Building A Second-Generation Data Quality Model

Deana M. Crumbling, M.S.
Technology Innovation Program
U.S. Environmental Protection Agency
Washington, D.C. USA
(703) 603-0643
crumbling.deana@epa.gov

Triad Training, EPA National Site Assessment Symposium
San Diego, June 28, 2004

Oversimplified 1980's (First-Generation) Data Quality Model

Methods = Data = Decisions

Screening Methods	→	**Screening Data**	→	**Uncertain Decisions**
"Definitive" Methods	→	**"Definitive" Data**	→	**Certain Decisions**

This model made sense for the 1980s, but fails to distinguish between important concepts
Analytical Methods : Data Quality : Decisions

First Generation Data Practices

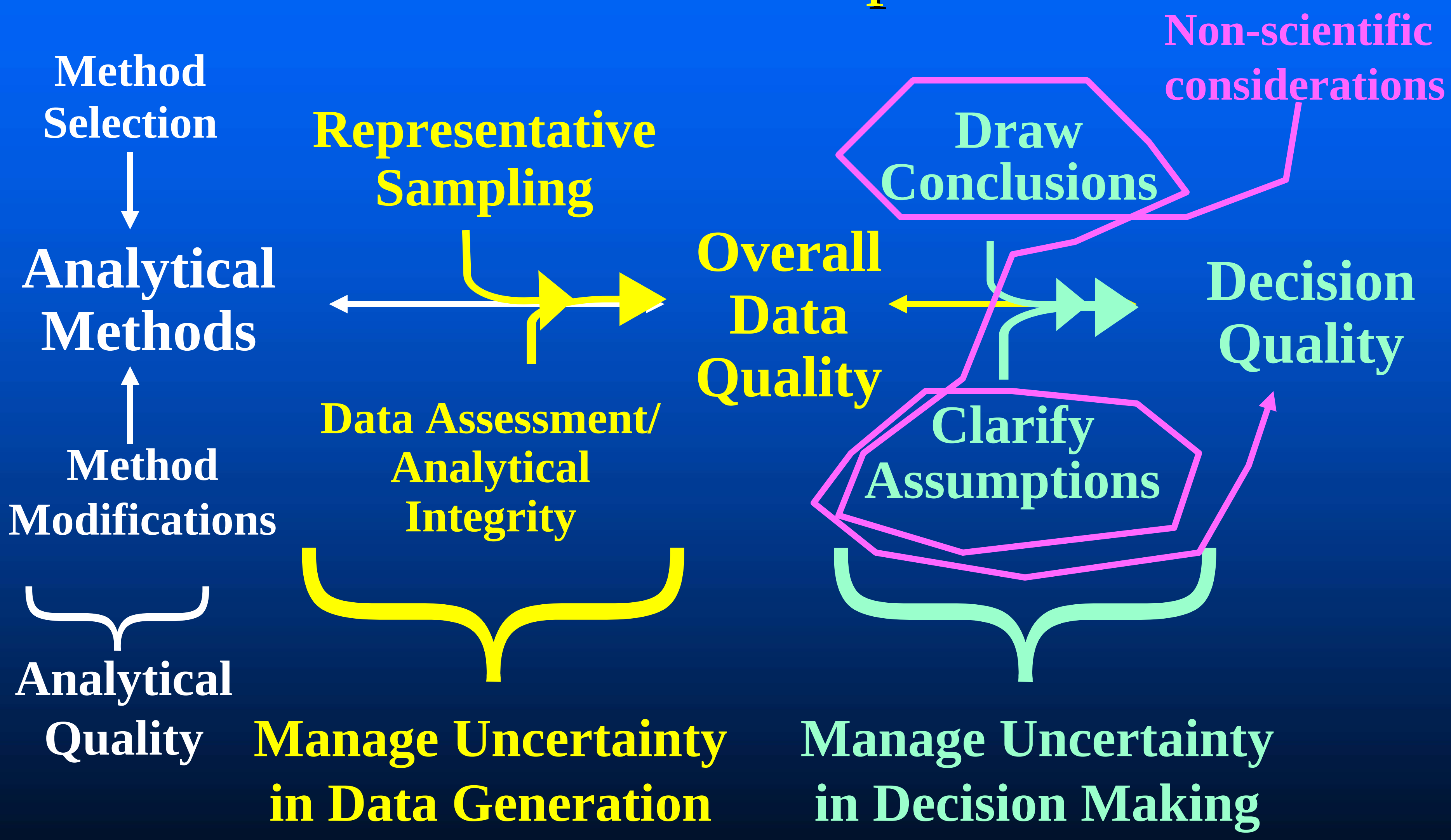
- “Data quality” judged by analytical methods & lab
- Regulator-approved methods = “definitive data”
- Uncertain extrapolations of data results
  - Results from tiny samples extrapolated to enormously larger volumes
  - Concentration values used as proxy for actual contaminant mass & bioavailability

Was OK as a Starting Point

- But 25+ years of experience doing things this way proves it not the “formula for success.”
- Does not produce confident decisions about nature & extent, exposure & site closure
- Does not lead to cost-effective cleanups & monitoring

Why not? What is missing?

1) Relationships between the Concepts is More Complex



2) Real-World is More Heterogeneous than the Old Data Quality Model Assumes

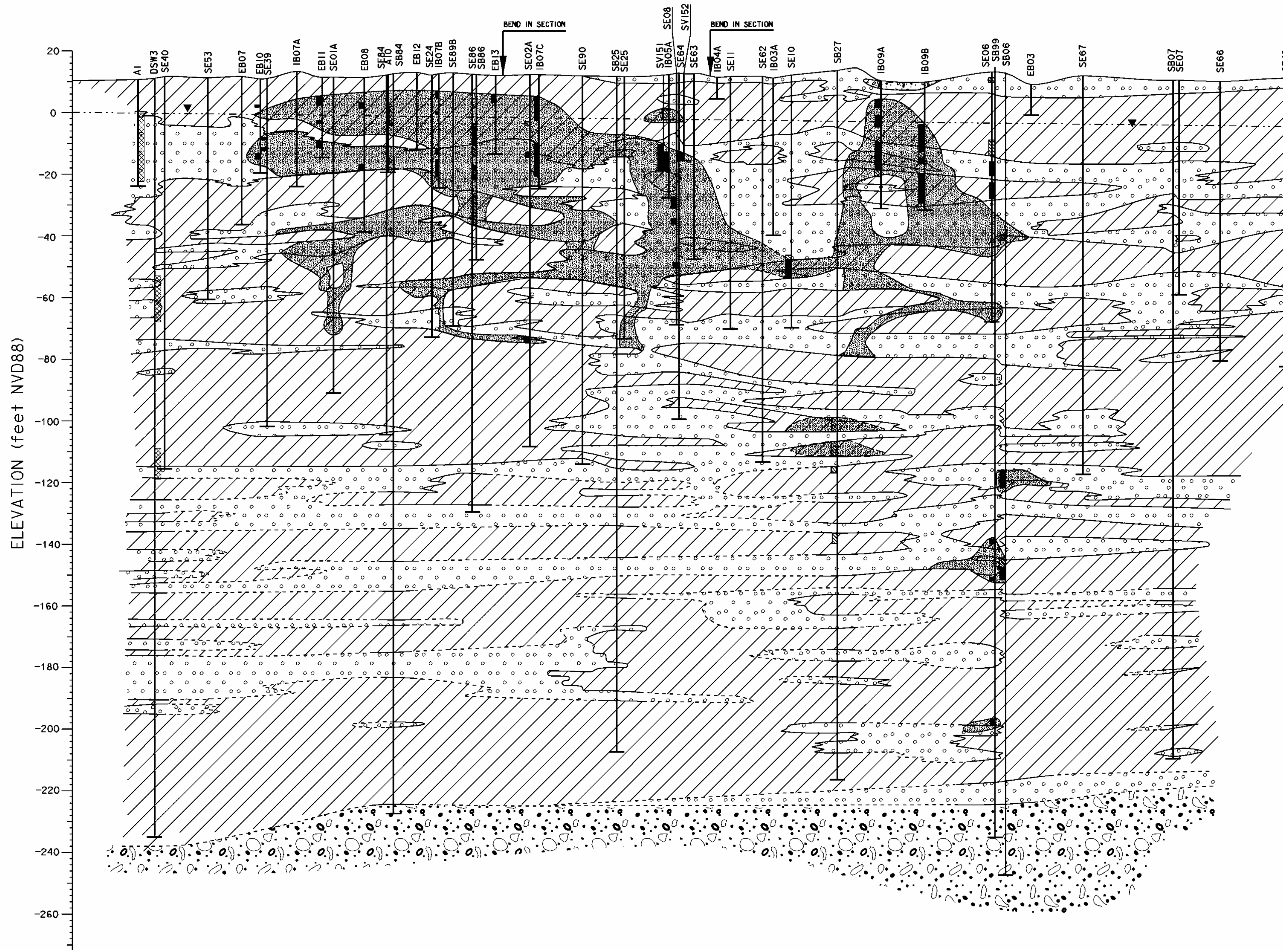
- Procedures assume contamination is relatively homogeneous (or is randomly variable)
- Simply not true for most sites
  - Release mechanisms create non-random spatial patterning at macro & micro scales
  - Physical transport → new spatial patterns, or reduce heterogeneity
  - Many contaminants behave like particulates
  - Degree of patterning depends on mechanism & scale of observation

You can't fool Mother Nature!

In a clash between a model & reality, reality always wins

A 2nd-Generation Environmental Data Quality Model

Write a new recipe for successful projects by changing the model to match the reality revealed by our new toys!


ELEVATION (feet NVD88)
20
0
-20
-40
-60
-80
-100
-120
-140
-160
-180
-200
-220
-240
-260
BEND IN SECTION
BEND IN SECTION
A1
DSW3
SE40
SE53
EB07
EB10
SE39
IB07A
EB11
SE01A
EB08
SE84
A10
SB84
EB12
SE24
IB07B
SE89B
SE86
SB86
EB13
SE02A
IB07C
SE90
SB25
SE25
SV151
IB05A
SE08
SE64
SV152
SE63
IB04A
SE11
SE62
IB03A
SE10
SB27
IB09A
IB09B
SE06
SB99
SB06
EB03
SE67
SB07
SE07
SE66

Building a New Data Quality Model
Definition

Data Quality = "a measure of the degree of acceptability or utility of data for a particular purpose." (USEPA QA/G-5, 2002)

- The "purpose" of data: make correct project decisions

- Then, data quality depends on the data providing accurate information about (i.e., representing) the "true state" (of the decision unit) <u>in the context of the decision</u> that the data user wants to make

Data Quality is about More than Just Chemical Analysis

Perfect Analytical Chemistry **+** **Non-Representative Sample(s)**

WRONG DECISION

"BAD" DATA

Data Quality =
Sample Representativeness + Analytical Quality

Need to Distinguish Analytical Quality from Data Quality

Triad Distinguishes between Analytical Quality & Data Quality

Data comes from samples

Therefore data quality must include sampling variables

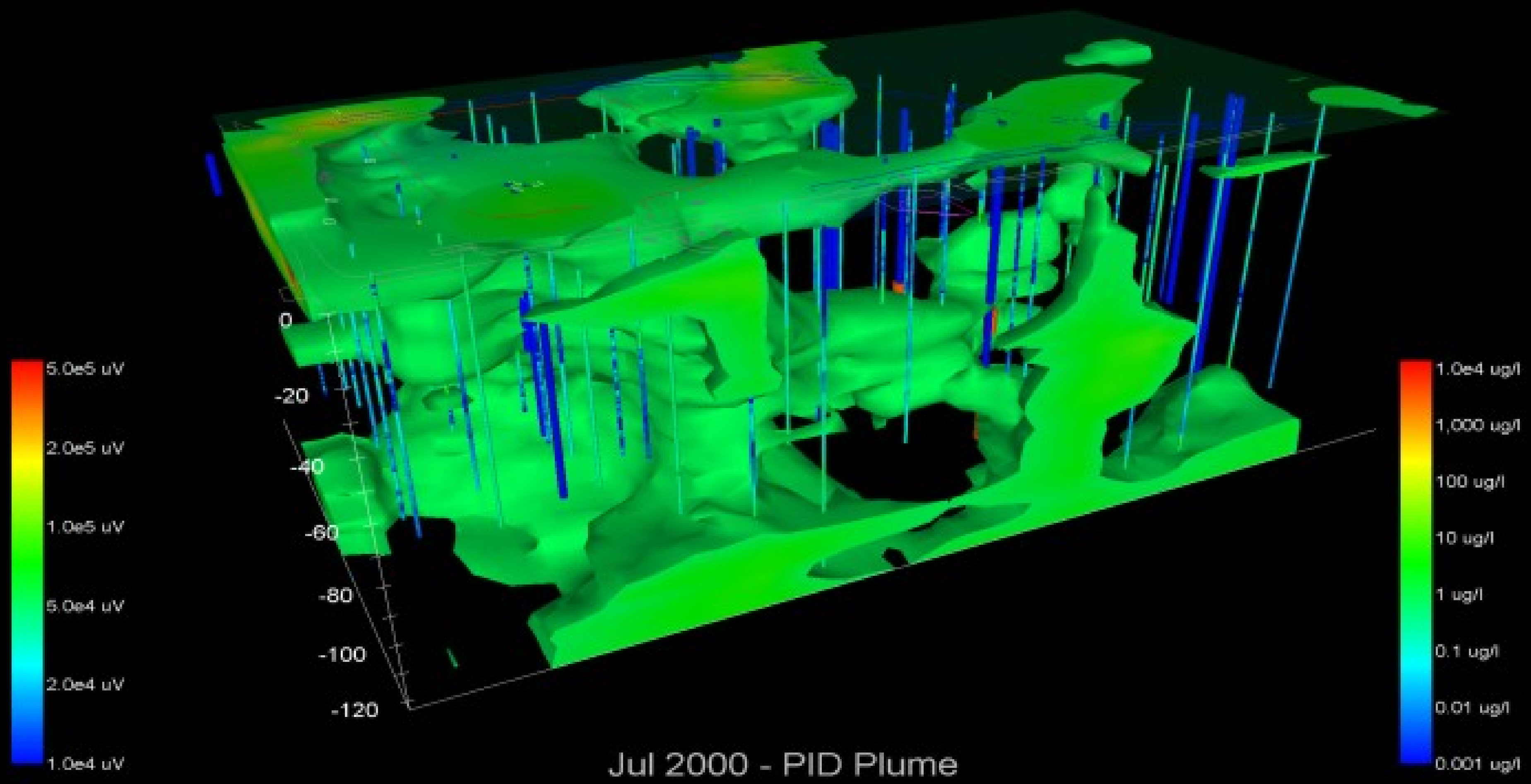
Be clear:
If intend to refer only to analytical side, then say "analytical quality"

But if say "data quality," be prepared to explain how sampling variables are managed.

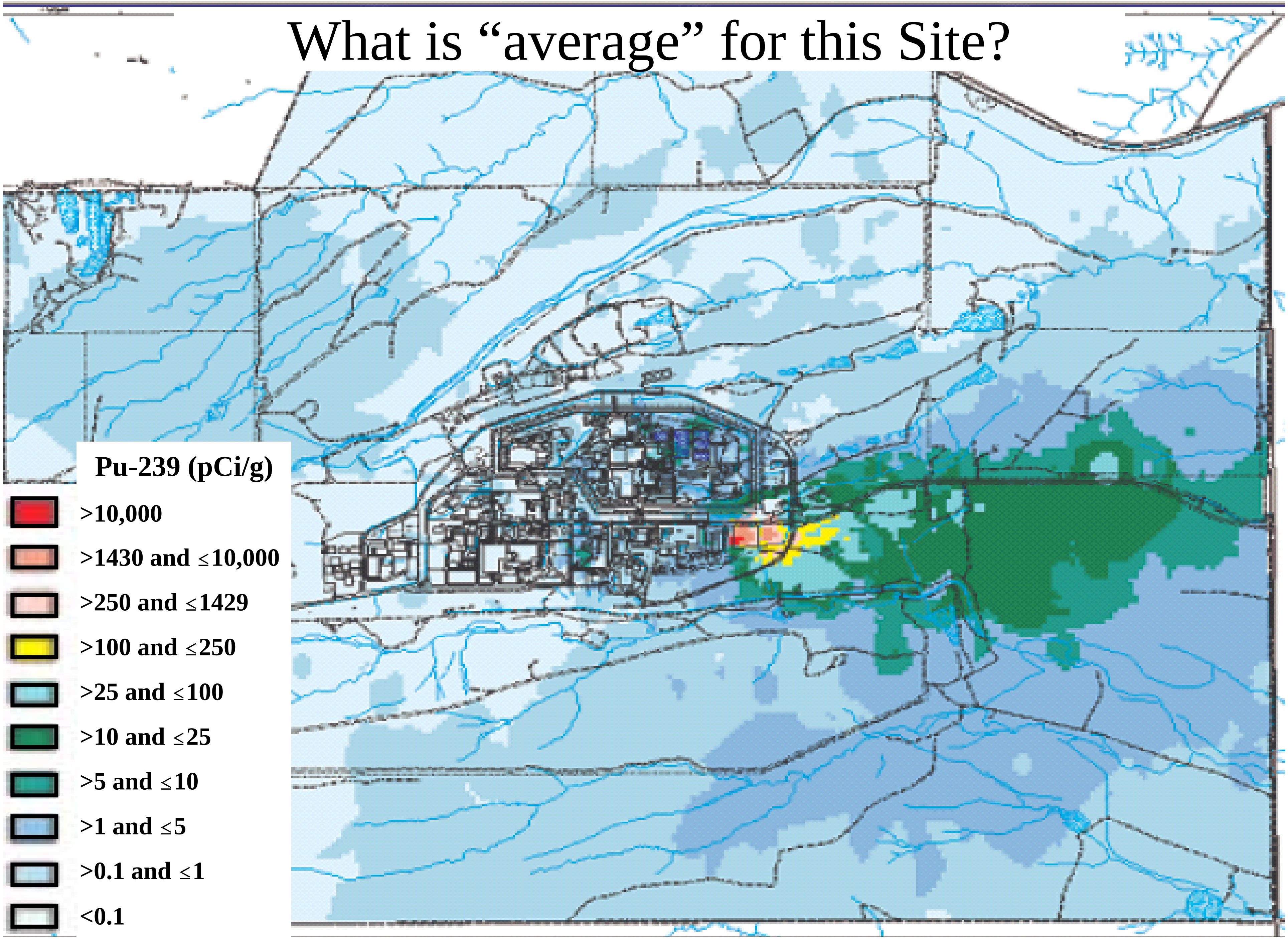
“Representativeness” Used to Mean “Average”

- Problem is heterogeneity. Concentrations can range orders of magnitude over very short distances
- If want an average, need to define “average over what”
  - Often not defined
  - When this variable not controlled, analytical results are variable
  - If undefined, no way to decide which result is “right”
- Even if could determine accurately determine the <u>average</u>, will it support good decisions about risk or remediation?

What is “average” for this Site?



What is “average” for this Site?



“Representativeness” as a Characteristic

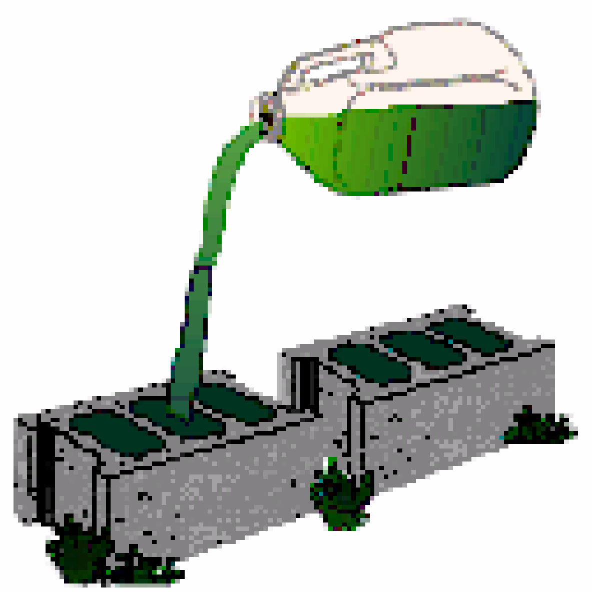
- ASTM & ANSI/ASQC define as
  - a characteristic of interest
  - of a population as defined by the project objectives
- Could be an average, but allows for characteristics other than the average
- Forces the desired characteristic to be defined
- Allows for the CSM to distinguish populations using a decision-based rationale

Data should be Representative of a "Target Population"

- Triad usage consistent with ASTM & ANSI/ASQC
- "Characteristics of interest" grounded jointly in the CSM & in the decision
- Decision defines what populations are of interest
- Different populations occur at both macro & micro scales
- Want to design sampling programs that can stratify populations
  - Ensures both representative data & cost-effective decisions

Macro Population Segregation

- Wenatchee site: 3 distinct soil decision-driven pop's
  - Compliant population (remain on site)
  - Mod non-compliant pop (landfill)
  - Severely contaminated pop (incinerate)



- No segregation = incinerate all: ~ $1.2 million (708 tons)
- Actual cost to clean closure using Triad = $589K
  - 56 tons incinerated, 334 tons landfilled
- Cost projected to be ~ $1.2 million if segregate by traditional data

How Was this Possible?

- Pesticide IA kits guided dynamic work plan: removed and segregated contaminated soil for disposal

230 **IA analyses** (w/ thorough QC) **+** 29 **fixed-lab** samples for 33 analytes

Managed **sampling uncertainty**: had very high confidence that all contamination above action levels was located and removed

Managed **field analytical uncertainty** as additional QC on critical samples: confirmed & perfected field kit action levels

- Clean closure data set
  - – 33 fixed lab samples for analyte-specific pesticide analysis
  - – Demonstrate full compliance with all regulatory requirements for all 33 pesticide analytes to >95% statistical confidence the first time!
- Field work completed: <4 months; single mobilization

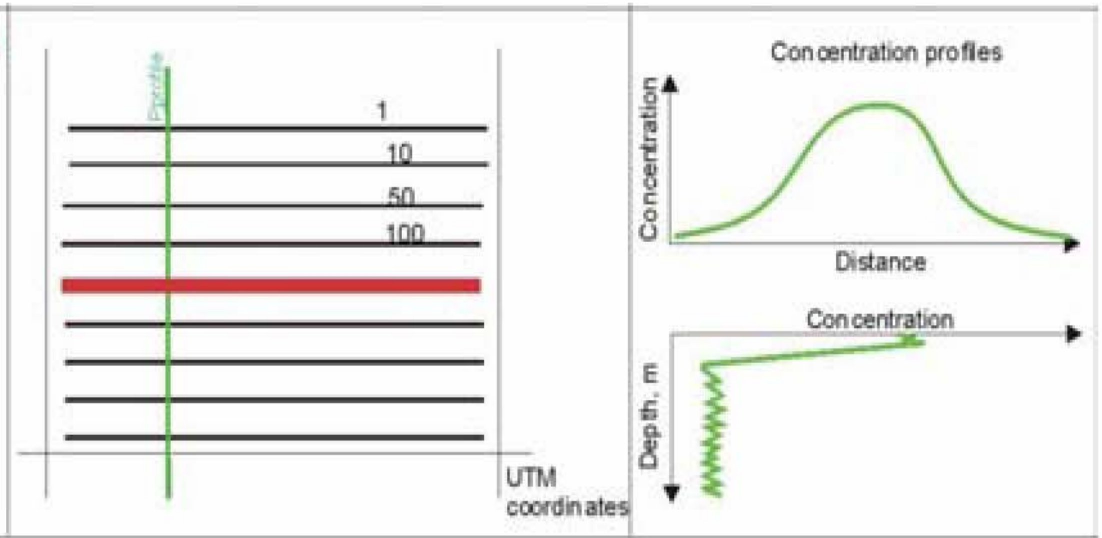
http://cluin.org/char1_edu.cfm#site_char

Populations with Different Spatial Distributions Need Different Sampling Strategies

Line Model

e.g. roads
railways

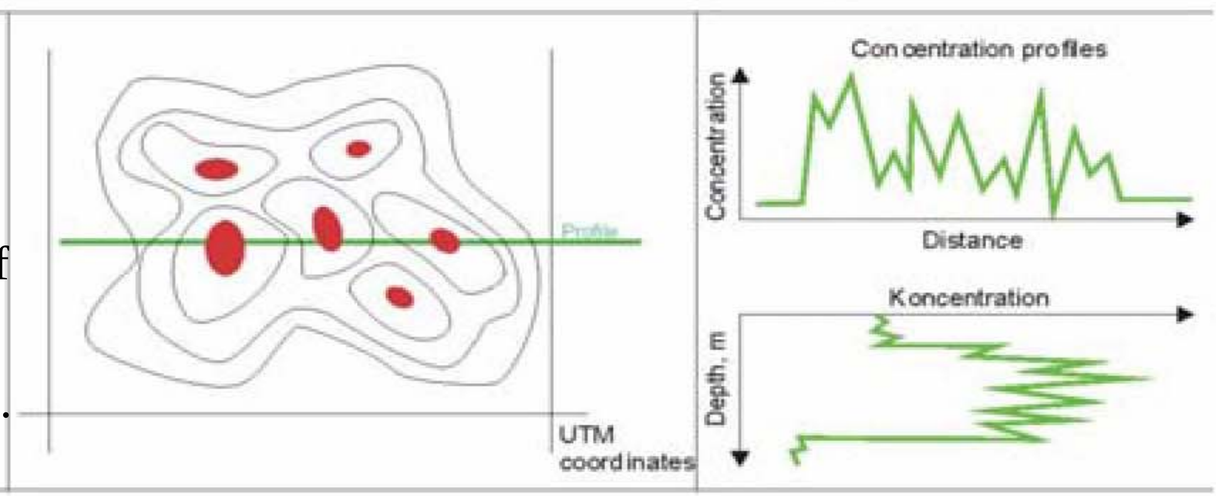
Soil impact from moving point sources aligned along a long linear element in the landscape, e.g. a road, railway.



Soil Fill Model

e.g. contaminated fill

Systematic dumping/filling with soil, waste, material of unknown origin variable and random load to the historic fill.



Falkenberg, et al., ConSoil (2003) Poster 1

Triad Representativeness Has 2 Stages

- 1st : Need to test & refine the CSM until confident that populations of interest are identified
  - “Populations” grounded in physical reality & decisions
- 2nd : Once CSM is refined enough to understand populations, then take samples known to represent those populations in order to measure the characteristics of interest
- Cost-effective and efficient when the 2 stages are blended together in real-time workflow

One-Size-Fits-All "Representativeness" Not Possible

Different decisions require different representativeness.

For example:

– Data set <u>representative of</u> contact over an exposure area should estimate the average concentration over the volume of the "exposure unit"

– Data set <u>representative of</u> an exposure pathway must detect & characterize a particular feature of interest (often not an average). For example:

» <200 mesh soil fraction is representative of dust exposure pathway, Pb conc = 2000 ppm

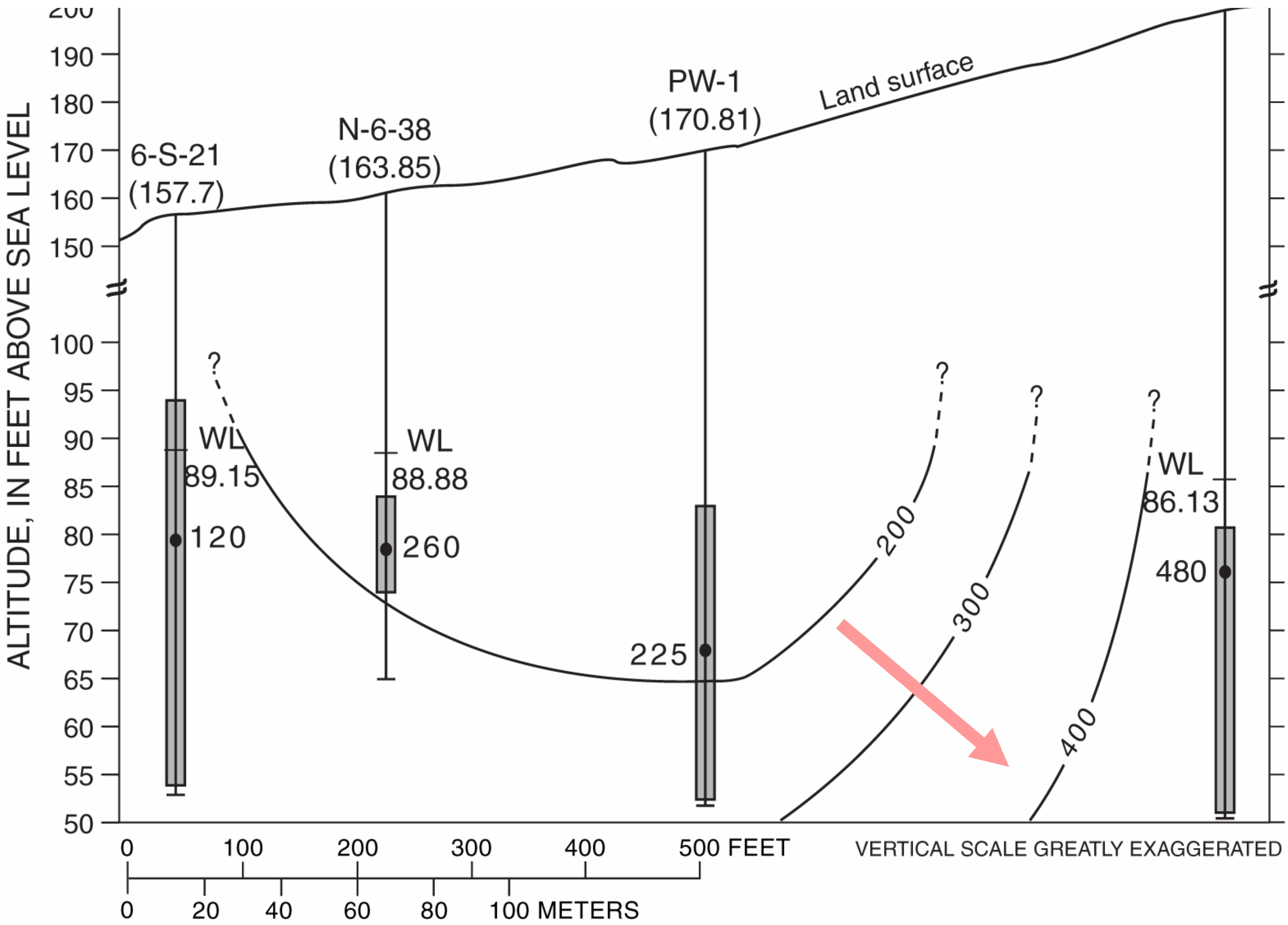
» "average" (homogenized bulk) soil Pb conc = 930 ppm (average will underestimate exposure)

1st-Generation Practices Allowed Data to be Collected without Considering the CSM

1) The possibility of different populations usually not considered
2) Sampling & analytical variables remain uncontrolled during data collection
3) Different populations can be unknowingly mixed → intermediate results
4) Data user doesn’t understand what population the data represent → misinterpret them

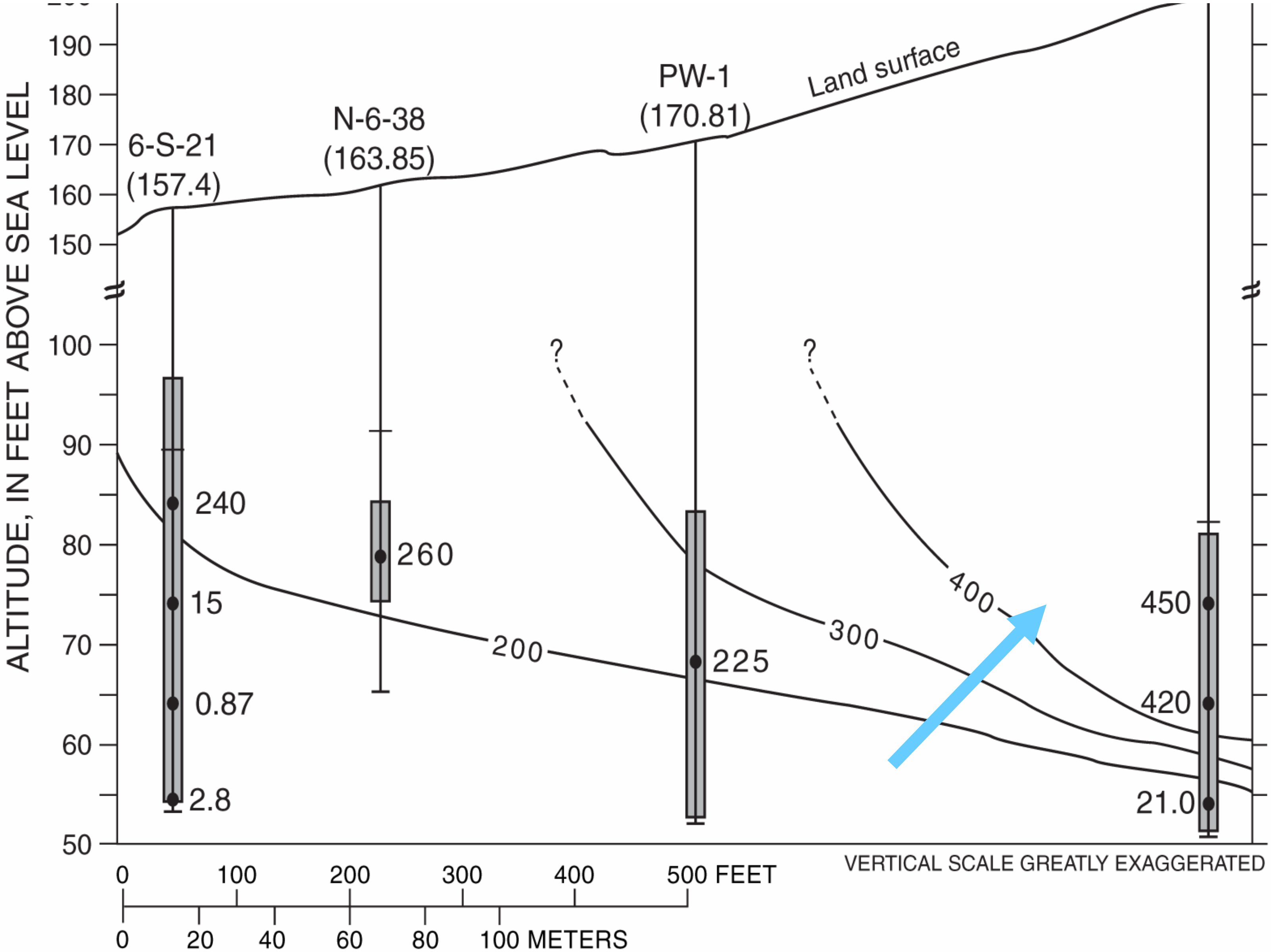
Heterogeneity Impacts Low-Flow Purge/Pumped Samples

TCA Conc. (ppb)



From USGS Report 02-4203 (2002) http://water.usgs.gov/pubs/wri/wri024203/

Same Well Field—Passive Diffusion Samplers Preserves Distinct Populations



From USGS Report 02-4203 (2002) http://water.usgs.gov/pubs/wri/wri024203/

Uncontrolled Sampling Variables Mix Different Populations to Produce Inaccurate CSMs

same well field…2 different sample collection techniques

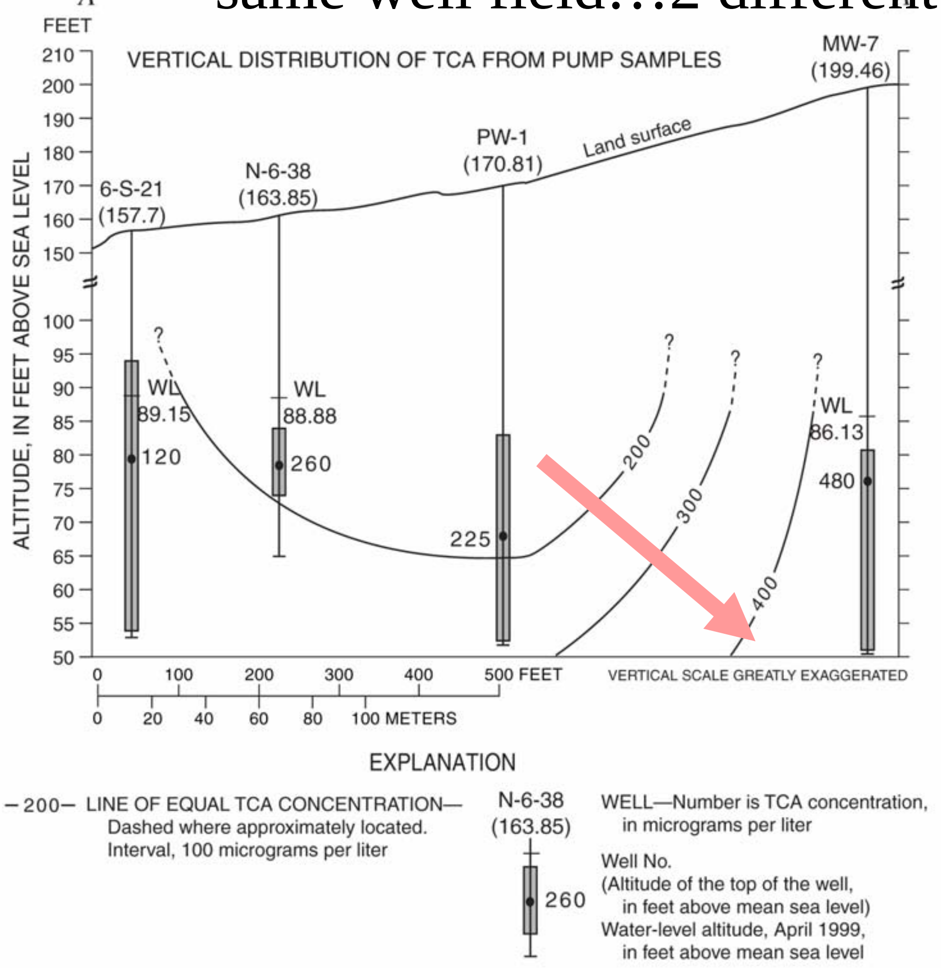


Figure 6. Vertical distribution of TCA concentrations in ground-water samples collected with the diffusion samplers and submersible pump.

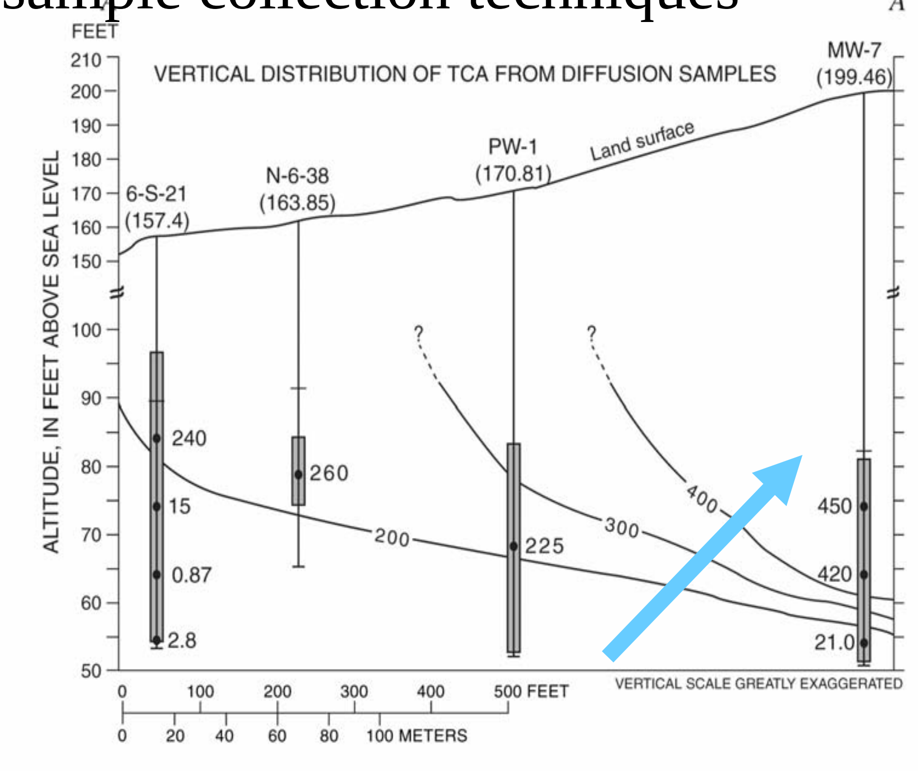


Figure 6.—*Continued.*

From USGS Report 02-4203 (2002)

http://water.usgs.gov/pubs/wri/wri024203/

No Triad project should <u>ever</u> be run without a CSM that articulates what is known or suspected about contaminant populations and environmental factors controlling fate & transport

This does NOT mean that you have to use expensive computer fate & transport models.

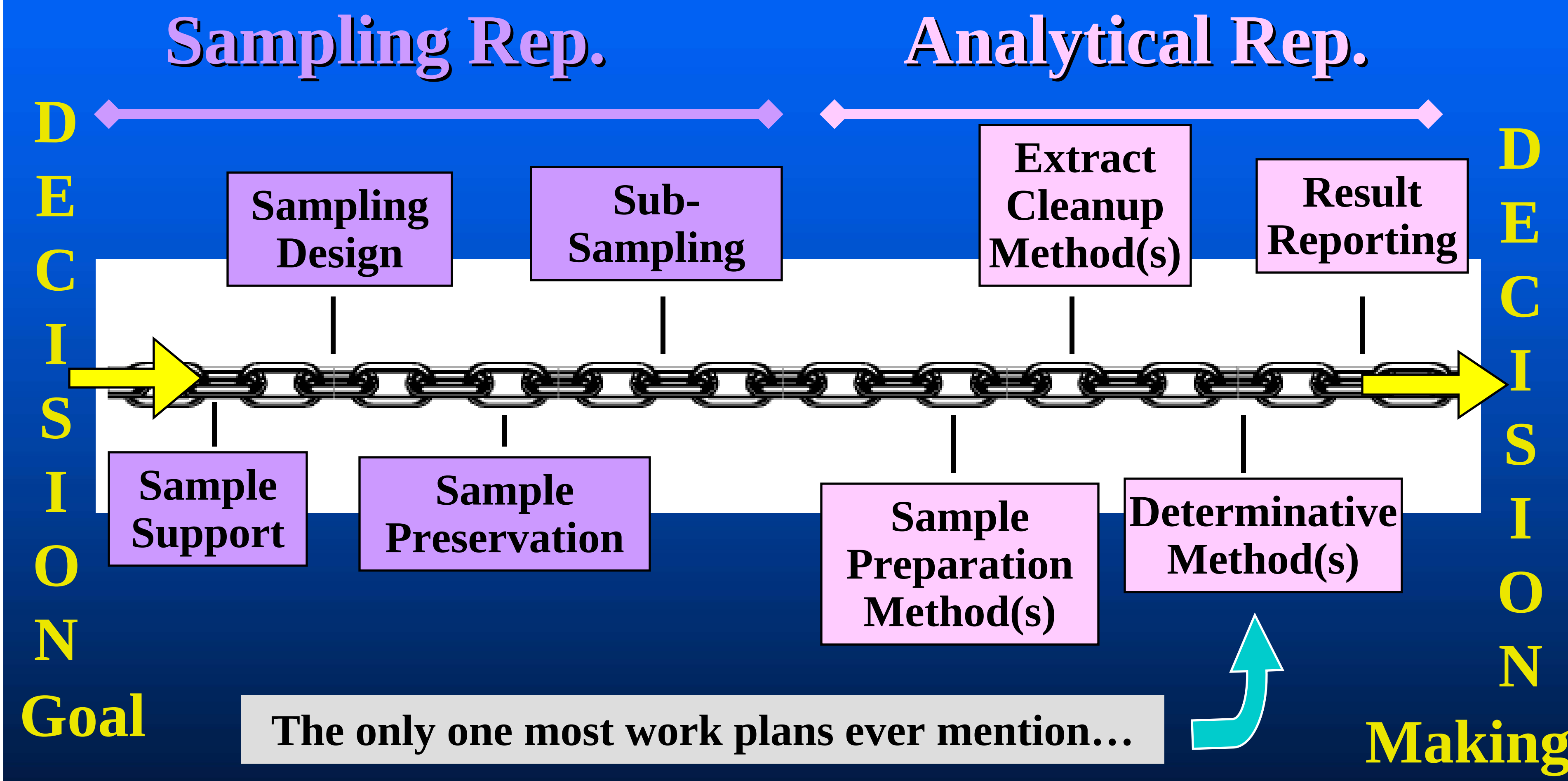
Triad's Nuts & Bolts of Data Representativeness

SAMPLING UNCERTAINTIES

Я

US

Variables that must be Controlled to Know what Population the Data Represent

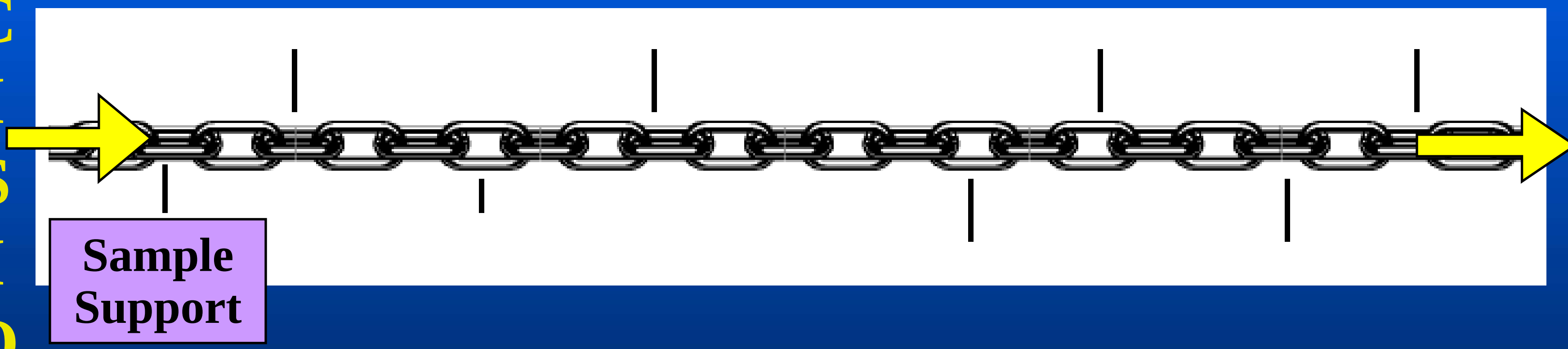


All links in the Data Quality chain must be intact for data to be representative of the decision!

The concept of sample support is critical to data representativeness, but least understood

Sampling Rep. **Analytical Rep.**

DECISION Goal



DECISION Making

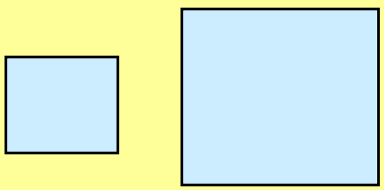
All links in the Data Quality chain must be intact for data to be representative of the decision!

Facets of “Sample Support”

- Physical properties of a sample (or subsample) that help determine what the analytical result will be
- Includes
  - Sample volume
  - Sample orientation
  - Particle size
  - Time

Sample Support: Size Matters!

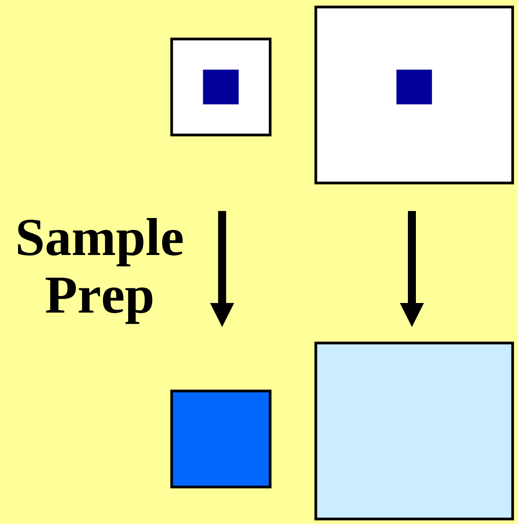
Typical regulatory and field practices assume that the size/volume of a sample has no effect on analytical results for contaminant concentrations.



That assumption doesn't hold true when environmental heterogeneity exists;

sample volume can determine the analytical result!

The Nugget Effect



Although there is the same contaminant mass in the captured nuggets, different volumes of cleaner matrix will produce different sample concentrations after sample homogenization.

Smaller supports are more variable because many contaminants behave like particulates

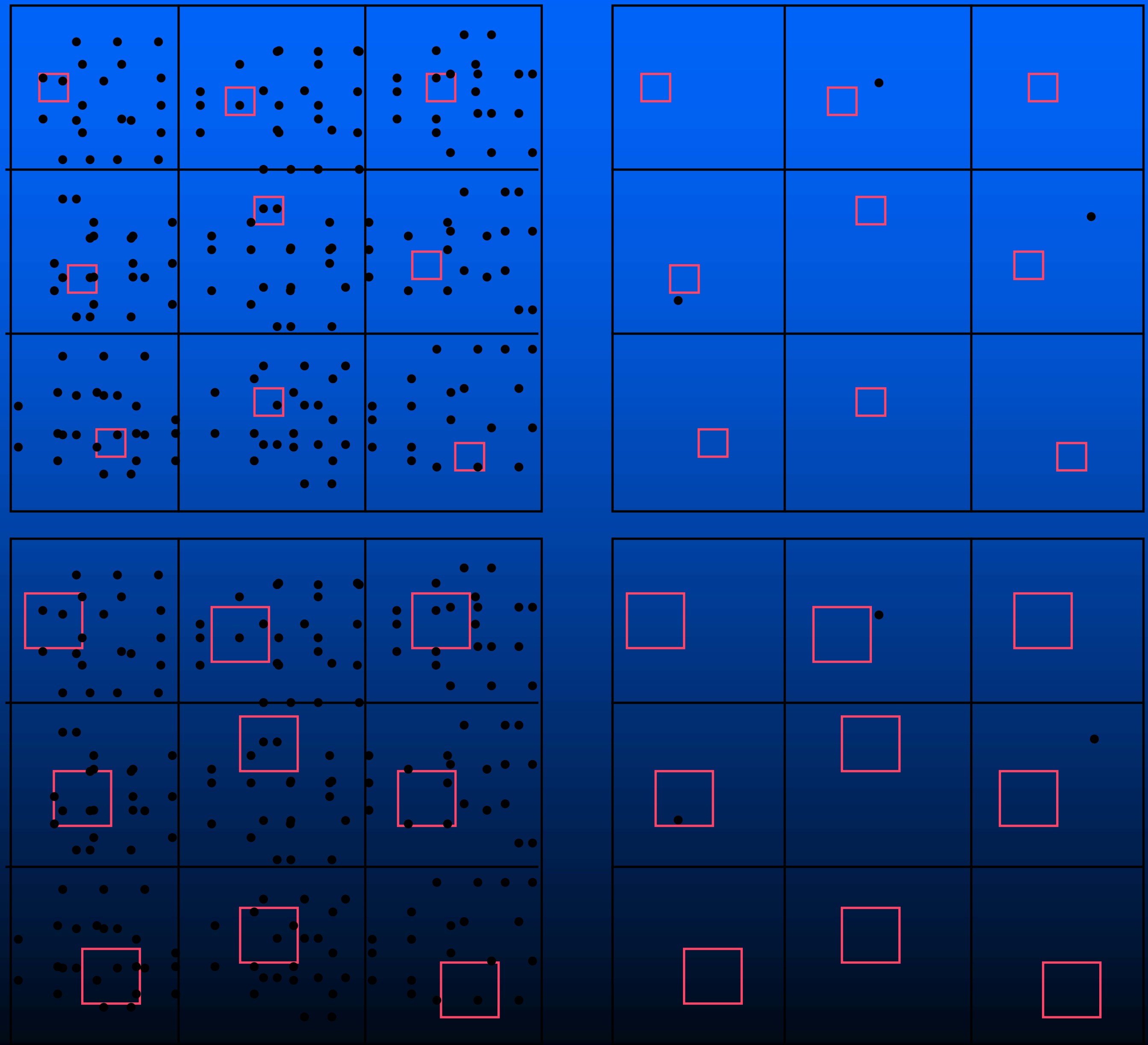


Figure adapted from Jeff Myers, 2001

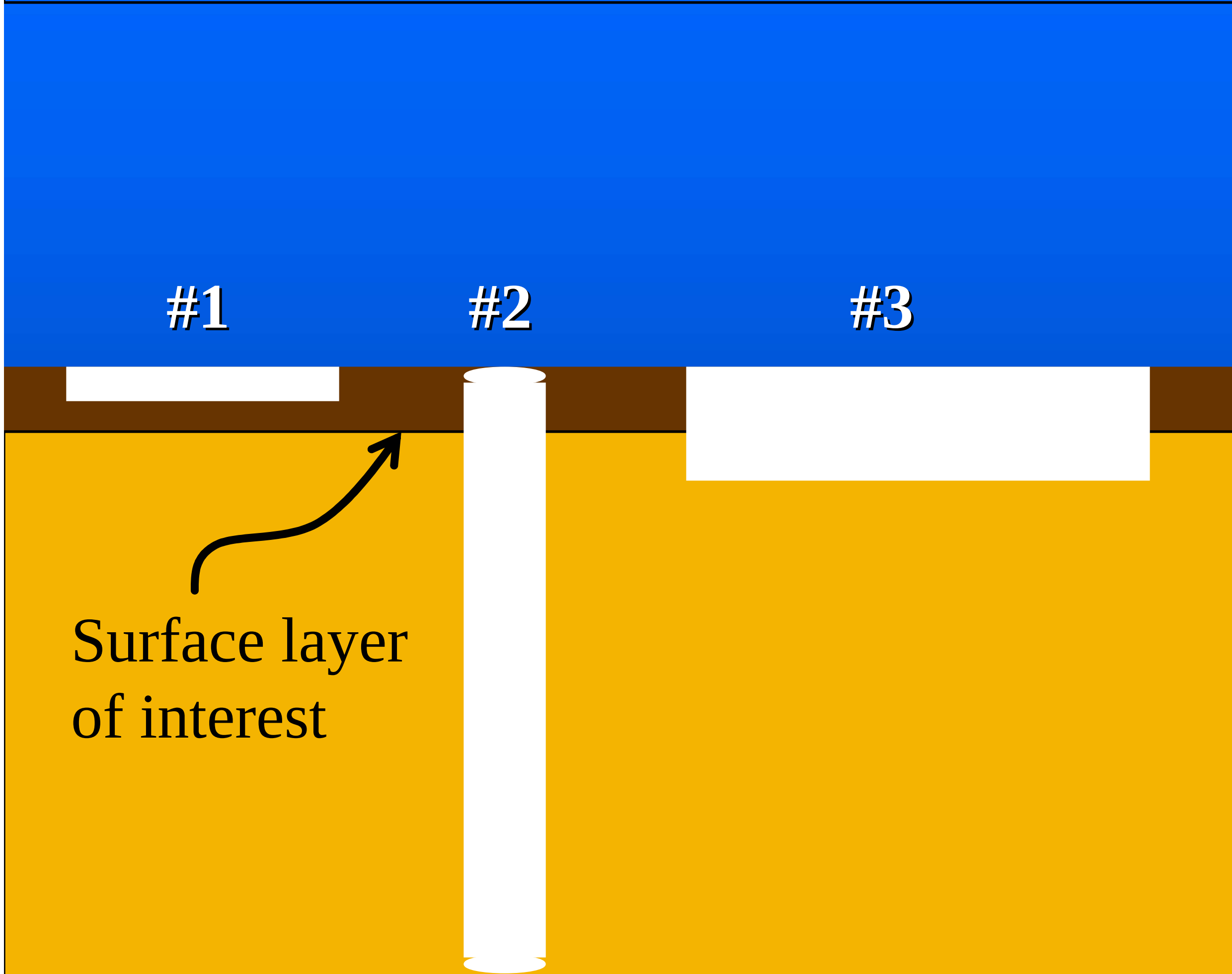
Black & Red boxes = different volume samples

Contrast different concentration and sample volume scenarios.

Left panels represent higher concentrations than right panel.

Top panels represent smaller sample supports than bottom panels

Sample Support: Includes Spatial Orientation

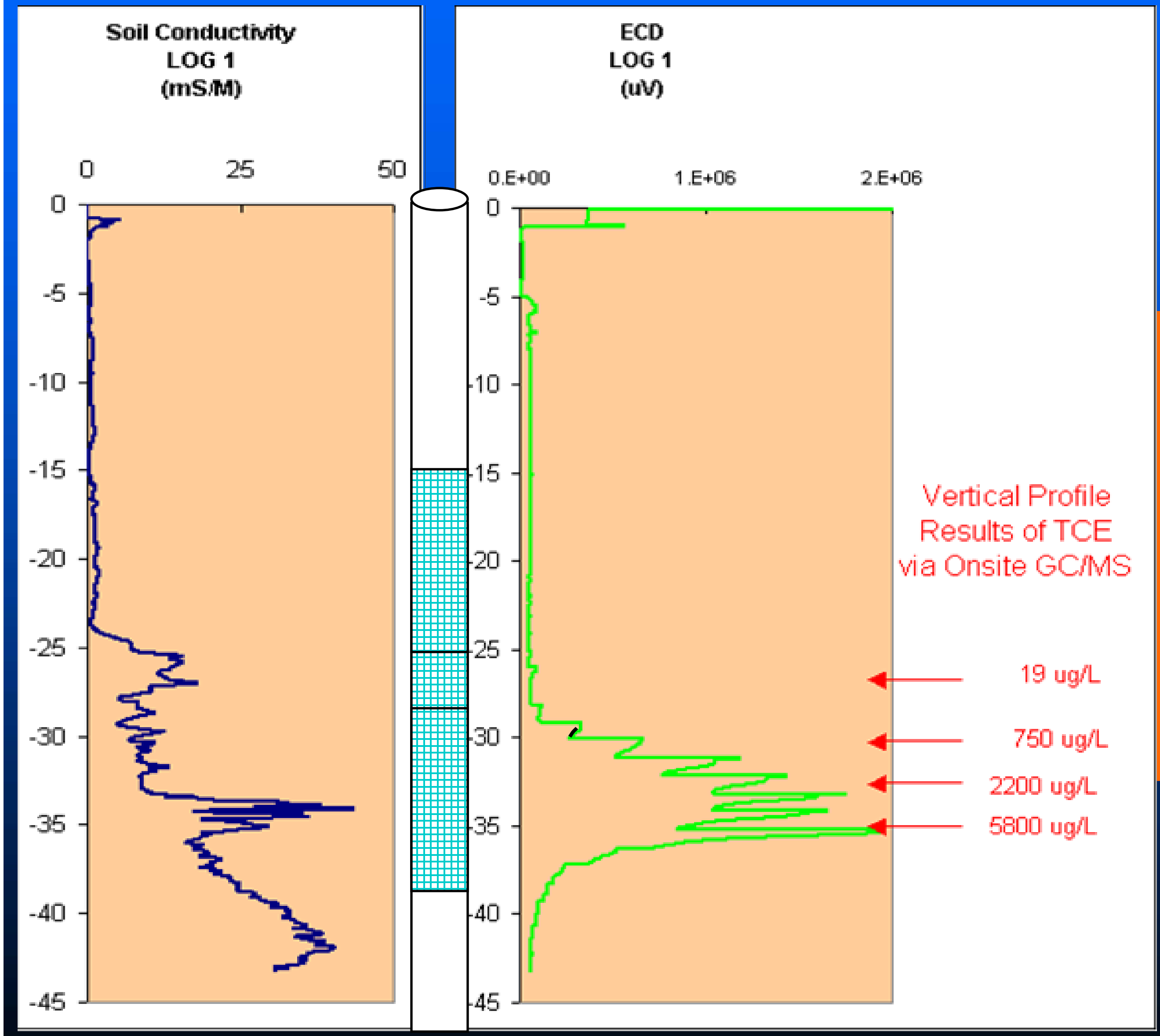


The decision driving sample collection:
Assess contamination resulting from atmospheric deposition

Given that the dark surface layer is the soil layer impacted by atmospheric deposition relevant to this project:

Which sample support (white areas #1, #2, or #3, each homogenized before analysis) provides a sample that is representative of atmospheric deposition for this site?

Different Sample Support Changes Analytical Results for GW



MIP = membrane-interface probe (w/ ECD detector)

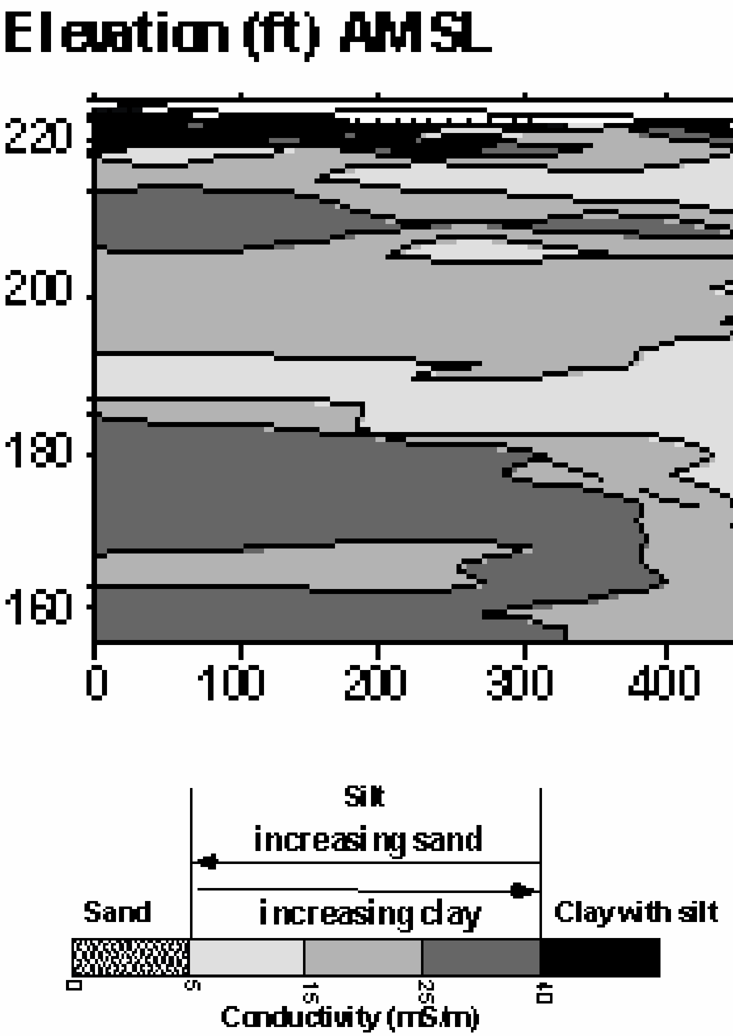
Sample support for MIP on scale of mm to inches

Sample support for discrete-depth GW samples on 6-in scale

Sample support for traditional well sampling on scale of feet

Graphic adapted from Columbia Technologies

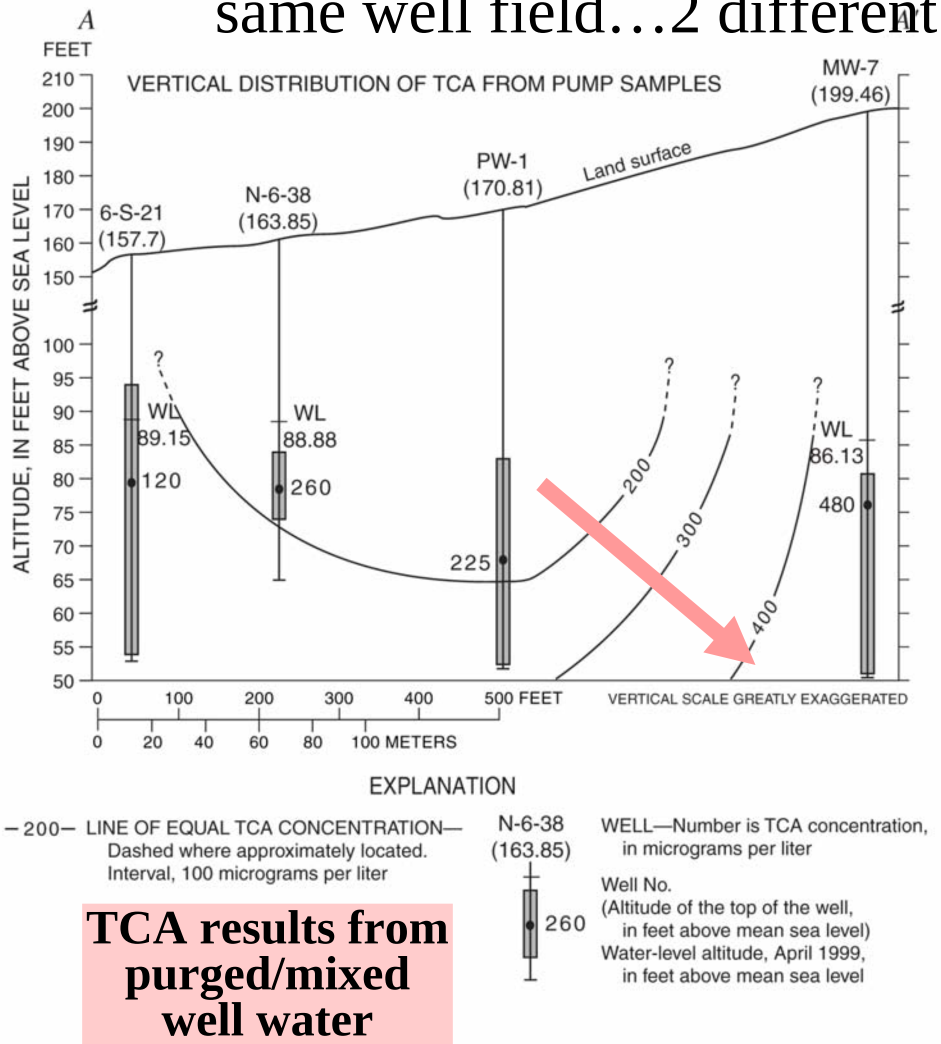
GW sample support is also a function of differential permeability of the stratigraphy



Stratigraphy determines the relative volume of water produced from different populations

Purging Creates a Different Sample Support than a Diffusion Sampler → Different CSMs

same well field…2 different sample collection techniques



TCA results from purged/mixed well water

Figure 6. Vertical distribution of TCA concentrations in ground-water samples collected with the diffusion samplers and submersible pump.

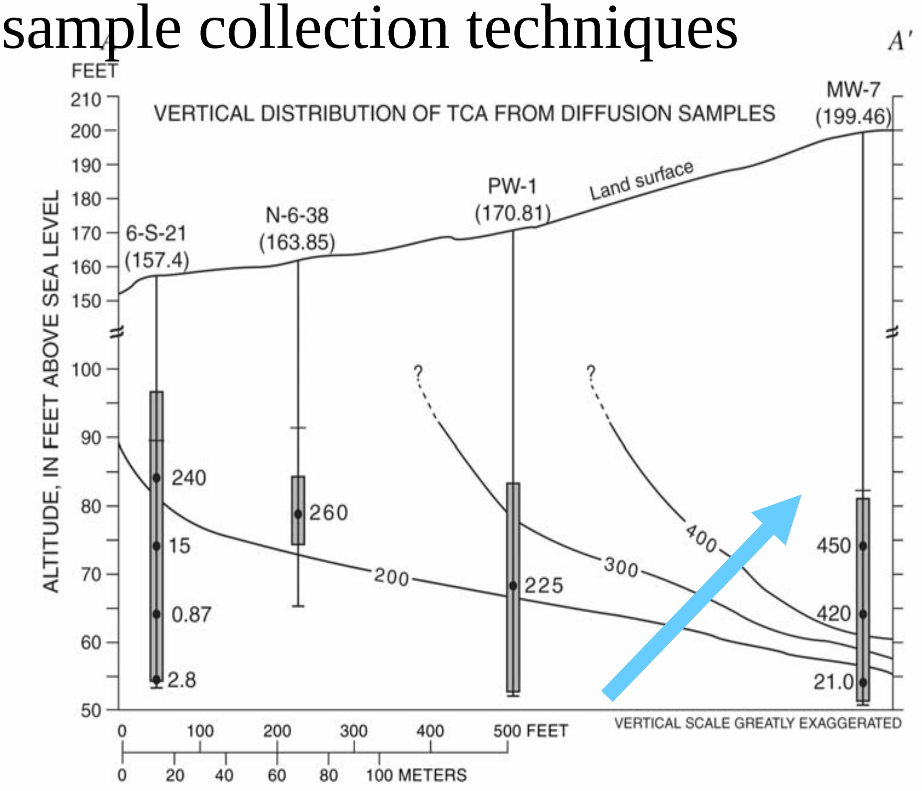


Figure 6.—*Continued.*

TCA results from depth-discrete well water

From USGS Report 02-4203 (2002)

http://water.usgs.gov/pubs/wri/wri024203/

Sample Support Can Spell the Difference Between Hits and NDs in the Same Well

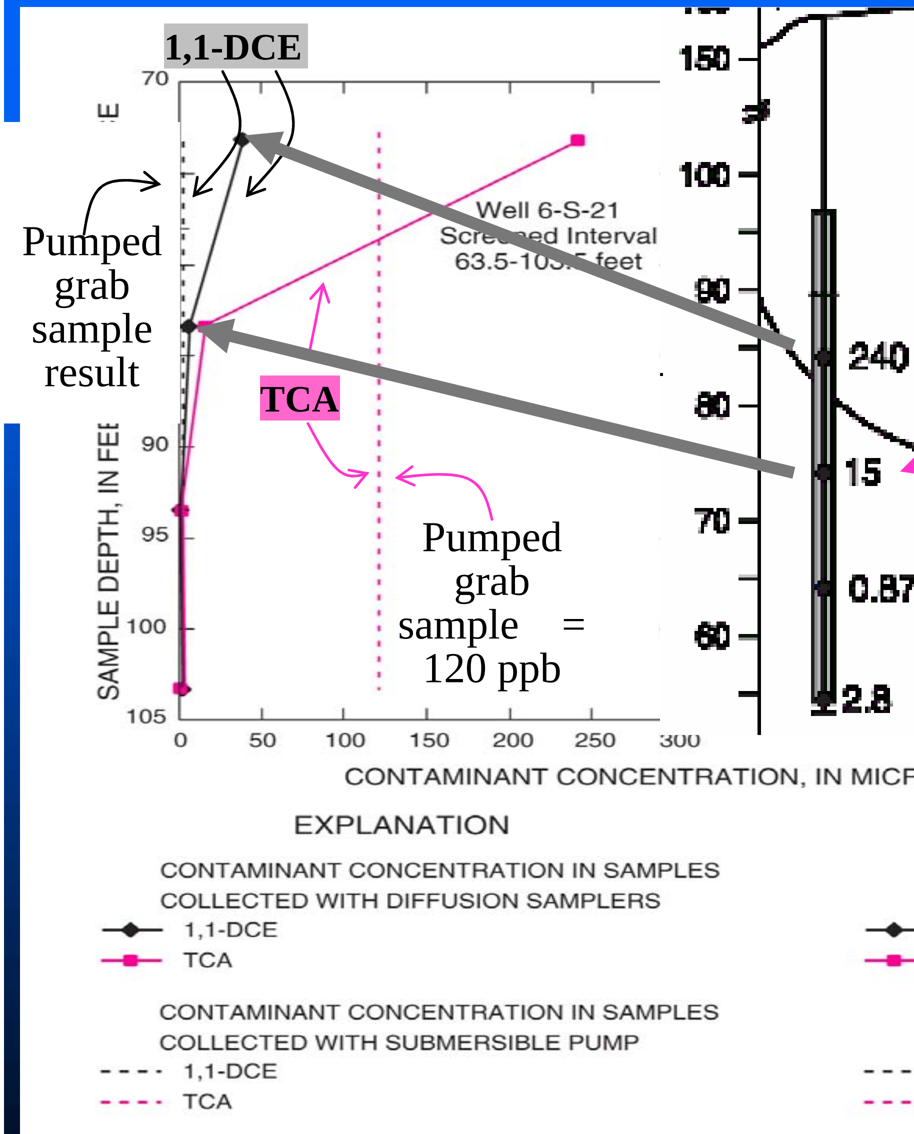


Figure 5. Comparison of selected volatile organic compound concentrations from and a submersible pump for wells with greater than 20-foot screened intervals in A

Zoom to Well 6-S-21

PDS TCA results

Vertical distribution pattern of DCE is same as TCA, but concentrations lower so that purging/mixing with cleaner water could dilute to ND, creating a misleading CSM

From USGS Report 02-4203 (2002); http://water.usgs.gov/pubs/wri/wri024203/

Different Particle Sizes Give Different Results

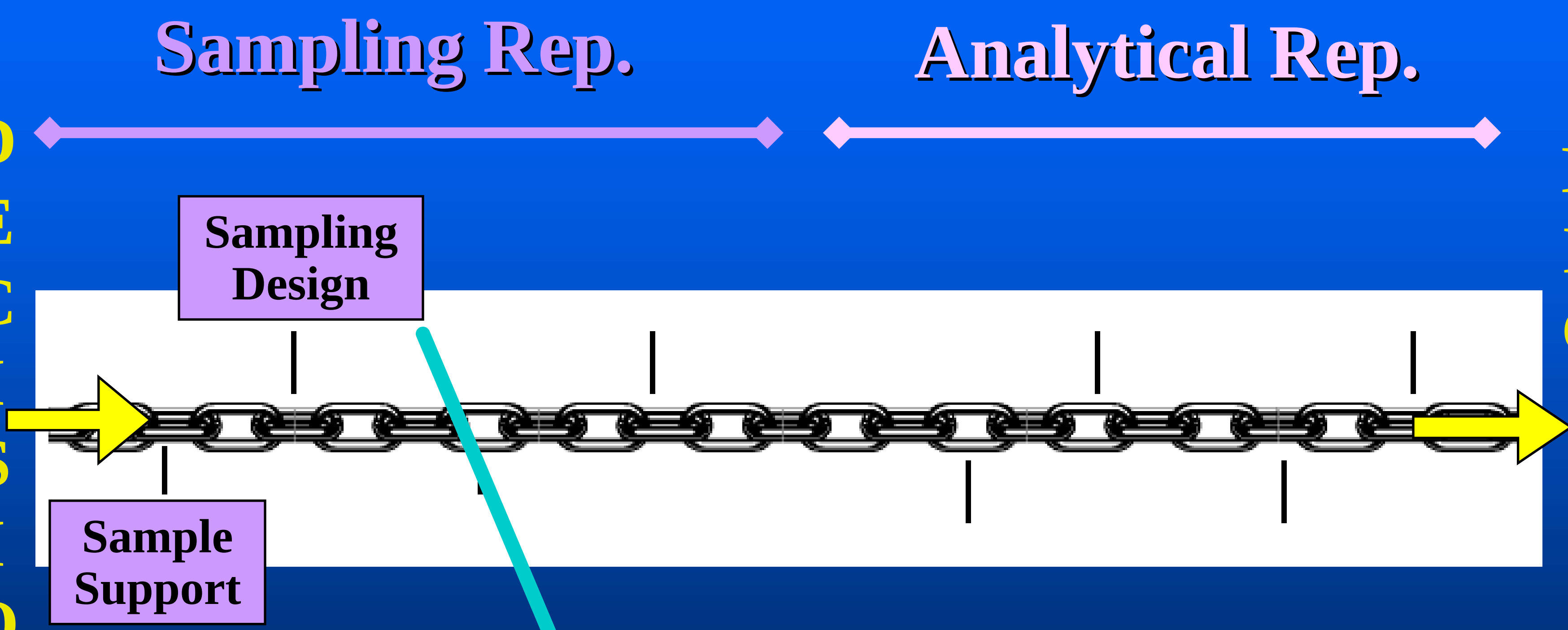
Adapted from ITRC (2003); http://www.itrcweb.org/SMART-1.pdf

Soil Grain Size (Standard Sieve Mesh Size)	Soil Fraction-ization (%)	Pb Conc. in fraction by AA (mg/kg)	Lead Distribution (% of total lead)
Greater than 3/8” (0.375”)	18.85	10	0.20
Between 3/8 and 4-mesh”	4.53	50	0.24
Between 4- and 10-mesh	3.65	108	0.43
Between 10- and 50-mesh	11.25	165	2.00
Between 50- and 200-mesh	27.80	836	25.06
Less than 200-mesh	33.92	1,970	72.07
Totals	100%	927 (wt-averaged)	100%

For this matrix, sampling/subsampling that captures larger particles will get lower results than procedures that get the smaller particles!! Cannot assume “average” is representative of the decision!

Macro Heterogeneity Affects Sampling Design

DECISION Goal



DECISION Making

e.g., number of samples, locations, grab vs. composite samples

Will Your Sampling Design Avoid Decision Errors from Misleading Grab Sampling?

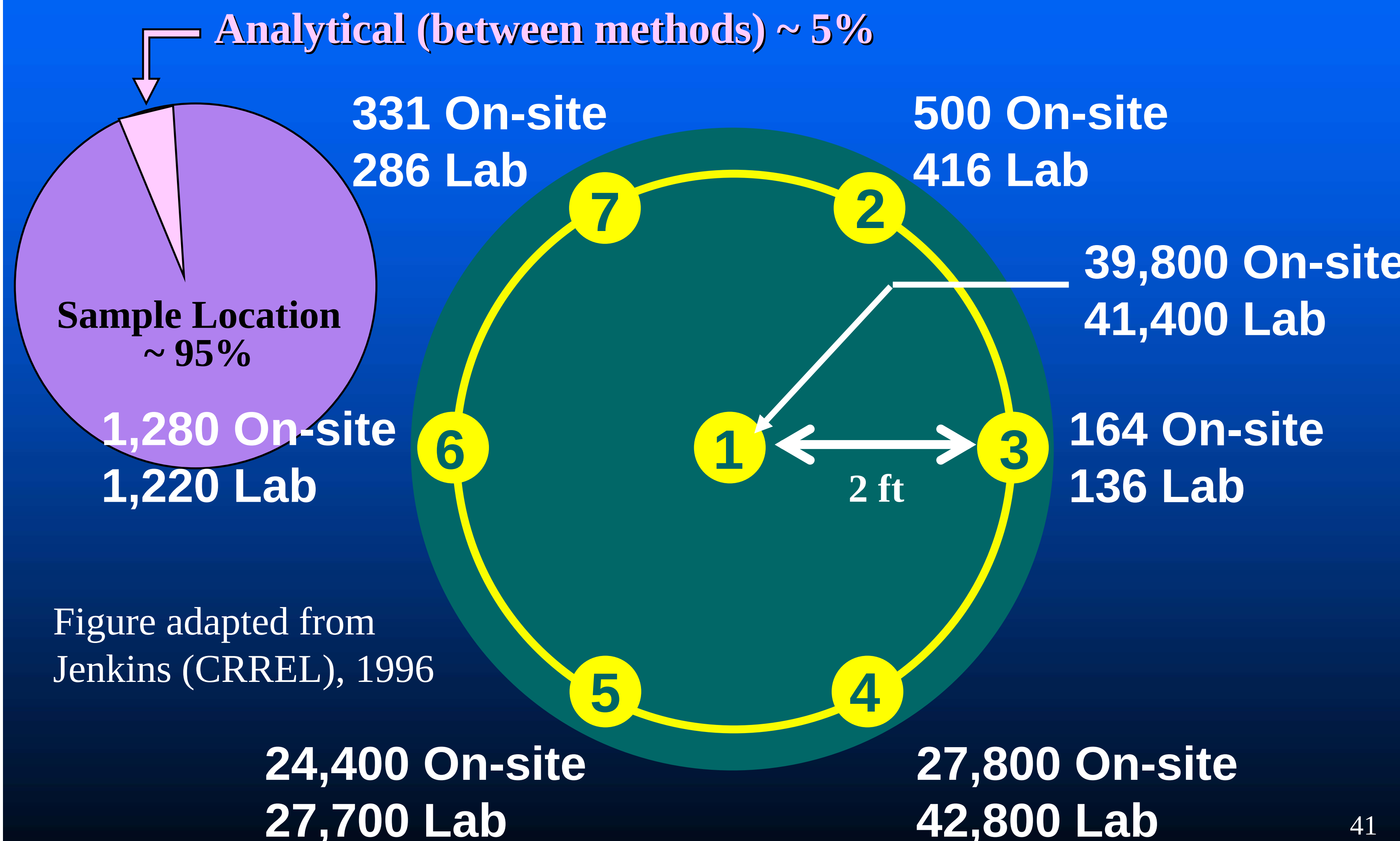
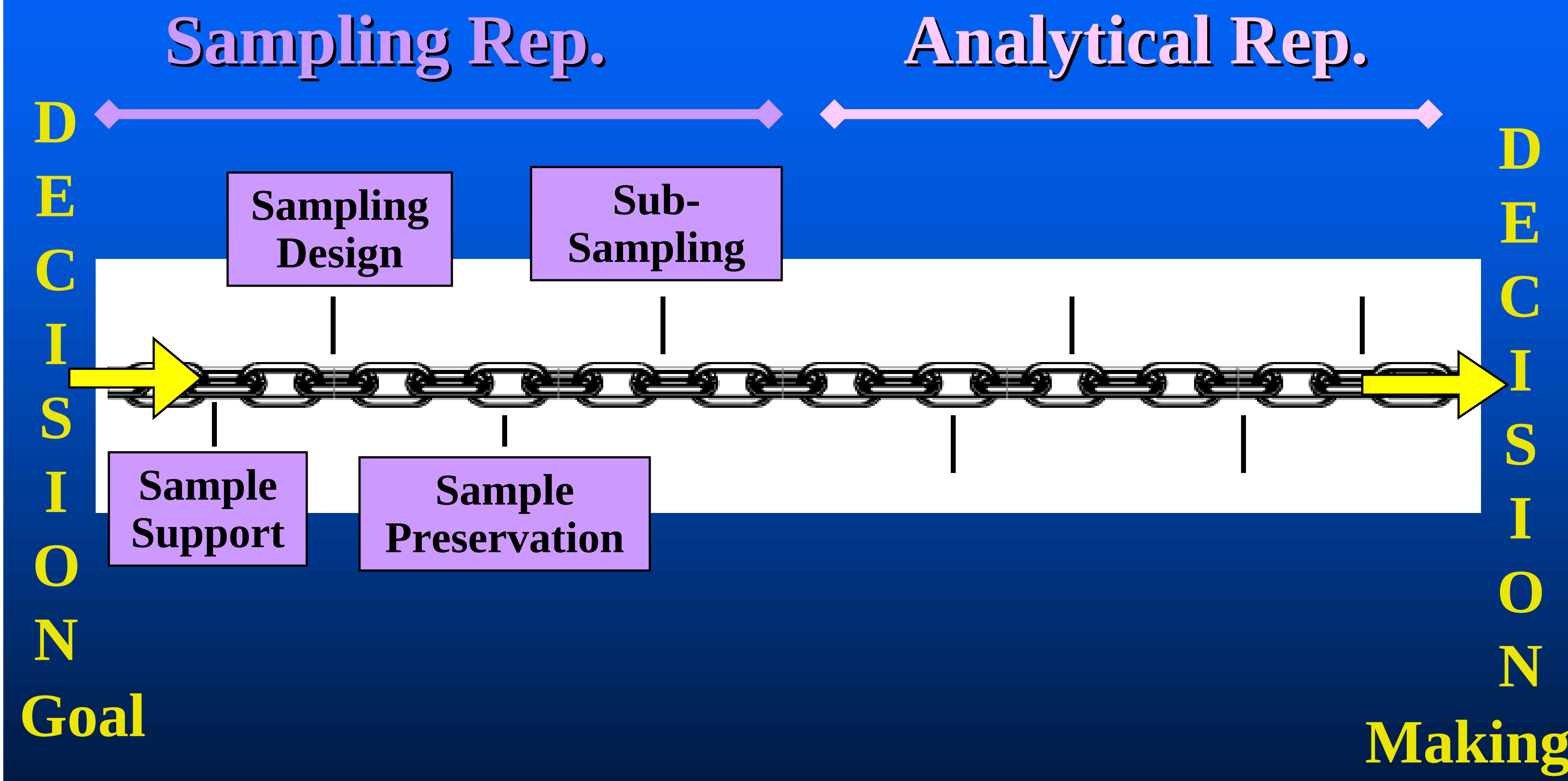


Figure adapted from
Jenkins (CRREL), 1996

Micro Heterogeneity Impacts Subsample Support



Smaller Subsamples Are More Variable
(^{241}Am in Soil Study)

Subsample Support (<u>after</u> sample was dried, ball-milled, sieved <10-mesh)	**Coefficient of Variation**	**Number of subsamples required to estimate the sample true mean ± 25% ***	**Number of subsamples required to estimate the sample true mean ± 10% ***
1 g	0.79	39	240
10 g	0.27	5	28
25 g	0.30	6	35
50 g	0.12	1	6
100 g	0.09	1	4

* Using classical parametric statistics at 95% confidence

Adapted from DOE (1978)

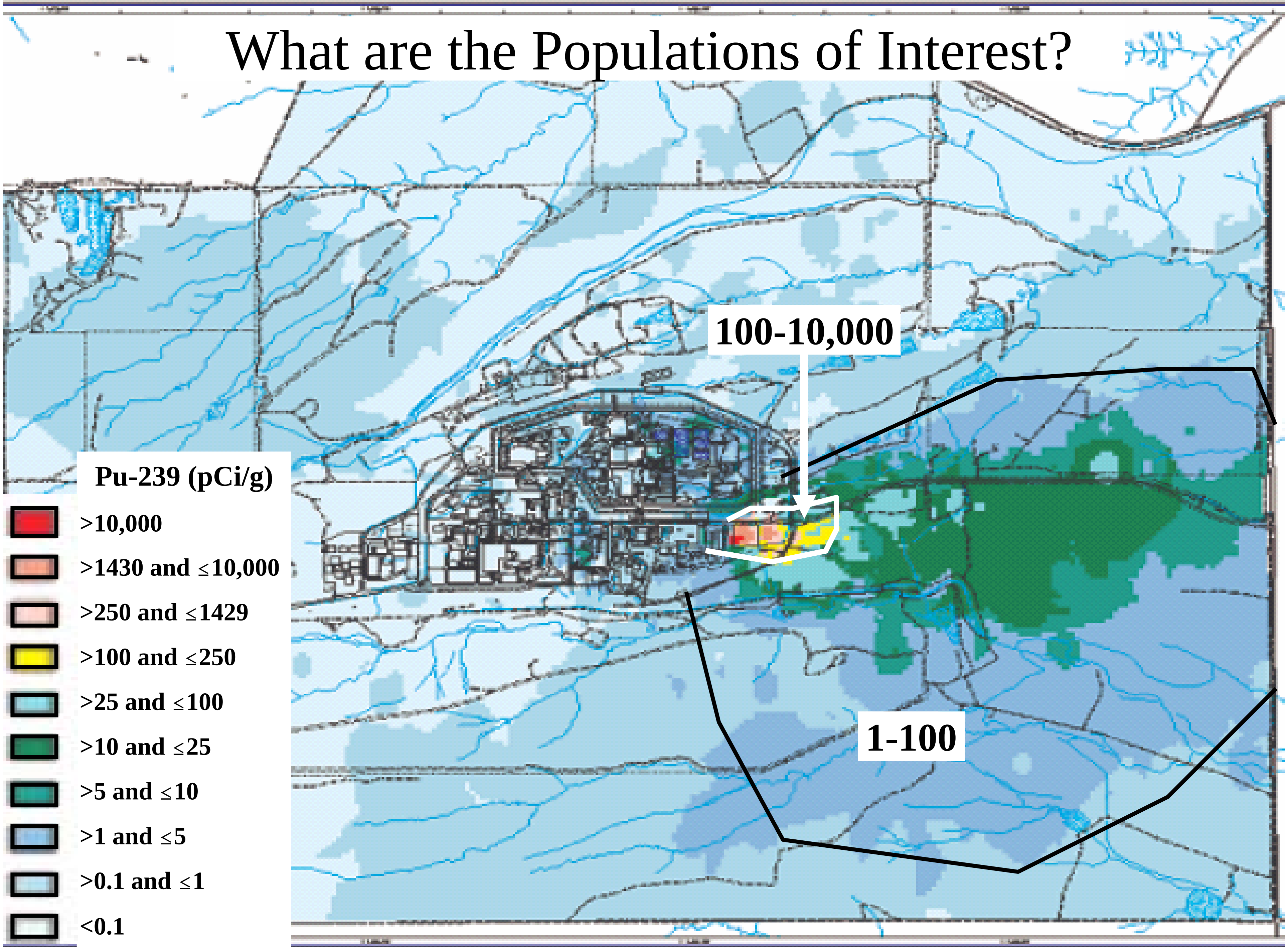
Major problem!! Advancing analytical science use smaller and smaller subsamples→ more variable results!

What is the Correct Sample Support?

- Sample support must represent or mirror the "decision (or population) support"
- **Decision/population support** = the physical characteristics of the "decision unit" (i.e., the population of interest)
  - Spatial properties of the population: 3-axis matrix dimensions, particle size
  - Time properties (if time is a variable)
- Sample collection & processing procedures must mirror these physical properties to maintain the data representativeness chain

If decision details unknown, then decision support unknown!
Then it's impossible to plan for representative data collection!

What are the Populations of Interest?



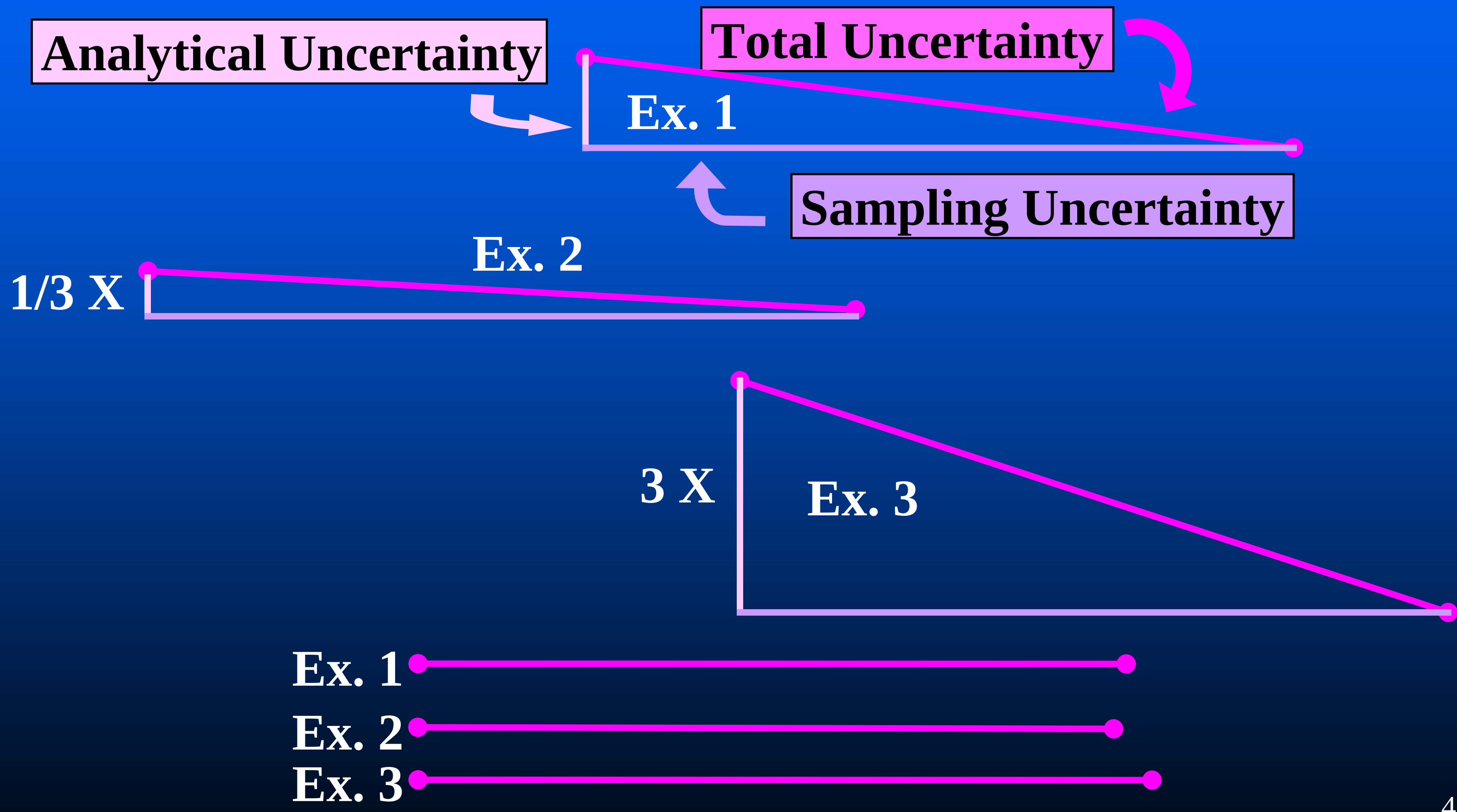
Integrating these Concepts into Practice Is What the Triad Approach is about

Systematic Project Planning

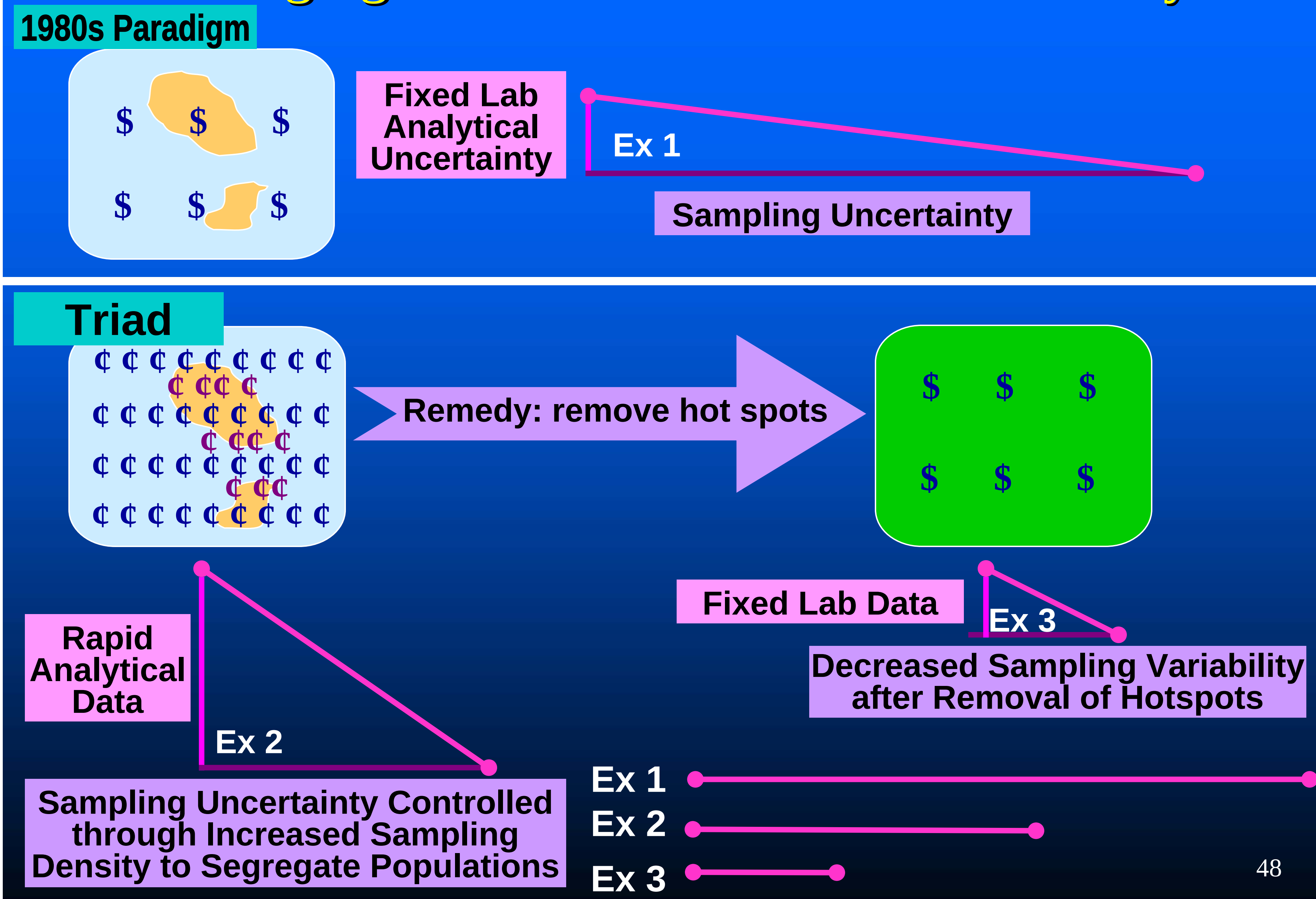
Dynamic Work Strategies

Real-time Measurement Technologies

Managing Data Uncertainty Means Managing the Components



Managing the CSM & Decision Uncertainty



Recognize Methods' Strengths & Limitations

Costly standard analytical methods

Low DL + analyte specificity

Manages analytical uncertainty
(analytical quality)

"Definitive" analytical quality
Screening sampling quality

Screening Quality Data

Cheaper, rapid analytical methods

High spatial density

Manages sampling uncertainty
(sampling representativeness)

Builds CSM

"Definitive" sampling quality
Maybe screening analytical quality

Updating the Data Quality Model to Cope with Heterogeneous Matrices

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

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

Collaborative data sets complement each other so that all sources of data uncertainty important to the decision are managed

Triad Projects Use Demonstrations of Methods Applicability

- A “pilot study” that helps to optimize tool selection and technical operations (both field tools & off-site analytics)
- “Kills many birds with 1 stone” when designed thoughtfully (see handouts)
- Critical if want realistic split sample comparisons

QC is an Important Triad Component

- Goal is to match project-specific QA/QC protocols for both field <u>and</u> fixed lab methods to <u>intended data use</u> to manage <u>decision uncertainty</u>.
- Difficult to achieve when oversight is checklist oriented.
- Purpose of QC is evaluate & demonstrate control over data generation variables
- Most powerful QC check of all = real-time evaluation of compatibility between data results and the CSM