1

Model-based criteria and for second-phase spatial sampling

Second-phase spatial sampling is the process of collecting additional measurements of a variable of interest within a two-dimensional framework. Different criteria exist to determine the ”optimal” location of these new observations. The inherent goal is to capitalize on spatial correlation information gathered during the first-phase sampling phase to improve spatial estimation. Several second-phase sampling objectives are discussed in this chapter. These are generally non-linear and require a method to find the set of additional locations among candidate locations, which will optimize a given sampling objective. In this chapter, the merits of simulated annealing (SA) are discussed, and an example using SA for the minimization of the is given with the the Austrian dataset.

1.1 Introduction

In sampling, the and time available at hand usually determines the number of samples that can be collected, and careful attention must be paid to design an appropriate of observations. It is generally recognized that a sampling configuration which will minimize the variance associated with the estimation is more desirable. For one-dimensional problems ($\mathbb{IR}$), Cochran (1946) has suggested that a *stratified random* design will always be more efficient than a *random* design to provide an unbiased estimate of the sampling variance. Cochran's initial work was later extended to sampling designs ($\mathbb{IR}^2$) by Quenouille (1949) and Das (1950) who used a linear model. With non-linear autocorrelation function however, *systematic* sampling is the most efficient technique, followed by *stratified random* sampling and *random* sampling (Zubrzycki 1958). Those results were later confirmed in a series of articles by Matérn (1960), Berry and Baker (1968), Bellehouse (1977), Ripley (1981) and Iachan (1985). In spatial sampling, measurements at specific areas must be acquired instead of trying to obtain information at every possible location (see, e.g.; Cochran 1963; Dalton *et al.* 1975; Ripley 1981; Arbia 1989; Haining 1990; Hedayat and Sinha 1991; Cressie 1991; Stehman and Overton 1996; Mueller 1998 and Thompson 2002 for various summaries). Although a full inventory will reflect the variation of the variable of interest, this process is rather time-consuming and constrained by the available . On the other hand, sparse sampling is less costly, and the of the variable may not be captured properly (Berry and Baker 1968). This chapter is concerned with (and) spatial sampling designs to acquire (and complement, respectively) information on the of a variable, for instance in the form of a map, or as a summary measure, which highlights the scales of variation.

In spatial sampling, the location of the samples is critical and may be influenced by the of the variable: for phenomena with little variation, samples can be spaced more evenly without the risk of not detecting variations at smaller ranges. Unfortunately, this variation must be estimated, and an objective is to design sampling patterns which will capture a maximum amount of information. If we in some areas, the of the variable of

interest is not captured. In spatially autocorrelated fields, oversampling may result in redundant data (Griffith 2005).

Once samples of the primary variable have been collected, it is possible to augment the initial set by collecting additional measurements at other locations, a method known as sampling (Cressie 1991, Muller 1998, van Groenigen and Stein 1998, and de Gruitjer *et al.* 2006). The inherent objective of a second-phase sampling strategy is to improve spatial prediction with added information. Following a first sampling phase, spatial covariance structure is summarized through a variogam function, and the computed at each location. Generally, are gathered away from existing points, that is where the is large. However, when the process under study is not stationary, sampling efforts should be directed in those strategic locations exhibiting strong locally (Rogerson *et al.* 2004, Delmelle and Goovaerts 2009). In this chapter, I discuss several sampling criteria are discussed. Whether the function accounts primarily for the minimization of the , or increase sampling efforts in areas of strong , the optimization problem is non-linear, and calls for robust methods, for instance . This chapter is structured as follows. **First**, geometric and geostatistical are presented in sections 1.2 and 1.3. **Second**, strategies to support a sampling design are discussed. **Third**, the framework to implement a procedure is presented in section 1.4, applied to a sampling design. This arises due to the non-linearity of the problem and that a complete enumeration of all possible solutions is not feasible in a timely manner. This is illustrated in section 1.5 with an application to the Austrian dataset.

1.2 Geometric and geostatistical designs

There exists different sampling schemes for the purpose of two-dimensional sampling with no prior information available (these are generally termed random, systematic or stratified approach). Assuming that a limited number of samples m are allocated in a study area denoted $\mathfrak{D}$, the spatial variable Z is then measured on m supports, $\{z(\mathbf{s}_i)|i = 1, 2, \dots m\}$. In a *simple random sampling* design, m sample points in $\mathfrak{D}$ are selected randomly (King 1969, Ripley 1981), and the selection of a sample should not influence the selection of any other one. Practically, the coordinates of the sample the pair $\{x_i, y_i\}$ are randomly drawn on the interval [(minX, maxX),(minY, maxY)]. For boundaries (or non rectangular region), a point can still be drawn at random in the minimum bounding rectangle, followed by a `inside` algorithm to determine whether the point will fall within the study region. In those situations, there is still a risk to experience edge effects. In a *systematic sampling* design, the population of interest is divided into m intervals of equal size (the same applies for non-squared areas). The first element is randomly or purposively chosen within the first interval $\frac{L}{\sqrt{m}}$ (as long that interval is within the boundary of the study region), starting at the origin, while the location of the remaining $m - 1$ elements are aligned regularly by the size of the interval.

$$x_i = x_1 + (i-1)\frac{L}{\sqrt{m}} \qquad y_i = y_1 + (j-1)\frac{L}{\sqrt{m}} \qquad \forall i, j = 1, \dots, \sqrt{m}. \tag{1.1}$$

The most common regular geometric configurations are the equilateral triangular grid, the rectangular (square) grid, and the hexagonal one (Cressie, 1991). The advantages of a systematic design lies in a good coverage of the observations. This design presents two inconveniences, however:

(i) the distribution of distances between points of $\mathfrak{D}$ is not represented fairly since many pairs of points are separated by the same distance,
(ii) there is a danger that the spatial process shows evidence of recurring that will remain uncaptured. This is a critical issue of systematic design that coincide in frequency with a regular pattern in the landscape (Griffith and Amrhein, 1997; Overton et al., 1993). A *systematic random method* approach, which combines both systematic and random procedures (Dalton et al. 1975, King 1969) can prevent the latter. *Stratified random* partitions the population (or $\mathfrak{D}$) into non-overlapping strata, and for each stratum, a specific set of samples

is collected, for instance a greater number of samples may be collected in a geographic area due to greater population.

Efficiency of spatial sampling designs.

Different criteria have been proposed to evaluate the merits of sampling designs (Muller 1988). An example is an estimator for the global mean that estimates the accuracy of the global mean estimates. Careful attention must be paid to the fact that some designs may be optimal for one criterion, yet not very efficient for other criteria. Another example is the minimization of the of the true surface (as discussed later). A design that leads to an accurate estimation of the global mean $z_{\mathfrak{D}}$ is desirable:

$$z_{\mathfrak{D}} = \frac{1}{[\mathfrak{D}]} \int_{\mathfrak{D}} z(\mathbf{s}) d\mathbf{s} \tag{1.2}$$

In this respect, it is desirable to select a configuration that minimizes the prediction error of $z_{\mathfrak{D}}$ for a given estimator, for instance the arithmetic mean (Equation 1.3):

$$\overline{z} = \frac{1}{m} \sum_{i=1}^{m} z(\mathbf{s}_i) \tag{1.3}$$

Efficiency is calculated for all possible realizations of the variable z by $var_{\xi}[z^*_{\mathfrak{D}} - z_{\mathfrak{D}}]$. This is generally implemented using σ^2_k, which is the geostatistical , defined in the next section. In terms of the sampling variance, *stratified random sampling* is at least always equally or more accurate than *random sampling*; its relative efficiency is a monotone increasing function of sample size.

Sampling spatial variables in a geostatistical context.

Spatial information from closeby measurements is generally not accounted for in classical sampling theory. Geostatistics describes the spatial continuity that is an essential feature of many natural phenomena (see, e.g., Matérn, 1960; Whittle, 1963). The variable z is modeled as a random process that can take a series of outcome values, according to some probability distribution (Goovaerts, 1997). Central to geostatistics is , an interpolation method which predicts the value of z at unsampled locations (usually on a set G of grid points $\{\mathbf{s}_g | g = 1, 2 \dots G\}$, keeping the (also called) to a minimum (Isaaks and Srivastava, 1989). is based on a , which summarizes the variance of values separated by a particular distance lag (h) is defined:

$$\widehat{\gamma}(h) = \frac{1}{2d(h)} \sum_{|\mathbf{s}_i - \mathbf{s}_j| = h} \Big(z(\mathbf{s}_i) - z(\mathbf{s}_j) \Big)^2, \tag{1.4}$$

where $d(h)$ is the number of pairs of points for a given lag value, and $z(\mathbf{s}_i)$ is the measured attribute value at location $\mathbf{s}_i$, a the and σ^2 the sill, where $\widehat{\gamma}(h)$ levels out (Cressie, 1991). Once the lag distance exceeds the range r, there is no spatial dependence between sample sites. The interpolated, kriged value at a location $\mathbf{s}$ in $\mathfrak{D}$ is a weighted mean of surrounding values; each value is weighted according to the model:

$$\widehat{z}(\mathbf{s}) = \sum_{i=1}^{I} w_i(\mathbf{s}) z(\mathbf{s}_i), \tag{1.5}$$

with I the set of neighboring points used to estimate the interpolated value at location $\mathbf{s}$, and $w_i(\mathbf{s})$ is the weight associated with each neighboring point. It is critical to have a wide range of distances in order to estimate the accurately, that is a good coverage of samples is desirable (Russo 1984, Van Groenigen *et al.* 1999). But this configuration should be supplemented by to guarantee that a few pairs of points are separated by very small distances, critical to estimate the . Webster and Oliver (1993) have indicated a minimum of $m = 150$ samples over the study area is necessary to estimate the , but this is also influenced by the over $\mathfrak{D}$ (phenomenon with less variation may need less samples). The strength of the is also partly dependent on the

number of pairs of points available within each distance class. The (WM) criterion attempts at reproducing an a priori defined ideal distribution of pairs of points for estimating the covariogram:

$$J_{w/m}(S) = \mathbf{a}\sum_{i=1}^{K}\mathbf{w}_i\left(\xi_i^* - \xi_i\right)^2 + \mathbf{b}\sum_{i=1}^{K}\sigma(m_i), \tag{1.6}$$

$$\sum_{i=1}^{K}\xi_i^* = \sum_{i=1}^{K}\xi_i^* = \frac{m\cdot(m-1)}{2}, \tag{1.7}$$

where i denotes a given lag class of the covariogram, K represents the total number of classes, the parameters **a**, **b**, and $\mathbf{w}_i$ are user-defined weights. The term ξ_i^* is a prespecified number of point-pairs for the i^{th}class, ξ_i is the actual number of distances within that class, and $\sigma(m_i)$ is the standard deviation from the median of the distance lag class (Warrick and Myers 1987 and Delmelle 2009). Equation 1.7 expresses the total number of possible distance pairs, given the number of samples. Another similar criterion suggested in the literature is the ***M**inimization of the **M**ean of the **S**hortest **D**istances*, requiring sampling points to be spread evenly over the study region (Van Groenigen *et al.* 1999).

Sampling designs minimizing the

The quantifies the prediction uncertainty at a particular location in space. This uncertainty is minimal at existing sampling points and increases with distance to the nearest samples. One common criterion is to design a sampling configuration, which minimizes the over $\mathfrak{D}$, with a known, a priori (or estimated) structure (Van Groenigen *et al.*, 1999). Equation 1.8 formulates the at a location **s**, where $\mathbf{C}_M^{-1}$ is the inverse of the $\mathbf{C}_M$, based on the covariogram function (Bailey and Gatrell, 1995). M denotes the set of initial samples and has cardinality m. The term **c** is a column vector and $\mathbf{c}^T$ the corresponding row vector, as given in Equation 1.10.

$$\sigma_k^2(\mathbf{s}) = \sigma^2 - \mathbf{c}^T(\mathbf{s})\cdot\mathbf{C}_M^{-1}\cdot\mathbf{c}(\mathbf{s}) \tag{1.8}$$

$$\mathbf{C}_M = \begin{bmatrix} \sigma^2 & C_{1,2} & \cdots & C_{1,m} \\ C_{2,1} & \sigma^2 & \cdots & C_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ C_{m,1} & C_{m,2} & \cdots & \sigma^2 \end{bmatrix} \tag{1.9}$$

$$\mathbf{c} = \begin{bmatrix} \sigma^2 \\ C_{21} \\ \vdots \\ C_{m1} \end{bmatrix}, \qquad \mathbf{c}^T = \begin{bmatrix} \sigma^2 & C_{12} & \cdots & C_{1m} \end{bmatrix} \tag{1.10}$$

Computationally, the study area $\mathfrak{D}$ is and the summed over all grid points $\mathbf{s}_g$. Alternatively, it can be computed at each sample candidate location. The (average) becomes (Delmelle 2009):

$$AKV = \frac{1}{G}\sum_{g\epsilon G}\sigma_k^2(\mathbf{s}_g) \tag{1.11}$$

The only requirement to calculate the is to have an initial covariogram and the locations of the m initial sample points. It then depends solely on the and configuration of the observations (Cressie, 1991). In a *first stage*, initial measurements of the variable are collected to calibrate the . However, this step may not always be required when a reliable *a priori* estimation of the is available. Designs minimizing the tend to spread samples evenly in the study region. An example optimizing initial samples for the minimization of the index is given in section 1.5.

1.3 Augmented designs: sampling

Second-phase sampling occur when there is a need to go out in the field to gather additional samples about the variable of interest. One important aspect is to capitalize on the spatial covariance information obtained during a first sampling phase, for instance under the form of a indexvariogram. Based on this covariance structure, the is computed at each grid node. The objective consists of locating those additional samples strategically to maximize the change in , that is locating those samples away from existing ones.

1.3.1 Additional Sampling Schemes to Minimize the kriging Variance

Our first objective $J[S]$ is to select a set of n additional points to our exisiting set of m samples, which will maximize the change in by as much as possible. This process can be thought as a simulation of what the change in is expected to be, without having to collect additional points, assuming the structure would remain constant (Burgess, Webster and McBratney 1981 as well as Cressie 1991). Specifically:

$$\underbrace{\text{MAXIMIZE}}_{\{\mathbf{s}_{m+1},\ldots,\mathbf{s}_{m+n}\}} \quad J[S] = \frac{1}{\lfloor G \rfloor} \sum_{g \epsilon G} \left(\sigma_k^{\text{old}}(\mathbf{s}_g)\right)^2 - \left(\sigma_k^{\text{new}}(\mathbf{s}_g)\right)^2, \tag{1.12}$$

with S denotes the sampling scheme. Objective 1.12 aims to collect new samples to reduce the or uncertainty by as much as possible. Equation 1.13 formulates the change in $\triangle\sigma_k^2$ over all grid points $\mathbf{s}_g$, with the addition of a new set N of size n. The change $\triangle\sigma_k^2$ is the difference between the of the initial set σ_k^{old} with the σ_k^{new} of the augmented set:

$$\triangle\sigma_k^2 = \frac{1}{G}\left[\sum_{g \epsilon G} \sigma_k^{\text{old}}(\mathbf{s}_g) - \sum_{g \epsilon G} \sigma_k^{\text{new}}(\mathbf{s}_g)\right] \tag{1.13}$$

$$\sigma_k^{\text{old}}(\mathbf{s}_g) = \sigma^2 - \underbrace{\mathbf{c}(\mathbf{s}_g)}_{[1,m]} \times \underbrace{\mathbf{C}^{-1}}_{[m]} \times \underbrace{\mathbf{c}^T(\mathbf{s}_g)}_{[m,1]} \tag{1.14}$$

$$\sigma_k^{\text{new}}(\mathbf{s}_g) = \sigma^2 - \underbrace{\mathbf{c}(\mathbf{s}_g)}_{[1,m+n]} \times \underbrace{\mathbf{C}^{-1}}_{[m+n]} \times \underbrace{\mathbf{c}^T(\mathbf{s}_g)}_{[m+n,1]}. \tag{1.15}$$

The objective function (Equation 1.16) is to find the optimal set S^* containing $m+n$ points that will maximize this change in (Christakos *et al.* 1992, Van Groenigen *et al.* 1999, Rogerson *et al.* 2004), where S is a specific sampling scheme.

$$\underbrace{\text{MAXIMIZE}}_{\{\mathbf{s}_{m+1},\ldots,\mathbf{s}_{m+n}\}} \quad J(S) = \frac{1}{G} \sum_{g \epsilon G} \triangle\sigma_k^2(\mathbf{s}_g; S) \tag{1.16}$$

The set of new points is selected from the set of candidate location P. With a total of $\binom{p}{n}$ possible sampling combinations, it is too time-consuming to find the optimal set using combinatorics and requires methods as defined in section 1.4. The initial sampling problem which minimizes the (or maximizes the change in) is a simplification of equation 1.16 with the set $n = 0$.

1.3.2 A weighted approach

Many authors have advocated the use of the as a measure of uncertainty. It can be misused as a measure of reliability of the kriging estimate, as noted by several authors (Deutsch and Journel 1992; Armstrong 1994). The rationale for this criticism is that the is merely a function of the sample pattern, sample density, the numbers of samples and their covariance structure (Delmelle and Goovaerts 2009). The assumes that the

errors are independent of each other, a situation referred to as . This means that the process is , an assumption violated in practice. Stationarity entails that the variation of the primary variable between two points remains similar at different locations in space, as long their separating distance remains unchanged. Mathematically, if we measure the difference in absolute value between two points $\mathbf{s}_i$ and $\mathbf{s}_j$ separated by a distance $d(\mathbf{s}_i, \mathbf{s}_j)$, and if we measure that difference again at two other points $\mathbf{s}_k$ and $\mathbf{s}_l$ separated by a similar distance $d(\mathbf{s}_k, \mathbf{s}_l)$, the results should be similar. In other words, $|y(\mathbf{s}_i) - y(\mathbf{s}_j)| \approx |y(\mathbf{s}_k) - y(\mathbf{s}_l)|$ for $d(\mathbf{s}_i, \mathbf{s}_j) = d(\mathbf{s}_k, \mathbf{s}_l)$. Bailey and Gatrell (1995) mentions that it is a matter of judgment whether to assume some stationarity in the variable of interest. Note that the letter $\mathbf{s}$ denotes a row vector containing the $\{x, y\}$ coordinate of the point. Figure 1.1 illustrates the problem in one dimensional, where ten hypothetical temperature values have been randomly simulated (after Goovaerts 1997). The figure on the right depicts the interpolated temperature values calculated using an exponential fitting model. From this graph it is possible to determine the values at locations $\mathbf{s}_1$ and $\mathbf{s}_2$. As Goovaerts points out, the variation in the close neighborhood of location $\mathbf{s}_1$ is much greater than $\mathbf{s}_2$, because it is surrounded by a very low and a very high value. However, the is similar at both $\mathbf{s}_1$ and $\mathbf{s}_2$, since their neighbors are at an equal distance. If a sample point is added to the initial set of 10 points, it preferably should be point $\mathbf{s}_1$, as it exhibits a greater amount of around its location than $\mathbf{s}_2$. Therefore, one issue pertains to choosing a good indicator to quantify the spatial uncertainty at a data point. The can certainly not be used as the sole indicator, but should be combined with other sampling criteria. This example

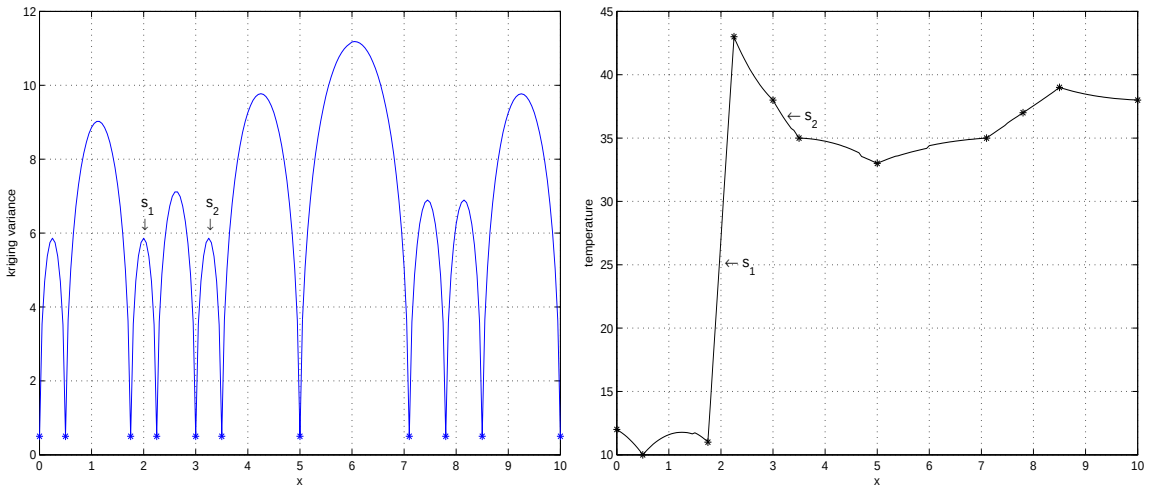

Figure 1.1 Example of nonstationarity. Points $\mathbf{s}_1$ and $\mathbf{s}_2$ have the same . However, the local variance at $\mathbf{s}_1$ is much greater than at $\mathbf{s}_2$ because there is much greater temperature variation among its neighboring points. After Goovaerts (1997).

illustrates some of the shortcomings of using the as the sole optimization criterion. Rogerson et al. (2004), Brus and Heuvelink (2007) and recently Delmelle and Goovaerts (2009) have proposed different criterion alternatives. For instance, Delmelle and Goovaerts (2009) suggest to weight the where the weights reflect the spatial variation. Their application to an exhaustive dataset lead to better reconstruction of the image.

There are two different approaches to formulate the sampling problem. Either the problem is defined as a single-weighted objective with no constraints, where the weights reflect the sampling objectives, or as a single objective where the weights become constraints. The single-weighted approach was suggested by Cressie (1991) and has been applied by Van Groenigen *et* al. (1999) and Rogerson *et al.* (2004) to weight the by a suitable weighting function $w(\cdot)$. The importance of a location to be sampled is represented by a weight $w(\mathbf{s})$, that is location-specific. The objective is to find the optimal sampling scheme S^* containing $m + n$

points that will maximize this change in weighted :

$$\underbrace{\text{MAXIMIZE}}_{\{\mathbf{s}_{m+1},\ldots,\mathbf{s}_{m+n}\}} \quad J(S) = \frac{1}{G}\sum_{g\epsilon G} w(\mathbf{s}_g)\cdot \triangle\sigma_k^2(\mathbf{s}_g;S) \tag{1.17}$$

Computational implementation

The inverse of the covariance matrix $\mathbf{C}_{M\cup N}$ is necessary to compute the . Since the initial matrix $\mathbf{C}_M$ (Equation 1.9) has been augmented by n rows and n columns. However, it is possible to calculate $\mathbf{C}_{M\cup N}^{-1}$ without having to invert the entire matrix. Consider in Equation 1.18 the augmented matrix $\mathbf{C}_{M\cup N}$

$$\mathbf{C}_{M\cup N} = \begin{bmatrix} \mathbf{C}_M & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C}_N \end{bmatrix} = \begin{bmatrix} \sigma^2 & \cdots & C_{1,m} & C_{1,m+1} & \cdots & C_{1,m+n} \\ \vdots & \ddots & \vdots & \cdots & \cdots & \cdots \\ C_{m,1} & \cdots & \sigma^2 & C_{m,m+1} & \cdots & C_{m,m+n} \\ C_{m+1,1} & \cdots & C_{m+1,m} & \sigma^2 & \cdots & C_{m+1,m+n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{m+n,1} & \cdots & C_{m+n,m} & C_{m+n,m+1} & \cdots & \sigma^2 \end{bmatrix} \tag{1.18}$$

where $\mathbf{B}$ is the covariance matrix between initial and new added points, $\mathbf{B}^T$ is the transpose of $\mathbf{B}$ ad $\mathbf{C}_N$ is the covariance matrix among the new, added samples. A new matrix of size $n \times n$ is introduced in Equation 1.19 for simplification purposes and called the $\mathbf{P}$-matrix.

$$\mathbf{P} = \left[\mathbf{C}_N - (\mathbf{B}^T \cdot \mathbf{C}_M^{-1} \cdot \mathbf{B})\right] \tag{1.19}$$

Keeping in mind this simplification of matrix notification, the inverse of the augmented matrix $\mathbf{C}_{M\cup N}$ can be formulated as (Horn and Johnson 1985):

$$\mathbf{C}_{M\cup N}^{-1} = \begin{bmatrix} \mathbf{C}_M^{-1} & \mathbf{0} \\ \mathbf{0}^T & 0 \end{bmatrix} + \begin{bmatrix} -\mathbf{C}_M^{-1}\cdot\mathbf{B} \\ I \end{bmatrix} \cdot \mathbf{P}^{-1} \cdot \begin{bmatrix} -\mathbf{C}_M^{-1}\cdot\mathbf{B} \\ I \end{bmatrix}^T \tag{1.20}$$

where I is an identity matrix of size $n \times n$. The $\mathbf{0}$-matrix is filled with zero elements. $\mathbf{C}_{M\cup N}^{-1}$ is a function of $\mathbf{C}_M^{-1}$ that remains constant, regardless of which new samples and how many of those are added. Only the matrices $\mathbf{P}$ and $\mathbf{B}$ vary when new points are added.

1.4 A simulated annealing approach

Whether for initial sampling or sampling, the objective is to find the optimal location of samples which will minimize an objective function. That set of optimal samples is selected from a set of candidate locations P, which is relatively large in practice, forbidding a total enumeration for the optimal set (Michalewicz and Fogel 2000). A H guides the search for an optimal sample set S^* (or near optimal S^+) $\subset P$. The set S^* is optimal to the objective function J defined in Equation 1.16. The efficiency of a depends on its capacity to give as often as possible a solution S^+ close to S^* (Grötschel and Lovàsz 1995). A *naïve optimization* selecting n points at random would not return a very good value for J, and is therefore not efficient. A *simultaneous* addition selects a set of samples one time. The major concern lies in the selection of those points: for an initial sampling set of, say $m = 50$ with a candidate set $p = 445$, there is a combinatorial explosion $\binom{p}{n} = \binom{445}{50}$, while in additional sampling, with an initial set $m = 50$ and an additional set of $n = 20$, the number of combinations becomes $\binom{p}{n} = \binom{445-50}{20}$. With such a high number of possible combinations, it is recommended to use intelligent search techniques, for instance simulated annealing, which is detailed in the

next paragraph.

Simulated annealing (SA) has been used as an optimization procedure in initial and sampling; Van Groenigen and Stein (1998) and Delmelle (2005) further discusses the implementation of in spatial sampling. SA is a method by which a metal cools and freezes into a minimum energy crystalline structure. The algorithm was originally proposed as a means of finding the equilibrium configuration of a collection of atoms at a given temperature. Kirkpatrick *et al.* (1983) made the connection between the cooling technique and the mathematical minimization problem. The major advantage of SA is its ability to avoid becoming trapped at local maxima. The algorithm employs a random search that accepts changes improving the objective function, but also non-improving moves. The latter is accepted with probability P_T where T stands for the current temperature. T cools down as the algorithm progresses, and so does the probability of acceptance. Different authors document the use of simulated annealing for spatial sampling (Van Groenigen *et al.* 1999, Lark 2002, Delmelle and Goovaerts 2009). The optimization starts with a randomly selected sampling scheme $S(i) = M \cup N$ where M is the initial set, and N the set of additional samples ($N \subset P \setminus M$). The set N can be selected at random or is determined from a greedy , for instance using samples corresponding to the peaks of the surface area (Delmelle and Goovaerts 2009). The objective function $J\big[S(i)\big]$ is evaluated and called the incumbent solution. $S(i) = \{\mathbf{s}_1 + \ldots + \mathbf{s}_m + \mathbf{s}^j_{m+i} + \ldots + \mathbf{s}^j_{m+n}\}$ is created as the union of the initial sample set M, augmented by a set of new sample points $\mathbf{s}^j_{m+i}$. From the first iterations, $S(i)$ becomes $S(j)$ and will serve as the initial sampling scheme for simulated annealing. Since the space of solution has not been explored yet, the value of $J\big[S(i)\big]$ is kept in memory as the best solution $S(i) = S^+ = S^\diamond$ found so far. $S(i)$ becomes $S(j)$. One sample $i \ni M$ is swapped for a point $i \in M$, and at that time $S(j)$ becomes $S(j+1)$.

Follows a sequence of random perturbations $S(j+1)$ of $S(j)$ that have a probability $P_T\big\{S(j) \to S(j+1)\big\}$ of being accepted, where $\mathbf{s}^j_{m+i}$ is swapped in favor of $\mathbf{s}^{j+1}_{m+i}$:

$$P_T\big\{S(j) \to S(j+1)\big\} = 1 \qquad \text{if } J\big[S(j+1)\big] \leq J\big[S(j)\big] \tag{1.21}$$

$$P_T\big\{S(j) \to S(j+1)\big\} = \frac{1}{1 + e^{\left(\frac{\triangle J}{T}\right)}} \qquad \text{if } J\big[S(j+1)\big] > J\big[S(i)\big] \tag{1.22}$$

The sampling scheme $S(j)$ becomes $S(j+1)$ when $J\big[S(j+1)\big] \geq J\big[S(j)\big]$, or when $J\big[S(j+1)\big] \leq J\big[S(j)\big]$ a test must be conducted as follows: $exp[-\frac{(candObj - incObj)}{\xi \times T}] >$*rand*(), with ξ a parameter reflecting changes in the objective function J. If $S(j+1)$ is accepted, it becomes the incumbent solution $S^\diamond = S(j+1)$, and serves as a starting point for a next scheme $S(j+2)$. The process continues in a similar fashion, until a certain level of iterations T_{fin} has been reached. Note that the temperature and step size κ decrease as the algorithm progresses. As the system cools down, the probability of accepting non-improving moves decreases with temperature decrease. The algorithm stops when the temperature T has reached its cutoff value T_{fin}. To find the set of optimal points $\mathbf{s}^*$ to be added-or nearly optimal $\mathbf{s}^+$, a high starting temperature T and a cooling factor T_{dec} close to 1 are necessary. In these conditions, simulated annealing can escape from a local maximum, but too slow of a cooling schedule will increase the running time of the algorithm. A trade-off between optimality gap and running time has to be determined. As a rule of thumb, it is advisable to have a large step size δ in the beginning of the algorithm, allowing wide jumps across $\mathfrak{D}$, when there are different maxima across the map. Simulated annealing is therefore sensitive to the choice of the cooling factor, that governs the search procedure. Additionally, it is suggested to conduct the simulated annealing in two stages, where the second one consists of conducting the search around the best solution found so far during the first stage (Delmelle 2005). During the second time however, a smaller step size κ is used. Note that the simulated annealing algorithm does not always converge. A pseudocode for the algorithm is given below:

def *kv* = function computing
set incSet;incObj←kv(incSet)
set T, K, T_{fin}
While T> T_{fin}
candSet←neighbor(incSet,K)
candObj←kv(candSet)
If candObj<bestObj
incObj←candObj, incSet←candSet
bestObj←candObj, bestSet←candSet
ElseIf exp[-$\frac{(candObj-incObj)}{\xi \times T}$] >*rand*()
incObj←candObj, incSet←candSet
EndIf
T=T×T_{dec}, K=K×K_{dec}
End While

1.5 Illustration:

In this section, we apply simulated annealing to the Austrian Data to gain insight into the problem structure and behavior. The goal is to find the location of (initial and additional) sample points which will keep the to a minimum. All computational results were obtained using Matlab v. 2010.

Initial sampling designs.

There are different sampling objectives exist in the literature when little (or no) a priori spatial information is known on the variable of interest. Assuming some spatial covariance information (range r, nugget a and sill σ^2), the sampling design can be optimized to minimize the . Information on the covariance structure from previous study can have a strong influence on the sampling , with large range r values spreading observations throughout the study region, while short ranges bring these closer to one another. In the Austrian dataset, the centroid of each statistical unit is used as a candidate location in the optimization, that is $p = 445$. We use an exponential covariogram model with parameters $r = 50000m$, $\sigma^2 = 200$ and a nugget $a = 1$. We optimize the location of initial samples ($m = 30$ and $m = 50$, respectively) for the minimization of the using randomly selected points as the starting set (see Figure 1.2 for an illustration) The σ_k^2 is measured and displayed on each statistical unit support $(\mathbf{s})$. The is measured on the points which can also serve as to be included in the sampling set. Since the statistical unit support are not exactly distributed in a grid pattern in the region (especially in the outer edge), careful attention must be paid when evaluating over the study region. The average (Equation 1.11) is lower when more measurements are used in the initial sampling phase. Due to a large range, the optimization algorithm tends to spread the points as far as possible from one another, guaranteeing a certain level of coverage. The graphs to the right illustrates the simulated annealing algorithm, which accepts several non-improving moves in the beginning of the optimization, but eventually converge to a minimum .

Augmented designs

We illustrate the example of augmented designs for the minimization of the . Additional samples are usually taken away from existing samples (van Groenigen and Stein 1998, Delmelle 2005), that is where the is maximum. The at unsampled locations can be determined with multiple design scenarios. An augmented design is said optimal to Equation 1.16 when the configuration maximizes the change in . Figure 1.3 illustrates the addition of $n = 10$ and $n = 25$ points to an initial randomly selected sample set $m = 50$. The large, black triangles are the locations of additional samples. In both cases, the initial sampling set (red triangles) is the same, but the locations of new samples change. The algorithm tends to these areas with additional samples away from initial samples. When the number of additional samples increases, the drops

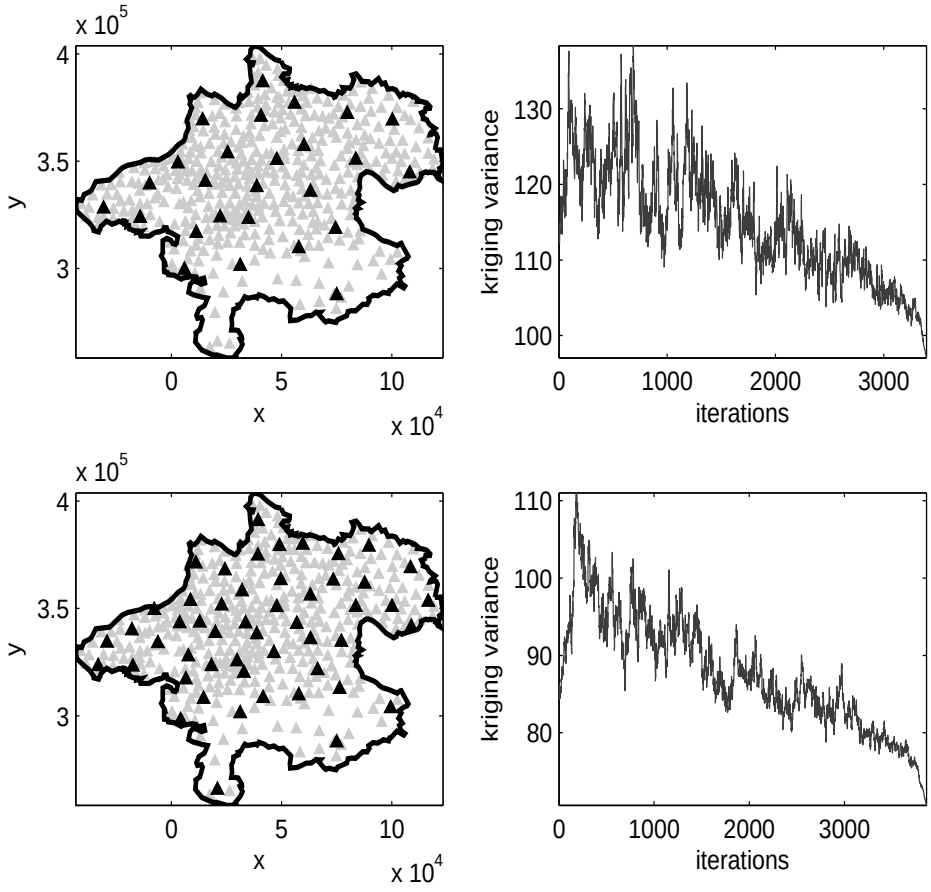

Figure 1.2 Two initial sampling designs with $m = 30$ samples (top figure) and $m = 50$ (bottom figure) from $p = 445$ candidate locations, minimizing the . Light grey triangles represent potential locations while black triangles are selected samples. The graph on the right illustrates the converge of the optimization process with simulated annealing.

significantly. For visualization purposes, the at each statistical support is assigned to the polygon its belong to, and a map in Figure 1.4 indicates that the remains high in the outer edges of the region.

1.6 Discussion:

In this chapter, we have illustrated a framework for initial and spatial sampling problems based on the change in . Different objectives exists to account for heterogeneity of the spatial variable, for instance by weighting the by locally-varying priorities. These weights can reflect local non-homogeneity. Another approach consists of using locally varying variograms (Haas 1990). Due to the non-linearity of the objective, it is recommended to use a technique to find a suitable set. Results of our application to the Austrian dataset illustrates the suitability of to optimize initial and augmented sampling set.

References/Further Reading and Bibliography

Arbia G. (1989). *Spatial Data Configuration in Statistical Abalysis, Regional Economics and Related Problems*. Kluwer, Dordrecht.

Armstrong M. (1994). Is research in mining geostats as dead as dodo? *In*: Dimitrakopoulos R. (Ed.) *Geostatistics for the Next Century*. KLuwer Academic Publisher. Dordrecht: 303-312.

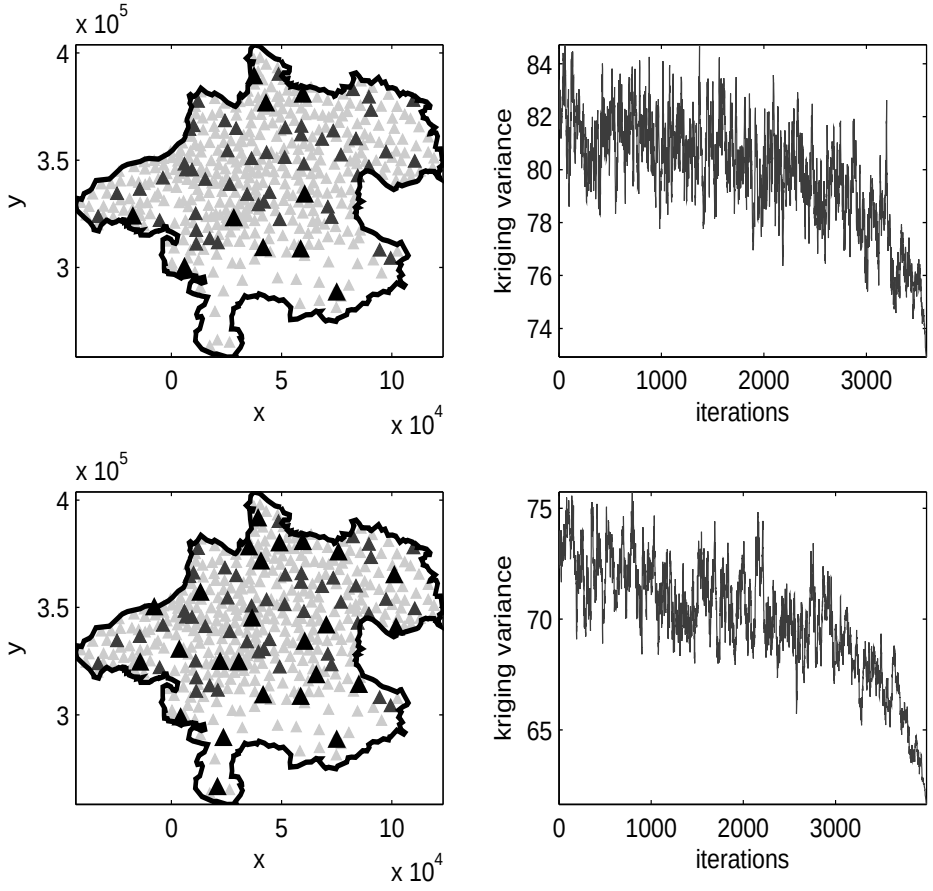

Figure 1.3 An example of two augmented sampling designs; $n = 10$ samples (top figure) and $n = 25$ (bottom figure) from $p = 445 - 50$ candidate locations, minimizing the . Light gray triangles represent potential locations while larger, black triangles are selected samples. The graph on the right illustrates the optimization process with simulated annealing.

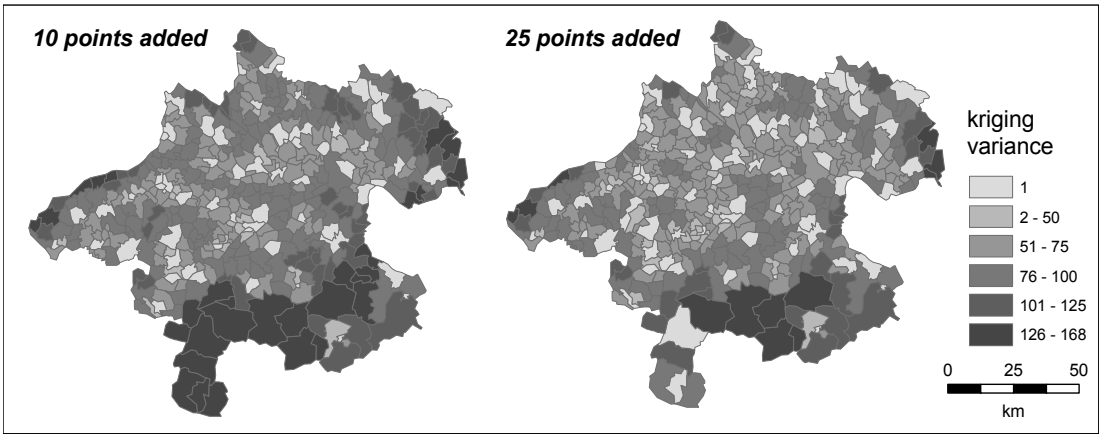

Figure 1.4 The after the addition of new samples to the initial set (use in conjunction with Figure 1.3). Darker areas denote regions of high uncertainty since no points were located in those regions.

Bailey T.C. and A. C. Gatrell (1995). *Interactive Spatial Data Analysis*. Longman, 413p.

Berry B.J.L. and A.M. Baker (1968). Geographic sampling. *In:* Berry B.J.L. and Marble D.F. (Eds.) *Spatial analysis: a reader in statistical geography*: 91-100. Englewood Cliffs, N.J., Prentice-Hall.

Brus D. and G.B.M. Heuvelink (2007). Optimization of sample patterns for universal kriging of environmental variables. *Geoderma*,vol. **138**: 86-95

Burgess T.M., Webster R. and A.B. McBratney (1981). Optimal interpolation and isarithmic mapping of soil properties: IV. Sampling strategy. *Journal of Soil Science*, vol. **32**: 643-659.

Cochran W.G. (1963). *Sampling techniques*. Second Edition. Wiley, New York, USA. 413p.

Christakos G. and R.A. Olea (1992). Sampling design for spatially distributed hydrogeologic and environmental processes. *Advanced Water Resources*, vol. **15**: 219-237.

Cressie, N. (1991). *Statistics for Spatial Data. Wiley*, New York, USA, 900p.

Dalton R., Garlick J, Minshull R. and Robinson A. (1975). *Sampling Techniques in Geography*. Goerges Philip and Son Limited, London, 95p.

Das A.C. (1950). Two-dimensional systematic sampling and the associated stratified and random sampling. *Sankhy* $\bar{a}$vol. **10**: 95-108.

De Gruijter, J., Brus, D.J., Bierkens, M.F.P. and Knotters M. (2006). *Sampling for Natural Resource Monitoring*. Springer, 332p

Delmelle E. (2005). *Optimizing Second-Phase Spatial Sampling Using Auxiliary Information*. Ph.D. dissertation, SUNY at Buffalo.

Delmelle E. (2009). Spatial Sampling. In: Fotheringham and Rogerson (eds.) *Handbook of Spatial Analysis*. Sage Publications. 528p.

Delmelle E. and P. Goovaerts (2009). Second-phase sampling designs for non-stationary spatial variables. *Geoderma*, vol. **153**: 205-216

Deutsch C.V. and A.G. Journel (1997) *Gslib: Geostatistical Software Library and User's Guide*. Oxford University Press, 2^{nd} edition, 369p.

Goovaerts P., 1997. Geostatistics for natural resources evaluation. 483p.

Griffith D. (2005). Effective geographic sample size in the presence of spatial autocorrelation. *Geographical Analysis*, vol.**95**: 740-760.

Griffith D. and C. Amrhein (1997). *Multivariate Statistical Analysis for Geographers*, Prentice Hall, 275p.

Grötschel M. and L. Lovàsz (1995). Combinatorial optimization. *In:* Graham R.L., Grötschel M. and L. Lovàsz (Eds.) *Handbook of Combinatorics*, vol 2: 1541-1579. Elsevier, Amsterdam, The Netherlands.

Haas T.C. (1990). kriging and automated variogram modeling within a moving window. *Atmospheric Environment*, vol.24: 1759-1769.

Haining R.P. (1990). Sampling spatial populations. *In:* Haining R.P. (Ed.) *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press.

Hedayat A.S. and B.K. Sinha (1991). *Design and inference in finite population sampling*. Wiley, New York, USA, 377p.

Iachan R. (1985). Plane sampling. *Statistics and Probability Letters*, vol. **3**: 151-159.

Horn R. A. and C. R. Johnson (1985). *Matrix Analysis*, Cambridge University Press. Cambridge, U.K.

Isaaks E.H. and R.M. Srivastava (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, New York, USA, 561p.

King L.J.(1969). *Statistical Analysis in Geography*. Chapter 4.

Kirkpatrick S., Gelatt C.D. and M.P. Vecchi (1983). Optimization by simmulated annealing. *Science*, vol. **220**: 671-680.

Lark R.M. (2002). Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. *Geoderma*, vol. **105**: 49-80.

Matérn B. (1960). *Spatial variation.* Springer-Verlag. Berlin, Germany, 151p.

Michalewicz Z. and D. Fogel (2000). *How to Solve It: Modern Heuristics*. Springer. 467p.

Muller, W., 1998. Collecting Spatial Data: Optimal Design of Experiments for Random Fields. Heidelberg: Physica-Verlag.

Overton W.S. and S.V. Stehman (1993). Properties of designs for sampling continuous spatial resources from a triangular grid. *Communications in Statistics - Theory and Methods*, vol. **21**: 2641-2660.

Quenouille M.H. (1949). Problems in plane sampling. *Annals of Mathematical Statistics*, vol. **20**: 355-375.

Ripley B.D. (1981). *Spatial statistics*. Wiley, New-York, USA. 252p.

Rogerson P.A., Delmelle E.M., Batta R., Akella M.R., Blatt A and G. Wilson (2004). Optimal sampling design for variables with varying spatial importance. *Geographical Analysis*, vol. **36**: 177-194.

Russo D. (1984). Design of an optimal sampling network for estimating the variogram. *Soil Science Society of America Journal*, vol. **48**: 708-716.

Stehman S.V. and Overton S.W. (1996). Spatial sampling. *In: Arlinghaus, S. (Ed.) Practical handbook of spatial statistics*:31-64. CRC Press, Boca Raton FL.

Thompson S.K. (2002). *Sampling*. Wiley. Second Edition. 367p.

Van Groenigen, J.W. and Stein, A., 1998. Constrained optimization of spatial sampling using continuous simulated annealing. Journal of Environmental Quality, vol. 27: 1078-1086.

Van Groenigen J.W., Siderius W. and A. Stein (1999). Constrained optimisation of soil sampling for minimisation of the . *Geoderma*, vol. **87**: 239-259.

Warrick A.W. and D.E. Myers (1987). Optimization of sampling locations for variogram calculations. *Water Resources Research*, vol. **23**: 496-500.

Webster R. and M.A. Oliver (1993). How large a sample is needed to estimate the regional variogram adequately. *In:* Soares A. (ed.), *Geostatistics Tróia '92*: 155-166. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Whittle P. (1963). Stochastic processes in several dimensions. *Bulletin of the International Statistical Institute*, vol. **40**: 974-994.

Zubrzycki S. (1958). Remarks on random, stratified and systematic sampling in a plane. *Colloquium Mathematicum*, vol. **6**: 251-264.

INDEX