SAND REPORT

SAND2002-3962
Unlimited Release
Printed January, 2003

Determining Optimal Location and Numbers of Sample Transects for Characterization of UXO sites

Roger L. Bilisoly and Sean A. McKenna

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

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

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865)576-8401
Facsimile: (865)576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800)553-6847
Facsimile: (703)605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2002-3962
Unlimited Release
Printed January 2003

Determining Optimal Location and Numbers of Sample Transects for Characterization of UXO sites

Roger L. Bilisoly and Sean A. McKenna
Geohydrology Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0735

Abstract

Previous work on sample design has been focused on constructing designs for samples taken at point locations. Significantly less work has been done on sample design for data collected along transects. A review of approaches to point and transect sampling design shows that transects can be considered as a sequential set of point samples. Any two sampling designs can be compared through using each one to predict the value of the quantity being measured on a fixed reference grid. The quality of a design is quantified in two ways: computing either the sum or the product of the eigenvalues of the variance matrix of the prediction error. An important aspect of this analysis is that the reduction of the mean prediction error variance (MPEV) can be calculated for any proposed sample design, including one with straight and/or meandering transects, prior to taking those samples. This reduction in variance can be used as a "stopping rule" to determine when enough transect sampling has been completed on the site. Two approaches for the optimization of the transect locations are presented. The first minimizes the sum of the eigenvalues of the predictive error, and the second minimizes the product of these eigenvalues. Simulated annealing is used to identify transect locations that meet either of these objectives. This algorithm is applied to a hypothetical site to determine the optimal locations of two iterations of meandering transects given a previously existing straight transect. The MPEV calculation is also used on both a hypothetical site and on data collected at the Isleta Pueblo to evaluate its potential as a stopping rule. Results show that three or four rounds of systematic sampling with straight parallel transects covering 30 percent or less of the site, can reduce the initial MPEV by as much as 90 percent. The amount of reduction in MPEV can be used as a stopping rule, but the relationship between MPEV and the results of excavation versus no-further-action decisions is site specific and cannot be calculated prior to the sampling. It may be advantageous to used the reduction in MPEV as a stopping rule for

systematic sampling across the site that can then be followed by focused sampling in areas identified has having UXO during the systematic sampling. The techniques presented here provide answers to the questions of "Where to sample?" and "When to stop?" and are capable of running in near real time to support iterative site characterization campaigns.

Table of Contents

List of Figures

List of Tables

Introduction

The questions of where to locate additional samples and when to stop sampling occur whenever an investigator collects spatial data, a situation that arises in a number of disciplines including mineral and petroleum exploration, environmental remediation, epidemiology, meteorology, forestry, ecology, hydrology, agriculture, and oceanography. The largest body of work focuses on the answers to these two questions from the perspective of point sampling. Considerably less attention has been given to determining the answers to these questions when the support, or shape and size, of the sample is a linear transect. One method of determining both optimal locations and optimal numbers of samples is the data worth approach.

Point Sampling

The fundamental ideas of spatial sampling design come from statistical sampling design. Following Back (2000), there is a hierarchy of types of sampling. *Haphazard sampling* collects samples at whim, which is convenient, but this allows no estimate of the size of the sampling error, which is the difference between the estimate of some statistical parameter calculated from the sample and the corresponding true population value. This type of sampling is not recommended. *Judgment sampling* collects samples according to an expert's opinion. Again error size is not quantifiable, so it is to be avoided. *Search sampling* is used for finding particular targets, e.g., search and rescue for lost individuals, or for detecting pollution sources. Search sampling often uses prior knowledge and is designed to provide exhaustive coverage of a site. Finally, *probability sampling* relies on random selection of sample locations. Although this type of sampling takes planning, it allows for estimates of error sizes, and it is the recommended approach for making conclusions about regions that have not been sampled.

Figure 1 displays an example of each of these four sampling schemes. In each case, a field believed to have been an industrial site is sampled to see if contamination exists. The upper left design arose from a person covering the site on foot. The upper right design was based on historical reports found by a researcher. The lower left design is exhaustive and will find any circular contamination spill with the diameter larger than the spacing interval. Finally, the lower right design uses randomly selected transects, that is, samples along a line, and is an example of probability sampling. Because this type of sampling allows error estimates and is cheaper than exhaustive sampling, we will focus on it below.

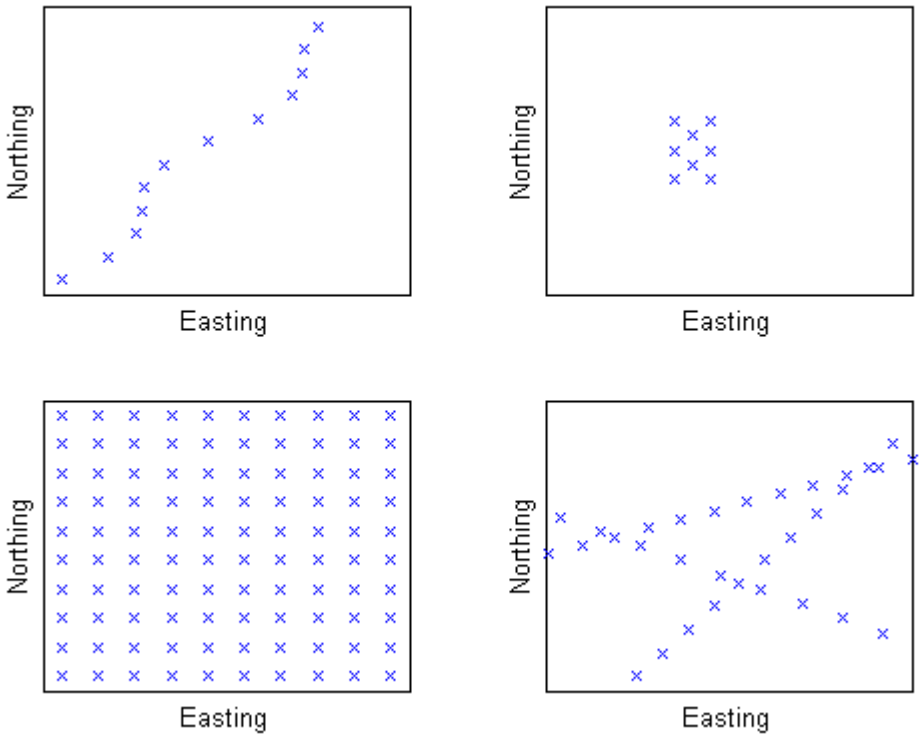


Figure 1. Examples of haphazard sampling (upper left), judgment sampling (upper right), search sampling (lower left), and probability sampling (lower right).

In probability sampling, most designs use one or more of the following four types. *Simple random sampling* uses sample locations that follow a Poisson distribution, that is, the samples are independently selected so that each position is equally likely to be picked. *Cluster sampling* uses independent groups or clusters of samples. The samples within each cluster are selected to be highly correlated in space. *Systematic sampling* uses a fixed pattern of samples, e.g., square or triangular grids. The starting point and orientation of the pattern is randomly assigned, but once this is done, locations of all the other samples are determined. Finally, *stratified sampling* breaks a region into sub-regions that are as homogenous as possible. Within these sub-regions, one of the previous three sampling techniques is used. See Figure 2 where the upper left plot has a simple random sample, the upper right plot a cluster sample, the lower left plot a systematic sample, and the lower right plot a stratified sample where samples have been located randomly within two previously selected north-south strips or strata.

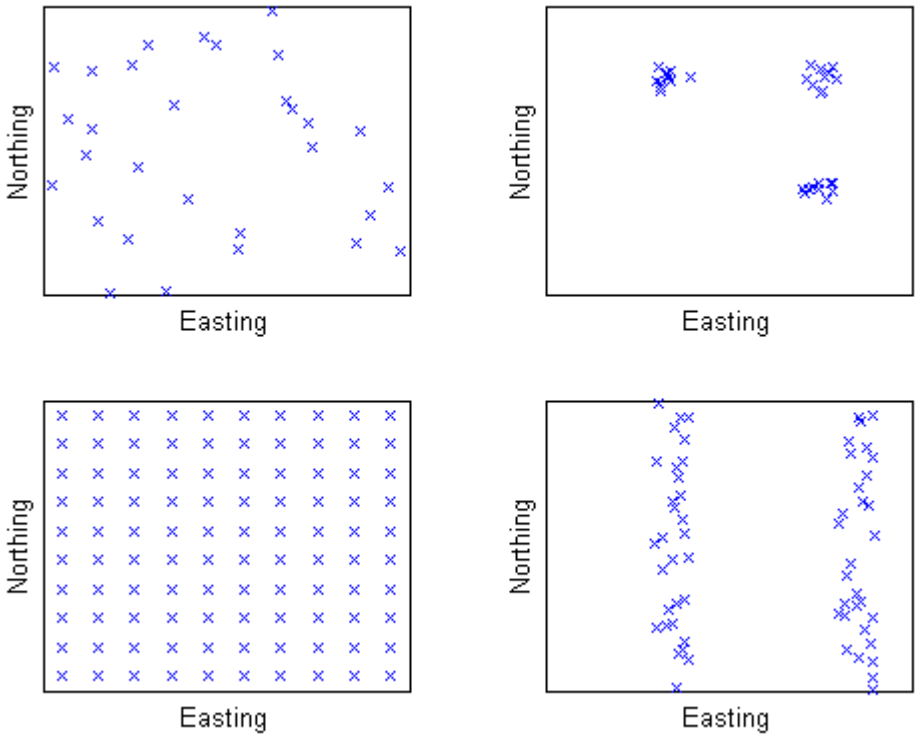


Figure 2. Examples of a simple random sample (upper left), cluster sample (upper right), systematic sample (lower left), and stratified sample (lower right).

Although there are many spatial sampling designs, e.g., Domburg et al. (1997) lists 15 main classes, most use a hierarchy of distance scales, each scale using one of the above three types of samples. These spatial scales stratify the population into subgroups, and the differing parts of the hierarchy are called stages. The choices of scales and sampling techniques are usually driven by both ease of execution and minimization of costs. We consider two examples to clarify these ideas.

For the first example, it is suspected that an old industrial site has been sampled for soil contamination and has localized spots of high contamination, but knowledge of the locations of the industrial processes are lost. Here a design that ensures a high probability of detecting all contaminated sites is needed, so we use systematic sampling in a grid. Since toxins starting in a small region tend to migrate in a plume over time, estimates of plume size need to be made from knowledge of diffusive and dispersive processes. This can be done even though the original locations are unknown. Gilbert (1987) discusses this problem and assumes that the contaminated spots are ellipses in shape. Let S be the ratio of the length of the minor axis to the length of the major axis, and let L be the length of the major axis. Let G be the size of the grid spacing. Finally, let β be the probability of not finding a contaminated site. Gilbert (1987) compares square, rectangular, and triangular grids and provides plots to determine L/G from β and S, so that grid spacing, G, can be determined from one's tolerance for missing a site and on the scale of contaminant spread.

The second example, Example 5-1 of EPA (2000), considers estimating arsenic levels in the soil within 500 m of a smoke stack. One knows that concentration in clayey soils will be more variable than in sandy soils, and downwind regions will have higher concentrations than regions perpendicular to the prevailing wind. A simple random sample would just choose points at random within 500 m of the smoke stack. However, using a stratified sampling design makes more sense here since we have identified four distinct regions with respect to arsenic level: clayey soils downwind, clayey soils perpendicular to the wind, sandy soils downwind, and sandy soils perpendicular to the wind. Using 60 samples, the simple random sample estimated the level of arsenic concentration as19.81 +/- 4.35 ppm, while the stratified random sample gave 22.94 +/- 0.48 ppm. Both methods are unbiased, but the second one is much more accurate. If accuracy of 4.35 ppm is sufficient, then one can use stratified sampling to reduce the number of samples to 8, a great savings in cost and effort. Note that stratification works here because the sub-regions differ from one another with respect to what we are measuring and accurate prior knowledge of the subregions is available.

The above discussion gives examples of spatial sampling designs that mimic traditional probabilistic sampling techniques. However, designs that are specific to the problem of spatial data have been developed. These tend to be computationally intensive, so as computing power has grown through the 1990s, so has the ability to apply these designs to real data. Two approaches will be discussed below. One generalizes the regular grid of systematic sampling to other types of space filling designs. The other involves picking locations to reduce MPEV.

To motivate the first approach, note two limitations of a regular grid. In practice all regions have boundaries, and where the grid meets the boundary irregularities occur. In addition, existing samples will generally not fit any regular grid, yet existing samples are common, and they often provide the impetus for an extensive site sampling. van Groenigen and Stein (1998) and van Groenigen et al. (2000) describe software created to tackle the problem of grid irregularities at the site boundaries. In the first paper, the Minimization of the Mean of Shortest Distances criterion (MMSD) is introduced. Define $d(x, S)$ as the distance from x to the nearest sampling point in a proposed design S. Then MMSD is satisfied when the expected value of $d(x, S)$ is a minimum. In practice this means minimizing his Equation 7 given here

$$\phi_{MMSD}(S) = \sum_{j=1}^{n_e} \frac{d(x_j, S)}{n_e}, \tag{1}$$

where the $\{x_j\}$ form a fine mesh in the region of interest and n_e is the number of locations in the fine grid. Minimization of this equation ensures that the samples are spread evenly across the site domain. van Groenigen used simulated annealing to find the optimal design S. In his second paper he generalized the MMSD to the Weighted Means of Shortest Distance criterion (WMSD), which adds a weighting function $w(x)$ to the MMSD:

$$\phi_{WMSD}(S) = \sum_{j=1}^{n_e} \frac{w(x_j) d(x_j, S)}{n_e}, \tag{2}$$

which allows the sample density to vary between different predefined strata.

He applied this to a remediation site by the River Maas. Weights were assigned to prioritize regions within the site. Regions with suspected higher concentrations were given more weight and thus more samples were located in these regions. Here previous samples existed as well as restricted regions, such as buildings, where no sampling could take place. Again simulated annealing was used to produce an optimal sampling scheme.

For the second approach, picking locations to reduce the predictive variance matrix, if a sample, *S,* taken to measure a property, *P,* in a region, *R*, is designed to give an estimate of that property throughout *R*, then this a prediction problem. If we can find the variance matrix, *C*, of the prediction of *P* in *R* from *S*, then optimal sampling can be formulated as minimizing *C* if it is a scalar, or minimizing a scalar function of *C* if it is a matrix. Note that *C* must be determined only from the locations of the sample, *S,* not from the values of *P* since we want to compare various proposed sampling locations to find the best one before we actually take any measurements. This evaluation of the sampling designs before any samples are taken is analogous to the preposterior step in the data worth approach to sample optimization (e.g., James and Gorelick (1994)).

For prediction of spatially correlated properties, kriging produces the best, linear, unbiased estimate, and the error associated with this estimate depends only on the locations of the sample, not the values of *P* at the samples, as is required for evaluating sampling designs. Minimizing the kriging variance is tested in van Groenigen et al. (1999) and van Groenigen (2000). The first paper minimizes the ordinary kriging variance:

$$\phi_{OK}(S) = \sum_{j=1}^{n_e} \frac{\sigma_{OK}^2(x_j \mid S)}{n_e}, \tag{3}$$

where $\{x_j\}$ is a fine grid of size n_e in region *R*. Since the prediction variance is calculated for one point at a time, the prediction variance is a scalar. Simulated annealing performs the minimization. The second paper also considers minimizing the maximum value of the ordinary kriging variance

$$\phi_{MAX}(S) = \max(\sigma_{OK}^2(x_j \mid S)), \tag{4}$$

and does a robustness analysis varying the variogram models used in the ordinary kriging estimate. Variograms model spatial autocorrelation and are used to determine *C*, the covariance matrix. Rouhani (1985) also considers the kriging variance in order to compute expected losses of various proposed sample locations to find the minimal expected loss.

Instead of predicting one point at a time, one can predict a vector of points all at once, such as a sampling transect, and work with the variance matrix, *C*, using the approach advocated in Fedorov and Hackl (1994). The details of this technique are given in the Approach Section below, but note that this technique still uses kriging, and that the predictive error is a covariance matrix, and so some scalar function of this matrix must be minimized. Two choices

of this scalar function are discussed in the Approach Section, each defining an optimality criterion.

Transect Sampling

Points are not the only shape possible for a sample. Point samples make sense if one has to go to a location and physically remove a homogenous sample for analysis. However, the development of remote sensing, the availability mobile sensor platforms, and ever increasing computing power and data storage allow for the collection of vast amounts of information over widespread regions. The dimensionality of the region leads to different types of sampling problems.

First, point samples are zero dimensional. Note that a point sample may be a region, but it will be homogenous with respect to what is being measured. For example, the theory of point sampling was originally developed to analyze agricultural experiments in the first quarter of the 20^{th} century. Although sections of a field would be a sample, these would be selected to be as homogenous as possible and can be treated as a point. Additionally, point samples typically only interrogate an extremely small fraction of the domain, e.g., 10^{-6} to 10^{-8} is not unusual. Since we have already discussed point sampling above, we move to the next dimension.

With the availability of off-road vehicles, airplanes, helicopters, and spacecraft, along with better remote sensing sensor technology and accurate global positioning systems, it has gotten progressively easier to collect information while the sensors are moving. With the ongoing development of cheaper, larger data storage capabilities, collecting large amounts of data is now routine. Together these trends have led to collecting data as the sampling instrument moves which produces one-dimensional samples called transects. In practice, data are collected at some temporal or spatial frequency, which makes any transect a collection of point samples. However, spatial autocorrelation makes the measurements at these locations dependent, which violates the assumptions of point sampling. Straightforward application of the point sampling techniques, such as gridded samples, will lead to inefficient sample designs and so a new sampling theory is required.

One can also take two and three-dimensional samples. As an example of the former, Landsat 7 has a swath width of 185 kilometers and pixel resolution of 30 m, and most properties are heterogeneous at this scale making the image two-dimensional. For a three-dimensional example, an archeological excavation over an entire site is heterogeneous in depth, deeper is older, and also in length and breadth. However, the theories appropriate for these two cases do not apply to transect sampling, and will not be discussed further here.

Finally, stereology will not be discussed here. It is the study of how lower dimensional samples can be used to estimate higher dimensional properties, e.g., using transects to estimate areal quantities. However, estimating the variability of these estimates is difficult and poorly developed. Without estimates of predictive error, it is not possible to find optimal designs, and so stereological techniques are not helpful for making transect sampling designs.

Theory of Transect Sampling

There are two main thrusts in taking transect samples. First, there are exhaustive search problems. For example, the Coast Guard uses transect searches to find people missing at sea. As a ship performs the search, common sense dictates that observers stay constantly on the lookout as the ship travels along its entire search path, and not just look for the missing person at specific locations along the ship's trajectory. In the second thrust, which is more closely related to point sampling, a limited number of transects in a region are used to estimate properties of the entire region. For instance, a helicopter with a magnetometer makes a north-south and an east-west pass over a field to estimate the total number of magnetic anomalies in the field. Although two passes are a small subset of the field, extrapolation to the entire field is possible.

Although searches are commonly done, e.g., mine sweeping or in a variety of search and rescue missions, there is a high premium on finding the target of the search as quickly as possible, and so only a few proven paths are used in practice. See Figure 3 for two examples used by the Coast Guard, taken from NSARC (2000). For both patterns, the width of the white swath equals the width of the ship. Observers on the ship can see halfway into the black region, so at the end of the search pattern, the entire square has been visually scanned. The first pattern is called the square single-unit pattern and spirals outwards from the last known position of the object that is the target of the search. The second is called the parallel track single-unit pattern and is used when the target could be anywhere within a specific region. Because finding the target is paramount, the designs search exhaustively. However, this paper considers sampling a subset of a region to make predictions outside the sample, and because exhaustive searching defeats this goal, we do not further consider exhaustive search patterns.

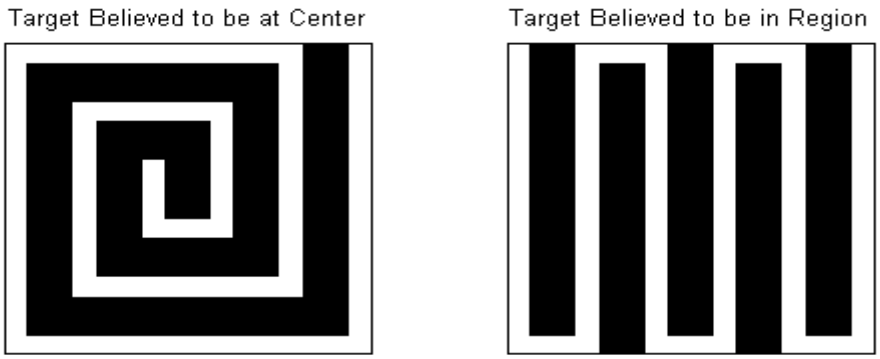


Figure 3. Two search and rescue paths used by the Coast Guard. White denotes the path which is as wide as the ship. Observers on the ship can see half way into the black region. The first is used when there is a last known position, the second when the region is believed to contain the target. From NSARC (2000).

As mentioned above, a transect sample consists of point samples along a path. If the samples were mutually independent, then one could use either the theories of systematic sampling or simple random sampling for point samples as discussed above. However, UXO data exhibit spatial correlation, so a more general sampling theory is needed. If one looks at practical recommendations from Burnham et al. (1980), transects are often deployed in one of several possible ways, the most basic being a linear transect in a random orientation. If several are deployed, then parallel segments filling a region, parallel segments deployed at random, or random segments are common; see Figure 4 which are modeled on figures from Burnham et al. (1980). Note that these transect patterns are suggested for biologists in the field, and that straight transects are easy to walk or drive or fly: just pick a feature in the distance and go toward it. Note that if one came to an obstacle like a cliff, the design would be modified in the field and the transect would end prematurely or be diverted. Similarly, a long transect of San Francisco Bay taken in a boat would bend so to stay in the water, e.g., see Jassby et al. (1997).

Like point sampling, large regions are easier to sample in stages, so in practice, a large region would be divided into subregions that in turn would be sampled by transects. The transect designs discussed above must consider orientation, but are natural extensions of systematic and simple random point sampling. Moreover, spatial sampling can be done with point samples without using transects. So for these two reasons, transect sampling has been dealt with less than point sampling in the literature.

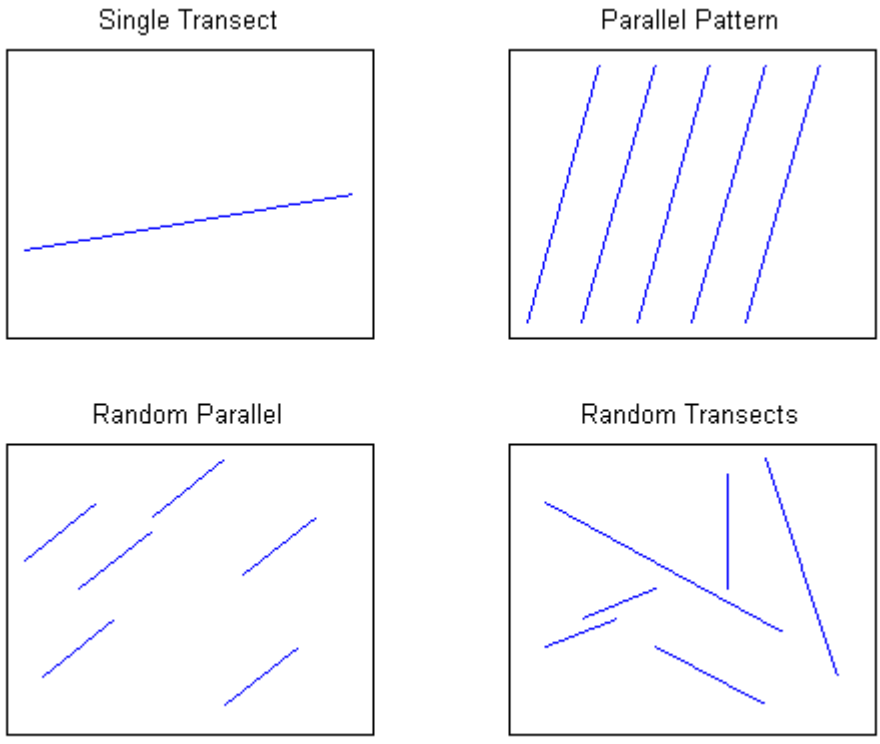


Figure 4. Typical transect sampling designs following Burnham et al. (1980).

With estimates of spatial autocorrelation, one can move beyond the simpler transect sampling models described above. In point sampling, one can go from systematic sampling to optimal designs by using predictive error covariances, and the same idea can be applied to transect sampling by adapting the techniques in Fedorov and Hackl (1994). All that must be added to their technique is optimization under constraints, which are used to enforce either linearity or bounded curvature. This allows us to find optimal designs for data collection while moving, e.g., measuring magnetic anomalies from a helicopter carrying a magnetometer. As noted above, the coupling of mobile instrument platforms with remote sensing has produced many opportunities to apply optimal transects in practice.

Prior to developing the techniques used to optimize transect sampling design, a detail regarding the application of the geostatistical estimation technique kriging to data collected along transects is presented. Using data values collected on a transect with kriging to make predictions at a point off of the transect can give unintuitive weightings as pointed out in Deutsch (1994). Performing kriging using a random function model implicitly assumes that the study region is infinite which has the effect of treating interior transect sites as clustered and thereby less informative than the sites at the edges of the transect. Several corrections for this problem have been suggested, but Deutsch (1994) provides a solution, finite domain kriging (FDK), that preserves positive definiteness. Figure 5 below demonstrates this effect in two examples. The top two plots compare ordinary kriging with FDK when the prediction site is beyond the range of the variogram model, in this example a spherical model. Since all the transect sites are farther than the range, they should all contribute equally, which is true in FDK. The bottom two plots show the same situation but the prediction site is now half the range of the spherical variogram. Here one expects the closest sites to be most important in prediction, but again this is only true in FDK.

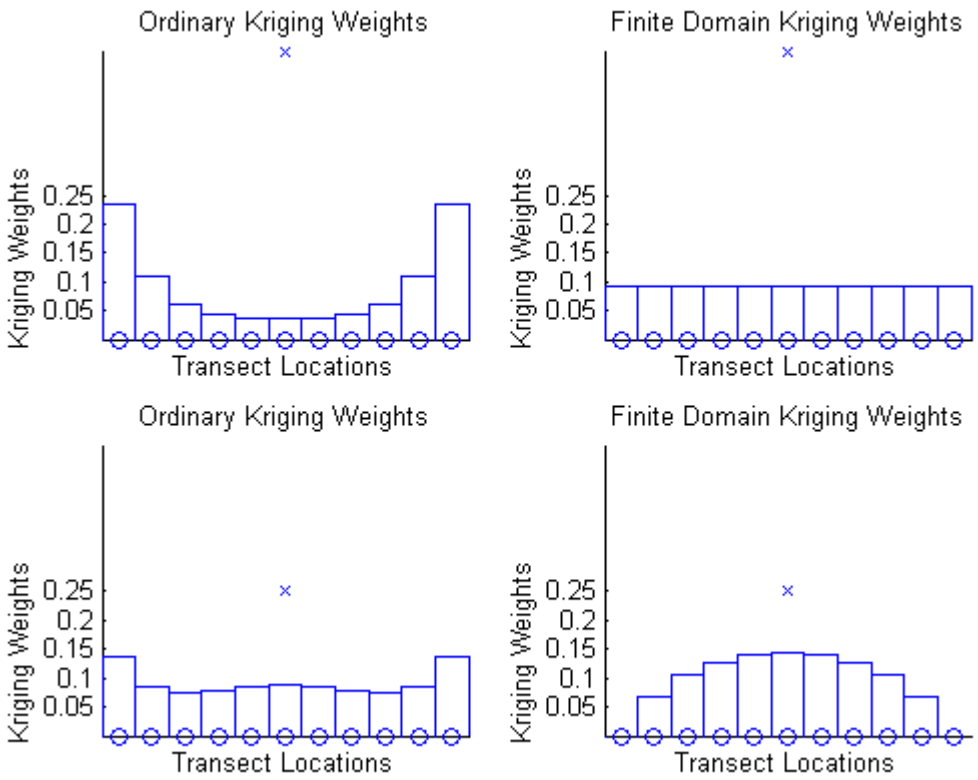


Figure 5. X marks the site being estimated, and O marks transect sample sites. The range of the spherical variogram equals the height of the plot, and a 20 percent nugget effect is used. This figure is based on Figure 3 of Deutsch (1994).

Data Worth Approaches

Taking additional samples always adds information, e.g., even taking repeated samples at the same location will provide information on measurement error and detector calibration drift. However, there is always some cost of gathering samples, if nothing else, it takes time to collect samples. Hence, one must cease sampling at some point. In general, stopping rules are based either on reduction of an economic objective function such that the number of samples that provide the lowest total project cost can be determined, or on the reduction of a measure of uncertainty or prediction variance to a prescribed level. First we consider the economic approach. We primarily follow Freeze et al. (1992), a collection of four papers republished by the National Ground Water Association. This collection of papers stimulated more papers exploring additional applications, e.g., James and Gorelick (1994), James et al. (1996), Zawadzki and Beckie (1996), and Freeze and Gorelick (1999).

The economic approach to determining when to stop sampling has two parts. First, we convert all costs, risks and benefits that occur in a typical characterization and remediation project into dollar amounts, which may not be easy. Consider the example of monitoring a landfill for possible contaminant plumes arising from leaks in the landfill liner (adapted from Freeze et al.

(1992)). Costs consist of the price of buying, placing and maintaining the monitoring sensors. The data collected are required and have no market value, but there are possible benefits, e.g., protection from frivolous lawsuits by satisfying regulatory requirements. Although one might be conservative and not assign a dollar value to this because it is hard to estimate, it would be a mistake to do the same with risks. A risk is the expected cost of a failure, that is, the probability of a failure times the cost of that failure occurring. In this example, a failure is a leak, and although the cost of fixing a leak that is detected is not hard to characterize, what is the cost of losing community goodwill? Protestors may be able to persuade politicians to shut down the facility far longer than required to fix a leak. Negative publicity may lead to lawsuits claiming damages for health problems, and estimating what damages might be awarded in court for, say, a case of cancer, is difficult. Even if one was willing to estimate the cost of settling cancer lawsuits, the mere existence of such a cost in a business plan may cause ill will in the community, and entail further costs. For the moment, however, ignore these complications and assume that dollar figures are computable and define the following objective function,

$$\varphi = \text{Benefits} - \text{Costs} - \text{Risks (in dollars)}. \qquad (5)$$

The second part of the economic approach splits the sampling into three stages. Note that since the *Costs* and *Benefits* are fixed, only the *Risks* term can be modified, so we focus solely on sample designs that minimize risk. First, one conducts a prior analysis of φ using all existing data and information. Second, for any proposed sampling scheme, one conducts a preposterior analysis. This analysis focuses on determining sampling locations that minimize the maximum risk prior to taking those samples. Returning to the example above, the better the monitoring network, the sooner a leak is detected, the less damage it can cause, and so the risk is lower. Hence minimizing measurement error also minimizes the *Risks* term, and thereby φ. The increased number of samples decreases the risk, but also increases the cost of monitoring. Finding the number of samples that provides the minimum total cost, φ, is the object of the economic approach to determining the stopping rule. Finally, after the data are collected, one can perform a posterior analysis. Note that evaluation of the sampling designs need only use the preposterior analysis.

The only troubling aspect of the economic approach is converting all terms into dollar amounts. This is especially difficult in sampling design for UXO sites where the costs of sampling design failure can result in death and/or bodily harm. In Freeze et al. (1992), decision trees are used and optimization methods only mentioned. With the greater computing power of today, optimization has the advantage in sensor deployment, e.g., Freeze and Gorelick (1999), p 950, advise using optimization if (1) optimal placement is important and (2) risks are hard to quantify. So in the Approach Section, we ignore computing dollar costs and decision trees, and concentrate on variance and optimization. That is, instead of minimizing φ, we concentrate on minimizing the variability of the predictive error. Note that our approach requires a decision maker to compare the cost of an additional sample with the benefit of its variance reduction, whereas the economic approach requires a decision maker to determine the change in the dollar amount of all risks. As seen above, the latter is complicated by hard to quantify social, political and legal issues.

Approach

Sampling of a site will be done in stages. Suppose that at the end of a stage there are n_s sites with measurements already taken. The discussion below specifies how to compare different samples consisting of n_p new sites where measurements have not yet been taken. Although details need to be explained, the basic idea is simple: comparisons are made based on the error arising from predicting the results of measurements on a rectangular grid of n_g sites using the $(n_s + n_p)$ current sites. Although estimating the n_g measurements on the grid requires knowing all the $(n_s + n_p)$ *values*, n_p of which are unknown, estimating the error only requires the *locations* of the $(n_s + n_p)$ sites, all of which are known. The sample of size n_p that has the least error is the best. Once the best sample is known, one can decide if it is worth carrying out, that is, if the information it provides outweighs the effort and expense needed to obtain the sample.

Optimizing Additional Sample Locations

Let R stand for the region of interest with respect to a spatial variable y. For example, let y be the signal strength that an electromagnetic sensor would read if flown over a field. We first consider the following problem: given that y has been measured at n_s locations in R, and given that we would like to take a fixed number, n_p, of additional measurements, where should the new measurements be taken?

The key idea to solve the above problem comes from the solution for the following simpler problem. What is the optimal spatial sample of size n_p for region R, given that no samples have been taken yet? Fedorov and Hackl (1994) solve this problem by minimizing the predictive error of unsampled sites given the sample of size n_p. We consider the case of the unsampled sites forming a rectangular grid of size n_g. Define $\mathbf{y}_p$ as the vector of measured values at the proposed locations (subscript p for *proposed*). Define $\mathbf{y}_g$ as the vector of values to be estimated in the rest of region R (subscript g for *grid*). Simple kriging gives the best linear, unbiased estimate

$$\hat{\mathbf{y}}_g = C_{gp} C_{pp}^{-1} \mathbf{y}_p , \tag{6}$$

where C_{pp} is the correlation matrix of $\mathbf{y}_p$ with itself, C_{gp} is the correlation matrix of $\mathbf{y}_g$ with $\mathbf{y}_p$, and the hat over $\mathbf{y}_g$ stands for the predicted value of $\mathbf{y}_g$. These correlation matrices require a correlation function between spatial locations, which is often modeled by a variogram in geostatistics. Since predicted error is the predicted value minus the true value, the variance matrix of the predicted error is

$$Var(\hat{\mathbf{y}}_g - \mathbf{y}_g) \equiv Var(C_{gp} C_{pp}^{-1} \mathbf{y}_p - \mathbf{y}_g) = C_{gg} - C_{gp} C_{pp}^{-1} C_{pg} . \tag{7}$$

We want to make this error as small as possible, but since the error is a g by g matrix, some scalar function of it must be taken first. In the section *Geometric Interpretation* below, the link between a matrix and its eigenvalues is explained. Since eigenvalues are scalars, it is natural to restrict consideration of scalar functions of a matrix to functions of the matrix's eigenvalues. In

practice, the product of the eigenvalues is most used (Myers and Montgomery (1995), p 364), and this choice corresponds to the D-optimality criteria. Recall that the product of the eigenvalues of a matrix equals the determinant of the matrix.

For the spatial design problem at hand, Fedorov and Hackl (1994) employ the D-optimality criterion. Hence the solution entails finding $\boldsymbol{y}_s$ so that

$$\min_{\mathbf{y}_s} | C_{gg} - C_{gp} C_{pp}^{-1} C_{pg} | \tag{8}$$

occurs. They also introduce a computational simplification. Let C be the correlation matrix of $\boldsymbol{y}$ with $\boldsymbol{y}$, where $\boldsymbol{y}$ is the vector of the sites $\boldsymbol{y_p}$ combined with $\boldsymbol{y_g}$. C can be represented by the block matrix

$$C = \begin{pmatrix} C_{pp} & C_{pg} \\ C_{gp} & C_{gg} \end{pmatrix}. \tag{9}$$

But we know from linear algebra that

$$| C | = | C_{pp} | \, | C_{gg} - C_{gp} C_{pp}^{-1} C_{pg} | \quad . \tag{10}$$

Since C is fixed, the two determinants on the right hand side are inversely proportional, so the D-optimality condition is equivalent to

$$\max_{\mathbf{y}_s} | C_{pp} |, \tag{11}$$

which is computationally more tractable if the sample size, n_p, is relatively small.

Note that the same criterion arises if one wants to minimize the matrix of predictive variance instead of the variance matrix of predictive error. Let μ_g be the expected value of $\boldsymbol{y}_g$, and μ_p be the expected value of $\boldsymbol{y}_p$, then the mean and variance of $\boldsymbol{y}_g$ given $\boldsymbol{y}_p$ are

$$\begin{aligned} E(y_g \mid y_p) &= \mu_g + C_{gp} C_{pp}^{-1} (y_p - \mu_p), \\ Var(y_g \mid y_p) &= C_{gg} - C_{gp} C_{pp}^{-1} C_{pg}. \end{aligned} \tag{12}$$

This conditional mean gives the best linear estimate of the unsampled values, and the conditional variance gives the variability of this estimate.

The above result can be extended to two-stage sampling. Let $\boldsymbol{y_p}$ be split into two vectors $\boldsymbol{y_o}$ and $\boldsymbol{y_n}$, where $\boldsymbol{y_o}$ are the values at the old locations of the first stage (subscript *o* for *old)*, and $\boldsymbol{y_n}$ are the values at the new locations of the second stage (subscript *n* for *new)*. Now C_{pp} can be split into the block matrix

$$C_{pp} = \begin{pmatrix} C_{oo} & C_{on} \\ C_{no} & C_{nn} \end{pmatrix}. \tag{13}$$

Using the same determinant identity above, we see that

$$| C_{pp} | \equiv | C_{oo} | | C_{nn} - C_{no} C_{oo}^{-1} C_{on} | . \tag{14}$$

Since C_{oo} is fixed, we conclude that

$$\max_{y_s} | C_{pp} | \equiv \max_{y_n} | C_{nn} - C_{no} C_{oo}^{-1} C_{on} | , \tag{15}$$

that is, in order to minimize the predictive variance matrix, the best new sample to add to the existing sample must maximize the above determinant.

Geometric Interpretation

Linear algebra proves that a determinant of a matrix is the product of the eigenvalues of that matrix; see section 1.5 of Strang (1986). Recall that the above discussion proves that finding the maximum of $|C_{pp}|$ also minimizes the determinant of the matrix

$$C_{gg} - C_{gp} C_{pp}^{-1} C_{pg} , \tag{16}$$

and so minimizes the product of the eigenvalues of this matrix; call them $\{\lambda_i\}$. If we are willing to assume that the predictive distribution of $\boldsymbol{y_g}$ given $\boldsymbol{y_p}$ is Gaussian, then the level curves of this distribution are hyperellipses (n-dimensional generalizations of two dimensional ellipses) with semi-major axes proportional to these eigenvalues λ_i; see Press (1972), section 3.3.2 for further discussion. However, these level curves are also confidence regions, that is, error regions for the predictive distribution. Moreover, the product of the λ_i is not just the determinant, but also proportional to the volume of this hyperellipse. Consequently, the condition of finding the minimum of the above determinant also minimizes the volume of the error region of the prediction distribution of $\boldsymbol{y_g}$ given $\boldsymbol{y_p}$. See Figure 6 below for a two dimensional example. The area of the ellipse is $\pi\lambda_1\lambda_2$, and the determinant equals $\lambda_1\lambda_2$, so minimizing the area of the ellipse also minimizes the predictive error.

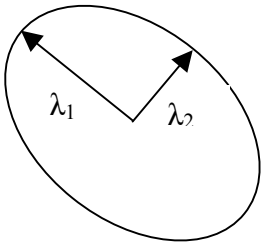


Figure 6. Level curve for a two-dimensional covariance matrix. Minimizing the product of the eigenvectors, $\lambda_1\lambda_2$, minimizes both the area of the ellipse and the predictive error of y_p given y_s.

Finding the Best New Locations

The above discussion proves that the following are equivalent

$$\min_{y_s} | C_{gg} - C_{gp}C_{pp}^{-1}C_{pg} | \equiv \max_{y_s} | C_{pp} | \equiv \max_{y_n} | C_{nn} - C_{no}C_{oo}^{-1}C_{on} | . \qquad (17)$$

Hence we can focus on the last condition if we have an existing sample of size n_o and want to compare differing arrangements of new samples to see which is best. Let S_1 be a proposed sample of size n, and let S_2 also be another proposed sample of size n. S_1 and S_2 are competing new sample designs. Once a correlation function is determined to model the spatial autocorrelation in the region R, C_{nn1}, C_{no1}, C_{on1} and C_{oo1} can be calculated for the sites in S_1. Calculate from these the determinant specified above to get D_1, that is,

$$D_1 \equiv \max_{y_n} | C_{nn1} - C_{no1}C_{oo1}^{-1}C_{on1} | . \qquad (18)$$

The analogous calculations can be done for S_2 to obtain D_2. Since S_1 and S_2 both have n sites, D_1/D_2 will be the ratio of predicted error regions. For example, if $D_1/D_2 = 0.8$, D_1 is 20 percent better than D_2. To remove the effect of proposed sample size we can consider

$$\left(\frac{D_1}{D_2}\right)^{1/n} = \text{Relative worth of } S_1 \text{ to } S_2 \text{ per site} . \qquad (19)$$

Note that the numerator and denominator above are geometric means of the eigenvalues that are associated with D_1 and D_2, respectively. Because of the geometric interpretation above, we have easily interpreted measures of relative worth of two proposed samples. One caution must be observed: this method is not robust if any of the eigenvalues are either zero or close to zero which can happen if any of the predicted sites are highly correlated to one of the sample sites.

Now define F to be the function that maps a proposed new sample S to the determinant D, e.g., $F(S_1) = D_1$ and $F(S_2) = D_2$. See Figure 7 below for several examples of this mapping where an

isotropic spherical variogram with range of 5 units is used. As the spacing of the proposed sample sites (denoted with *x's* exceeds the range of the variogram, *F* increases to 1, so the lower right configuration is the best among the four examples. This makes intuitive sense, as the samples are located at the corners of the domain thus providing the maximum possible sample spacing, or greatest coverage of the domain. Since *F* maps a proposed new sample to a number such that bigger is better, we can make *F* a fitness function for simulated annealing, an optimization technique.

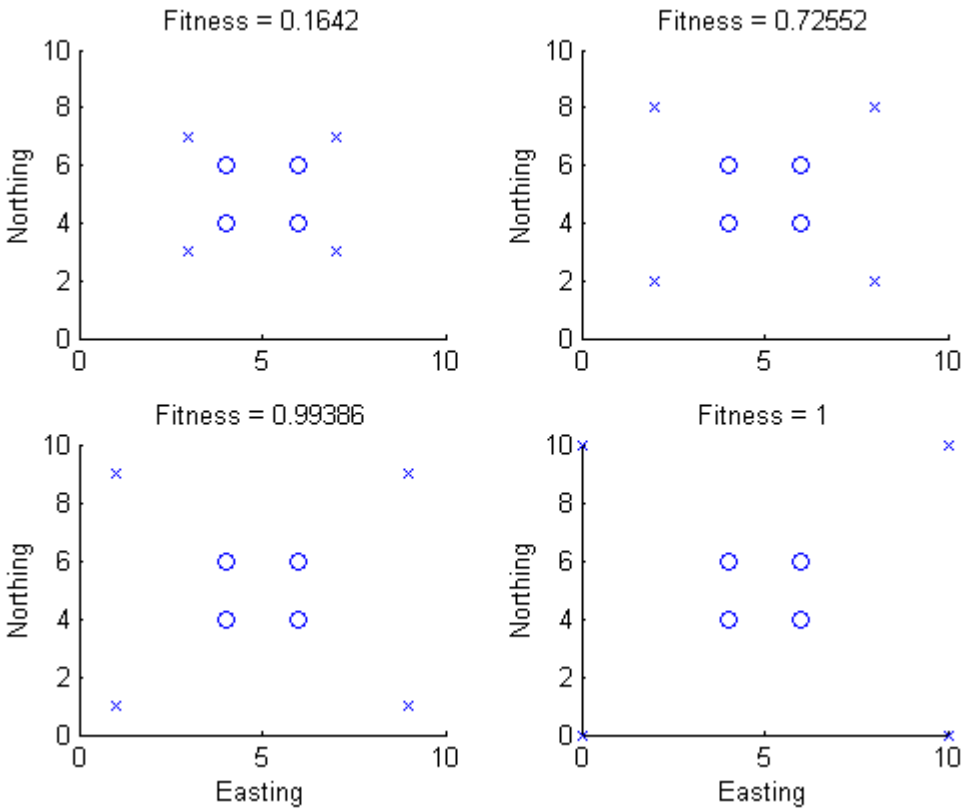


Figure 7. O's represent the existing sample, X's represent a new, proposed sample. An isotropic, spherical variogram is used with a range of five.

Simulated annealing mimics the process of cooling a molten crystal (see Section 5.1 of Michalewicz and Fogel (2000)). If the cooling schedule is slow enough, the result will be a well-formed crystal that has achieved its minimum energy configuration. If cooling is too fast, then the process is likely to end up in a local minimum that is not a global minimum. If cooling is too slow, then the process is inefficient. In practice erring toward the latter is more desirable than erring toward the former since taking extra time computing a good answer is better than saving time computing an inferior local minimum. Figure 8 shows pseudocode for simulated annealing for optimizing a sample. The algorithm takes a solution *S* and perturbs it to S^* try to find a better solution. However, to avoid being stuck in a local minimum, it will sometimes take solutions that are less fit. The probability of accepting inferior solutions decreases to 0 as the temperature parameter *T* decreases to 0.

```
Randomly initialize the sample S
Compute F(S), the fitness of S
While T > T_min
     Perturb S to S*
     If F(S*) > F(S)
          Replace S by S*
     Else If random(0,1) < exp((F(S*) - F(S))/T)
          Replace S by S*
     End If
     Decrease T
End While
```

Figure 8. Pseudocode for simulated annealing applied to picking an optimal additional sample. S is the proposed new sample, T is the temperature, and random(0,1) produces uniformly distributed random numbers between 0 and 1.

When to Stop Sampling

The above discussion offers a mechanism for placing new samples, but does not address the question of whether or not more samples should be taken. That is, should we stop sampling because sufficient information is already known? In theory, adding samples will always add some information, but is it adding enough information to justify the effort or expense of taking the sample? Taking an exhaustive survey is time consuming and costly, so it is worth considering if we can stop prior to that.

The above discussion, however, gives us the key tool to decide when to stop sampling: the predictive covariance matrix. We consider two different ideas, and for both we use the following set-up. Impose on R a fixed grid of points G. We will choose samples anywhere in R, in particular, not restricted to the grid G. These samples will be used to find the predictive covariance matrix for the points in G. Let S_o be the old, existing sample, let S_n be the new, proposed sample, and let S_b be the union of S_o and S_n (subscript b for *both)*. Define $\boldsymbol{y_o}$ as the vector of measured values for S_o, and similarly for $\boldsymbol{y_b}$ and $\boldsymbol{y_n}$. S_b has to do better than S_o, but we need to quantify the superiority and judge whether it is large enough to justify taking S_n. Finally, we know from the above discussion that the predictive covariance matrices of $\boldsymbol{y_g}$ given $\boldsymbol{y_o}$ and $\boldsymbol{y_g}$ given $\boldsymbol{y_b}$ are

$$\begin{aligned} Var(y_g \mid y_o) &= C_{gg} - C_{go}C_{oo}^{-1}C_{og} \\ Var(y_g \mid y_b) &= C_{gg} - C_{gb}C_{bb}^{-1}C_{bg} \end{aligned}. \qquad (20)$$

Recalling that these are measures of error of prediction we can now specify the two methods of determining when to stop sampling.

First, since the determinant of these covariance matrices is proportional to the volume of the error region for a Gaussian distribution, the following ratio compares the two

$$\frac{|\, C_{gg} - C_{gb} C_{bb}^{-1} C_{bg} \,|}{|\, C_{gg} - C_{go} C_{oo}^{-1} C_{og} \,|} < 1 - r \,. \qquad (21)$$

The constant r is the factor of improvement deemed necessary to go ahead with the sample. For example, $r = 0.20$ means that we require a 20 percent reduction of predictive error in order to proceed with the taking the sample S_n. The improvement in this ratio will be large for the initial samples, but will begin to decrease with iterative application and the addition of more samples until eventually the reduction in the predictive error goes below r. At this point we are done sampling. Finally, note that since the determinant is used, this criterion is built on D-optimality.

Second, the diagonal of a covariance matrix contains the variances for each of the locations. So instead of reducing the two matrices above to a number, we could plot the diagonals of each so that the variance of each point of G is plotted at its location. For example, let G be a 4 by 3 grid with x and y coordinates (1, 1), (2, 1), (3, 1), (4, 1), (1, 2), (2, 2), …, (4, 3) as in the grid shown in Figure 9.

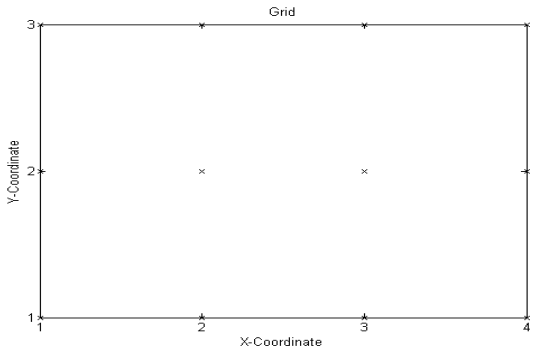


Figure 9. Example of a 4 by 3 reference grid.

Then $C_{gg} - C_{go}C_{oo}^{-1}C_{og}$ will be a 12 by 12 matrix, and its *(i,i)* entry will be the predictive error variance of the *i*th site in G. For example, the (2, 2) entry will be the variance of the location at coordinates (2, 1), and the (12, 12) entry will be the variance of the location at coordinates (4, 3). Moreover, the sum of all 12 variances will be the trace of $C_{gg} - C_{go}C_{oo}^{-1}C_{og}$, which is easy to compute and one way to summarize the variances over the grid G. Note that the MPEV is the trace divided by 12. In this second approach, the trace of $C_{gg} - C_{go}C_{oo}^{-1}C_{og}$ is minimized with the addition of new samples until it is reduced to a value below a prescribed threshold d. Using the trace instead of the determinant of the error matrix links this second method to A-optimality.

Comparison of D-optimality and A-optimality

In Figure 10 we compare the fitness of two transects parallel to the x-axis with y-values of 4 and 8 with the fitness of two centered orthogonal transects. Both pairs of transects have no

prior samples to build on, that is, both are the first stage of sampling. G consists of the grid of all points with x-coordinates and y-coordinates in the range {1.5, 2.5, …, 10.5}. Let o stand for *orthogonal transects* and p stand for *parallel transects.* Then we have:

$$C_{gg} - C_{go}C_{oo}^{-1}C_{og} = \text{Predictive Error of Orthogonal Transects on Grid, and}$$
$$C_{gg} - C_{gp}C_{pp}^{-1}C_{pg} = \text{Predictive Error of Parallel Transects on Grid.}$$

Since these are predictive errors, smaller is better, and so if the ratio is greater than one, it is the denominator that is better. In the two ratios below, recall that the denominator comes from the parallel transects. Spatial correlation is modeled by a spherical variogram with range of 3 units. Below we see that both D-optimality (which uses the Det( ) of the predictive covariance matrix) and A-optimality (which uses the Trace( ) of the predictive covariance matrix) agree that the parallel transects are better. The parallel transects do better because the orthogonal transects have highly correlated samples near the crossing point. However, the measure of improvement is quite different: 88 percent vs. 1.6 percent between the two different calculations.

$$\frac{Det(C_{gg} - C_{go}C_{oo}^{-1}C_{og})}{Det(C_{gg} - C_{gp}C_{pp}^{-1}C_{pg})} = 1.8813$$
$$\frac{Trace(C_{gg} - C_{go}C_{oo}^{-1}C_{og})}{Trace(C_{gg} - C_{gp}C_{pp}^{-1}C_{pg})} = 1.0160 \qquad (22)$$

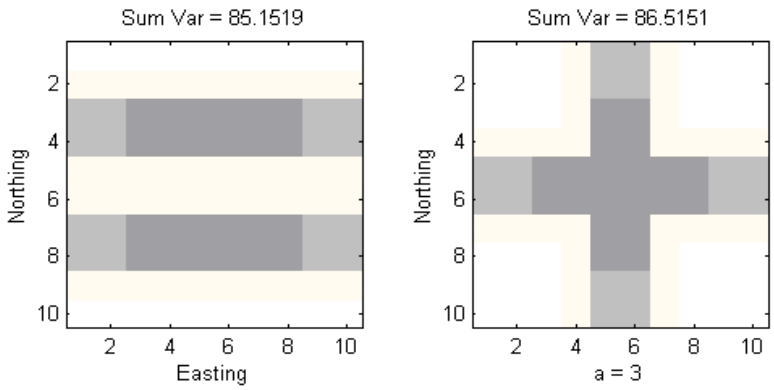


Figure 10. Comparing two parallel to two orthogonal transects by the D-optimality and A-optimality criteria. Maps show variance from two, one-dimensional sample transects.

To summarize, we have developed a technique for determining the optimal locations of future transect samples by minimizing the predictive error across the domain after those samples are obtained. This minimization of the prediction error can be coded as an objective in a simulated annealing algorithm and used to determine these optimal locations. Two techniques have been

presented to determine the amount of variance reduction that will occur due to any proposed design. These two approaches are based on D-optimality and A-optimality criteria and a comparison of these two techniques shows that the D-optimality criterion is more difficult to compute but provides a more sensitive measure of predictive error. D-optimality is used in only the first example below because it uses a small (11x11) reference grid. Since achieving A-optimality is equivalent to minimizing MPEV, any time we consider a variance map, we are implicitly considering A-optimality, which is done for all three examples below.

Example Applications

Three example applications are discussed. The first two consider spatial transect sampling design using hypothetical data, and the third shows a reliability analysis using magnetometer data collected by Oak Ridge National Laboratories (ORNL) at the Pueblo of Isleta in New Mexico as ground truth.

Optimal Transect Locations

The first example illustrates the results of using simulated annealing (see Figure 8) to determine the best transect given all the existing samples, transects or not. Simulated annealing is used to minimize the determinant of the predictive error matrix, which is equivalent to maximizing Equation (11) as discussed above. Note that repeating the process of adding transects multiple times need not be optimal, e.g., let T_1, T_2 be two transects. Finding the optimal way to add both T_1 and T_2 at once need not be the same as finding the optimal T_1, and then finding the optimal T_2 given that T_1 has already been collected. However, this simulated annealing approach is ideal if there are already existing samples, or if one wants to decide whether or not to collect the next sample.

The hypothetical site is 1100x1100 meters and Figure 11 has four plots of the site where red denotes the transect, which can meander, and the background color scheme shows predictive variance as a function of location. The mean of these predictive error variances is the MPEV. White denotes the maximum possible variance of 1.0, and black denotes zero variance, the lowest possible value. The upper left plot has no samples, so the variance map shows that every site has the maximum variance and the mean variance is 1. The upper right plot shows an initial transect, one that might be done as part of an initial survey of the site to determine if the region shows evidence of requiring further, more detailed sampling. In practice, this initial transect also allows an initial variogram model to be determined that is used in the simulated annealing code to optimize future transect locations. However, in this hypothetical example, a fixed isotropic, spherical variogram with a range of 300 meters and a nugget value of zero was used for all stages of sampling.

The lower left and lower right plots of Figure 11 show the results of using simulated annealing to determine the position of the added transects. In both cases the new transect was constrained to start at the bottom of the region, the length of both transects was fixed, and the maximum change of direction of the transect was constrained so that sharp turns were penalized in the fitness function of the simulated annealing program. The lower left plot shows the optimal meandering transect of length 3500 meters given the original straight transect of

length 1100 meters. The lower right plot shows the optimal meandering transect of length 2000 meters given both the original straight transect and the meandering transect of length 3500 meters.

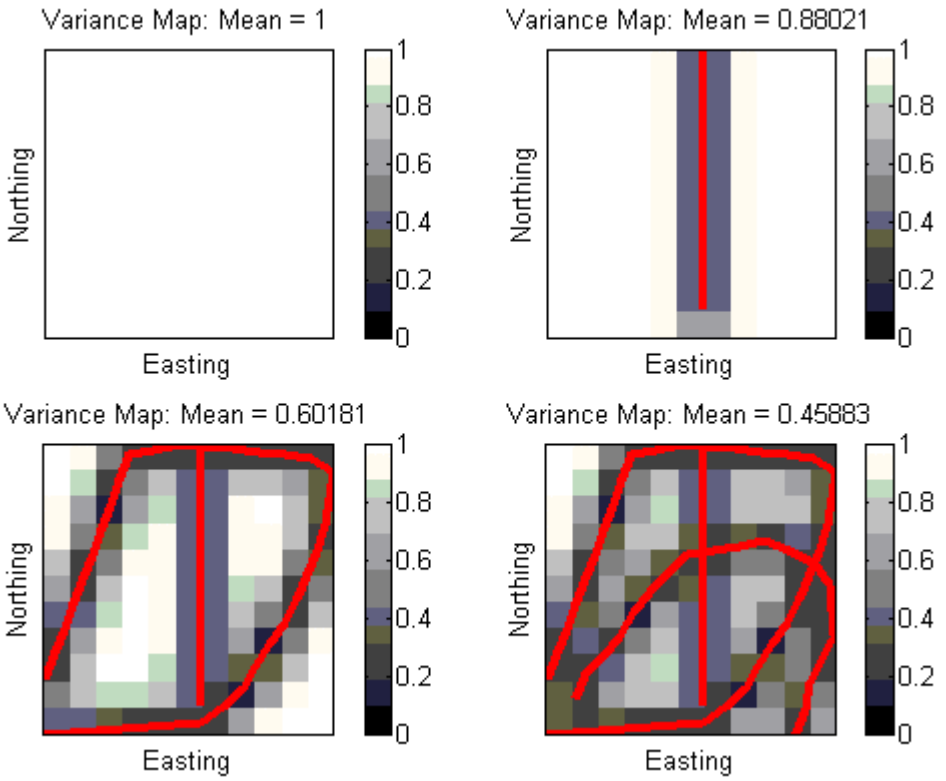


Figure 11. Four stages of meandering transect sampling where 1 is the maximum possible variance. The upper left plot has no samples and no variance reduction. The upper right plot has the initial transect. The lower left plot shows the best meandering transect given the first transect. The lower right plot shows the best meandering transect given the first two transects. An isotropic spherical variogram with range of three is used.

For this example, a fixed cost of sampling is assumed to be $300/acre. Assuming a sensor footprint width of 8 meters and converting all units to metric, results in a sampling cost of $0.54/linear meter of transect. This cost assumption may not fit for all sites, but it does allow for calculation of the cost of reducing the predictive variance to certain levels as more transect samples are taken in this example. Figure 12 shows the cost of obtaining the transects and the reduction in the predictive variance as a function of total transect length.

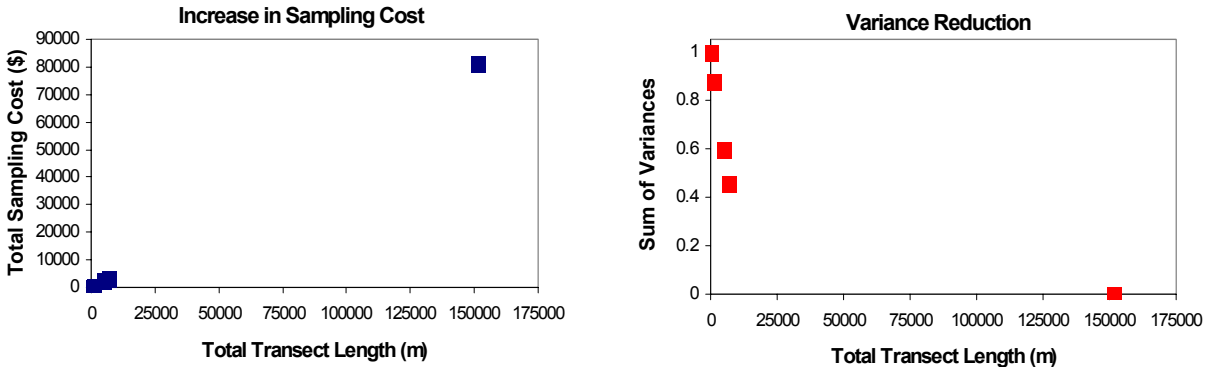


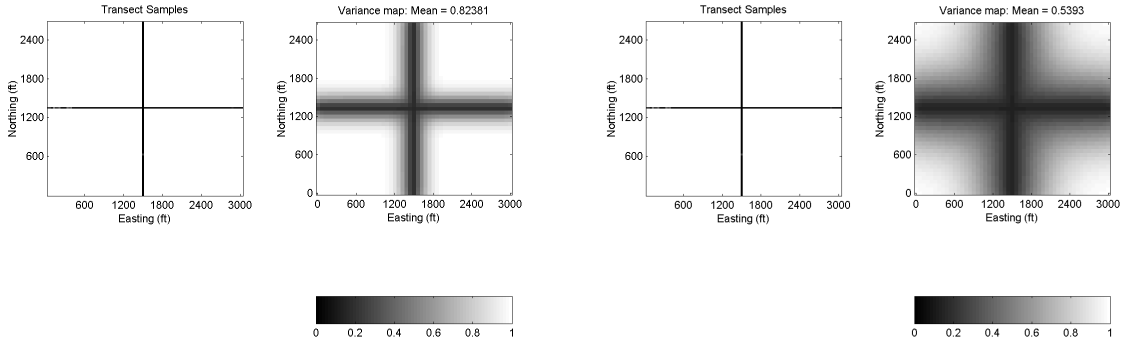
Figure 12. The cost of obtaining the transect samples as a function of sample length (left image) and the reduction in predictive variance as a function of sample length (right image).

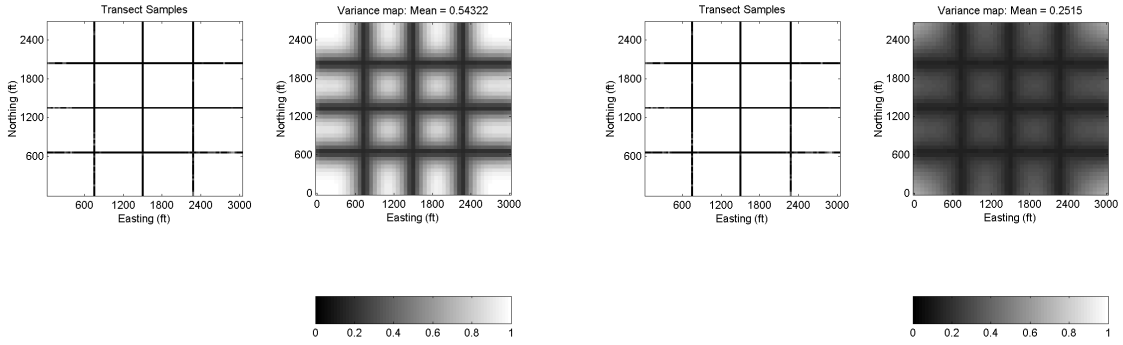
In order to sample the site exhaustively, a total of 151,250 meters of sampling transect with an 8 meter wide footprint are necessary. The total cost of this sampling is $81,675. This cost is shown as the upper right point in the left image of Figure 12. The results in Figure 12 clearly show that the cost of sampling is simply a linear transform of the total sample length. However, the reduction in variance is certainly non-linear. The total variance at the site without any sampling is normalized to 1.0 (Y-axis of right image in Figure 12). With the addition of only three transects, or sampling 4.4 percent of the total site, the predictive variance has been reduced to less than 50 percent of the original value. These results indicate that the rate of variance reduction will decrease drastically with initial sampling and then decrease more slowly before it reaches a value of 0.0 when the site is exhaustively sampled.

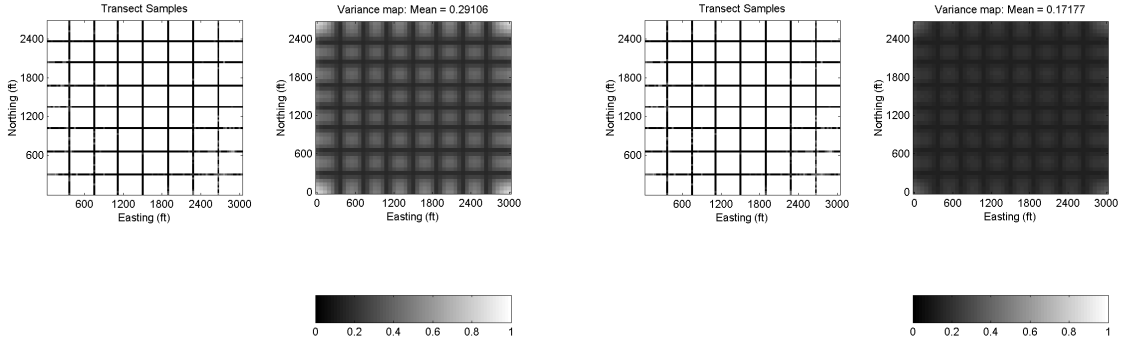
Variance Reduction

The second example illustrates the relationship between sampling density and reduction of MPEV at all sites. The goal of this exercise is to demonstrate on a hypothetical example site how the reduction in the MPEV could be used as a stopping rule to determine when enough samples have been taken at the site. The grid used measures 3030 ft in the east-west direction, 2700 ft in the north-south direction. In Figure 13, all the variance map plots in the left column are based on an isotropic, spherical variogram with range of 600 ft and a 10 percent nugget effect, while the variance map plots in the right column use the same form of variogram, but one with a range of 1800 ft. The first row of Figure 13 has two transects, each dividing the region into half, and together having a total length of 5730 ft. The second row has six transects, where the spacing is 675 ft in the north-south direction and 750 ft in the east-west direction. The total length of the six transects is 17,200 ft. The third row has 14 transects, with spacing of 340 ft (north-south) and 380 ft (east-west). The total length of the 14 transects is 40,100 ft. Finally, the fourth row has 30 transects, with spacing of 170 ft (north-south) and 190 ft (east-west). The total length of the 30 transects is 171,900 ft. All these designs are systematic sampling designs and are well suited for guaranteeing detection of target areas with dimensions greater than that of the transect spacing. The disadvantage to systematic sampling

designs is that the cost of taking measurements nearly doubles as the spacing between transects halves.







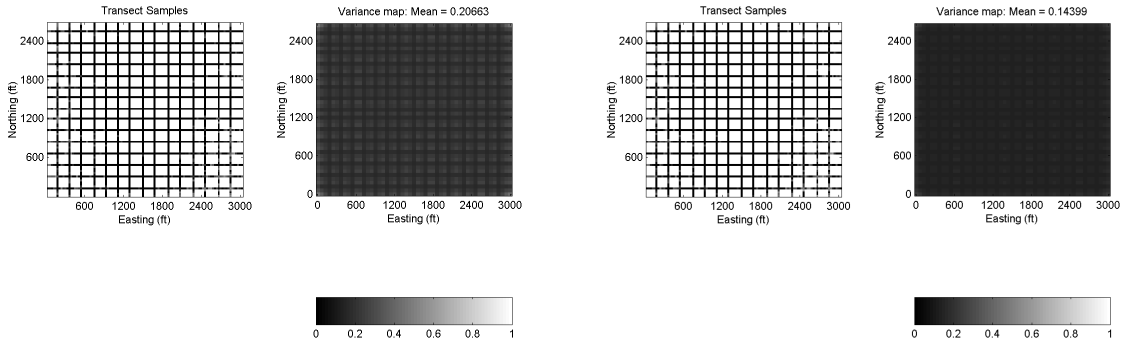


Figure 13. Demonstration of variance reduction as a function of density of transect sampling. The left column uses a spherical variogram with a range of 600 ft, and the right column uses a spherical variogram with a range of 1800 ft. Both variograms are isotropic. The mean variance has been normalized to a value of 1.0 (white) in these images.

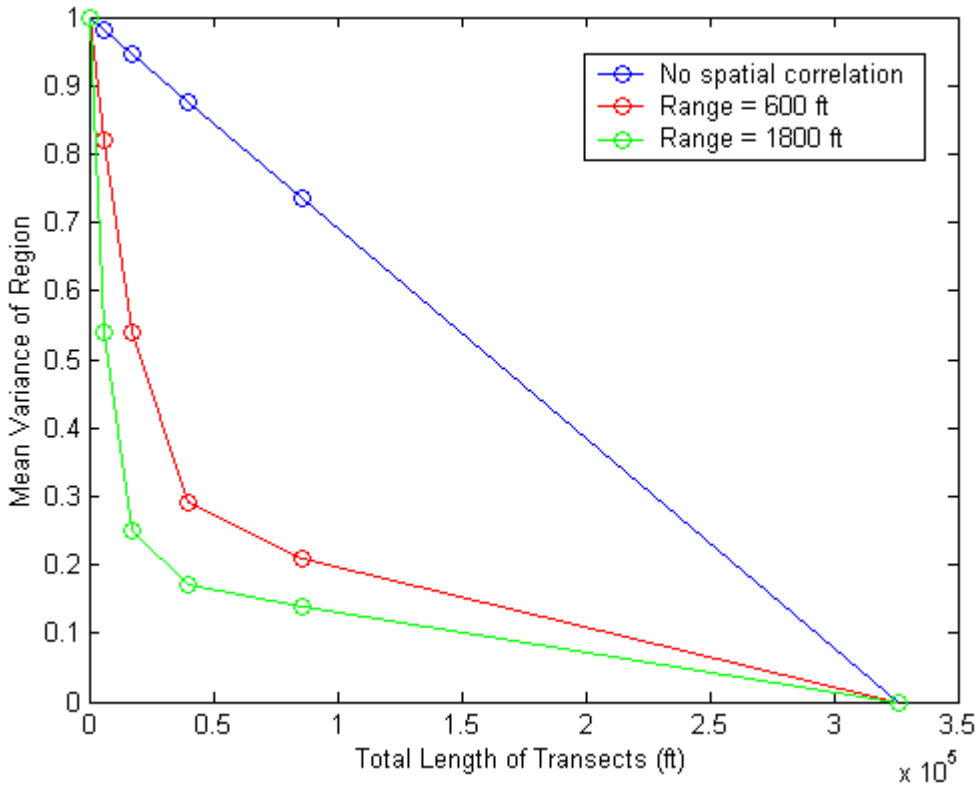


Figure 14. Mean variance of region versus total length of transects for plots in Figure 13.

Figure 14 summarizes the results in Figure 13 by plotting the relationship between the total length of all transects and MPEV for the region. Assuming a 25 ft footprint, zero MPEV is

obtained at 327,000 ft of sampling. For sampling at 10 percent coverage, the MPEV is less than 40 percent of the original when the variogram range is 600 ft, and the MPEV is less than 20 percent of the original when the variogram range is 1800 ft. Since 10 percent sampling coverage, roughly 14 transects, produces a 90 percent cost reduction relative to surveying the entire site, it may be worth considering ceasing the systematic sampling when the MPEV is reduced to 20-30 percent of its original value

If there is no spatial correlation, sampling provides no information about the region outside of the footprint. In this case the MPEV is only linearly proportional to the amount of sampling done as seen in Figure 15. In the case of no spatial correlation, sampling design makes no difference in the MPEV reduction as long as samples are not repeated.

Variance Reduction at Isleta Pueblo Site

Magnetometer data collected at the Isleta Pueblo "Isleta7" site are used to demonstrate the variance reduction techniques applied in the previous hypothetical examples on an actual UXO site. Additionally, the sampling results are used with probability indicator kriging to define decisions made regarding the future disposition of each location at the site. The decision results are determined by comparison to the known sample values and the amount of correct, false positive, and false negative decisions at each stage in the sampling are compared.

The Pueblo has sites that were once used as bombing ranges and would like to determine the extent of unexploded ordnance (UXO) at these locations. Consequently, exhaustive surveys were done at some of the locations using the airborne magnetometer system developed by ORNL. These data are used here as ground truth in testing systematic sampling designs. The original sample spacing across the site was 3 ft, and the goal in this exercise is to determine if regions with high magnetic signals can be defined with significantly fewer sample transects. The original data set has a mean value of 1.77 nT/m, but a maximum value of 16,964 nT/m, so a $\log_e$ transform was performed. In addition, the original data set was aggregated to reduce the number of records so that the below techniques could be conducted. Each 30 ft by 30 ft region, which contains 100 observations, was replaced by the maximum value of the 100 observations relocated to the center of the 30 x 30 foot region. By preserving only the maximum of the 100 data values, this data reduction process will result in a more conservative set of decisions than would using all of the data. In addition, two subregions on the east and west edges of the study region were removed due to sparse sampling. The original data are shown in Figure 15 along with a map indicating locations where the signal is above 3.0 on the $\log_e$ scale.

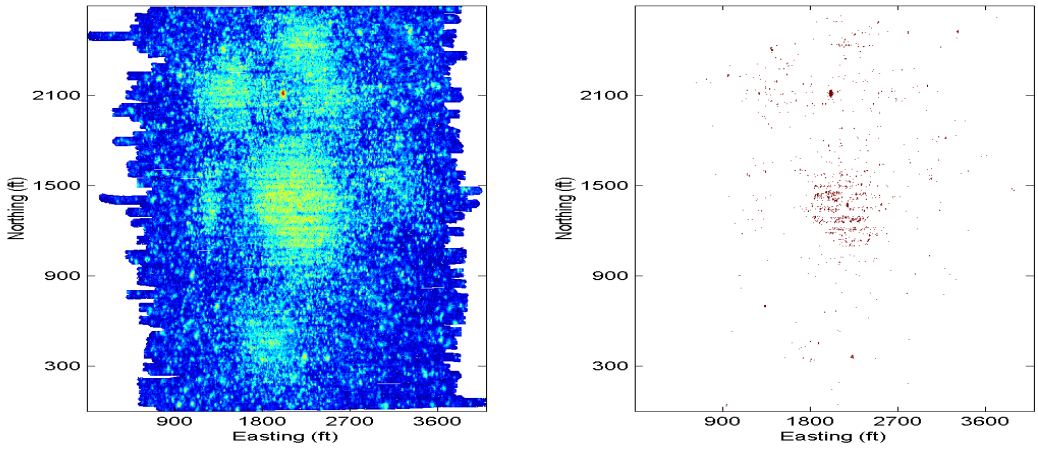


Figure 15. Map of original Isleta7 data with map of samples above threshold.

The results of the airborne magnetometer survey are shown in the left image of Figure 15 and the locations of the anomaly values above and below the 20 nT/m threshold are shown in the right image of Figure 15. In this example, we are interested in accurately and efficiently determining the location of the magnetic anomalies that have signals greater than 20 nT/m. In the absence of excavation information, these "anomalies of interest" serve as proxies for the UXO. The maps of the Isleta site shown in Figure 15 indicate that this site presents a difficult characterization problem. There is not a clearly defined single target area that contains all anomalies greater than 20 nT/m. At least half of the anomalies of interest are randomly distributed in areas surrounding the center area of the site.

The analysis is done in several stages. Figure 16 shows four levels of sampling that use 2, 6, 14 and 30 transects (the series is given by 2^n-2, for n = 2, 3, 4, 5). Black denotes unsampled regions, red denotes samples above the threshold of 20 nT/m, and blue denotes samples below the threshold of 20 nT/m. These transect data are the basis of making predictions through a geostatistical estimation algorithm and these predictions are checked against the complete survey of the region

Figure 17 shows probability maps based on the data displayed in Figure 16. Note that the transect data are honored exactly in the creation of the probability maps: only probabilities of 0 or 1 occur along the transects. The probabilities in the regions beyond the transects have been calculated using indicator kriging using a variogram model fit to the empirical variogram calculated from the data collected on the transects. The modeled variogram based on two transects has range 400 ft in the north-south direction and 700 ft in the east-west direction with nugget of 0.123 and sill of 0.117. For 6 transects, the modeled variogram has range 700 ft in the north-south direction and 800 ft in the east-west direction, with a nugget of 0.094 and sill of 0.026. Note that the spacing interval of the transects is roughly the same as the range of the variogram , so all points in the region are well within the range of sampled locations. Figure 18 shows the fit of this variogram to the data obtained from the software package VarioWin. For 14 and 30 transects the ranges are similar and the transect spacing is less, so the probability predictions are more heavily constrained than are the sampling designs with fewer transects.

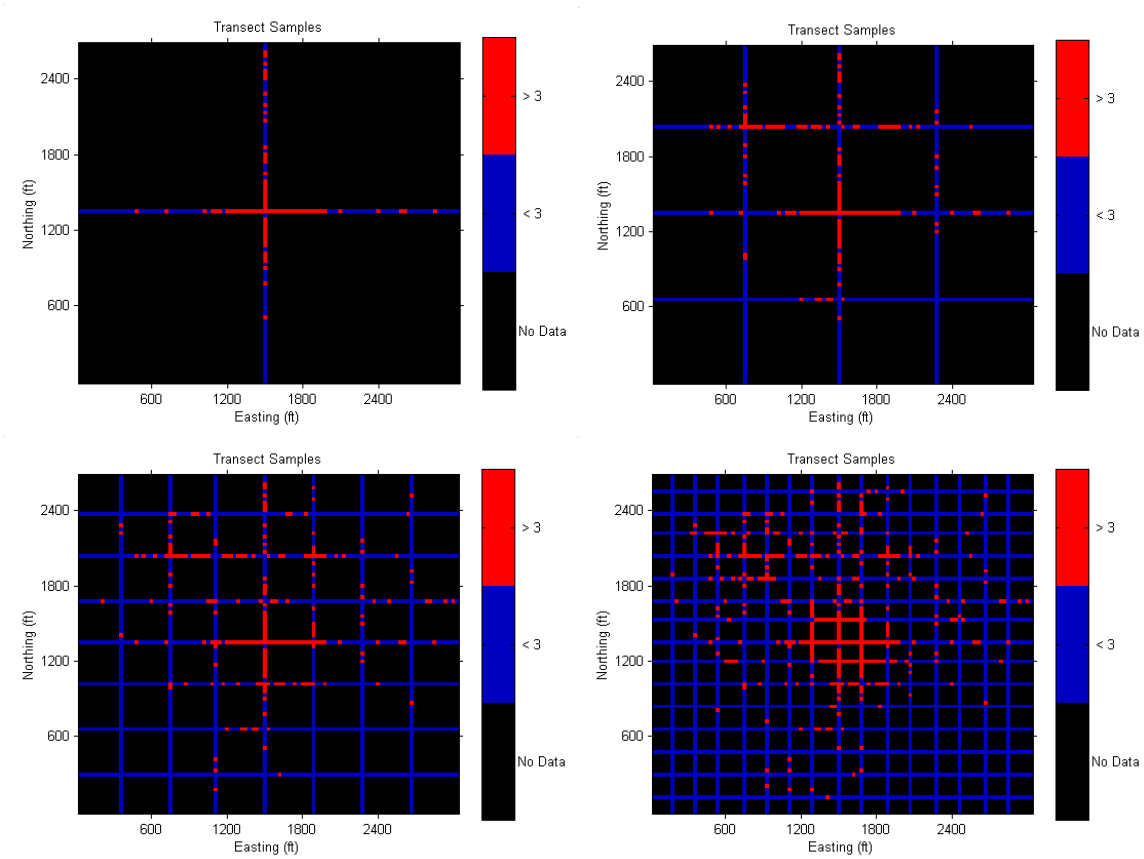


Figure 16. Result of transect sampling with differing densities of transects. For all plots, red indicates signal strength above 20 (nT/m), blue signal strength below 20, and black indicates no data collected. Note that the colorbar is on a log scale (3 = $\log_e(20)$).

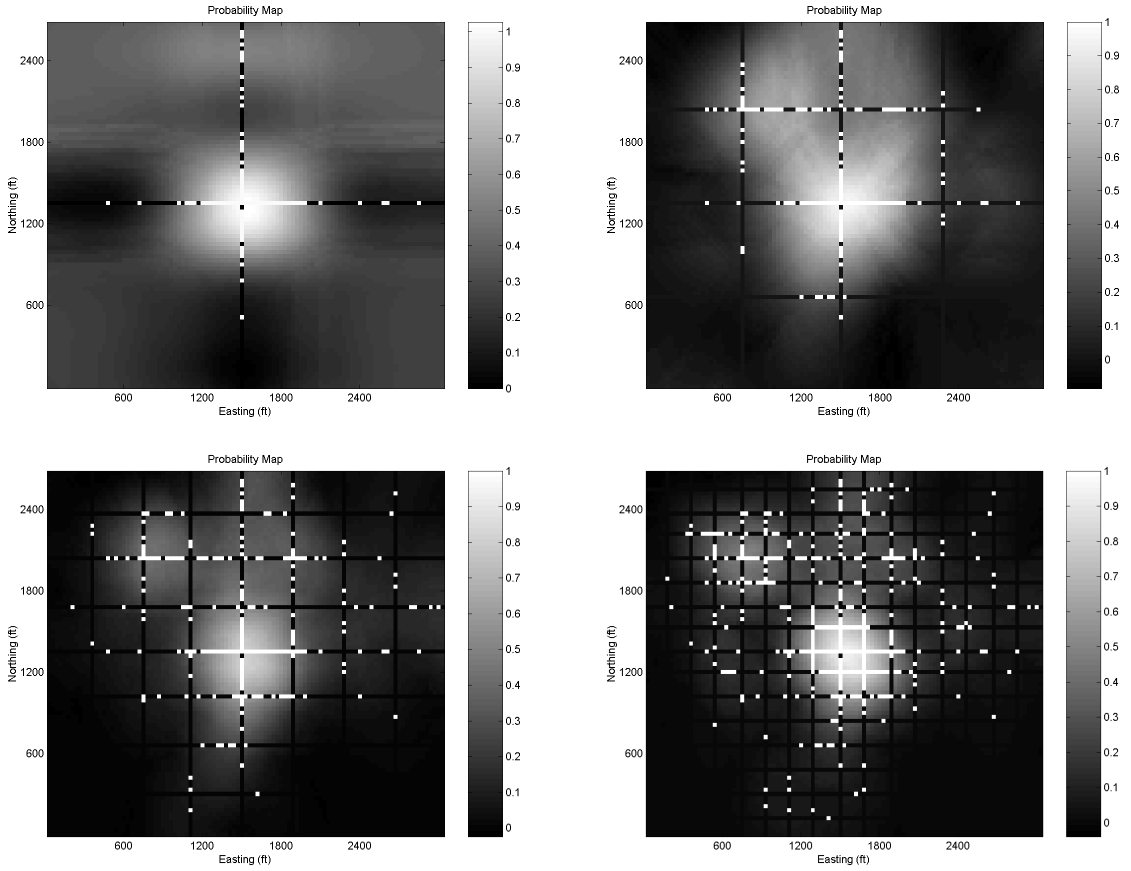


Figure 17. Probability map constucted by indicator kriging using the transect data given in Figure 16.

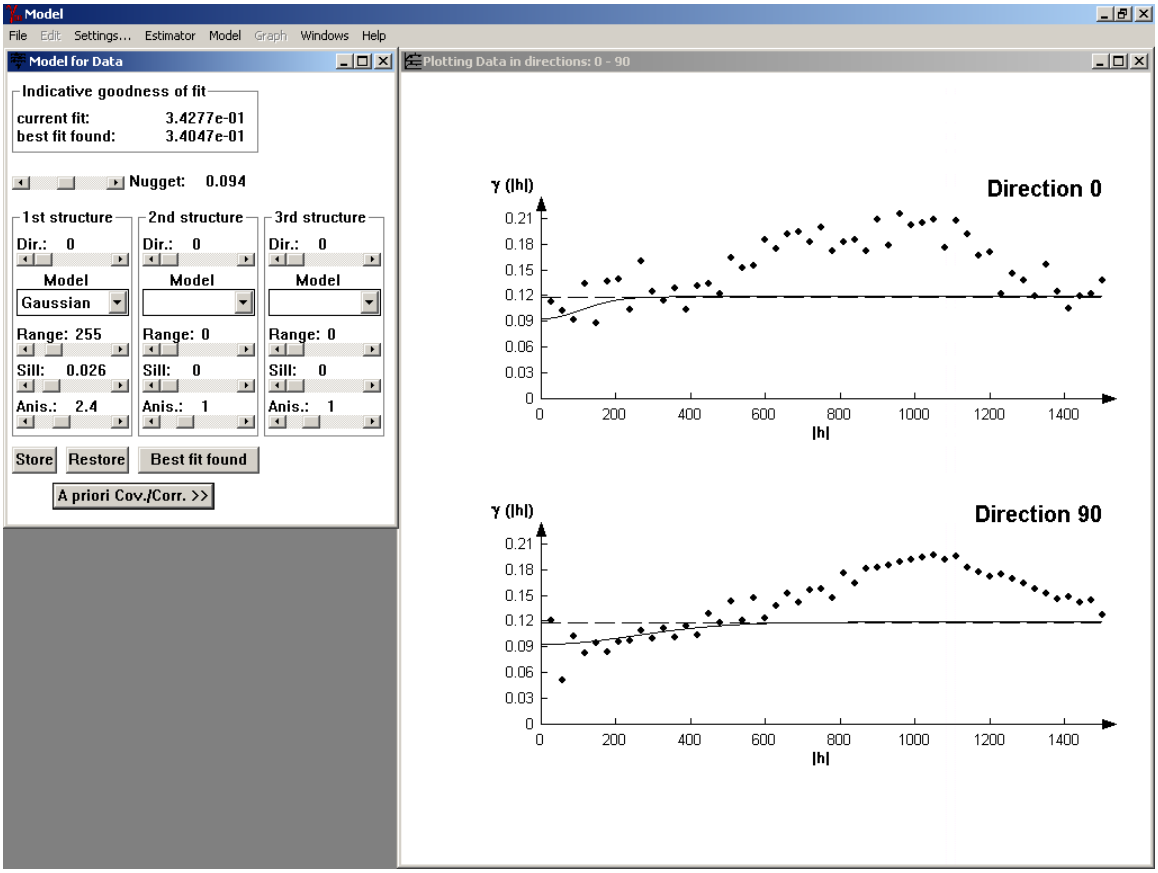


Figure 18. Variogram fit done in VarioWin.

The probability maps along with a level of design reliability determine the decision as to whether or not additional surveying, or excavation, at a location is necessary or to classify the location as needing no further action. More details on using probability maps to make decisions at UXO sites can be found in McKenna (2001) and McKenna et al. (2002). Table 1 shows the two types of error possible. One can excavate a location that did not have UXO, and therefore did not need excavation, and this type of error is called a false positive. Otherwise, a location that actually has UXO can be left as is; this second type of error is a false negative. A false negative is dangerous since it leaves UXO in the ground, while a false positive only adds to the total cost of the excavation. False negatives can be eliminated if the site is completely excavated, though this is generally too expensive to do. False positives can be eliminated if no excavation is performed, though this inaction is not responsible and does not remove any of the UXO. In practice, it is necessary to decide on a level of design reliability that can reasonably balance costs and risks.

Action/Ground Truth	**UXO exists**	**No UXO**
Remediate	Correct Decision	False Positive
Do Nothing	False Negative	Correct Decision

Table 1. Results of decision made given the ground truth.

Figure 19 shows the decisions made at four different levels of sampling for a design reliability of 0.95. Note that there is a large increase in correct decisions going from a sampling design of two transects (upper left image of Figure 19) to one with 6 transects (upper right images of Figure 19). However, adding additional transects beyond six does not produce similar large reductions in false positives. See Table 2 for exact numbers of the different decision results.

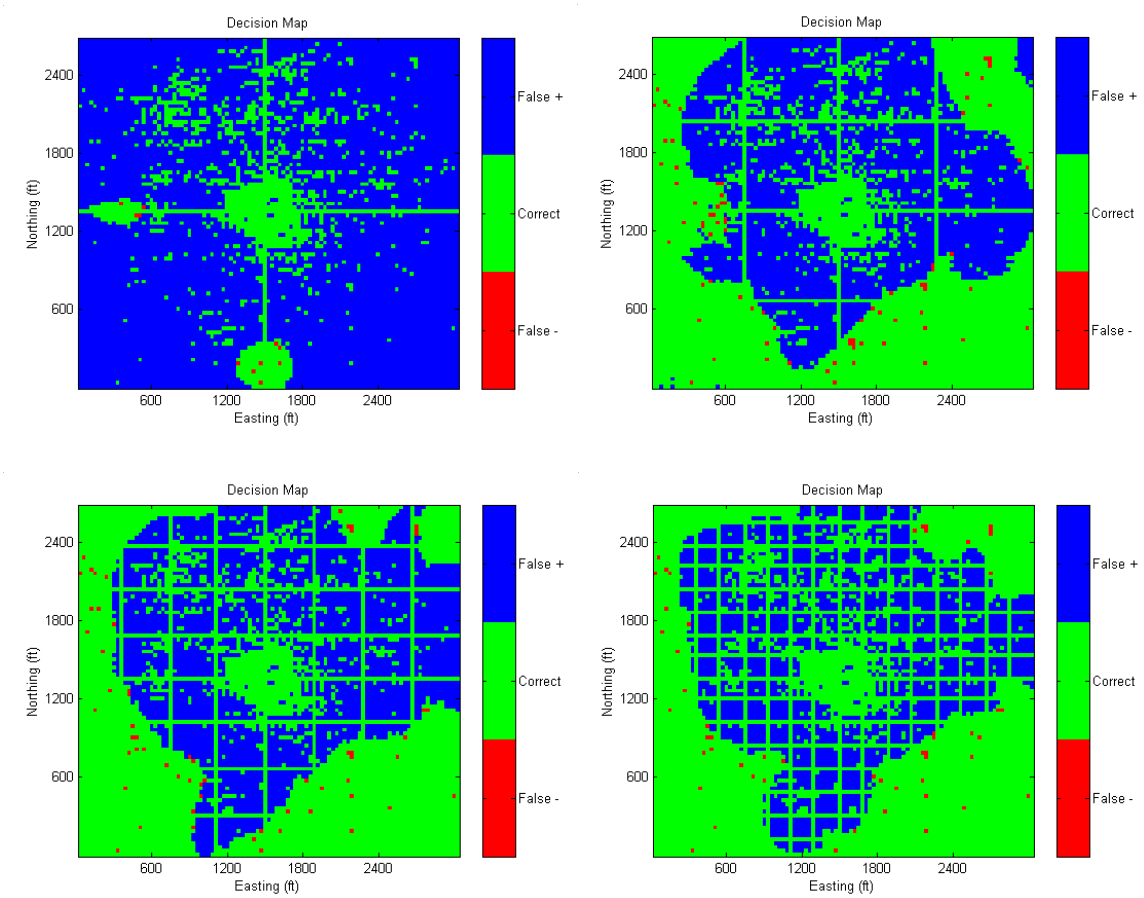


Figure 19. Classification of decisions to remediate where blue denotes a false positive, red denotes a false negative, and green denotes a correct decision.

# Transects	# Correct Decisions	# False Positives	# False Negatives
2	1461	7618	11
6	5063	3948	79
14	5258	3772	60
30	6082	2955	53

Table 2. Comparison of sampling density and decisions made.

The results shown in Figure 19 and Table 2 indicate that the lowest number of false negative decisions occurs when only two sample transects are taken. However, only taking two transects also results in the highest number of false positives. Adding four more transects to the sample design decreases the number of false positives by nearly 50 percent, but these less conservative decisions based on the six transects also result in an increase of the false negatives from 11 to 79. Additional sampling transects further reduce the number of false positive and false negative decisions, but the return on the sampling investment is not as great as it was for the initial sampling transects.

Table 2 and Figure 19 both are based on a design reliability of 0.95, and it is natural to consider other design reliabilities. Figure 20 plots the proportion of correct decisions (green line), false positives (blue line) and false negatives (red line) as a function of the chosen design reliability. Note that the decision results generally improve across all design reliabilities as the number of transects increases, but, as mentioned above, much of this improvement is seen in the transition from 2 to 6 transects and only incremental improvements are seen with the addition of more transects.

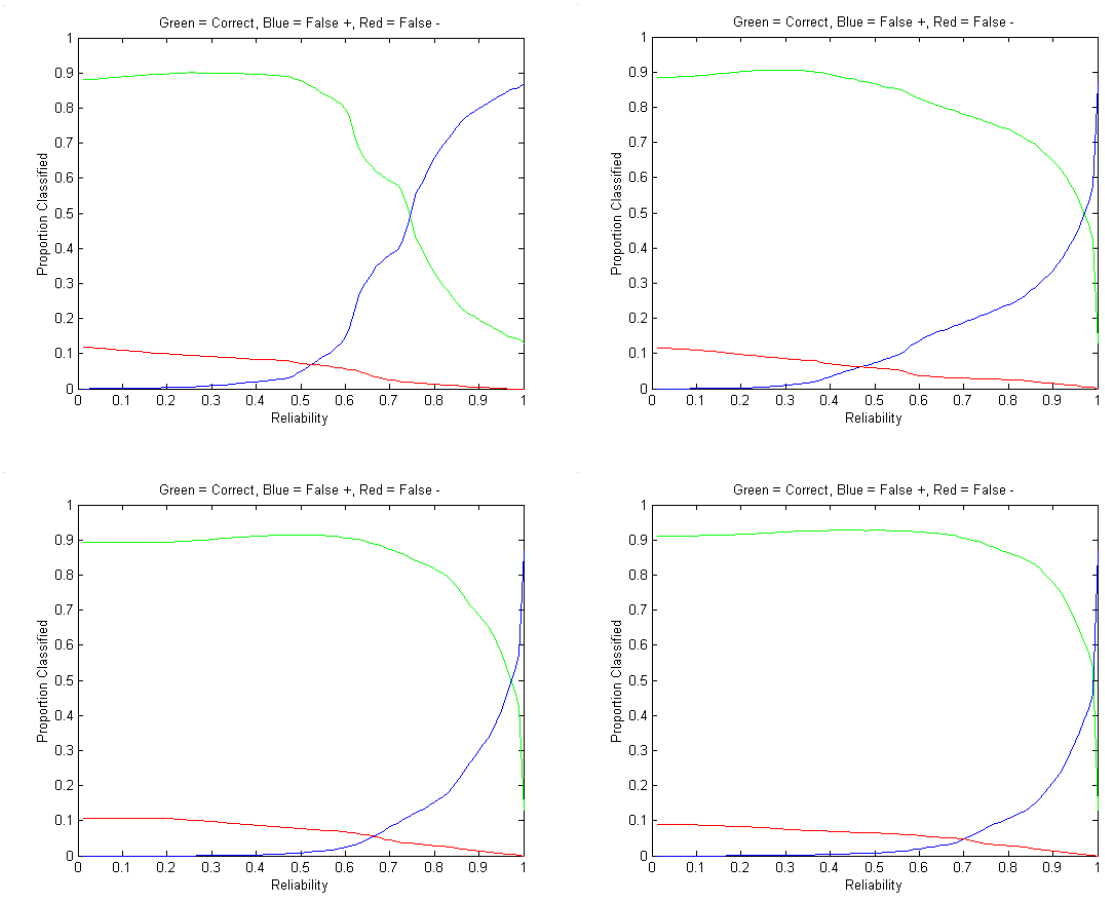


Figure 20. Decision results curves corresponding to the respectively placed plots in Figure 19.

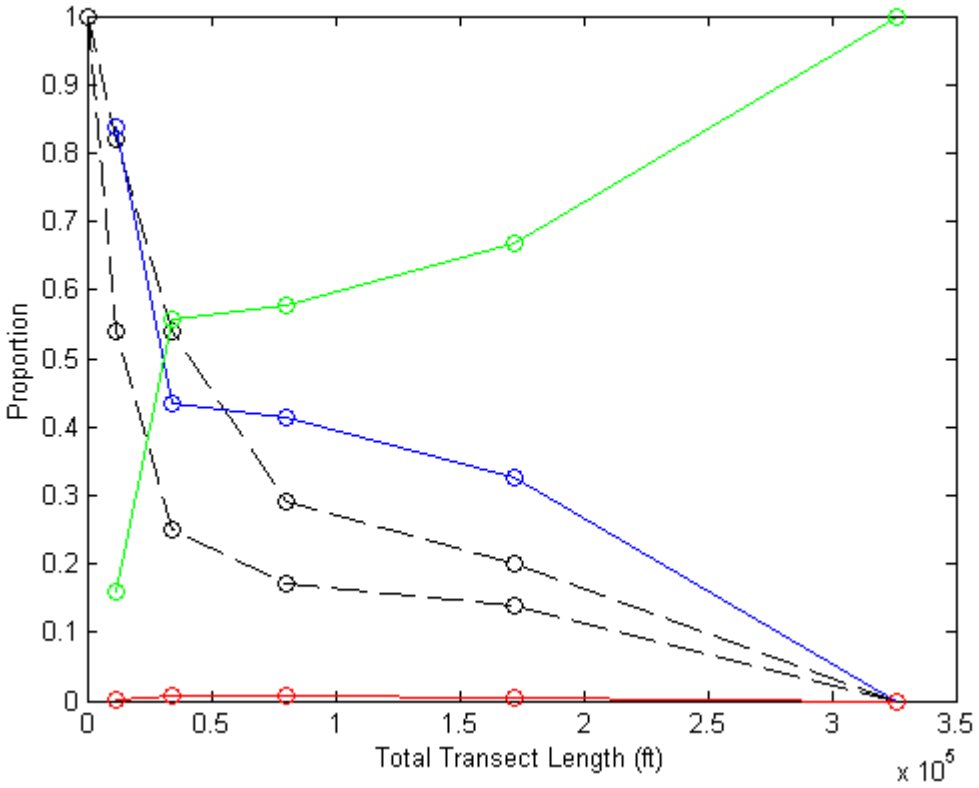


Figure 21. Proportion of correct decisions (green), false positives (blue), false negatives (red) and the proportion of MPEV for two different variogram models relative to the original level as a function of the amount of transect sampling.

If the reduction in predictive variance is used to determine when to stop sampling, ideally there would be a direct relationship between the amount of predictive variance for a given sampling design and the results of the decisions that would be made at that level of sampling. Unfortunately, this relationship will be different for every site and sampling pattern and it is not possible to determine this relationship prior to excavation of the site. Although the results are only applicable to the Isleta 7 site, it is instructive to examine the relationship between the variance reduction and the decision results at the 95 percent design reliability. This relationship is shown in Figures 21 and 22. As MPEV decreases note that both the proportion of false positives and false negatives go to zero. Although the proportion of false negatives monotonically decreases as MPEV decreases, this is not true for the false positives. For very short total sampling length, the false positives initially go up (Figure 22), and then decrease monotonically to zero. This behavior gives a concave shape to the plot in Figure 22.

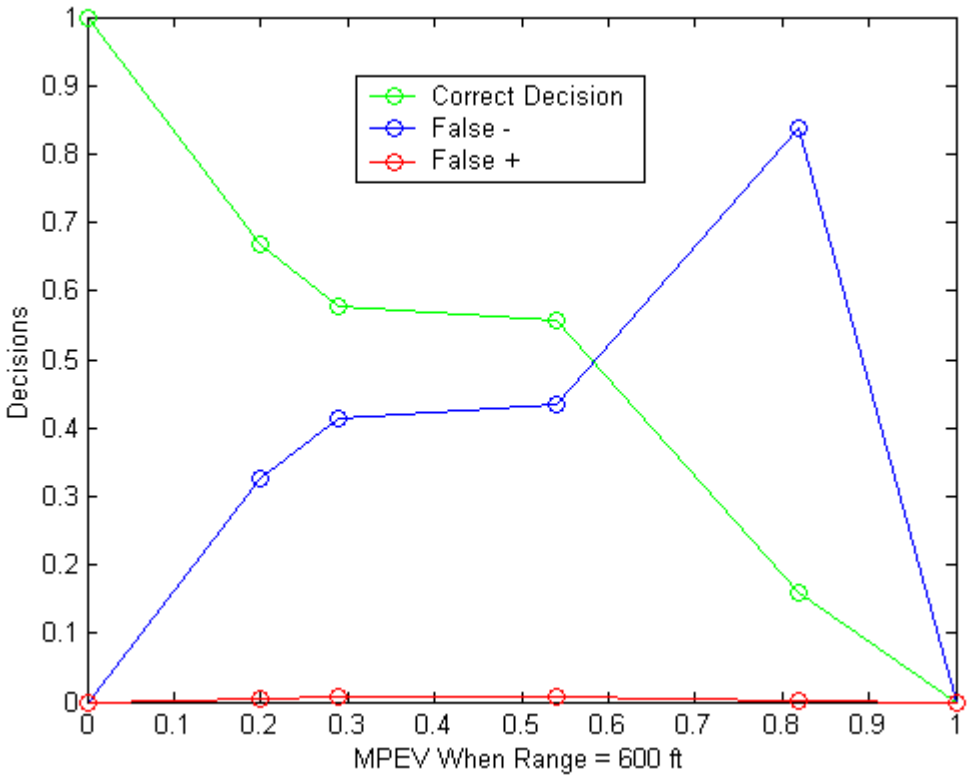


Figure 22. Decisions results as a function of MPEV for the variogram with a 600-foot range.

Discussion of Results

Each of the three examples discussed above illustrates the power of spatial sampling design to make conclusions beyond the existing samples in a region. The first example utilizes the theory of predictive error variance and D-optimality to construct optimal meandering transects for a region with a signal that obeys a known variogram model by using simulated annealing. The variance map shown in Figure 11 shows the ability of each additional transect to reduce variability, and so increase the information known about the region.

The second example shows how the mean variance of a region decreases as a systematic transect sample design is made more and more dense. However, Figure 14 shows decreasing returns as the number of transects grows which suggests that limited sampling will allow reasonable prediction of the signal beyond the samples taken and that to reduce the prediction error variance to levels below 10 percent may require sampling of 50 percent or more of the site area. The second example also demonstrates the effect of spatial correlation on the ability to predict the presence or absence of UXO away from the sample transects. Three cases were examined in this example: no spatial correlation and variogram ranges of 600 and 1800 feet, or roughly 0, 20 and 60 percent of the width of the site. At 5 percent of the region sampled

(approximately two transects), MPEV is 0.95, 0.54 and 0.25 of the original MPEV for no spatial correlation, variogram range of 600 ft, and variogram range of 1800 ft, respectively. These results show that if spatial correlation is not taken into account, then predictions of the occurrence of UXO can only be made at the sampled locations and 100 percent sample coverage will be necessary. Fortunately, the distribution of UXO at most sites does display spatial correlation and this can be used to advantage in making decisions on actions to be taken across the site.

The third example using data collected from the Pueblo of Isleta also shows decreasing returns on the amount of variance reduction achieved for increasing amounts of sampling transects. Although the amount of data more than doubles each time the transect sampling density increases, Table 2 shows that the most improvement is achieved in going from just two to six transects. The additional denser transect samples do show further improvement in the reduction of the variance, but by much smaller percentages. Since collecting samples from six transects gives comparable results to 14 and 30 transects, it makes sense to consider stopping the systematic sampling after the six transects are collected. At this stage in the sampling, it might be advantageous to switch from systematic sampling of the entire site area to focused sampling in those areas where the systematic sampling has indicated the presence of UXO.

These three example problems demonstrate techniques for the optimal location of meandering transects and the calculation of the reduction in prediction error variance that will occur for any sampling transect design prior to conducting that sampling design. Although not shown here, the optimal location of straight transects can also be calculated using the same simulated annealing approach demonstrated for meandering transects. The second and third examples show the reduction of the prediction error variance obtained by the proposed addition of straight transects to an initially unsampled site. These calculations can also be completed for proposed meandering transects and can be done for any combination of existing and proposed straight and/or meandering transects. Both the optimal transect location and the predictive error variance calculations can be done relatively quickly on a typical laptop computer, thus making field-based, near real time, iterative sample design a possibility.

Conclusions

The availability of mobile sensor platforms has increased the amount of data collected along transects, both straight and meandering. Recently developed towed and airborne sensor platforms for the survey of UXO sites are a prime example of sensors that provide large amounts of data along transects. The techniques and algorithms available to fully exploit this wealth of transect data have not kept pace with the development of the sensors. This report presents a new approach to determining the optimal location of additional transects that uses the minimization of MPEV as the optimization objective function. Additionally, this report extends previously developed approaches for the calculation of prediction error variance to use with data collected along transects. These approaches use either D or A optimality criteria to calculate the reduction in variance that will occur for any proposed sampling design. The approach based on D-optimality is used in the simulated annealing algorithm used in the first

example. A-optimality, which is less computationally intensive, is used in the second and third example, and is equivalent to minimizing MPEV.

The three example problems demonstrate the techniques developed for optimization of transect sampling on two hypothetical sites and the Isleta Pueblo site. The first example problem demonstrates the ability of a simulated annealing algorithm to optimally locate meandering transects under the objective of minimizing the variance of the prediction error across the site. This same approach could be used to locate straight sampling transects as well. The second and third examples show how the reduction in the variance of the prediction error can be calculated for any proposed sampling design. Results from the second and third examples show that the reduction in mean prediction error variance is not a linear function of the amount of transect sampling. Results show that for several different variogram models on both a hypothetical site and an actual site, systematic sampling of 30 percent or less of the site with straight, parallel transects can reduce the mean prediction error variance to 10-20 percent of the initial mean prediction error variance. Impractically large amounts of additional sample transects are necessary to reduce the mean prediction error variance below 10-20 percent of the initial values.

The relationship between the MPEV and the quality of the decisions made on whether or not to excavate certain locations or leave them as is, is not easily defined. The results for the Isleta site show that as the MPEV is reduced, the number of correct decisions increases and the number of false positives and false negative decisions decrease, but these results are specific to the variogram, sample data and the true distribution of the UXO at the Isleta site and will not directly transfer to other sites.

The reduction in predictive error can be used as a stopping rule to decide when enough samples have been taken. For example, it may be decided to take enough systematic sample transects to reduce the MPEV across the site less than or equal to 20 percent of the initial MPEV. However, the lack of a direct connection between the MPEV and the eventual decision results based on the sampling pattern that achieves that MPEV may preclude the use of the prediction error variance alone as a stopping rule. One new idea that has come up during this work is to initiate site characterization with systematic sampling (e.g., parallel transects with a fixed transect spacing) using a transect spacing designed to reduce the prediction error variance to a prescribed level, say 20 percent of the initial MPEV. After these samples are obtained, the site characterization can switch from systematic sampling to exhaustive sampling within small regions of the site where the systematic sampling identified the presence of UXO. Such a hybrid sampling design would provide adequate coverage of the entire site as evidenced by the reduction in the MPEV under systematic sampling and would also allow for focused exhaustive sampling in small areas of the site. Only exhaustive sampling of the entire site can guarantee no false negatives, but this level of sampling is impractical for large sites. If a design reliability can be chosen in coordination with the stakeholders for a given site, quality results can be obtained at a much lower cost relative to exhaustive sampling.

Acknowledgements

This work was done under funding provided by the U.S. Department of Defense, Strategic Environmental Research and Development Program (SERDP). The authors thank Bill Doll

and Jeff Gamey of Oak Ridge National Laboratories for providing the geophysical sensor data obtained on the Isleta 7 site. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

References

Back, P., 2000. Sampling Strategies and Data Worth Analysis for Contaminated Land. SGI Varia 500. Swedish Geotechnical Institute, Göteborg.

Burnham, K. P., Anderson, D.R. and Laake, J.L., 1980. Estimation of Density from Line Transect Sampling of Biological Populations. Wildlife Monographs No. 72, Supplement to The *Journal of Wildlife Management*, 44.

Deutsch, C. V., 1994. Kriging with Strings of Data. *Mathematical Geology*, 26, 623-638.

Domburg, P., de Gruijter, J. and van Beek, P., 1997. Designing Efficient Soil Survey Schemes with a Knowledge-Based System Using Dynamic Programming. *Geoderma*, 75, 183-201.

EPA Quality Staff, 2000. Guidance for Choosing a Sampling Design for Environmental Data Collection. EPA QA/G-5S. Washington, D.C.

Fedorov, V. V. and Hackl, P., 1994. Optimal Experimental Design. *Calcutta Statistical Association Bulletin*, 44.

Freeze, R.A. and Gorelick, S.M., 1999. Convergence of Stochastic Optimization and Decision Analysis in the Engineering Design of Aquifer Remediation. *Ground Water*, 37, 934-954.

Freeze, R.A., Massmann, J., Smith, L., Sperling, T. and James, B., 1992. Hydrological Decision Analysis. National Ground Water Association, Dublin, OH.

Gilbert, R. O., 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, New York.

James, B.R. and Gorelick, S.M., 1994. When Enough is Enough: The Worth of Monitoring Data in Aquifer Remediation Design. *Water Resources Research*, 30, 3499-3513.

James, B.R., Gwo, J-P, and Toran, L., 1996. Risk Cost Decision Framework for Aquifer Remediation Design. *Journal of Water Resources Planning and Management*, 122, 414-420.

Jassby, A.D., Cole, B. E. and Cloern, J. E., 1997. The Design of Sampling Transects for Characterizing Water Quality in Estuaries. *Coastal and Shelf Science*, 45, 285-302.

McKenna, S.A., 2001. Application of a Doubly Stochastic Poisson Model to the Spatial Prediction of Unexploded Ordnance, in Proceedings of the Annual International Association of Mathematical Geology (IAMG) Annual Conference, Cancun, Mexico, Sept., 9-12.

McKenna, S.A., H. Saito and P. Goovaerts, 2002. Estimating the Spatial Distribution of UXO from Limited Sample Data Using Geostatistics, in: Proceedings of the UXO/Countermine Forum, Sept. 3rd-6th, Orlando, Florida, 11 pp.

Michalewicz, Z. and Fogel, D., 2000. *How to Solve It: Modern Heuristics, 2nd Printing*. Springer-Verlag, New York.

Myers, R. H. and Montgomery, D. C., 1995. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. John Wiley & Sons, New York.

National Search and Rescue Committee (NSARC), 2000. United States National Search and Rescue Supplement. Washington D.C.

Press, S. J., 1972. *Applied Multivariate Analysis*. Holt, Rinehart and Winston, New York.

Rouhani, S., 1985. Variance Reduction Analysis. *Water Resources Research*, 31, 937-846.

Strang, G., 1986. *An Introduction to Applied Mathematics*. Wellesley-Cambridge Press, Wellesley, MA.

van Groenigen, J. and Stein, A., 1998. Constrained Optimisation of Spatial Sampling Using Continuous Simulated Annealing. *Journal of Environmental Quality*, 27, 1078-1086.

van Groenigen, J., Siderius, W. and Stein, A., 1999. Constrained Optimisation of Soil Sampling for Minimisation of the Kriging Variance. *Geoderma*, 87, 239-259.

van Groenigen, J., 2000. The influence of Variogram Parameters on Optimal Sampling Schemes for Mapping by Kriging. *Geoderma*, 97, 223-236.

van Groenigen, J., Pieters, G. and Stein, A., 2000. Optimizing Spatial Sampling for Multivariate Contamination in Urban Areas. *Environmetrics*, 11, 227-244.

Zawadzki, W. and Beckie, R., 1996. The Effect on Measurement Scale on the Worth of Hydraulic Conductivity Data: Slug Tests and Pumping Tests. Uncertainty in the Geologic Environment: From Theory to Practice, ed. Shackelford, C.D., Nelson, P.P. and Roth, M.J.S, Geotechnical Special Publication No. 58.

Distribution

Internal:

7	MS 0735	(6115) Roger Bilisoly
3	MS 0735	(6115) Sean McKenna
1	MS 0735	(6115) Erik Webb
1	MS 0719	(6131) James Phelan
10	MS 0731	(6805) NWM Library
1	MS 9018	(8945-1) Central Technical Files
2	MS 0899	(9616) Technical Library
1	MS 0612	(9612) Review and Approval Desk for DOE/OSTI

External:

Richard O. Gilbert
Pacific Northwest National Laboratory
6 Clemson Court
Rockville, MD 20850

Brent Pulsipher
Battelle,
Pacific Northwest National Lab
P.O. Box 999 / MS K5-12
Richland, WA 99352

Evan Englund
USEPA National Exposure Research Laboratory
Environmental Sciences Division/ORD
P. O. Box 93478
Las Vegas, NV 89193-3478

Bill Doll
Oak Ridge National Laboratory
PO BOX 2008 MS6038
Oak Ridge TN 37831-6038

Jeff Gamey
Oak Ridge National Laboratory
PO BOX 2008 MS6038
Oak Ridge TN 37831-6038

Les Beard
Oak Ridge National Laboratory
PO BOX 2008 MS6038
Oak Ridge TN 37831-6038

George Ostrouchov
Oak Ridge National Laboratory
PO BOX 2008 MS6367
Oak Ridge TN 37831-6367

Dennis Wolf
Oak Ridge National Laboratory
PO BOX 2008 MS6418
Oak Ridge TN 37831-6418

Dr. Anne Andrews
SERDP Program Office
901 North Stuart Street, Suite 303
Arlington, VA 22203

Mr. Bradley P. Smith
SERDP Program Office
901 North Stuart Street, Suite 303
Arlington, VA 22203

Dr.Jeffrey Marqusee
SERDP Program Office
901 North Stuart Street, Suite 303
Arlington, VA 22203

Mr. Jeff Fairbanks
SERDP Program Office
901 North Stuart Street, Suite 303
Arlington, VA 22203

Mr. Matthew Chambers
SERDP Program Office
901 North Stuart Street, Suite 303
Arlington, VA 22203

Dr. Pierre Goovaerts
Senior Chief Scientist
Biomedware, Inc.
710 Ridgemont Lane
Ann Arbor, MI, 48103-1535

Mr. Hirtoaka Saito
The University of Michigan
Department of Environmental Engineering
1351 Beal Avenue (13 EWRE)
Ann Arbor, MI 48109-2125

Anita Singh
Lockheed Martin Environmental Services
1050 E. Flamingo Road
Suite E120
Las Vegas, NV 89119