The (Un)Importance of Soft Data Calibration in Collocated Cokriging:

Implications for Environmental Assessment and Remediation

(An Example from a Soil-Contamination Site)

Chris Rautman
Sandia National Laboratories
Albuquerque, New Mexico

Sandia National Laboratories

Work sponsored by the U.S. Department of Energy
Accelerated Site Technology Deployment Program

Outline

Background on the Example Site

Alternative Remediation Plans

-- The Segmented Gate System technology

Geostatistical Modeling to Support Deployment of the SGS

-- Data: Hard and Soft

-- Variography

-- Creating an "Exhaustive" Secondary Data Set

-- Simulation with Collocated Cokriging

-- Probabilistic Summaries of Contaminant Nature & Extent

Theory Underlying Collocated Cokriging and Why Things Worked Out the Way They Did

Conclusions and Implications

Background Information

Landscape soils at Brookhaven National Laboratory on Long Island, New York, have been contaminated with radioactive cesium-137, a fission product from reactor research.

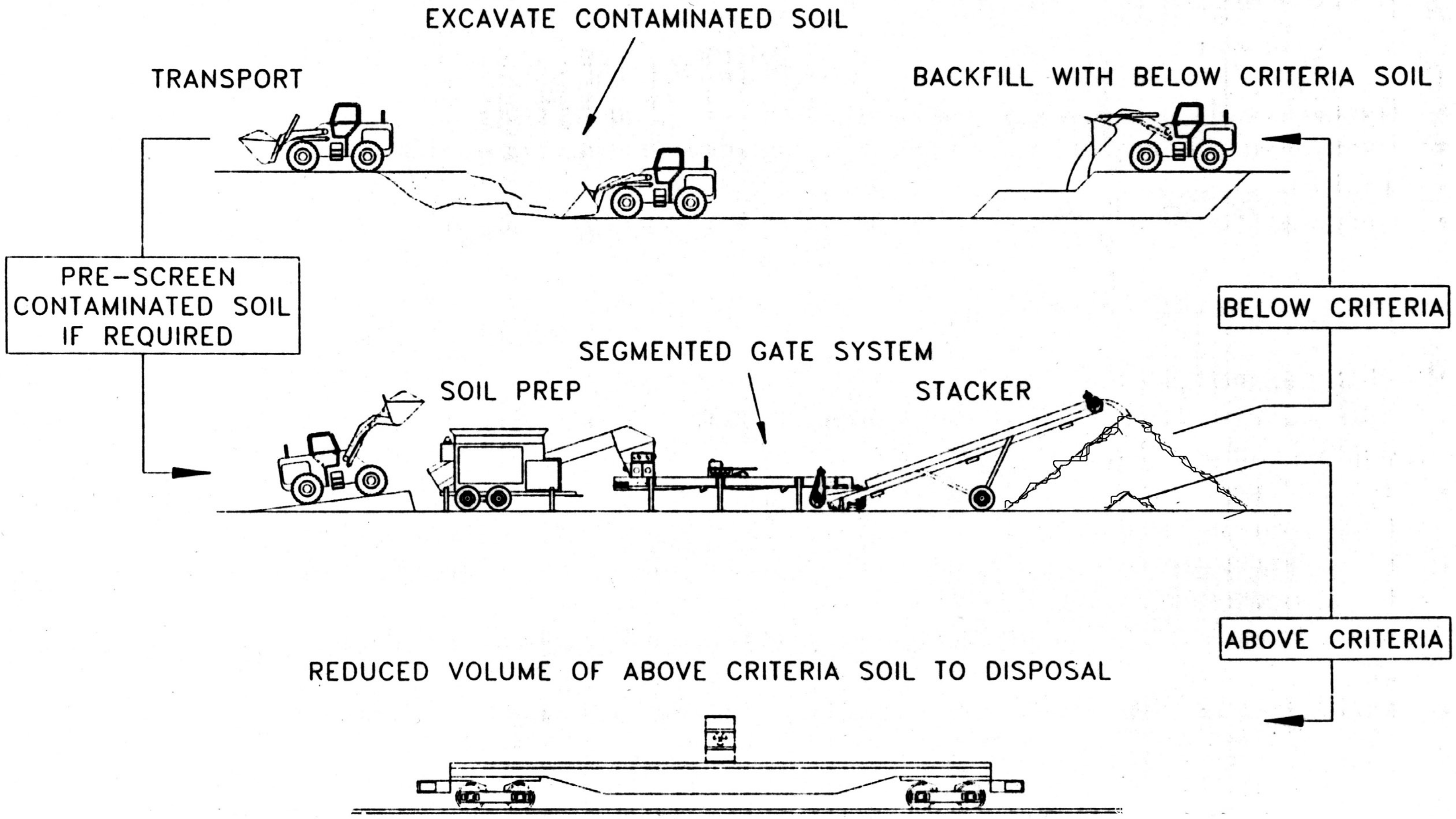
Preliminary gamma-ray surveys outlined regions for remediation. A baseline remediation plan was proposed that essentially involved excavation to a depth of 2 feet of material with activities exceeding 23 pCi/g followed by off-site disposal of contaminated soil.

A proposal has been made to apply a novel remediation technology, the Segmented Gate System, to separate contaminated from clean materials in real time during excavation to reduce the volume needed to be shipped.

Purpose: Evaluate likelihood of successful use of SGS.

The Segmented Gate System (SGS)

The SGS is designed to discriminate radioactively contaminated soils from material below a specified threshold activity during excavation:



The Segmented Gate System

Field Set-up

Hydraulically
Activated
"Segmented Gates"

Contaminant Data

Two types of soil data (^{137}Cs activity) were available:

-- Hard: Laboratory-measured in a controlled environment using a standardized radiometric-counting configuration in a mock-up of the SGS.

-- Soft: Field radiometric measurements using a portable device known as the "mower."

Hard Data: 17 soil profiles of four 6-inch depth increments; for a total of 68 total measurements in 3-D.

Soft Data: 5,527 closely spaced values, but only in 2-D.

Field Sampling and Measurement

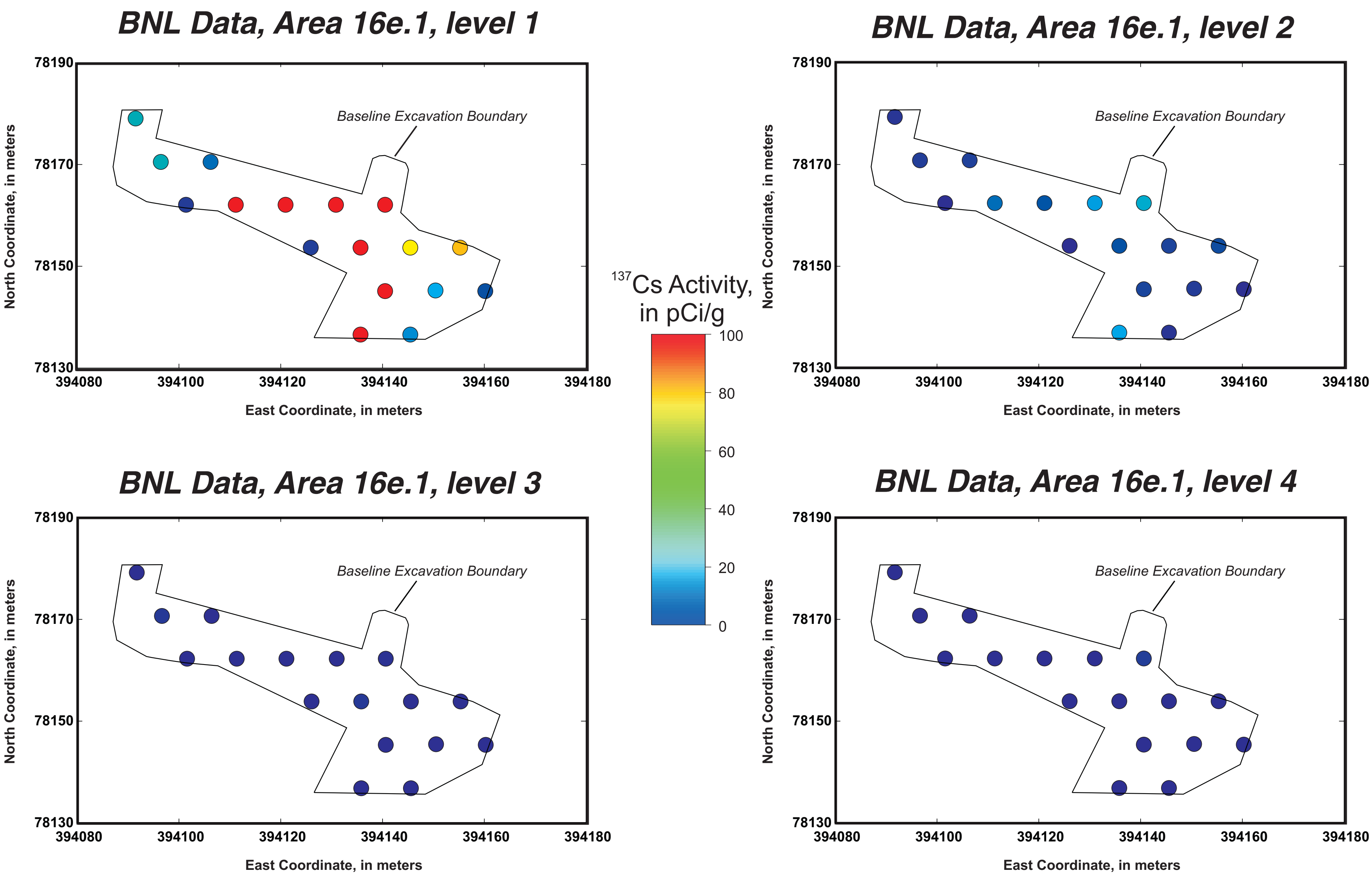
Soil Sample
Collection

Sample Size
(Support Volume)

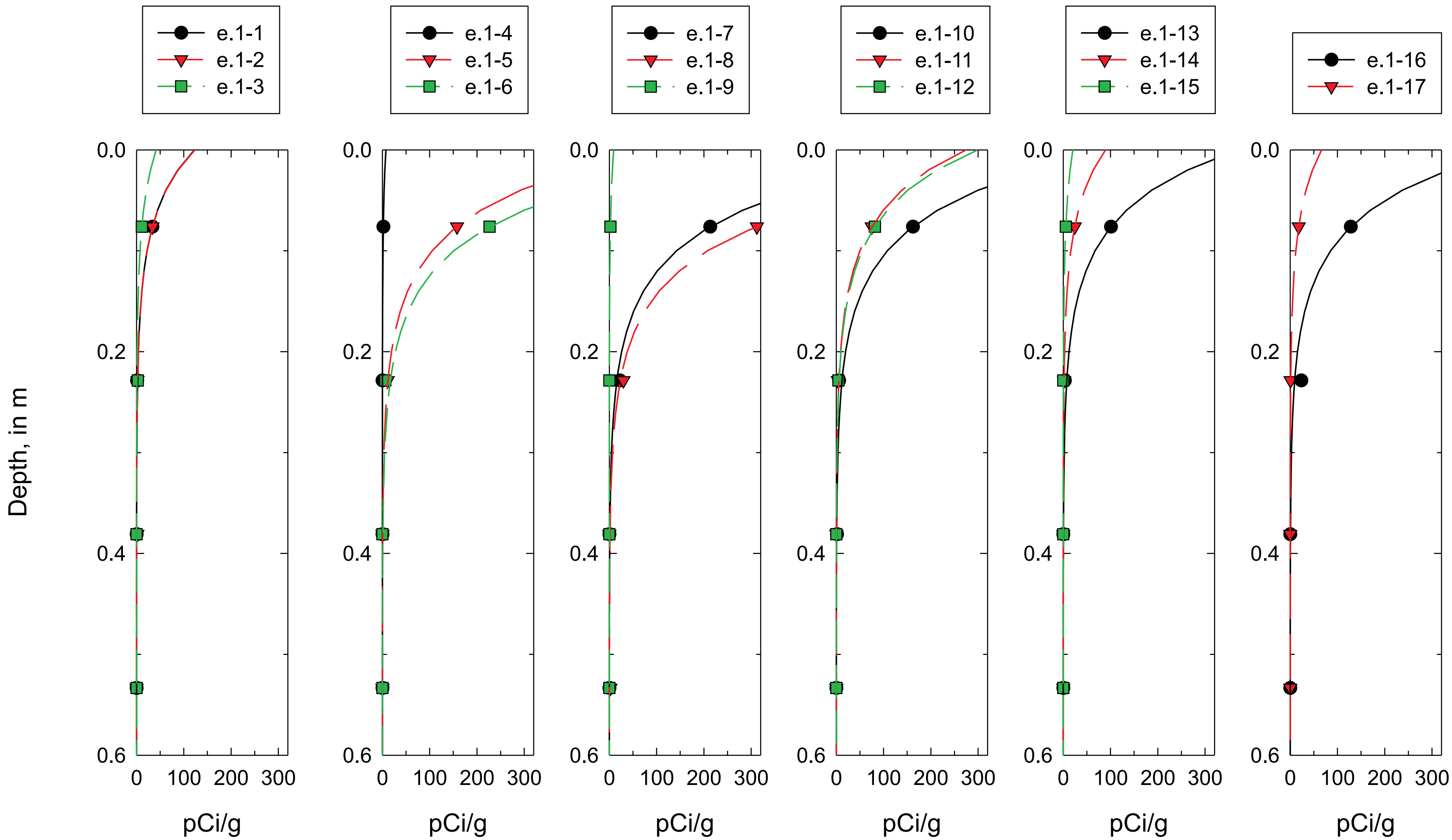
The "Mower"

Calibration Source
"Sample" Size
(Support Volume)

Soil Profile Data by Level

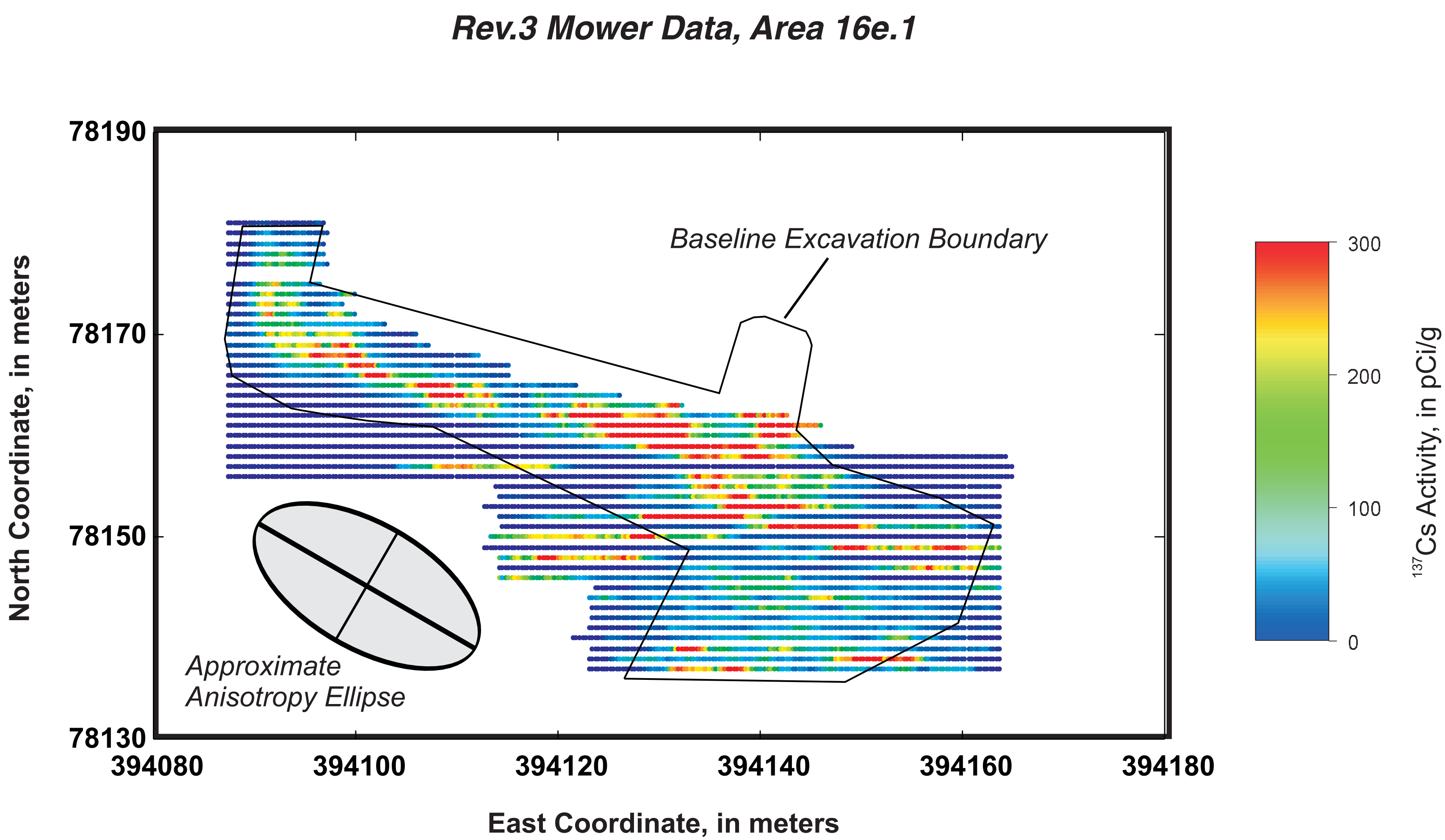


Soil Sample Data in Profile View

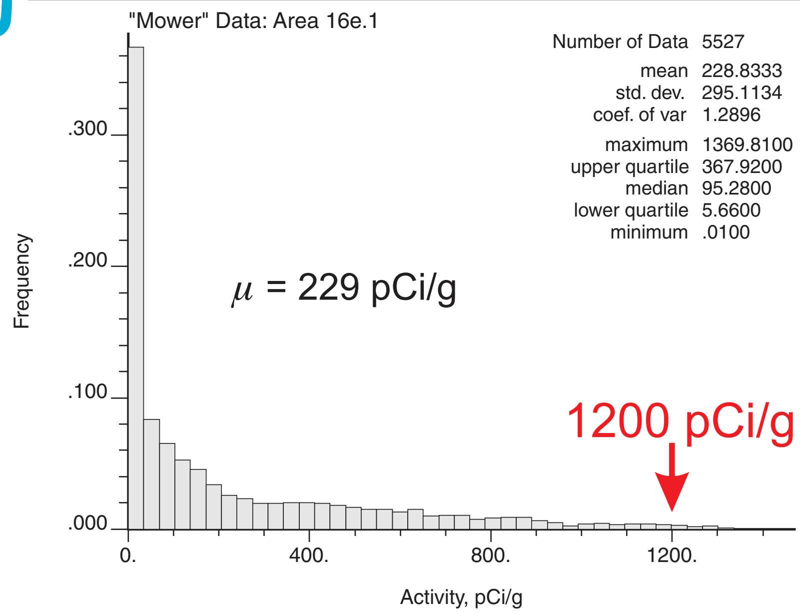


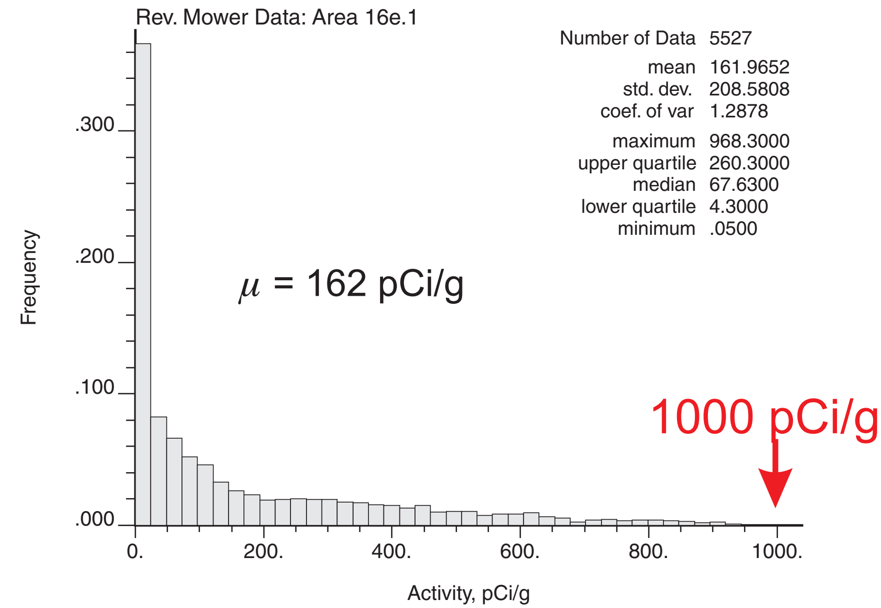
Prediction curves are of the form: $Y = ae^{-bx}$

Mower Radiometric Data

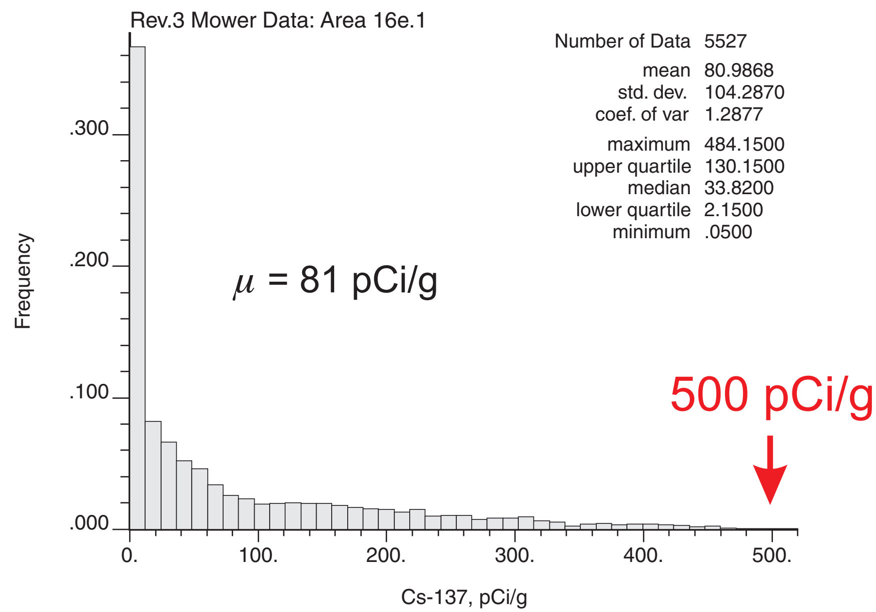


Mower Data Calibration

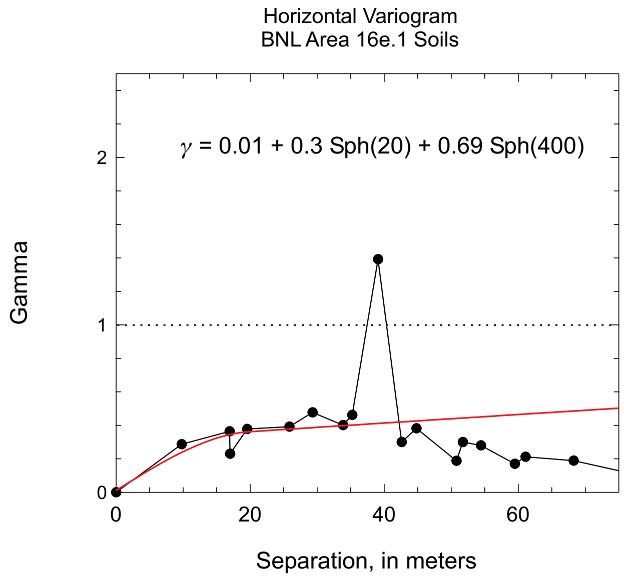


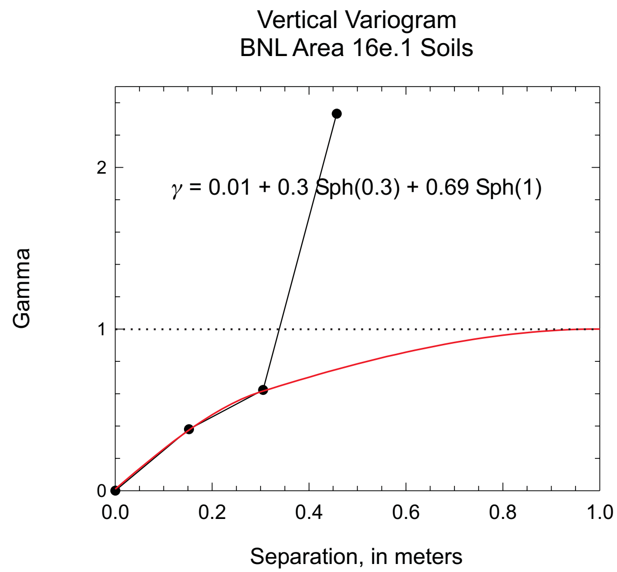


Successive Revisions of Soft-Data Calibration



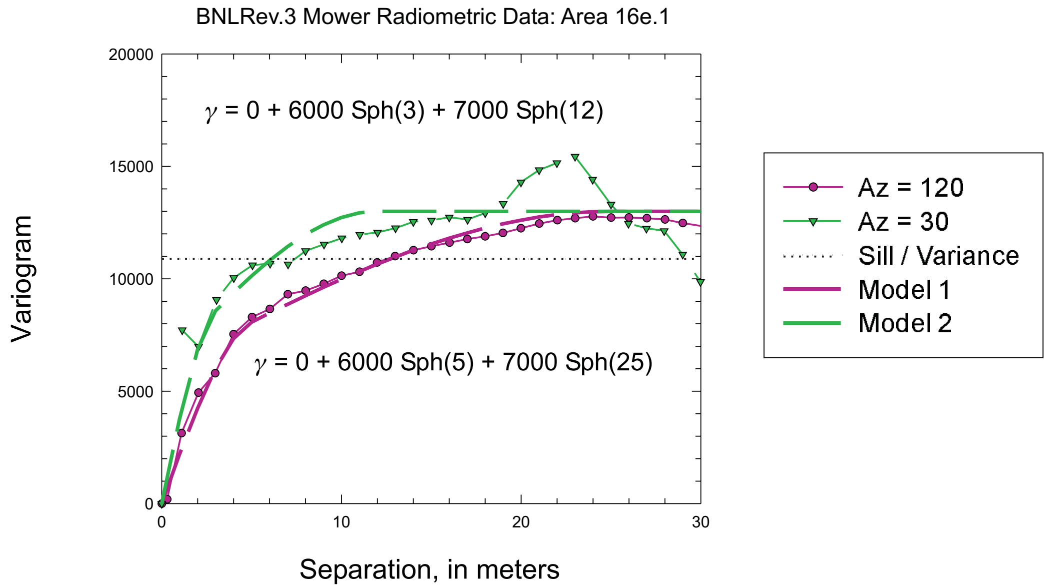
Variography



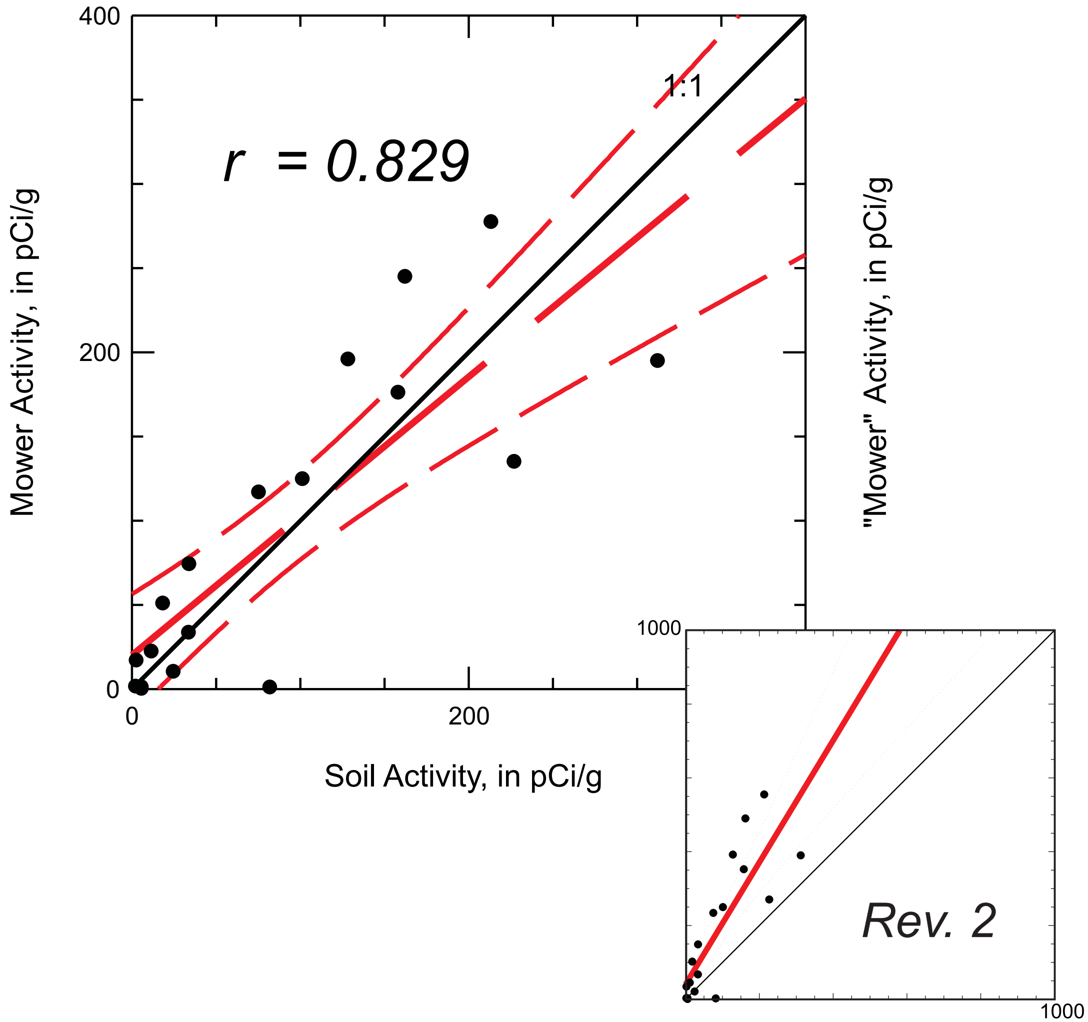


Soil Data in 3-D

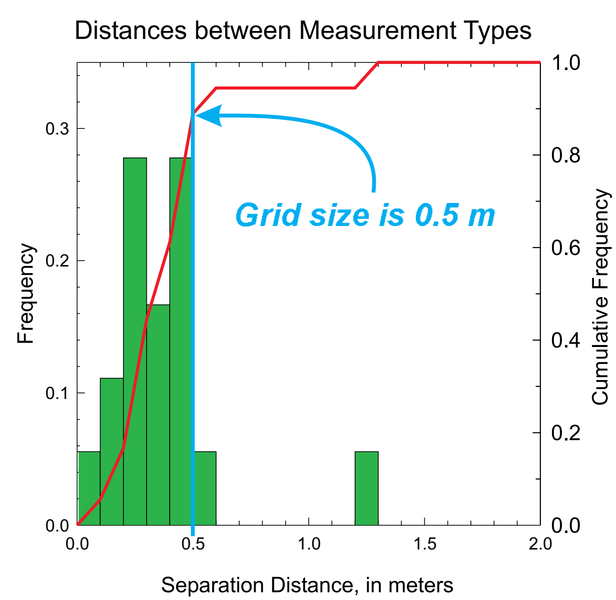
Mower Data in 2-D with Anisotropy



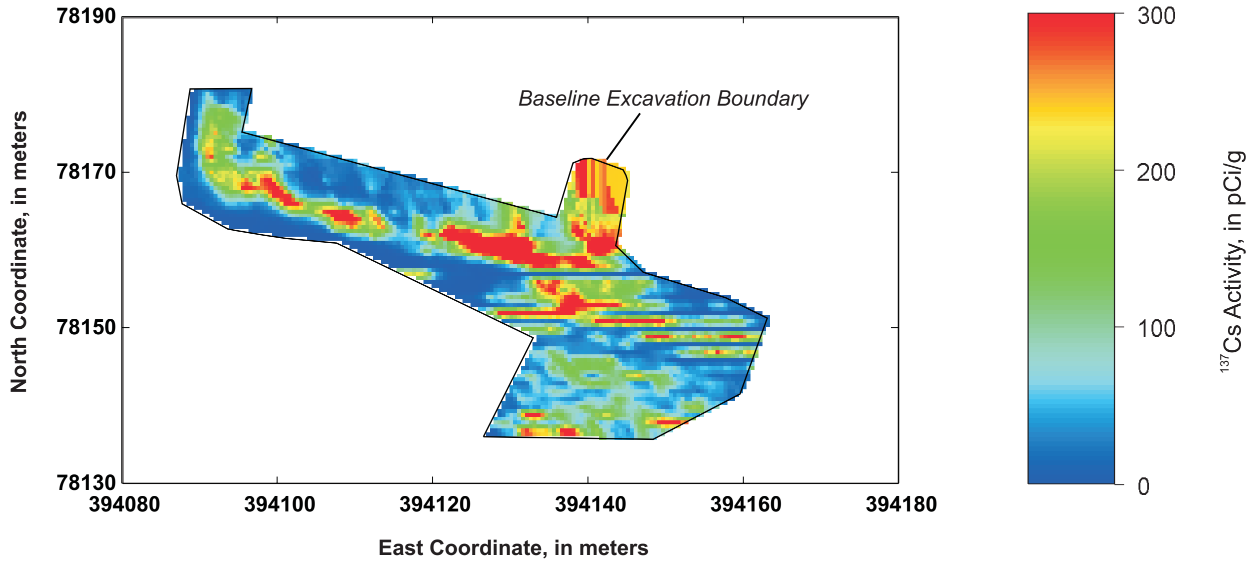
Correlation of Hard and Secondary Data



Nearest Neighbors Cross-plot



Generating "Exhaustive" Secondary Data



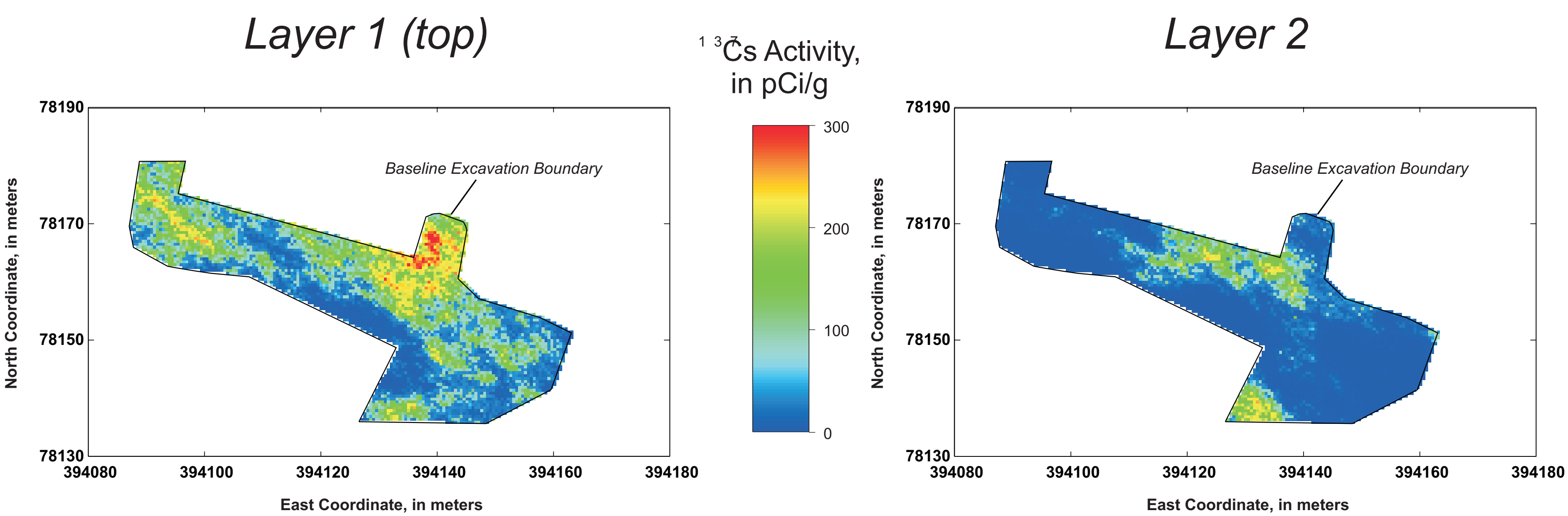
The 5,500 mower data were kriged onto the top layer of a regular 3-D grid and extrapolated to depth using the negative exponential decay function:

$$Y = ae^{-bx}$$

where Y is the secondary value on layers 1-4, a is the decay secondary value on the surface, x is depth, and b is a fitted coefficient representing the "depth-decay" of *all* soil values.

Replicate Simulations

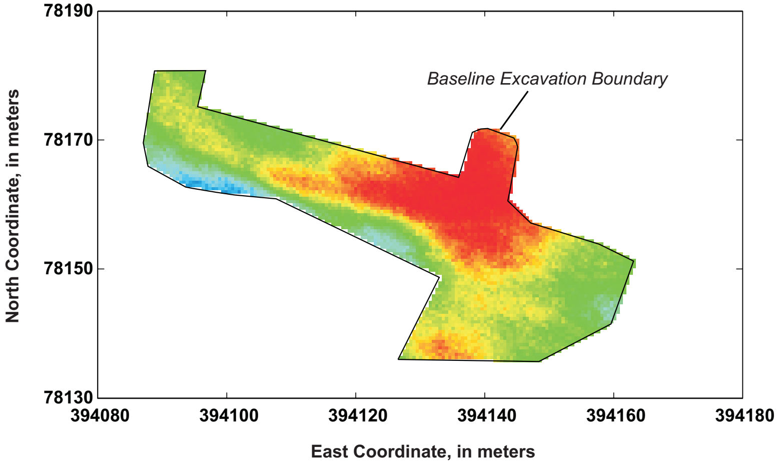
A suite 100 of statistically indistinguishable simulations was generated using sequential gaussian simulation conditioned to the 68 soil-profile measurements with collocated cokriging for the secondary data



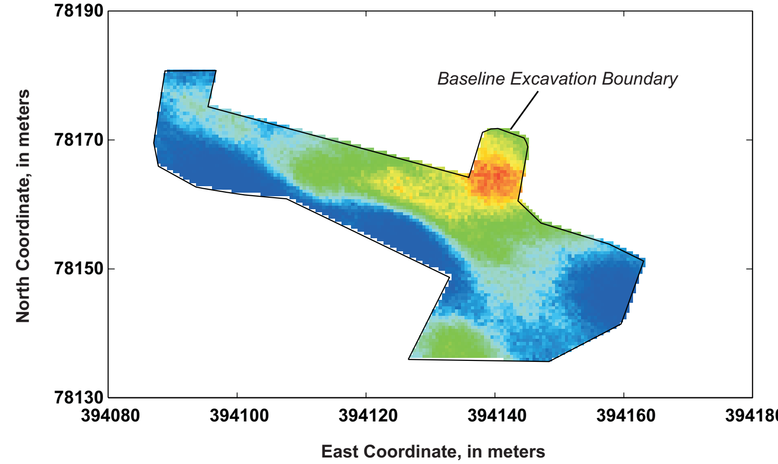
Layers 3 and 4 are essentially all blue (but not zero)

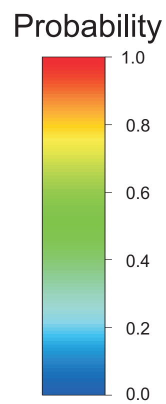
Probabilistic Summaries

BNL 16e.1; Pr{>23 pCi/g}, level 1

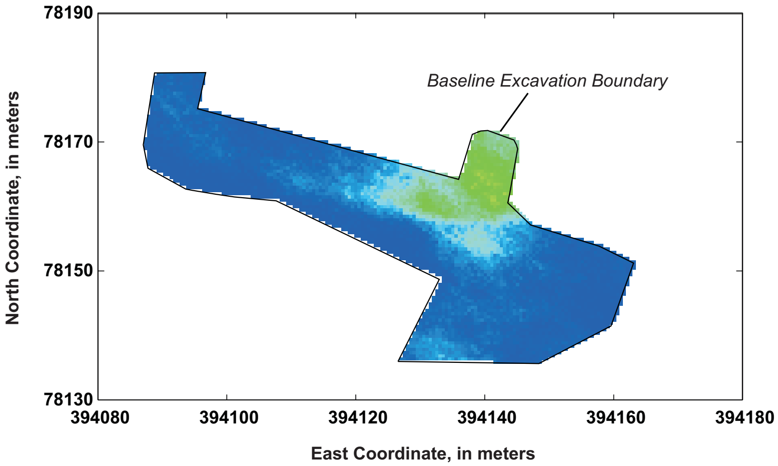


BNL 16e.1; Pr{>23 pCi/g}, level 2

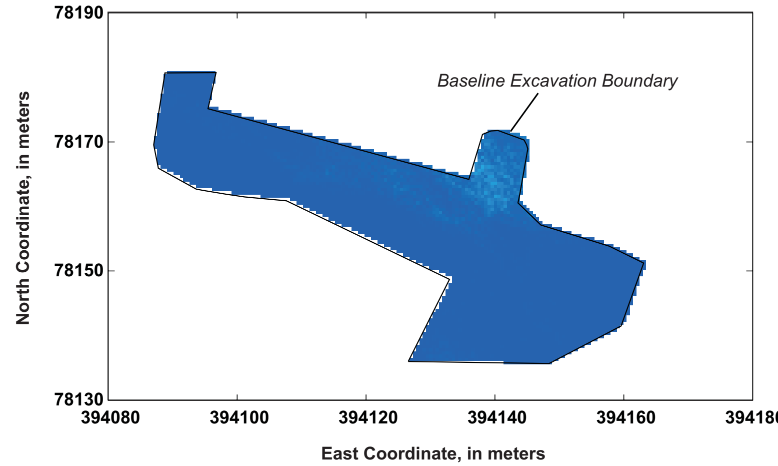




BNL 16e.1; Pr{>200 pCi/g}, level 1

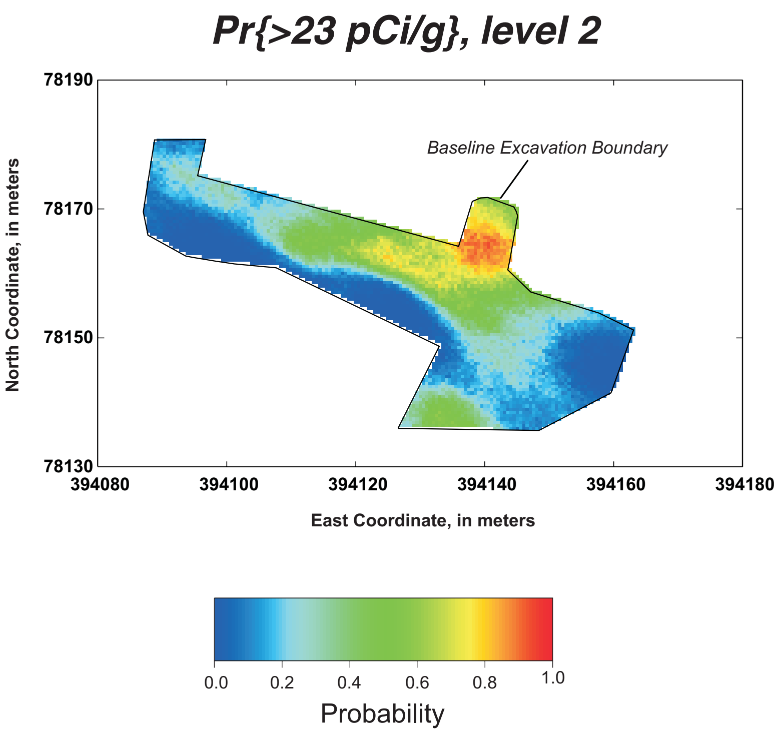


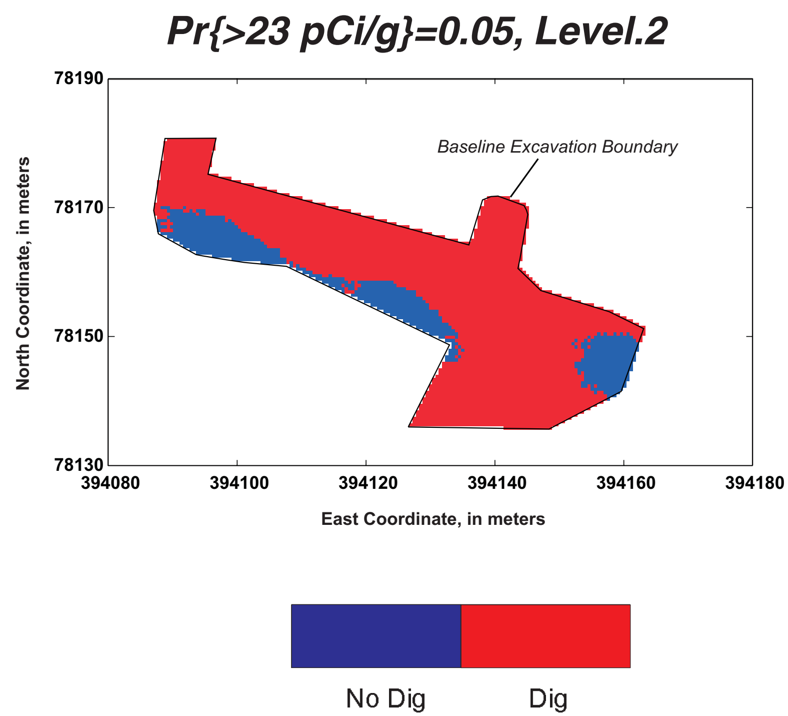
BNL 16e.1; Pr{>200 pCi/g}, level 2



Conversion to Excavation Maps

Excavation Maps represent concrete plans for remedial action, but require the determination of a ***Reliability Level*** acceptable to stakeholders



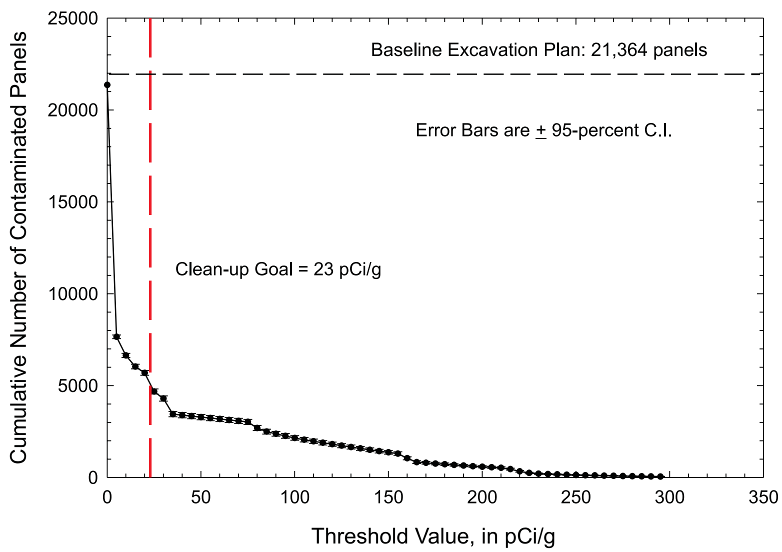


Level 2: Probability of Failure = 0.05; Reliability = 0.95

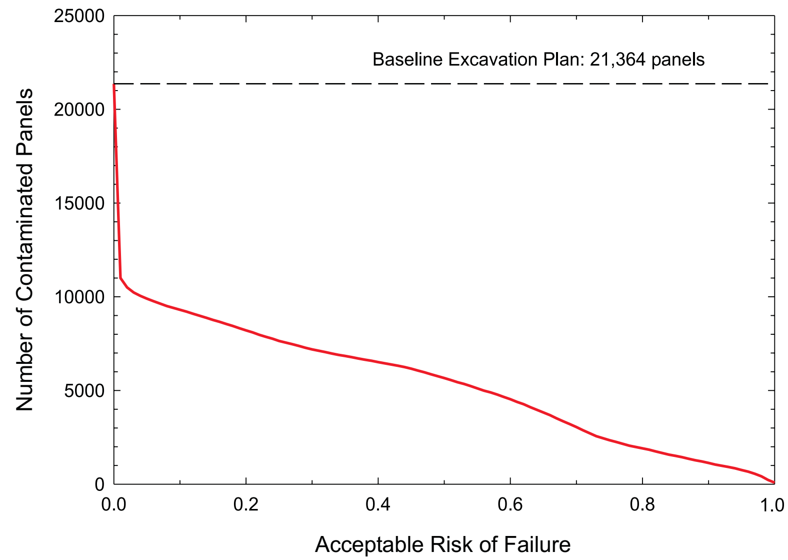
Extent of Contamination

Notice that the "Extent" of Contamination is a function both of the activity threshold selected ***and*** of the acceptable Risk of Failure (or Reliability) level demanded.

Remediation Panel: 0.5x0.5x0.1524 m
Region within excavation trimline only



by Threshold Activity



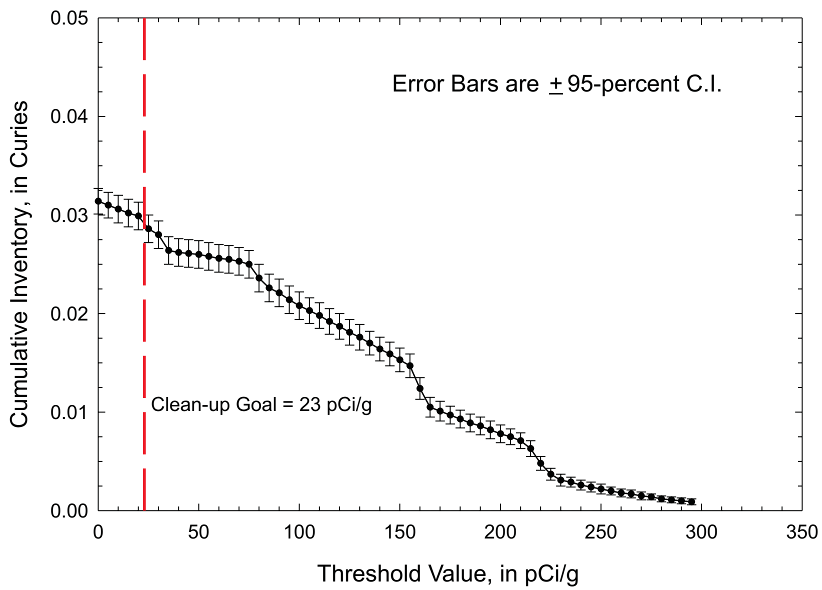
by Risk of Failure at 23 pCi/g

Number of 3-D Contaminated Panels (volumes)

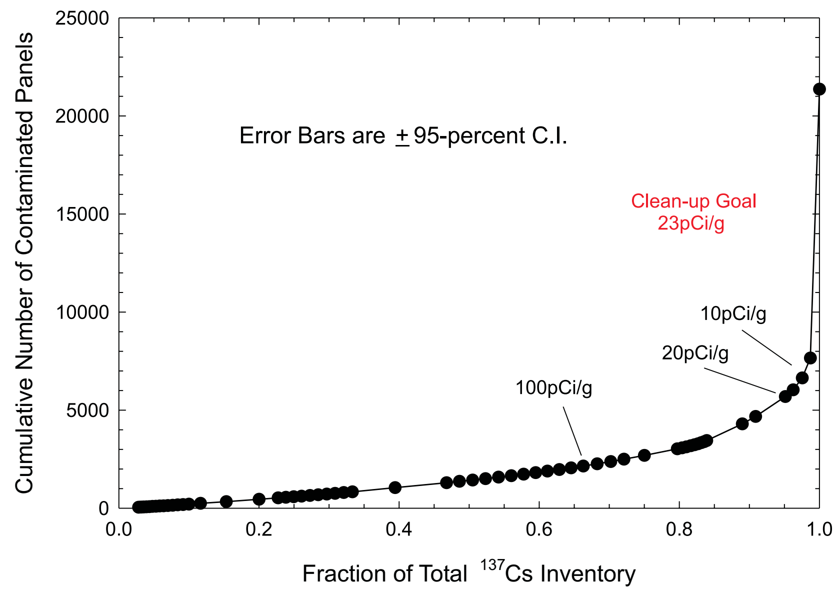
Quantitative Inventory Estimates

Conversion of Each Simulation to an Inventory Estimate:

$$\textit{TotalActivity} = \textit{PanelActivity} \times \rho_b \times \textit{PanelVolume}$$

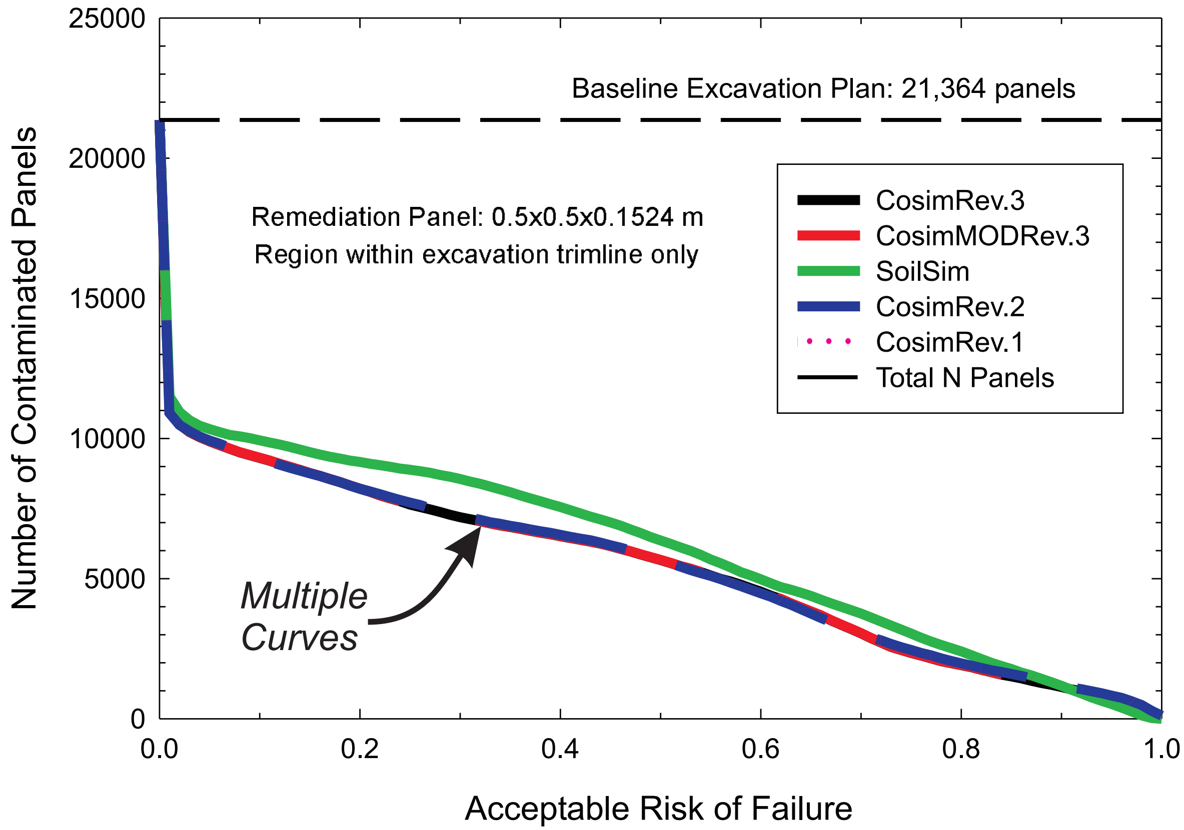


Inventory Curve



Fractional Inventory Curve

An Interesting Observation .....,



Because of the "recalibration" of the mower data, the models were computed three (3) different times. Yet the "extent of contamination" remained the same! ***Why?***

Q: Who Dunnit?

A: Collocated Cokriging

An examination of the formulation of collocated cokriging immediately reveals why the absolute magnitude of the secondary data does not and cannot influence the final magnitudes of the modeled values

Cokriging

The cokriging estimator is typically given as:

$$Z^{*}_{cok} = \underbrace{\sum_{\alpha 1 = 1}^{n1} \lambda_{\alpha 1}(\mathrm{u}) \cdot Z(\mathrm{u}_{\alpha 1})}_{\textit{Primary Data}} + \underbrace{\sum_{\alpha 2 = 1}^{n2} \lambda'_{\alpha 2}(\mathrm{u}) \cdot Y(\mathrm{u}_{\alpha 2})}_{\textit{Secondary Data}}$$

where one has knowledge of the joint spatial covariance functions:

$$C_Z(h),\ \ C_Y(h),\ \ and\ \ C_{ZY}(h)$$

after Deutsch and Journel, 1998, p. 73

Collocated Cokriging

The collocated cokriging estimator, which is used in computing the expected value of the conditional probability distribution in a simulation environment, is:

$$Z^*_{cok} = \sum_{\alpha 1}^{n1} \lambda_{\alpha 1}(\mathrm{u}) \cdot Z(\mathrm{u}_{\alpha 1}) + \lambda' \cdot Y(\mathrm{u})$$

Collocated Value

where the Z- Y cross covariance is estimated as:

$$C_{ZY}(h) = B \cdot C_Z(h)$$

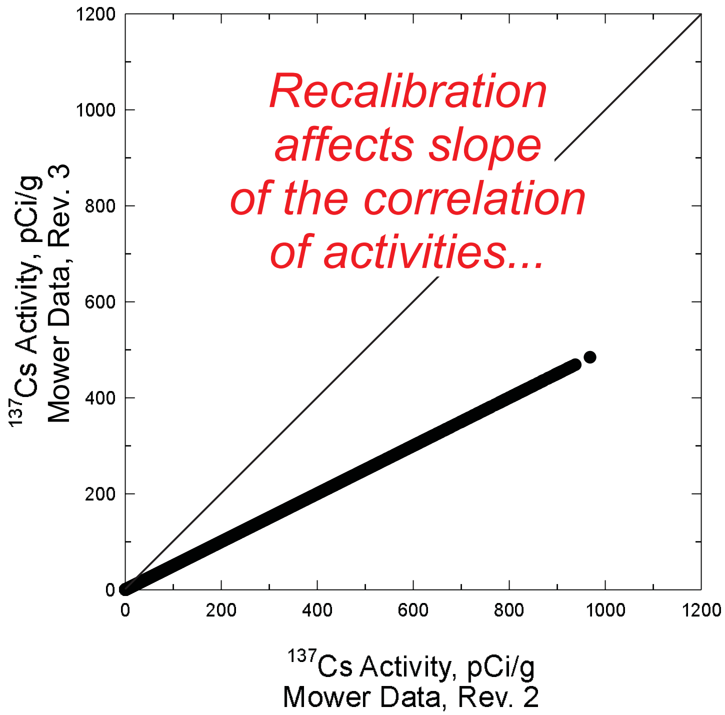
and where

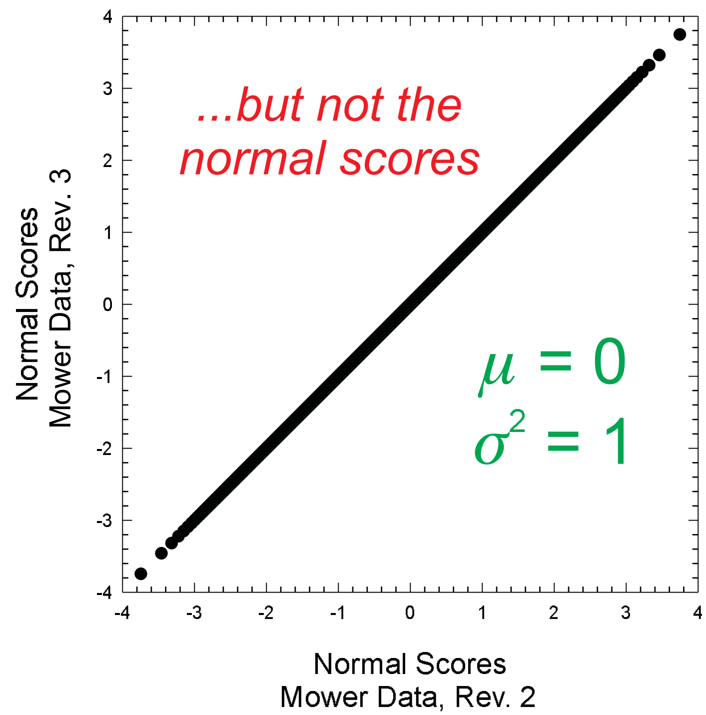
$$B = \sqrt{C_Y(0) / C_Z(0)} \cdot \rho_{ZY}(0)$$

after Deutsch and Journel, 1998, p. 74-75

Soft Data Used Only as Normal Scores

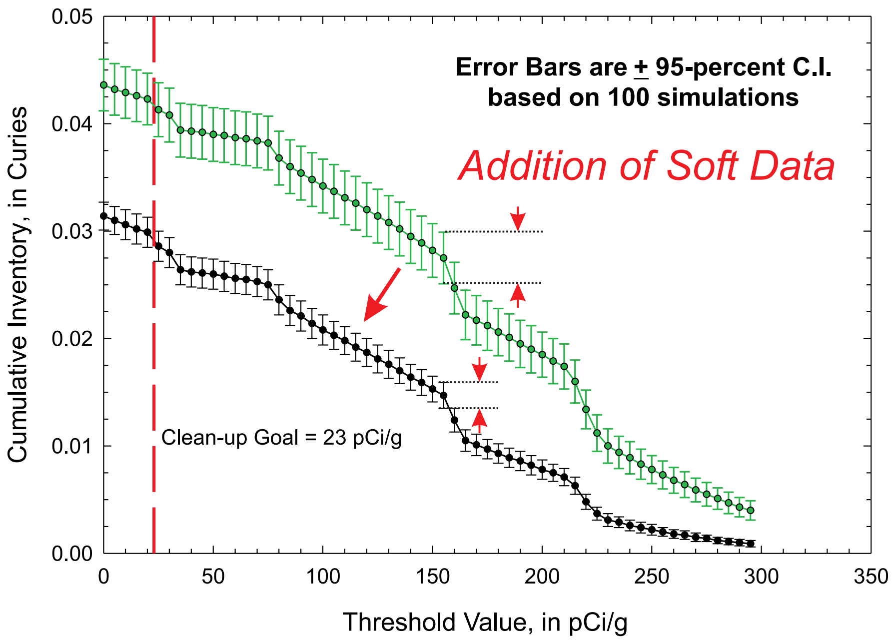
The several "recalibrations" of the mower data were simply linear rescalings of the count rates
==> *Normal-score values are unchanged.*



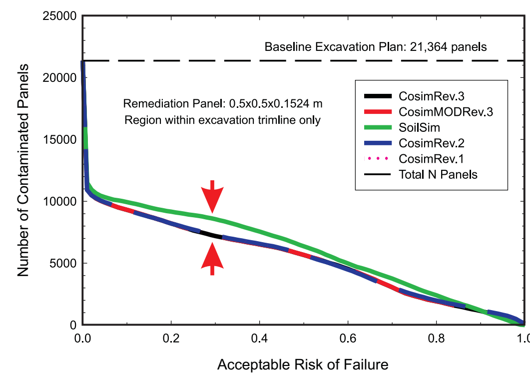


Comparison of Mower Data, Revs. 2 and 3

What Did We Gain with the Soft Data?



The simulation exercise also was performed using only the 68 soil-profile data (although the anisotropic variogram using the mower data was retained).



The information added by the secondary measurements has produced a smaller spread of simulated total inventory estimates (a smaller space of uncertainty).

The total estimated inventory was also reduced, although this modeling conclusion has not been validated

Conclusions

Re: the SGS -- a broad range of contaminant values are predicted to exist at the site of interest. This suggests that deployment of the SGS would be successful in reducing volumes of contaminated soil for offsite disposal.

"Soft" or secondary data can be incorporated into modeling of various physical quantities regardless of the "absolute accuracy" or precise calibration of the secondary quantity being measured.

This conclusion follows directly from the formulation of the collocated cokriging estimator of the expected value of the pdf in a simulation environment.

It also provides a mechanism for incorporation of ***indirect*** measurements relevant to the primary variable of interest.

Implications for Environmental Studies

It is probably "better" to invest scarce characterization resources in areally and/or volumetrically ***extensive*** measurements of lower precision than to allocate those same resources to spatially more sparse, ***intensive*** measurements of extreme precision.

These conclusions do required a certain degree of linearity to the response function of the secondary data *and* meaningful cross-variable correlation.

Incorporation of secondary data can be demonstrated to reduce the spread of uncertainty across a suite of otherwise similar simulated models.

Theory suggests that adding information would also increase accuracy of the models (closeness to the true value), but this has not been validated here.