicates for both 5-pt composites & 30-pt incremental samples so **within-DU variability was controlled**; rest of the DUs had triplicate 5-pt only; all samples processed using incremental procedures* so **within-sample heterogeneity was controlled** (compare 5-pt vs 30-pt performance)
- Furnace Creek: EPA R7 study of 23 residential yard DUs within Caledonia, MO (Pb from smelter operations); surficial soil only; each DU had triplicates for 5-, 9-, and 30-pt composites so **within-DU variability was controlled**; all samples processed using incremental procedures so **within-sample heterogeneity was controlled** (compare 5-pt vs 9-pt vs 30-pt performance)

*Description of incremental procedures available at http://www.itrcweb.org/ISM-1

Summary of Sampling Types Covered by the Data Set Comparisons Among the 3 Field Studies

Soil Sampling Type Replication within Each Project DU	**CO-Pueblo** (reps/DU)	**EPA-Pueblo** (reps/DU)	**Furnace Creek** (reps/DU)
Type 1: Grab samples [no comparison to other sampling types]	5 per DU N = 81 DUs	---	---
Type 2: 5-point composites [no comparison to other sampling types]	---	3 per DU N = 135 DUs	---
Type 3: 5-point composites [compared to 9-pt and/or 30-pt composites]	---	3 per DU N = 20 DUs	3 per DU N = 23 DUs
Type 4: 9-point composites [compared to 5-pt and 30-pt composites]	---	---	3 per DU N = 23 DUs
Type 5: 30-point composites [compared to 5-pt and/or 9-pt composites]	---	3 per DU N = 20 DUs	3 per DU N = 23 DUs

During data analysis, data sets were trimmed where needed to ensure comparability across the 3 studies. The final "N" for each measure is provided in the data summary tables that follow.

Summary of Data Variability/Imprecision
(as mean %RSD for each of the project's sampling types)

%RSD as Performance Measure of Data Variability	**CO-Pueblo** mean %RSD	**EPA-Pueblo** mean %RSD	**Furnace Creek** mean %RSD
Type 1: %RSD calc'd for each DU's set of 5 grab samples, then averaged over N sets	43.5% RSD (N = 81 DUs)	---	---
Type 2: %RSD for each DU's triplicate set of 5-pt composites, then averaged over N sets	---	26.7% (N = 135 DUs)	---
Type 3: %RSD for each DU's triplicate set of 5-pt composites, then averaged over N sets	---	21.6% (N = 18 DUs)†	24.3% (N = 22 DUs)*
Type 4: %RSD for each DU's triplicate set of 9-pt composites, then averaged over N sets	---	---	14.9% (N = 22 DUs)*
Type 5: %RSD for each DU's triplicate set of 30-pt composites, then averaged over N sets	---	7.2% (n = 18 DUs)†	8.2% (N = 22 DUs)*

† 2 DUs removed from summary calculation because extremely low DU Pb concentrations (<25 ppm) mathematically caused anomalously high RSDs (>50%).

* 1 DU removed from summary calculation due to presence of small garbage burn pit.

Statistically Significance Differences Among the Variability of Different Sampling Techniques

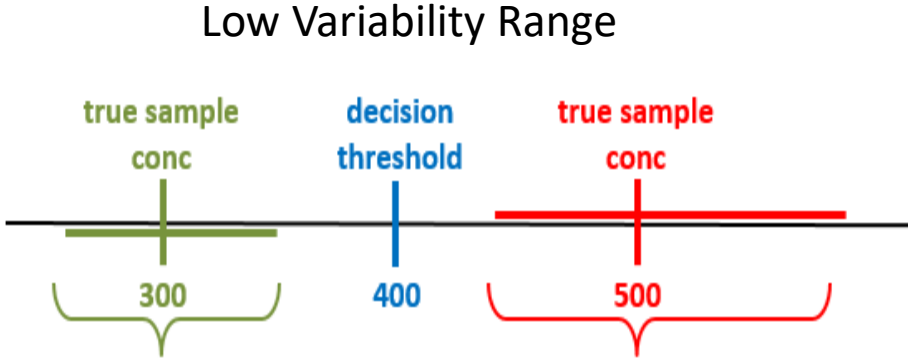
- All field samples of the EPA-Pueblo & Furnace Creek studies were processed & analyzed exactly the same way
  - So differences in variability among the different sampling techniques is due to different numbers of increments per field sample
- The Wilcoxon Signed-Rank test (WSR)
  - A non-parametric test of medians for paired measures (e.g., paired 5-pt triplicate RSD and 30-pt triplicate RSD from the same DU).
  - The WSR tests whether medians are the same, so it is insensitive to outliers; allows inclusion of all relevant project data.

Summary of WSR Tests

- An on-line calculator for the WSR test (www.socscistatistics.com) was used to test for statistically significant differences between paired RSDs from different sampling techniques (1-sided tests were used).
- Since the CO-Pueblo has discrete samples only (no paired composite samples), there are no WSR tests for that project.
  - The median RSD for CO-Pueblo was 30.6% (N = 81 DUs).
  - Note: number of increments for a discrete sample = 1

Project	RSD medians tested 5-pt	9-pt	30-pt	DUs containing paired comparison data	WSR test p-value
EPA-Pueblo	7.4%		4.9%	N = 20	0.0014
Furnace Creek (burn pit excluded)	14.1%	12.5%		N = 22	0.07
		12.5%	6.1%	N = 22	0.0001

Affect of Data Variability on Decision Error when a 400 ppm Pb Decision Is Based on a Single Sample Result



Study & Sample Type	Imprecision Level %RSD ⟺ %RPD	Range of Possible Results if True DU Conc = 300 ppm Pb	Range of Possible Results if True DU Conc = 500 ppm Pb
T5: EPA-Pueblo/30-pt	7% RSD = 10% RPD	272 - **300** - **331**	**453** - **500** - 552
T5: Furnace Creek/30-pt	8 11	268 - **300** - **336**	**446** - **500** - 560
T4: Furnace Creek/9-pt	15 21	242 - **300** - **371**	**404** - **500** - 619
T3: EPA-Pueblo/paired 5-pt	20 28	226 - **300** - **400**	**376** - **500** - 665
T3: Furnace Creek/5-pt	24 34	213 - **300** - **423**	**350** - **500** - 715
T2: EPA-Pueblo/solo 5-pt	27 38	204 - **300** - **442**	**340** - **500** - 736
T1: CO-Pueblo/grabs	43.5 61	159 - **300** - **567**	**265** - **500** - 944

If a DU's true concentration is **±100** from the **400 ppm** action level (**300** or **500**), and data generation has 20%RSD imprecision or greater, false positive & negative decision errors are possible.

FP/FN Decision Error Rates at 2 Pb Thresholds
(based on XRF results for N sets of replicate DU field samples)

Soil Sampling Type FP/FN Rates over N	CO-Pueblo @400 (ppm) @200		EPA-Pueblo @400 (ppm) @200		Furnace Creek @400 (ppm) @200	
Type 1: Grab samples/ compiled over 73 DUs*	FP = 42.1% FN = 5.5%	FP = 6.3% FN = 40.4%	---	---	---	---
Type 2: 5-pt composites/ compiled over 111 DUs*	---	---	FP = 7.6% FN = 4.2%	FP = 6.3% FN = 40.4%	N = 21 for Furnace Creek comparison DUs	
Type 3: 5-pt composites/ compiled over N DUs*	---	---	FP = 21.7% FN = 0.0%	FP = 0.0% FN = 100%	FP = 7.1% FN = 10.2%	FP = 2.7% FN = 21.1%
Type 4: 9-pt composites/ compiled over 21 DUs*	---	---	N = 16 for EPA-Pueblo comparison DUs		FP = 22.7% FN = 2.4%	FP = 5.1% FN = 8.3%
Type 5: 30-pt composites/ compiled over N DUs*	---	---	FP = 5.3% FN = 0.0%	FP = 0.0% FN = 0.0%	FP = 0.0% FN = 0.0%	FP = 4.9% FN = 0.0%

*DUs with average concentrations <100 or >800 ppm Pb were excluded from FP & FN decision error rate tallies so counts reflected the same conc range across all 3 studies.

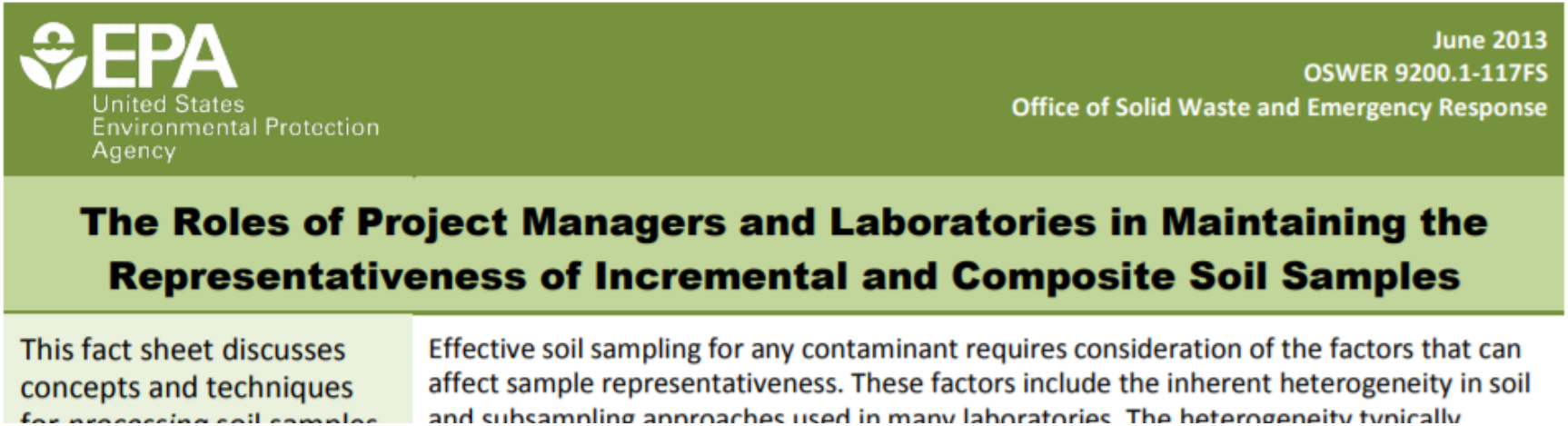
Conclusions

- Data variability & the DU concentration both influence FP & FN decision error rates.
  - More data variability can be tolerated when DU concentrations are much lower, or much higher, than the decision threshold
- In others words, the ability of a sample to represent the true condition of a DU with respect to a specific decision depends on at least 4 things:
  - Project design that defines the proper population to be sampled
  - DU concentration
  - The value used as the decision threshold
  - Data variability (combines field heterogeneity & subsampling error)
    - ✓ Both RSD means & medians decrease as numbers of increments (per composite field sample) increase
    - ✓ Indicates increasing ability of a sample to represent the true DU Pb concentration.

The Role of Laboratories

- The Lab can only control for one of those factors
  - ➢Subsampling error can be controlled with sample processing, matching the analytical mass to soil particle size, and incremental subsampling (all are considered in incremental sampling procedures)
- The heaviest responsibility for sample representativeness rests on the remedial project manager
  - ➢They need to be taught this!

Teaching aid: https://clu-in.org/download/char/RolesofPMsandLabsinSubsampling.pdf



EPA
United States Environmental Protection Agency

June 2013
OSWER 9200.1-117FS
Office of Solid Waste and Emergency Response

The Roles of Project Managers and Laboratories in Maintaining the Representativeness of Incremental and Composite Soil Samples

This fact sheet discusses concepts and techniques for processing soil samples

Effective soil sampling for any contaminant requires consideration of the factors that can affect sample representativeness. These factors include the inherent heterogeneity in soil and subsampling approaches used in many laboratories. The heterogeneity typically