FY'03 Progress Report:

Strategic Environmental Research

and Development Program

Project Number: UX-1200

Bayesian Approach to UXO Site Characterization with Incorporation of Geophysical Information

Sean A. McKenna, Hirotaka Saito
Geohydrology Department
Sandia National Laboratories
PO Box 5800 MS 0735
Albuquerque, NM 87185-0735

December 19, 2003

Abstract

UXO site characterization approaches are developed to assist decision makers in determining where additional characterization efforts need to be expended and where additional characterization is not effective. These decisions are based on limited transect data and require that without 100 percent site characterization there is a finite probability of leaving some UXO behind. One theoretical limitation of geostatistical approaches to estimation is the assumption that sample data exist in an unbounded domain. Contiguous transect data, because of their close proximity to each other violate this assumption and can produce unwanted results in the estimates. The extent of these unwanted results are checked for a variety of transect sample designs on three simulated sites and the results show that the effects of the finite domain associated with transect data are negligible and traditional geostatistical estimation techniques can be applied.

Four example calculations are used here to demonstrate aspects of this spatial-statistics based approach to UXO site characterization. The first example demonstrates the ability of the spatial estimation techniques to estimate different attributes that might be of interest in a site characterization activity: the total number of anomalies, the number of anomalies of interest and the probability of at least one anomaly of interest. Prior to making the estimates, a cross-validation process is used to check the applicability of the variogram and kriging model for the specific application. These cross-validation results provide excellent predictions of the results of the final estimations. The second example demonstrates the ability of probability mapping to define the edge of a UXO target from a limited number of parallel transects representing three percent of the entire site. The third example extends the results of the first example by adding prior information in the form of discretizing the site into different strata each of which is thought to have undergone a similar site history. This stratified approach is perhaps the simplest way to incorporate prior information into the site characterization approach and has not previously been applied to UXO sites. Incorporation of the strata into the estimation procedure allows for extending the estimates made in the first example further away from the limited sample data to cover all portions of the site under the assumption that the mean values assigned to the strata are representative of each stratum all the way across the site. Results show that the estimation results of the number of anomalies, anomalies of interest and probability of at least one anomaly of interest across the site are consistent with the estimates made on smaller portions of the site in the first example. The fourth example demonstrates two methods for locating additional samples in a second phase of sampling. The two approaches are meandering paths in the area of highest uncertainty to better define the edge of the target areas and infill sampling to better characterize the entire site. Decision results between the two approaches are nearly identical even through the latter approach uses almost twice as many samples. However, both sets of results are not significantly different from the decision results made with just the original set of samples.

Table of Contents

Abstract 2
Table of Contents 3
Table of Figures 4
Table of Tables 6
Introduction 7
  Variogram 10
  Kriging 13
    Ordinary Kriging 13
    Indicator Kriging 14
    Finite Domain Kriging 15
    Finite Domain Kriging Example 18
Model Evaluation 23
Site Characterization Process 25
Demonstration Applications 26
  Application 1 26
    True Site and Sample Data 26
    Variograms 29
    Model Validation 32
    Estimation Results 36
    Application 1: Summary 40
  Application 2 42
    Pueblo of Laguna Site and Sample Data 42
    Mapping Signal Strength 43
    Probability Mapping 46
    Application 2 Summary 49
  Application 3 50
    Incorporating Secondary Information 50
    Residual Variograms 52
    Estimation 55
    Application 3 Summary 61
  Application 4 62
    Second-Phase Samples: Meandering Path 62
    Second-Phase Samples: Straight Transects 64
    Second-Phase Sampling Resutls 67
    Application 4 Summary 68
Conclusions 71
Acknowledgements 72
References 72

Table of Figures

Figure 1. Conceptualization of the transect sampling and different approaches to the aggregation of information at the sample cell scale. 9
Figure 2. Three different variogram models with the same parameters: nugget =0.0, sill = 1.0 and range = 100.0. 12
Figure 3: An example of experimental direct semivariogram of primary data with a spherical model fitted. 12
Figure 4. Kriging weights along a transect of data for estimation of a point off of the transect at a Y coordinate of 2500.0. The upper figure shows the kriging weights calculated using ordinary kriging and the lower image shows the kriging weights calculated with the finite domain kriging algorithm. 19
Figure 5. Distribution of anomalies simulated for testing finite domain kriging against ordinary kriging. The three sites are referred to as homogeneous (top), isotropic (middle) and anisotropic (bottom). The red dot in the lower two images shows the center of the target. 20
Figure 6. Comparison of the mean error (bias) and the mean squared error, top and bottom images repectively for estimates made with OK and FDK as a function of the percent of the site sampled. Results are presented for the homogeneous, isotropic and anisotropic sites shown in Figure 5. 22
Figure 7. Distribution of anomalies within the simulated site. The axes dimensions are in meters. 27
Figure 8. Histogram showing the distribution of the simulated signal strengths for all objects within the site domain. 28
Figure 9. Locations of sampling transects on the simulated site. The color scale denotes the total number of anomalies within each 25x25 meter cell along the transects 29
Figure 10. Experimental and model variogram fit to the total number of anomaly data obtained on the transects. 31
Figure 11. Experimental and model variogram fit to the number of anomaly data with geophysical signals above 3.0 nT/m obtained on the transects. 31
Figure 12. Experimental and model variogram fit to indicator data created for a geophysical signal threshold of 3.0 nT/m. 32
Figure 13. Proportions of different decision results as a function of the design reliability for the cross-validation results. 36
Figure 14. Estimates of the total number of anomalies within each model cell. 37
Figure 15. Estimated number of anomalies of interest. 38
Figure 16. Estimated values of the probability of at least one anomaly of interest. 39
Figure 17. Proportions of different decision results based on the map in Figure 16 for a range of RD. 40
Figure 18. Survey data for the Pueblo of Laguna N-11 target area site as sampled on the 15x48 foot decision cells. The maximum signal value within each decision cell is shown here. .. 43
Figure 19. Sample transect data for the Pueblo of Laguna N-11 site. Along each transect, the maximum analytic signal within each 15x48 foot decision cell is shown. 44
Figure 20. Variogram of the maximum analytic signal as calculated from the N-11 target area transect data. 45

Figure 21. Estimated maximum analytical signal values made using ordinary kriging and the transect data and variogram shown in Figures 19 and 20 respectively. Compare to the actual sampled data in shown in Figure 18. 45
Figure 22. Indicator variogram for the N-11 target area site and a geophysical signal threshold of 5.0 nT/m. 46
Figure 23. Probability of at least one geophysical anomaly with a signal strength greater than 5.0 nT/m across the site as estimated through indicator kriging. 47
Figure 24. The estimated target boundaries (blue) for four different levels of R_D. The anomalies of interest lying outside of the estimated target area are shown in red. 48
Figure 25. Proportions of decision results and anomalies of interest found and left behind for R_D values between 0.70 and 1.00. 49
Figure 26. Schematic representation of the hypothetical site with three different strata representing: low (blue), medium (green) or high (red) suspected relative intensities of anomalies based on historical data. 52
Figure 27. Residual variograms for the total number of anomalies (upper image), anomalies of interest (middle image) and indicator (lower image) values. 54
Figure 28. Estimated residual values for the total number of anomalies. 55
Figure 29. Estimated residuals for the anomalies of interest. 56
Figure 30. Estimated residuals for the probability at least one anomaly of interest. 56
Figure 31. Estimates of the total number of anomalies as created using prior information. 58
Figure 32. Estimates of the number of anomalies of interest as created using prior information. 59
Figure 33. Estimates of the probability of at least one anomaly of interest as created using prior information. 60
Figure 34. Proportions of different decision results based on the map in Figure 33 for a range of R_D. 61
Figure 35. Locations of maximum uncertainty in the presence or absence of an anomaly of interst at the Laguna N11 site. The red regions contain cells with probabilities of at least one anomaly of interest between 0.4 and 0.6. 62
Figure 36. Original straight transects and the meandering path transect obtained in the second round of sampling. The log10 color scale shows the maximum analytic signal strength in nT/m within each sample cell. 63
Figure 37. Updated probability map showing the probability of at least one anomaly of interest at every location. This map was updated from the original map by incorporating a meandering path transect in the region of greatest uncertainty. 64
Figure 38. Straight and parallel sampling transects obtained in the original and second round of sampling. The log10 color scale shows the maximum analytic signal strength in nT/m within each sample cell. 65
Figure 39. Updated indictor variogram calculated from the original and second round of straight transects. 66
Figure 40. Updated map showing the probability of at least one anomaly of interest everywhere at the site based on the original and second round of straight transects. 66
Figure 41. Decision results as a function of R_D for the initial sampling and two different approaches to the second iteration of sampling for the Laguna N-11 site. 69

Figure 42. Decision maps for four different levels of R_D: 0.70 (top); 0.80 (second from top), 0.90 (second from bottom), 0.95 (bottom). The results in the left column are for the straight transects. The right column has the results for the meandering path. 70

Table of Tables

Table 1. Variogram model parameters for the three different data sets..... 32
Table 2. Cross-validation results of estimation of total number of anomalies. 34
Table 3. Cross-validation results of estimation of the number of anomalies of interest. 34
Table 4. Results of the jackknifing assessment of the total number of anomalies estimation..... 37
Table 5. Estimation results for the anomalies of interest..... 38
Table 6. Means of each quantity within the three different strata. 52
Table 7. Variogram model parameters used to fit the variograms of the residuals. 53
Table 8. Results of the estimation of the total number of anomalies using prior information as determined through jackknifing..... 57
Table 9. Results of the estimation of the number of anomalies of interest using prior information as determined through jackknifing..... 59

Introduction

This report documents work done in the third year of the UX-1200 SERDP project investigating the use of spatial statistical, or geostatistical, techniques to accomplish more efficient UXO site characterization. This work is aimed at providing approaches to decreasing the amount of a large site that must undergo detailed geophysical surveys by extrapolating information from a limited amount, typically less than 10 percent of the site, of detailed surveying. The extrapolation of this information must also include some indication of the confidence with which that extrapolation can be made. Approaches to detailed survey area reduction done through geostatistical estimation are demonstrated in this report using four different example applications. These example applications were chosen to demonstrate a broad range of calculations that can be made in support of site characterization decisions. Results of these example estimations are assessed through comparison to additional data that were held back from the initial estimations.

The techniques and the example applications presented in this report build on previous work done for this project. The focus of the work done in 2001 was to define an approach to making site characterization decisions based on mapping the probability of at least one anomaly of interest with geostatistical techniques and coupling that map with a specified design reliability to make characterization decisions. Prior information was incorporated into the mapping using both cokriging and kriging with a locally varying mean. Example applications were done on simulated data sets, the N-10 target area at the Pueblo of Laguna in New Mexico and the Stronghold site in South Dakota. Work in 2002 focused on the effect of the ROC curve on the final geostatistical estimations. This work used both kriging with a locally varying mean and collocated cokriging to integrate prior information. Example calculations were done using simulated sites and a surveyed target area from the Pueblo of Isleta in New Mexico.

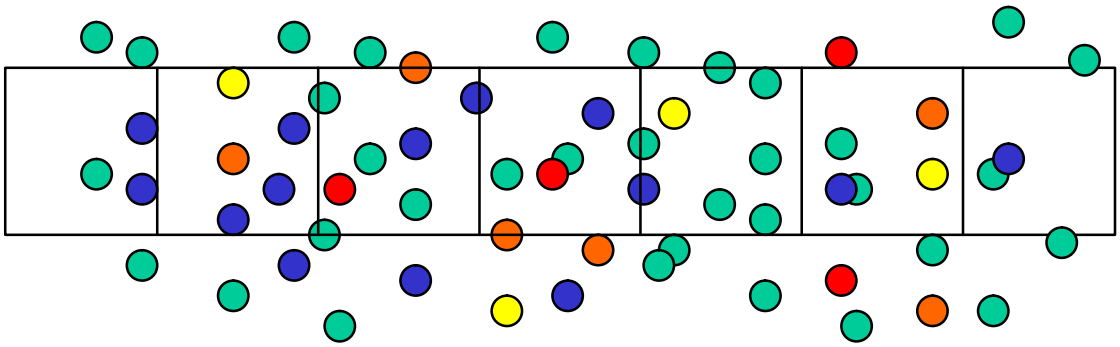
The techniques presented in this report are general in the sense that they can be applied to estimation of any measured attribute of interest. In the characterization of UXO sites, the measured attribute of interest may be the total number of anomalies, the number of "anomalies of interest" or the probability of having at least one anomaly of interest at any unsampled location (Figure 1). An anomaly of interest is defined here as some geophysical anomaly that has been identified as worthy of additional investigation. These anomalies may be identified as being above some threshold geophysical signal (e.g., a magnetic gradient equal to or greater than 10 nT/m) or possessing some degree of fit to a physical or statistical model of an ordnance item believed to be at the site (e.g., the measure of fitness computed through inverse modeling with comparison to the true UXO shape). Research on defining these threshold or fitness measures for UXO discrimination is outside the scope of this work, but examples of such efforts are available (e.g., Tantum and Collins, 2001; Paison, et al., 2002).

The samples collected at UXO sites are done along transects that have a fixed width, or footprint, and may be of varying length in a single direction along a straight transect or along a meandering path. For the analyses presented here, the transect is broken up into equal area cells, either square or rectangular, where at least two sides of each cell are equal in length to the transect footprint. Within each of these cells, the different attributes can be measured and a value assigned to the cell. Unlike nearest neighbor approaches (e.g., Byers and Raftery, 1997) that utilize the spatial relationships between the locations of the individual objects, the geostatistical

approach presented here aggregates or summarizes all sub-cell information to the size of the sample cell.

Figure 1 shows a conceptual model of this sampling and aggregation approach. The sampling transect can be represented as a series of contiguous square or rectangular cells arranged in a line. The geophysical anomalies in the area of the transect are shown as colored circles in Figure 1. The color of each circle represents the strength of the geophysical signal or the fitness of the anomaly with respect to its similarity to an ordnance model. Here cooler colors represent lower signal strength / fitness and warmer colors represent higher values. The same sample cells are shown in the middle of Figure 1 where the number in each cell represents the total number of anomalies within that cell. The sample cells are shown again in the bottom of Figure 1 where the number within each cell now refers to the number of anomalies of interest found within each cell. In Figure 1, anomalies of interest are those colored red at the top of the figure. The information aggregated to the cell size is assigned spatial coordinates equal to those of the center of the cell. These spatially referenced data can now be used in the geostatistical analyses conducted to characterize the site.

There are three main areas of focus for the work done in 2003: 1) The controlled testing of the site characterization approaches developed in the project on simulated data sets created by Mitretek; 2) demonstration of the flexibility of the site characterization approach to create spatial estimates of different attributes that might be measured on a site; and 3) the use of probability mapping to define the edges of targets and to locate additional sampling transects. The first area of focus is still ongoing and will be documented in a separate report. The other two areas of focus are presented here through the use of four example calculations using a simulated site and data collected at the Pueblo of Laguna N-11 site in New Mexico. The focus of these calculations is to provide a wide range of examples in which geostatistical methods can be applied to UXO site characterization problems. In addition to these example applications, a theoretical limitation of the basic geostatistical estimation algorithm, kriging, caused by redundant data as collected along transects is investigated. Compared to work of the past two years, more emphasis is placed on target edge delineation and locating second-phase sample transects and less emphasis is placed on techniques for integration of prior information, although a very simple and previously untested approach for prior information integration is examined.



Total number of anomalies within each cell

3	5	6	4	8	5	2

Number of anomalies of interest within each cell

0	0	1	1	0	0	0

Figure 1. Conceptualization of the transect sampling and different approaches to the aggregation of information at the sample cell scale.

Geostatistics

Geostatistics is the study of spatially correlated data as well as a set of tools developed to quantify the spatial variability of sample data and adaptations to regression theory to allow information on spatial variability to be used in estimating values at unsampled locations based on a finite number of existing sample data. Originally developed for ore reserve estimation in the mining industry, geostatistical techniques have become widely accepted and deployed throughout the earth and environmental sciences. Details on the theory and application of geostatistics to a wide variety of problems can be found in: Deutsch and Journel (1989) Goovaerts (1997) Isaaks and Srivastava (1989) Journel and Huijbregts (1978) and Olea (1999).

Variogram

The fundamental building block of a geostatistical analysis is the quantification of the variability of the sample data as a function of the average distance between any two data points. This distance to variability relationship is captured by the variogram, or more formally, the semi-variogram. The variogram has been used for description of spatial patterns (Western and Blölschl, 1999), variance mapping (Rouhani, 1985), and generation of realizations of spatial processes (stochastic simulation) (McKenna, 1998).

In practice the experimental variogram is computed as one-half the average squared difference between the components of data pairs:

$$\hat{g}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h})]^2 \qquad (1)$$

where N(**h**) is the number of pairs of data locations a vector **h** apart. The result of applying equation 1 to a data set is a set of discrete points that define γ as a function of the separation distance, **h**, or $\gamma(\mathbf{h})$. Multivariate geostatistics, which is an extension of univariate geostatistics, allows for incorporation of secondary variables into the variogram calculation and modeling, and the cross-semivariogram is usually required for further analysis (e.g., cokriging). The cross-semivariogram is a measure of joint variation of two attributes z_i and z_j, and the experimental cross semivariogram is computed as:

$$\hat{g}_{ij}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{a=1}^{N(\mathbf{h})} [z_i(\mathbf{u}_a) - z_i(\mathbf{u}_a + \mathbf{h})] \cdot [z_j(\mathbf{u}_a) - z_j(\mathbf{u}_a + \mathbf{h})] \qquad (2)$$

The experimental (direct- or cross-) semivariograms are used to describe spatial patterns of attributes, but it is rarely a final goal.

While the discrete points of the variogram calculated with a single or with multiple variables, define the semi-variance of the data as a function of separation distance, they alone cannot be used for spatial estimation in kriging algorithms. Spatial estimation requires that the variogram be defined at all separation distances. Therefore a continuous model of the spatial variability is fit to the points of the experimental variogram. Automatic model fitting algorithms exist (e.g.,

Wingle et al., 1999); however, most practitioners rely upon fitting a variogram model to the points of the experimental variogram by hand.

The choice of variogram model is, in practice, limited to a small number of analytical expressions. A constraint on the function used to model the variogram is that it must produce a positive-definite covariance matrix in the system of kriging equations. A positive definite covariance matrix ensures that there is a unique solution to the kriging equations and that the solution will produce non-negative kriging variance.

The three most commonly used variogram models are the spherical, exponential and Gaussian models. Each of these models is defined by two parameters: the sill, C, and the range, a. Theoretically, the sill is equal to the variance of the entire data set and represents the amount of variability between any two points that are uncorrelated. Alternative notation for the sill is the covariance of the data at h=0, C(0). The range is the separation distance beyond which the data are no longer positively correlated.

The spherical model is:

$$?(\mathrm{h}) = \mathrm{C} \cdot \left[1.5\frac{\mathrm{h}}{\mathrm{a}} - 0.5\left(\frac{\mathrm{h}}{\mathrm{a}}\right)^3 \right] \qquad for \quad h < a$$

$$?(\mathrm{h}) = \mathrm{C} \qquad for \quad h \geq a$$

The Exponential model is:

$$?(\mathrm{h}) = \mathrm{C} \cdot \left[1 - \mathrm{e}^{-\frac{3\mathrm{h}}{\mathrm{a}}} \right]$$

The Gaussian model is:

$$?(\mathrm{h}) = \mathrm{C} \cdot \left[1 - \mathrm{e}^{\left(-\frac{(3h)^2}{a^2}\right)} \right]$$

Examples of all three variogram models with a range of 100 and a sill of 1.0 are compared in Figure 2. The spherical model reaches the sill value at a distance equal to the range. The exponential and Gaussian models reach 95 percent of the sill value at a distance equal to the range and then asymptotically approach the full value of the sill. Typically, when referring to the exponential and Gaussian models, the distance at which the model reaches 95 percent of the sill is referred to as the “practical range” of the model. The Gaussian model applies to data that vary smoothly at short separation distances. The exponential model applies to data that exhibit stronger variability with increasing separation distance.

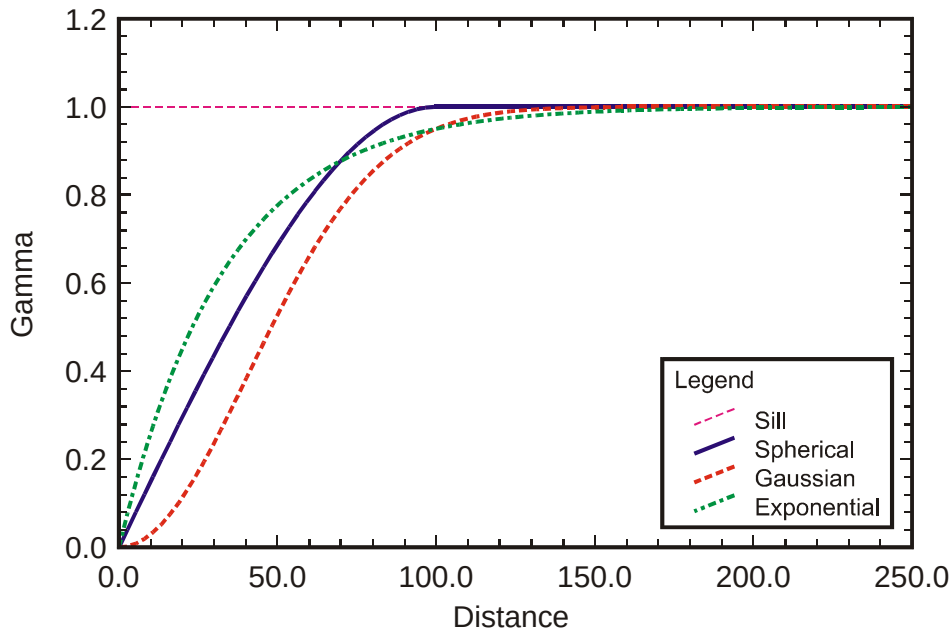


Figure 2. Three different variogram models with the same parameters: nugget =0.0, sill = 1.0 and range = 100.0.

Figure 3 shows an example of an experimental variogram with a spherical model fitted. Often the spatial correlation varies with direction, and such a case requires calculation of variograms in different directions and fitting of anisotropic (direction-dependent) models.

The nugget is a third parameter often used to define a variogram model. The nugget is a non-zero value of $\gamma(\mathbf{h})$ when $\mathbf{h} = 0$. In Figure 2, all variogram models intercept the Y-axis at $\gamma(0) = 0.0$. In many cases, there will be some level of variability at zero separation distance. This variability may be due to repeatability issues with the measurements and/or some level of spatial variability occurring at a scale that is smaller than the minimum sample spacing. The value of the nugget ranges between 0.0 and the value of the sill. If the nugget is equal to the sill value, then there is no spatial correlation in the data and traditional statistical approaches, those that do not account for spatial correlation, can be applied.

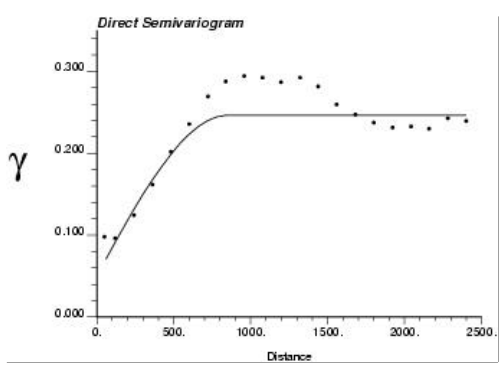


Figure 3: An example of experimental direct semivariogram of primary data with a spherical model fitted.

Kriging

In general, kriging is the process of using the information on spatial correlation contained in the variogram to make estimates of the sampled attribute at unsampled locations. Kriging is a "BLUE" process meaning that the kriging equations are formulated to provide the Best Linear Unbiased Estimate of a property at an unsampled location. Specifically, "best" refers to the kriging estimates having a minimum variance about the unknown true value at each unsampled location and "ubiased" meaning that the kriging estimate is centered on the unknown true value at the unsampled location. Solution of the kriging equations at any specific location may produce an estimate with a larger variance or some bias, but across all estimates, the minimum variance and unbiasedness properties hold true.

There are a number of variations on the kriging estimator and how it is applied to different problems. Those variants that are most applicable to UXO problems are discussed briefly below.

Ordinary Kriging

Consider the problem of estimating the value of a continuous attribute z (e.g. UXO intensity) at an unsampled location $\mathbf{u}$, where $\mathbf{u}$ is a vector of spatial coordinates. The information available consists of measurements of z at n locations $\mathbf{u}_\alpha$, $z(\mathbf{u}_\alpha)$, $\alpha = 1,2, ..., n$, that may have been obtained on a set of sample transects. Kriging is a form of generalized least square regression. All univariate kriging estimates are variants of the general linear regression estimate $z^*(\mathbf{u})$ defined as:

$$z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{a_1=1}^{n(\mathbf{u})} l_{a_1}(\mathbf{u})[z(\mathbf{u}_{a_1}) - m(\mathbf{u}_{a_1})] \qquad (3)$$

where $l_{\alpha 1}(\mathbf{u})$ is the weight assigned to the datum $z(\mathbf{u}_{\alpha 1})$ and $m(\mathbf{u})$ is the trend component of the spatially varying attribute. In practice only the observations closest to $\mathbf{u}$ being estimated are retained, that is the $n(\mathbf{u})$ data within a given neighborhood or window $W(\mathbf{u})$ centered on $\mathbf{u}$. If there is no trend in the data across the site, m is no longer a function of the spatial location $\mathbf{u}$ but is now the global mean of the data set, then Equation 3 defines the simple kriging, SK, estimator. In most practical applications of kriging, SK has proven to be overly restrictive and ordinary kriging is the preferred choice.

The most common kriging estimator is ordinary kriging (OK), which estimates the unsampled value $z(\mathbf{u})$ as a linear combination of neighboring observations without enforcing a global mean onto the estimate:

$$z^*_{OK}(\mathbf{u}) = \sum_{a_1=1}^{n(\mathbf{u})} l^{OK}_{a_1}(\mathbf{u})\, z(\mathbf{u}_{a_1}) \qquad (4)$$

OK weights l_α are determined so as to minimize the error or estimation variance $s^2(\mathbf{u})$ = Var$\{Z^*(\mathbf{u})$-$Z(\mathbf{u})\}$ under the constraint of unbiasedness of the estimate (4). These weights are obtained by solving a system of linear equations, which is known as the "ordinary kriging system":

$$\begin{cases} \sum_{b=1}^{n(\mathbf{u})} l_b(\mathbf{u})\, g(\mathbf{u}_a - \mathbf{u}_b) - m(\mathbf{u}) = g(\mathbf{u}_a - \mathbf{u}) \qquad a = 1,\ldots,n(\mathbf{u}) \\ \sum_{b=1}^{n(\mathbf{u})} l_b(\mathbf{u}) = 1 \end{cases} \tag{6}$$

The unbiasedness of the OK estimator is ensured by constraining the weights to sum to one, which requires the definition of the Lagrange parameter $m(\mathbf{u})$ within the system of equations. The addition of the Lagrange parameter can be thought of as the addition of another unknown to the system to balance the additional equation added to the system to ensure unbiased estimates. The only information required by the system are the variogram values for different lags, and these are readily derived from the variogram model fit to experimental values.

Indicator Kriging

In the earth and environmental sciences, problems often arise where it is not necessary to estimate the value of an attribute directly at an unsampled location but only to estimate whether that value is above or below some threshold level. For example, in a soil contamination problem it may only be necessary to estimate whether or not the contaminant concentration at an unsampled location exceeds, or does not exceed, the regulatory threshold. Similarly in a UXO site characterization, the investigation team may only be interested in knowing if there is at least one UXO in an unsampled portion of the site. These types of binary, yes or no, variables are referred to as indicators and indicator kriging is the application of a kriging algorithm to these indicator data.

Indicator data are determined through an indicator transform of the original, continuously defined, sample data. The resulting indicator datum, $i(\mathbf{u},z_k)$ is determined as:

$$i(\mathbf{u}, z_k) = \begin{cases} 1 & if \quad z \le z_k \\ 0 & if \quad z > z_k \end{cases} \tag{7}$$

where z_k is a threshold value against which each sample value is compared. The indicator transform as defined above is consistent with the definition of the cumulative probability distribution function for a discrete variable (e.g., Conover, 1980). However, in many environmental applications, the inverse of the indicator transform is of more interest. This inverse transformation is defined as:

$$i(\mathbf{u}, z_k) = \begin{cases} 1 & if \quad z \ge z_k \\ 0 & if \quad z < z_k \end{cases} \tag{8}$$

This inverse statement of the indicator transform places emphasis on those values that exceed the threshold value z_k. Christakos and Hristopulos (1996) provide discussion of the utility of the indicator transform shown in 8 for characterizing the spatial distribution of a contaminant.

Data that have been indicator transformed can be directly interpreted as the probability of a certain condition being true at location **u**. At a sample location where the value of the sampled attribute is known and can be compared directly to the threshold value, z_k, this probability can only be 0 or 1. In order to determine the probability of the condition being true at any unsampled location, the observation $z(\mathbf{u}_{\alpha 1})$ can be replaced by its indicator transform $i(\mathbf{u}_{\alpha 1};z_k)$ in the simple or ordinary kriging systems. This requires that the variogram also be calculated using the indicator transformed data. Kriging with this type of indicator data that define probabilities is known as probability indicator kriging, PIK, and the resulting estimates are a map of the condition being true at any location.

Finite Domain Kriging

One of the inherent assumptions in the development of the kriging equations is that the domain in which the kriging estimates are made is essentially infinite. This assumption stems from the use of a random function model as the basis for the kriging equations. One of the advantages of kriging as an estimator is that it not only accounts for the distance between all existing data points used in the estimation and the location being estimated, but that it also accounts for clustering of the existing data in an estimation. The covariances from each existing datum and the point being estimated are contained in a vector on the right hand side of the kriging equation (Equation 6). The further the distance between an existing datum and the location being estimated, the smaller the covariance between these points as defined by the complement of the variogram. The left hand side of the kriging equations contains a matrix with the spatial covariances between all existing data points. When this matrix is inverted during solution for the values of the kriging weights, values that are close together and have high spatial covariance are inverted to provide a lower overall weight to these points. This function of the left hand side of the kriging equations serves to filter out redundant information caused by data clustering.

Data collected in close proximity, such as along a linear transect, provide redundant information due the fact that the data are close together and spatially correlated. The kriging equations work to filter this redundancy by giving data at the far ends of the transect larger weights than those in the center of the transect. These data at the far end of the transect are seen as being “less redundant” and are therefore more heavily weighted. However, these data are also further from the estimation point and provide less direct information on the point being estimated. Therefore, the kriging equations provide a counter-intuitive set of results for estimations made using transect data.

The problem of kriging with transect data, or more generally, “strings of data” has been examined previously in two papers by Deutsch (1993 and 1994) where attention was focused on the application of kriging using data collected along vertical boreholes in subsurface investigations. The solution to the kriging weights problem developed by Deutsch (1994) is briefly outlined here:

The solution to the problem of kriging with a string of data (Deutsch 1994) is to replace the correlation function, *r(h)*, used to populate the covariance matrix on the left-hand side of the kriging equations with a redundancy measure calculated as:

$$r_{(n)}(\mathbf{u}-\mathbf{u}') = r(\mathbf{u}'-\mathbf{u}) + [\bar{r}(\mathbf{u}',(n)) - \bar{r}(\mathbf{u},(n))] \qquad (9)$$

where *(n)* is the set of transect data used in the kriging, **u** and **u**′ are the coordinate vectors defining the location to be estimated and an existing data point, respectively and the overbar indicates the average value of a property – here the average correlation between an estimation location or data point and all other *n* data points along a transect. The right hand side of the kriging equations containing the estimation location to existing data locations covariance values remains unchanged. Deutsch (1994) points out that the redundancy measure in Equation 9 is a positive, semi-definite function as long as r is a positive, semi-definite correlation function.

The estimator for the finite domain kriging (FDK) formulation is essentially the same as that for the OK estimator in equation 4;

$$z^*_{FDK}(\mathbf{u}) = \sum_{a_1=1}^{n(\mathbf{u})} v_{a_1}^{FDK}(\mathbf{u})\, z(\mathbf{u}_{a_1}) \tag{10}$$

with the replacement of the correlation function by the redundancy measure defined in 9 within the FDK system:

$$\begin{cases} \sum_{b=1}^{n(\mathbf{u})} v_b(\mathbf{u})\, r_{(n)}(\mathbf{u}_a - \mathbf{u}_b) - m(\mathbf{u}) = r(\mathbf{u} - \mathbf{u}_a) \qquad a = 1,\ldots,n(\mathbf{u}) \\ \sum_{b=1}^{n(\mathbf{u})} v_b(\mathbf{u}) = 1 \end{cases} \tag{11}$$

Because $r_{(n)}$ is a positive semi-definite function, the solution of the system in 11 is unique as long as each data point has a unique location. One disadvantage of the FDK system, relative to the traditional kriging equations, is that it is not an exact interpolator (i.e., it will not necessarily return the measured data value at a measured location).

The estimation variance, or kriging variance, of the FDK system is given by:

$$s^2_{FDK} = 1 - \sum_{a=1}^{n(u)} v_a\, r(\mathbf{u} - \mathbf{u}_a) - m(\mathbf{u}) \tag{12}$$

The calculated value of the FDK estimation variance is always positive, but due to the inexactitude of the FDK estimator, the FDK estimation variance is not necessarily equal to zero at the data locations.

In the case where there are L transects each containing n_l samples, the FDK estimator is modified from equation 10 and applied to each string of data:

$$z^{*l}_{FDK}(\mathbf{u}) = \sum_{a_1=1}^{n_l} l^l_a(\mathbf{u})\, z(\mathbf{u}^l{}_a) \tag{13}$$

where the n_l locations, u^l_a, in string l are used to make the estimate. The corrected kriging system is:

$$\begin{cases} \sum_{b=1}^{n_l} v_b^{(l)}(\mathbf{u})\, r_{(n)}(\mathbf{u}_a^l - \mathbf{u}_b^l) - m^{(l)}(\mathbf{u}) = r(\mathbf{u}_a^l - \mathbf{u}) \qquad a = 1,\ldots,n_l(\mathbf{u}) \\ \sum_{b=1}^{n_l} v_b^{(l)}(\mathbf{u}) = 1 \end{cases} \tag{14}$$

The average value along a given transect is:

$$\bar{z}_{(l)} = \frac{1}{n_l}\sum_{a=1}^{n_l} z(u_a^l) \qquad l = 1,\ldots,L \tag{15}$$

and an unsampled location can be estimated by considering each transect as a single data point represented by the average value of the data on that transect. The kriging estimator for this type of estimate is:

$$\hat{z}(\mathbf{u}) = \sum_{a_1=1}^{L} w_l(\mathbf{u})\, \bar{z}_{(l)} \tag{16}$$

The uncorrected kriging system for these estimates made with the transect averages is:

$$\begin{cases} \sum_{l'=1}^{L} w_l(\mathbf{u}) \cdot \bar{r}(l'-l) - m(\mathbf{u}) = \bar{r}(\mathbf{u} - l) \qquad l = 1,\ldots,L \\ \sum_{l'=1}^{L} w_l(\mathbf{u}) = 1 \end{cases} \tag{17}$$

where:

$$r(\mathbf{u} - l) = \frac{1}{n_l}\sum_{a=1}^{n_l} r(\mathbf{u} - \mathbf{u}^{(l)}) \tag{18}$$

and

$$r(l' - l) = \frac{1}{n_l' n_l}\sum_{b=1}^{n_l'}\sum_{a=1}^{n_l} r(\mathbf{u}_b^{l'} - \mathbf{u}_a^l) \tag{19}$$

The final estimation step is to employ the weights calculated for each transect to recombine the corrected estimates made using values along each individual transect (equation 13) into a final estimate:

$$z_{FDK}^{**}(\mathbf{u}) = \sum_{l=1}^{L} ?_l(\mathbf{u}) \cdot z_{FDK}^{*l}(\mathbf{u}) \tag{20}$$

This final estimate at a given location, **u**, can also be written as the sum of the estimate using data along each transect multiplied by the sum of the weight assigned to each transect:

$$z^{**}_{FDK}(\mathbf{u}) = \sum_{l=1}^{L} ?_{l}(\mathbf{u}) \cdot \sum_{a=1}^{n_l} l^{l}_{a} Z(\mathbf{u}^{l}_{a}) \tag{21}$$

As an example of the FDK correction for an estimate at a single location using data from a single transect, Figure 4 shows the calculated kriging weights for data along the transect used to estimate a point with an X coordinate that is at the center of the transect and that is 875 units off of the transect. The weights calculated with the OK system (top image) and the FDK system (lower image) are compared. The variogram model used in this example is a Gaussian model with a range of 3900 units and no nugget effect. The results of the FDK system are to significantly increase the values of the kriging weights assigned to the data closest to the estimation location and reduce the weight assigned to the values at the ends of the transect.

Finite Domain Kriging Example

The effect of the FDK formulation relative to using a straightforward OK approach in a UXO site characterization setting is examined systematically using three different 5000 × 5000 m Poisson fields, one homogeneous and two non-homogeneous (Figure 5). The homogeneous field has a uniform intensity and variance of the point distribution over the site. For the two non-homogeneous fields, a non-uniform point distribution characterized by a single feature in the center of the domain having increased intensity is superimposed on the homogeneous field. This feature represents a UXO target in a UXO site (McKenna et al., 2001) and the two non-homogeneous fields have different shapes in the area of spatially varying intensity centered at the feature (i.e., target): the isotropic target and the anisotropic target (Figure 5). In the remainder of this report, they are referred to as the “homogeneous,” “isotropic,” and “anisotropic” fields, respectively.

The three different simulated anomaly fields are sampled with parallel, north-south transects. For a given sampling event, a fixed number of transects are selected at random locations along the east-west axis. Four different transect widths are used: 10, 50, 100 and 200 meters. Along each transect, the anomalies are counted within equal area sampling cells that are always 50 meters long by 10, 50, 100 or 200 meters wide depending on the transect width selected. Different numbers of transects with randomly chosen locations are selected for the different sampling widths to sample different percentages of the total site. A total of 180 different randomly selected transects are examined for each number of transects and transect width. The sample cells along these transects provide the input data for both OK and FDK. The variogram calculated from complete knowledge of the true distribution of the anomalies is used with both of the kriging algorithms. Use of the same variogram for all sample designs ensures that differences in the results will be solely to differences in the kriging algorithms

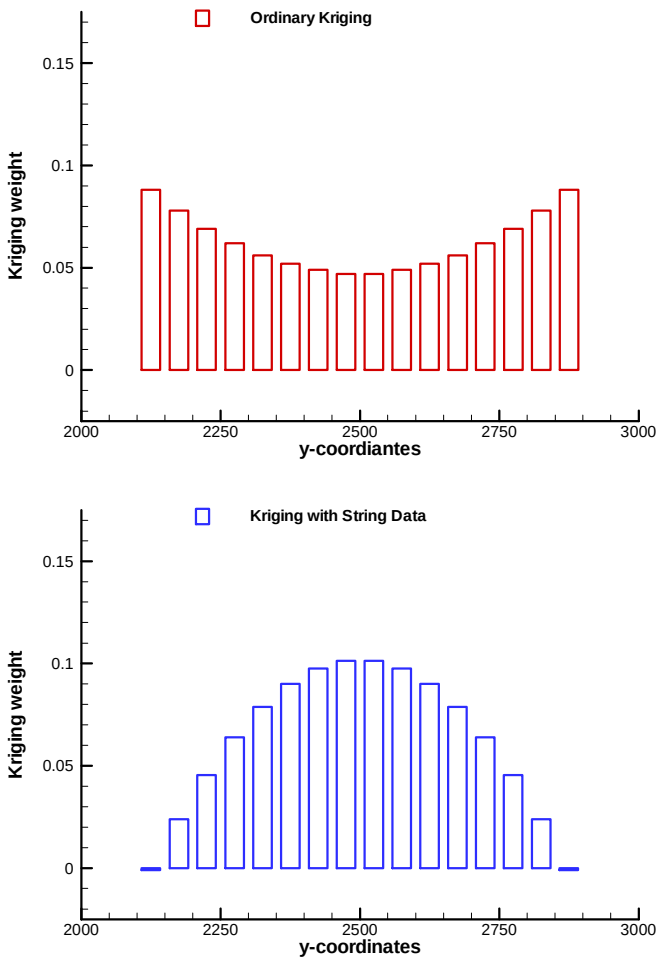


Figure 4. Kriging weights along a transect of data for estimation of a point off of the transect at a Y coordinate of 2500.0. The upper figure shows the kriging weights calculated using ordinary kriging and the lower image shows the kriging weights calculated with the finite domain kriging algorithm.

For each site, the performance of the OK and FDK algorithms and sampling designs with respect to sampling density and transect width was investigated by estimating the object intensities at the unsampled sites and then comparing them to the true intensities through jackknifing. Across all estimates, the mean error, ME, or bias, the mean absolute error (MAE), and the mean square error (MSE) were computed. Since the variable of interest is the actual number of objects, errors are not comparable if the size of the cell is different. To make the results comparable, all results are normalized to errors associated with 50 × 50 m cells.

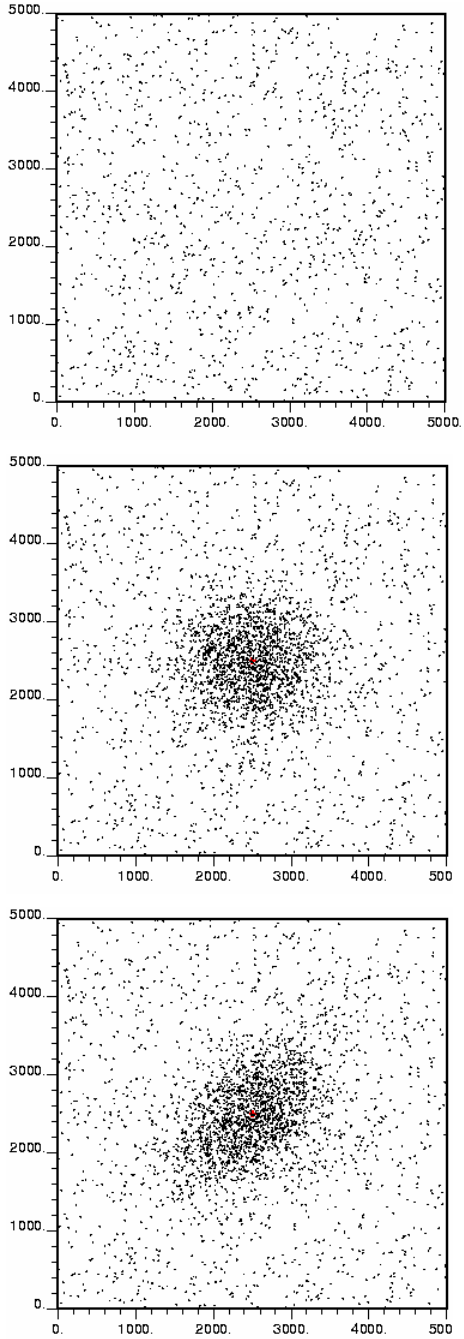


Figure 5. Distribution of anomalies simulated for testing finite domain kriging against ordinary kriging. The three sites are referred to as homogeneous (top), isotropic (middle) and anisotropic (bottom). The red dot in the lower two images shows the center of the target.

Figure 6 shows, for each spatial field and kriging algorithm, the average normalized ME and MSE statistics as a function of the sampling intensity. For the example results in Figure 6, 50-m wide sampling transects are shown. As expected, when the homogenous field is examined, there is almost no impact of sampling intensity on either ME or MSE regardless of the kriging algorithm used. The exhaustive semivariogram computed from the homogeneous shows a pure nugget effect and it makes the kriging estimate simply a local average. On the other hand, on

average, object intensities are overestimated when three to ten transects (i.e. 3 to 10 % of the site) are sampled from either isotropic or anisotropic field. Overestimation of object intensities happens when most sampling transects are located over the target area or near the target where the object intensity is higher than that of the background. When enough numbers of transects are sampled to locate some transects in the region of background intensity the estimated intensities become unbiased. This leads to higher MSE in this range of sampling intensity (3 to 10 % of the area sampled) but the MSE remains nearly constant as more than 10% of the site is sampled. Again there is no profound difference between OK and FDK in terms of MSE; FDK produces slightly higher MSE over the entire sampling intensity than OK. The results obtained for the isotropic site and the anisotropic sites are similar and there is no significant difference between them in terms of estimation errors. In addition, the same trend is observed for the MAE (not shown in this report).

In summary, the difference in estimation errors between OK and FDK is much smaller than the impact of the choice of the sampling intensity or the transect width for all three sites. One of the main reasons is that the impact of FDK is lessened when multiple transects are considered, as the number of data from the same transect decrease and data used in the kriging estimated are come from multiple transects. In that case, data manipulation by FDK does not help improve the estimation and using OK appears to be justified. Based on these results, the OK algorithm is used for the remainder of the estimation problems discussed in this report.

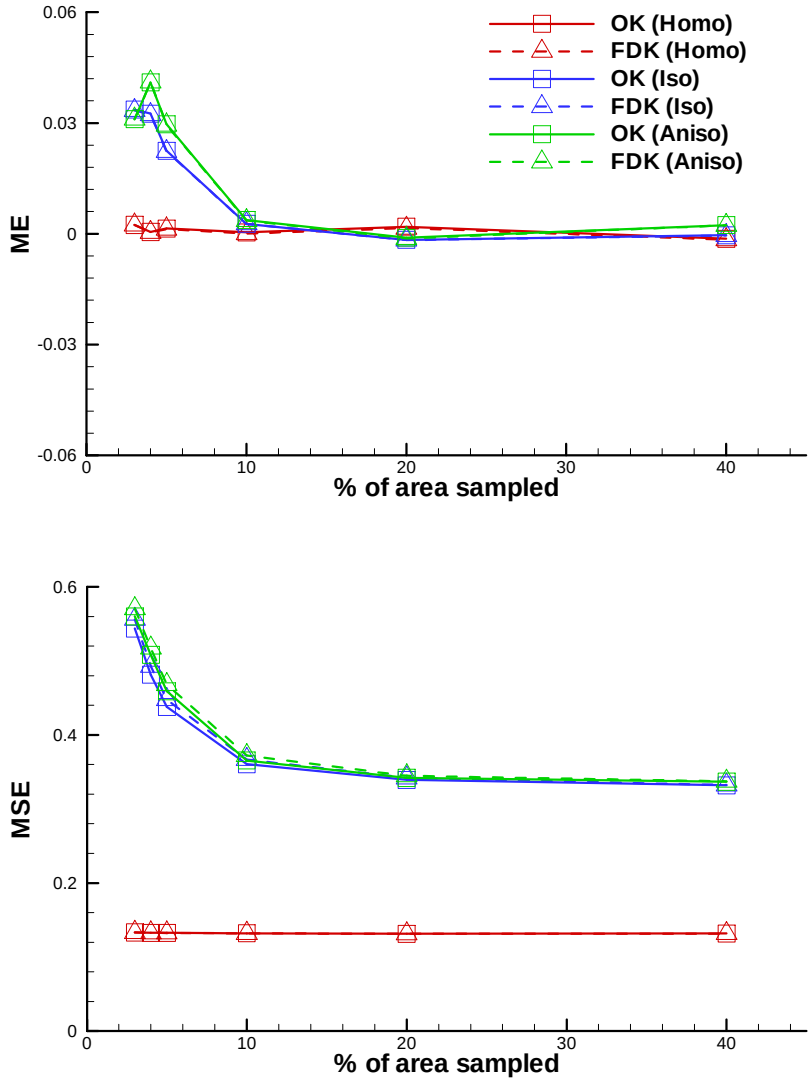


Figure 6. Comparison of the mean error (bias) and the mean squared error, top and bottom images repectively for estimates made with OK and FDK as a function of the percent of the site sampled. Results are presented for the homogeneous, isotropic and anisotropic sites shown in Figure 5.

Model Evaluation

For the example applications examined in this report it is desirable to assess the model predictions against the true spatial distribution of anomalies. This is done through the general process of jackknifing where a large proportion of the available data are held back from the analysis and model building and only used to assess the results of the model after it is created. The only data that are used to make the estimations are those sampled along the transects.

For the estimation of the total number of anomalies or the number of anomalies of interest at any location, it is possible to directly compare the estimates to the true values through the jackknifing procedure. However, another approach that appears to offer benefits to the decision makers at a UXO site is to map the probability of having at least one anomaly of interest at any location. Estimates of probability cannot be directly compared to true probability values as those are only 0.0 or 1.0. Therefore, it is necessary to check the results of a decision that would be made using the probability map against the true presence or absence of anomalies of interest at every location to assess the results of this probability estimation. In order to make a decision, it is necessary to determine at what probability the results will be compared. This determination is done by examining the problem from the perspective of engineering reliability (e.g. Harr, 1987). The decision made at any spatial location is a function of both the estimated reliability, R_E and the design reliability, R_D, as specified by the decision maker. The value of R_E at every location is calculated directly from the estimated probabilities as:

$$R_E = 1.0 - P(\text{at least one anomaly of interest}) \tag{22}$$

The value of R_D is set by the decision maker to a level that is acceptable by all involved parties. The specific meaning of R_D at any location is the probability of <u>not</u> having one or more anomalies of interest at that location.

The four different results of decisions that can be made and their relationship to the actual presence or absence of at least one anomaly of interest are:

$$\text{Correct (A): } (R_E \geq R_D) \quad \textit{and} \quad (\#_{anomaly} < 1) \tag{23}$$

$$\text{Correct (B): } (R_E < R_D) \quad \textit{and} \quad (\#_{anomaly} \geq 1) \tag{24}$$

$$\text{False Positive: } (R_E < R_D) \quad \textit{and} \quad (\#_{anomaly} < 1) \tag{25}$$

$$\text{False Negative: } (R_E \geq R_D) \quad \textit{and} \quad (\#_{anomaly} \geq 1) \tag{26}$$

The two types of correct decisions (A and B) occur when the location is correctly left as is ($R_E \geq R_D$ and $\#_{anomaly} < 1$) or when the site is correctly assigned to an area requiring further investigation ($R_E < R_D$ *and* $\#_{anomaly}$ ³ *1*). The two types of incorrect decisions arise when the location is unnecessarily slated for further surveying (False Positive) or when the location is not assigned to the region requiring further surveying, but in fact has at least one anomaly of interest

(False Negative). It is the latter of the two incorrect decisions that can be of severe consequence in UXO remediation.

For a given value of R_D, there are two steps to assessing the results of the estimated probability values. The first assessment step consists of using the estimated values of R_E at each location to determine whether or not the R_E is less than or greater than R_D. This determination is then combined with the true number of anomalies of interest at each model cell and the decision result (Equations 23 through 26) is determined. The result of the decision is recorded for each location within the site and the final proportions of each type of result across the site are tabulated.

It is recognized that the design reliability can be set to optimize different remediation objectives. For example, it may be desirable to select a value of R_D that minimizes the number of false positive and false negative errors under a loss function that counts each false negative as being of equal importance to three false positives. As another example, the objective may be simply to maximize the number of correct decisions. In order to examine the changes in the decisions made as a function of R_D, the model assessments shown in this report are conducted over a range of R_D values.

Site Characterization Process

The steps in the site characterization process used here are outlined below. These steps assume that the goals of the site characterization have already been defined and agreed upon by the different parties involved in making decisions at the site. The type of sample data that go into these analyses can be any combination of data collected at points, along straight transect and/or along meandering path transects

1) Assembly of the spatially referenced site data. These include the geophysical survey data, the site specific calibration of the geophysical instrument and any prior information on the site history. Ideally these data are already contained in a data base and a GIS and are all spatially referenced and the quality of the spatial referencing has been assessed.
2) The sample data and the site are discretized into sample cells and decision cells each of a finite areal extent. The easiest approach is to make the sample cell size and the decision cell size equal; however, this is not always possible as the sample cell size is generally limited by the sensor footprint.
3) The attribute of interest that will be used to make site characterization decisions is determined and aggregated to the size of the sample cell. The spatial correlation of these data are calculates as an experimental variogram and a variogram model is fit to the points of the experimental variogram.
4) The applicability of the variogram and the kriging procedure to the specific estimation problem is checked through a cross-validation procedure. This cross-validation procedure is demonstrated in Example Application 1.
5) Estimates of the attribute of interest are made for distances out to the range of the variogram. These estimates are made at the scale of the decision cell using ordinary or probability indicator kriging.
6) Decisions are made for each decision cell that has been estimated. Typically, these decisions are to either apply more detailed surveying to the decision cell or to leave it as is. The type of decision made and its effect on the final results will depend on the attribute that has been estimated (number of anomalies or probability).
7) Based on the results of the estimation, locations for additional transect sampling can be identified. These locations can be chosen to provide the greatest reduction in uncertainty or to meet other site characterization objectives. After these additional samples are collected, the site characterization process starts over with step number 1.

The calculations involved in this site characterization process are not overly time consuming and after the seven steps above have been completed once, it should be possible for a trained analyst to redo the steps with additional data in less than a day. This short time frame will allow for near real time iteration of the data collection and decision making process.

Demonstration Applications

Four different demonstration applications are presented. Each application highlights a different aspect of the geostatistical approach to making UXO site characterization decisions:

1) Application 1: This application demonstrates the flexibility of the kriging algorithm to estimate different properties that might be sampled on a UXO site such as the total number of anomalies per sample cell, the number of anomalies of interest within a sample cell, or the probability of at least one anomaly of interest within each sample cell. The ability to check the model setup prior to making estimates through a cross-validation process is demonstrated. The sample data are extracted from a simulated site.
2) Application 2: This example uses the Pueblo of Laguna N-11 target area as surveyed by Oak Ridge National Laboratories (ORNL) as the basis for demonstrating the probabilistic estimation of the target boundary from a limited number of parallel transects. This transect sampling design is, in general, similar to what would be used to detect a target area of a known shape and size with a specified level of confidence.
3) Application 3: This example is an extension of Application 1 that incorporates prior information on the site to extend the kriging estimates further away from the sample data. The stratum approach to incorporate prior information used here is probably the simplest means of incorporating prior information into kriging estimates and is well-suited for the types of prior information typically encountered at UXO sites.
4) Application 4: compares the decision results made for the Laguna N-11 site for two different approaches to locating the second round of samples. The two approaches considered are: 1) to locate a single meandering transect in the area of highest uncertainty surrounding the target area; 2) to infill with a new set of transects exactly midway between the existing transects.

Application 1

The first example application uses a simulated UXO site to demonstrate the ability of geostatistical estimation to predict the spatial distribution of three different attributes: the total number of anomalies, the number of anomalies of interest and the probability of having at least one anomaly of interest at all unsampled locations. These estimations are made from initial transect data that were collected on two separate straight transects and on a single continuous meandering path transect.

True Site and Sample Data

The true site used in this example is created using the UXO simulator developed previously for this project (McKenna et al., 2001). The distribution of all objects at the site is shown in Figure 7. Figure 7 shows a total of 59,467 objects within a 25 km^2 area. The average anomaly intensity is 2.38E-03 m^{-2}. As is true with the majority of UXO sites that have been geophysically surveyed, the anomalies are not distributed uniformly throughout the site, but show higher intensities near the two target areas. The target area in the northeastern corner of the site was created to represent the result of two separate mortar firing locations sending ordnance onto two distinct yet overlapping target areas. The higher anomaly intensity in the southwest corner of the site is representative of the anomaly distribution resulting from repeated aerial bombing of a target point with a preferential southwest to northeast flight path.

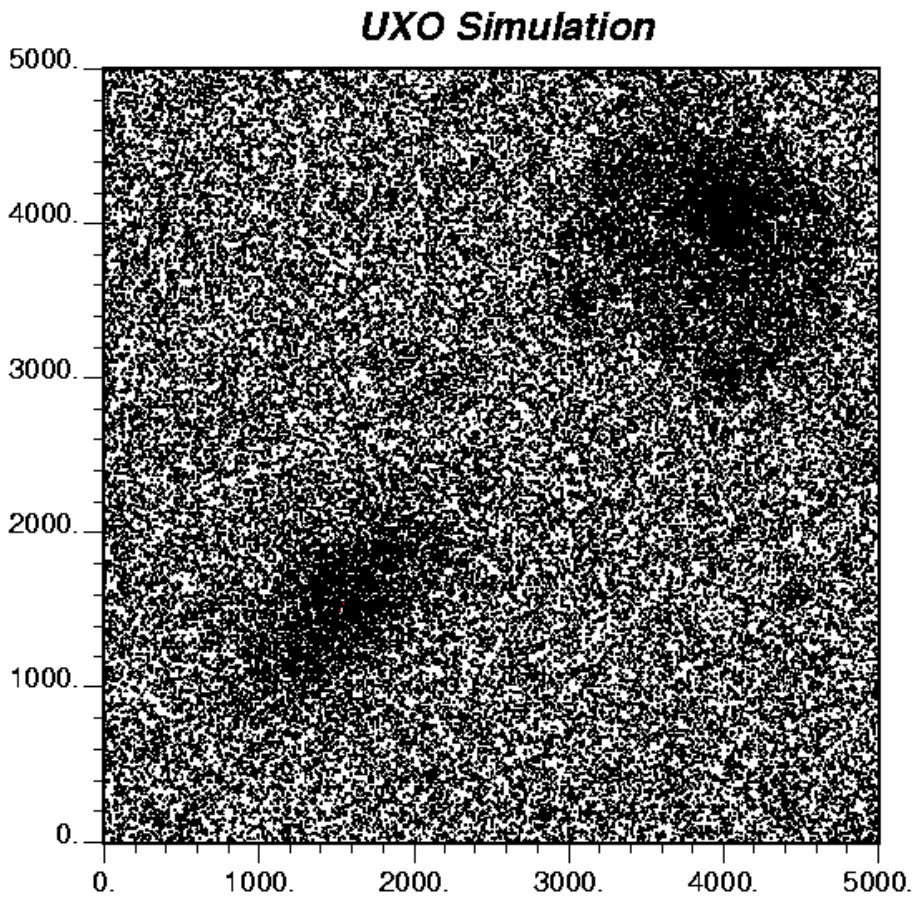


Figure 7. Distribution of anomalies within the simulated site. The axes dimensions are in meters.

In the UXO simulator, each simulated anomaly is assigned a signal strength from one of two different, overlapping, log-normal distributions. The distribution of signal strengths for the clutter and fragments has a geometric mean of 1.0 and ranges from 0.1 to 10.0. The distribution of signal strength for the anomalies that are true UXO ranges from 1.0 to 100.0 with a geometric mean of 10.0. It is noted, that there is significant overlap in the two distributions. The distribution of signal strengths across all simulated objects is shown in Figure 8.

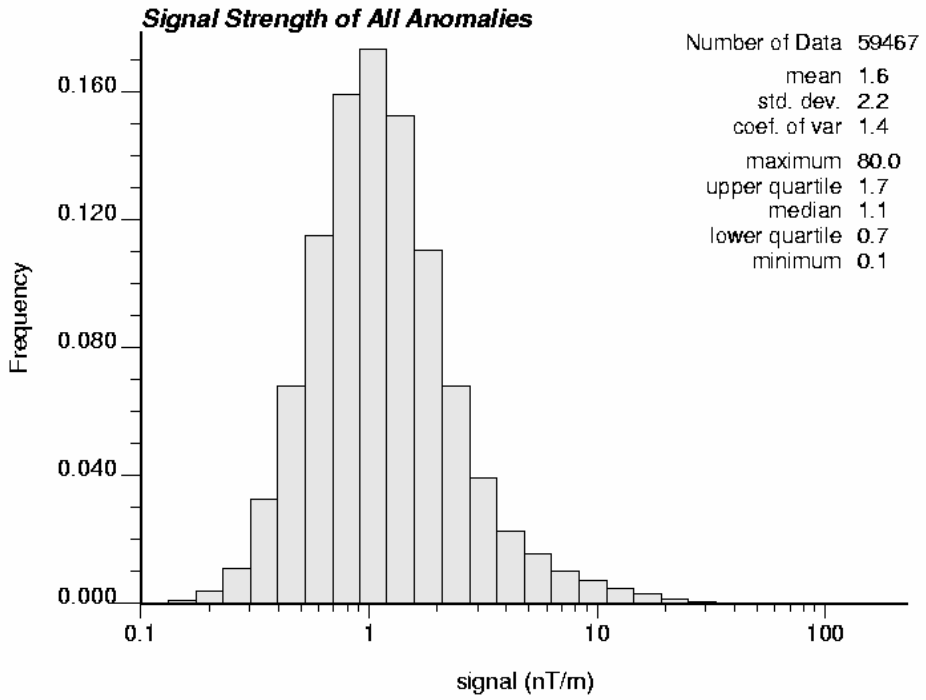


Figure 8. Histogram showing the distribution of the simulated signal strengths for all objects within the site domain.

The sample data are collected along three different transects: two straight transects that are orthogonal to each other and one meandering path transect in the southwestern portion of the site. These transects have been located to intersect suspected target areas as well as to provide some coverage across the site. The location of the meandering transect was set up to be limited by vegetation or other obstacles at the site. Each transect has a constant width of 25 meters as might be obtained from back and forth passes of a helicopter mounted sensor system. Figure 9 shows the location of the transect data. The color scale in Figure 9 indicates the total number of anomalies within each 25 x 25 meter cell along the transect. A total of 706 cells, or 1.77 percent of the total site, are sampled along the three transects.

Inherent in the geostatistical estimation approach is the conceptualization of the site as a discrete set of equal size model cells. Ideally, the scale of the samples and the scale at which the variables are estimated will be the same. For this work, all samples have a size of 25x25 meters and the estimates are also made on cells of similar size. The site considered in this application is 5000 by 5000 meters and has a total of 40,000 cells. A total of 706 of these cells have been sampled and estimates are made at the remaining 39,294 cells.

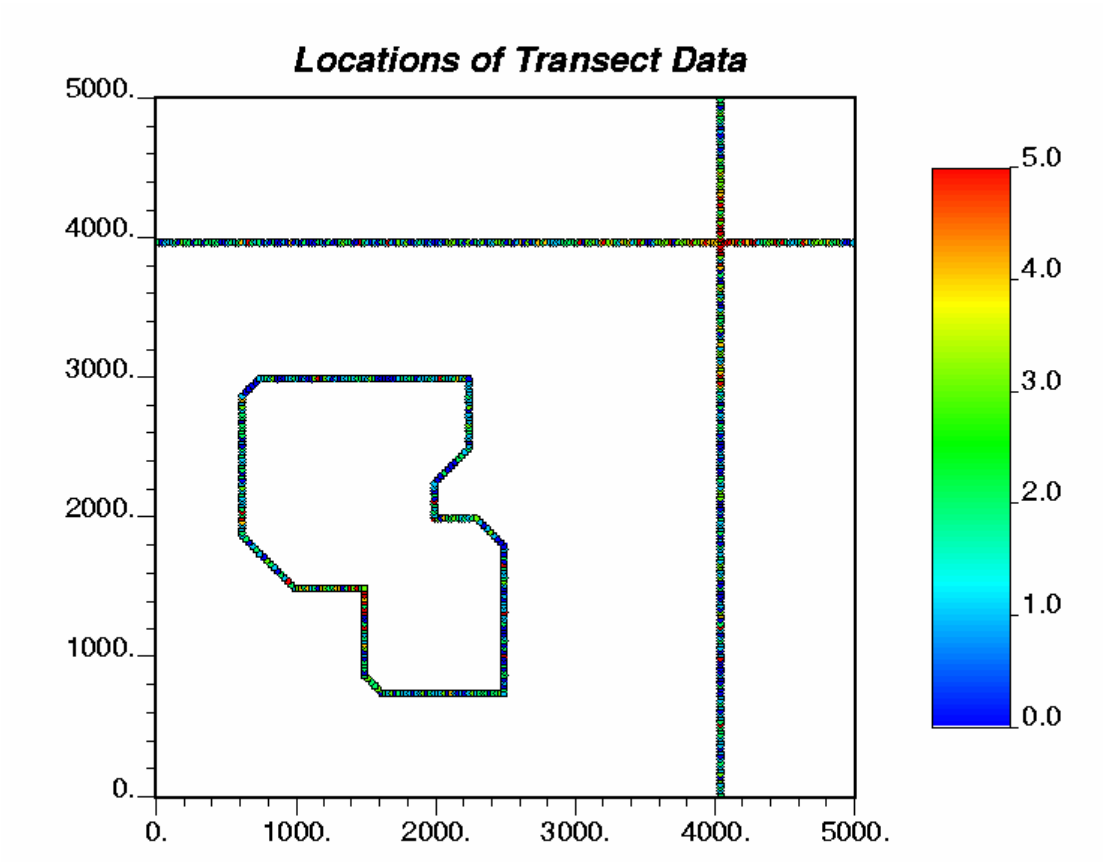


Figure 9. Locations of sampling transects on the simulated site. The color scale denotes the total number of anomalies within each 25x25 meter cell along the transects

Variograms

The sample data collected along the transects are used as input for the variogram calculations. For each of the 706 sample cells contained in the transects, several summary data are recorded including the total number of anomalies, the total number of UXO (assumes that all anomalies in each cell were excavated), the number of anomalies above signal strength thresholds of 3, 4 and 10 nT/m and the average signal strength of all anomalies within the cell. At this point in the site characterization process, the site characterization team must determine the variable in which they are most interested in mapping. Ideally, all anomalies in the sample transects would be excavated and the spatial distribution of UXO would then be estimated across the site. In reality, especially at a large site, this amount of excavation will not occur during the site characterization phase.

In lieu of the exact information provided by excavation, the spatial distribution of all anomalies, or all anomalies of interest (e.g., those with a signal strength above 3 nT/m) is considered. Estimating the former will provide information on the locations of increased anomaly density that correlate with target areas. Estimation of the latter property, anomalies of interest, is done under the assumption that, in general, the set of anomalies with larger signal strengths include

the true UXO. In this report, uncertainty in the signal strength and its relation to true UXO, for example as defined through a ROC curve, is not considered. The effect of different ROC curves on the site characterization process has been discussed in (REFERNECE FOR LAST YEAR REPORT) and is also being examined in a set of controlled tests administered by SERDP through the Mitretek Corporation that will be documented in a future report.

Three variograms are calculated on the transect data using: 1) the total number of anomalies; 2) the number of anomalies with signal strengths greater than or equal to 3 nT/m (an *anomaly of interest*) and 3) the probability of having at least one anomaly of interest at every location. The variograms calculated for this exercise are done using the VarioWin software (Pannatier, 1996) that is freely available at: http://www-sst.unil.ch/research/variowin/index.html. Analysis of the sample data did not reveal any preferential direction of spatial correlation and therefore all variograms were calculated as omnidirectional variograms (i.e., spatial correlation in all directions is averaged into a single variogram value for each separation distance).

The variogram of the total number of anomalies is shown in Figure 10. The parameters of the model variogram fit to the experimental variogram points are given in the first line of Table 1. Figure 10 shows that there is considerable small-scale variability in the number of anomalies from one cell to the next. The nugget value of 1.9 accounts for 67 percent of the total variance (total variance = nugget + sill = 2.82). The relative value of the nugget is controlled by the intensity of the anomalies and the sample cell size with larger cells providing a greater smoothing effect and thus a relatively lower nugget. The range of the total anomaly variogram is 425 meters, or roughly $1/12^{th}$ of the length of the site.

The variogram for the anomalies of interest, those above 3.0 nT/m, is shown in Figure 11. The variogram model parameters for this variogram are given in the middle row of Table 1. Similar to the variogram of the total number of anomalies, there is a relatively large nugget effect. For the anomalies of interest variogram, the nugget represents 55 percent of the total variability. The range of the anomalies of interest variogram is 250 meters, or five percent of the total domain length. This range is shorter than that of the total number of anomalies as expected due to the more localized nature of the anomalies of interest.

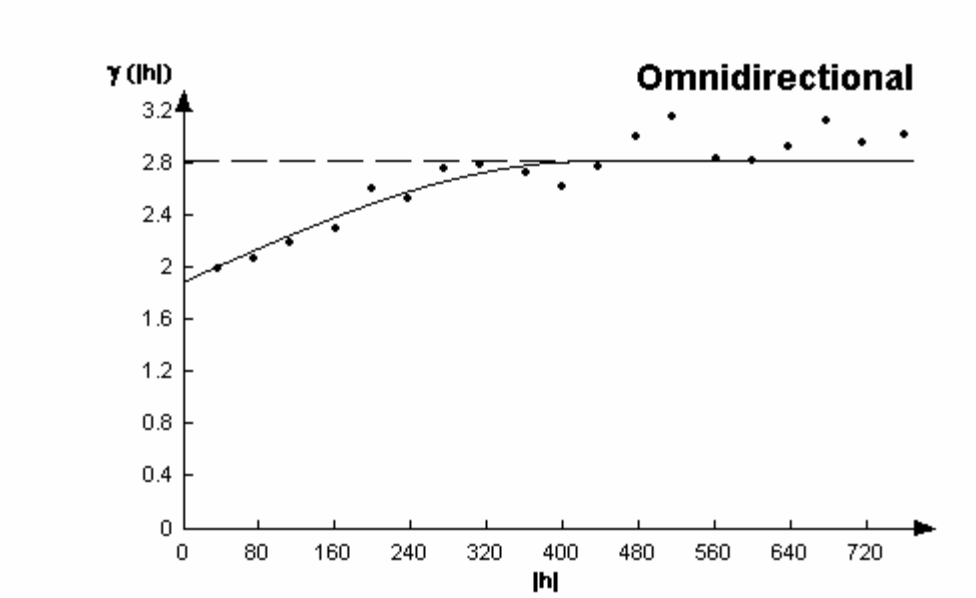


Figure 10. Experimental and model variogram fit to the total number of anomaly data obtained on the transects.

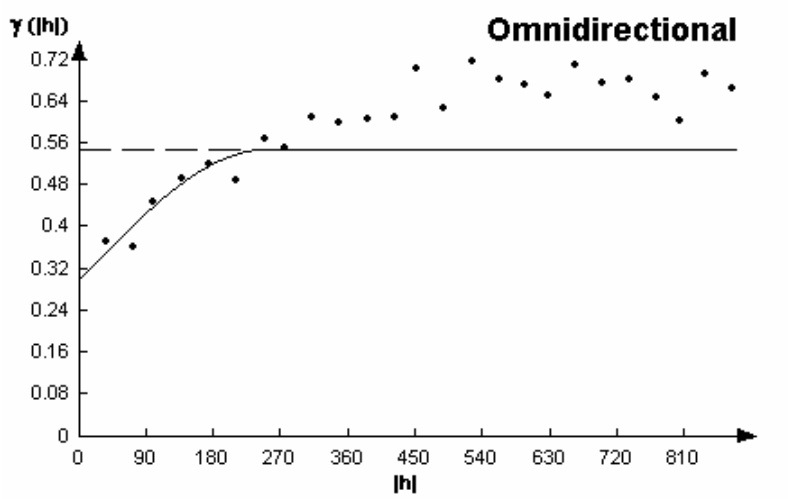


Figure 11. Experimental and model variogram fit to the number of anomaly data with geophysical signals above 3.0 nT/m obtained on the transects.

The third variogram is calculated on an indicator transform of the sample data. Those cells containing at least one anomaly with a signal strength of 3.0 nT/m or greater are assigned a 1.0 and those without such an anomaly are assigned a value of zero. These indicator values are directly interpretable as the probability of at least one anomaly of interest existing within each cell. A total of 121 cells, or roughly 17 percent of the sample data, contain at least one anomaly with a signal strength of 3.0 nT/m or larger. The variogram calculated on these indicator data is

shown in Figure 12 and the variogram model parameters are given in the bottom line of Table 1. This indicator variogram has a nugget that accounts for 52 percent of the total variability in the data and has a range of 800 meters. This longer range value, relative to the other two variograms is, in a large part, due to the large number of contiguous cells with 0.0 indicator values where the sample transects cover areas of the site without any anomalies of interest.

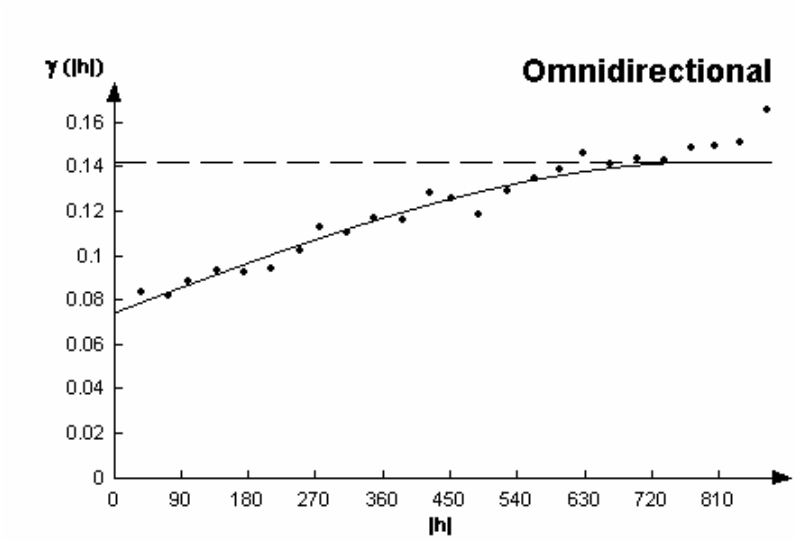


Figure 12. Experimental and model variogram fit to indicator data created for a geophysical signal threshold of 3.0 nT/m.

Table 1. Variogram model parameters for the three different data sets.

	Model Type	Nugget	Sill	Range (m)
Total Anomalies	Spherical	1.9	0.92	425
Number of Anomalies >= 3 nT/m	Spherical	0.30	0.25	250
Probability of one anomaly >= 3 nT/m	Spherical	0.075	0.068	800

Model Validation

A considerable amount of literature has appeared in recent years arguing that it is impossible to validate or verify predictive models and that such models can only be disproved and even then they can only be disproved for the specific application that is considered (e.g., Konikow and Bredehoeft, 1992). Several authors have argued that such phrases as model "validation" or "verification" be replaced with less philosophically loaded terms such as "model checking" or "model evaluation". Arguing the philosophy of such issues and the correct semantics is beyond the scope of this report. Here we use the historically popular form of "cross-validation" as a model checking technique and use the consistent term "model validation" to describe the process.

A standard procedure for checking the applicability of any statistical prediction model to a given problem is the cross-validation technique. Cross-validation consists of deleting one datum from the original data set containing *n* points, estimating the value of the deleted datum using the remaining (*n-1*) data, comparing the estimated value to the deleted true value, calculating one or more performance measures on the results of this comparison (e.g., the error or absolute error) and then averaging the value of the performance measure over all *n* deletions of a datum. An excellent review of cross-validation and other model evaluation techniques is provided by Efron and Gong (1983). As pointed out by Efron and Gong (1983), a distinct advantage of the cross-validation technique is that it can be applied to predictions made through any type of estimation algorithm of arbitrary complexity.

In the assessment of estimates made through kriging, an existing datum at a location is removed and the kriging algorithm is used to estimate the value of that datum using the remaining data. The use of a search window that includes a finite number of sample data less than the total number of sample data means the same number of data are used to estimate the value at the deleted location as at any other location when a value is not deleted. In other words, in the definition of cross-validation given by Efron and Wong (1983) the *n-1* remaining data are used to make the estimate. In cross-validation of the kriging results in this report, the full number of data in the search window are used to make this estimate; however, the datum at the estimation location is not included in this estimate.

Cross-validation is conducted for estimates of all three quantities being considered. The results of the cross-validation are presented as tables showing the proportions of different type of results for the estimation of the total number of anomalies and the number of anomalies of interest and as a graph for decisions made from the probability map.

The kriging program, *kt3d*, developed by Deutsch and Journel (1998) is used to create estimates of all three properties from the sample data. In the case of estimating the total number of anomalies, *kt3d* produces non-integer estimates of the number of anomalies at each cell. A fraction of an anomaly is not a realistic estimate and therefore the estimates are adjusted to be whole integer estimates of the number of anomalies by rounding each estimate up to the next integer value. This adjustment imparts additional conservatism to the estimates by increasing the estimated number of anomalies in any one cell by as much as 0.99 and forcing there to be at least one anomaly in every cell. The results of the estimation of total number of anomalies are summarized in the matrix in Table 2. In Table 2, the different estimated numbers of anomalies are shown in the rows and the different numbers of true anomalies are shown in the columns. In Table 2, values on the diagonal are the number cells in the domain where the total number of anomalies was correctly estimated. Values in the cells above the diagonal are where the estimated number of anomalies is less than the true number of anomalies and these cells can be thought of as false negative results. Entries in the matrix of Table 2 below the diagonal contain the number of cells where the estimated number of total anomalies is greater than the true number. These results are, in a sense, false positive results

Table 2. Cross-validation results of estimation of total number of anomalies.

	0	1	2	3	4	5	6	7
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.047	0.033	0.031	0.004	0.001	0.003	0.001	0.000
2	0.166	0.200	0.144	0.068	0.016	0.013	0.003	0.001
3	0.016	0.034	0.042	0.025	0.025	0.008	0.006	0.003
4	0.006	0.003	0.007	0.020	0.006	0.006	0.007	0.003
5	0.000	0.003	0.004	0.007	0.008	0.006	0.001	0.001
6	0.000	0.000	0.001	0.004	0.003	0.004	0.000	0.004
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The results in Table 2 show how well the chosen variogram model and kriging algorithm work in cross-validation mode. These results represent a self-consistent check on the estimation process without having to collect any additional data. The percent of the estimates that are correct are 21.4, 57.5 percent of the estimates are false positives and 20.7 percent are false negatives. Table 2 shows that no cells were reestimated as containing zero anomalies and this result is due to the post-estimation modification of rounding the estimated fractional anomaly amount up to the next highest whole integer. This upward adjustment also biases the overall estimates towards the conservative end of the spectrum as evidenced by the larger number estimates below the diagonal (false positives) than above the diagonal (false negatives). Evidence of the smoothing nature of the kriging algorithm is seen by the results tending to overestimate lower true values and underestimate higher true values (Table 2).

The results of the cross-validation estimation of the number of anomalies of interest are summarized in Table 3. Again, the proportion of correct decisions for each number of anomalies are shown in the diagonal of the matrix, false negatives are above the diagonal and false positive results are below the diagonal. For these estimates, 8.9 percent were correct, 86.8 percent are false positives and 4.2 percent are false negatives. The large number of false positives is influenced by the conservative decision of rounding up the fractional estimates to the next integer value. Almost all of the false positives occur at locations where one anomaly of interest is estimated but none exist.

Table 3. Cross-validation results of estimation of the number of anomalies of interest.

	0	1	2	3	4	5	6	7
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.807	0.082	0.021	0.004	0.001	0.000	0.000	0.000
2	0.017	0.024	0.004	0.006	0.003	0.001	0.000	0.001
3	0.004	0.011	0.004	0.003	0.001	0.003	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

For the estimation of the probability of having at least one anomaly of interest, the cross-validation step is not as straightforward as in the case of the total number of anomalies and the number of anomalies of interest. For probability mapping, there can only be two correct

answers: “0” or “1” corresponding to either none or at least one anomaly of interest existing at a location, respectively. However, the estimation procedure produces a full range of probabilities within [0,1] and therefore it is necessary to apply some sort of decision rule to the estimated probabilities before comparing them to the true values. The decision rule approach that has been developed and tested in this work is that of design reliability, R_D, as defined above.

The estimated probabilities of having at least one anomaly of interest are interpreted as the complement of the estimated reliability: $R_E = 1.0 - P(anomaly)$. Any location on the site where $1.0 - P(anomaly)$ is less than a specified design reliability, R_D, must be surveyed in more detail, and any location where $1.0 - P(anomaly)$ exceeds R_D can be left as is. The higher the specified value of RD, the more conservative are the site characterization decisions. The results of the decisions made using this strategy can be evaluated through cross-validation in the same way the estimates of total number of anomalies and number of anomalies of interest have been evaluated.

Results of decisions made using the cross-validation estimates of the probability of at least one anomaly of interest are shown in Figure 13 for a range of R_D values. At higher values of R_D, the proportion of correct decisions decreases and the proportion of false positive decisions increases. For all values of R_D shown in Figure 13, the proportion of false positive decisions remains less than or equal to 0.05. These results demonstrate the ability of the probability mapping to make relatively refined decisions compared to just estimating the total number of anomalies or the number of anomalies of interest. As an example, the proportion of correct decisions made with the probability map remains over 0.50 for all values of R_D, whereas in the estimation of the total number of anomalies and the number of anomalies of interest, the proportion of correct decisions were 0.214 and 0.089 respectively.

The cross-validation process is meant to recreate the mechanics of the estimation process as closely as possible. Therefore, the results of the actual estimation should be consistent with the results observed in the cross-validation step. In a practical application, it is possible to examine the cross-validation results and determine whether or not the variogram model and kriging are adequate for the application. Different variogram models and options within the kriging algorithm can be test ed using cross-validation before the final estimates are made. In a practical application, the cross-validation results are the only thing that will be available; however, for this hypothetical site, the actual results of the estimation can be compared to the underlying “true” distribution of the anomalies. This comparison to the true data held back from the estimation process is termed “jackknifing”.

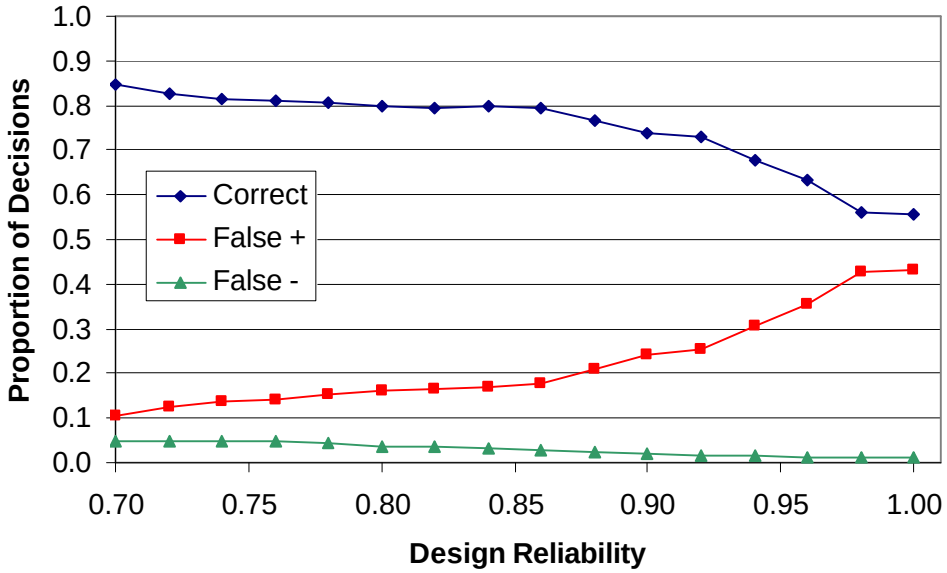


Figure 13. Proportions of different decision results as a function of the design reliability for the cross-validation results.

Estimation Results

After the variogram model has been estimated and has been deemed satisfactory through a cross-validation exercise, it can be used in a kriging process to estimate values of the chosen attribute at unsampled locations. The three variograms modeled above are now used to estimate the total number of anomalies, the number of anomalies of interest and the probability of at least one anomaly of interest at unsampled locations within the domain. The estimations are limited to being within the distance equal to the range of the variogram from the sample data.

The estimates of the total number of anomalies are shown in Figure 14. A total of 23,582 estimates are made. The estimated values show the highest number of anomalies in the upper left (NE) corner corresponding to one of the suspected target areas. The estimates made from the meandering path data also show relatively high numbers of anomalies, but not to the same level as in the NE corner of the site. The discontinuities along the transects are caused by the relatively large nugget effect in the variogram. Each sample data point is honored exactly, but this is accomplished in the kriging algorithm by creating a relatively sharp inflection in the estimated values at the data points.

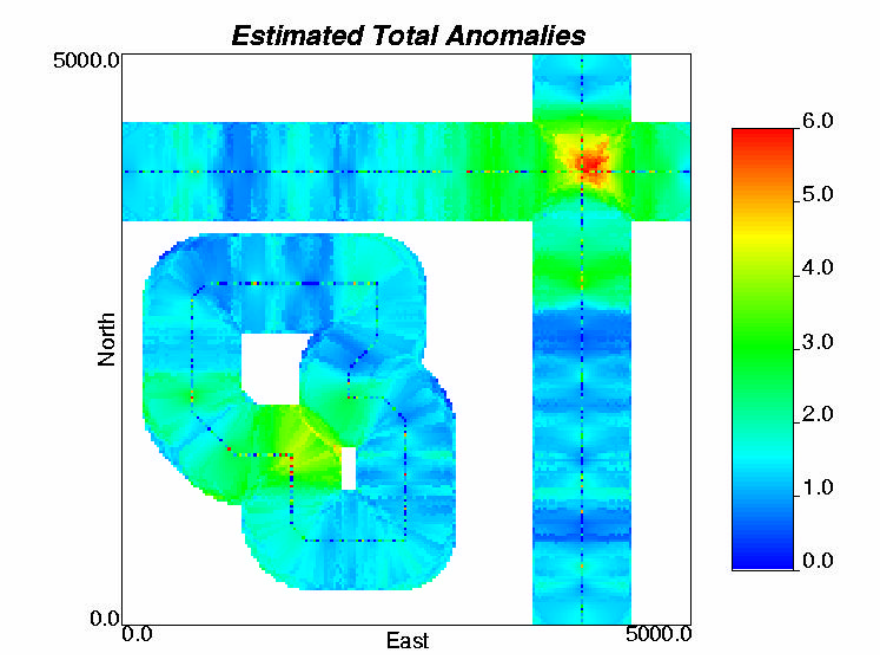


Figure 14. Estimates of the total number of anomalies within each model cell.

The accuracy of the estimates of the total number of anomalies are evaluated in Table 4. These results are in the same format as those used in the cross-validation exercise above and can be directly compared with those results. Across the 23,582 estimations, 22.9 percent are correct; 59.1 percent are false positive and 18.0 percent are false negatives. Each of these results are within +/- three percent of the values obtained in the cross-validation.

The results in Table 4 show that the majority of the cells are estimated to contain two anomalies and this creates the majority of false positives in areas where the actual number of anomalies of interest are zero or one. The majority of the false negatives, those cells above the diagonal, occur when kriging under estimates the number of total anomalies by one when there are two or three total anomalies actually in the cell.

Table 4. Results of the jackknifing assessment of the total number of anomalies estimation.

	0	1	2	3	4	5	6	7
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.039	0.042	0.025	0.011	0.003	0.001	0.000	0.000
2	0.164	0.224	0.147	0.064	0.022	0.007	0.002	0.001
3	0.023	0.044	0.049	0.030	0.017	0.009	0.003	0.001
4	0.003	0.008	0.010	0.015	0.007	0.005	0.003	0.002
5	0.000	0.001	0.002	0.002	0.004	0.002	0.001	0.001
6	0.000	0.000	0.000	0.001	0.000	0.002	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000

A similar set of results is produced for the estimation of the number of anomalies of interest within each cell. The estimated values are shown in Figure 15, and the estimation results are

summarized in Table 5. The shorter range of the anomalies of interest variogram leads to considerably fewer locations being estimated, 13,275 estimates, when compared to the estimates of the total number of anomalies (Figure 14). These estimates are in a relatively tight band around the data locations on the straight and meandering transects. Of the 13,275 total estimations, 9.2 percent are correct, 87.6 percent are false positives and 3.2 percent are false negatives. All of these estimated percentages are within +/- 1 percent of the values obtained in the cross-validation.

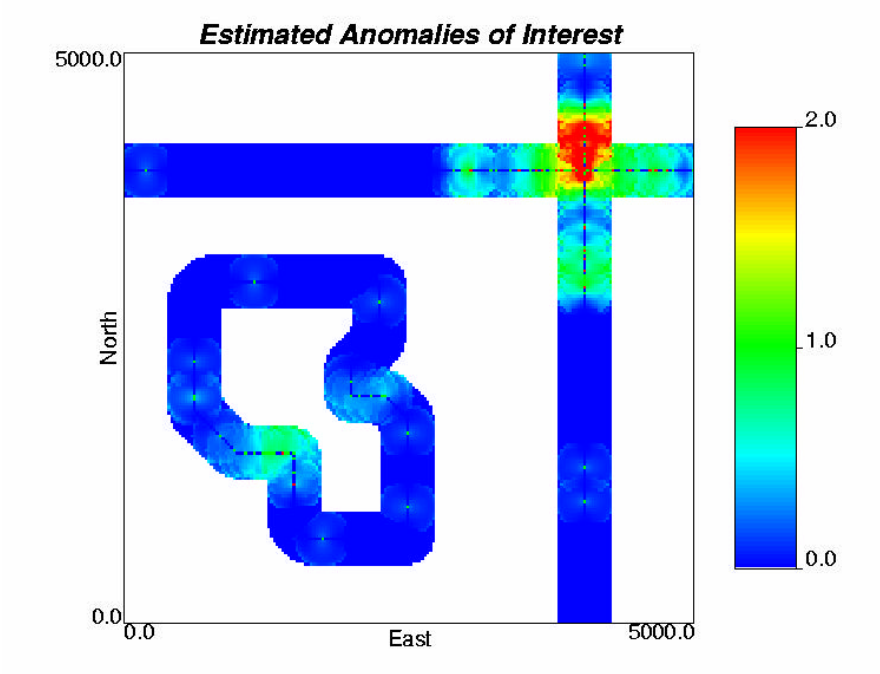


Figure 15. Estimated number of anomalies of interest.

Examination of Table 5 shows that almost all of the false positive estimates (83.5 percent of all estimates) occur when one anomaly of interest is estimated, but no anomalies of interest exist at that location. The high number of false positives, in this case are a direct consequence of the decision to round all fractional estimates up to the next highest integer number of anomalies. The majority of the false negative results (1.7 percent of all estimates) occur when one anomaly of interest is estimated and two actually exist at that location. This type of false negative result goes away when the estimation problem is turned to estimating the probability of at least one anomaly of interest as the actual number of anomalies of interest becomes irrelevant as long as there is at least one of them.

Table 5. Estimation results for the anomalies of interest.

	0	1	2	3	4	5	6	7
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.835	0.083	0.017	0.005	0.001	0.000	0.000	0.000
2	0.011	0.021	0.008	0.004	0.002	0.001	0.000	0.000
3	0.002	0.002	0.004	0.002	0.001	0.000	0.000	0.000
4	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The final attribute to be estimated in this exercise is the probability of at least one anomaly of interest at every location. The results of this estimation cannot be shown in a simple table, as were the previous two sets of estimations, because the results are in the form of a decision that is a function of the specified value of the design reliability. The estimates of the probability of at least one anomaly of interest are shown in Figure 16 and the decision results as a function of R_D are shown in Figure 17.

The indicator variogram calculated on 0,1 values defining the absence or presence of at least one anomaly of interest at every sample location provides the longest range of any of the three variograms calculated in this example application. The effect of this longer range is that 36,814 probability estimates are made which is a significantly larger number than those created in estimating either the total number of anomalies or the number of anomalies of interest. From Figure 16, the highest probability of at least one anomaly of interest occurs in the NE corner of the site where a target area is suspected. The decision results as a function of R_D show declining proportions of correct values as R_D increases made up for by an increasing proportion of false negatives. These results show that even at a relatively high R_D of 0.95, the proportion of decisions that are false negatives is still less than 0.4. The proportion of decisions that are false negatives at this same level of R_D is approximately 2 percent. Figure 17 can be compared to the same curve created in the cross-validation exercise (Figure 13). The two graphs are nearly identical.

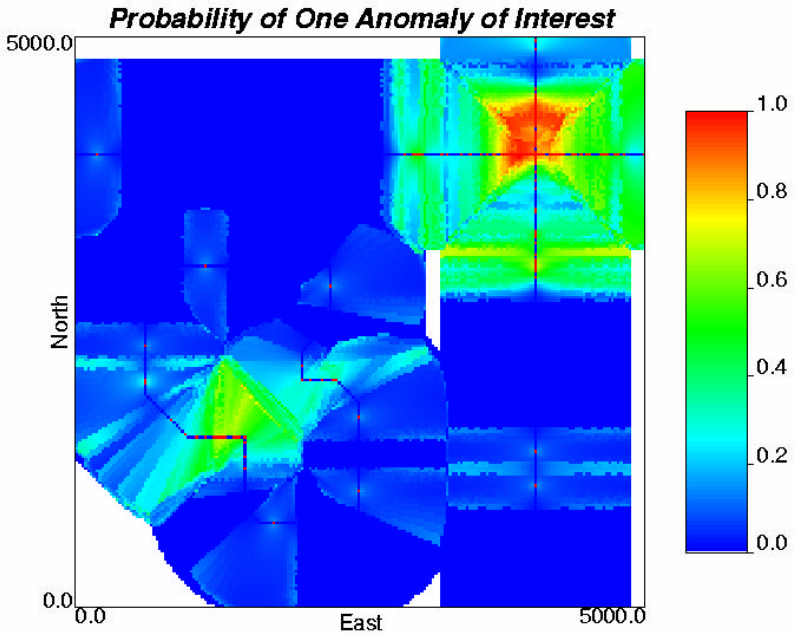


Figure 16. Estimated values of the probability of at least one anomaly of interest.

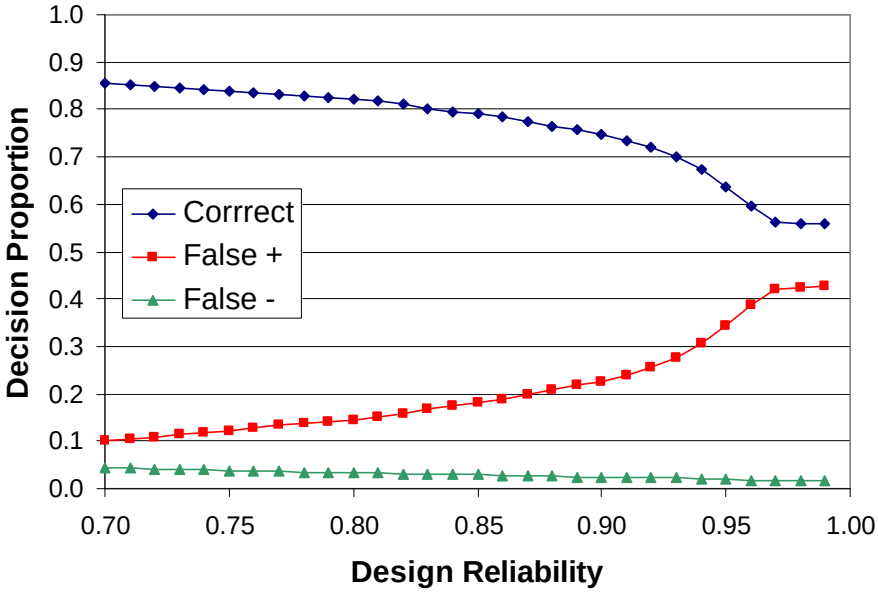


Figure 17. Proportions of different decision results based on the map in Figure 16 for a range of R_D.

Application 1: Summary

The main conclusions from this example application are: 1) the probability mapping approach provided superior decision making results when compared to the two different anomaly mapping approaches; and 2) that the cross-validation exercise done using only the existing sample data provided excellent predictions of the results of the actual estimations.

The probability mapping approach provided higher levels of correct decisions and lower levels of false positive decisions when compared to mapping the total number of anomalies or mapping the number of anomalies of interest. To some extent this result is due to the decision to round up any fractional estimates of number of anomalies to the next integer value in the two anomaly mapping approaches. This decision creates artificially high levels of false positive results. However, mapping the number of anomalies requires that some type of fairly arbitrary decision be made as to what fraction of an estimated anomaly should be counted as a full anomaly. For this work, the most conservative possible approach was taken. The probability mapping approach avoids having to make the decision as to what fractional value needs to be counted as a true anomaly by requiring a decision on the acceptable reliability to which a site needs to be characterized. Another approach to estimating numbers of anomalies that avoids the arbitrary decision making would be to use indicator kriging to estimate the probability of having a particular integer value of anomalies at any location. However, the decision making focus is generally on whether or not there is at least one anomaly of interest, not whether there are three, four or five and therefore, the current probability mapping approach gets at the issue of importance directly.

The cross-validation step provided excellent predictions of the results that were obtained in the actual estimations. This step is generally used to compare different variogram models and

options in the setup of the kriging algorithm. Additionally, cross-validation could be used to assess the impacts of different decisions made in the site characterization. For example, the choice of the optimal value of R_D for the making decisions off the probability map, or the effect of choosing different fractions of a true anomaly above which to round up to next integer value. Different values of these two thresholds could be compared in the cross-validation stage and then applied in the decision making that uses the final estimations.

The excellent correlation between the cross-validation results and the final prediction results obtained in this example depend on the available samples being fully representative of the site conditions.

Application 2

The second example application uses magnetometer data collected at the Pueblo of Laguna site in New Mexico in the spring of 2002 to demonstrate the ability of geostatistical estimation to estimate the probability of having at least one anomaly of interest at all unsampled locations. These estimations are made from initial transect data that were collected on multiple straight and parallel transects and the final goal is to determine the extent of the target areas as defined by the locations of the anomalies of interest.

Pueblo of Laguna Site and Sample Data

The N-11 target area in west-central New Mexico is selected from several different target areas that have been geophysically surveyed at the Pueblo of Laguna to serve as an example site for this application. The N-11 site is arid grassland at an elevation of approximately 1750 meters above sea level. The site is roughly 6900 feet by 6500 feet (area of 1030 acres or 417 ha). The N-11 Target Area was used as a practice range for precision aerial bombing training for an indeterminate amount of time between the end of WW II and 1990. Previous work (McKenna, 2001) used the N-10 target area at the Pueblo of Laguna to develop initial ideas on the use of spatial statistics to characterizing UXO sites. The surveyed area of the N-11 site is nearly an order of magnitude larger than that of the previously examined N-10 area.

The data used in this example were collected in the spring of 2002 by the geophysics group at Oak Ridge National Laboratory (ORNL) using a set of eight helicopter mounted magnetometers. More details on the surveying equipment and its capabilities can be found in Doll et al. (2003). The magnetic anomaly data were reported as the analytical model signal in nT/m on a 3x3 foot grid. The original footprint of the airborne sensor platform is approximately 25 feet, yet this footprint width is not seen in the final data as reported on the 3x3 foot grid. The original dimensions of the surveyed area at the site are roughly 7000x7000 feet resulting in over 5.5E+06 grid nodes for the analytical magnetometer signals. Not all of these locations contained actual signal data, as much as 50 percent of these node locations were not surveyed and were reported as “missing” data. For this exercise, the site is trimmed to dimensions of 6900x6500 feet and all of the grid nodes with missing data are removed. This smaller domain and removal of the nodes without a magnetometer signal data results in approximately 2.2E+06 grid nodes with magnetometer signals remaining.

The basis of this approach is that characterization decisions will be made over areas of the site with a finite size. While it is possible to make geostatistical predictions on 3x3 foot cells, corresponding directly to the support of the survey data, this spatial resolution would most likely be too fine for practical use in making characterization decisions. For this example, we use a 15x48 foot cell size for making decisions. This decision cell size is the minimum area over which a decision (i.e., schedule for detailed surveying or leave as is) will be made. This rectangular shape is chosen to be consistent with a sensor having a 15-foot wide footprint that is considerably smaller than that used by ORNL in the actual survey, but is somewhat larger than the widths of other sensor platforms often used in UXO characterization studies.

The initial survey data are resampled onto the 15x48 foot decision cells. Each decision cell contains a maximum of 720 original 3x3 foot survey cells. Some decision cells will have fewer

original survey cells due to the less than 100 percent survey coverage. For each decision cell, the maximum signal value across all 720 or fewer survey cells is found and shown in Figure 18. The data shown in Figure 18 serve as the ground truth for this exercise.

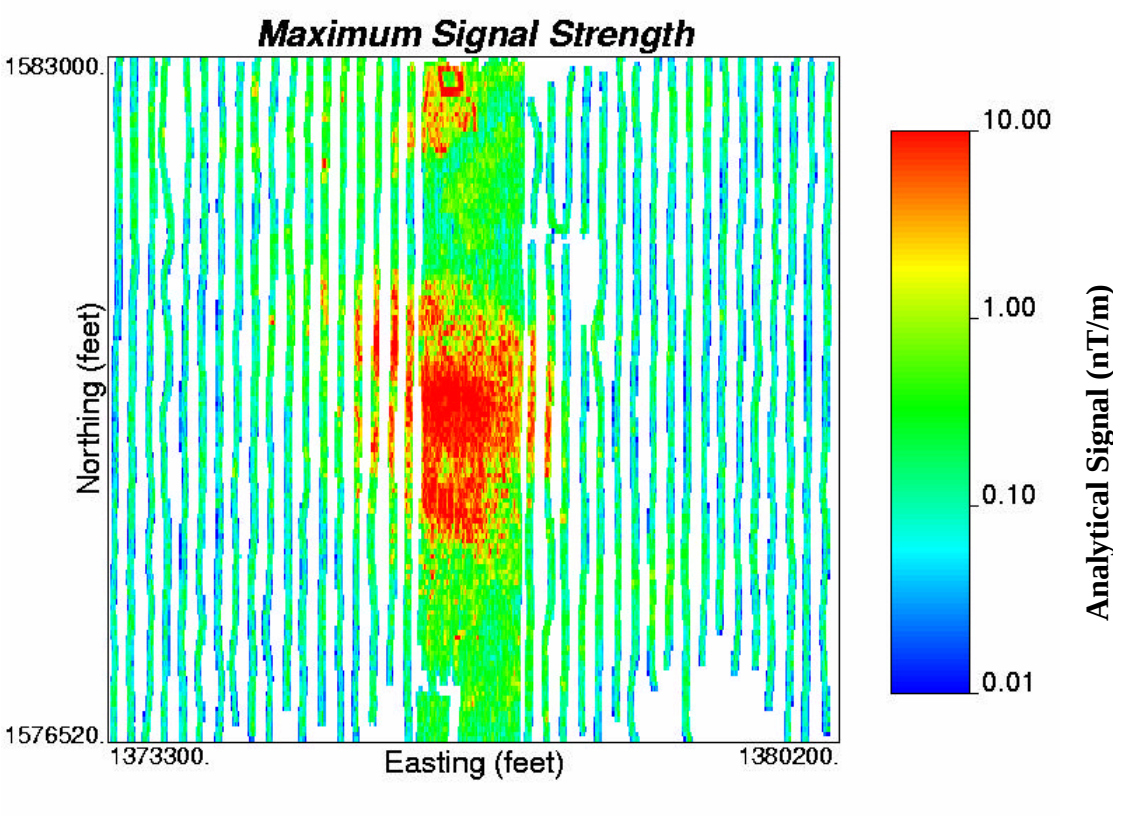


Figure 18. Survey data for the Pueblo of Laguna N-11 target area site as sampled on the 15x48 foot decision cells. The maximum signal value within each decision cell is shown here.

The N-11 target area data upscaled to the 15x48 foot scale are then sampled along 14 parallel transects with a constant spacing of 480 feet. The sampling transect locations and the maximum signal value for each decision cell along those transects are shown in Figure 19. Note that not all of the transects in Figure 19 are continuous across the domain. This is a result of the original sampling coverage shown in Figure 18 above, but it is also consistent with a parallel transect survey design where topography, vegetation or infrastructure precludes surveying in some areas of the site. These 14 transects cover roughly 3 percent of the site.

Mapping Signal Strength

As an initial test, or validation, of the effectiveness of the geostatistical approach to mapping from limited transect information, the data shown in Figure 19 are used to compute the variogram of the maximum signal strength. This variogram is fit using a spherical model with a range of 2800 feet and a nugget value of 3.0 (Figure 20). This variogram is used with the transect data and ordinary kriging to estimate the maximum signal value at the same locations for which the original sampled data were obtained. The results of this estimation are shown in Figure 21. Visual comparison of Figures 18 and 21 indicates that geostatistical techniques are

capable of making accurate predictions of spatially distributed anomaly signal values from limited transect data.

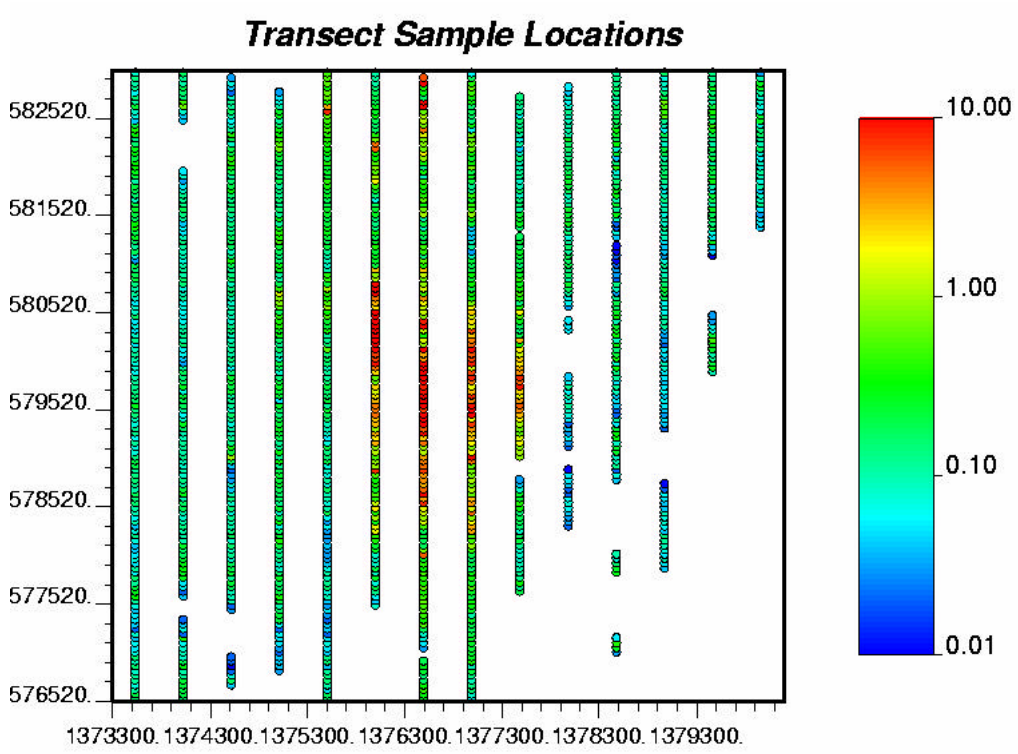


Figure 19. Sample transect data for the Pueblo of Laguna N-11 site. Along each transect, the maximum analytic signal within each 15x48 foot decision cell is shown.

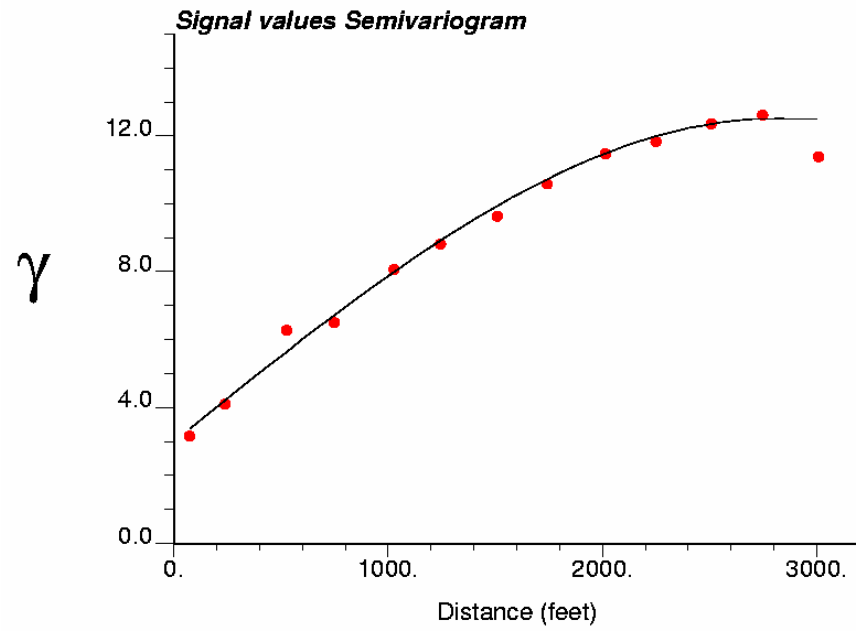


Figure 20. Variogram of the maximum analytic signal as calculated from the N-11 target area transect data.

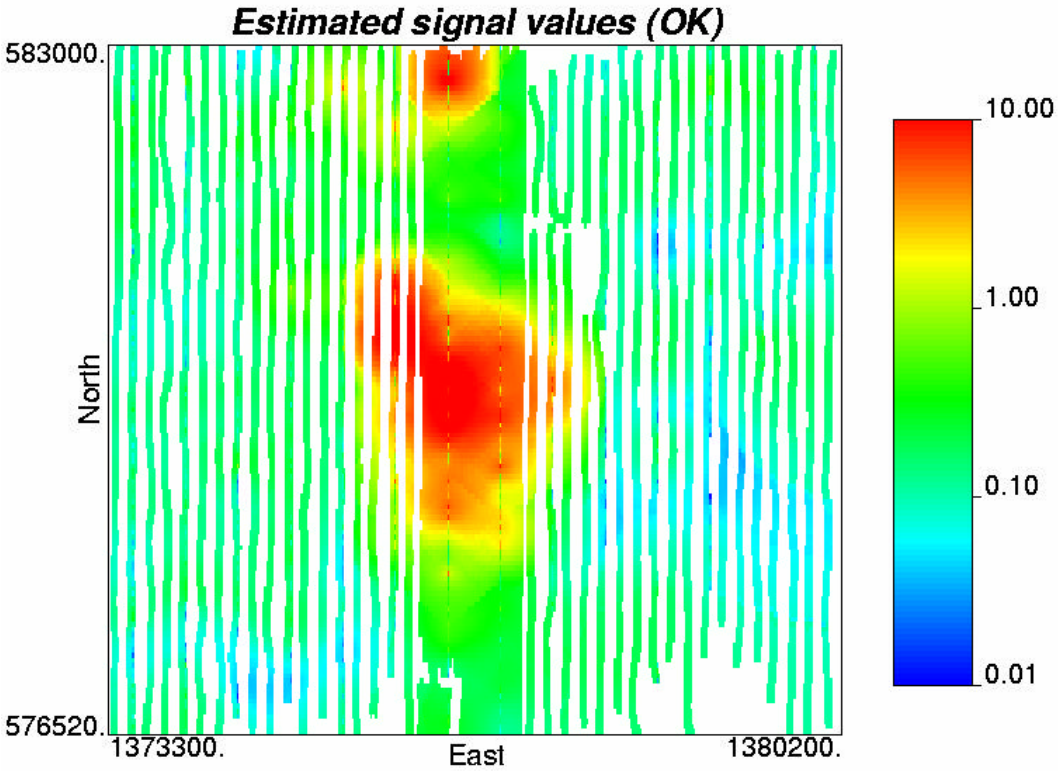


Figure 21. Estimated maximum analytical signal values made using ordinary kriging and the transect data and variogram shown in Figures 19 and 20 respectively. Compare to the actual sampled data in shown in Figure 18.

Probability Mapping

The estimation of the maximum signal strength across the site from the limited transect data is one type of geostatistical mapping that can be employed in UXO studies. This type of map could then be used to determine the locations of all estimates above a certain geophysical signal threshold (e.g., 5 nT/m) and these locations could be scheduled for more detailed geophysical surveying. The drawback of this approach is that is does not directly account for uncertainty in the estimates of the maximum signal. Accounting for uncertainty in these estimates would require combining the estimated signal values with the kriging variance map. However, this approach would only account for uncertainty as a function of the distance away from any existing transect data and does not directly take into account any uncertainty with respect to the decision being made (e.g., uncertainty with respect to being above or below the 5 nT/m threshold). This type of uncertainty could also be accounted by kriging under the multiGaussian model (e.g., Goovaerts, 1997), but a simpler approach that does not require the transformation of the data to a Gaussian distribution and that has proven useful in previous studies is to map the probability of having at least one anomaly of interest across the site. This is the same approach as the third demonstration used in the first example application.

The transect data are reclassified as binary indicators (equation 8) using a threshold of 5.0 nT/m. An indicator value of 1.0 defines a location with an anomaly above 5.0 nT/m and is also the probability of at least one anomaly above this threshold existing within the decision cell at that location. An indicator value of 0.0 denotes a location with no anomalies, zero probability, above the 5.0 nT/m threshold. These indicator data are used to calculate and model an indicator variogram for the site. This variogram is shown in Figure 22. The indicator variogram was fit with a nugget value of 0.013 and two spherical models having ranges of 600 and 2800 feet and sills of 0.009 and 0.043.

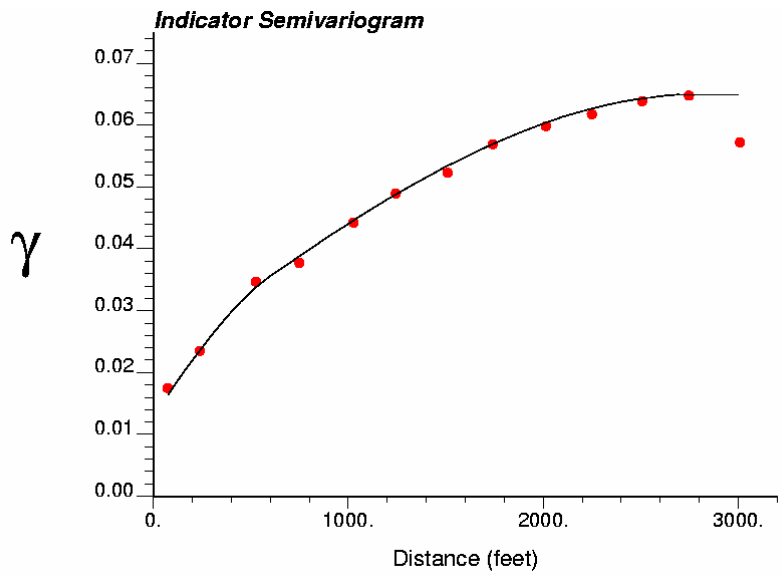


Figure 22. Indicator variogram for the N-11 target area site and a geophysical signal threshold of 5.0 nT/m.

The indicator data along the transects and the indicator variogram in Figure 22 are used to map the probability of having at least one geophysical anomaly with an analytic signal value above 5.0 nT/m within all decision cells across the site. The resulting estimated probability map is shown in Figure 23. Probabilities range from 0.0 to 1.0 and can only be one of these two values at the sample locations along the sample transects. At unsampled locations within the domain, the probability value can range anywhere in the [0,1] range. The map in Figure 23 shows high probability of at least one anomaly of interest in the center of the site as well as along the northern boundary near the center of the site.

Relative to the map of maximum signal strength in Figure 21, the probability map in Figure 23 provides the basis for a probabilistic interpretation of the extent of the boundary of the target area. The extent of the target boundary is determined by selecting an acceptable value of R_D and then slating every point within the chosen R_D contour for detailed surveying.

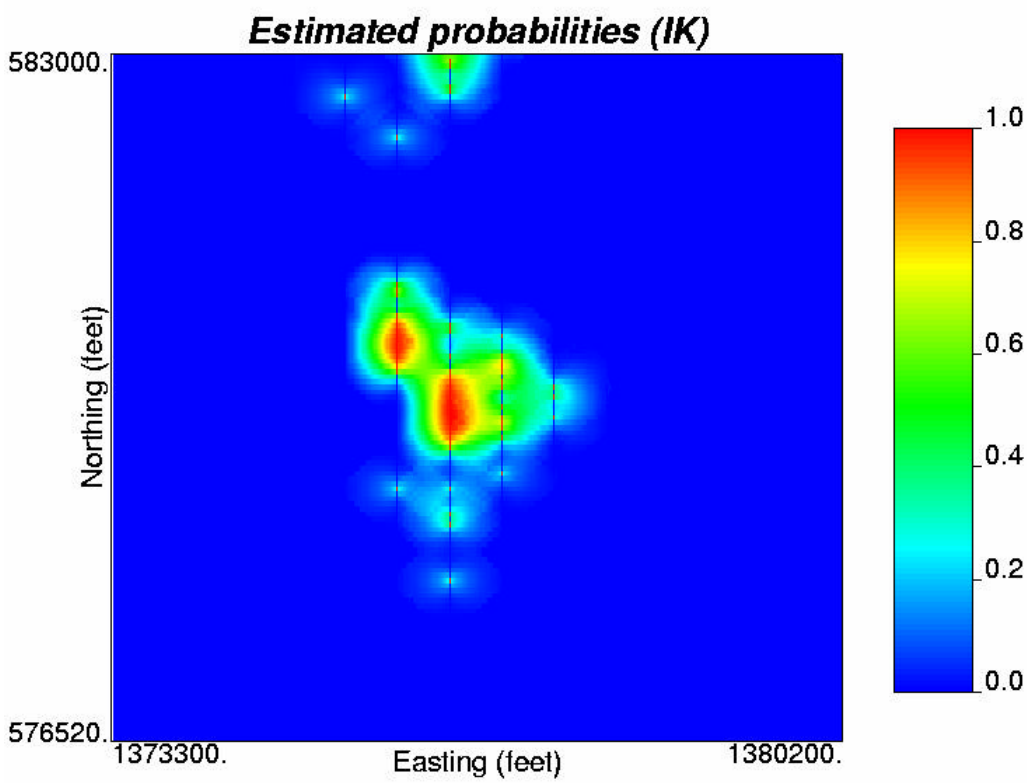


Figure 23. Probability of at least one geophysical anomaly with a signal strength greater than 5.0 nT/m across the site as estimated through indicator kriging.

Figure 24 shows example target boundaries for four different values of R_D. In addition, the location of all anomalies of interest outside of the estimated target area are shown in red in Figure 24. The locations of these anomalies of interest are known from the original geophysical survey data collected by ORNL. As the design reliability increases, the characterization decision

becomes more conservative and the size of the estimated target region increases to cover more of the site. This increase in size results in more of the anomalies of interest being contained within the estimated target region, but also increases the amount of the site that must undergo a detailed survey. These two results tend to decrease the number of false negative characterization decisions and increase the number of false positive decisions, respectively.

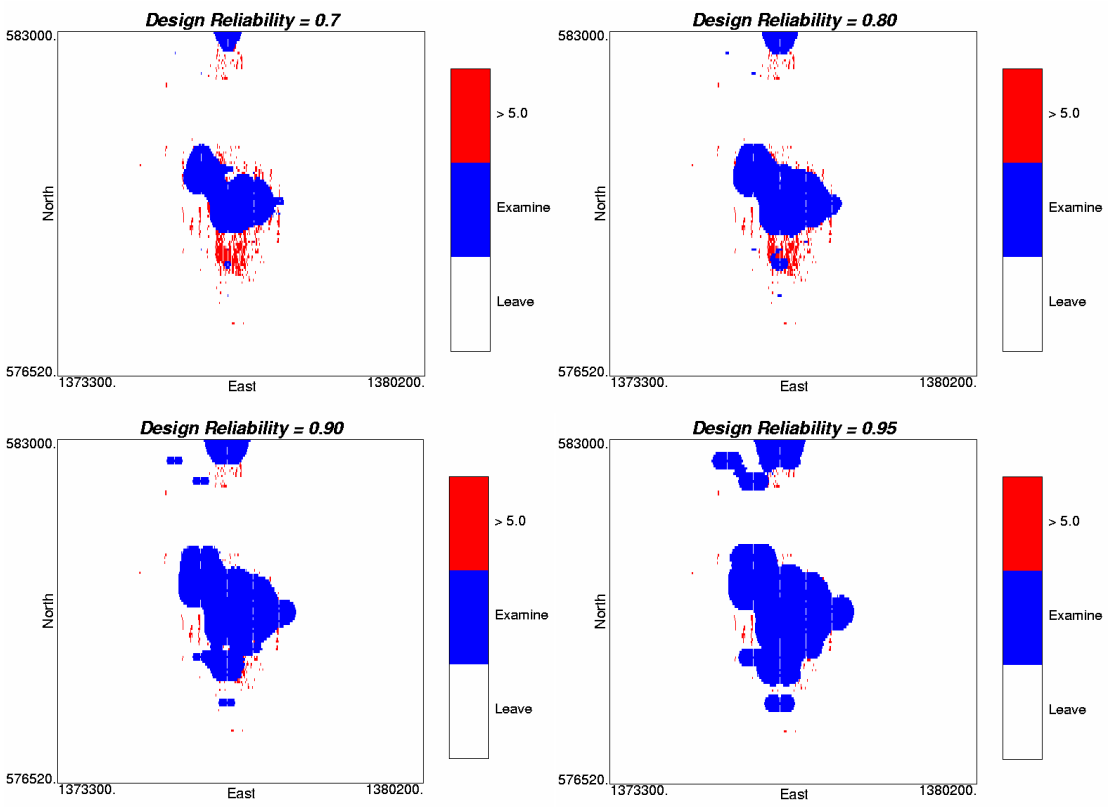


Figure 24. The estimated target boundaries (blue) for four different levels of R_D. The anomalies of interest lying outside of the estimated target area are shown in red.

The results of the decisions made for all values of R_D between 0.7 and 1.0 are calculated and shown in Figure 25. The image on the left side of Figure 25 shows the proportion of correct decisions, false positive and false negative decisions as well as the proportion of the anomalies of interest that are found as well as those that still remain. The decisions results are at the scale of the decision cells (15x48) while the proportion of anomalies found and remaining are calculated for the individual anomalies. The right image of Figure 25 shows an expanded view of the lower portion of the image on the left. From Figure 25, it can be seen that as the value of R_D increases, the proportion of correct decisions decreases while the proportion of false positive decisions increases. The proportion of false negatives decreases monotonically as a function of increasing R_D and is 2 percent or less for all values of R_D shown. The false negatives define the proportion of decision cells remaining that contain at least one anomaly of interest while the number of anomalies left behind refers to the number of individual anomalies remaining. The reason that the proportions of the false negatives and the anomalies of interest are not equal for a given R_D, is that more than one anomaly of interest can remain in a single decision cell.

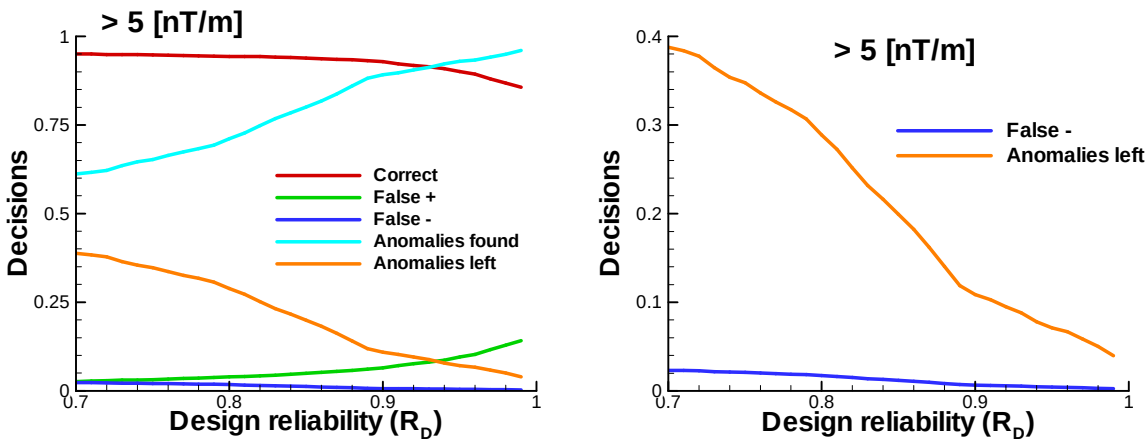


Figure 25. Proportions of decision results and anomalies of interest found and left behind for R_D values between 0.70 and 1.00.

Application 2 Summary

This example application demonstrated the ability of geostatistical mapping techniques to estimate a fourth attribute, maximum signal strength, from a limited number of equally spaced parallel transects with a 15 foot width. These same data were also used to map the probability of one anomaly of interest across the site and these results were used to map the extent of the target area where the target area is defined by the locations of the anomalies of interest. The results show (Figure 24) that this technique is able to efficiently identify the outlines of the target without including large areas of the site without anomalies of interest within the estimated target region. Additionally, the definition of the target acknowledges the uncertainty inherent in making decisions across a large site from limited information. The decision maker determines the reliability that is necessary for the characterization decision and this reliability defines the extent of the target areas. As seen in Figure 24, this approach is not limited to defining only a single target, but will define the individual extent of multiple targets provided there are some sample data within those targets. Results for this example, show that fewer than three percent of the remaining decision cells contained anomalies of interest for any value of R_D above 0.70.

Application 3

The third example application builds on the first application by extending the estimates with secondary information. The stratified approach to developing the secondary information has not previously been applied to UXO site characterization problems and may prove extremely useful because of its simplicity.

Incorporating Secondary Information

A distinct advantage of geostatistical approaches to estimating spatially distributed properties is the flexibility with which these approaches can incorporate secondary information into the estimation of the primary variable. These approaches include cokriging (Wackernagel, 1998; Goovaerts, 1997), kriging with a locally varying mean (Deutsch and Journel, 1998), kriging with an external drift (Journel and Huijbregts, 1978), collocated cokriging (Xu, et al., 1992; Almeida and Journel, 1994) and stratified kriging (Stein et al., 1988; Stein, 1994). For the problem of UXO site characterization, we would like to incorporate secondary information that is quantified from an archival search report, or airborne imagery, with the primary data collected along sample transects. Previous work on this project has investigated the use of cokriging, kriging with a locally varying mean and collocated cokriging to integrate prior information with transect sample data.

The choice of which approach to use to incorporate secondary information into kriging estimates is based mainly on the character of this secondary data relative to the primary data. Three aspects of the secondary data must be considered: 1) the spatial extent of the secondary data; 2) the units of measurement of the secondary data as compared to the units of the primary variable; and 3) the number of secondary variables. The practical application of traditional cokriging approaches is generally limited to the case where the secondary data are oversampled with respect to the primary variable but are not available at all locations to be estimated. This situation of the secondary data existing at the same locations as the primary data as well as at additional locations is the “partially heterotopic” case as defined by (Wackernagel, 1998). If the secondary data are available at all estimation locations, then the spatial correlation of the secondary variable will filter that of the primary variable under traditional cokriging. Additionally, numerical instability in the solution of the cokriging system can occur (Goovaerts, 1997).

If the units of the secondary data match those of the primary variable, then kriging with a locally varying mean is an excellent choice for data integration. Kriging with a locally varying mean replaces the stationary mean in the simple kriging formulation (equation 8) with one that varies spatially. The variation in this locally varying mean must be smooth without sharp discontinuities (Deutsch and Journel, 1998). Generally if the secondary data are obtained from a geophysical sensor that locally averages the signals of the subsurface anomalies, then this smoothly varying condition will be met. Kriging with a trend (KT) does not require that the units of the primary and secondary data be the same. KT limits the trend model to two terms that can be thought of as the value of the trend when the value of the primary data is equal to zero and a scaled version of the secondary data. As in the case of kriging with a locally variable mean, the secondary data need to be smoothly varying. Collocated cokriging allows for the incorporation of more than one secondary variable in the estimation of the primary variable.

For the problems of UXO site characterization, secondary data that exist at all locations to be estimated are of most interest. The utility of both kriging with a locally varying mean and collocated cokriging with spatially exhaustive secondary information have been previously examined for applications to UXO site characterization problems as part of this project (Saito et al., 2002).

For this demonstration, we use a prior classification of the site into three strata as the secondary information. This type of classification could readily be generated from archival search reports available at many sites. This approach does not require any prior knowledge of the data or site history within the strata, but only the geometric definition of portions of the site that may have different histories. Estimation within strata proceeds as follows. The site is divided into a number of strata and the mean value of the sample data lying within each stratum is calculated. For every data location, the residual between the sampled value and the stratum mean are calculated. These residuals are used to construct a single residual variogram for the whole site and the spatial estimation of the residuals is then completed across the site. For each location, the estimated residual is added back to the corresponding stratum mean to get the final estimate of the primary variable. This technique is chosen because it is well suited to information contained in archival search reports. Previous applications of this stratum method for spatial estimation include Stein (1994) and Stein et al. (1998). The use of prior information to simply define different strata with assumed differences in anomaly/UXO intensity is perhaps the most simple use of secondary data to aid in spatial estimation.

Based on the available historical information, the example hypothetical site is divided into three strata defining the probability of UXO within each: Low, Medium and High. The spatial extent of these three zones are shown schematically in Figure 26. These three zones are used to define the secondary information for each of the three quantities being estimated: total number of anomalies, number of anomalies of interest, and the probability of having at least one anomaly of interest. It is noted that it is not necessary to determine the actual values defining “low”, “medium” and “high” but only to determine the spatial extent of the regions with presumed similar behavior.

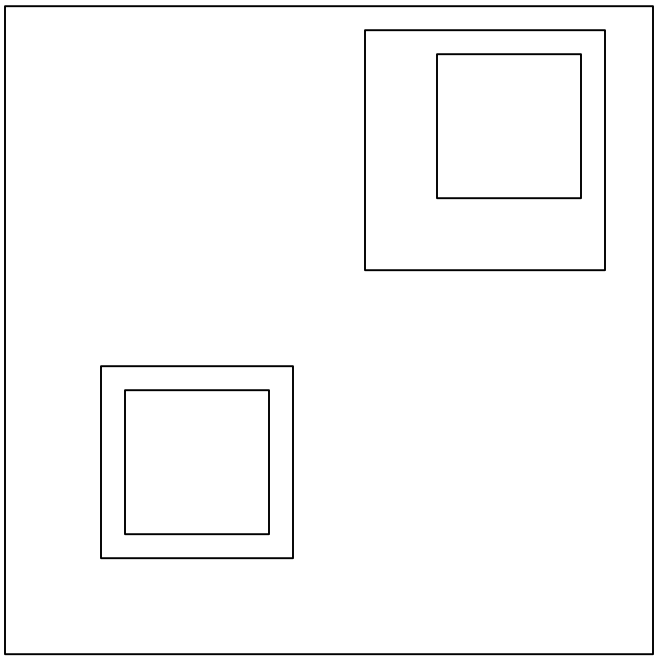


Figure 26. Schematic representation of the hypothetical site with three different strata representing: low (blue), medium (green) or high (red) suspected relative intensities of anomalies based on historical data.

The mean value of each quantity being estimated within each of the strata is calculated from the sampling transects as they intersect the three different strata. These mean values and the total number of 25x25 meter transect sample cells lying within each zone are shown in Table 6. An inherent assumption in using the stratum approach to secondary data is that there are enough samples available within each stratum to estimate a mean value that is representative of the true distribution of anomalies within the stratum.

Table 6. Means of each quantity within the three different strata.

Stratum	Total Anomalies	Anomalies of Interest	Indicator	Number of Samples
Low	1.357	0.066	0.057	457
Medium	1.966	0.291	0.217	175
High	4.149	1.473	0.77	74

Residual Variograms

Each sample value along the transect is subtracted from the corresponding stratum mean to produce a residual value. These residuals are then used to calculate and model residual variograms. The residual variograms for each of the three attributes being modeled are shown in Figure 27 and the parameters of the models fit to these residual variograms are given in Table 7. For the residual values, the indicator variogram is fit with two nested models to better

approximate the experimental variogram data. These residual variograms show significantly higher nugget effects and shorter ranges than the variograms calculated on the raw sample data. This behavior is expected as the residuals should represent the uncorrelated random variability about the subtracted model. The small amount of spatial correlation remaining is due to the model, the mean values within each zone, being an approximation and not a perfect description of the sample data. This behavior is what would be expected for transect sampling at a field site.

Table 7. Variogram model parameters used to fit the variograms of the residuals.

	Model Type	**Nugget**	**Sill**	**Range (m)**
Total Anomalies	Spherical	1.8	0.31	90
Number of Anomalies >= 3 nT/m	Spherical	0.25	0.117	40
Probability of one anomaly >= 3 nT/m	Spherical	0.058	0.016	30
	Exponential	NA	0.022	350

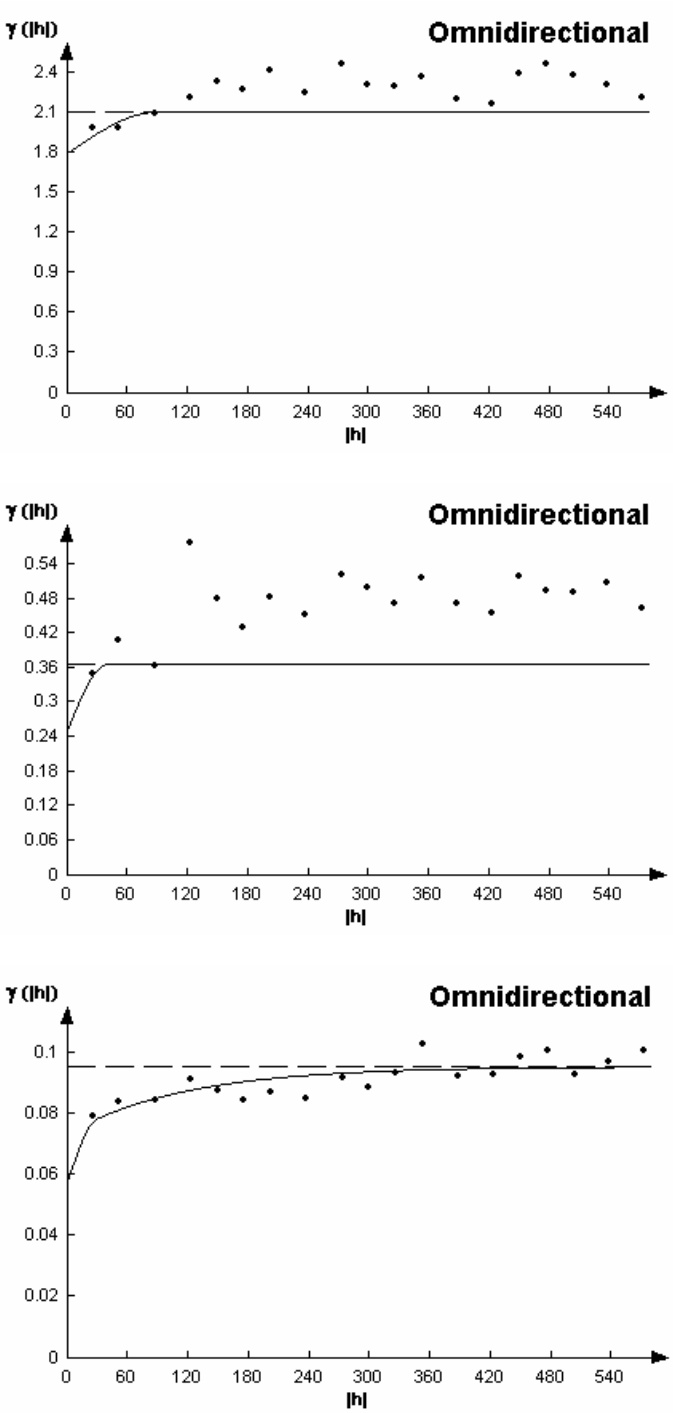


Figure 27. Residual variograms for the total number of anomalies (upper image), anomalies of interest (middle image) and indicator (lower image) values.

Estimation

The estimates of the residuals are shown in Figures 28, 29 and 30. The most notable difference between these estimates and the estimates shown in Figures 14, 15 and 16 is that the area of estimation is smaller. This is due to the shorter range values of the residual variograms relative to the raw value variograms.

The residual values at any location, ***x***, are calculated as:

$$residual(\mathbf{x}) = stratum_mean(\mathbf{x}) - data(\mathbf{x}) \qquad (27)$$

therefore a positive residual means that the average within the stratum over estimates the actual value at that location and a negative residual under estimates the actual value at that location. The majority of the areas of estimated residuals shown in Figures 28, 29 and 30 have values near 0.0 meaning that at these locations the mean value of the stratum will prevail as the estimate.

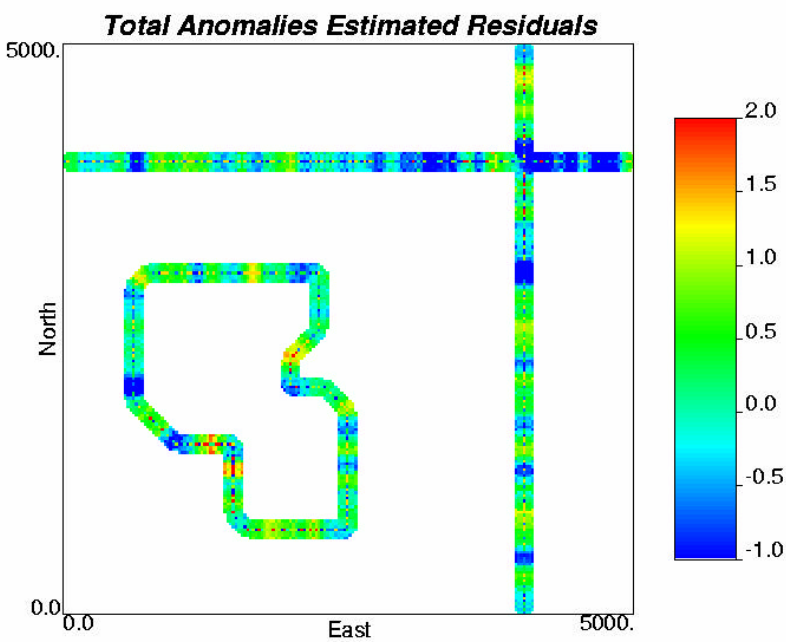


Figure 28. Estimated residual values for the total number of anomalies.

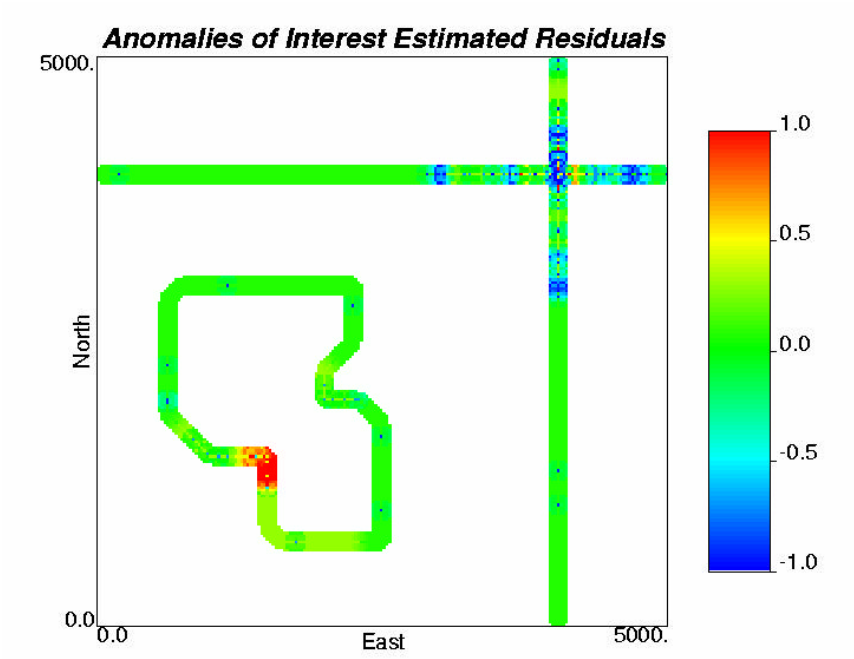


Figure 29. Estimated residuals for the anomalies of interest.

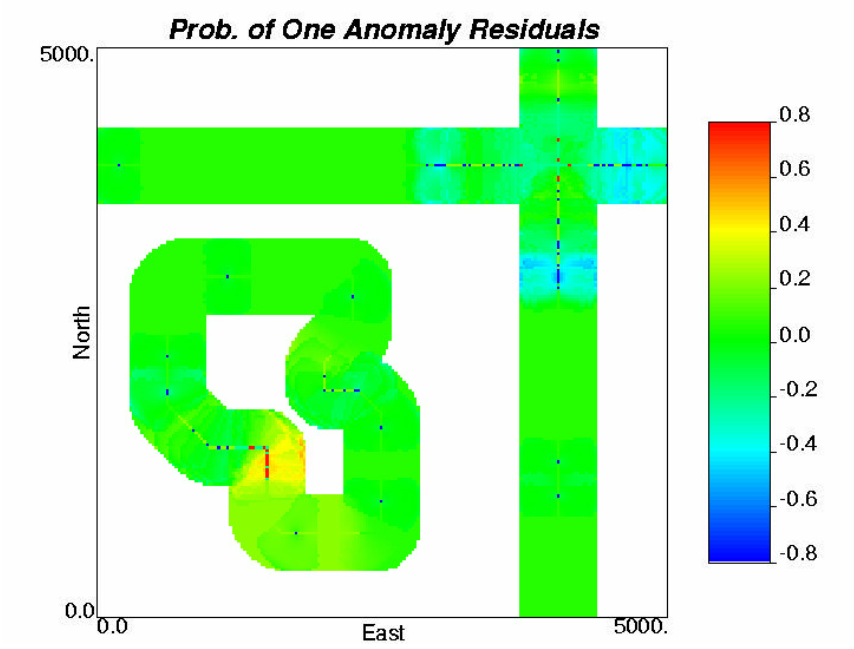


Figure 30. Estimated residuals for the probability at least one anomaly of interest.

The estimated residuals are added back to the mean value of each property within each of the zones defined as prior information. The final estimates are the sum of the estimated residuals and the zone means. These final estimates are shown in Figures 31, 32 and 33. As was done for the cross-validation and the estimation results in example application 1, the results of the estimation are summarized in tables for the total number of anomalies and the number of anomalies of interest and in a graph for the probability of at least one anomaly of interest.

The estimation results in Figures 31, 32 and 33 clearly show the locations of the different stratum and the effect that the mean values in each stratum have on the resulting estimates. The incorporation of the secondary information provided by using the strata allows for estimates at all locations. However, for locations that are further away from a sample point than the range of the residual variogram, the estimated value reverts to the stratum mean. This effect places considerable importance on being able to determine representative mean stratum values from the available transect data. For the “low” and “medium” strata, the mean values are 1.36 and 1.97 respectively. For locations that are beyond the variogram range from any sample point, these means are used to make the estimate and any values corresponding to these means will then be rounded up to the next highest integer, 2, prior to evaluating the results of the estimation through jackknifing.

The jackknifing results for the estimation of the total number of anomalies are given in Table 8. Across all estimates, 22.3 percent are correct, 59.9 percent are false positive and 17.7 percent are false negative. These results are within +/- one percent of the results obtained for the estimates in the first example application with the major difference being that the use of the strata allows for estimations across the entire site. The majority of the false positive estimates are where two anomalies were estimated but there were zero or one actual anomalies in those locations. These false positives are a direct result of the mean values in the “low” and “medium” strata (Table 8) and the decision to round each estimate up to the next highest integer value. All estimates in these two strata that are beyond the variogram range from the nearest sample point are estimated as having two anomalies. However, the reality is that much of this area has only one, or no anomalies at all and therefore a large number of false positives are estimated. The majority of the false negative results occur where two anomalies are estimated in the low and medium strata, but three or four anomalies actually exist.

Table 8. Results of the estimation of the total number of anomalies using prior information as determined through jackknifing.

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.004	0.006	0.007	0.004	0.001	0.001	0.000	0.000	0.000	0.000
2	0.237	0.314	0.212	0.098	0.037	0.014	0.004	0.002	0.000	0.000
3	0.008	0.008	0.006	0.003	0.002	0.000	0.000	0.000	0.000	0.000
4	0.001	0.001	0.001	0.001	0.000	0.001	0.000	0.000	0.000	0.000
5	0.001	0.003	0.005	0.006	0.004	0.002	0.002	0.001	0.001	0.000
6	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

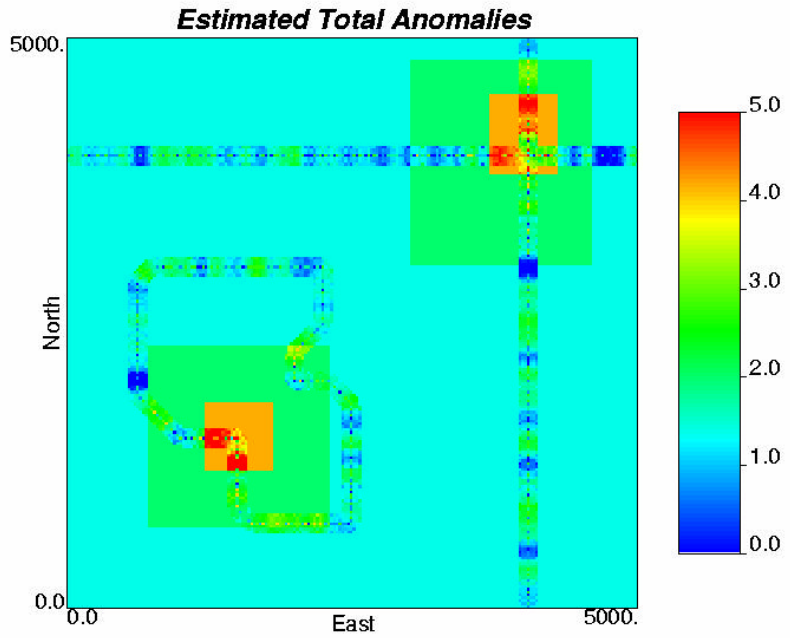


Figure 31. Estimates of the total number of anomalies as created using prior information.

The results of the estimation of the number of anomalies of interest using the prior information are shown in Figure 32 and Table 9. The effect of dividing the site into strata is clearly evident in the map of estimates in Figure 32. Across all 40,000 estimates, the correct estimates account for 7.0 percent, the false positives account for 91.1 percent and the false negatives account for just 1.9 percent. These results are within +/- four percent of the estimates made without using the strata, but using the strata allows for estimates across the entire site. The large number of false positives is due to the mean value estimated for the low and medium strata and the conservative decision to round each estimate up to the next highest integer value. The means of the low and medium strata are 0.07 and 0.29 respectively (Table 6) and for any location where the stratum mean serves as the estimate, these values will both be rounded up to 1.0. These means and the rounding up decision lead to a large number of locations were one anomaly of interest is estimated, but no anomalies of interest exist (Table 9). The larger percentage of false positives results in a low percentage of false negative decisions. The majority of these false negatives occur when one anomaly is estimated and two actually exist (Table 9).

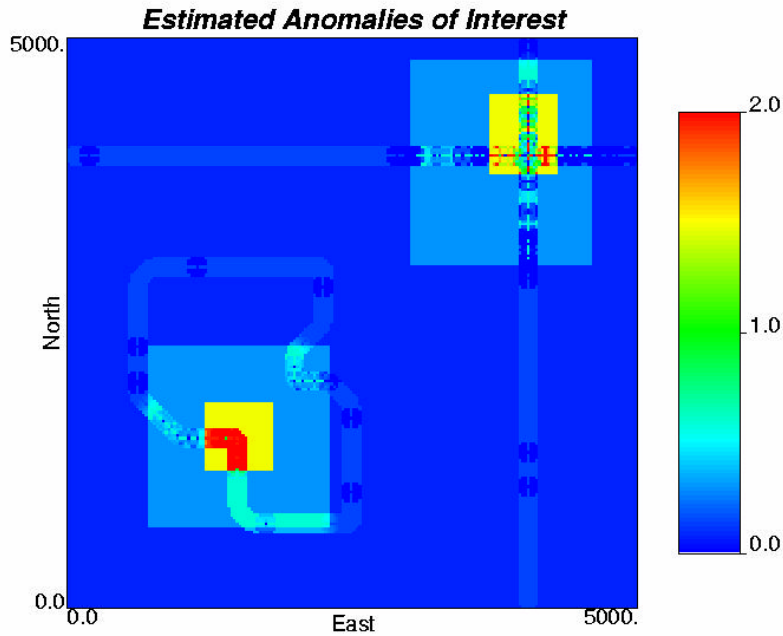


Figure 32. Estimates of the number of anomalies of interest as created using prior information.

Table 9. Results of the estimation of the number of anomalies of interest using prior information as determined through jackknifing

	0	1	2	3	4	5	6	7	8	9
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.889	0.067	0.012	0.003	0.001	0.000	0.000	0.000	0.000	0.000
2	0.010	0.009	0.003	0.002	0.001	0.000	0.000	0.000	0.000	0.000
3	0.002	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The final attribute to be estimated is the probability of having at least one anomaly of interest at every location. This is done using the variogram calculated on the residuals of the indicator data and the discretization of the site into three strata. The results of the estimation are shown in Figure 33 and the evaluation of these estimates using jackknifing is summarized across a range of R_D values in Figure 33. As seen in the estimation results for the other two attributes, the segmentation of the site domain into strata has a strong effect on the final estimates. High probabilities of at least one anomaly of interest occur mainly in the suspected target area in the southwest portion of the site. The suspected target region in the northeast corner of the site has probabilities in the 0.5 to 1.0 range.

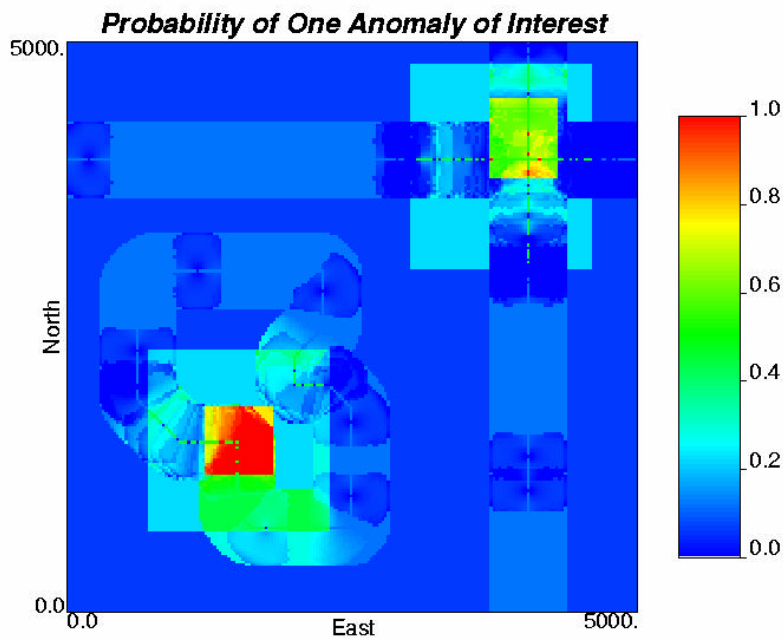


Figure 33. Estimates of the probability of at least one anomaly of interest as created using prior information.

The results of the decisions made at different levels of R_D (Figure 34) show a stepped behavior that has not been seen in previous results (e.g., Figure 17). The discrete jumps or drops in the proportions of different decisions are the result of stratifying the site into three distinct zones. For example near an R_D value of 0.95, there is a large increase in the number of false positive results and a corresponding decrease in the number of correct decisions. Below this level of R_D, the "low" stratum is not slated for additional surveying, but at R_D values about approximately 0.95, it becomes necessary to do detailed surveying throughout the low stratum area. Detailed surveys across the entire low area do find additional anomalies of interest (drop in false negatives for $R_D > 0.95$ in Figure 34) but they are generally inefficient and the number of false positives increases dramatically as this region is surveyed. The other jumps in the curves correspond with the values of R_D at which the high and medium strata are also slated for detailed surveying.

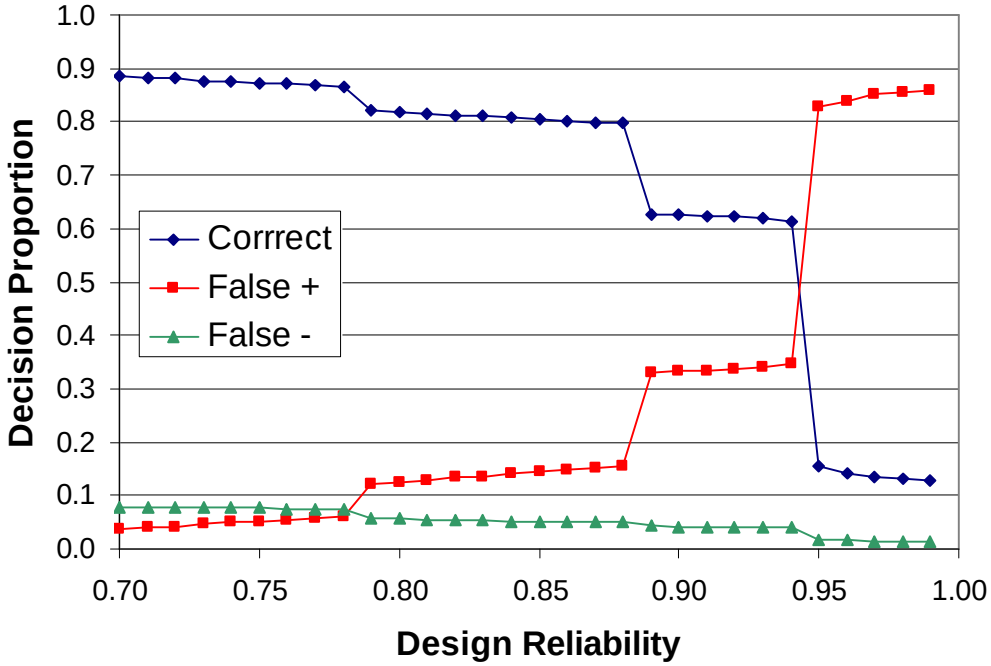


Figure 34. Proportions of different decision results based on the map in Figure 33 for a range of R_D.

Application 3 Summary

This example application demonstrates a simple approach for incorporating prior information into UXO site characterization activities. Results of dividing the site into strata based on archival information allows for estimations across the entire site, but the majority of these estimates are simply the stratum mean. The results of this example show that estimates can be made across the entire site and the results are comparable to estimates that can only be made across much smaller portions of the site when prior information is not used. Unfortunately, these results only hold for the estimates of the total number of anomalies and the number of anomalies of interest while the proportion of correct and false positive results based on the probability estimates are degraded relative to results when the strata are not used.

Application 4

The fourth example application builds on the second application by examining two different approaches to identifying the second iteration of sampling. These two approaches are: 1) take a single meandering path transect within the regions of highest target boundary uncertainty as defined using the initial sampling and probability indicator kriging; and 2) locate additional straight, parallel transects in areas midway between the existing transects. The results of the decisions are again compared to the underlying true values using a jackknifing procedure to determine how well both cases of additional sampling did in improving the results.

Second-Phase Samples: Meandering Path

The first approach uses the results of using PIK on the original 14 transect samples to estimate the probability of at least one anomaly of interest at every location (Figure 23). For these examples, anomalies of interest are those above 5.0 nT/m and the locations with high probability of these anomalies are believed to correspond to target locations. In the original probability map (Figure 23), regions of maximum uncertainty correspond to areas with a probability near 0.50 of having at least one anomaly of interest. For this example, the maximum uncertainty regions are defined as those with probabilities between 0.4 and 0.6 and these are shown in Figure 35.

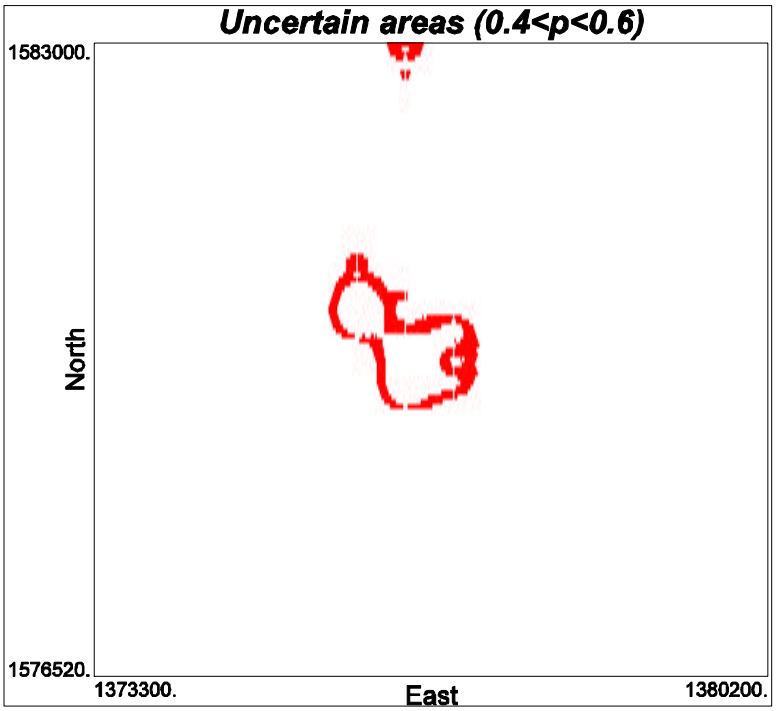


Figure 35. Locations of maximum uncertainty in the presence or absence of an anomaly of interst at the Laguna N11 site. The red regions contain cells with probabilities of at least one anomaly of interest between 0.4 and 0.6.

The areas shown in red in Figure 35 correspond to the areas in which the second round of samples should be located to achieve the maximum reduction in uncertainty with respect to the location of anomalies of interest. This entire high uncertainty region, including the area near the northern boundary of the site is sampled with meandering path transects as shown along with the original straight sampling transects in Figure 36.

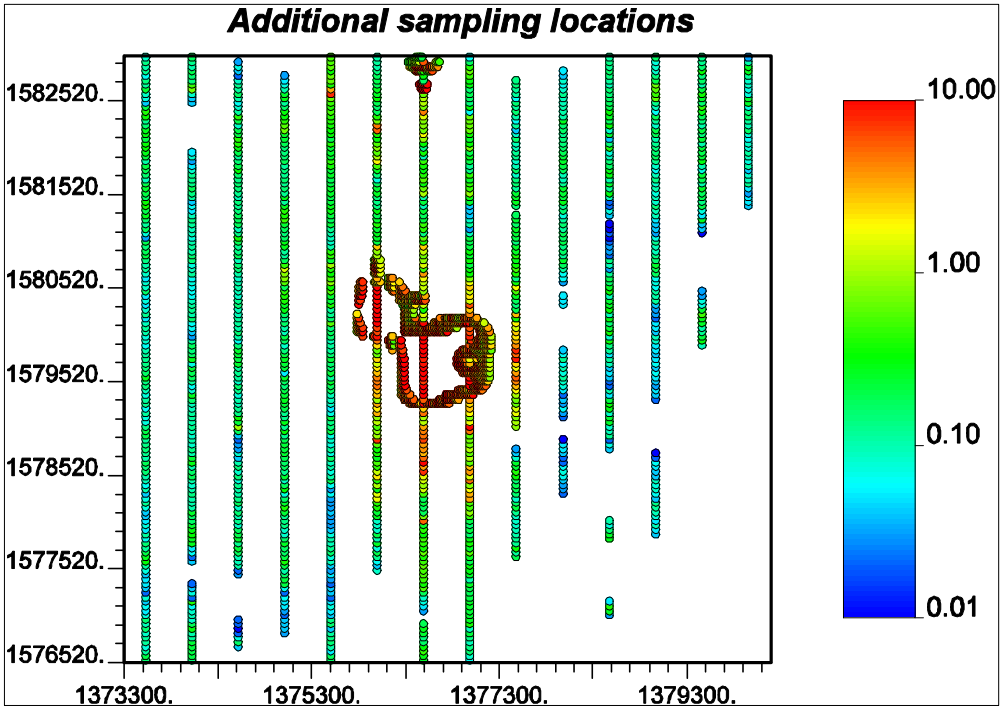


Figure 36. Original straight transects and the meandering path transect obtained in the second round of sampling. The log10 color scale shows the maximum analytic signal strength in nT/m within each sample cell.

The samples obtained by surveying this high uncertainty region show little spatial correlation when transformed to the indicator [0,1] space. This result is expected as these samples were specifically located in an area with estimated probabilities of at least one anomaly of interest near 0.5. Given these estimates and the fact that a sample can only take the values 0.0 or 1.0, it is expected that nearly every other sample cell would be a 1.0 with the other half of the sample cells containing a 0.0. This independence from one cell to the next along the meandering path creates a nugget effect variogram (i.e., no spatial correlation) and therefore the original indicator variogram (Figure 22) as calculated off of the straight transects is used in the PIK process to create the updated estimates of the probability of at least one anomaly of interest. This updated probability map is shown in Figure 37 and this map can be compared to Figure 23 to asses the impact of the meandering path transect on the overall estimation of probabilities. The location of the larger meandering path transect in Figure 36 is readily apparent near the center of Figure 37.

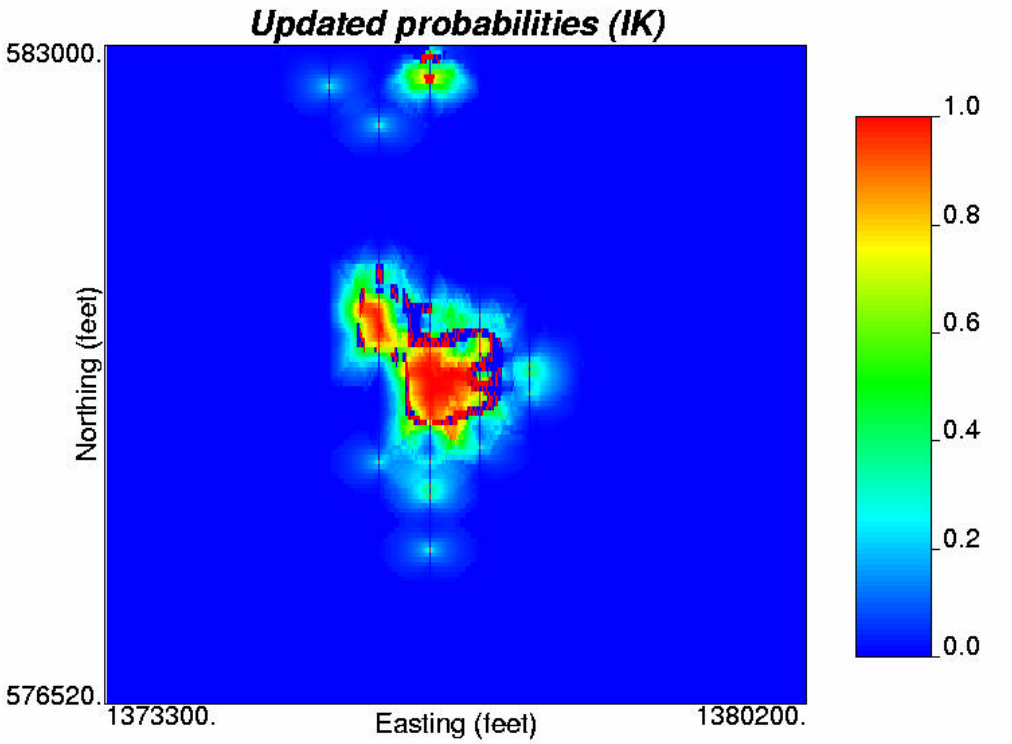


Figure 37. Updated probability map showing the probability of at least one anomaly of interest at every location. This map was updated from the original map by incorporating a meandering path transect in the region of greatest uncertainty.

Second-Phase Samples: Straight Transects

A second approach to locating additional transects is to locate a single straight transect midway between the existing straight and parallel transects obtained in the initial round of sampling. This approach would result in an increase in the probability of detecting a target of a given size, or a large decrease in the size of a target area that could be detected for the same level of confidence used in the initial transect design. This approach of "infilling" with additional straight transects was applied to the Laguna N11 site and the transects resulting from the initial and the second iteration of sampling are shown in Figure 38. The additional transects could not be located exactly midway between the original transects for this example due to the gaps in the original field survey data, but they were located as close to midway between the existing transects as possible.

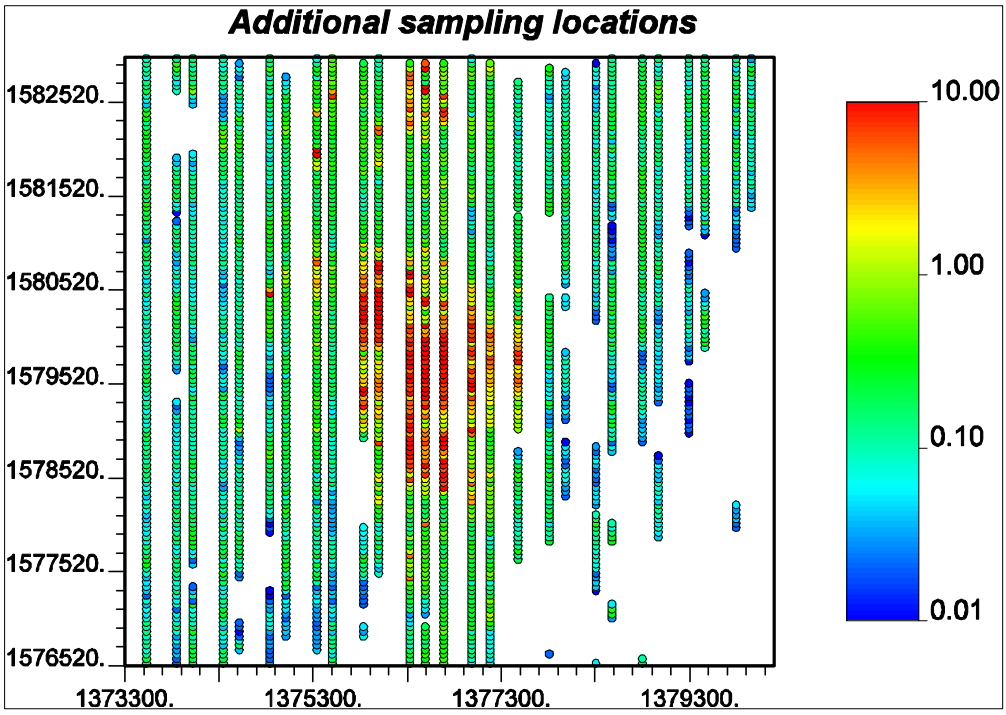


Figure 38. Straight and parallel sampling transects obtained in the original and second round of sampling. The log10 color scale shows the maximum analytic signal strength in nT/m within each sample cell.

The additional transect data are transformed to the 0,1 indicator values based on whether or not at least one anomaly of interest is found within each sample cell and the indicator variogram is calculated and modeled using both the original and second-round sample data. Unlike the first approach of locating a meandering path in the area of highest uncertainty, this second approach of infilling additional straight transects across the entire site does not preferentially sample regions with low spatial correlation. Therefore, the indicator variogram calculated using all the sample data does display spatial correlation and can be used directly in the estimation of the probability values for the updated probability map. The indicator variogram constructed using all the straight transect data and the updated probability map created with this variogram and the second round of straight transects are shown in Figures 39 and 40 respectively.

The variogram fit to the full set of indicator data has a nugget value of 0.015, and is fit with two nested spherical variogram models having ranges of 400 and 2800 feet and sill values of 0.013 and 0.05. The variogram model fit to both sets of straight transect data is very similar to that fit to the original transect data. This similarity indicates that the 14 original transects were adequate to recover the spatial variation across the entire site.

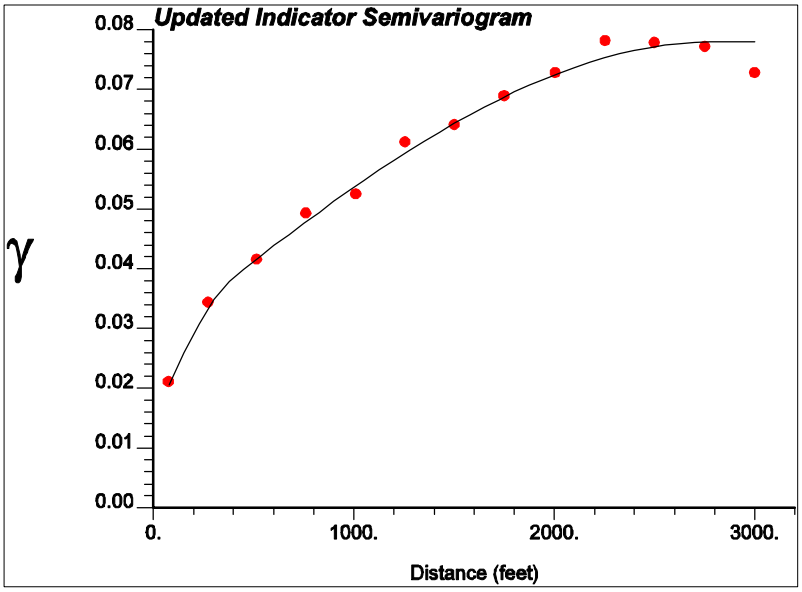


Figure 39. Updated indictor variogram calculated from the original and second round of straight transects.

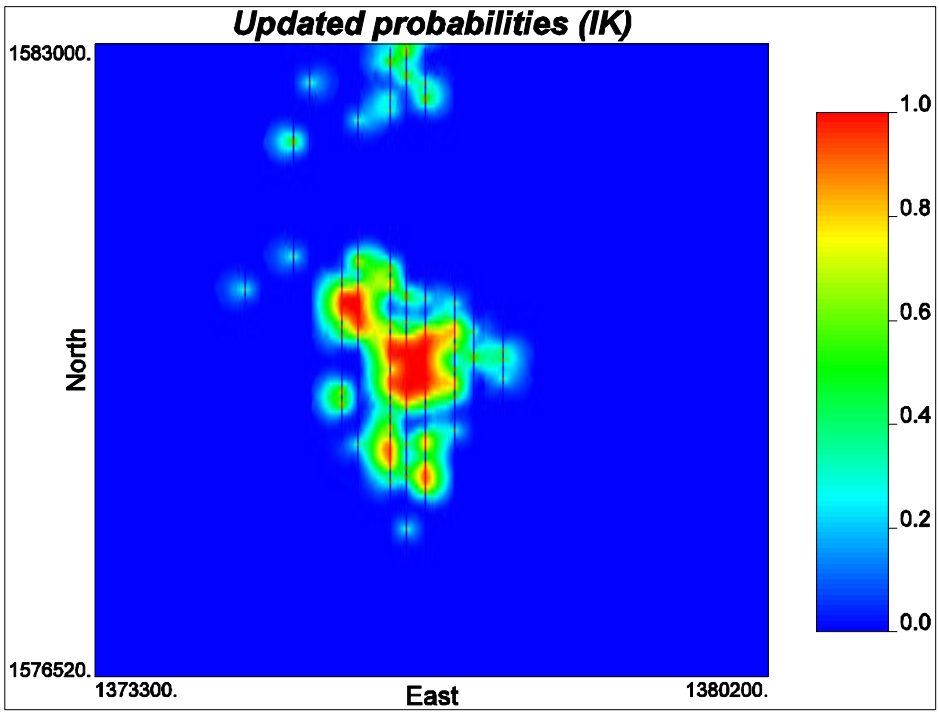


Figure 40. Updated map showing the probability of at least one anomaly of interest everywhere at the site based on the original and second round of straight transects.

Second-Phase Sampling Results

The final comparison between the two methods of locating second-phase samples is done by comparing the results of the site characterization decisions that would be made using the two different probability maps (Figures 37 and 40). This is done in the same way as in the previous example applications by making decisions for a range of R_D values and comparing the decisions to the known data that were held back from the analysis and then reporting the proportions of the different decision results. Also, the proportions of anomalies of interest found or left behind using the different sampling designs are examined. These comparisons are shown in Figure 41.

The top graph in Figure 41 shows the proportion of the total anomalies that are found and that are left behind as a function of the chosen R_D. These graphs are calculated using the actual anomalies and their locations. Although the decisions are made at the 15x48 decision cell scale, no aggregation of the true anomalies into the decisions cells is considered in this top graph. Therefore, if a false negative decision is made for a decision cell and that cell contains 5 anomalies of interest, this will lead to 5 anomalies being left behind. The results in the top graph of Figure 41 show that at levels of R_D >= 0.95, 93 percent or more of all anomalies of interest are found and 7 percent or less are left behind. The sampling design that uses two sets of straight transects produces the highest proportion of anomalies found across all values of R_D, although at the highest values of R_D, there is little difference in the results when using just the original 14 transects or the original transects combined with either the meandering path or the 13 additional straight transects.

The middle graph on Figure 41 shows the proportion of correct and false positive decisions made as a function of R_D for each of the three sampling designs. These results are calculated by comparing the decisions made at the decision cell scale to the true decision that would be made at the same decision cell scale if the site were perfectly characterized. For these results, the original 14 transects combined with the meandering path give the highest proportion of correct decisions and the lowest proportion of false positive decisions across values of R_D from 0.70 to approximately 0.97. For design reliabilities above 0.97, the 14 original transects combined with the 13 additional straight transects produce the highest proportions of correct decisions and the lowest proportions of false positive decisions. However, the differences in the proportions of correct and false positive decisions across the three different sampling designs are minimal for all values of R_D considered. These results indicate that little improvement in the decisions resulted from adding the second round of sampling. This result is true for either the meandering path or the additional straight transect second-phase sampling designs.

The bottom image in Figure 41 shows the proportion of false negative decisions made for each of the three different sampling designs across the full, 0.0 to 1.0, range of R_D values. This lower graph shows that the addition of second phase samples, either the single meandering path or multiple straight transects, decreases the number of false negative decisions relative to just using the original 14 straight transects. For different values of R_D, the second-phase sampling approach that gives the lowest proportion of false negative results varies as a function of R_D. For values of R_D greater than 0.95, the choice of sampling design is nearly inconsequential with respect to the number of false negative decisions with the original sampling design and the two second-phase designs all achieving a false negative proportion of approximately 4.0E-03 or less.

Another way to view the decision results for this example application is to map the extent of the site that is slated for additional surveying versus that which is left as requiring no further action. These maps are compared for four different levels of R_D: 0.70, 0.80, 0.90 and 0.95 in Figure 42. The results for the second iteration samples being obtained along thirteen straight transects are shown on the left and the results obtained with a single meandering path as the second iteration are shown in the right column. The two approaches are fundamentally different in that the additional straight transects are designed to learn more about the entire site while the meandering path is specifically designed to better resolve the extent of the target area. As the R_D increases, the straight transect sampling results in more isolated areas outside the main target area being scheduled for detailed surveying, while the results based on the additional meandering transect generally produces changes in and around the main target areas. The lack of spatial correlation in the samples along the meandering path transect limits the ability of these data to inform nearby locations that were not sampled and thus limits the effect of these samples to the actual sample locations.

Application 4 Summary

The major result of the fourth example application is that the original 14 transects provide enough information to characterize the site and in particular, these transects provide enough data to adequately define the boundaries of the target regions. The 13 additional transects, or the meandering path transect, taken as a second-phase of sampling do not significantly alter the results from those obtained with the original 14 transects. The different approaches to the second-phase of sampling do define slightly different areas for additional surveying; however, the decision results across the entire site are essentially identical.

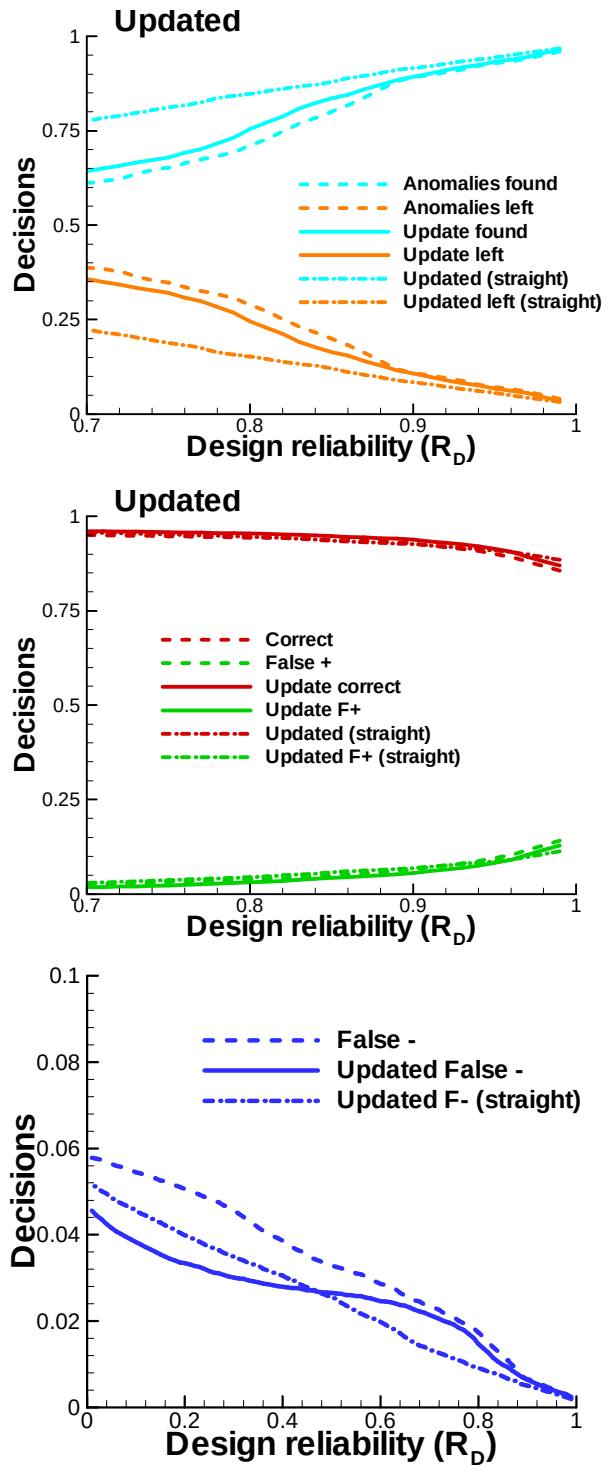


Figure 41. Decision results as a function of R_D for the initial sampling and two different approaches to the second iteration of sampling for the Laguna N-11 site.

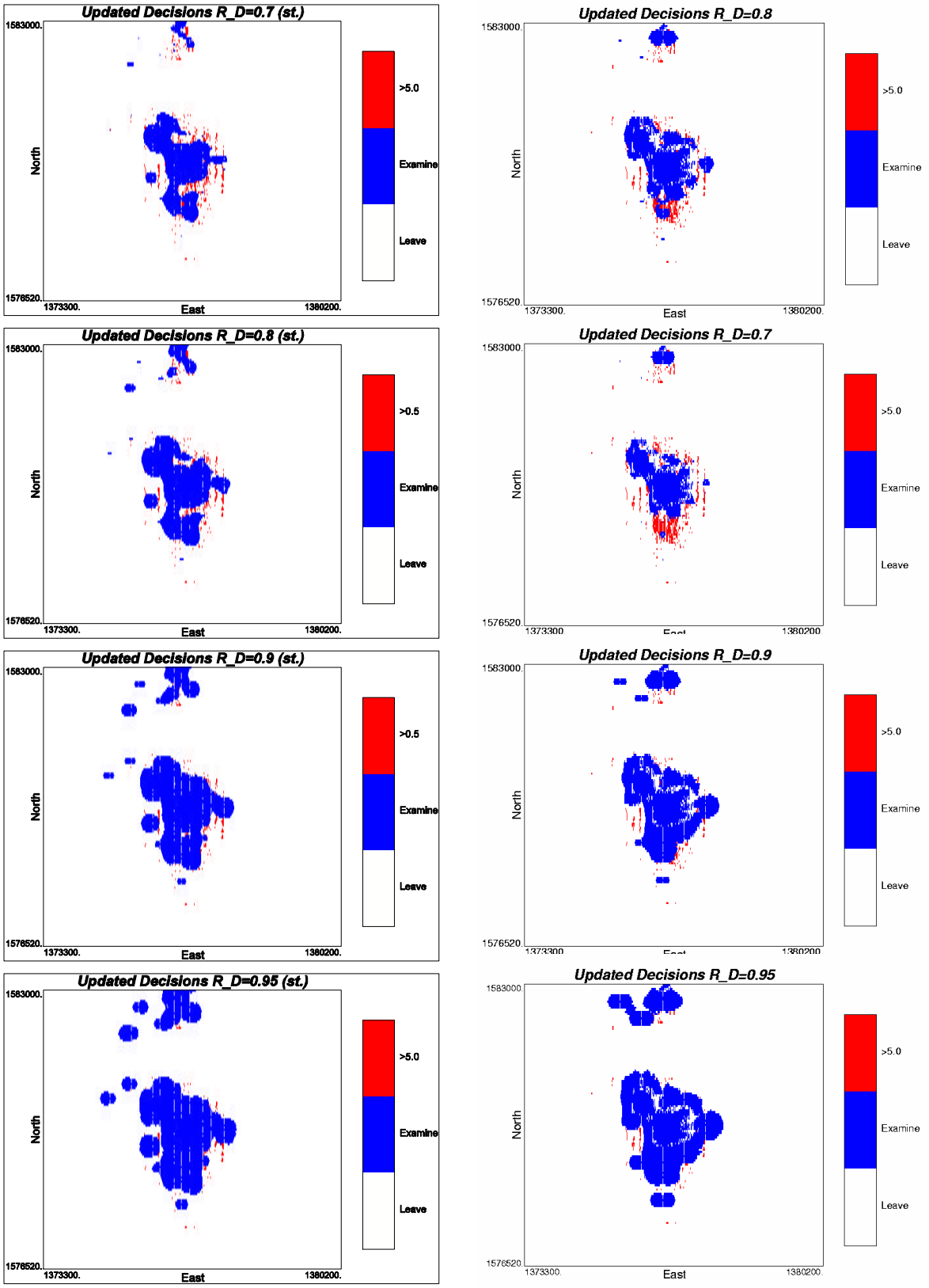


Figure 42. Decision maps for four different levels of R_D: 0.70 (top); 0.80 (second from top), 0.90 (second from bottom), 0.95 (bottom). The results in the left column are for the straight transects. The right column has the results for the meandering path.

Conclusions

This report has addressed one potential limitation in the spatial estimation procedure and has provided a number of simple applications of different aspects of the site characterization approach across four different examples. The results of the investigation into whether or not it is necessary to correct the kriging procedure to account for the finite domain effect associated with transect data has been answered. All results examined over multiple different transects, transect widths and proportions of site sampling coverage show that there is no appreciable difference between finite domain kriging and traditional ordinary kriging. Therefore it is possible to make accurate estimates from transect data using the readily available ordinary kriging algorithm.

In general all four example applications looked at sites where less than 10 percent of the was sampled and the site characterization tools were able to produce decision results that limited the number of false negatives to 5 percent or less. Significant differences in the amount of correct and false positive decisions exist across the different attributes that are estimated and the different approaches to decision making. The specific conclusions that can be drawn from each of the example applications are:

1) The probability mapping approach provided superior decision making results when compared to mapping the total number of anomalies or mapping the number of anomalies of interest. To some extent this result was due to the decision to round up any fractional estimates of number of anomalies to the next integer value in the two anomaly mapping approaches. This decision creates artificially high levels of false positive results. However, mapping the number of anomalies requires that some type of fairly arbitrary decision be made as to what fraction of an estimated anomaly should be counted as a full anomaly. For this work, the most conservative possible approach was taken. The probability mapping approach avoids having to make the decision as to what fractional value needs to be counted as a true anomaly by requiring a decision on the acceptable reliability to which a site needs to be characterized. The cross-validation step provided excellent predictions of the results that were obtained in the actual estimations. This step can is generally used to compare different variogram models and options in the setup of the kriging algorithm, but the results obtained here demonstrate that cross-validation can also be used to gain confidence in the accuracy of the estimates across a site using just the data that have already been collected.

2) This example application demonstrated the ability of geostatistical mapping techniques to estimate a fourth attribute, maximum signal strength, from a limited number of equally spaced parallel transects with a 15 foot width. These same data were also used to map the probability of one anomaly of interest across the site and these results were used to map the extent of the target area where the target area is defined by the locations of the anomalies of interest. The results show that this technique is able to efficiently identify the outlines of the target without including large areas of the site without anomalies of interest within the estimated target region. Additionally, the definition of the target acknowledges the uncertainty inherent in making decisions across a large site from limited information. The decision maker determines the reliability that is necessary for the characterization decision and this reliability defines the extent of the target areas. This approach is not limited to defining only a single target, but will define the individual extent of multiple targets provided there are some sample data within those targets.

3) This example application demonstrated the utility of a simple approach for incorporating prior information into UXO site characterization activities. Results of dividing the site into strata based on archival information allows for estimations across the entire site, but the majority of these estimates are simply the stratum mean. The results of this example show that estimates can be made across the entire site and the results are comparable to estimates that can only be made across much smaller portions of the site when prior information is not used. These results only hold for the estimates of the total number of anomalies and the number of anomalies of interest while the proportion of correct and false positive results based on the probability estimates are degraded relative to results when the strata are not used.

4) The major result of the fourth example application is that the original 14 transects provide enough information to characterize the site and in particular, these transects provide enough data to adequately define the boundaries of the target regions. The 13 additional transects, or the meandering path transect, taken as a second-phase of sampling do not significantly alter the results from those obtained with the original 14 transects. The different approaches to the second-phase of sampling do define slightly different areas for additional surveying; however, the decision results across the entire site are essentially identical.

Acknowledgements

The authors acknowledge Jeff Gamey, Bill Doll and David Bell of Oak Ridge National Laboratories for providing the magnetometer data for the Pueblo of Laguna N-11 site. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed-Martin Company, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000

References

Almeida, A.S. and A.G. Journel, 1994, Joint simulation of multiple variables with a Markov-type coregionalization model, *Mathematical Geology*, 26 (5), pp. 565-588.

Byers, S. and A. E. Raftery, 1998, Nearest-neighbor clutter removal for estimating features in spatial point processes, *Journal of the American Statistical Association*, 93 (442) pp. 577-584.

Christakos, G and D.T. Hristopulos, 1996, Stochastic indicators for waste site characterization, *Water Resources Research*, 32 (8), pp. 2563-2578.

Conover, W.J., 1980, *Practical Nonparametric Statistics, Second Edition*, Wiley and Sons, New York, 493 pp.

Deutsch and Journel, 1998, *GSLIB: Geostatistical Software Library and User's Guide, Second Edition*, Oxford University Press, New York, 369 pp.

Deutsch, C.V., 1994, Kriging with strings of data, *Mathematical Geology*, 26 (5), pp. 623-638.

Deutsch, C.V., 1993, Kriging in a finite domain, *Mathematical Geology*, 24 (1), pp.41-52.

Doll, W.E., T.J. Gamey, L.P. Beard, D.T. Bell and J.S. Holladay, 2003, Recent advances in airborne survey technology yield performance approaching ground-based surveys, The Leading Edge, May, pp. 420-425.

Efron, B. and G. Gong, 1983, A leisurely look at the bootstrap, the jackknife and cross-valdiation, *The American Statistician*, 37 (1), pp. 36-48.

Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford Univ. Press, New York.

Harr, M.E., 1987, Reliability-based design in civil engineering, Dover, Mineola, New York, 290 pp.

Journel, A.G., 1983, Non-parametric estimation of spatial distribution. *Mathematical Geology*, 15(3): 445-468.

Journel, A.G., and Ch. J. Huijbregts, 1978, *Mining Geostatistics*, Academic Press, London, 600 pp.

Konikow, L.F. and J.D. Bredehoeft, 1992, Ground-water models cannot be validated, *Advances in Water Resources*, 15, pp. 75-83.

McKenna, S.A., 1998, Geostatistical Approach for Managing Uncertainty in Environmental Remediation of Contaminated Soils: Case Study, *Environmental and Engineering Geoscience*, Vol. IV, No. 2, Summer 1998, pp. 175-184.

McKenna, S.A, H. Saito and P. Goovaerts, 2001, Bayesian approach to UXO site characterization with incorporation of geophysical information, SERDP Project UX-1200 deliverable, Dec. 30th, 51 pp.

McKenna, S.A., 2001, Application of a Doubly Stochastic Poisson Model to the Spatial Prediction of Unexploded Ordnance, In: *Proceedings of the 2001 Annual Meeting of the International Association of Mathematical Geology*, Cancun, Mexico, Sept. 6-12, pp. 21.

Olea, R.A., 1999, *Geostatistics for Engineers and Earth Scientists*, Kluwer Academic Publishers, Boston, 303 pp.

Pannatier, Y., 1996, *VarioWin: Software for Spatial Data Analysis in 2D*, Springer-Verlag, New York, 91 pp.

Pasion, L.R., S.D. Billings and D.W. Oldenburg, 2002, Evaluating the effects of magnetic susceptibility in UXO discrimination problems, SAGEEP, Las Vega, Nevada, Feb. 12th, 12 pp.

Rouhani, S., 1985. Variance Reduction Analysis, Water Resources Research, 21 (6) (June), pp. 837-846.

Saito, H., S.A. McKenna and P. Goovaerts, 2002, Accounting for exhaustive secondary information in geostatistical characterization of UXO sites, SERDP deliverable, December 2002.

Stein, A., 1994, The use of prior information in spatial statistics, *Geoderma*, 62, pp. 199-216.

Stein, A., M. Hoogerwerf and J. Bouma, 1988, Use of soil map delineation to improve (co)-kriging of point data on moisture deficits, *Geoderma*, 43, pp. 163-177.

Wackernagel, 1998, *Multivariate Geostatistics, 2nd Completely Revised Edition*, Springer, Berlin, 291 pp.

Tantum, S.L. and L. M. Collins, 2001, A comparison of algorithms for subsurface target detection and identification using time-domain electromagnetic induction data, IEEE Transactions on Geoscience and Remote Sensing, 39 (6), pp. 1299-1306.

Western A.W. and G. Blöschl, 1999, On the spatial scaling of soil moisture, *Journal of Hydrology*, 217 (3-4), pp. 203-224.

Wingle, W.L, E.P Poeter and S.A. McKenna, 1999, UNCERT: Geostatistics, uncertainty analysis and visualization software applied to groundwater flow and contaminant transport modeling, *Computers and Geosciences*, Vol. 25, pp. 365-376.

Xu, W., T. Tran, R.M. Srivastava and A.G. Journel, 1992, Integrating seismic data in reservoir modeling : The collocated cokriging alternative, Society of Petroleum Engineers (SPE) paper # 24742.