10

Spatial Sampling

Eric Delmelle

10.1. INTRODUCTION

When trying to make inferences about a phenomenon, 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; Dalton *et al.*, 1975; Hedayat and Sinha, 1991; and Thompson 2002 for various summaries). A full inventory would yield a clear picture of the variability of the variable of interest, although this process is very time-consuming and expensive. Haining (2003) underlines that the cost of acquiring information on each individual may rule out a complete census. Sparse sampling on the other hand is cheap, but misses important features. However, there are instances where the level of precision may be the major motivation of the sampling process, especially when sampling remains relatively inexpensive. As a rule of thumb, it is generally desirable to have a higher concentration of samples where exhaustive and accurate information is needed, keeping in mind that the number of samples should always be as representative as possible of the entire population (Berry and Baker, 1968).

AQ: Thompson (2002) not in refs

When surveying a phenomenon characterized by spatial variation, it is necessary to find optimal sample locations in the study area D. This problem is referred to spatial or two-dimensional sampling and has been applied to many disciplines such as mining, soil pollution, environmental monitoring, telecommunications, ecology, geology, and geography, to cite a few. Specific studies on spatial sampling can be found in Ripley (1981), Haining (2003), Cressie (1991), Stehman and Overton (1996) and Muller (1998). Spatial and non-spatial sampling strategies share common characteristics:

1 the size m of the set of samples;
2 the selection of a sample design, limited by the available budget;
3 an estimator (e.g., the mean) for the population characteristic; and
4 an estimation of the sampling variance to compute confidence intervals.

Following Haining, spatial sampling challenges can be divided into three different categories. The first pertains to problems concerned with estimating some non-spatial characteristics of a spatial population; for example, the average income of households in a state. The second category deals with problems where the spatial variation of a variable needs to be known, in the form of a map, or as a summary measure that highlights scales of variation. The third category includes problems where the objective is to obtain observations that are independent of each other, allowing classical statistical procedures to assist in classifying data.

10.1.1. Spatial structure

A common objective in both spatial and non-spatial approaches is to design a sampling configuration that minimizes the variance associated with the estimation. In this regard, the location of the samples is very critical and depends heavily on the structure of the variable. In non-spatial problems, it may be crucial to stratify the sampling scheme according to important underlying covariates. This holds for spatial phenomena as well. Unfortunately, this variation is often unknown, and an objective is to design an optimal sampling arrangement, to obtain a maximum amount of information. If we undersample in some areas, the spatial variability will not be captured. Oversampling on the other hand can result in redundant data. Consequently, not only the quantity of the samples is important, but also their locations. This chapter is concerned primarily with the second category of sampling challenges, i.e. capturing the spatial structure of the primary variable.

10.1.2. Structure of the chapter

In this chapter, spatial sampling configurations are reviewed along with their benefits and drawbacks. Second, the influence of geostatistics on sampling schemes is discussed. Sampling schemes can be designed to capture the spatial variation of the variable of interest. Two common objectives therein are the estimation of the covariogram and the minimization of the kriging variance. Third, methods of adaptive sampling and second-phase sampling are presented. Such methods are of nonlinear nature, and appropriate optimization techniques are necessary to solve such problems. Finally, salient sampling problems such as sampling in the presence of multivariate information, and the use of heuristics are discussed.

10.2. SPATIAL SAMPLING CONFIGURATIONS

This section reviews significant sampling schemes for the purpose of two-dimensional sampling. In the following subsections I will assume that a limited number of samples m is collected within a study area denoted D. The variable of interest Z is sampled on m supports, generating observations $\{z(\mathbf{s}_i) \mid _i = 1, 2, \ldots m\}$. For ease of illustration, a square study area is used.

10.2.1. Major spatial sampling designs

Random sampling

A *simple random sampling* scheme consists of choosing randomly a set of m sample points in D, where each location in D has an equal probability of being sampled (Ripley, 1981). The selection of a unit does not influence the selection of any other one (King, 1969). Figure 10.1(a) illustrates the random configuration. This type of design is also called *uniform random sampling* since each point is chosen independently uniformly within D. Practically, two random numbers K_i and K_i' are drawn

AQ: King (1969) publishing details

from the interval [0, 1]. Then the point $\mathbf{s}_i$, defined by the pair $\{x_i, y_i\}$ is selected such that:

$$x_i = K_i L, \quad y_i = K_i' L, \tag{10.1}$$

where L denotes the length of the study area D (Aubry, 2000). The process is repeated m-times. According to Griffith and Amrhein (1997), the distribution of the points may not be representative of the underlying geographic surface, because

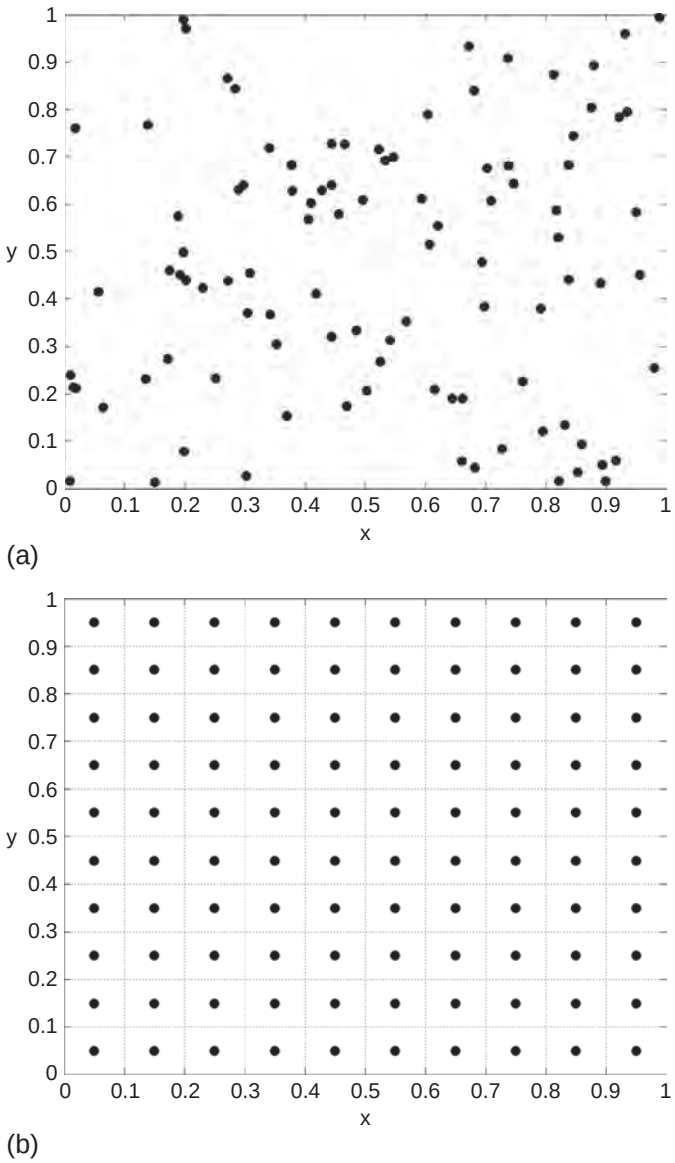
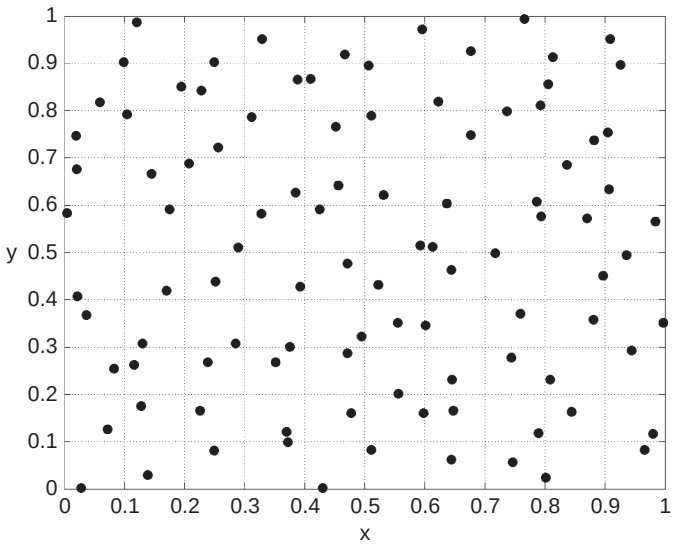
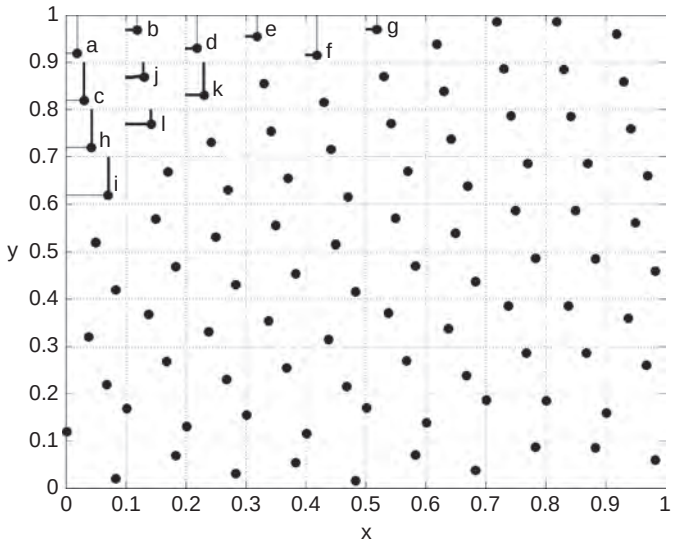


Figure 10.1 From left to right, top to bottom: random, centric systematic, systematic random, and systematic unaligned sampling schemes. Sampling size $m = 100$.



(c)



(d)

Figure 10.1 Continued

for most samples drawn, some areas will be oversampled while other will be undersampled. The advantages of this design however reside in its operational simplicity, and its capacity to generate a wide variety of distances among pairs of points in D.

Systematic sampling

The population of interest is divided into m intervals of equal size. The first element is randomly or purposively chosen within the first interval, starting at the origin. Depending on the location of the first sample, the remaining $m - 1$ elements are aligned

regularly by the size of the interval Δ. If the first sample is chosen at random, the resulting scheme is called *systematic random sampling*. When the first sample point is not chosen at random, the resulting configuration is called *regular systematic sampling*. A *centric systematic sampling* occurs when the first point is chosen in the center of the first interval. The resulting scheme is a checkerboard configuration. The most common regular geometric configurations are the equilateral triangular grid, the rectangular (square) grid, and the hexagonal one (Cressie, 1991). Practically, consider the case where D is divided into a set of small, square cells of size $\Delta = L/\sqrt{m}$. A first point $\mathbf{s}_1 = \{x_1, y_1\}$ is selected within the first cell in the bottom left of D. The coordinates of $\mathbf{s}_1$ are subsequently used to determine the following point $\mathbf{s}_i = \{x_i, y_i\}$ (Aubry, 2000):

$$x_i = x_1 + (i-1)\Delta, \quad y_i = y_1 + (j-1)\Delta$$
$$\forall i,\ j = 1, \ldots, \sqrt{m}. \tag{10.2}$$

To locate sample points along the x- and y-directions, it is imperative to have a desired number of samples m for which $\sqrt{m}$ must be an integer value. The benefits of a systematic approach reside in a good spreading of observations across D, guaranteeing a representative sampling coverage. Additionally, the spreading of the observations prevents sampling clustering and redundancy. This design however presents two inconveniences:

1 the distribution of distances between points of D is not sampled adequately because many pairs of points are separated by the same distance; and
2 There is a danger that the spatial process shows evidence of recurring periodicities that will remain uncaptured, because the systematic design coincides in frequency with a regular pattern in the landscape (Griffith and Amrhein, 1997; Overton and Stehman, 1993).

The second drawback can be lessened considerably by use of a *systematic random* method that combines systematic and random procedures (Dalton *et al*., 1975). One sample point is randomly selected within each cell. However, sample density needs to be high enough to have some clustering of observations or the spatial relationship between observations cannot be built. From Figure 10.1(c), some patches of D remain undersampled, while others regions show evidence of clustered observations. A *systematic unaligned* scheme prevents this problem from occurring by imposing a stronger restriction on the random allocation of observations (King, 1969).

Stratified sampling

According to Haining (2003), there are cases when local-area estimates are to be examined, causing stratification to be built into the sampling strategy. In *stratified sampling*, the survey area (or D) is partitioned into non-overlapping strata[1]. For each stratum, a set of samples is collected, where the sum of the samples over all strata must equal m. The knowledge of the underlying process is a determining factor in defining the shape and size of each stratum. Some subregions of D may exhibit stronger spatial variation, ultimately affecting the configuration of each stratum (Cressie, 1991). Smaller strata are preferred in non-homogeneous subregions. When points within each stratum are chosen randomly, the resulting design is named *stratified random sampling*. In Figure 10.2(a), six strata are sampled in proportion to their size. For instance, stratum A represents 30% of D, therefore if $m = 100$, 30 sample points will be allocated within A. Figure 10.2(b) illustrates the allocation of one sample per stratum (*in casu* the centroid), undersampling larger strata.

10.2.2. Efficiency of spatial sampling designs

The sampling efficiency is defined as the inverse of the sampling variance.

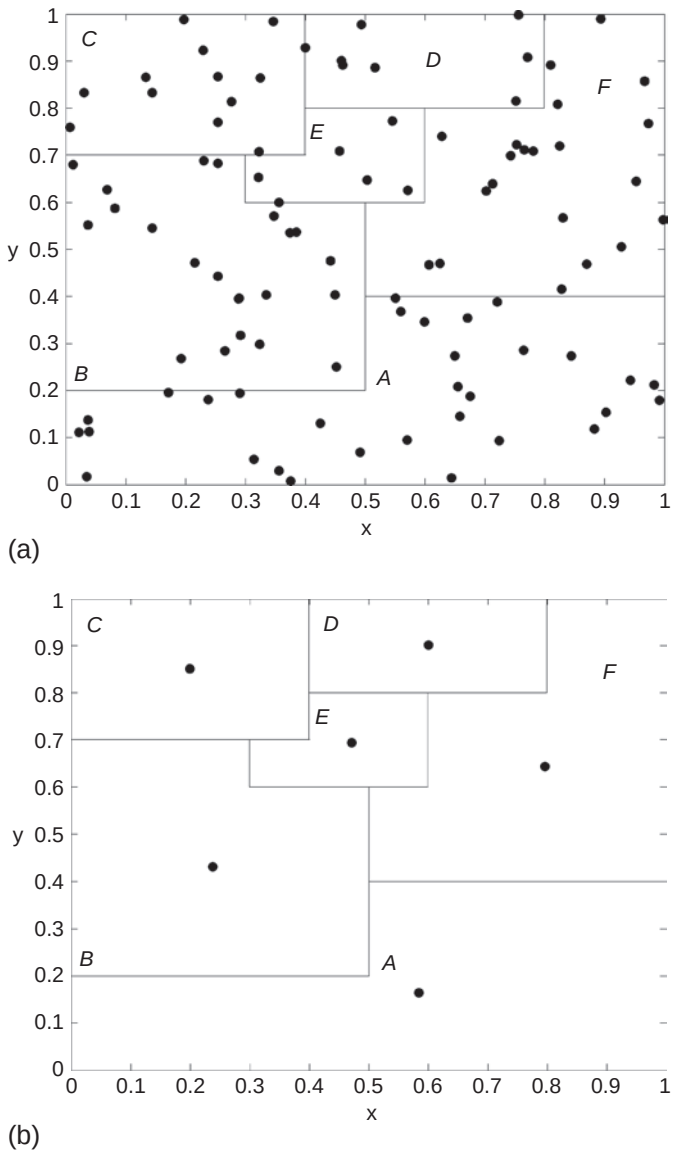


Figure 10.2 Stratified sampling designs with six strata of different sizes ($m = 6$ on the right figure and $m = 100$ to the left).

According to Aubry (2000), the most efficient design leads to the most accurate estimation. Consider the estimation of the global mean z_D:

$$z_D = \frac{1}{[D]} \int_D z(\mathbf{s})\, \mathrm{d}\mathbf{s}. \qquad (10.3)$$

It is desirable, from a statistical standpoint to select a configuration that minimizes the prediction error of z_D for a given estimator, for instance the arithmetic mean:

$$\bar{z} = \frac{1}{m} \sum_{i=1}^{m} z(\mathbf{s}_i). \qquad (10.4)$$

Efficiency is calculated for all possible realizations of the variable Z by $Var_{\xi}\left[Z_D^* - Z_D\right]$ using σ_k^2, which is the geostatistical prediction error, defined later. 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.

Spatial autocorrelation. Ideally, the density of sample points should increase in locations exhibiting greater spatial variability. Values of closely spaced samples will show strong similarities and it may be redundant to oversample in those areas. The spatial autocorrelation function summarizes the similarity of the values of the variable of interest at different sample locations, as a function of their separating distance (Gatrell, 1979; Griffith, 1987). Moran's I (Moran, 1948, 1950) is a measure of the degree of spatial autocorrelation among data points:

$$I = \frac{m}{\mathbf{W}} \frac{\sum_{i,j} w(\mathbf{s}_{ij})(z(\mathbf{s}_i) - \bar{z})(z(\mathbf{s}_j) - \bar{z})}{\sum_i (z(\mathbf{s}_i) - \bar{z})^2} \tag{10.5}$$

with $\mathbf{W}$ defined as a weight matrix $w(\mathbf{s}_{ij})$, m is the number of observations, the mean of the sampled values is denoted by $\bar{z}$ and $z(\mathbf{s}_i)$ is the measured attribute value at location $\mathbf{s}_i$. The weight $w(\mathbf{s}_{ij})$ is a measure of spatial proximity between points $\mathbf{s}_i$and $\mathbf{s}_j$; for example:

$$w(\mathbf{s}_{ij}) = \exp(-\beta \mathrm{d}(\mathbf{s}_{ij})^2) \tag{10.6}$$

where $d(\mathbf{s}_{ij})^2$ is the squared distance between location $\mathbf{s}_i$ and point $\mathbf{s}_j$. Moran's I value is not implicitly constrained within the interval $[-1, +1]$. Spatial autocorrelation generally decreases as the distance among sample points increases. A positive autocorrelation occurs when values taken at nearby samples are more alike than samples collected further away. When the autocorrelation is a linear decreasing function of distance, *stratified random sampling* has a smaller variance than a *systematic design* (Quenouille, 1949). If the decrease in autocorrelation is not linear, yet concave upwards, systematic sampling is more accurate than stratified random sampling, and a centered systematic design, where each point falls exactly in the middle of each interval, is more efficient than a random systematic sampling configuration (Madow, 1953; Zubrzycki, 1958; Dalenius *et al*., 1960; Bellhouse, 1977; Iachan, 1985).

10.2.3. Other sampling designs

Nested or hierarchical sampling

Nested or hierarchical sampling designs require the study area D to be partitioned randomly into sample units (or blocks) creating the first level in the hierarchy, which is then further subdivided into sample units nested within level 1, and so forth (Haining, 2003). These units can be systematically or irregularly arranged. As the process progresses, the distances between observations decreases (Corsten and Stein, 1994). One advantage of a nested sampling design is that it allows for multiple scale analysis and supports quadrat analysis. Spatially nested sampling designs may work well for geographic phenomenon that are naturally clustered and for exploring multiple scale effects. Hierarchical sampling is also possible at the discrete level. In such cases, it is desirable to select at first randomly one or more counties in a state. Then within these counties we might sample a number of quadrats, or say, townships and finally, within the latter, randomly select some farmsteads (King, 1969).

In a multivariate case, dependent and independent variables are hierarchically organized and are thus not collected at the same sampling frequency (Haining, 2003). The primary variable may exhibit rapid change in spatial structure while the secondary variables are much more homogeneous. A hierarchical sampling design captures such variation by collecting one variable at points nested within larger

sampling units so it can be collected more intensively than another variable.

Clustered sampling

This type of sampling consists of the random selection of groups of sites where sites are spatially close 'within' groups (Cressie, 1991). Clusters of observations are drawn independently with equal probability. In a first stage, when the population is grouped into clusters, the clusters are first sampled (Haining, 2003). Either all of the observations in the clusters are included, or only a random selection from it. Cluster sampling is essentially useful in a discrete case, when a complete list of the members of a population cannot be obtained, yet a complete list of *groups* (i.e., clusters) of the variable is available. The method is also useful in saving sampling cost.

10.3. SAMPLING RANDOM FIELDS USING GEOSTATISTICS

Most classical statistical sampling methods make no use of the spatial information provided by nearby samples. Geostatistics describes the spatial continuity that is an essential feature of many natural phenomena. It can be seen as a collection of statistical methods, describing the spatial autocorrelation among sample data. In geostatistics, multidimensional random fields are formalized and modeled as stochastic processes (see, e.g., Matérn, 1960; Whittle, 1963). In other words, the variable of interest is modeled as a random process that can take a series of outcome values, according to some probability distribution (Goovaerts, 1997). Kriging is an interpolation technique that estimates the value of the primary variable at unsampled locations (usually on a set G of grid points $\{\mathbf{s}_g \mid {}_g = 1, 2, \ldots, G\}$, while minimizing the prediction error. Using data values of Z, an empirical semivariogram $\hat{\gamma}(h)$ summarizing the variance of values separated by a particular distance lag (h) is defined:

$$\hat{\gamma}(h) = \frac{1}{2\mathrm{d}(h)} \sum_{|s_i - s_j| = h} (z(\mathbf{s}i) - z(\mathbf{s}j))^2 \tag{10.7}$$

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 $\mathbf{s}_i$. The semivariogram is characterized by a nugget effect a, and a sill σ^2 where $\hat{\gamma}(h)$ levels out. The nugget effect is the spatial dependence at micro scales, caused by measurement errors at distances smaller than the possible sampling distances (Cressie, 1991). Once the lag distance exceeds a certain value r, called the range, there is no spatial dependence between the sample sites. The variogram function $\hat{\gamma}(h)$ becomes constant at a value called the sill σ^2. A model $\gamma(h)$ is fitted to the experimental variogram (e.g., an exponential model). With the presence of a nugget effect a:

$$\gamma(h) = a + (\sigma^2 - a)(1 - e^{-3h/r}). \tag{10.8}$$

The corresponding covariogram $C(h)$ that summarizes the covariance between any two points is:

$$C(h) = C(0) - \gamma(h) = \sigma^2 - \gamma(h). \tag{10.9}$$

The interpolated, kriged value at a location $\mathbf{s}$ in D is a weighted mean of surrounding values; each value is weighted according to the covariogram model:

$$\hat{z}(s) = \sum_{i=1}^{I} w_i(\mathbf{s}) z(\mathbf{s}_i) \tag{10.10}$$

where I is the set of neighboring points that are used to estimate the interpolated

value at location **s**, and $w_i(\mathbf{s})$ is the weight associated with each surrounding point. The optimization of spatial sampling in a geostatistical context first requires the estimation of a model to express the spatial dependence at different pairs of distances. This is summarized in the covariogram function. Secondly, such a model is then used for optimal interpolation of the variable under study (Van Groenigen, 1997).

10.3.1. Optimal geometric designs for covariogram estimation

To compute the most representative covariogram and to capture the main features of spatial variability, a good spreading of sample points across the study area is necessary (Van Groenigen *et al*., 1999). In that context, *systematic sampling* (Figure 10.1(b)) performs well. However, such a sampling design does not guarantee a wide range of separating distances (which is necessary to estimate the covariogram), because:

1 distances are not evenly distributed; and
2 there are few pairs of points at very small distances to estimate the nugget effect.

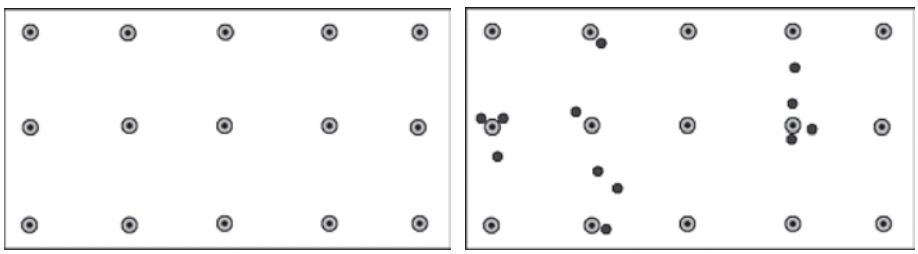
A *systematic random* or *systematic unaligned* will generate a greater variety of distance pairs. Another solution consists of designing a sampling arrangement where a subset of the m observations are evenly spread across the study area D and the remaining points are somewhat more clustered (Figure 10.3), to capture the covariance at very small distances.

Sample size and sample configuration issues

Optimizing the sampling configuration to estimate the parameters of the covariogram is not an easy task. Webster and Oliver (1993) suggested that a total of at least $m = 150$ samples over the study area is necessary. Moreover, the reliability of the covariogram is partly dependent on the number of pairs of points available within each distance class. In this context, the Warrick/Myers (WM) criterion tries to reproduce an *a priori* defined ideal distribution of pairs of points for estimating the covariogram. The procedure allows one to account for the variation in distance. Following Van Groenigen (1997), the WM-criterion is defined as:

$$J_{w/m}(S) = \mathbf{a}\sum_{i=1}^{K}\mathbf{w}_i(\xi_i^* - \xi_i)^2 + \mathbf{b}\sum_{i=1}^{K}\sigma(m_i) \tag{10.11}$$

$$\sum_{i=1}^{K}\xi_i^* = \frac{m(m-1)}{2} \tag{10.12}$$



(a) (b)

Figure 10.3 A systematic sampling scheme of $m = 36$ points in D is improved by the introduction of $n = 12$ additional samples (•) clustered among the initial samples.

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 $\xi_i^*$is a prespecified number of point-pairs for the ith 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). Equation (10.12) expresses the total number of possible distance pairs, given the number of samples. So for instance, when $m = 4$, six pairs of points are generated.

Presence of anisotropy

Anisotropy (as opposed to isotropy) is a property of a natural process, where the autocorrelation among points changes with distance and direction between two locations. In other words, spatial variability is direction-dependent. Spatial variables may exhibit linear continuity, such as in estimating riparian habitat along rivers, aeolian deposits, and soil permeability along prevailing wind directions. We talk about an isotropic process however when there is no effect of direction in the spatial autocorrelation of the primary variable. It is generally desirable to augment the sampling frequency in the angle of minimum continuity, since the spatial gradient of variation is maximum in that direction.

Impact of the nugget effect

Bogaert and Russo (1999) made an attempt to understand how the covariogram parameters are influenced by the choice of particular sampling locations. Their objective was to limit the variability of the covariogram estimator. When the covariogram has no nugget effect, the benefits of the optimization procedure are somewhat diminished. In the presence of a nugget effect, a *random sampling* configuration will score poorly, because of the limited information offered by random sampling for small distances.

Using nested designs

A nested design allows good estimation of the nugget effect at the origin. However, *nested sampling* configurations produce inaccurate estimation of the covariogram in comparison to *random* and *systematic sampling*. This occurs due to the rather limited part of an area covered by the sampling scheme, yielding a high observation density in subregions of the area, and a low observation density for other parts of the area. This in turn generates only a few distances for which covariogram values are available. Nested sampling designs are especially unsuitable when the observations collected according to such a design are used subsequently to estimate values at unvisited locations (Corsten and Stein, 1994).

10.3.2. Optimal designs to minimize the kriging variance

Kriging provides not only a least-squares estimate of the attribute but also an attached error variance (Isaaks and Srivastava, 1989), quantifying the prediction uncertainty at a particular location in space. This uncertainty is minimal, or zero when there is no nugget effect, at existing sampling points and increases with the distance to the nearest samples. A major objective consists of designing a sampling configuration to minimize this uncertainty over the study area. This can be achieved when the covariogram, representing the spatial structure of the variable, is known *a priori* or has been estimated. In this regard, optimal sampling strategies have been suggested to reduce the prediction error associated with the interpolation process (Pettitt and McBratney, 1993; Van Groenigen *et al*., 1999). Equation (10.13) formulates the kriging variance at a location **s**, where $\mathbf{C}_M^{-1}$ is the inverse of the covariance matrix $\mathbf{C}_M$ 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 (10.15):

$$\sigma_k^2(\mathbf{s}) = \sigma^2 - \mathbf{c}^T(\mathbf{s}) \cdot \mathbf{C}_M^{-1} \cdot \mathbf{c}(\mathbf{s}) \tag{10.13}$$

$$C_M = \begin{bmatrix} \sigma^2 & C_{1,2} & \dots & C_{1,m} \\ C_{2,1} & \sigma^2 & \cdots & C_{2,m} \\ \vdots & \vdots & \ddots & \\ C_{m,1} & C_{m,2} & \cdots & \sigma^2 \end{bmatrix} \tag{10.14}$$

$$\mathbf{c} = \begin{bmatrix} \sigma^2 \\ C_{2,1} \\ \vdots \\ C_{m,1} \end{bmatrix}, \quad \mathbf{c}^T = \begin{bmatrix} \sigma^2 & C_{1,2} & \dots & C_{1,m} \end{bmatrix}. \tag{10.15}$$

The total kriging variance *TKV* is obtained by integrating Equation (10.13) over *D*:

$$TKV = \int_D \sigma_k^2(\mathbf{s}) \mathrm{d}\mathbf{s}. \tag{10.16}$$

Computationally, it is easier to discretize *D* and sum the kriging variance over all grid points $\mathbf{s}_g$. The average kriging variance *AKV* over the study area is defined as:

$$AKV = \sum_{g \in G} \sigma_k^2(\mathbf{s}_g). \tag{10.17}$$

The only requirement to calculate the kriging variance is to have an initial covariogram and the locations of the *m* initial sample points. It then depends solely on the spatial dependence and configuration of the observations (Cressie, 1991).

Illustration

Since continuous sampling is not feasible, it is necessary to discretize the area into a set of potential points. Seeking the best sampling procedure becomes a combinatorial problem. Figure 10.4 illustrates the kriging variance associated with a random sampling and a systematic random sampling from an exponential model. Darker areas denote a higher interpolation uncertainty, which is increasing away from existing points. The estimation error is low at visited points.

Distance-based criteria

It is possible to design sampling configurations considering explicitly the spatial correlation of the variable (Arbia, 1994). What would you do if you were in a dark room with candles? You would probably light the first candle at a random location or in the middle of the room. Then you would find it convenient to light the second candle somewhere further away from the first. How far away will depend on the luminosity of the first candle. The stronger the light, the further it can be located from the first candle. You would then light the third candle far away from the two first ones. Such an approach – known as ***D**epending areal **U**nits **S**equential **T**echnique* (DUST) – is an infill sampling algorithm, and very suitable to locate points in minimizing the kriging variance over *D*. Another method, known as the ***M**inimization of the **M**ean of the **S**hortest **D**istances* (MMSD) requires all sampling points spread evenly over the study area, ensuring that unvisited locations are never far from a sampling point. Both MMSD and DUST methods assume:

1 prior knowledge of the spatial structure of the variable; and
2 a stationary variable – an assumption violated in practice.

Both criteria are purely deterministic, resulting in spreading pairs of points evenly across the study area, similar to the systematic configuration. Van Groenigen (1997) notes that the area *D* is a continuous, infinite plane. In reality, it is not physically possible

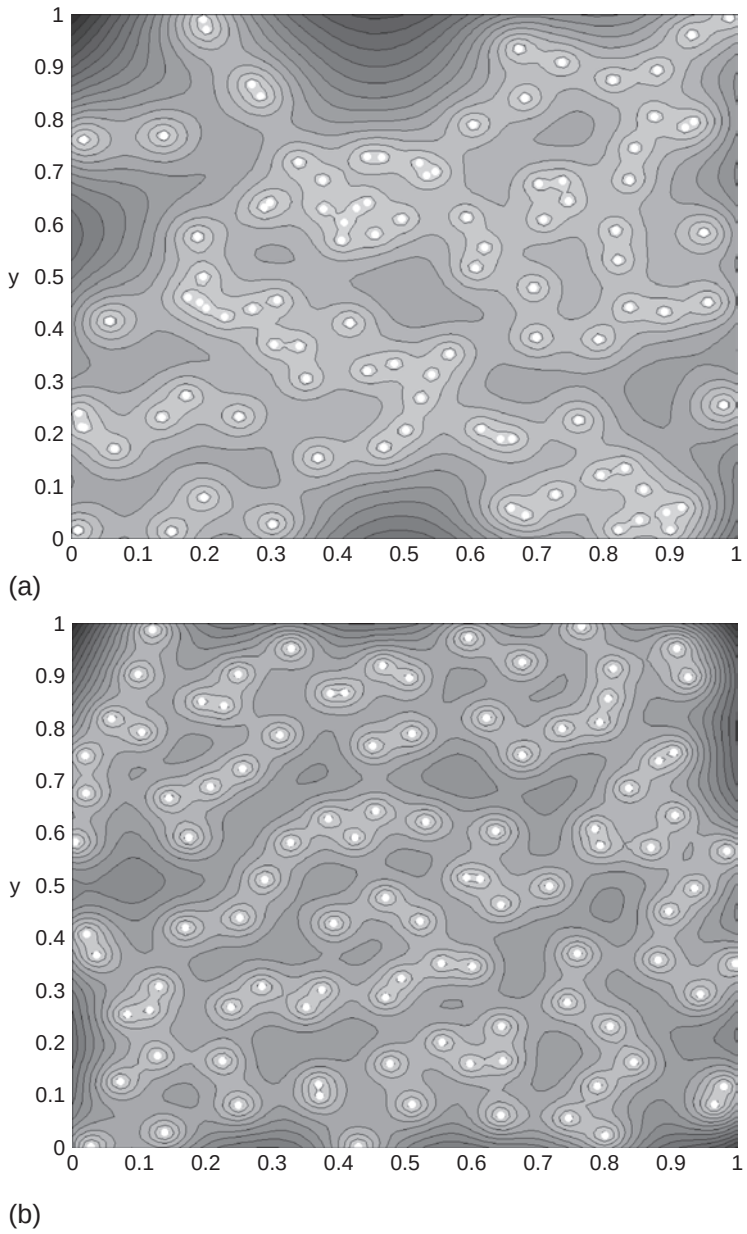


Figure 10.4 The kriging variance of a systematic random pattern (right figure) reduces the value of Equation 17 by 20% from a random pattern. Sample patterns are similar to the ones in Figure 10.1.

to sample everywhere; the presence of spatial barriers such as roads, buildings or mountains restricts the sampling process and limits the number and location of potential points.

Impact of the nugget effect

What is the influence of the nugget effect and sampling densities on the final sampling configuration? As the ratio *nugget*/*sill* increases, a different sampling configuration

is reached, placing more observations near the boundaries of the study area, because of the high variance at short distances. In that case, more samples are needed to obtain the same level of objective function (equation (10.17)) over D (Burgess *et al*., 1981). When the nugget effect is maximum ($\approx$ sill), the covariogram is pure noise, and the resulting optimal sampling scheme is purely random, because no spatial correlation is present. At maximum sampling density, the estimation variance can never be less than the nugget effect. When the variance among pairs of points at very small distances ($\approx$ nugget effect) is very high, a hexagonal design will perform best.

Presence of anisotropy

Which type of sampling design performs better in reducing the maximum kriging variance, when anisotropy is present? When the process is isotropic, a systematic equilateral triangle design will keep the variance to a minimum, because it reduces the farthest distance from initial sample points to points that are not visited. A square grid performs well, especially in the case of isotropy (McBratney and Webster, 1981; McBratney *et al*., 1981). When anisotropy is present on the other hand, a square grid pattern is preferred to a hexagonal arrangement, although the improvement is marginal (Olea, 1984).

Choice of a covariogram fitting model

Does the choice of a covariogram fitting model affect the value of equation (10.17)? According to Van Groenigen (2000), an exponential model generates a point-symmetric sampling configuration that is identical to a linear model. However, the use of a Gaussian model tends to locate sample points very close to the boundary of D. This is explained by the large kriging weights assigned to small distance values (parabolic behavior at the origin).

10.3.3. Sampling reduction

Sampling density reduction of an existing spatial network is a problem related to sampling designs and is relevant in many regions of the world where funding for environmental monitoring is decreasing. The process entails lowering the number of required samples to reach an effective level of accuracy. Technically, it consists of selecting existing samples from the original data set that will, in combination with a spatial interpolation algorithm, produce the best possible estimate of the variable of interest, in comparison with the results obtained if all sample points were used (Olea, 1984). Usually, it is assumed that the residuals come from a stationary process, and that the covariogram is linearly decreasing, with no nugget effect, and that the process is isotropic. In a study aimed at predicting soil water contents, Ferreyra *et al*., (2002) developed a similar sampling density reduction method, from 57 observations to 10 observations. With an optimal arrangement of 10 samples, over 70% of the predicted water contents had an error within $\pm 10\%$, showing that a similar level of confidence is reached with a limited number of samples.

10.4. SECOND-PHASE AND ADAPTIVE SAMPLING

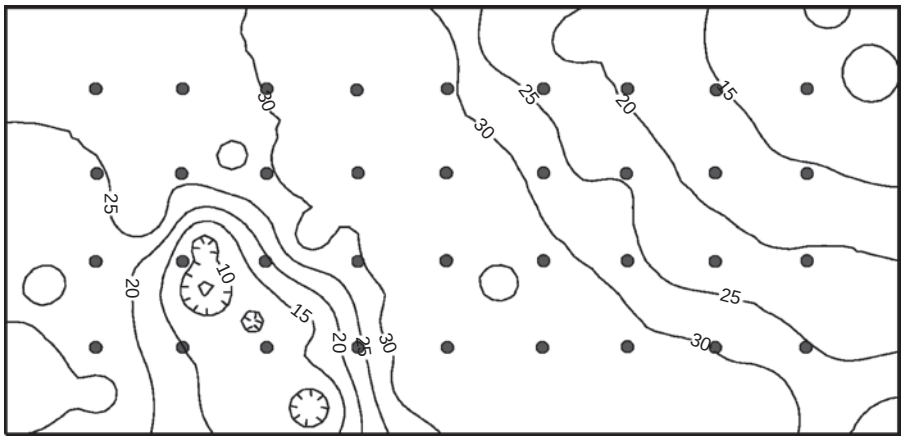
When there is a need or desire to go out in the field to gather more information (i.e., additional samples) about the variable of interest, we talk about adaptive and second-phase sampling, depending on the study objective. In the following subsections, both techniques are discussed.

10.4.1. Adaptive sampling

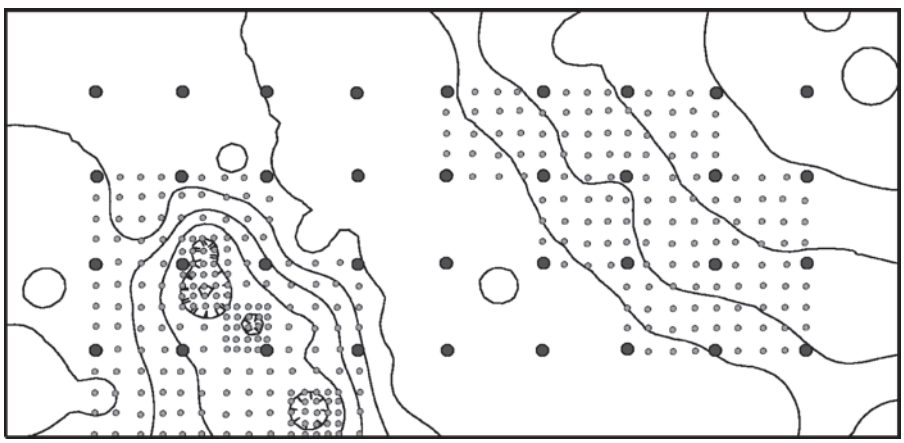
Adaptive sampling finds its roots in the concept of *progressive sampling* (Makarovic, 1973). It provides an objective and automatic method for sampling, for example, terrain of

varying complexity when sampling altitude variation. As illustrated in Figure 10.5, progressive sampling involves a series of successive runs, beginning with a coarse sampling grid and then proceeding to grids of higher densities. The grid density is doubled on each successive sampling run and the points to be sampled are determined by a computer analysis of the data obtained on the preceding run. The analysis proceeds as follows: a square patch of nine points on the coarsest grid is selected and the height differences between each adjacent pair of points along the rows and columns are computed. The second differences are then calculated. The latter carries information on the terrain curvature. If the estimated curvature exceeds a certain threshold, it becomes necessary on the next run to increase the sampling density and sample points at the next level of grid density.

A similar study was carried out by Ayeni (1982) to determine the optimum number and spacing of terrain elevation data points to produce a Digital Elevation Model (DEM). The importance of evaluating



(a)



(b)

Figure 10.5 Initial systematic sampling of altitude is performed over the study in the top figure. When a strong variation in elevation is encountered, the sampling density is increased up to a desirable threshold is met.

the adequate number of data points as well as the appropriate sampling distribution of such points, that in turn constitute a good match to characterize a given terrain. Determining a sufficient number of points is not straightforward, since it depends on terrain roughness in relation to the size of the area occupied by the terrain. The ideas suggested in progressive sampling were later carried over to the field of *adaptive sampling* (see Thompson and Seber, 1996). A major difference with conventional designs lies in the selection of additional samples in adaptive designs, because the location of a new sample will depend upon the value of the points observed in the field. In other words, the procedure for selecting additional samples depends on the outcome of the variable of interest, as observed during the survey of an initial sampling phase. The addition of a new sample contributes to improve the confidence of the sampling distribution. Adaptive sampling is very efficient in the context of soil contamination (Cox, 1999). How should a risk manager decide where to re-sample in order to maximize the information of contamination? In this particular context it is generally recommended to sample in locations above a particular threshold and draw a fixed number of additional samples around them until subsequent measurement values are below a pre-specified contamination threshold. Figure 10.6 illustrates the procedure of adaptive cluster sampling, where sample points represent measurement locations of hypothetical contamination rates. On the left, contamination rates have been measured at seven locations. A geographic location is said to be at risk (and needs remediation) when its value is above 0.7 or at 70% of the contamination threshold. Call a property fathomed if samples have been taken from its immediate neighbors. A common choice is to define new neighbors of a contaminated zone to the North, South, East, and West: fathom each property on the list by sampling and remove it from the risk list when it has been fathomed. In other words, the procedure re-samples four neighboring locations of a contaminated site. Once a site shows a contamination rate under the threshold value, it is fathomed. Otherwise, the procedure continues until a trigger condition is satisfied (e.g., a maximum number of additional samples is reached). This approach has some limitations however, because there is little rationale in taking additional samples in areas where we know that the probability of exceeding a particular threshold is maximal.

10.4.2. Second-phase sampling

In second-phase spatial sampling, a set M of m initial measurements has been collected, and a covariogram $C(h)$ has been calculated. In a second-phase, the scientist goes out to the field to augment the set of observations, guided by the covariogram. The objective function aims to collect new samples to reduce the kriging variance or uncertainty by as much as possible. Equation (10.18) formulates the change in kriging variance $\Delta\sigma_k^2$ over all grid points $\mathbf{s}_g$, when a set N of size n containing new sample points is added to our initial sample set M. The change $\Delta\sigma_k^2$ is the difference between the kriging variance calculated with initial sample points and the kriging variance of the augmented set $M \cup N$ containing $[m+n]$ samples:

$$\Delta\sigma_k^2 = \left[TKV^{\text{old}} - TKV^{\text{new}}\right] = \frac{1}{G}\left[\sum_{g\in G}\sigma_{k,\text{old}^2}(\mathbf{s}_g) - \sum_{g\in G}\sigma^2_{k,\text{new}}(\mathbf{s}_g)\right] \tag{10.18}$$

$$\sigma^2_{k,\text{old}}(\mathbf{s}_g) = \sigma^2 - \underbrace{c(\mathbf{s}_g)}_{[1,m]} \cdot \underbrace{\mathbf{C}^{-1}}_{[m]} \cdot \underbrace{\mathbf{c}^T(\mathbf{s}_g)}_{[m,1]} \tag{10.19}$$

$$\sigma^2_{k,\text{new}}(\mathbf{s}_g) = \sigma^2 - \underbrace{c(\mathbf{s}_g)}_{[1,m+n]} \cdot \underbrace{\mathbf{C}^{-1}}_{[m+n]} \cdot \underbrace{\mathbf{c}^T(\mathbf{s}_g)}_{[m+n,1]}. \tag{10.20}$$

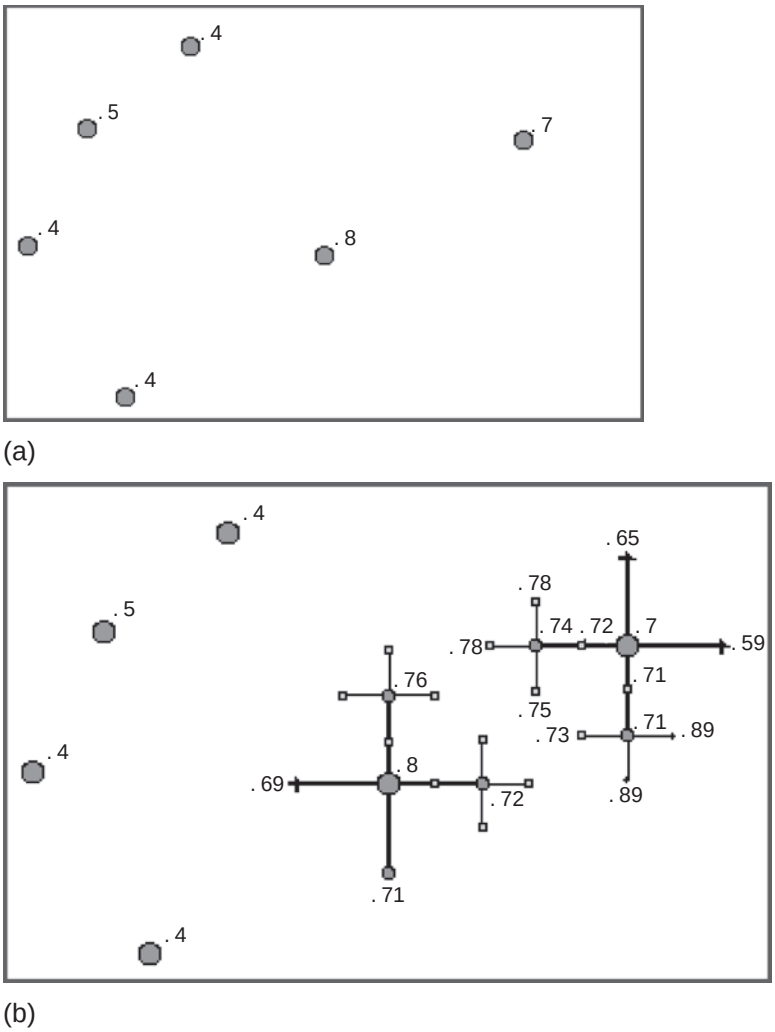
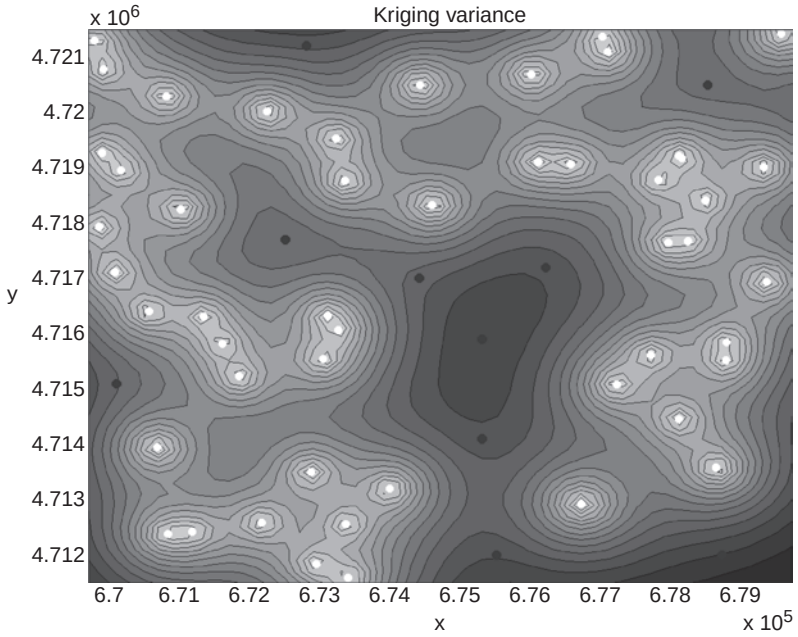


Figure 10.6 The cluster adaptive sampling procedure, illustrated in the context of toxic waste remediation. A site is fathomed (+) when its toxicity rate does not exceed the contamination value.

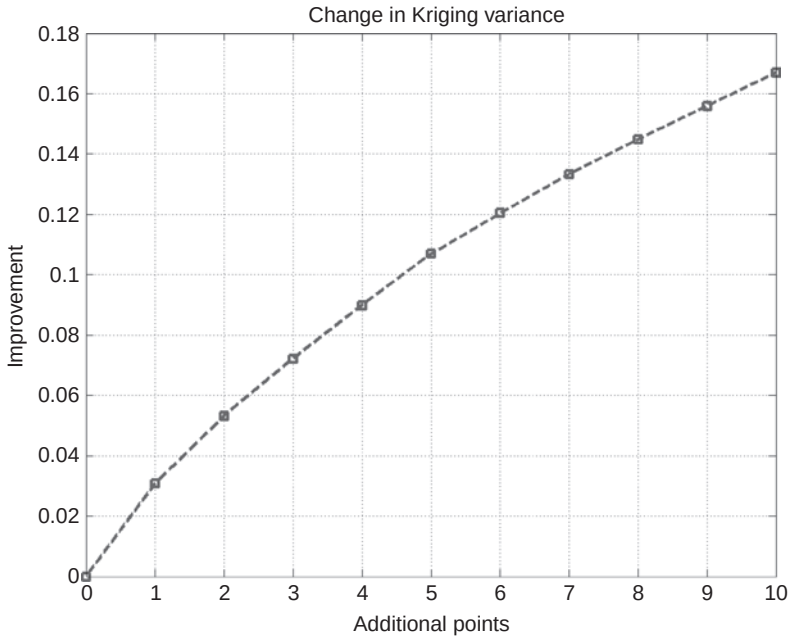
The objective function (equation (10.21)) is to find the optimal set S^* containing $m + n$ points that will maximize this change in kriging variance (Christakos and Olea, 1992; Van Groenigen *et al*., 1999), where S is a specific sampling scheme:

$$\underbrace{MAX}_{\{\mathbf{s}_{m+1},\ \ldots,\mathbf{s}_{m+n}\}} \quad J(S) = \frac{1}{G}\sum_{g\in G}\Delta\sigma_k^2(\mathbf{s}_g; S). \tag{10.21}$$

For simplicity, the continuous region D is usually approximated by a finite set P of p points (Cressie, 1991). The set of new points is selected from the set of potential points P. Hence, there is a total of $\binom{p}{n}$ possible sampling combinations and it is too time-consuming to find the optimal set using combinatorics. Figure 10.7 illustrates the case where 50 sample points have been collected in a first stage, leading to an



(a)



(b)

Figure 10.7 An initial sampling network of $m = 50$ points (in white) has been augmented with the addition of $n = 10$ new samples (in blue). The figure to the right displays the improvement.

exponential covariogram, with the sequential addition of $n = 10$ new points and an improvement in the objective function of nearly 20%.

Weighting the kriging variance? This use of a weighting function $w(\bullet)$ for the kriging variance was originally suggested by Cressie (1991) and has been applied by Van Groenigen *et al*., (2000), Rogerson *et al*., (2004), and Delmelle (2005). The importance of a location to be sampled is represented by a weight $w(\mathbf{s})$. The objective is to find the optimal sampling scheme S^* containing $m + n$ points that will maximize this change in weighted kriging variance. From equation (10.21):

$$\underbrace{MAX}_{\{\mathbf{s}_{m+1},\ldots,\mathbf{s}_{m+n}\}} \quad J(S) = \frac{1}{G}\sum_{g\in G} w(\mathbf{s}_g)\Delta\sigma_k^2(\mathbf{s}_g; S). \tag{10.22}$$

In an effort to detect contaminated zones in the Rotterdam harbor, Van Groenigen *et al*., (2000) introduced the ***W**eighted **M**eans of **S**hortest **D**istance* (WMSD) criterion, offering a flexible way of using prior knowledge on the variable under study. However, the weights do not reflect the spatial structure of the variable, but rather the scientist's perception of the risks of contamination. In the first sampling phase, sampling weights are assigned to sub-areas based on their risks for contamination. In the second phase however, a greater weight is assigned to locations expected to exhibit a higher priority for remediation. Four weighting factors are considered with weights $w = 1$, 1.5, 2, and 3, leading to more intensive sampling where the weight is higher. In a more recent study, Rogerson *et al*., (2004) have developed a second-phase sampling technique, allowing re-sampling in areas where there is some uncertainty associated with a variable of interest, and hence not in areas where the probability of an event occurring is near 0 or 1. A greedy algorithm was proposed to locate the points that would maximize the change in weighted kriging variance.

Shortcoming of the use of the kriging variance

Many authors have advocated the use of the kriging variance as a measure of uncertainty. It is unfortunately misused as a measure of reliability of the kriging estimate, as noted by several authors (Deutsch and Journel, 1997; Armstrong, 1994). It is solely a function of the sample pattern, sample density, the numbers of samples and their covariance structure. The kriging variance assumes that the errors are independent of each other. This means that the process is stationary, an assumption usually 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. Figure 10.8 illustrates non-stationarity in two dimensions (Armstrong, 1994). The objective in this particular example is to interpolate the value of the inner grid point, highlighted with a question mark. The interpolation depends on the values of the four surrounding points. Two scenarios are presented. The one in *b* shows three very similar values and an extreme one. The one in *a* however shows four values in a very narrow range. Assuming the spatial structure is similar in both cases and since the configuration of the data points is the same, the kriging variances are identical. However, we have more confidence in the left-hand side scenario since there is less variation among the neighbors. This illustrates that the prediction error is not suitable for setting up confidence intervals and should not be used as an optimization criterion for additional sampling strategies.

10.5. CURRENT RESEARCH DIRECTIONS

10.5.1. Incorporating multivariate information

Sample data can be very difficult to collect, and very expensive, especially in monitoring air or soil pollution for instance (Haining,

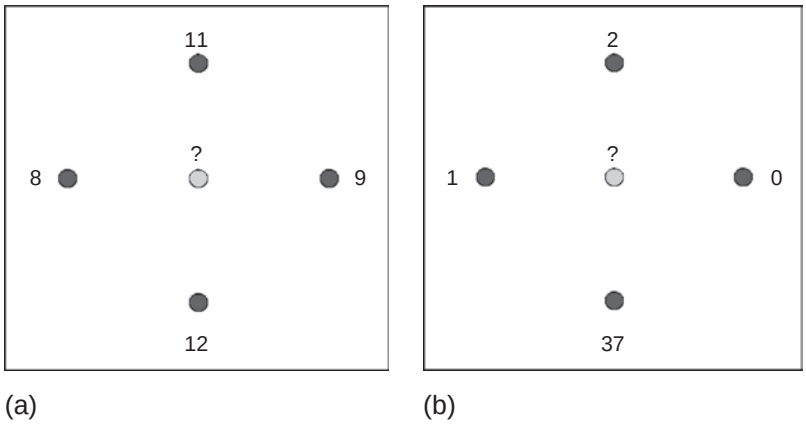


Figure 10.8 Example of two-dimensional non-stationarity. Dark points are used as data values to interpolate the center point (light gray). After Armstrong (1994).

2003). Secondary data can be a valuable asset if they are available all over a study area and combined within the primary variable (Hengl *et al*., 2003). Secondary spatial data sources include maps, national, socioeconomic, and demographic census data, but also data generated by public sources (local and regional). This is very valuable and there had been a dramatic growth in the availability of secondary data associated with DEMs and satellites (for environmental data). Such secondary data is easily integrated within a GIS framework (Haining, 2003). In multi-phase sampling, for instance, research has been confined to the use of covariates in determining the locations of initial measurements, whereby sample concentration is increased where covariates exhibit substantial spatial variation (Makarovic, 1973). Ideally, secondary variables should be used to reduce the sampling effort in areas where their local contribution in predicting the primary variable is maximum (Delmelle, 2005). If a set of covariates predicts accurately the data value where no initial sample has been collected yet, there is little incentive to perform sampling at that location. On the other hand, when covariates perform poorly in estimating the primary variable, additional samples may be necessary. The general issue pertains to quantifying the spatial contribution given by covariates.

10.5.2. Weighting the kriging variance appropriately

Some current research has looked at ways to weight the kriging variance. Intuitively, one would like to sample at unvisited locations, far away from existing ones. This is accomplished using the kriging variance as a sampling criterion. However, the spatial variability of the primary variable is not accounted for. It is recommended to weight the kriging variance where the gradient of the primary variable is maximum, i.e., where contour lines come close to one another, because there is a rapid change in the variable (Delmelle, 2005). It is also desirable to reduce sampling effort by using information provided by auxiliary variables, when available.

10.5.3. The use of heuristics in sampling optimization

In second-phase sampling, the set N of additional samples will be chosen from

a set P of candidate sampling locations. This set is relatively large in practice, and hence the number of possible solutions forbids an exhaustive search for the best answer (Michalewicz and Fogel, 2000). A total enumeration of all the solutions ($\approx$ naive approach) is not possible, because of the combinatorial explosion. Suppose, for instance, that there is a potential set of $N = 900$ points and that we are willing to take $n = 100$ additional points within that set. This generates $\binom{900}{100}$ 900 $100 \approx 9.384 \times 10^{134}$ solutions, and hence the use of a naive approach is not recommended (Goldberg, 1989; Grötschel and Lovàsz, 1995). The search for an approximate solution of complex problems is conducted using a suitable heuristic method H. The use of a heuristic is necessary to assist in the identification of an optimal sample set S^* (or near optimal S^+) $\subset P$. The heuristic controls a process that intends to solve this optimization problem. The set S^* is optimal to the objective function J defined in equation (10.22). The efficiency of a heuristic depends on its capacity to give as often as possible a solution S^+ close to S^* (Grötschel and Lovàsz, 1995). In second-phase sampling, there are two different ways of supplementing an initial set. Either n points are selected at one time and added to the initial set all together, or one point at a time is added n-times to the initial set. The former is defined as simultaneous addition and the latter is known as sequential addition and is suboptimal. Note that a hybrid approach that would combine both techniques is possible as well. In spatial sampling, limited research has been devoted to comparing the benefits and drawbacks of these heuristics. The greedy (or myopic) algorithm has been used by Aspie and Barnes (1990), Christakos and Olea (1992) and Rogerson *et al*., (2004). Simulated annealing has been applied to spatial sampling problems in Ferri and Piccioni (1992), Van Groenigen and Stein (1998), and Pardo-Igúzquiza (1998).

10.5.4. Spatio-temporal sampling issues

Spatial sampling optimization as discussed in this chapter is based on the assumption of stationarity of the variable itself over time ($\approx$ no temporal variation). Variables such as rainfall, temperature, and snowfall vary over time and it is not possible to take a second set of samples to improve the prediction of these variables without affecting the stability of the model. Work in this context has been carried out by Lajaunie *et al*., (1999).

NOTE

1 Note that a *systematic sampling scheme* is a special case of a *stratified design* in that the strata are all squares of equal size.

REFERENCES

Arbia, G. (1994). Selection techniques in sampling spatial units. *Quaderni di Statistica e Matematica Applicata Alle Scienze Economico-Sociali*, **16**: 81–91.

Armstrong, M. (1994). Is research in mining geostats as dead as dodo? In: Dimitrakopoulos R. (ed.). *Geostatistics for the Next Century*, pp. 303–312. Dordrecht: Kluwer Academic Publisher.

Aubry, P. (2000). Le Traitement des Variables Régionalisées en Ecologie: Apports de la Géomatique et de la Géostatistique}. Thèse de doctorat. Université Claude Bernard – Lyon 1.

Aspie, D. and Barnes, R.J. (1990). Infill-sampling design and the cost of classification errors. *Mathematical Geology*, **22**: 915–932.

Ayeni, O. (1982). Optimum sampling for digital terrain models: A trend towards automation. *Photogrammetric Engineering and Remote Sensing*, **48**: 1687–1694.

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

Bellhouse, D.R. (1977). Optimal designs for sampling in two dimensions. *Biometrika*, **64**: 605–611.

Berry, B.J.L. and Baker, A.M. (1968). Geographic sampling. In: Berry B.J.L. and Marble D.F. (eds),

Spatial Analysis: a Reader in Statistical Geography; pp. 91–100. Prentice-Hall: Englewood Cliffs, N.J.

Bogaert, P. and Russo, D. (1999). Optimal sampling design for the estimation of the variogram based on a least squares approach. *Water Resources Research*, **35**(4): 1275–1289.

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

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

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

Corsten, L.C.A. and Stein, A. (1994). Nested sampling for estimating spatial semivariograms compared to other designs. *Applied Stochastic Models and Data Analysis*, **10**: 103–122.

Cox, L.A. (1999). Adaptive spatial sampling of contaminated soil. *Risk Analysis*, **19**: 1059–1069.

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

Dalenius, T., Hajek, J. and Zubrzycki, S. (1960). On plane sampling and related geometrical problems. *Proceedings of the Fourth Berkeley Symposium*, **1**: 125–150.

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

Delmelle, E.M. (2005). *Optimization of Second-Phase Spatial Sampling Using Auxiliary Information*. Ph.D. Dissertation, Department of Geography, SUNY at Buffalo.

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

Ferreya, R.A., Apezteguía, H.P., Sereno, R. and Jones, J.W. (2002). Reduction of soil water sampling density using scaled semivariograms and simulated annealing. *Geoderma*, **110**: 265–289.

Ferri, M. and Piccioni, M. (1992). Optimal selection of statistical units. Computational Statistics and Data Analysis, **13**: 47–61.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press. 483p.

Gatrell, A.C. (1979). Autocorrelation in spaces. *Environmental and Planning A*, **11**: 507–516.

Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley.

Griffith, D. (1987). *Spatial Autocorrelation: A Primer*. Washington, DC: AAG.

Griffith, D. and Amrhein, C. (1997). *Multivariate Statistical Analysis for Geographers*. New Jersey: Prentice Hall. 345p.

Grötschel, M. and Lovàsz, L. (1995). Combinatorial optimization. In: Graham R.L., Grötschel and Lovàsz (eds), *Handbook of Combinatorics*, Vol. 2; pp. 1541–1579. Amsterdam, The Netherlands: Elsevier.

Haining, R.P. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge University Press. 452p.

Hedayat, A.S. and Sinha, B.K. (1991). *Design and Inference in Finite Population Sampling*. New York: Wiley. 377p.

Hengl, T., Rossiter, D.G. and Stein, A. (2003). Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Australian Journal of Soil Research*, **41**: 1403–1422.

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

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

King, L.J. (1969). *Statistical Analysis in Geography*. 288p.

Lajaunie, C., Wackernagel, H., Thiéry, L. and Grzebyk, M. (1999). Sampling multiphase noise exposure time series. In: Soares A., Gomez-Hernandez J. and R. Froidevaux (eds), *GeoENV II – Geostatistics for Environmental Applications*; pp. 101–112. Dordrecht: Kluwer Academic Publishers.

Madow, W.G. (1953). On the theory of systematic sampling. III. Comparison of centered and random start systematic sampling. *Annals of Mathematical Statistics*, **24**: 101–106.

Makarovic, B. (1973). Progressive sampling for digital terrain models. *ITC Journal*, **15**: 397–416.

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

McBratney, A.B. and Webster, R. (1981). The design of optimal sampling schemes for local estimation and

mapping of regionalized variables: II. Program and examples. *Computers and Geosciences*, **7**: 331–334.

McBratney, A.B., Webster, R. and Burgess, T.M. (1981). The design of optimal sampling schemes for local estimation and mapping of regionalized variables: I. Theory and method. *Computers and Geosciences*, **7**: 335–365.

McBratney, A.B. and Webster, R. (1986). Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. *Journal of Soil Science*, **37**: 617–639.

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

Moran, P.A.P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society, Series B*, **10**: 245–251.

Moran, P.A.P. (1950). Notes on continuous phenomena. *Biometrika*, **37**: 17–23.

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

Olea, R.A. (1984). Sampling design optimization for spatial functions. *Mathematical Geology*, **16**: 369–392.

Oliver, M.A. and Webster, R. (1986). Combining nested and linear sampling for determining the scale and form of spatial variation of regionalized variables. *Geographical Analysis*, **18**: 227–242.

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

Pardo-Igúzquiza, E. (1998). Optimal; selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *Journal of Hydrology*, **210**: 206–220.

Pettitt, A.N. and McBratney, A.B. (1993). Sampling designs for estimating spatial variance components. *Applied Statistics*, **42**: 185–209.

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

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

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

Stehman, S.V. and Overton, S.W. (1996). Spatial sampling. In: Arlinghaus, S. (ed.) *Practical Handbook of Spatial Statistics*; pp. 31–64. Boca Raton, FL: CRC Press.

Thompson, S.K. and Seber, G.A.F. (1996). *Adaptive Sampling*. New York: Wiley. 288p.

Van Groenigen, J.W. (1997). Spatial simulated annealing for optimizing sampling – different optimization criteria compared. In: Soares, A., Gómez-Hernández, J. and Froidevaux, R. (eds). *GeoENV I – Geostatistics for Environmental Applications*. Dordrecht: Kluwer Academic Publishers.

Van Groenigen, J.W. (2000). The influence of variogram parameters on optimal sampling schemes for mapping by kriging. *Geoderma*, **97**: 223–236.

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

Van Groenigen, J.W., Siderius, W. and Stein, A. (1999). Constrained optimisation of soil sampling for minimization of the kriging variance. *Geoderma*, **87**: 239–259.

Van Groenigen, J.W., Pieters, G. and Stein, A. (2000). Optimizing spatial sampling for multivariate contamination in urban areas. *Environmetrics*, **11**: 227–244.

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

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

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

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