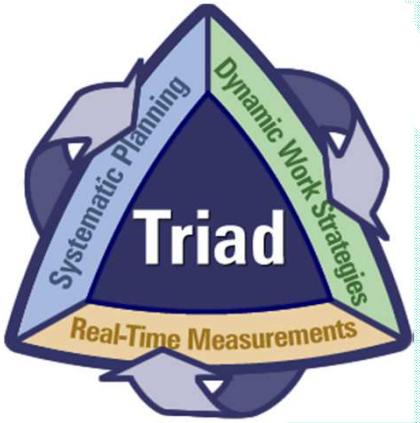
Module 3

Basic Terms and Concepts



"To be [averaged]
or not to be [averaged],
that is the question…"

From William Shakespeare's Hamlet, Prince of Denmark, Act III, Scene I, as translated by course instructor

The Triad Approach

Systematic Project Planning

Dynamic Work Plan Strategy

Real-time Measurement Technologies

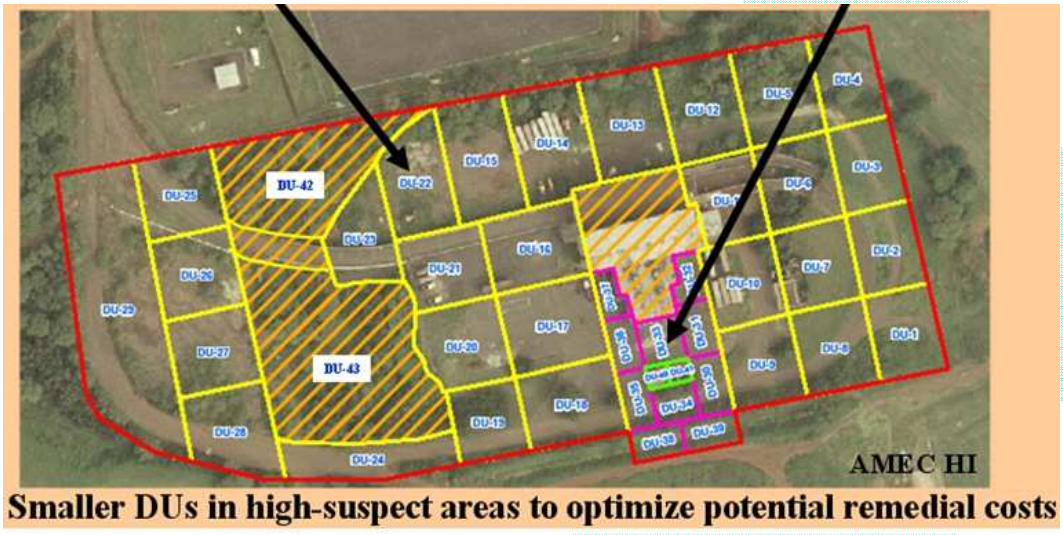
Basic Concepts

- **Decision Unit (DU):** the volume of soil or set of objects treated as **single unit** for decision-making
  - Such as ¼-acre area to 2-inch depth, a bin of soil, a set of drums
  - Examples: exposure units, survey units, remediation units…
- **Population:** Set of objects or material volumes sharing a common characteristic
  - Can be same object set as DU, but doesn't need to be
  - A fuzzy concept when referring to soils and water

Example of Setting Decision Units (DUs)

Future residential lots, DUs sized as exposure areas (EUs)

Pesticide mixing area, DUs sized to assist remediation



Smaller DUs in high-suspect areas to optimize potential remedial costs

More Basic Concepts

- **Sample:** A portion of a population/decision unit collected to characterize a population/decision unit parameter of interest
  - Disconnect between the definition of a sample in statistics and the way the term is used in hazardous waste site characterization.
- **Sample Support:** Physical dimensions and characteristics of a (sub)sample
  - Well-defined concept in other settings such as human studies, much more ambiguous with respect to soils and water.

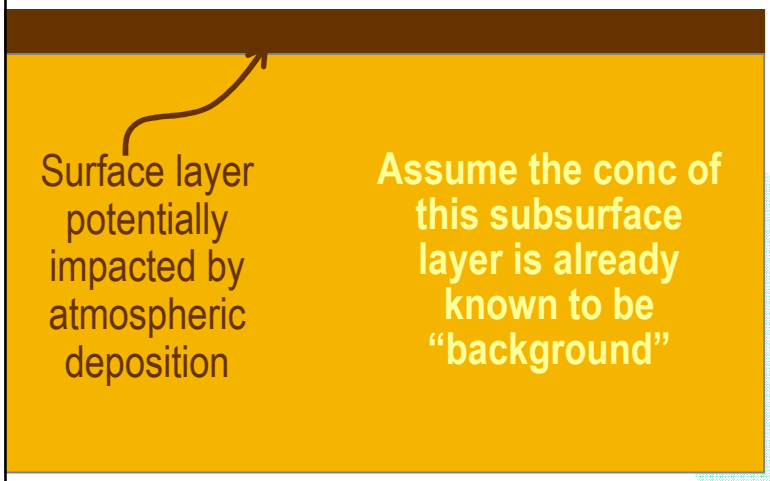
More Concepts

- **Representativeness:**
  - Degree to which a sample reflects original population in context of decision
  - Ability to confidently extrapolate concentration results from a tiny sample to represent the concentration of the much larger volume of soil (area of inference) from whence it came
  - Example:
    - MARSSIM FSS unit – more than 300 metric tons of soil
    - Typical discrete sample – 400 grams of soil
    - Typical alpha spectroscopy sub-sample – a few grams of soil

Illustration of Sample Support and Representativeness

This cartoon is ONLY to illustrate a concept: don't make it overly complex!

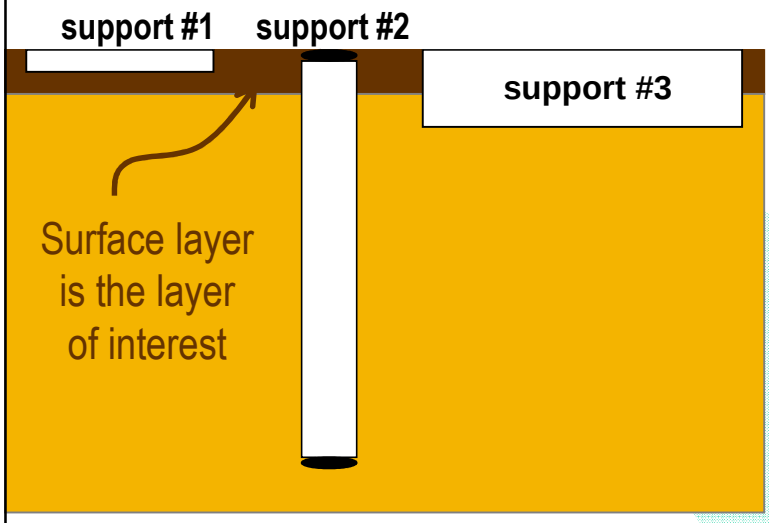
Soil Cross-Section



Has the surface soil layer been impacted by atmospheric deposition?

Which of these sample supports is <u>most</u> representative of the soils of interest?

All soil within the sample support outline (white area) is considered to be a single sample. The sample is *thoroughly* homogenized before analysis.



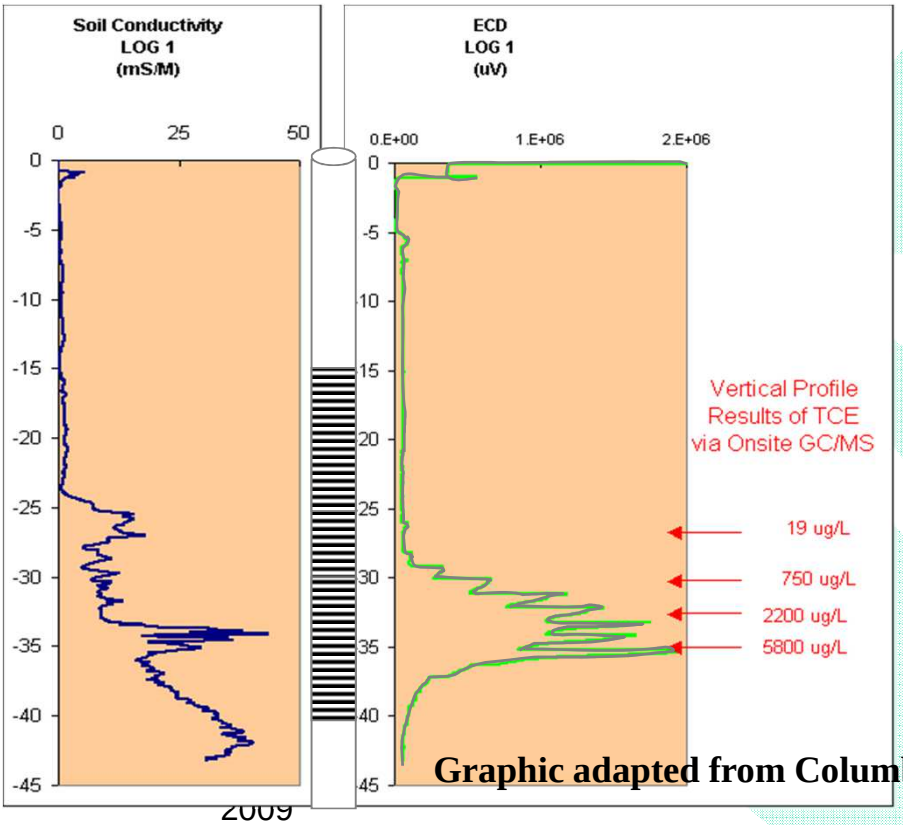
<u>Answer:</u> Support #1.

Why?: This sample contains only soil from the layer of interest.

BUT…

How could sample support #1 be modified to *better* represent the layer of interest?

Advances in Sampling & Measurement Technologies Highlight **Representativeness Issues**



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

GW data results HIGHLY dependent on sample support

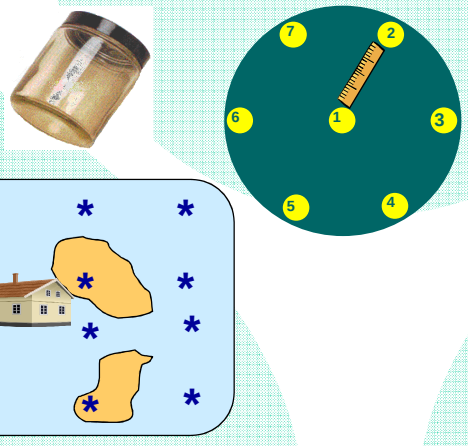
Graphic adapted from Columbia Technologies

10

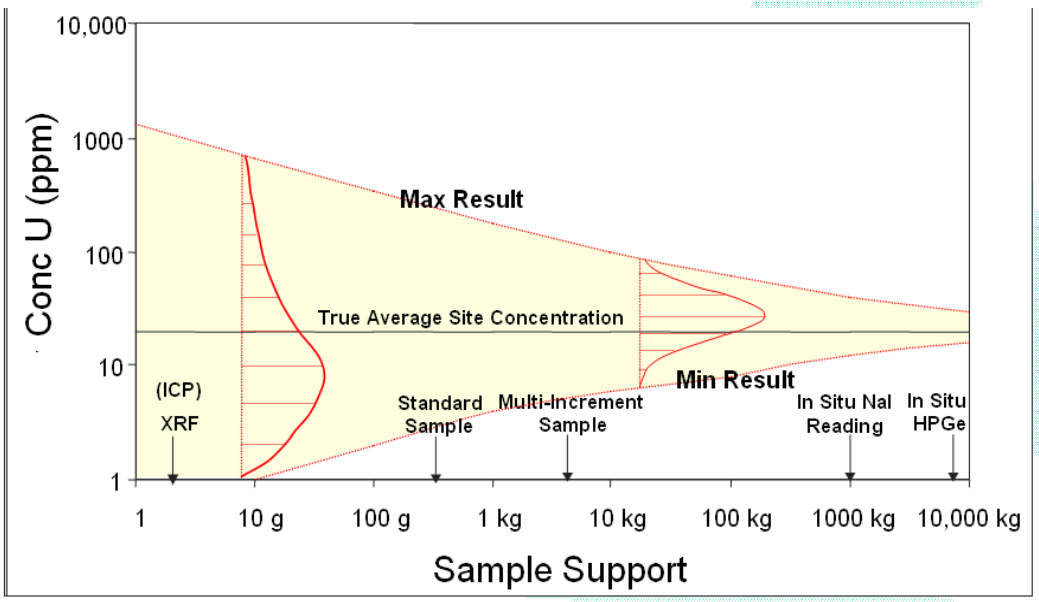
2009

Heterogeneity/Variability

- **Heterogeneity:** Variations in the value of a parameter throughout an area or volume (typically as observed in sample results)
- **Variability:** Variations in measured concentrations observed in (sub)sample results
  - **Within-sample** heterogeneity
  - **Short-scale between-sample** heterogeneity (can affect agreement between co-located samples)
  - **Long-scale between-sample** heterogeneity (on scale of conventional distances between samples)



Sample Support Determines the Degree of Variability Observed Among Samples from a Decision Unit



Basic Statistical Terms

- **Mean:** Average concentration for a given decision unit
- **Median:** Concentration at which half of a decision unit would be below and half above
- **Range:** Concentration interval defined by the minimum and maximum concentration values
- **Variance:** A measure of the "spread" of concentration values for a set of samples or measurements
- **Standard Deviation:** Square root of the variance

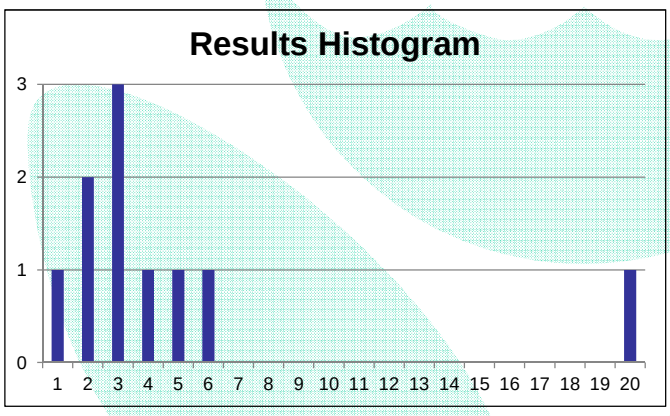
Easy Quiz

A Decision Unit was sampled 10 times systematically, with the following results:

2, 6, 3, 4, 20, 3, 5, 3, 1, and 2 ppm.

What is the:

- Mean?
- Median?
- Range?
- Standard Dev?



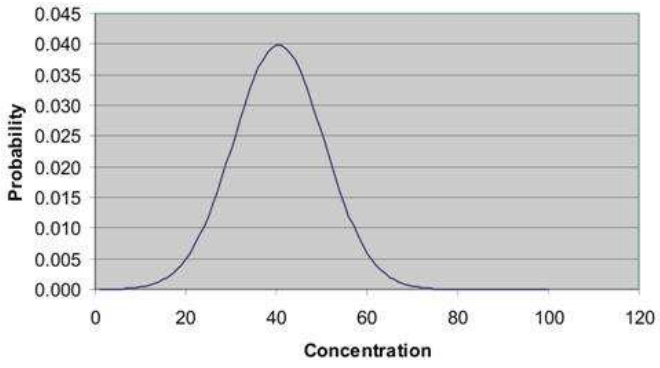
Sampling Unit (SU)

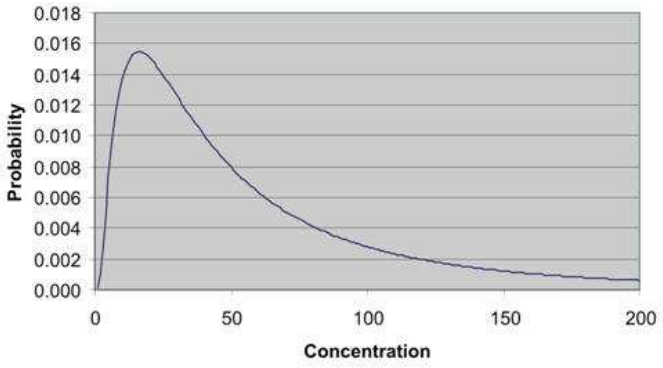
Sampling Unit = the volume of soil represented by a sample and its data result; the "area of inference" for a sample

- If 10 individual samples for analysis are taken from Area A, then there are 10 SUs in Area A
- If a single sample for analysis is taken for Area A, then there is 1 SU for Area A (Area A = the SU)
- If 10 individual samples are systematically taken from Area A and used to form one composite that is analyzed, there is 1 SU for Area A

A Good Time to Introduce Common Statistical Data Distributions

- Normal (Gaussian) distribution:
  - Can take on any (+ or -) value
  - Symmetric
- Lognormal distribution:
  - No negative values
  - Skewed to right





More Basic Statistical Terms

- **Coefficient of Variation (CV):** ratio of standard deviation to mean, a measure of relative variability
  - also called relative standard deviation (RSD)
- **Skewness:** the degree to which one end of a statistical distribution is pulled out to one side
- **Confidence Interval:** range of values that estimates the uncertainty around a point estimate, such as a data result or a mean.
  - the true value is expected to be somewhere within that range some given amount (such as 95%) of the time

Using Confidence Intervals to Communicate Uncertainty

- **Confidence level** is usually set at **95%** by default, but others can be used
- The values at each end of the interval are called the **confidence limits**: lower (LCL) and upper (UCL)

- Values between the confidence limits make up the **confidence interval** around the mean
- Width of confidence interval driven by: the confidence level, variability present in the data, assumptions about underlying data distribution & number of data points (n).

UCL Value Applications

- Used as conservative estimates of the mean concentration for risk assessments (exposure point concentrations)
- Used as points of comparison for determining whether mean values within decision units have achieved clean up criteria that are averaged

What Drives UCL Values?

- The number of samples contributing to the mean estimate
- The variability observed in the sample data
- The confidence level required
- The assumed underlying data distribution
- ProUCL can be used to determine the appropriate UCL value to use for a data set

Back to Easy Data

- Results Mean: 4.9 ppm
- 95% UCL on estimated mean using ProUCL:
  - Assuming normality: 8.1 ppm
  - Assuming lognormal: 9.9 ppm
  - Assuming gamma: 8.5 ppm
  - Assuming nonparametric: 7.6 to 22.2 ppm

Any Questions?