Pacific Northwest National Laboratory
Operated by Battelle for the Energy
SAND2007-7001P
Sandia National Laboratories

Introduction to Kriging

Sean A. McKenna and Barry Roberts
Sandia National Laboratories

VSP-UXO Course
Huntsville, Alabama
November 6th-7th, 2007

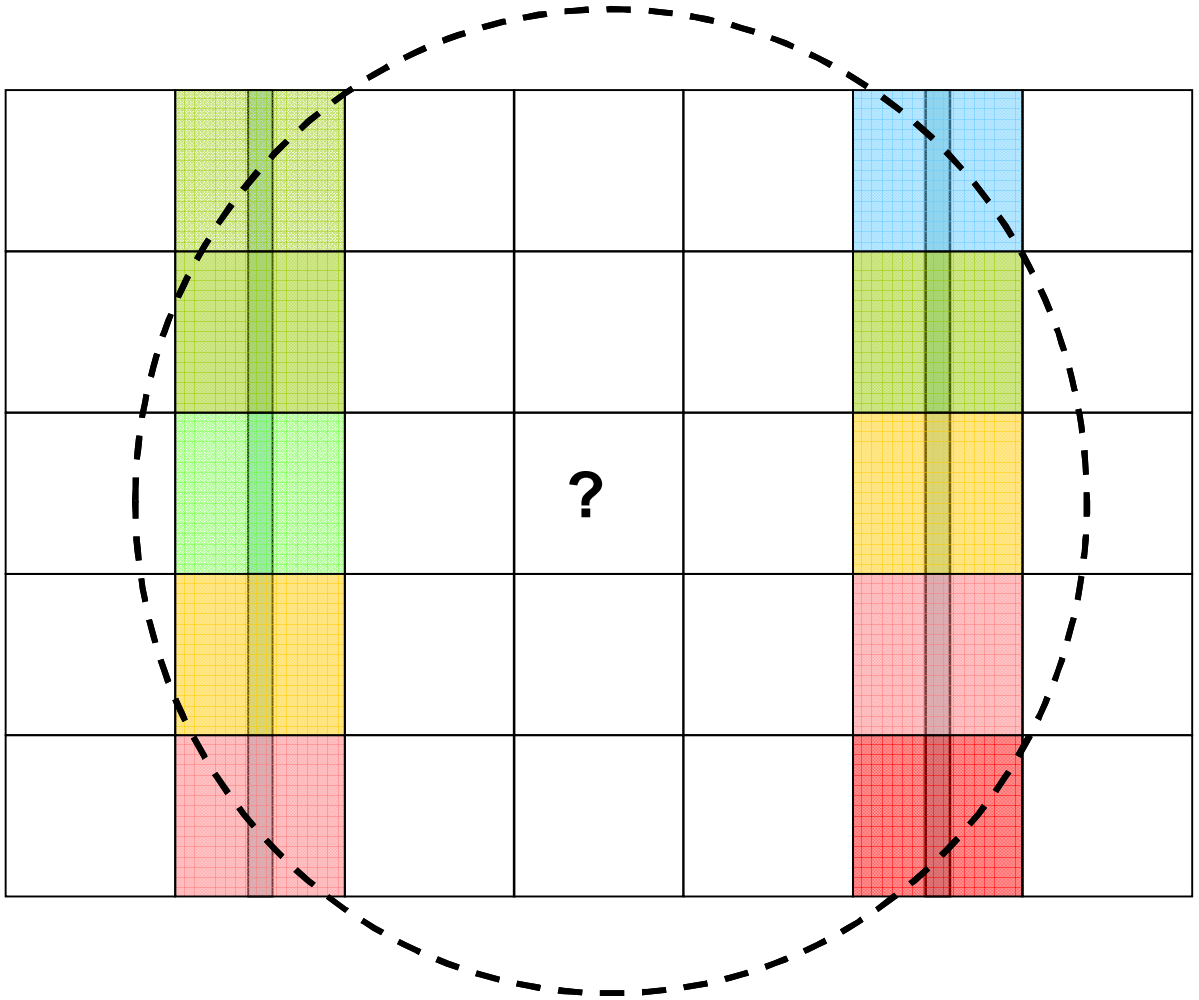
Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Spatial Statistics

- Two parts to spatial estimation:
  - 1) Define a model of spatial correlation, or spatial variation, for the data set (variogram)
  - 2) Use that model in an estimation algorithm to estimate the data values at unsampled locations

Geostatistics: Developed in mining industry beginning in 1950's. Now has wide application in the mining, petroleum and environmental areas for estimation of spatially varying properties

Spatial Estimation



Transect data have been scaled up to represent anomaly density at the scale of the estimation grid.

Each estimate is a weighted linear average of the surrounding transect data and provides the anomaly density estimate for that unknown cell

Kriging provides the estimate and the estimation variance

Kriging vs. Other Estimation

- Of spatial estimation algorithms, only Kriging provides:
  - Exact estimator (returns measured values at their measurement locations)
  - Non-convex estimator (Estimates are not constrained to minimum and maximum measurements)
  - Provides both an estimate and the estimation variance (what confidence do I have in my estimate)
  - Weights assigned to data in estimation use model of spatial variation (variogram – allows for anisotropy)
  - Weights assigned to data account for data redundancy

Kriging is B.L.U.E.

- Best
  - Distribution of residuals between estimates and true values is tight (minimum variance)
- Linear
  - Each estimate is a linear weighted average of surrounding data values
- Unbiased
  - Average residual about true values is zero
- Estimator

Kriging Process

- Define estimated value as a weighted linear average of the surrounding measurements

$$Z^*_{OK}(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i)$$

- Subject to:

$$\sum_{i=1}^{N} \lambda_i = 1.0$$

Calculation of Weights

Covariance between all pairs of existing data points

Unknown vector of weights

Covariance between each existing data point and the estimation location

$$C \bullet \lambda = d$$

$$\begin{bmatrix} C_{11} & C_{12} & : & C_{1N} & 1 \\ C_{21} & C_{22} & : & C_{2N} & 1 \\ : & : & : & : & 1 \\ C_{N1} & C_{N2} & : & C_{NN} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ : \\ \lambda_N \\ \mu \end{bmatrix} = \begin{bmatrix} C_{01} \\ C_{02} \\ : \\ C_{0N} \\ 1 \end{bmatrix}$$

$$\lambda = C^{-1} d$$

Extra row and column of 1's ensures that weights sum to 1.0

Adding row and column creates N+1 equations, for N unknowns, add Lagrange parameter, m, to be able to solve

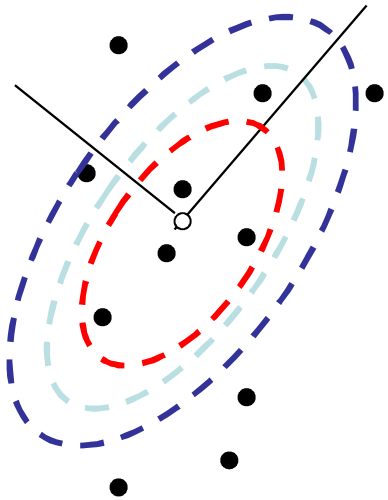
Calculation of Estimation Variance

- After solution for the optimal, minimum variance, weights, the OK kriging variance is:

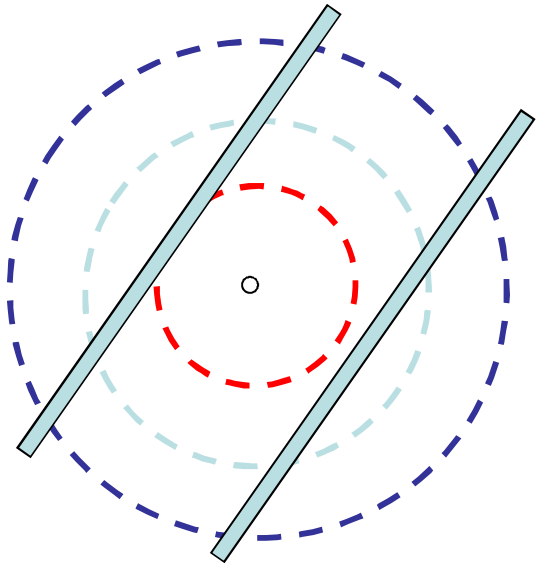
$$\sigma^2_{OK}(x_0) = Cov(0) - \sum_{i=1}^{N} \lambda_i Cov(x_0, x_i) - \mu$$

- A few points:
  - Measured data values do not directly enter into the calculation of the estimation variance (a function of variogram model and data configuration)
  - At large distances, Cov(*h*) goes to zero and estimation variance is equal to the variance of the data set (Cov(0))

Search Neighborhood Definition



Generally, define the search neighborhood as an ellipse oriented along the direction of maximum correlation



For geophysical transect data at UXO sites, try to get data from several transects into the search neighborhood.

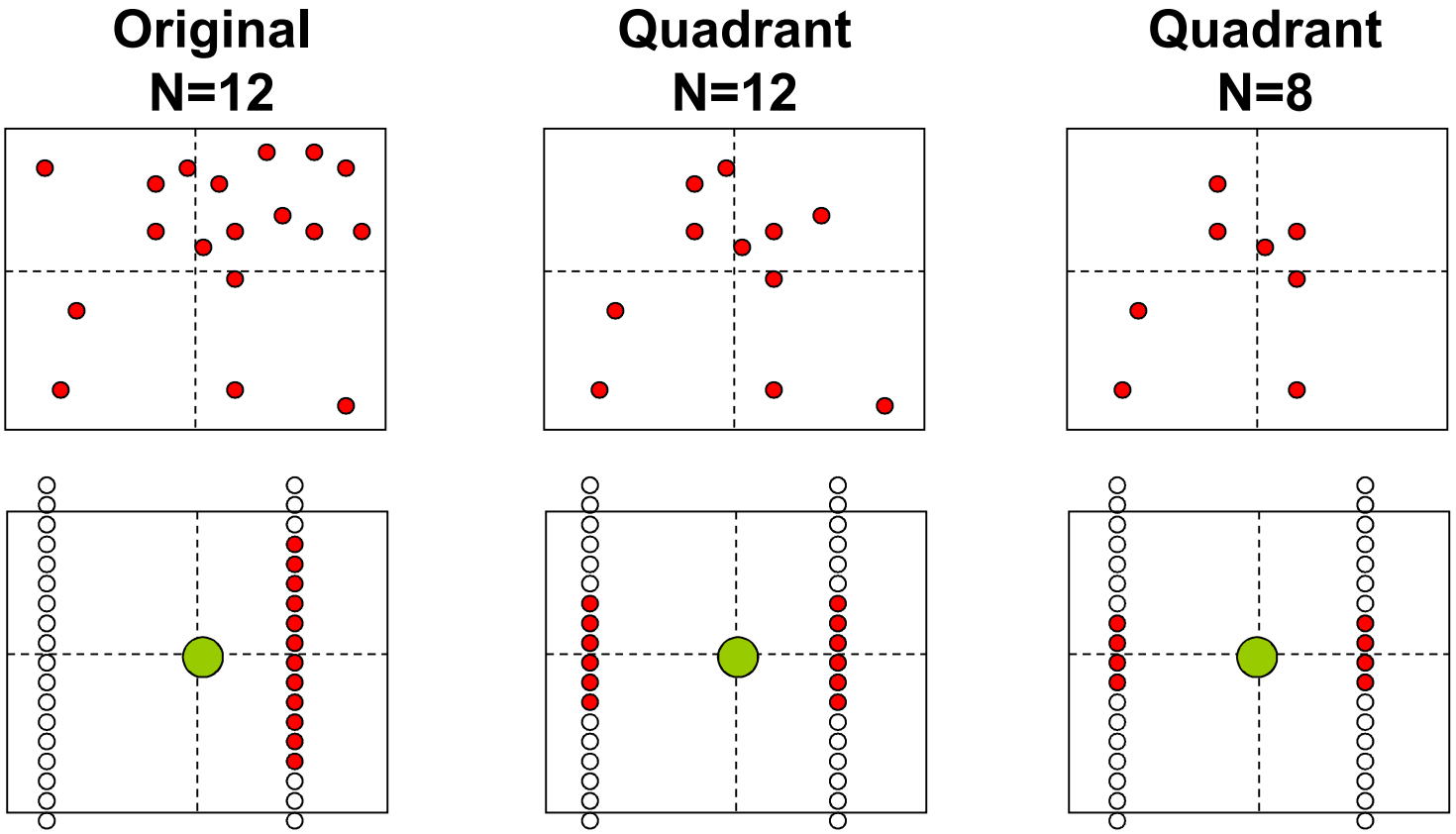
Isotropic search neighborhood has been appropriate for every site we have looked at so far

Too many samples?

- Also need to define maximum number of samples in neighborhood
  - Limit amount of computation
  - Maintain appropriateness of stationary random function model

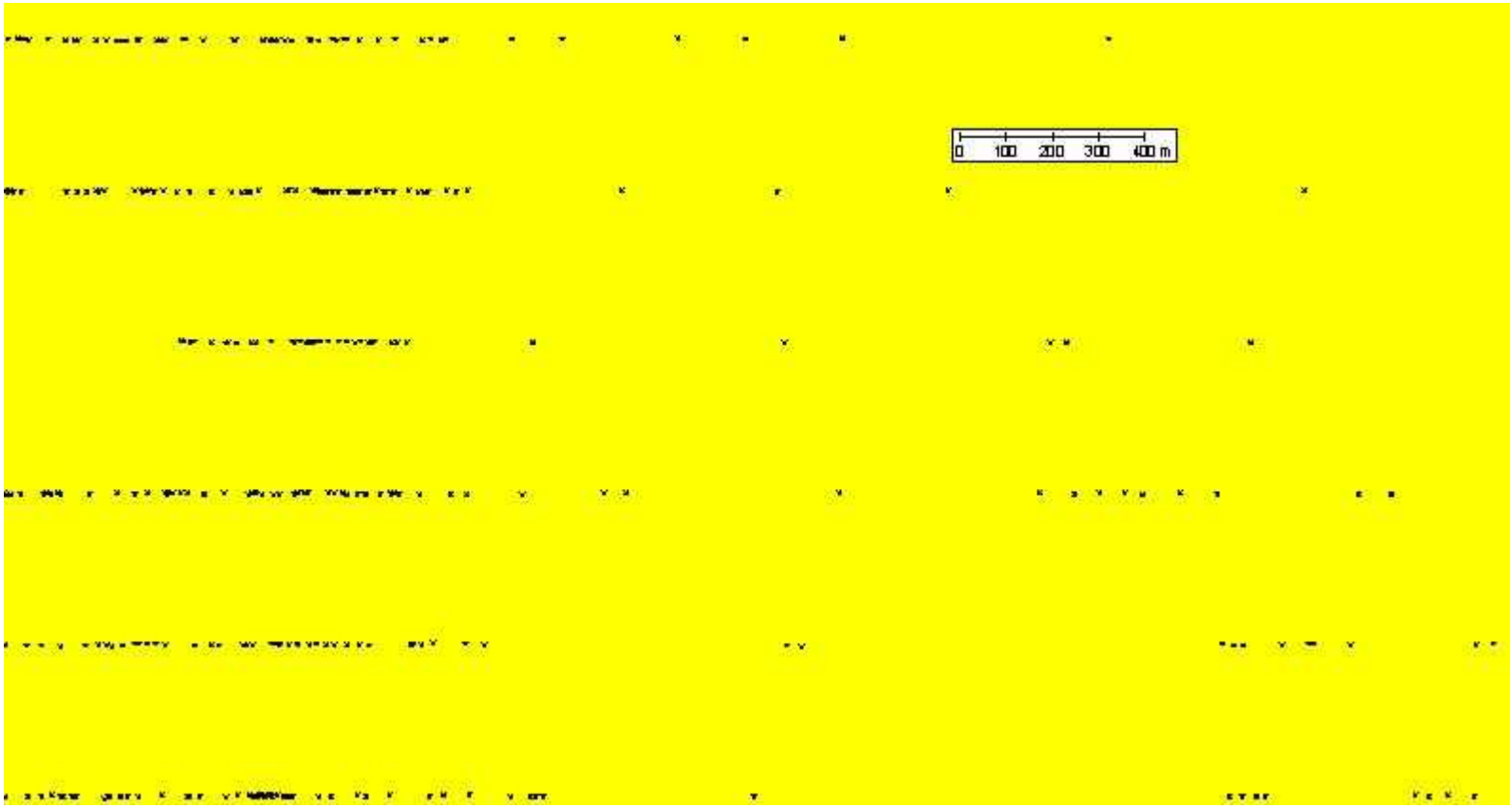
Redundancy

- Kriging takes care of redundant samples
- Quadrant/Octant searches can also be used



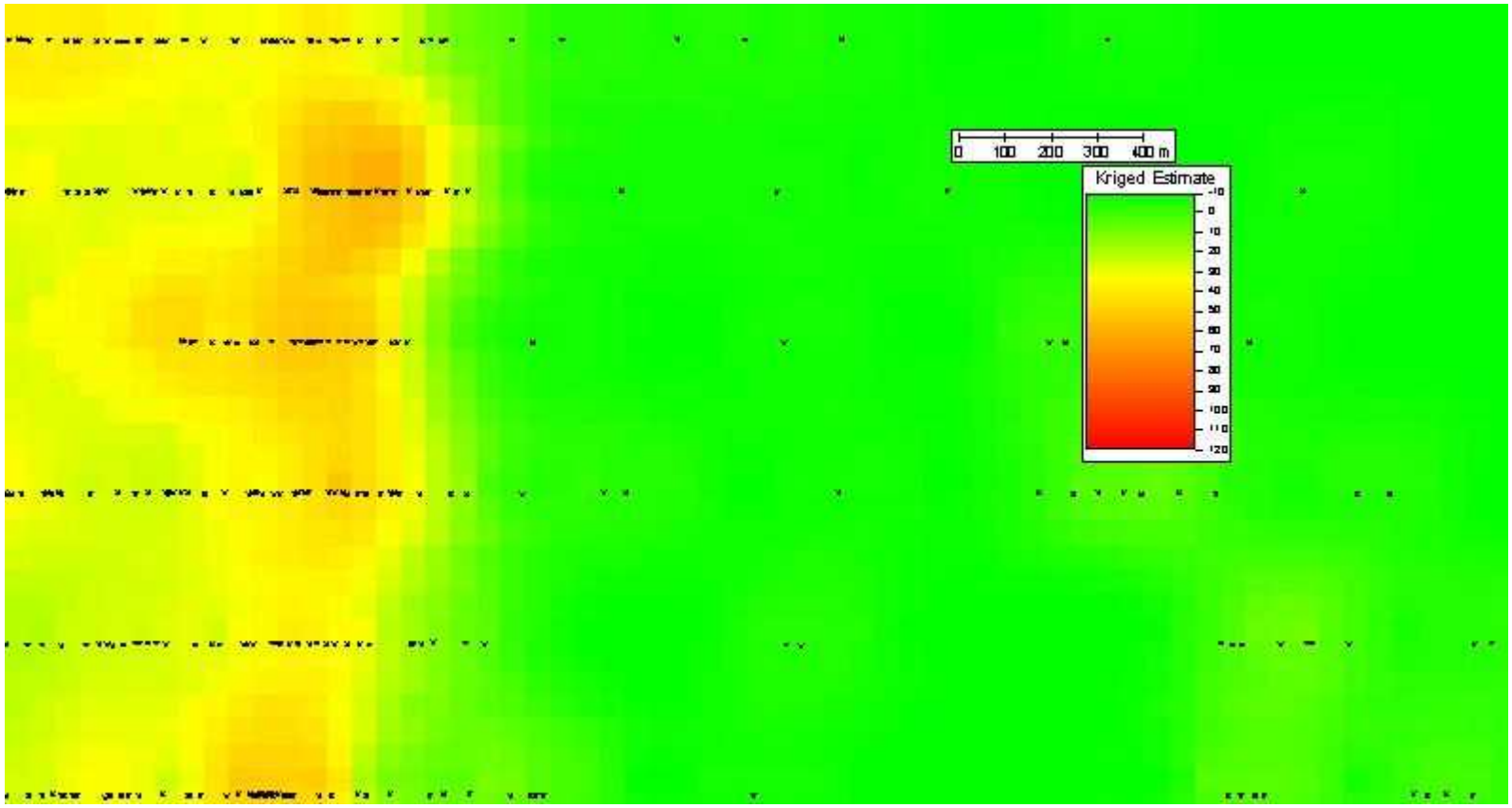
Estimation

From the anomaly locations along the transects, create a continuous estimate of anomaly density on the underlying grid



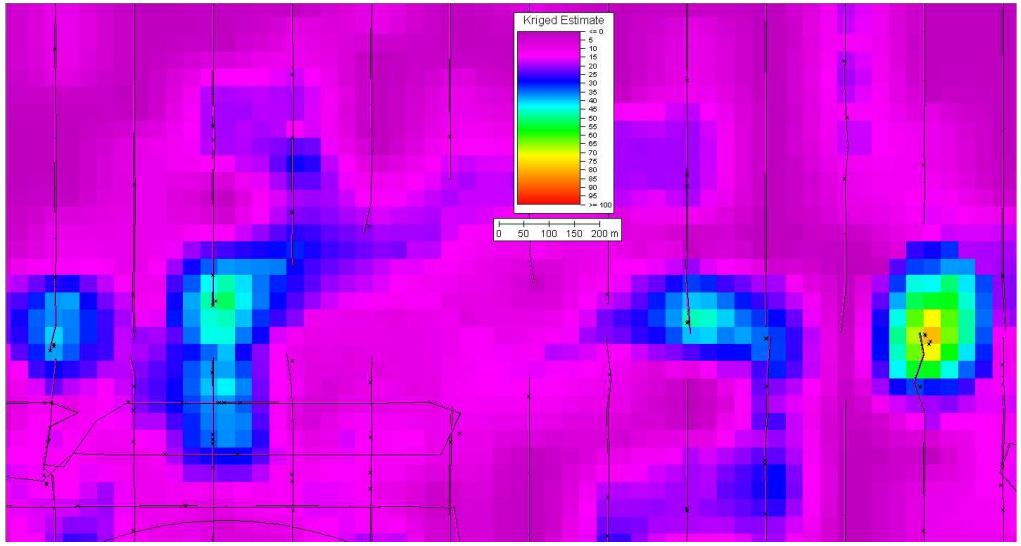
Estimation

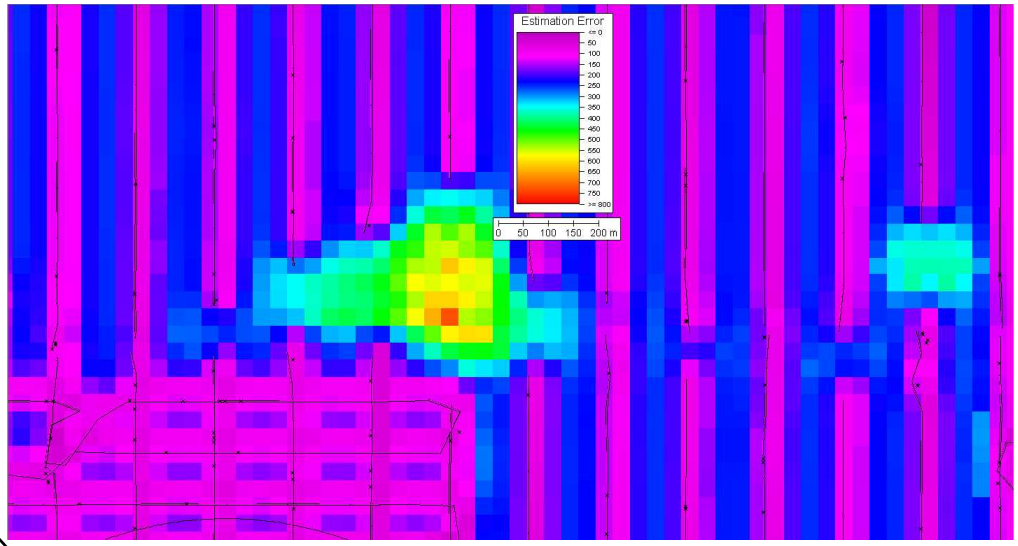
Anomaly estimates (per acre) for each 50x50m grid cell



Example 2

Kriged estimates on 50x50m grid, possible issues with estimates at end of transect





Estimation, or kriging, variance provides inverse measure of confidence in kriging estimates

How large is too large?

Summary

- Kriging provides optimal spatial interpolation
  - Estimated values at all locations
  - Estimation variance (uncertain of the map)
- Customization of kriging tools for UXO sites where data occur in transects
  - Data processing to get representative measure of anomaly density at all transect locations
  - Search neighborhood – will test in exercises