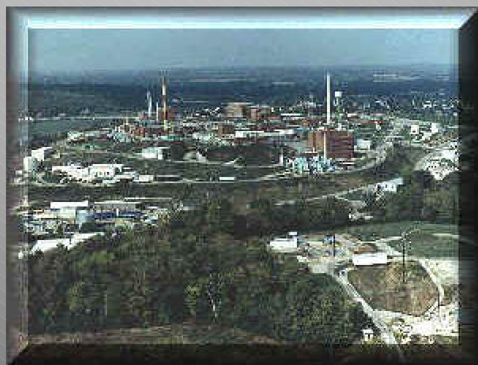
Smart Sampling

Section 1

Case Study: Mound PRS-379

Smart Sampling

Mound



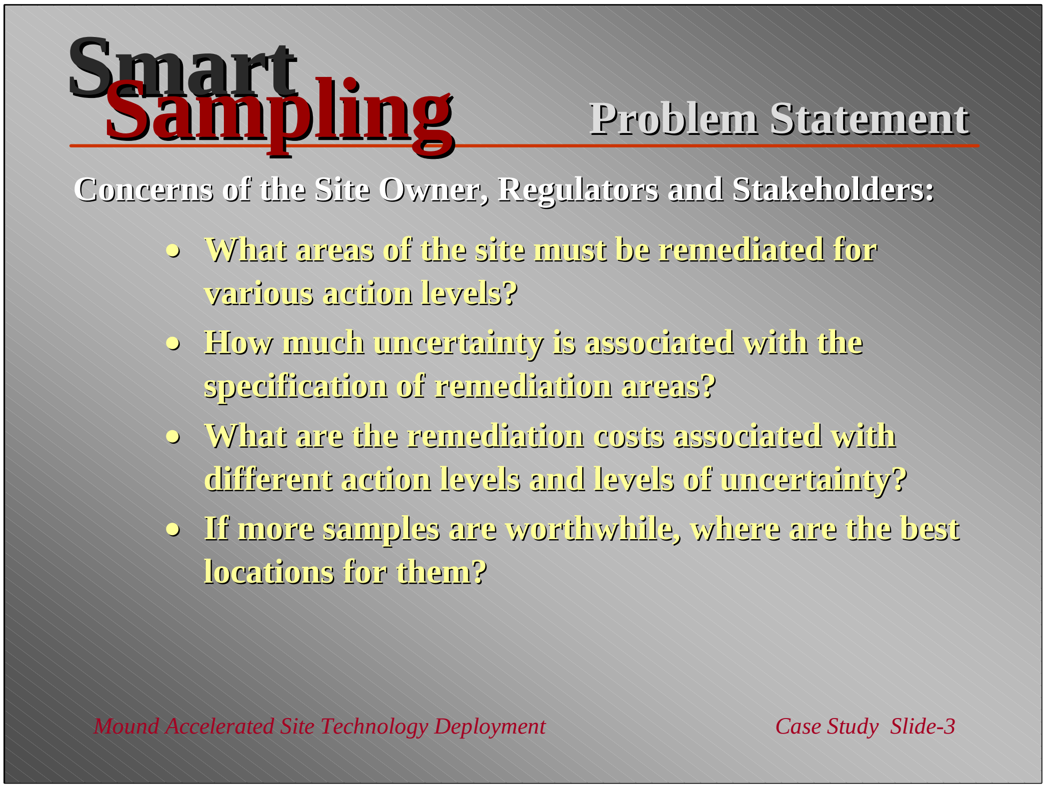
The Mound site is a former DOE Plutonium processing plant

Final goal: Release portions of the site back to the public for use as parks, industrial sites, or residential areas

PRS-379 has been sampled and remediation decisions need to be made.

Mound Accelerated Site Technology Deployment

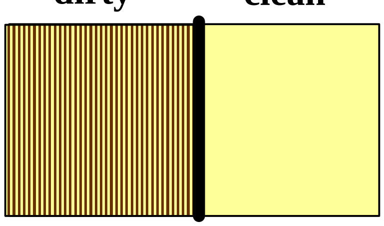
Case Study Slide-2



SmartSampling employs a probabilistic approach rather than the typical deterministic methodology. Proposed remediation maps are generated for specific action levels and a % probability of failing to remove all "dirty" soil.

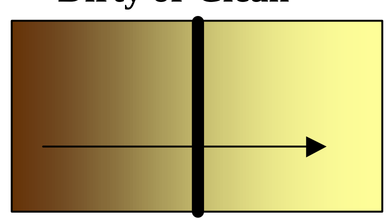
Deterministic

dirty **clean**



Probabilistic

More to Less Probability of being Dirty or Clean

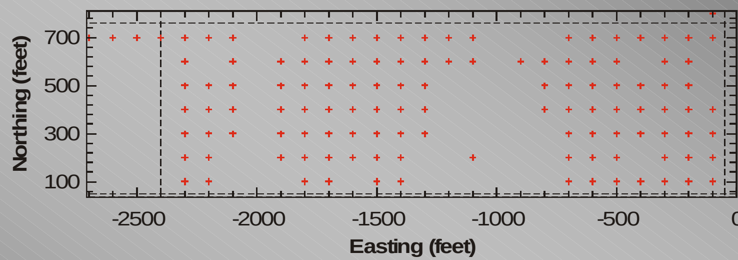


Sites frequently want to look at multiple action levels, each of which has different economic consequences.

SmartSampling generates graphic displays of cost consequences for different levels of uncertainty as well as different action levels.

Smart Sampling

PRS-379 Initial Samples



- The domain @2400 ft by 800 ft
- Contaminant of concern is Pu-238
- 127 samples distributed on a grid with 100 ft spacing
- 10 percent of data are non-detect

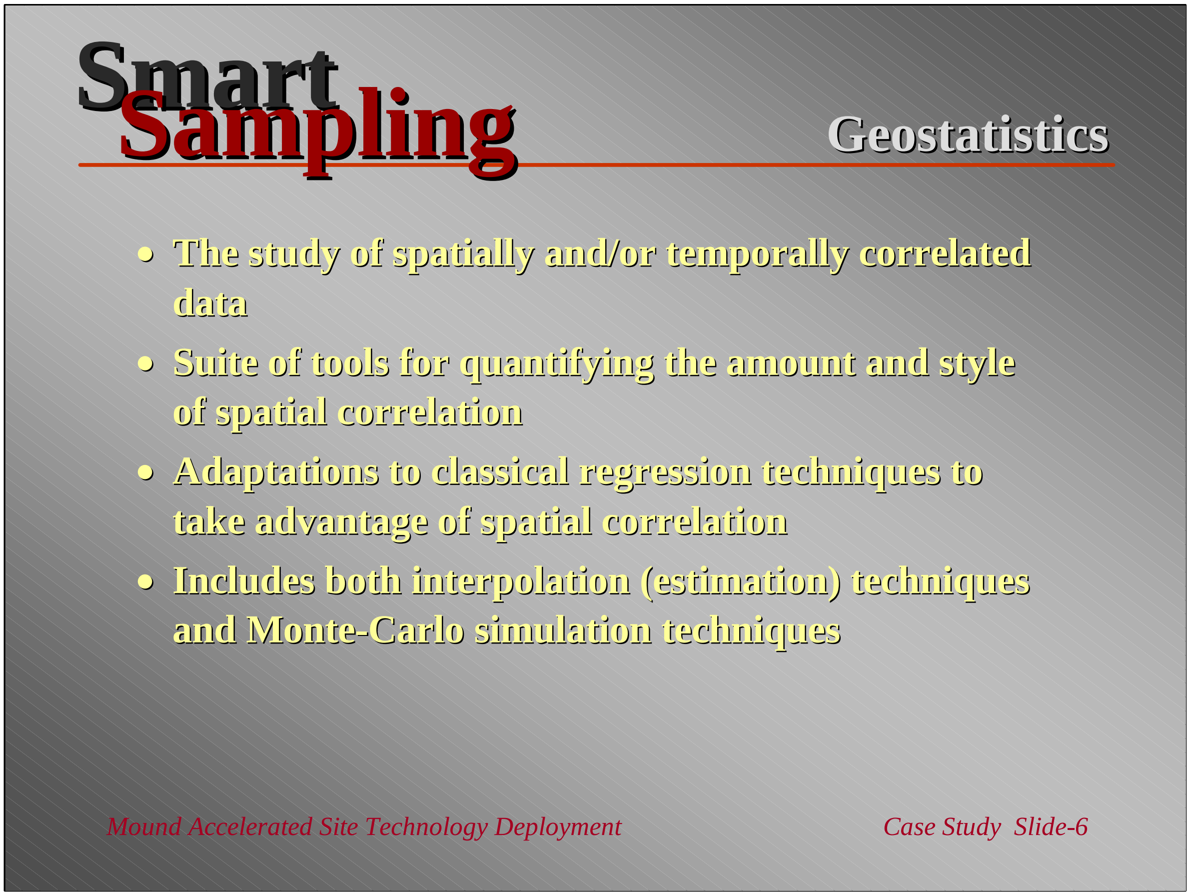
This is a fairly typical data set for a site.

SmartSampling

Approach to Problem

After the concentration and uncertainty parameters of interest have been determined, the technical core of SmartSampling, Geostatistical Simulation, is used to:

- accurately estimate the concentration at unsampled locations
- provide a measure of the uncertainty in that estimate
- create maps of the probable spatial distribution of contaminants with associated uncertainty levels
- provide sampling locations that will contribute the most to reducing uncertainty



Samples of contaminants, sediments, porosity, etc. tend to be more similar when the samples are closely spaced and less similar as the distance between the sample locations increases.

Smart Sampling

Spatial Correlation

Variogram: functional relationship between sample distance and variability

$$g(h) = \frac{1}{2n}\sum_{i=1}^{n}(z_i(x) - z_i(x+h))^2$$

- **Equation determines the amount of variability, *g* at each lag spacing, *h*.**
- **The results are plotted as a series of points (*h*, *g*) on a graph.**
- **A model is then fit through the points defining the experimental variogram**

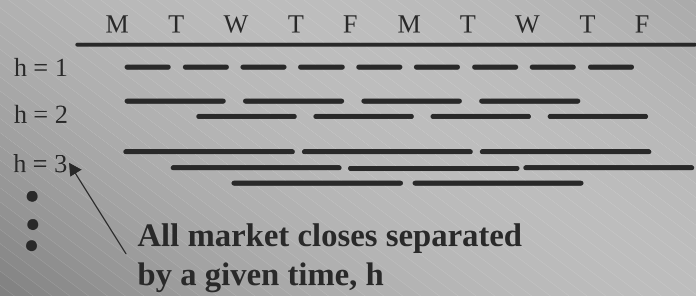
Mound Accelerated Site Technology Deployment *Case Study Slide-7*

Variography provides a quantitative means of predicting concentrations between samples.

Smart Sampling

Problem Example

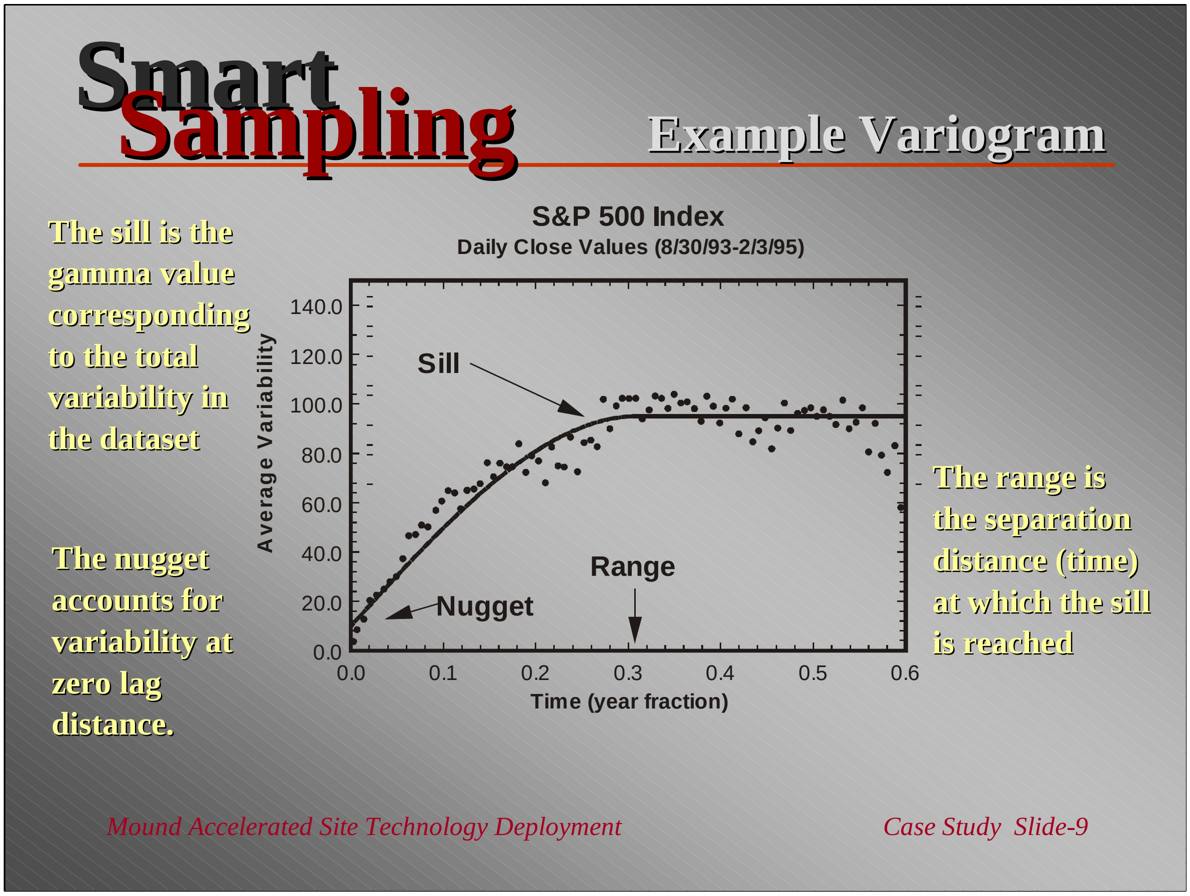
Daily Close values of S&P Index:



As the time between market closes becomes further and further apart, the variability in closing values increases until there is no correlation between them.

For each h, the experimental variogram, g is one-half the average value of the sum of the squared differences.

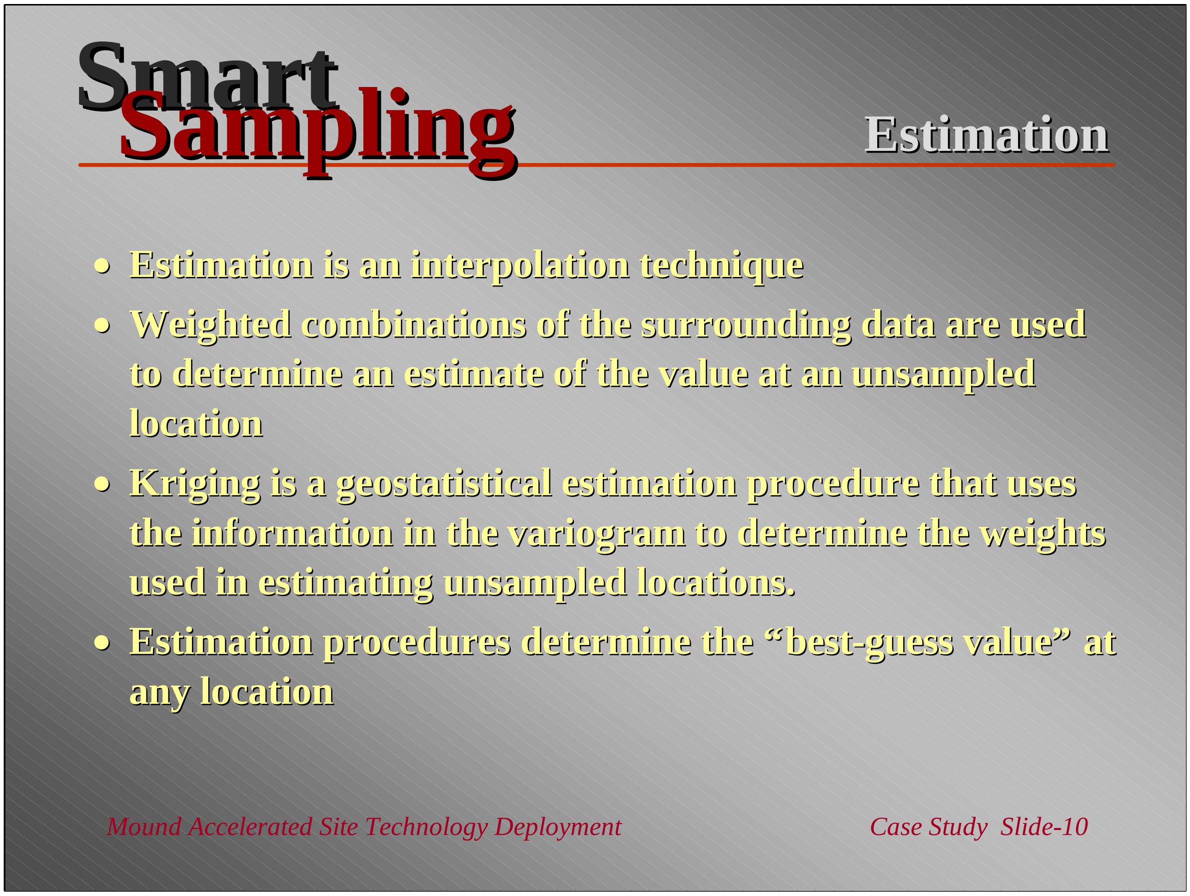
Mound Accelerated Site Technology Deployment *Case Study Slide-8*



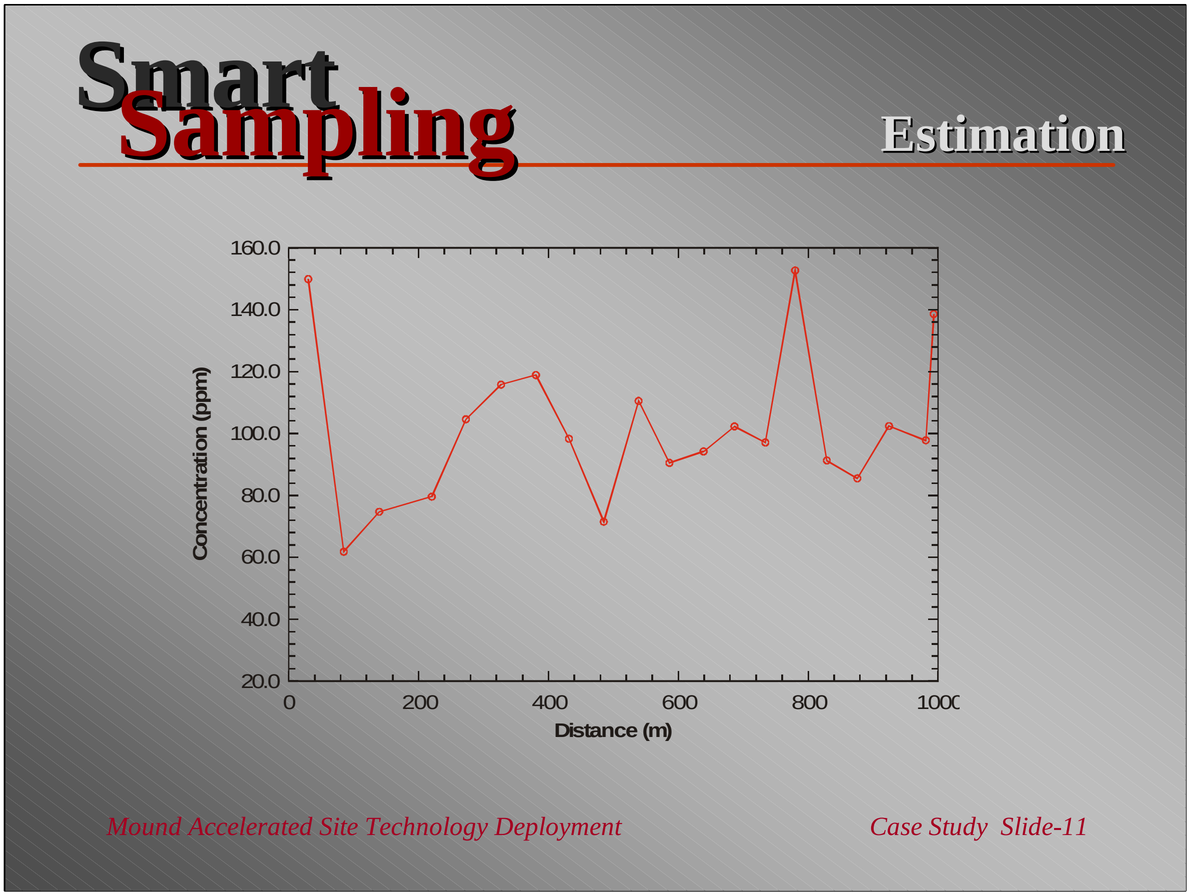
There are multiple models to fit a smooth function to the data created by the equation.

This variogram shows that you cannot make a prediction of the Close Value beyond a three month period with any accuracy.

A nugget can be a representation of measurement error or property correlation at distances smaller than the sample spacing.



- Kriging is essentially the process of determining the expected value of concentration at a given location by calculating a weighted least-squares mean of other surrounding data points.
- The weights used in the least-squares estimation are calculated using the model of spatial correlation as defined by the variogram. These weights account for the distance each data point is away from the location being estimated and the clustering of the data points.
- Since kriging is an estimation technique, the concentration map derived from kriging will contain less variability than the actual sample data (lower variance).
- This smoothing effect will also ensure that the minimum and maximum of the estimated map do not fall outside the bounds of the sample data.
- A kriged estimate of concentration along the transect would look very similar to the best guess drawn by hand.

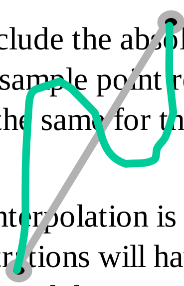


20 samples have been collected from a line across the site.

- Need a continuous model of concentration between all data points
- Straight lines or optimally smooth curves are commonly put through the points.

Some of the problems with Estimation are:

- No quantifiable range of uncertainty in the estimates (no error bars along the lines connecting the points)
- Estimation assumes that the samples include the absolute minimum and maximum in that domain. (The lowest sample point remains the lowest point even after connecting the points, the same for the highest sample point.)
- Estimation is a form of interpolation. Interpolation is a smoothing process. The histogram of the estimated concentrations will have much less variability than the histogram of the original data points.



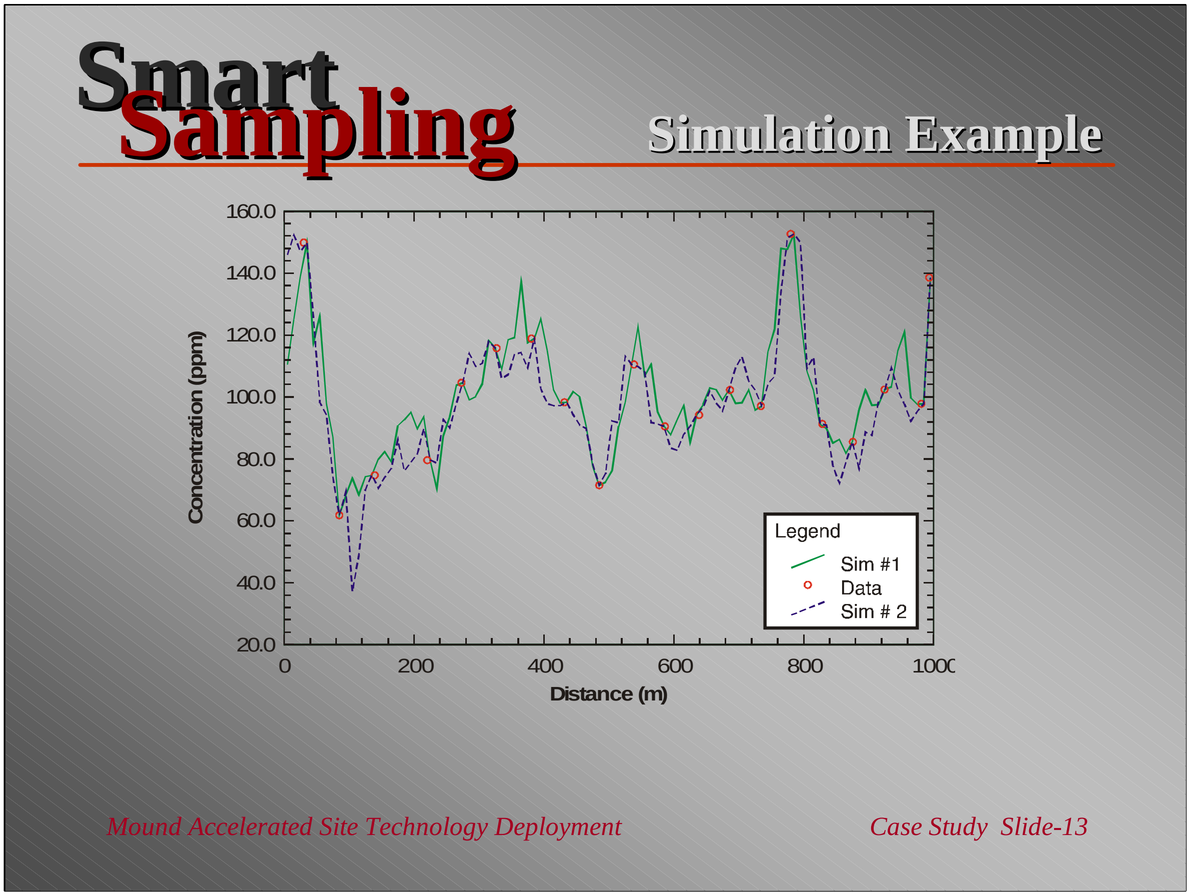
Smart Sampling

Simulation

Geostatistical simulation is a Monte-Carlo technique for producing multiple, equally probable, realizations of a sampled variable.

- **Simulation reproduces the variability of the initial data and does not have the smoothing effect of estimation**
- **Simulation reproduces the actual values, histogram, and variogram of the input data**
- **Each realization is a plausible model of the reality from which the samples were obtained**

Geostatistical Simulation is an alternative to traditional estimation techniques. Many different maps, all of which honor the sample data, can be generated.



You can generate 2, 100, or 1,000 possible values connecting the data points. These will create a "cloud" of potential realities.

This multiple map technique allows you to do post-processing of the data.

A vertical slice across the cloud gives a range of possible concentrations at that point.

Rubber Band Analogy: Imagine a peg board with pegs at the data points. Rubber bands stretched between the pegs represent the best-fit line of estimation. When the rubber bands are plucked, causing them to vibrate, the vibrations represent the cloud of lines in simulation.

Smart Sampling

Probability Mapping

Rather than mapping the best estimate of concentration, this process maps the probability of exceeding an action level.

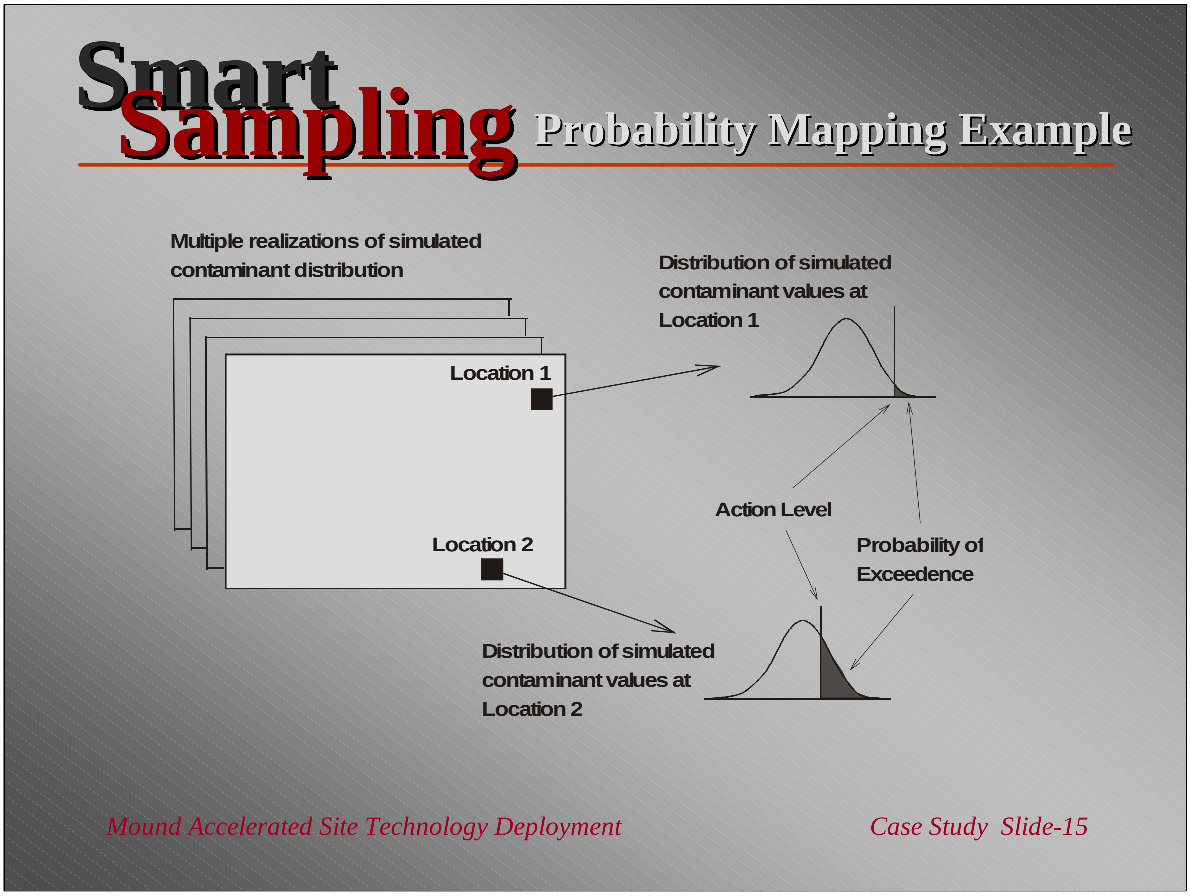
- **Realizatations can be post-processed to provide information on the probability of exceeding an action level at every location.**
- **Probability mapping is based on the concept that without exhaustive sampling, there will always be uncertainty in the extent of contamination**

The question is: **How much uncertainty is acceptable?**

Mound Accelerated Site Technology Deployment *Case Study Slide-14*

Probability mapping is the use of geostatistical simulation to determine the probability of exceeding a specified level of a contaminant at each location in the simulation domain.

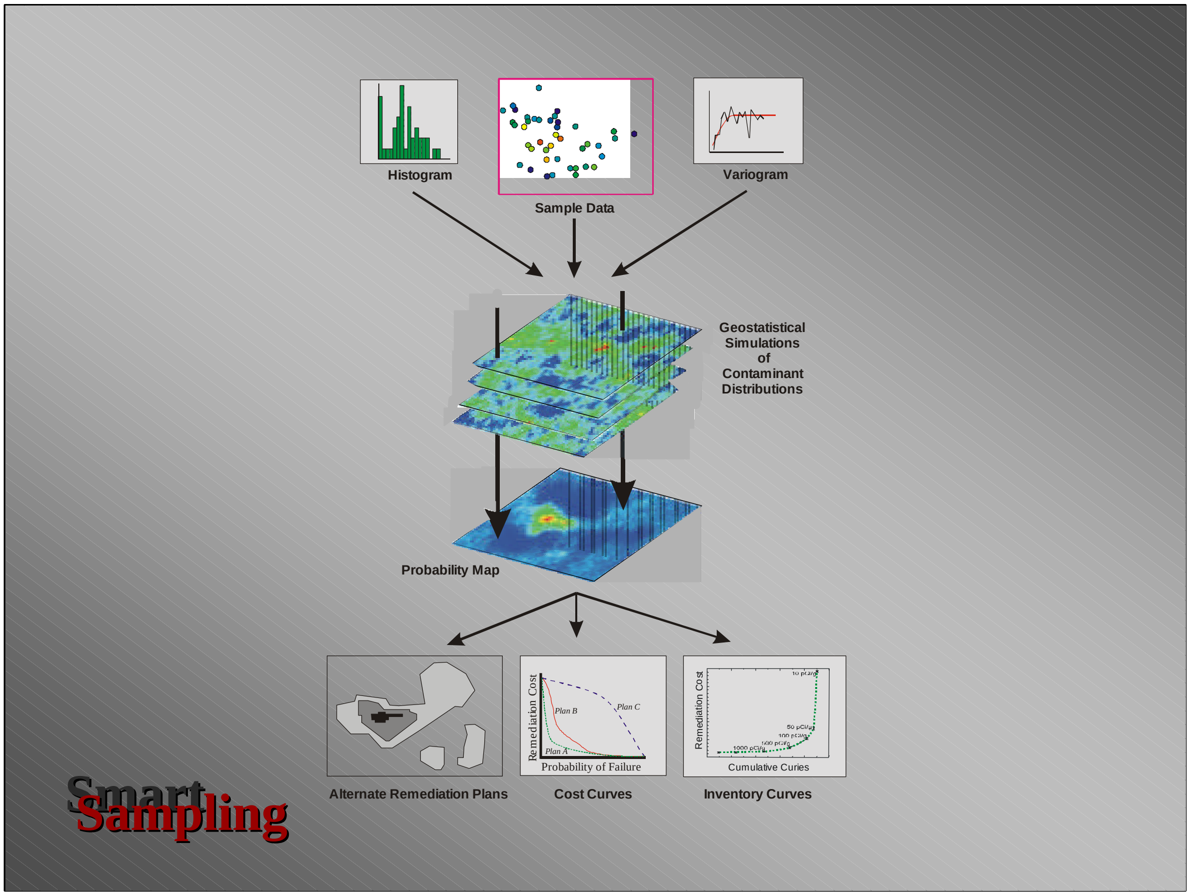
For 100 realizations of a contaminant distribution with an action level of 25 pCi/g, if 30 of the 100 realizations show concentrations greater than 25 pCi/g at a given location, then the probability of exceeding the action level at that location is 0.30, or 30%.

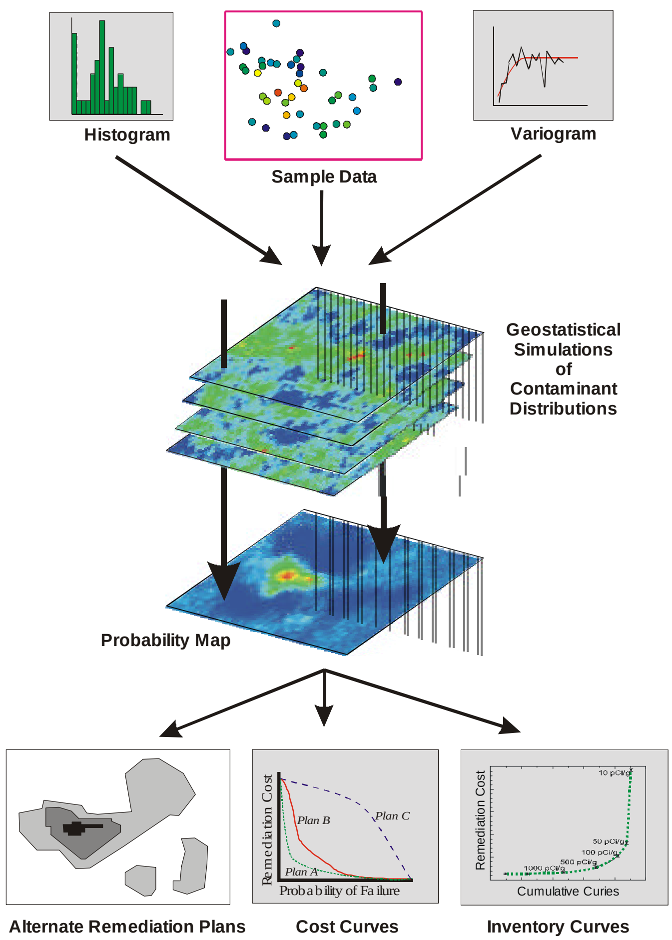


With geostatistical simulation, we can easily create many two- or three-dimensional maps of concentration.

Stack all those maps up and at any point "drill through" the realizations to get a distribution of concentrations at that point.

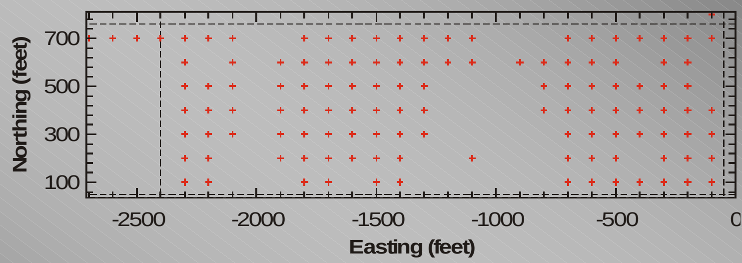
Of the distribution obtained by "drilling though" the maps, all we're really interested in is how many of those realizations exceeded the action level at that point. The output of doing this at every location is a single map that shows the probability of exceeding the action level.


Histogram
Sample Data
Variogram
Geostatistical Simulations of Contaminant Distributions
Probability Map
Alternate Remediation Plans
Cost Curves
Inventory Curves
Smart Sampling

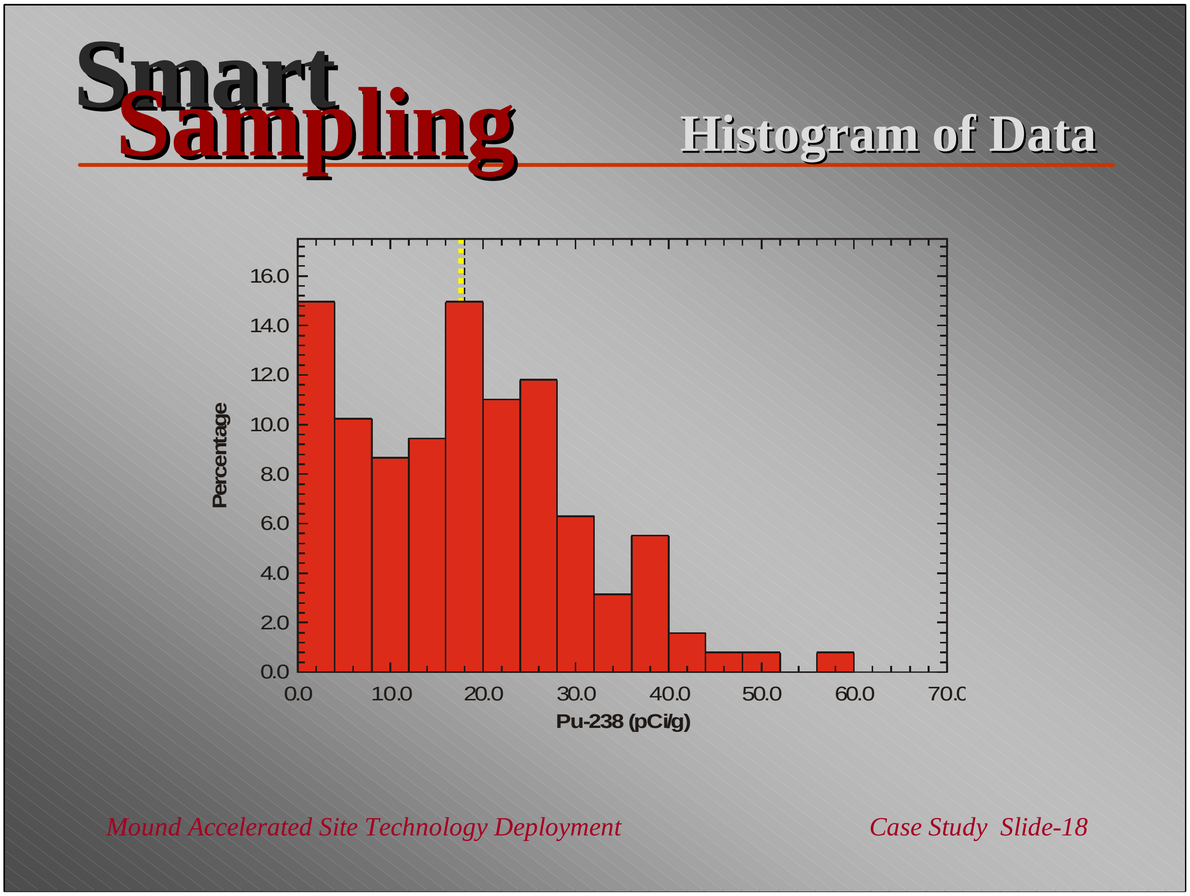

Histogram
Sample Data
Variogram
Geostatistical Simulations of Contaminant Distributions
Probability Map
Remediation Cost
Plan B
Plan C
Plan A
Probability of Failure
10 pCi/g
50 pCi/g
100 pCi/g
500 pCi/g
1000 pCi/g
Cumulative Curies
Alternate Remediation Plans
Cost Curves
Inventory Curves

Smart Sampling

Site Characterization



- •127 samples taken for Pu-238 with a field survey device
- •Sampling grid of 100’ sauare with some “holes”
- •Pu-238 activity is reported in pCi/g
- •10% of data are non-detect (set to detection limit of 0.01 pCi/g)



Need a good, short definition of histogram here.

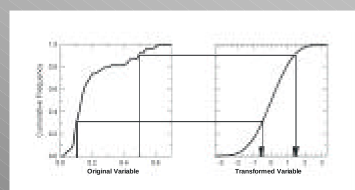
The histogram of the data is often log-normally distributed with lots of very small activity levels, tailing off as the activity level increases.

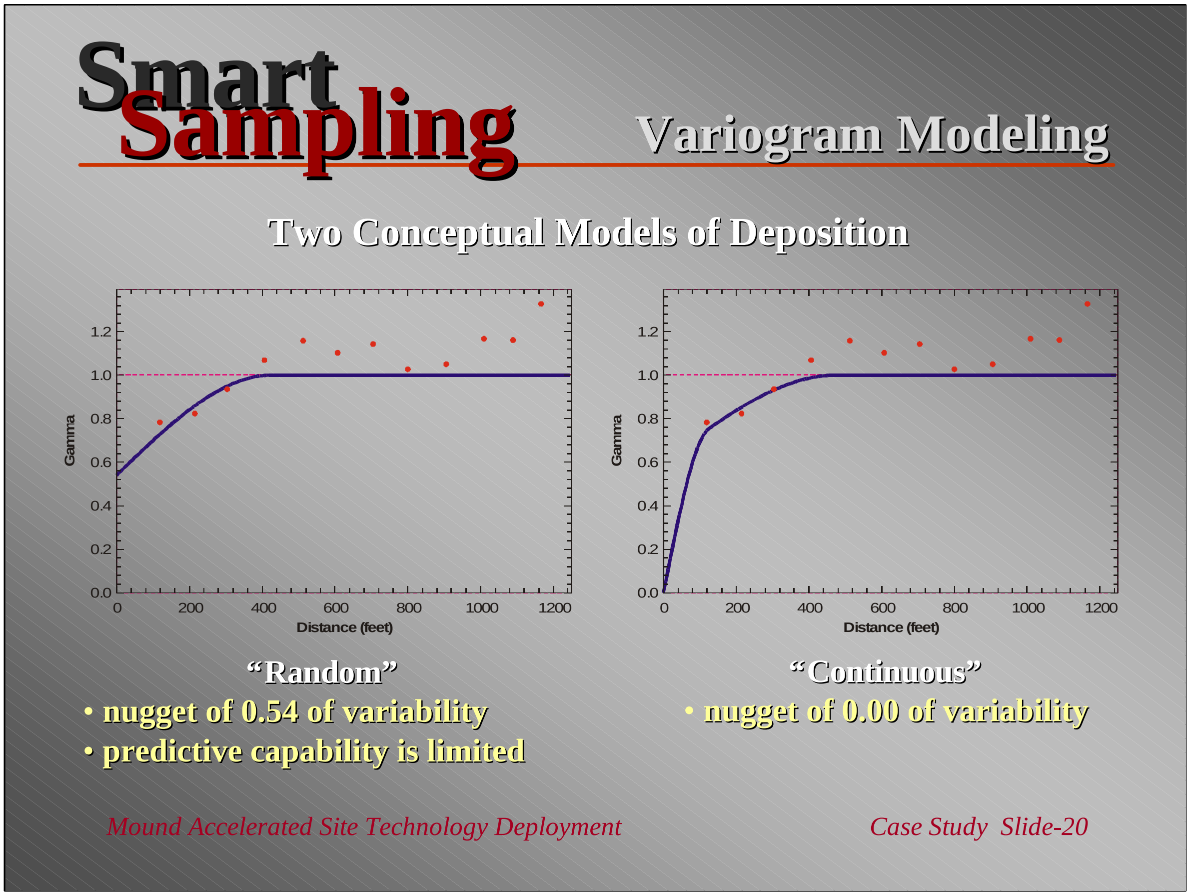
Smart Sampling

Normal-Score Transform

- Transformation of raw data into a standard-normal distribution (Gausian distribution, mean =0.0, std. dev. = 1.0)
- Quantile preserving transformation that does not change spatial correlation.
- Allows for kriging and simulation in gaussian space
- Also allows for easier understanding of variogram components

(Chris will provide better image of transform)





QUESTION: Do we want to use a variogram without fitted model to show the variability at the first sample?

Because the sampling was performed on a grid with 100 ft spacing, we don't have any points below 100' and almost 80% of the total variability is reached by the first point. As a result, there is quite a bit of variability in how the model fits through those points.

We need to look at the method of deposition of the contaminants. Was it a guy with a wheelbarrow randomly dumping ("random")? Or was it deposited evenly as if settling out of the discharge of a smokestack ("continuous")? If you move 5 ft from a sample, how well can we predict the level of contamination?

For the purpose of this training, we will run both models all the way through the case study and note differences as we go.

Smart Sampling

Modeling

The creation of a good model requires subjective interpretation. The data alone are rarely enough.

(After Deutsch and Journel, 1992)

Mound Accelerated Site Technology Deployment *Case Study Slide-21*

In practice, you never have enough information. Fitting a variogram through the points is very much an art.

In order to effectively fit a model you need to take into account all that you know about the site and the contaminant distribution.

In the case of this study, the single or multiple process of deposition are unknown.

Smart Sampling

Geostatistical Simulation

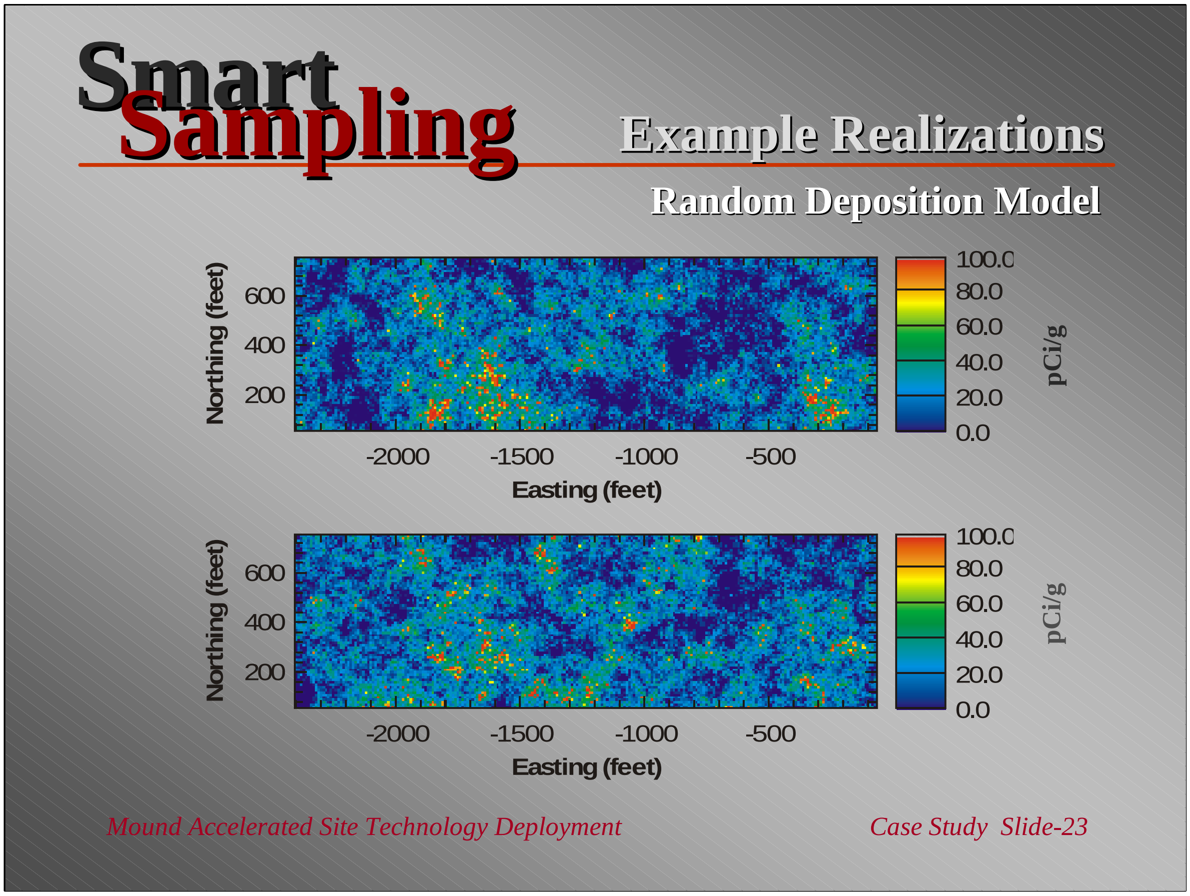
The site is discretized into 16,380 cells; each is 10’x10’

- Assumption made that the smallest area considered for remediation is 10x10, the approximate size of a bulldozer blade.

Conceptualized as a 16,380 potential samples on 10 foot centers

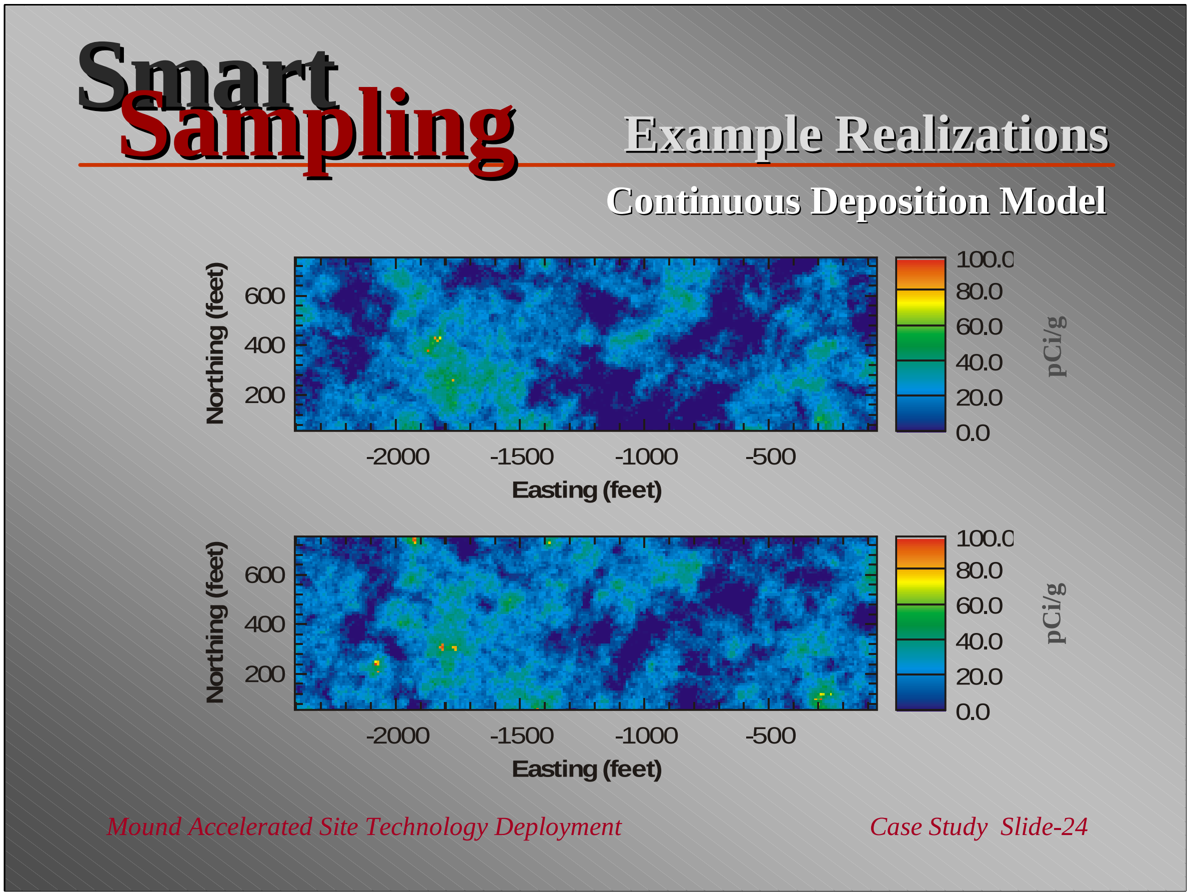
- The samples were not collected as composites across any area

100 realizations created for PRS 379



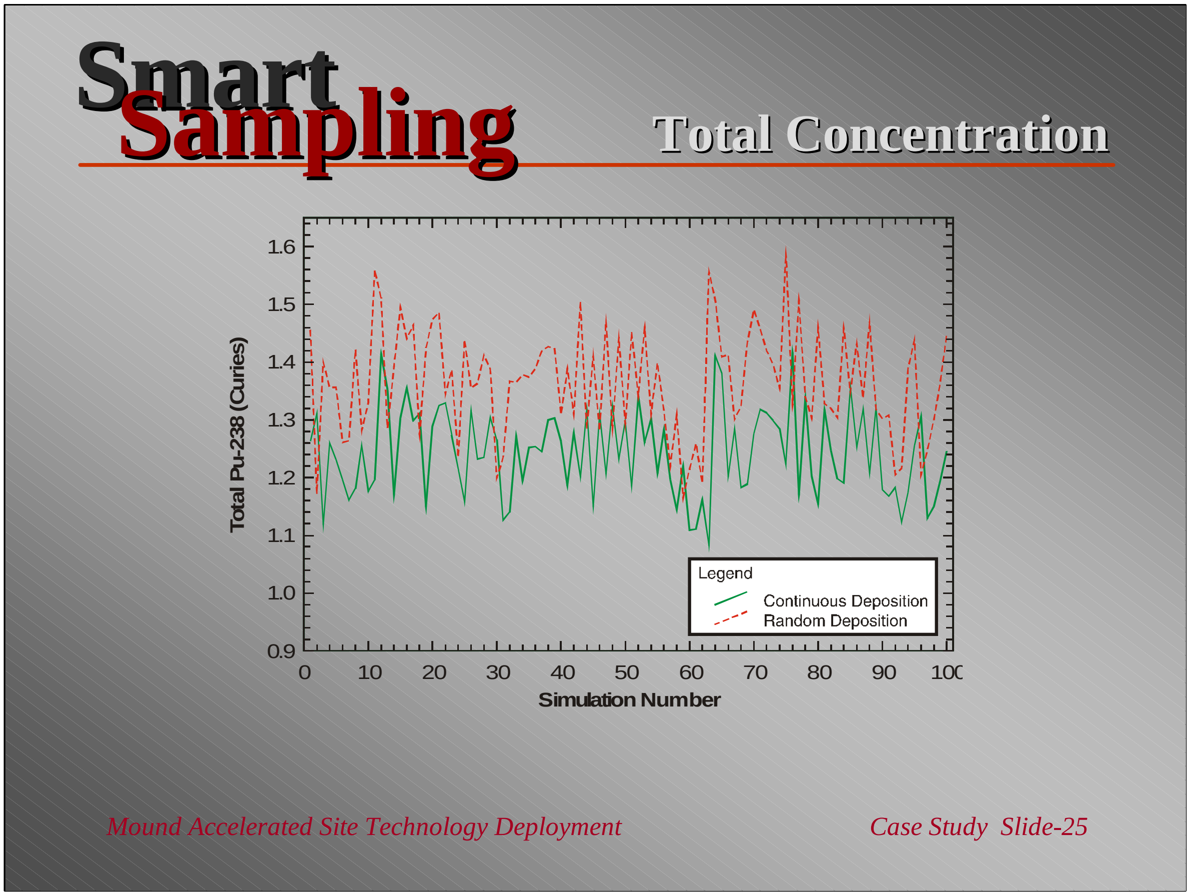
These are 2 of the 100 equally probable images of the concentration across the site created by geostatistical simulation

Based on what we know about the site, either of these could be the actual concentration distribution



These are 2 of the hundred realizations created by simulation. You can see the smoothing effect caused by the zero nugget.

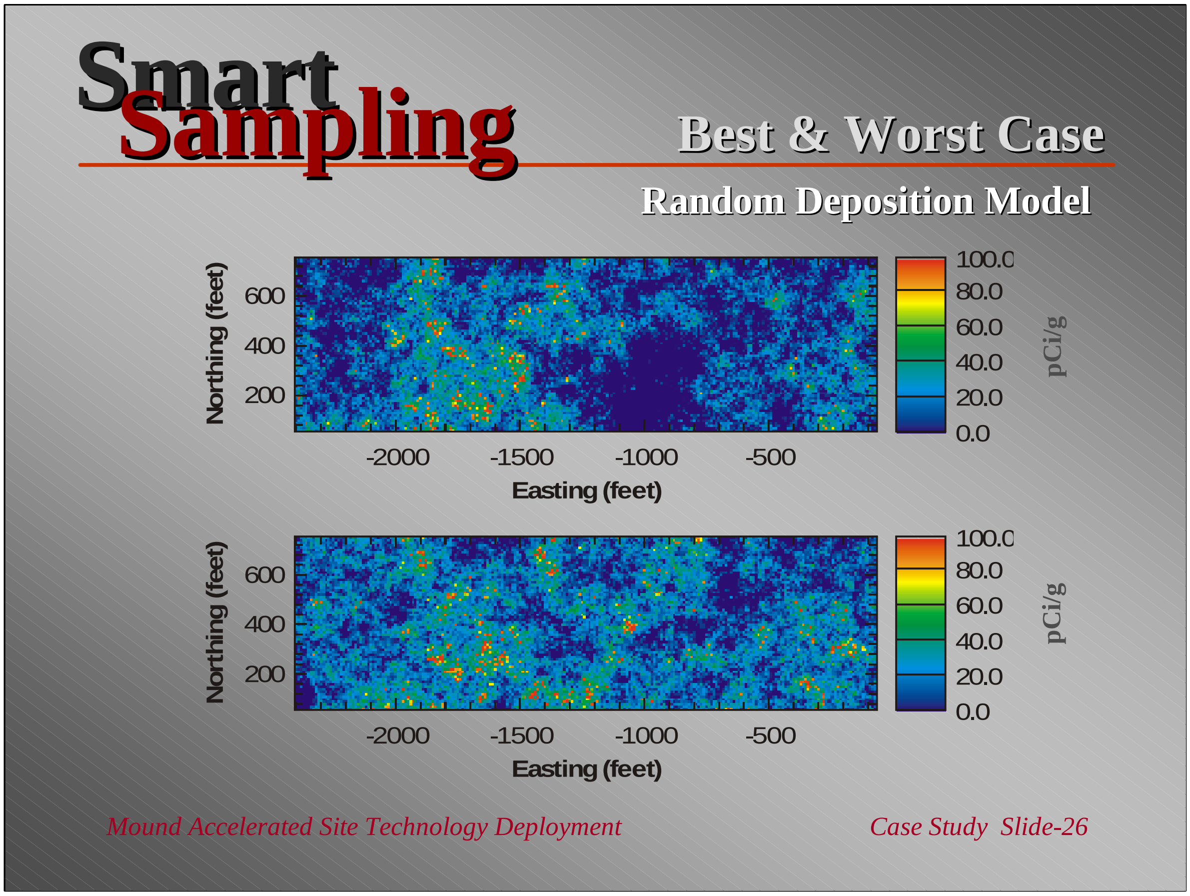
There are similarities between the two, but the realizations are different.



This is the total Curies of plutonium for each realization plotted against the realization (1-100), showing values for both the continuous and random deposition models.

The random model gives us a little higher total activity than the continuous model on average. For some idea of the variability, note that any single realization is somewhere between 1.1 and 1.6 curies of activity for either set of models.

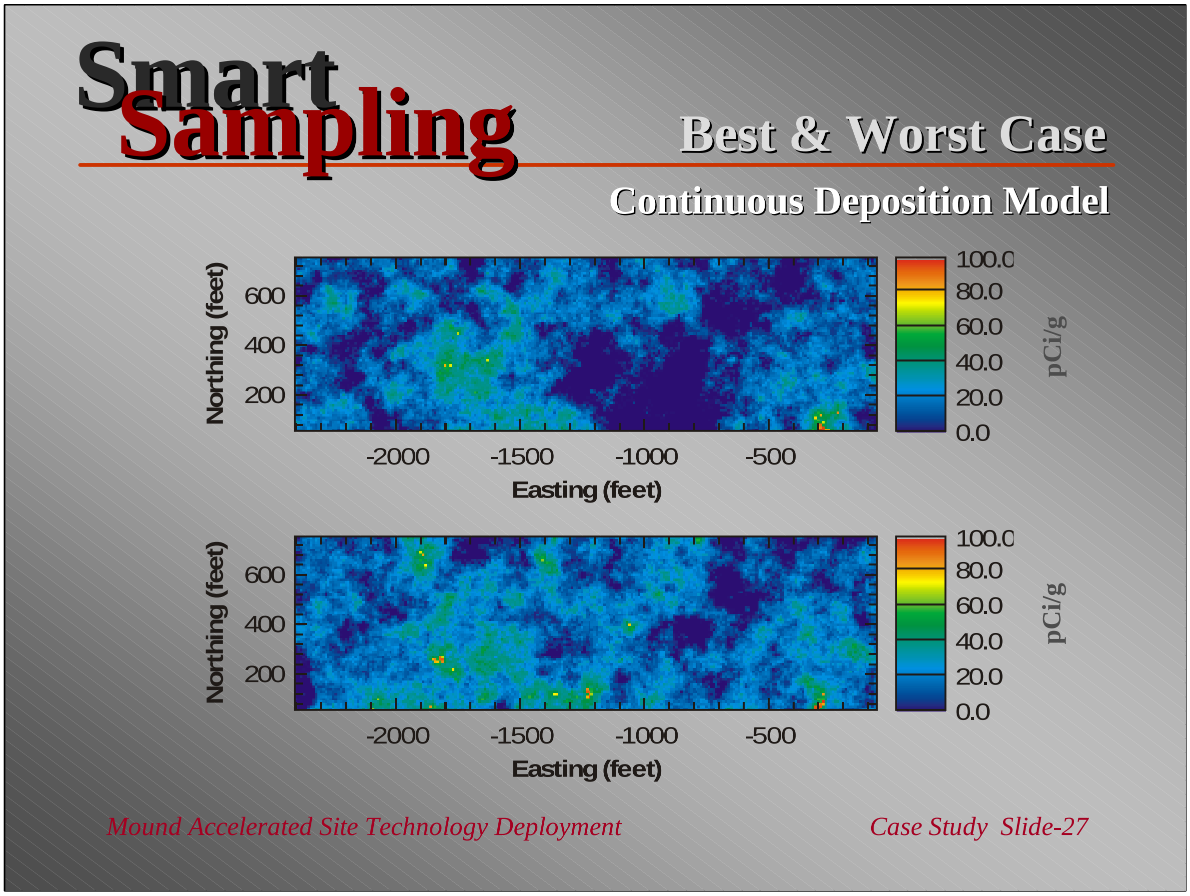
If we were to design a remediation process, (e.g., soil on a conveyer belt moving into a processor) that was able to remove the plutonium, we'd want to have some idea what capacity, in terms of plutonium activity, we would need the processor to be able to handle.



It can be useful to look at the realizations by total activity, pulling the best and worst cases.

The top realization images the "Best Case" with the lowest total curies.

The bottom "Worst Case" realization displays the highest total curies.



As with the “random” realizations, what drives the difference between the best and worst cases, the area of large uncertainty, is the data hole around -1000.

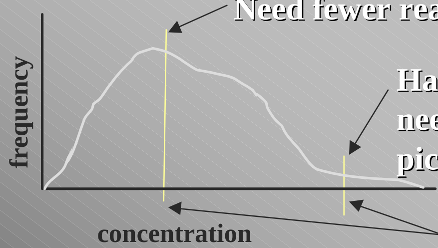
Smart Sampling

What is Enough?

How many realizations are enough to make an informed decision?

- **The answer is dependent on the decision rule**
  (the action level relative to the sampling distribution)

Need fewer realizations to get accurate picture

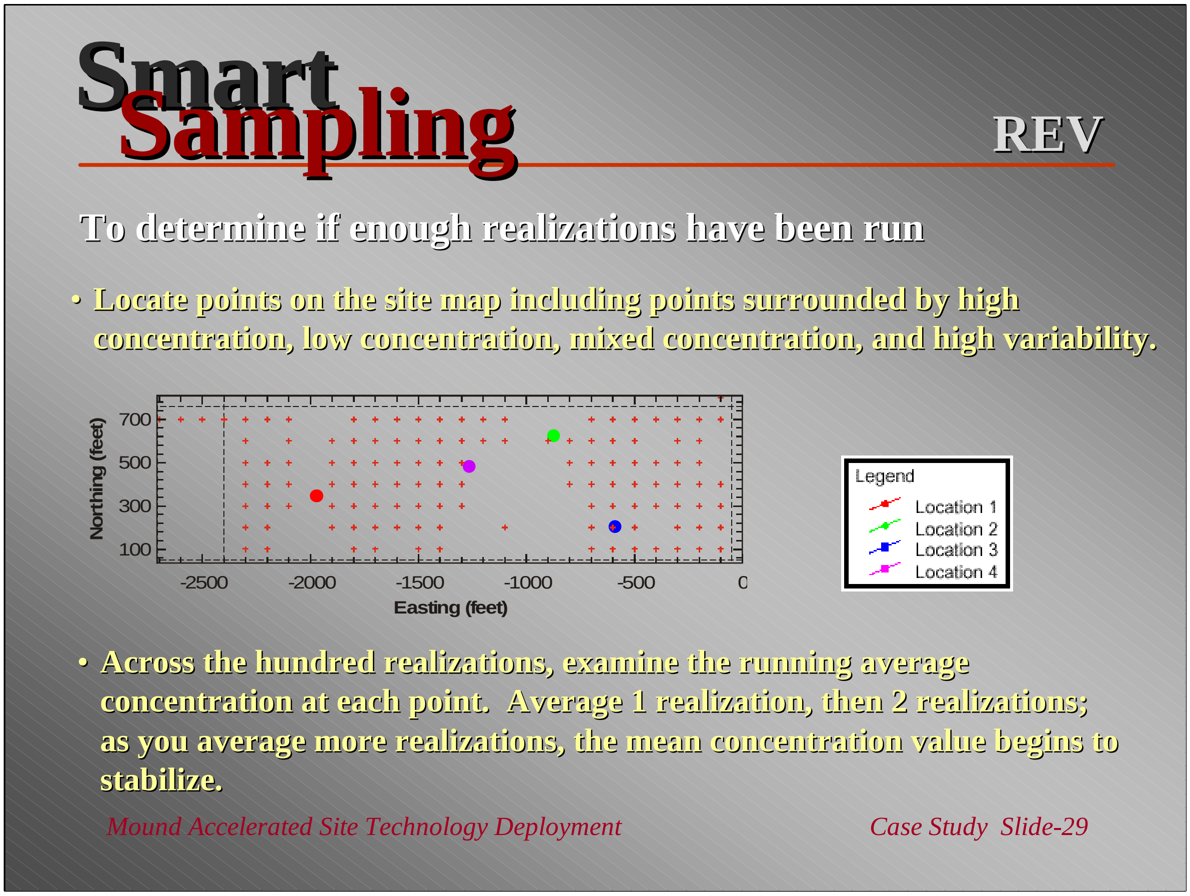


Hard to define the tail to the pixel, so need more realizations to get an accurate picture

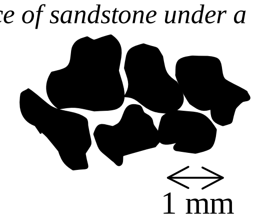
Action levels

- **Is an idea of the best, worst, and expected conditions enough?**
- **Test how many is "enough" with concept of a representative elementary volume (REV) from groundwater hydrology**

how much precision is necessary to make



A slice of sandstone under a microscope looks something like this:

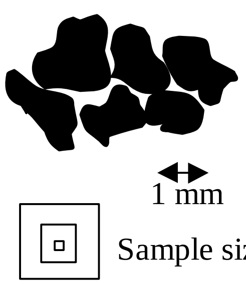


☐ Sample size

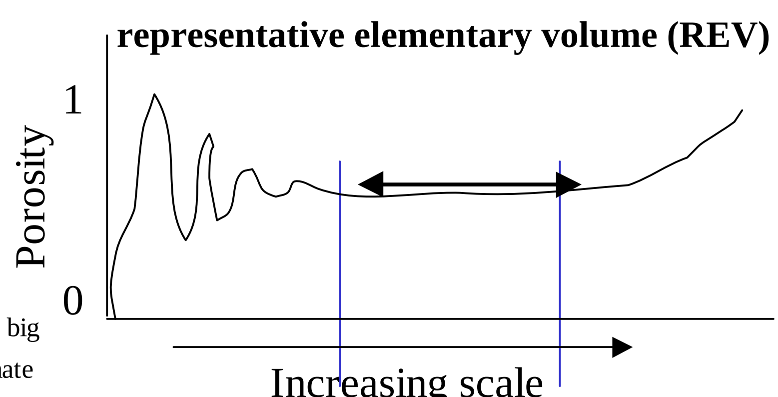
A randomly located tiny sample will either have a porosity of 1 (grain) or 0 (pore) as the sampler moves around the slice.

As the sampler increases in size, the sampled porosity becomes an average of both grains and pores.

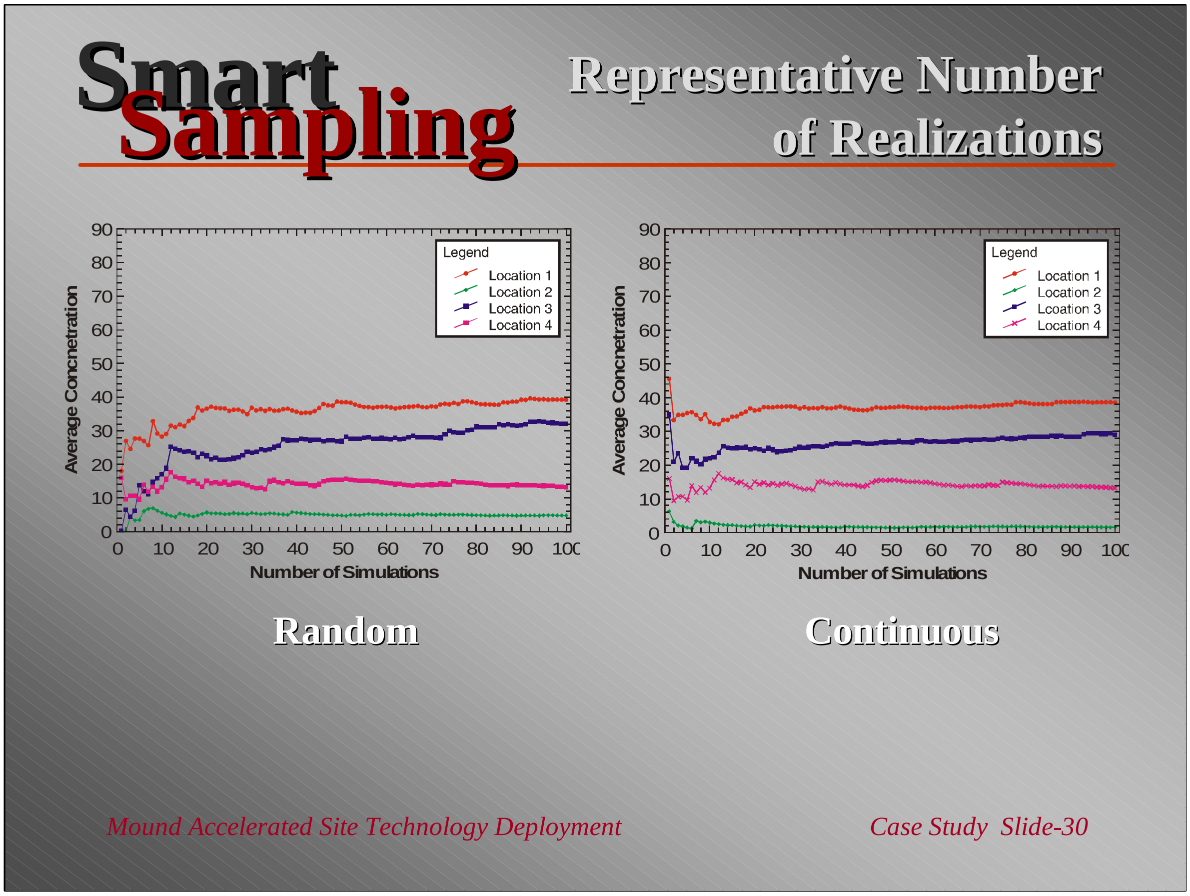
A plot of the results it looks like this:



Sample size



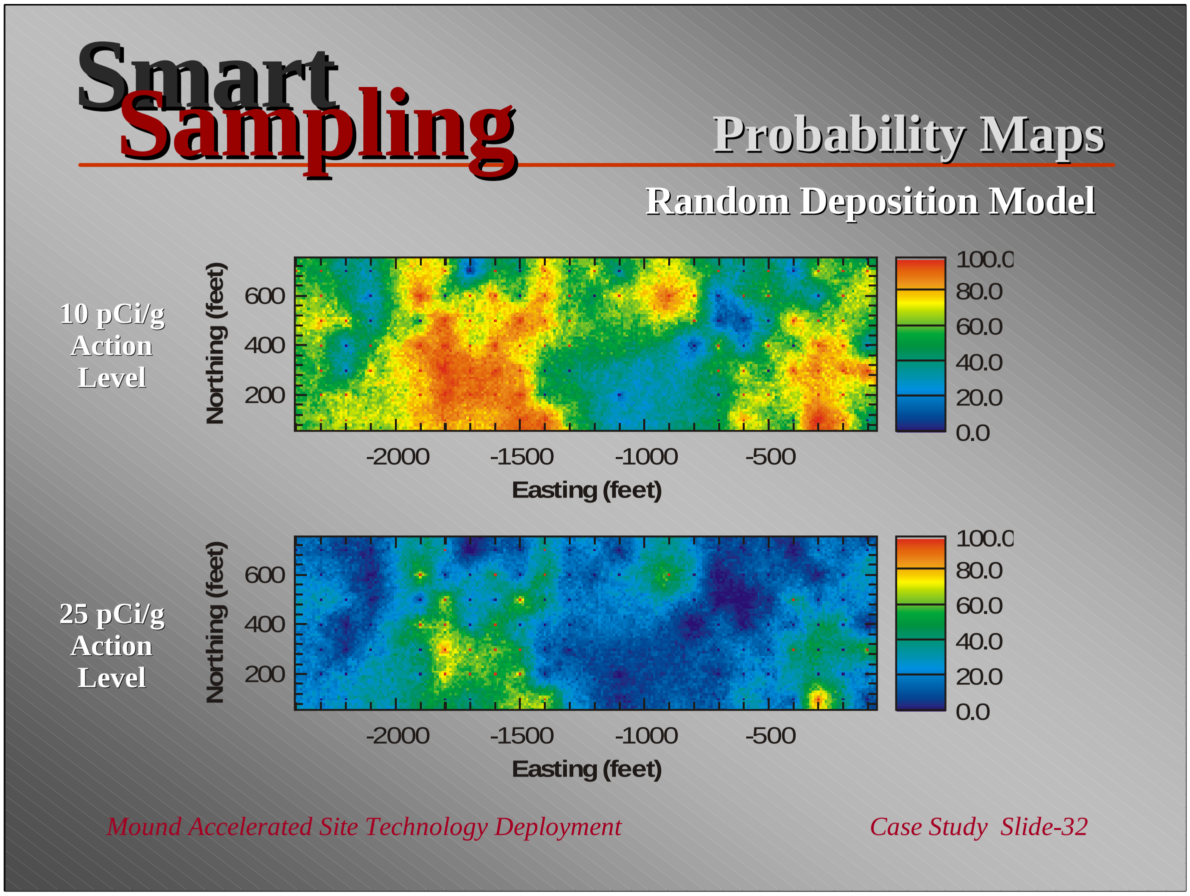
At some point the average is over a big enough area that the porosity estimate stabilizes. As the sample size continues to increase, at some point another soil type gets averaged in and the estimate changes again.



- With just a couple simulations, the average concentration can be quite variable, but as the running average contains more realizations, the average concentration value begins to stabilize.
- Different points may need a different number of realizations to stabilize to a constant mean concentration.
- This technique provides an easy way to make a defensible argument on whether or not you’ve done enough realizations

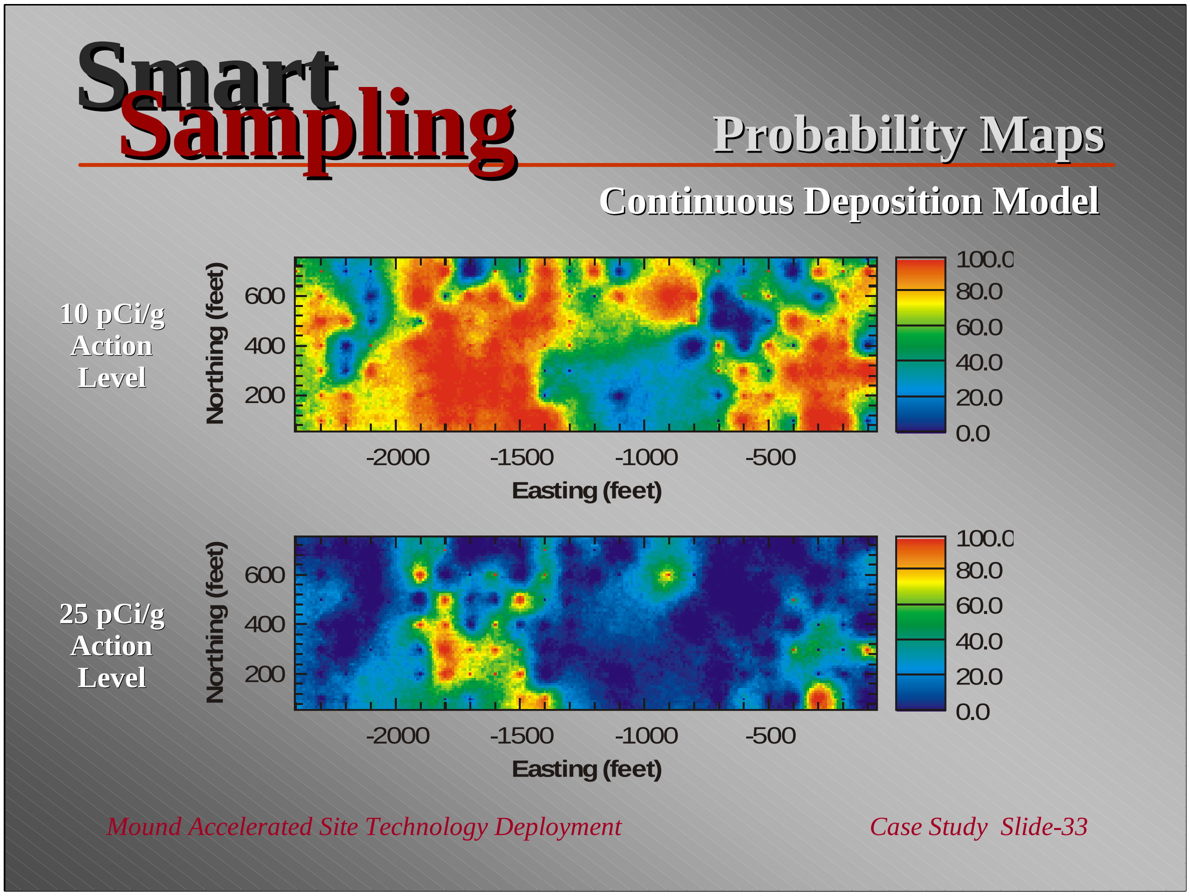
Smart Sampling Action Levels & Reliability

- The site is negotiating with regulators and stakeholders regarding final action level
- Three action levels are being considered: 10, 25 and 50 pCi/g
- All parties are considering cleaning up to 95% reliability, but they also want to see what clean up to 99% reliability would require (0.05 and 0.01 probabilities of failure)



The probability of exceeding an action level is shown in %.

The sample locations show up as 100% or 0% probability of failure, because they are exact values and they are honored by every simulation. They are either above or below the action level.



The sample values influence the values of surrounding cells here so they appear as blobs instead of isolated dots.

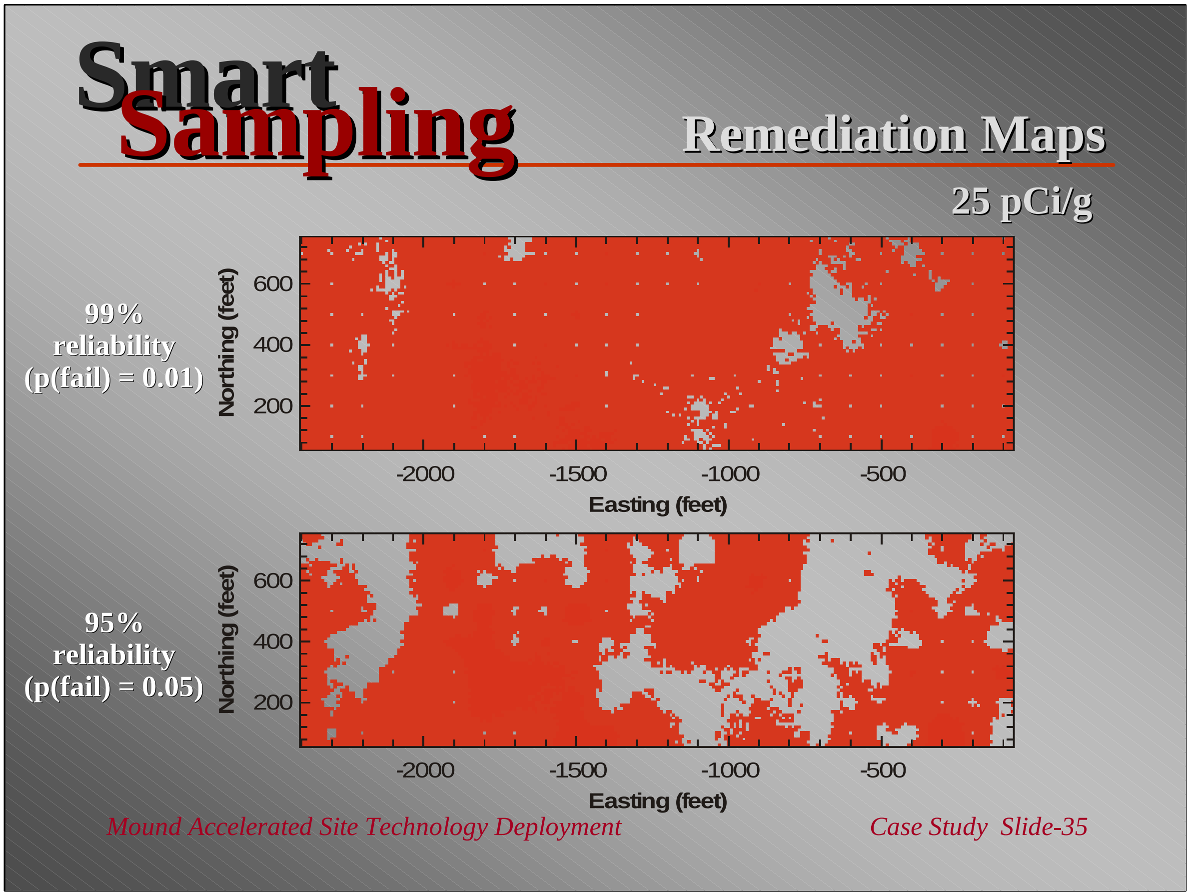
At the low action level, there is a large area of the site that has a good chance of being over 10 pCi/g. If the action level rises to 25pCi/g, the chances of being higher than the action level occur in a much smaller area of the site.

Smart Sampling — Remediation Mapping

- Remediation maps are generally created by contouring the acceptable reliability on the probablility map

  For example, to clean up to a 40% probability of failure, draw the line where the **blue** turns to **green** on the probability map.

- Areas with reliability less than that requested are remediated
- Areas with reliability equal to or greater than that requested are left in place



Per the decision rule, the red colored cells require remediation. Each square is a 10x10 cell.

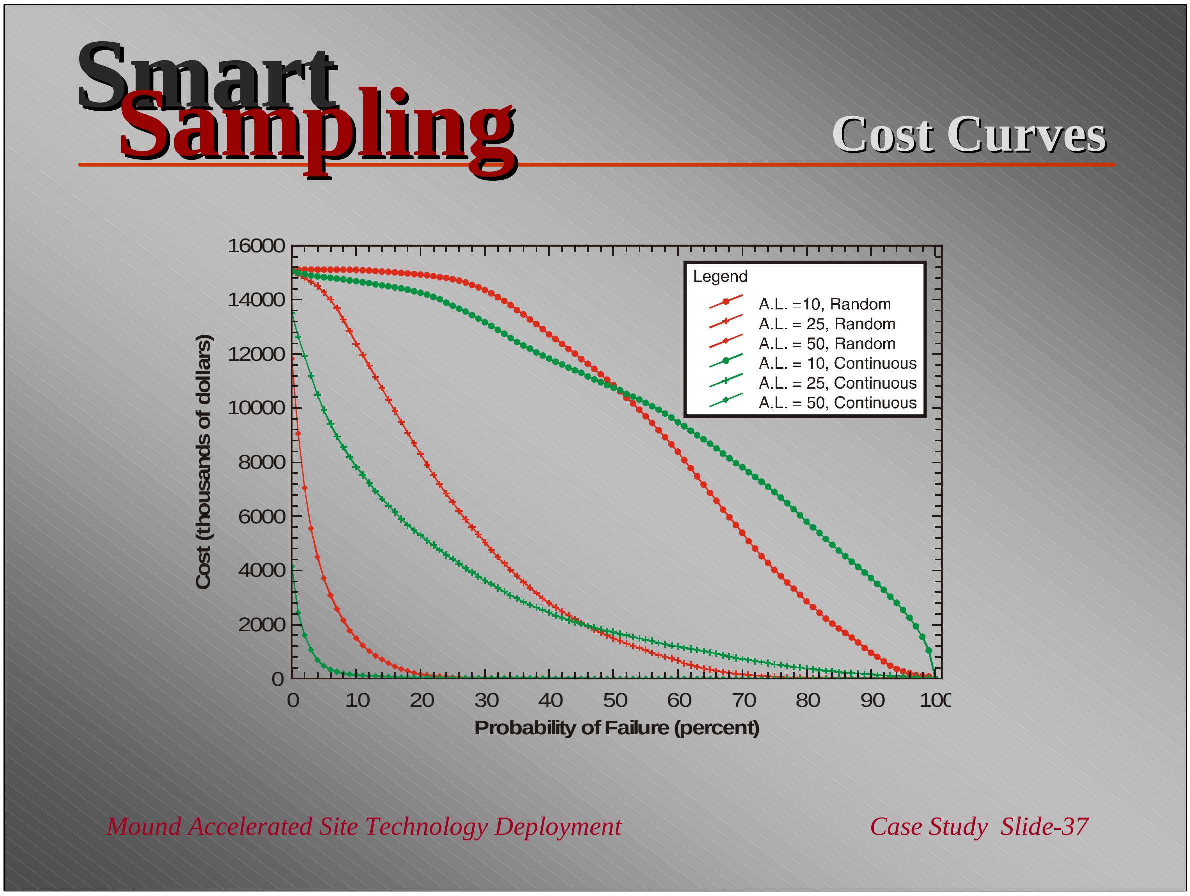
Smart Sampling

Cost of Remediation

Cost estimates for remediation at three action levels and two reliabilities are requested

- **Depth of remediation is 6 inches**
- **Cost of remediation is $500/yd^3.**

It is possible to determine costs at all three action levels for all reliabilities



The plot shows the cost of remediation in thousands of dollars vs. the probability of failure for both deposition models

The decision on which model to use for the final remediation design must be based on knowledge of the site history and deposition method, not just cost.

At the 10pCi/g action level, backing off on the reliability won't make much difference, but dropping the action level to 25pCi/g makes a substantial change in cost.

Smart Sampling

Follow Up Sampling

Site has budget for 39 more samples, where should they go?

- **Use probability maps to determine optimal locations to reduce uncertainty**
- **Highest uncertainty = area of 50% probability of exceeding action level**

It is advisable to set an action level before proceeding with location of additional sampling.

- **The greatest variability between realizations may be in areas that are reliably much above or below the action level and where more sampling won't make a difference.**

Mound Accelerated Site Technology Deployment *Case Study Slide-38*

At locations with 50% reliability, you know as much about contaminant concentration as you would by flipping a coin 100 times. Other areas are either basically above or below the reliability level, and more sampling data will not change the remediation maps.

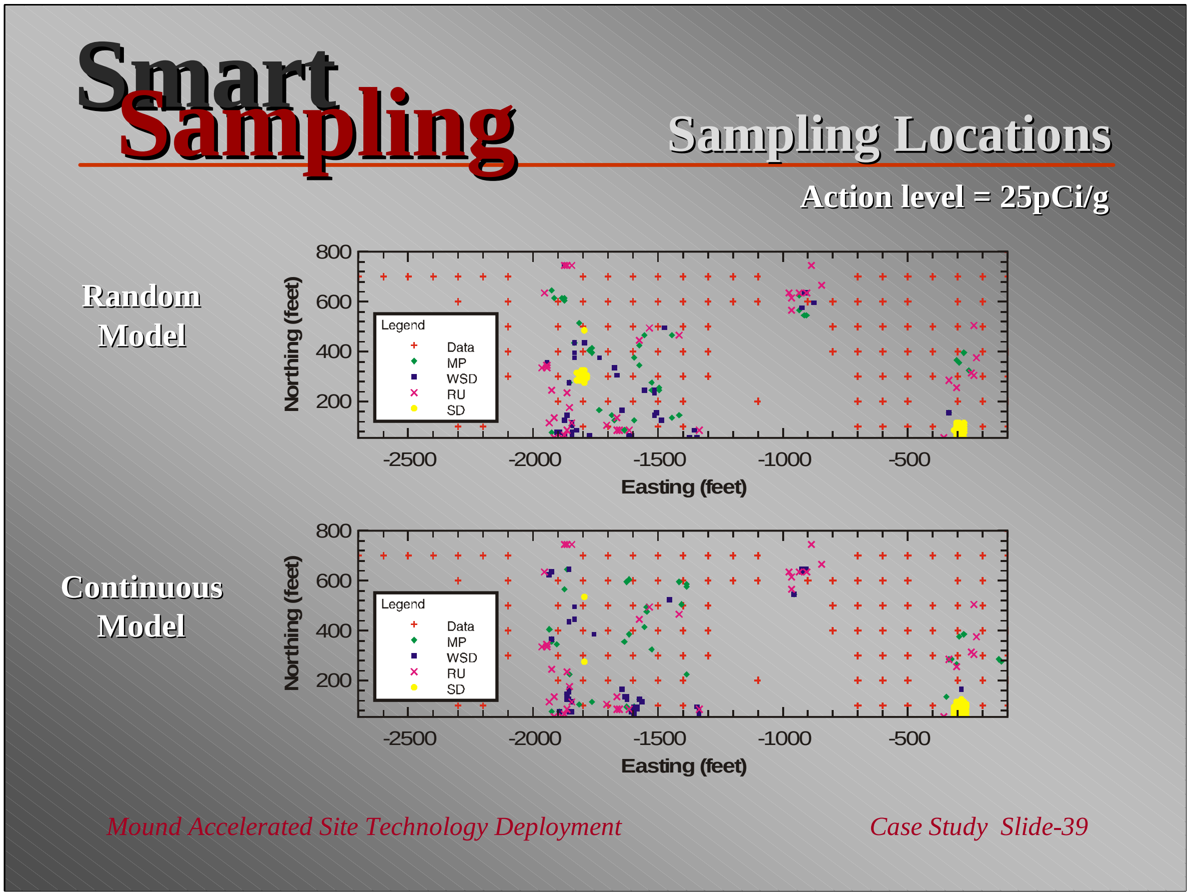
The point of geostatistical simulation is to use the spatial continuity information to help make predictions into unsampled areas to prevent too much sampling.

Additional sampling has no "worth" or "information content" if it does not: a) change the decision or b) reduce uncertainty.

Examples of "worthless" additional sampling:

- continued sampling in the immediate vicinity of several samples all of which are markedly over the relevant action level.
- continued sampling in regions of extensive background levels.

Additional sampling should emphasize regions of maximum uncertainty.



These maps show a few different methods for determining where to sample

•Standard deviation technique does not look at an action level, but at areas where the greatest variability occurred between realizations.

•The other techniques take the action level and uncertainty into account, and are different ways of hitting the "green zone" of 50% uncertainty.

Smart Sampling

Summary

- Uncertainty is due to limited sampling of a spatially heterogeneous variable
- Spatial uncertainty creates uncertainty in remediation designs and costs
- SmartSampling uses geostatistical simulation to provide a technique for examining and quantifying the amount uncertainty.
- Estimates of uncertainty can be propagated through the costs

Smart Sampling

Summary

- In case study assumptions regarding the conceptual model have a large impact on cost estimates
- As in any engineering design, 100% reliability (if attainable) increases costs
- Follow up sampling designs that incorporate the action level are superior to those that just examine variability