Published in final edited form as:

Geoderma. 2009 October 15; 153(1-2): 205–216. doi:10.1016/j.geoderma.2009.08.007.

Second-Phase Sampling Designs for Non-Stationary Spatial Variables

Eric M. Delmelle^{1,2} and Pierre Goovaerts³

¹Department of Geography and Earth Sciences, and Center for Applied Geographic Information Systems, University of North Carolina at Charlotte, Charlotte, NC 28223, U.S.A

³BioMedware, Inc., Ann Arbor, MI, 48104 U.S.A

Abstract

In spatial sampling, once initial samples of the primary variable have been collected, it is possible to take additional measurements, an approach known as second-phase sampling. Additional samples are usually collected away from observation locations, or where the kriging variance is maximum. However, the kriging variance (also known as prediction error variance) is independent of data values and computed under the assumption of stationary spatial process, which is often violated in practice. In this paper, we weight the kriging variance with another criterion, giving greater sampling importance to locations exhibiting significant spatial roughness that is computed by a spatial moving average window. Additional samples are allocated using a simulated annealing procedure since the weighted objective function is non-linear. A case study using an exhaustive remote sensing image illustrates the procedure. Combinations of first-phase systematic and nested sampling designs (or patterns) of varying densities are generated, while the location of additional observations is guided in a way which optimizes the proposed objective function. The true pixel value at the new points is extracted, the semivariogram model updated, and the image reconstructed. Second-phase sampling patterns optimizing the proposed criterion lead to predictions closer to the true image than when using the kriging variance as the main criterion. This improvement is stronger when there is a low density of first-phase samples, and decreases however as the initial density increases.

Keywords

Geostatistics; Grid Spacing; Heuristic Methods; Optimization; Spatial Sampling; Sequential Addition; Stationarity

1 Introduction

When trying to make inferences on soil properties, we are forced to collect a limited number of samples instead of trying to acquire information at every possible location (see, e.g. Cochran 1963, Hedayat *et al.* 1991, Thompson 2002). An extensive inventory yields a clear picture of the variability of the phenomenon, but is time-consuming and expensive. Sparse sampling is cheap, but a paucity of samples may miss important features. As a rule of thumb, sampling is often denser in areas deemed to be critical (Berry and Baker 1968). When a soil property varies spatially, another challenge consists of finding optimal sample locations (Bellehouse 1977, Cressie 1991, Stehman and Overton 1996, Muller 1998, Haining 2003 and more recently de Gruijter *et al.* 2006, Delmelle 2009).

²Corresponding author. Tel: (704) 687-5991. Eric.Delmelle@uncc.edu.

Two distinct techniques are generally recognized in the field of spatial sampling (Haining 1990, Brus and de Gruijter 1997). In a design-based approach the population of values in a region is considered as unknown, but fixed values. Sampling is carried out by taking a number of randomly selected locations and the population values are observed. Design-based sampling patterns are suitable in estimating the mean of soil property over a study region. Repeating a similar sampling pattern numerous times generates a distribution around the estimate of the mean. On the other hand, the model-based approach in this paper considers the population of values as but one realization of a stochastic process; the mean is seen as random variable. The model-based approach is suited to predict the outcome of soil property at a particular location. In *first-phase sampling*, one objective is to collect spatial samples at short distances to estimate the variogram. As such, it is recommended to cluster a few observations in order to quantify the variability at short distances. Recent developments have been suggested to estimate the variogram, based on (1) least-squares through the minimization of the covariance matrix (Bogaert and Russo 1999), and (2) maximum likelihood coupled with simulated annealing (Lark 2002). Capitalizing on the property that kriging provides an estimate of the variance (Cressie 1991), another objective consists of optimally locating initial samples to minimize the kriging variance. When the covariance structure is unknown however, space-filling criteria represent an alternative solution (Royle and Nychka 1998, Brus *et al.* 2007). Since the kriging variance is minimal at observation locations and increases away from them, it is desirable to spread all observations evenly, ensuring that unvisited locations are never far from a sampling point. For that purpose, a triangular or square sampling configuration performs well (Yfantis *et al.* 1987). Once initial samples have been collected, the variable of interest (or soil property) can be estimated at unsampled locations using interpolation techniques, such as kriging. If the number of initial samples is deemed insufficient, or if the interpolation results are judged unsatisfactory, the initial sampling set can be augmented, an approach known as *second-phase sampling*. The essential question is how to construct an efficient design of additional measurements, capitalizing on the spatial information of the soil property, obtained during a first-phase sampling phase. In second-phase sampling, the first sampling stage consists of collecting initial measurements to estimate the variogram and perform interpolation, while the second stage capitalizes on the variogram structure to allocate additional samples. Van Groenigen *et al.* (1999) have applied this technique to estimate sand percentage content on a river terrace in Thailand, by minimizing the maximal kriging variance. Di Zio *et al.* (2004) calibrated the variogram by maximum likelihood estimation, and used simulated annealing to allocate additional observations, minimizing the average kriging variance. Second-phase sampling has been performed in the context of soil contamination (Cox 1999, Hsia *et al.* 2000). It is recommended to sample in locations above a particular threshold and draw a fixed number of additional samples around them until subsequent measurements are below that threshold.

Recently, Van Groenigen *et al.* (2000) applied the concept of spatial weights in the optimization process. Weights were used to detect contaminated zones in the Rotterdam harbor, to distinguish between areas with different sampling priorities. In the first phase, greater sampling weights were assigned to regions exhibiting higher contamination risks. In the second phase, locations expected to exhibit a higher priority for remediation were weighted accordingly. Rogerson *et al.* (2004) have applied a similar technique, increasing the sampling weights where the uncertainty associated with the variable of interest was high, and hence not in areas where the probability of an event occurring is either 0 or 1. The authors used a greedy algorithm to minimize the maximum weighted kriging variance, but more suitable search algorithms were not evaluated in the study.

The kriging variance is unfortunately often misused as a measure of reliability of the kriging estimate (Armstrong 1994, Deutsch and Journel 1997). It assumes that the process is stationary, an assumption often violated in practice. The kriging variance is solely a function of the sample

pattern, sample density, the numbers of samples, along with their covariance structure (Cressie 1991). Optimizing second-phase sampling patterns to minimize the kriging variance will locate additional samples at intermediate positions between existing samples, ignoring the underlying spatial variation of the soil property. There is an incentive to combine the kriging variance with another criterion —under the form of weights— reflecting the magnitude of the spatial variation, even if the density of previously sampled points is low (high kriging variance). This paper attempts to develop a method for selecting additional sampling locations which minimizes the overall prediction-error variance and accounts for non-constant variance of the variable of interest.

In this paper, temporal stationarity is assumed. Collecting additional measurements for the purpose of improving the prediction of time-dependent variables has been carried out by Lajaunie *et al.* (1999), and more recently de Gruijter *et al.* (2006). Additionally, designing spatial sampling patterns of a primary variable in a multivariate context has been addressed by Bueso and Angulo (1999), Hengl *et al.* (2003), Delmelle (2005) as well as Brus and Heuvelink (2007). In the next section, the second-phase sampling problem is modeled mathematically. Section 3 reviews suitable heuristic techniques to locate additional samples. Section 4 illustrates the suggested sampling methodology on an exhaustive remote sensing dataset. In the final section, a summary is provided.

2 Second-phase sampling

In a first sampling phase, a soil property Y is measured at m locations within a study area, D . Measurements are denoted $y(\mathbf{s}_i)$, $\forall i = 1 \dots m$. Kriging is generally performed on a set of grid nodes $\mathbf{s}_g$ ($g = 1, 2 \dots G$) where each neighboring observation is weighted according to a variogram model. Kriging yields an associated kriging variance (prediction error) which measures the uncertainty of the prediction. Following Bailey and Gatrell's notation (1995), the kriging variance at a gridpoint $\mathbf{s}_g$ is as follows:

$$\left(\sigma_k(\mathbf{s}_g)\right)^2 = \sigma^2 - \mathbf{c}^T(\mathbf{s}_g) \cdot \mathbf{C}^{-1} \cdot \mathbf{c}(\mathbf{s}_g), \quad (1)$$

where $\mathbf{C}^{-1}$ is the inverse of the covariance matrix $\mathbf{C}$ based on the covariogram function. The term $\mathbf{c}$ is a column vector and $\mathbf{c}^T$ its corresponding row vector. The Total Kriging Variance (TKV) is obtained by integrating Equation 1 over D . Computationally, it is easier to discretize D over a fine grid of points (set G). The Average Kriging Variance (AKV) over D becomes:

$$\text{TKV} = \frac{\int_D \left(\sigma_k(\mathbf{s}_g)\right)^2}{|D|} \approx \text{AKV} = \frac{1}{|G|} \sum_{g \in G} \left(\sigma_k(\mathbf{s}_g)\right)^2 \quad (2)$$

A general sampling objective is to reduce the kriging variance by as much as possible (Van Groenigen *et al.* 1999). If a set of size n containing new sample points is added to our prediction sample set of size m , the change in kriging variance over all grid points $\mathbf{s}_g$:

$$\text{AKV}^{\text{old}} - \text{AKV}^{\text{new}} = \frac{1}{|G|} \sum_{g \in G} \left| \left(\sigma_k^{\text{old}}(\mathbf{s}_g)\right)^2 - \left(\sigma_k^{\text{new}}(\mathbf{s}_g)\right)^2 \right|, \quad (3)$$

where $\left(\sigma_k^{\text{old}}(\mathbf{s}_g)\right)^2$ and $\left(\sigma_k^{\text{new}}(\mathbf{s}_g)\right)^2$ are the kriging variance calculated with the set $\mathbf{M}$ of m initial sample points and the set $\mathbf{N}$ of n additional set of points respectively. The absolute value of the

difference is calculated since the screening effect could result in negative kriging weights and greater kriging variance. Using Equation 3, the sampling objective $Q[S]$ is defined as:

$$\underset{\{s_{m+1}, \dots, s_{m+n}\}, m \in \mathbf{P}}{\text{Maximize}} \quad Q[S] = \text{AKV}^{\text{old}} - \text{AKV}^{\text{new}} = \frac{1}{[G]} \sum_{g \in G} \left| \left(\sigma_{\mathbf{k}}^{\text{old}}(s_g) \right)^2 - \left(\sigma_{\mathbf{k}}^{\text{new}}(s_g) \right)^2 \right|, \quad (4)$$

where S stands for the sampling pattern. The set $\mathbf{P}$ of p potential points is obtained by discretizing D , generating a minimum of $[(p-m) + (p-m-1) + \dots + (p-m-n+1)]$ to a maximum of $\binom{p-m}{n}$ possible sampling combinations depending on the addition strategy, and assuming $n \neq m, \forall m \in \mathbf{M}, \forall n \in \mathbf{N}$. Since additional observations are not collected at initial sample locations, the first term is $p-m$.

2.1 Weighting the kriging variance

The kriging variance does not account for the roughness of the kriging estimates that is reflected by differences in data value between nearby gridpoints. Let $\hat{y}(s_g)$ be the interpolated value of the soil property Y at a grid node s_g . The objective consists of estimating by how much that grid node is different in value from its surrounding points s_j defined by a neighborhood J . From Figure 1, a circular filter is constructed around each grid node s_g that encompasses its neighbors. The squared difference in interpolated values between the central grid node $\hat{y}(s_g)$ and the surrounding ones $\hat{y}(s_j)$ is computed. The process moves from one gridpoint to another and is repeated for each grid point. The squared difference is then summed over the set G . To regulate the importance of nearby points, a distance factor $d(s_j, s_g)$ as well as a parameter β are introduced. The weight $\lambda(s_g)$ is defined as:

$$\lambda(s_g) = \sum_{j \in J, j \neq g} \frac{d(s_j, s_g)^{-\beta} \cdot \left(\hat{y}(s_j) - \hat{y}(s_g) \right)^2}{\sum_{j=1, j \neq g}^J d(s_j, s_g)^{-\beta}} \quad (5)$$

If the neighborhood J is kept constant and for fixed values of $\hat{y}(s_j)$, $\lambda(s_g)$ will exhibit great values when $\beta < 1$, as more weight is given to far away data points which are likely to be dissimilar. As β increases, $\lambda(s_g)$ decreases since more importance is given to nearby observation values and eventually $\lambda(s_g)$ flattens out for values of $\beta > 4$. An appropriate neighborhood J would encompass adjacent nodes in order to capture changes of variation. If J is too large, zones of rapid changes may go undetected. The weighted kriging variance is computed by multiplying Equation 5 with Equation 2:

$$\text{WPE} = \frac{1}{[G]} \sum_{g \in G} \left(\frac{\lambda(s_g)}{\arg\max_{s_g \in G} \lambda(s_g)} \right)^\alpha \cdot \left(\sigma_{\mathbf{k}}^{\text{old}}(s_g) \right)^2 \quad (6)$$

where α is a parameter controlling the importance given to the weights. Although a value of $\alpha = 1$ is used in this paper, note that when $\alpha = 0$ for instance, Equation 6 reduces to Equation

2. The weights $\left(\frac{\lambda(s_g)}{\arg\max_{s_g \in G} \lambda(s_g)} \right)$ reflecting the spatial roughness are relative weights from 0 to 1.

2.2 Problem formulation

To account for the spatial variation of the primary variable, Equation 4 is modified by introducing a location-specific weighting factor, defined in Equation 5. The second-phase sampling problem is then formulated as a single-weighted objective (Cressie 1991):

$$\underset{\{s_{m+1}, \dots, s_{m+n}\}}{\text{Maximize}} \quad Q[S] = \frac{1}{|G|} \sum_{g \in G} \left(\frac{\lambda(s_g)}{\arg\max_{s_g \in G} \lambda(s_g)} \right)^\alpha \cdot \left| (\sigma_K^{\text{old}}(s_g))^2 - (\sigma_K^{\text{new}}(s_g))^2 \right| \quad (7)$$

When $\alpha = 0$, Equation 7 reduces to an objective which finds the best augmented sampling pattern maximizing the change in kriging variance.

3 Spatial search strategies

In second-phase sampling, the set of additional samples $\mathbf{N}$ is chosen from a set of candidate locations $\mathbf{P}$, relatively large in practice. Since the objective function Q in Equation 7 is non-linear, the search for an optimal sample set $S^* \subset \mathbf{P}$ (or near optimal S^+) must be conducted using a suitable heuristic method H (Michalewicz and Fogel 2000). The efficiency of a heuristic depends on its capacity to give as often as possible a solution S^+ close to S^* (Grötschel and Lovász 1995). In the context of spatial sampling, some research has been devoted to comparing the benefits and drawbacks of different heuristics. The greedy (or myopic) algorithm has been used by Aspie and Barnes (1990), Christakos and Olea (1992), Christakos and Killam (1993) and Rogerson *et al.* (2004). Spatial simulated annealing has been applied by Sacks and Schiller (1988), Ferri and Piccioni (1992), Van Groenigen and Stein (1998) and Pardo-Igúzquiza (1998). Alternative strategies consist of capitalizing on the benefits of different heuristics to reach an optimal solution quicker.

Two different approaches exist to augment an existing initial sampling set. Either a total of n points supplement the existing set, or one point at a time is added, n -times. The latter is defined as sequential addition and is suboptimal while the former is known as simultaneous addition. The sequential approach evaluates the reward of each potential point to the objective function, while the simultaneous approach assesses the merit of a new set of points as a whole. A sequential algorithm is relatively fast, but the simultaneous approach will return results closer to the optimal. In this paper, we use a sequential greedy algorithm, which serves as a starting set to the simultaneous simulated annealing.

3.1 Sequential addition

In the sequential addition, once additional samples have been selected and added to the initial set $\mathbf{M}$, $n - 1$ additional locations are chosen in a similar, sequential fashion. There are $[(p - m) + (p - m - 1) + \dots + (p - n + 1)]$ solutions to the problem, which can be prohibitive when $\mathbf{P}$ is large. Sequential addition is particularly desirable when a specific level of reduction in the weighted kriging variance has to be obtained; new samples are collected sequentially until the level of reduction has been reached. The sequential addition method used in this paper is the greedy (or myopic) algorithm. Greedy supplements the initial sampling set $\mathbf{M}$ sequentially by adding the point corresponding to the location on the weighted kriging variance surface exhibiting the maximum value s_{m+1}^+ (peak). The process continues in a similar fashion n -times.

The augmented set S^+ is given by $S^+ = \mathbf{M} \cup \{s_{m+1}^+ + \dots + s_{m+n}^+\}$. The implementation of the algorithm is relatively trivial, and the algorithm returns a first, good solution (Christakos *et al.* 1992). However, it leads to a suboptimal solution $S^+ \leq S^*$.

3.2 Simultaneous addition

Simultaneous addition consists of supplementing the initial set $\mathbf{M}$ with a set $\mathbf{N}$ of n additional points in one time. Due to the the combinatorial explosion $\binom{p-m}{n}$, we advocate the simulated annealing algorithm using multiple-locations as perturbations. The *greedy* algorithm usually get trapped in a local optimum while simulated annealing escapes to find the optimal solution S^* . Simulated annealing is a method of optimization which emulates how a metal cools and freezes into a minimum energy crystalline structure (Kirkpatrick *et al.* 1983). The algorithm employs a random search that always accepts changes improving the objective function $Q[S]$, but also accepts non-improving perturbations with a certain probability. A non-improving perturbation is accepted with probability P_T where T stands for the current temperature, cooling down as the algorithm progresses, and so does the probability of acceptance (Van Groenigen *et al.* 1999).

After a certain number of iterations, the current temperature and step size are decreased. A high starting temperature T and a slow cooling factor are necessary to find an optimal solution. This will allow the algorithm to escape from a local maximum, at the cost of a greater running time. The algorithm terminates when the temperature of the system T has reached a cutoff-value T_{fin} , which is too cold to allow any unfavorable perturbations. At that final stage however, an additional search is carried out in the vicinity of the best solution found so far, to investigate if better solutions could be hidden at nearby candidate points. When complete, the initial set $\mathbf{M}$ is augmented by the best sampling solution found overall.

4 Case study

In this section, a case study is presented to gain insight into the problem structure. The goal is to find a set of new samples using a combination of sequential greedy and simultaneous simulated annealing in order to maximize the change in the weighted kriging variance (Equation 7). An exhaustive remote sensing image is used to illustrate the procedure but also to test whether augmented sampling patterns which optimize Equation 7 lead to better predictions than designs maximizing the change in kriging variance alone. The sampling strategy is tested using an exhaustive SPOT High Resolution Visible (HRV) scene of a 16km^2 area covered by tropical forests and savannah (Goovaerts 2002). Figure 2 illustrates the exhaustive dataset. The image is divided into 200 rows and columns, yielding a set of 40000 pixels. All computational results were obtained using Matlab v. 7.6. running on a Linux desktop computer.

4.1 Initial sampling patterns

To guarantee a coverage over the study region as well as estimating the variation of the variogram at small distances, initial patterns are designed using a combination of systematic and nested sampling. Additionally, different scenarios of sampling densities are generated to investigate the impact of the sampling density on the results. Figure 3 illustrates the 6 different first-phase sampling scenarios. A 4 by 4 pattern for instance divides the study region into 16 cells or intervals (4 rows and columns), generating 1 systematic sample of 16 sample points. The coordinates of the first sample is purposively chosen within the first interval and corresponds to a location close to the origin. Depending on the location of the first sample, the remaining 15 samples are aligned regularly by the size of the cell (Delmelle 2009). Next, nested sampling is carried out by adding 4 more clustered observations around 4 randomly chosen systematic samples, yielding a total sampling size of 32 observations. Practically, a fixed neighborhood is drawn around each of the 4 randomly selected samples and 4 samples selected at random within this neighborhood. The nested sampling algorithm is flexible in controlling the size of the neighborhood which will influence how clustered nested samples can be. In this

paper, a neighborhood 4 times smaller than the dividing interval is used. The same procedure applies in the 6 by 6 design, 1 systematic sample of 36 observations is generated, while 4 more clustered observations around 6 randomly chosen systematic observations are added. This yields a total of 60 observations. The same procedure is repeated to designs of higher order densities.

For each of the 6 scenarios, 25 total different sets of gridded and clustered data are selected to attenuate the impact of sampling fluctuations. This is implemented by shifting the origin of the systematic sampling pattern while new clustered data are selected at random, resulting in 150 realizations of sampling patterns of varying densities and origins. Figure 4 illustrates the procedure for 2 different scenarios of the 8 by 8 design.

The pixel value at each sampling location $y(s_i)$ is extracted. Due to the high number of realizations, as well as the size of the exhaustive dataset, variograms were automatically fitted by an ordinary least squares approach. Exponential models with varying variogram parameters are tested, and the model exhibiting the lowest sum of squares is kept. The variable is interpolated over the set of grid points G (40000points) using ordinary kriging. The average absolute error of prediction (MAE) is reported:

$$MAE = \frac{1}{|G|} \sum_{g \in G} |\hat{y}(s_g) - y(s_g)| \quad (8)$$

where $\hat{y}(s_g)$ is the interpolated value of the primary variable at a grid node s_g , and $y(s_g)$ its true value. Ultimately, the goal is to find an augmented sampling pattern which reduces Equation 8 by as much as possible, leading to a better reconstruction of the “true” image.

4.2 Computing the spatial roughness

The spatial roughness of the primary variable is the relative weighted squared difference in value of an a interpolated grid node from its neighboring grid nodes. Figure 2b shows the weights from the exhaustive dataset for instance. Interestingly, the river band in the Northeastern part of the image is characterized by low weights, while its embankments reflect stronger spatial variation. For each of the 150 realizations, local variations in interpolated values are computed with Equation 5 for each grid point s_g using its 4 nearest neighbors ($L/J = 4$), and parameter $\beta = 1.5$. Careful attention must be paid when selecting J , as too large of a radius may smooth out local variations. Figure 5a illustrates the interpolated map from a 8 by 8 sampling pattern, case 23, Figure 5b its corresponding kriging variance while Figure 5c maps the spatial roughness. Areas where contours of the interpolated map (Figure 5a) are far apart do not present any spatial roughness, and their weights are close to 0, while regions where contours of the interpolated map come close together exhibit higher weight values. The kriging variance and spatial roughness are slightly negatively correlated; areas close to existing samples have a lower kriging variance but characterized by higher weights. This negative correlation decreases with increasing initial sampling densities.

Figure 5d shows the weighted kriging variance which is computed by multiplying the weights with the kriging variance map (Equation 6). The weighted kriging variance map contrasts from the map of the weights in that more importance is given to location away from existing points. For instance, at location $x = 40, y = 100$ the importance of the weighted kriging variance is noticeably higher than on the map of the weights. Interestingly, note how the region east of location $x = 87, y = 167$ increases its sampling attraction after the kriging variance is combined with weights. As a consequence of combining the kriging variance, some locations loose their sampling importance, for instance in the direct vicinity of locations $x = 32, y = 167, x = 32, y = 30, x = 169, y = 112, x = 87, y = 140$. Combining the kriging variance with spatial roughness

avoids resampling at existing locations during first and second phase sampling, while simultaneously forcing to concentrate sampling efforts in areas of spatial variation. Note that we evaluated combining the kriging variance with weights calibrated from different neighborhood sizes J . As J increases, greater weighted kriging variance values are observed farther away from existing sample points as initial weights $\lambda(s_g)$ tend to be smoother and more distributed across the study region.

4.3 Second-phase sampling

Given an initial sampling set $\mathbf{M}$, the objective is to select an additional set $\mathbf{N}$ of $n = 30$ points at which the exhaustive dataset will be sampled. Instead of simply gathering more information on the spatial variation in the primary variable, our underlying goal is to gather information in those strategic locations, (a) away from existing points, and (b) where the primary variable exhibits great geographical changes. To test whether augmented designs maximizing the change in weighted kriging variance would lead to better predictions (i.e. reconstruction of the “true” image), we compare the merits of this approach to an objective maximizing the change in kriging variance. Essentially, we are seeking the objective which reduces Equation 8 the most. To find the set of points that would maximize the change in weighted kriging variance, summed over all grid points, we use a combination of a sequential greedy algorithm with a simultaneous simulated annealing. The size of the potential set $\mathbf{P}$ is equal to 40000 minus the size of the initial sampling set $\mathbf{M}$. Table 1 summarizes the size of the set $\mathbf{P}$ in the sequential and simultaneous cases. There are $[(p - m) + (p - m - 1) + \dots + (p - m - n + 1)]$ in the sequential case where m is the number of initial sample points, while there are $\binom{p - m}{n}$ solutions to the simultaneous case which is very substantial.

Sequential greedy algorithm—A dynamic greedy algorithm is implemented by selecting iteratively the point s_{m+1}^+ from $\mathbf{P}$ exhibiting the highest weighted kriging variance (surface peak). Once found, s_{m+1}^+ is added to the initial set $\mathbf{M}$, and the new objective function computed with the augmented set, $S = \{\mathbf{M} \cup s_{m+1}^+\}$. The weighted kriging variance surface is recomputed. The remaining $n - 1$ points are then added in a similar fashion. Of very interesting nature is the tendency of the algorithm to cluster additional observations in the vicinity of contours where the weighted kriging variance reaches local optima (see Figure 6a). Although the solution remains sub-optimal, the advantage of the myopic approach resides in its ability to return a very good sampling pattern S^+ in time-critical situations. As an example of the 8 by 8 design, case 23, the addition of the $n = 30$ new points yields an overall reduction in the weighted kriging variance of 13.17%

Simultaneous simulated annealing—A total of 850 iterations are performed on the set $\mathbf{N}$. Figure 6b displays the location of additional points obtained using a simultaneous addition with a greedy start for the 8 by 8 design, case 23. In comparison with Figure 6a, new samples are more evenly spaced across regions exhibiting high weights. For instance, the area nearby location $x = 35, y = 155$ receives 3 additional observations with the greedy algorithm, but only one when using simulated annealing. The same is true for the area around location $x = 90, y = 150$. The use of simulated annealing provides a reduction of 17.23% on the weighted kriging variance, or an improvement of 4.7% on the solution set obtained using a greedy algorithm. Figures 6c and d display additional sampling locations minimizing the kriging variance, using a sequential greedy algorithm and simultaneous simulated annealing, respectively. Due to the nature of the optimization criterion (Equation 4), the sample pattern of the augmented set is geometric regardless of the heuristic, although simulated annealing has the tendency to allocate additional samples towards the center of the study region. Figures 7a and b display the interpolation maps for the same initial design (8 by 8, case 23) using additional sampling

locations minimizing the kriging variance and weighted kriging variance, respectively, while Figures 7c and d summarize the reconstruction error. Figure 7e summarizes the difference between maps 7c and d. Positive areas indicate that the augmented pattern which optimizes Equation 6 leads to prediction closer to the true image, while negative regions indicate that the augmented pattern optimizing Equation 4 leads to prediction closer to the true image

4.4 Reconstruction of the true image

For comparison purposes, the exact same procedure consisting of combining greedy and simulated annealing, is applied on the initial sets to find additional locations which would maximize the change in the kriging variance, this is to say when $\alpha = 0$ in Equation 7. Once the locations of additional observations which optimize Equation 7 (as well as when $\alpha = 0$) have been determined, the pixel value at each new observation $y(s_{m+1} \dots s_{m+n})$ is extracted and variograms automatically fitted by an ordinary least squares approach. For each of the 150 realizations, the augmented set is used to interpolate the variable Y over the grid points by ordinary kriging. The average absolute error of prediction (MAE) is then reported. To assess the quality of the designs maximizing the change in weighted kriging variance as well as kriging variance alone, 30 Monte-Carlo simulations for each of the 150 sampling realizations are carried out. Alternatively stated, the simulations consist of augmenting the initial set by randomly allocating m additional samples, repeating this 30-times for each sampling realization. The average mean absolute error for 30 random simulations $\overline{MAE}_R$, is reported.

$$\overline{MAE}_R = \frac{\sum_{R=1}^{30} \frac{1}{|G|} \sum_{g \in G} |\hat{y}(s_g) - y(s_g)|}{30} \quad (9)$$

Figure 8 graphs the mean reduction in MAE for the different sampling strategies and patterns of varying initial sampling densities.

Regardless of the strategy adopted to allocate additional samples, the overall reduction in MAE is the greatest for sampling patterns of low density. This reduction decreases as the initial sampling pattern becomes larger, due to the relative size of the augmented set in comparison to the initial set. For denser sampling patterns (12 by 12 and 14 by 14), differences in MAE values among the 3 strategies are marginal. Sampling designs augmented to maximize the change in weighted kriging variance strongly contrast with augmented designs maximizing the kriging variance alone in that they lead to lower levels of MAE values, which is very significant. This difference attenuates as initial sampling densities increase however. When a comparison is made between designs augmented with randomly allocated observations and designs reducing the kriging variance, there is marginal reduction in MAE for the last 2 patterns : 12 $\times$ 12 and 14 $\times$ 14.

Figures 9a—f map the average benefits of using one strategy over another for the 6 different sampling scenarios. Darker regions indicate positive differences in MAE. Alternatively stated, these regions corresponds to areas where the MAE value is smaller when using a strategy maximizing the change in weighted kriging variance than when using a strategy maximizing the change in kriging variance alone. For low-density designs, the geographical variation in gains follow closely the spatial pattern of the exhaustive image (figure 2b). As the sampling pattern density increases, the patches of gains and losses loose intensity and the local benefits of one strategy over another are less apparent. As attested by declining Moran's I values (Moran 1950), spatial autocorrelation in the gains decrease with increasing sampling densities.

5 Concluding remarks

In this paper, we have suggested a new criterion for the second-phase spatial sampling problem which accounts for non-stationarity. The kriging variance is weighted by the spatial variation of the primary variable. This approach guarantees that additional samples will be collected away from existing ones, but also in areas characterized by strong spatial roughness, or where contours from the interpolated variable come close together. The use of a filter-process is suggested, reflecting the difference in interpolated value of a grid point from its neighbors. The greater the variation in a region, the more likely that location will be selected during a second stage sampling.

The objective function—which tries to allocate new samples to maximize the change in weighted kriging variance—is non-linear and calls for heuristic techniques. In this paper, a combination of a sequential greedy (myopic) algorithm with a simultaneous simulated annealing using multiple-locations as perturbations is suggested. Although the greedy algorithm is quick at finding a sub-optimal solution, by nature it often gets trapped in a local optimum while simulated annealing escapes to find the optimal solution S^* .

A case study illustrated the methodology using an exhaustive remote sensing scene. Augmented designs optimizing Equation 7 significantly contrasted with designs minimizing the kriging variance alone as the MAE was consistently lower. However, this disparity lessened as the initial sampling density increased. Interestingly enough, augmented sets minimizing the kriging variance provided MAE reductions marginally lower than random monte-carlo simulations. The case study empirically reveals that a lower number of additional samples, optimizing the weighted kriging variance, is necessary to achieve a similar MAE reduction level that would be reached using augmented samples that minimize the kriging variance alone.

As a methodological improvement for future research, another objective function that uses locally determined variogram models to obtain kriging variances should be considered as they reflect non-stationarity (Haas 1990). An alternative is to rescale the kriging variances according to a locally determined variance from the data. It is suggested to postulating a different model of spatial variation with non-constant variance. Additionally, the weights λ in the objective function Equation 5 could be modified by applying a local Moran's I . Along these lines, it would be worthwhile to investigate how varying values of β and neighborhood sizes J affect the location of additional samples as well as the quality of the solution obtained. Secondly, we have assumed that the primary variable Y behaved isotropically, which means that the spatial variability is direction-dependent. Anisotropy should be accounted for, and sampling intensity increased in directions of minimum continuity. Finally, additional improvements are recommended in applying stochastic optimization techniques such as genetic algorithms or tabu search in the simultaneous addition strategy (Michalewicz and Fogel 2000).

Acknowledgments

We acknowledge helpful comments and guidance given by Dr. Peter Rogerson (Geography Department, University at Buffalo), and we also thank three anonymous reviewers, whose suggestions helped us improving this manuscript. The research by the second author was funded by grant 1R43CA135814-01 from the National Cancer Institute. The views stated in this publication are those of the author and do not necessarily represent the official views of the NCI.

References

- Armstrong, M. Is research in mining geostats as dead as dodo?. In: Dimitrakopoulos, R., editor. *Geostatistics for the Next Century*. Kluwer Academic Publisher; Dordrecht: 1994. p. 303-312.

- Aspie D, Barnes RJ. Infill-sampling design and the cost of classification errors. *Mathematical Geology* 1990;22:915–932.
- Bailey A, Gatrell T. *Interactive Spatial Data Analysis*. Longman. 1995
- Bellhouse DR. Optimal designs for sampling in two dimensions. *Biometrika* 1977;64:605–611.
- Berry, BJL.; Baker, AM. Geographic sampling. In: Berry, BJL.; Marble, DF., editors. *Spatial analysis: a reader in statistical geography*. Englewood Cliffs, N.J.: Prentice-Hall; 1968. p. 91-100.
- Bogaert P, Russo D. Optimal sampling design for the estimation of the variogram based on a least squares approach. *Water Resources Research* 1999;35(4):1275–1289.
- Bueso MC, Augulo JM, Cruz-Sanjulian J, Garcia-Arostegui JL. Optimal Spatial Sampling Design in a Multivariate Framework. *Mathematical Geology* 1999;31:507–525.
- Brus D, de Gruijter J. Random sampling or geostatistical modeling? Choosing between design-based and model-based sampling strategies for soil. *Geoderma* 1997;80:1–59.
- Brus, D.; de Gruijter, J.; van Groenigen, JW. Designing spatial coverage samples using the k-means clustering algorithm. In: Lagarchie, P.; McBratney, AB.; Voltz, M., editors. *Digital Soil Mapping: An Introductory Perspective*. Elsevier; Amsterdam: 2007.
- Brus D, Heuvelink GBM. Optimization of sample patterns for universal kriging of environmental variables. *Geoderma* 2007;138:86–95.
- Christakos G, Killam BR. Sampling design for classifying contaminant level using annealing search algorithms. *Water Resources Research* 1993;29:4063–4076.
- Christakos G, Olea RA. Sampling design for spatially distributed hydrogeologic and environmental processes. *Advanced Water Resources* 1992;15:219–237.
- Cochran, WG. *Sampling techniques*. Second Edition. Wiley; New York, USA: 1963. p. 413
- Cox LA. Adaptive spatial sampling of contaminated soil. *Risk Analysis* 1999;19:1059–1069. [PubMed: 10765446]
- Cressie, N. *Statistics for Spatial Data*. Wiley; New York, USA: 1991. p. 900
- De Gruijter, J.; Brus, DJ.; Bierkens, MFP.; Kotters, M. *Sampling for Natural Resource Monitoring*. Springer; 2006. p. 332
- Delmelle, E. Ph.D. dissertation, SUNY at Buffalo. 2005. *Optimizing Second-Phase Spatial Sampling Using Auxiliary Information*.
- Delmelle, E. Spatial sampling. In: Rogerson, P.; Fotheringham, S., editors. *The SAGE Handbook of Spatial Analysis*. Sage Publication; 2009. p. 528
- Deutsch, CV.; Journel, AG. *Gslib: Geostatistical Software Library and User's Guide*. 2. Oxford University Press; 1997. p. 369
- Di Zio S, Fontanella L, Ippoliti L. Optimal spatial sampling schemes for environmental surveys. *Environmental and Ecological Statistics* 2004;11:397–414.
- Ferri M, Piccioni M. Optimal selection of statistical units. *Computational Statistics and Data Analysis* 1992;13:47–61.
- Goovaerts P. Geostatistical modelling of spatial uncertainty using *p*-field simulation with conditional probability fields. *International Journal of Geographical Information Science* 2002;16(2):167–178.
- Grötschel, M.; Lovász, L. Combinatorial optimization. In: Graham, RL.; Grötschel, M.; Lovász, L., editors. *Handbook of Combinatorics*. Vol. 2. Elsevier; Amsterdam, The Netherlands: 1995. p. 1541-1579.
- Haas TC. Kriging and automated variogram modeling within a moving window. *Atmospheric Environment* 1990;24:1759–1769.
- Haining, RP. *Sampling spatial populations*. In: Haining, RP., editor. *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press; 1990.
- Hengl T, Rossiter DG, Stein A. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Australian Journal of Soil Research* 2003;41:1403–1422.
- Hsia CK, Juang KW, Lee DY. Estimating the second-stage sample size and the most probable number of hot spots from a first stage sample of heavy-metal contaminated soil. *Geoderma* 2000;95:73–88.
- Kirkpatrick S, Gelatt CD, Vecchi Mp. Optimization by simulated annealing. *Science* 1983;220:671–680. [PubMed: 17813860]

- Lajaunie, C.; Wackernagel, H.; Thiry, L.; Grzebyk, M. Sampling multiphase noise exposure time series. In: Soares, A.; Gomez-Hernandez, J.; Froidevaux, R., editors. *GeoENV II - Geostatistics for Environmental Applications*. Kluwer Academic Publishers; Dordrecht: 1999. p. 101-112.
- Lark RM. Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. *Geoderma* 2002;105:49–80.
- Michalewicz, Z.; Fogel, D. *How to Solve It: Modern Heuristics*. Springer; 2000. p. 467
- Moran PAP. Notes on continuous stochastic phenomena. *Biometrika* 1950;37:17–23. [PubMed: 15420245]
- Muller, W. *Collecting Spatial Data: Optimal Design of Experiments for Random Fields*. Heidelberg: Physica-Verlag; 1998.
- Pardo-Igúzquiza E. Optimal; selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *Journal of Hydrology* 1998;210:206–220.
- Rogerson PA, Delmelle EM, Batta R, Akella MR, Blatt A, Wilson G. Optimal sampling design for variables with varying spatial importance. *Geographical Analysis* 2004;36:177–194.
- Royle A, Nychka D. An algorithm for the construction of spatial coverage designs with implementation in SPLUS. *Computers and Geosciences* 1998;24:479–488.
- Sacks, J.; Schiller, S. Spatial designs. In: Gupta, SS.; Berger, JO., editors. *Statistical decision theory and related topics IV*. Springer-Verlag; New-York, USA: 1988. p. 385-399.
- Stehman, SV.; Overton, SW. Spatial sampling. In: Arlinghaus, S., editor. *Practical handbook of spatial statistics*. CRC Press; Boca Raton FL: 1996. p. 31-64.
- Thompson, SK. *Sampling*. Second Edition. Wiley; 2002. p. 367
- Van Groenigen JW, Stein A. Constrained optimization of spatial sampling using continuous simulated annealing. *Journal of Environmental Quality* 1998;27:1078–1086.
- Van Groenigen JW, Siderius W, Stein A. Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma* 1999;87:239–259.
- Van Groenigen JW, Pieters G, Stein A. Optimizing spatial sampling for multivariate contamination in urban areas. *Environmetrics* 2000;11:227–244.
- Yfantis EA, Flatman GT, Beha JV. Efficiency of kriging estimation for square, triangular and hexagonal grids. *Mathematical Geology* 1987;19:183–205.

Figure 1.
A 3×3 moving window: a circle is passed around a grid node within a specific distance.

Figure 2. Reference SPOT image in *a* and its corresponding weights in *b* using parameters $LJJ = 4$ and $\beta = 1.5$.

Figure 3.
Initial sampling: combinations of systematic and nested sampling patterns of varying densities.
Above each figure, the title indicates the number of rows and columns and how many systematic sample points are randomly selected for additional, nested sampling.

Figure 4.
Impact of the coordinates of the first sample on the location of the remaining sampling pattern (*8 by 8 design*). A total of 8 samples are then selected randomly and 4 additional samples randomly clustered around them.

Figure 5. Interpolation of initial sample points for the 8 by 8 design, case 23 (see Figure 4(b)) using ordinary kriging with nugget $a = 0$, sill $\sigma^2 = 140$ and range $r = 38$ and associated kriging variance (b). Figure (c) represents the relative weight from Equation 5, while figure (d) is the weighted kriging variance using Equation 7.

Figure 6.

Figures (a) and (b) display the locations of the augmented set minimizing the weighted kriging variance using a greedy and a simulated annealing algorithm respectively. Similarly, figures (c) and (d) display the locations of the augmented set minimizing the kriging variance using greedy and simulated annealing respectively.

Figure 7. Interpolation of the augmented sample patterns (in (a) the pattern which optimizes Equation 4 and in (b) the pattern optimizing Equation 6) for the *8 by 8 design*, case 23 (see Figure 4(b)) using ordinary kriging. Figures (c) and (d) represent the error from the true image following the interpolation with the augmented sampling pattern optimizing the kriging variance and weighted kriging variance, respectively. Note the mean absolute error $MAE = 5.88$ and 5.77 in (c) and (d), respectively. Figure (e) displays the difference in gain (decrease in the Mean Absolute Error). Regions of positive values indicate areas where sampling patterns optimizing Equation 7 are closer to the true image.

Figure 8.

Decrease in the mean absolute error (MAE) from the initial set for varying sampling densities. On the x -axis, the number before the parenthesis is the size of the initial sampling set, while the number in parenthesis is the cardinality of the augmented set. KV and WKV stand for kriging variance and weighted kriging variance, respectively.

Figure 9. Difference in gain (decrease in the Mean Absolute Error) by using sampling pattern which optimize the weighted kriging variance from designs minimizing the kriging variance. Regions of positive values indicate areas where designs optimizing Equation 7 are closer to the true image. Moran's I value reaches .7373 for figure *a* and .5951 for figure *f*.

Table 1

Number of sampling permutations according to the design density. The set $\mathbf{M}$ is the initial sampling set, $\mathbf{M} \cup \mathbf{N}$ the augmented set with $|\mathbf{N}| = 30$, while $\mathbf{P}$ is the set of potential locations.

Design	$ \mathbf{M} $	$ \mathbf{M} \cup \mathbf{N} $	$ \mathbf{P} $ sequential	$ \mathbf{P} $ simultaneous
4 by 4	32(16)	62	1158666	4.20E+105
6 by 6	60(36)	90	1157854	4.11E+105
8 by 8	96(64)	126	1156810	4.00E+105
10 by 10	140(100)	170	1155534	3.87E+105
12 by 12	192(144)	222	1154026	3.72E+105
14 by 14	252(196)	282	1152286	3.56E+105