ELSEVIER

Geoderma 105 (2002) 49–80

GEODERMA

www.elsevier.com/locate/geoderma

Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood

R.M. Lark *

Mathematics and Decision Systems Group, Silsoe Research Institute, Wrest Park, Silsoe, Bedford, MK45 4HS, UK

Received 29 September 2000; accepted 13 June 2001

Abstract

Recent studies have attempted to optimize the configuration of sample sites for estimation of the variogram by the usual method-of-moments. This paper shows that objective functions can readily be defined for estimation by the method of maximum likelihood. In both cases an objective function can only be defined for a specified variogram so some prior knowledge about the spatial variation of the property of interest is necessary.

This paper describes the principles of the method, using Spatial Simulated Annealing for optimization, and applies optimized sample designs to simulated data. For practical applications it seems that the most fruitful way of using the technique is for supplementing simple systematic designs that provide an initial estimate of the variogram.

Keywords: Spatial variation; Geostatistics; Soil sampling; Optimization

1. Introduction

1.1. The variogram and its estimation

Geostatistics has been applied widely in soil science to solve the problem of estimating soil properties at unvisited sites from limited sample data (Webster and Oliver, 2000). Central to any geostatistical analysis is the variogram, $\gamma(\mathbf{h})$, which describes the spatial dependence of a random function that is assumed to

* Fax: +44-1525-860156.

E-mail address: murray.lark@bbsrc.ac.uk (R.M. Lark).

0016-7061/02/$ - see front matter
PII: S0016-7061(01)00092-1

be realized in our soil variable. The variogram must be obtained from sample data.

Once a variogram of a soil property has been obtained it may be used to design an optimized sample grid for mapping the variable (McBratney et al., 1981), for optimal estimation of the variable at a point or over a block (Burgess and Webster, 1980), for simulating random fields with the same spatial properties as the variable (Papritz and Webster, 1995), for understanding the spatial structure of the variable (Oliver, 1999) and for exploring the scaling properties of models for which the variable is an input (Lark, 2000a). For this reason much attention has been focussed on the problems of estimating the variogram.

In most soil studies the variogram is obtained by a method of moments estimator (Matheron, 1962). This is applied to the $N(\mathbf{h})$ pairs of observations that are separated by the lag $\mathbf{h}$. The estimate of the variogram at this lag is given by

$$\hat{\gamma}(h) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} \{z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})\}^2. \quad (1)$$

An appropriate parametric model of the variogram function is then fitted to the point estimates. Various aspects of the method-of-moments estimator have been studied by soil scientists, including the sample requirements (Webster and Oliver, 1992) and the effects of outlying data (Lark, 2000b).

An alternative to the method-of-moments is the maximum likelihood estimator. This treats a set of n data as a realization of an n-variate Gaussian process with a covariance matrix that may be written in terms of the parameters of the variogram that we wish to estimate (Pardo-Iguzquiza and Dowd, 1998). This method has its advantages and disadvantages, which are summarised by Lark (2000c).

1.2. Sampling to estimate the variogram—scope for optimization

How should we sample a region in order to estimate the variogram? The first point to note is that sampling does not have to be random. This is because the statistical inference of the variogram by either of the above methods uses model-based statistics rather than design-based methods (although the method-of-moments can be used in design-based analysis, see Brus and de Gruijter, 1994). In model-based statistics the assumption of randomness attaches to the realizations of the random function—of which we assume the data are one—and not to the inclusion or exclusion of a particular site from the sample (that is the assumption in design-based statistics).

In practice scientists often use some form of systematic sampling to estimate the variogram—either on grids or on transects. Such designs are simple to implement, and provide a basis for the final sample grid to be used for mapping

by kriging. In some circumstances, however, we may wish to make an initial reconnaissance survey of an area that will provide information on its spatial variability to plan more detailed sampling. In this situation the estimation of the variogram is a key concern because it can be used to design a sample grid such that the error of the kriged estimates is limited to acceptable values (McBratney et al., 1981). The question therefore of how best to sample to estimate the variogram arises.

It is evident that some configurations of sample points will be better for estimating the variogram than others. If all pair-comparisons among a data set are over lag intervals longer than the range of spatial dependence, for example, then spatial dependence will not be resolved. Intuitively it is clear that we require a sample which permits comparisons over lag intervals large, small and intermediate relative to the scale of spatial variation. Viloria (1989) showed this empirically. It is also known that we cannot obtain good estimates of the variogram for lags larger than half the distance across the sampling grid. Such considerations are the basis of rules-of-thumb to be used when planning sampling (e.g. Webster and Oliver, 2000). They also prompt the question of whether the disposition of sampling points can be optimized for the purpose of estimating the variogram so that a relatively small number of costly samples can be used most efficiently. Some attention has been paid to this problem, and the literature is now briefly reviewed.

Optimizing a sample lay-out requires some objective function to be minimized or maximized. Various alternative objective functions have been proposed for sampling for the variogram (all assuming that the method-of-moments estimator is to be used). The earlier criteria were based on practical considerations rather than a rigorous analysis of the estimation problem. Russo (1984) noted that in practice the estimator in Eq. (1) is not usually applied to $N(\mathbf{h})$ pairs of observations separated by an exact lag $\mathbf{h}$, but rather to a broader 'lag class'—observations separated by lags within a range of distances and bearings about the central lag $\mathbf{h}$. Russo (1984) proposed that the variation of lag intervals within all the lag-classes should be minimized to improve the precision of the final variogram model. This will reduce the smoothing effect of estimating the variogram for lag classes rather than exact lag intervals.

Warrick and Myers (1987) suggested another practical criterion for optimizing the distribution of sampling sites. They proposed that the frequency distribution of lag distances between observations in the sample should be controlled by maximizing its conformity to a prespecified function. Bresler and Green (1982) conjectured that a uniform distribution of lag distances is to be preferred, i.e. the values of $N(\mathbf{h})$ should be as similar as possible over all central lags $\mathbf{h}$. However, van Groenigen (1999) found that sample arrays that optimized the Warrick–Myers criterion for a uniform distribution of lag distances gave less precise estimates of the variogram than regular sample grids with the same number of point and Müller and Zimmerman (1999) also cast doubt on the criterion. Since

it is known (e.g. Stein, 1988) that the error in kriging estimates is most sensitive to error in the variogram at short distances, the a priori basis for Bresler and Green's (1982) criterion is not clear.

One interesting study on sampling design for spatial analysis is that of Pettitt and McBratney (1983). They showed how transects with exponentially distributed spacings were most efficient for estimating spatial variance components. This analysis is rather different to conventional methods of variogram estimation, although monotonically increasing variograms (i.e. those without periodic components) can be estimated from spatial variance components.

Two recent studies have attempted to identify a rigorous criterion for optimizing a sample design for estimating the variogram by the method-of-moments. Bogaert and Russo (1999) obtained an approximation to the covariance matrix of the parameters of a variogram model fitted by generalized least-squares to the method-of-moments estimates. This was based on a first-order approximation to linearize the variogram model with respect to its parameters. The determinant of the covariance matrix was then treated as the objective function to be minimized with respect to the location of the sample points by a special case of simulated annealing. Bogaert and Russo (1999) showed by simulation that their sample designs had advantages over alternatives, particularly when applied to realizations of random functions with a nugget effect.

Since the variogram is a non-linear function, the covariance matrix of its parameters for any given sample array depends on the parameters themselves. Therefore, we can only strictly optimize the sample array when we know the variogram parameters in advance. Bogaert and Russo (1999) suggested that a Bayesian solution to this problem be sought, perhaps by specifying a prior distribution for the variogram parameters which is uniform between upper and lower bounds.

Müller and Zimmerman (1999) developed a similar approach based on finding the information matrix, $\mathbf{I}$, corresponding to the generalized least-squares solution for the parameters of a variogram model fitted to point estimates. They envisaged a situation in which an initial sample (e.g. a grid or transects) is supplemented by additional sites located to maximize the determinant $|\mathbf{I}|$. An approximation to the variogram may be estimated from the initial sampling points and used to find the information matrix for any sample array. Optimized sample configurations were found to out-perform both random and systematic sampling of simulated data.

1.3. Optimizing sampling for estimation by maximum likelihood

The studies reviewed above all assumed that the method-of-moments is used to estimate the variogram from data. Maximum likelihood is an alternative to the method-of-moments for estimating the variogram. This paper studies the problem of optimizing sample arrays for this method of estimation for two reasons.

First, there is interest among environmental scientists in the method of maximum likelihood for variograms because of advantages both practical and theoretical that have been claimed for it (Pardo-Iguzquiza and Dowd, 1998; Diggle and Ribeiro, 1999). Lark (2000c) reviews the arguments. Second, the method of maximum likelihood lends itself to the optimization problem because the covariance matrix of the variogram parameters can be obtained simply and directly for a specified model and sample configuration. This is not the case with the method-of-moments, where an approximation to the covariance matrix has to be found.

The next section of this paper sets out the theory of optimizing the sampling configuration of estimating the variogram by maximum likelihood. Spatial simulated annealing (van Groenigen, 1999) was used as the optimization method. Some critical questions about the feasibility of the method are then addressed using simulation.

2. Theory

2.1. The objective function

Consider a set of n observations $\mathbf{z} = \{z(\mathbf{x}_1), z(\mathbf{x}_2), \ldots, z(\mathbf{x}_n)\}$ When inferring the variogram by maximum likelihood we treat $\mathbf{z}$ as a realization of an n-variate Gaussian process with variance σ^2 and with a covariance matrix $\mathbf{V}$ which may be factored

$$\mathbf{V} = \sigma^2 \mathbf{A}, \tag{2}$$

where $\mathbf{A}$ is the correlation matrix. Any element of $\mathbf{A}$, $\{i, j\}$ may be written in terms of the parameters of the underlying variogram and the lag which separates the two locations $\mathbf{x}_i$ and $\mathbf{x}_j$. Our goal is to estimate the vector $\boldsymbol{\theta}$ of p variogram parameters $\theta_1, \theta_2, \ldots, \theta_p$.

A simple variogram function is the isotropic exponential model with a nugget effect:

$$\gamma(h) = c_0 + c_1\left(1 - \mathrm{e}^{-\frac{h}{a}}\right), \tag{3}$$

where h is a lag distance, c_0 and c_1 are variance components and a is a distance parameter. The vector $\boldsymbol{\theta}$ for this variogram contains two parameters which can be used to define the correlation matrix $\mathbf{A}$. These are the distance parameter a and the ratio of spatial dependence

$$s = \frac{c_1}{c_0 + c_1}, \tag{4}$$

since $c_0 + c_1$ equals the a priori variance σ^2.

A negative log-likelihood function can then be written with respect to σ^2, $\boldsymbol{\theta}$ and *either* an estimate of the mean of $\mathbf{z}$, β, *or* $\boldsymbol{\beta}$—a vector of drift parameters. Estimates of these parameters may then be obtained by minimizing the negative log-likelihood function numerically. Details of the procedure are given by Pardo-Iguzquiza and Dowd (1998) and by Lark (2000c). Pardo-Iguzquiza (1997) provides Fortran code. Residual Maximum Likelihood (REML) may be preferred over ordinary maximum likelihood in some circumstances, and is strongly recommended when drift parameters need to be estimated. It is not considered further in this paper.

The maximum likelihood estimates of the variogram parameters in $\boldsymbol{\theta}$ have uncertainty that is quantified by a covariance matrix. This covariance matrix can be obtained from the estimation theory (Pardo-Iguzquiza, 1997). The variance of a maximum likelihood estimate of one of the variogram parameters, $\hat{\theta}$, achieves the lower bound set by the Cramér–Rao inequality:

$$\mathrm{var}(\hat{\theta}) \geq [\mathbf{I}(\theta)]^{-1}, \tag{5}$$

where $\mathbf{I}$ is the Fisher information matrix:

$$\mathbf{I}(\theta) = -E\left[\frac{\partial^2 L}{\partial\theta\partial\theta'}\right], \tag{6}$$

L is the log-likelihood function. The term in square brackets in Eq. (6) is the Hessian of the log-likelihood function that is maximized in the process of estimation. The covariance matrix may be obtained by inverting the information matrix.

Rather than obtaining the information matrix via the Hessian in Eq. (6), Kitanidis (1987) showed that each element, $[\mathbf{I}(\theta)]_{ij}$, could be obtained with the following equation:

$$[\mathbf{I}(\theta)]_{ij} = \frac{1}{2}\mathrm{Tr}\left[\mathbf{A}^{-1}\mathbf{A}_i\mathbf{A}^{-1}\mathbf{A}_j\right], \tag{7}$$

where $\mathbf{A}_i \equiv (\partial/\partial\theta_i)\mathbf{A}$ and $\mathrm{Tr}[\cdot]$ denotes the trace of the matrix in brackets.

For a given variogram and set of sample points at known locations an information matrix for the maximum likelihood estimates of the variogram parameters may be obtained using Eq. (7) and, by inversion, their covariance matrix $\mathbf{C}(\theta)$. Since this covariance matrix quantifies our uncertainty about the estimated variogram parameters, and depends on the spatial distribution of the sample data through the correlation matrix $\mathbf{A}$, it would seem to be a natural starting point for defining an objective function for variogram estimation by maximum likelihood.

The choice of an objective function will depend on the goal of the sampling. If we are concerned primarily with estimating the correlation scale of the variable of interest, then the variance of the distance parameter a could be extracted from $\mathbf{C}(\theta)$ as the objective function to be minimized, but it is hard to

imagine a practical situation in which this would be of most concern. A generalized measure of uncertainty about $\boldsymbol{\theta}$ is the determinant of $\mathbf{C}(\theta)$ as used by Bogaert and Russo (1999) and Müller and Zimmerman (1999)—the latter actually maximized the determinant of $\mathbf{I}(\theta)$. The interpretation of this criterion raises questions, however, since it is a combination of variances and covariances of variables with different units.

The objective function considered in this paper is the precision of the estimated kriging variance calculated from the estimate of the variogram for a specified distribution of sample sites around a target site. The kriging variance is calculated by

$$2\sum_{i=1}^{n}\lambda_i\gamma(\mathbf{x}_i-\mathbf{x}_0)-\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\gamma(\mathbf{x}_i-\mathbf{x}_j), \tag{8}$$

where $\mathbf{x}_0$ is the target site for kriging and the $\mathbf{x}_i$ $(i=1,\ldots,n)$ are neighbouring sites used to derive the kriged estimate at $\mathbf{x}_0$ with kriging weights λ_i. The kriging variance is seen to be a linear combination of values of the variogram, so will depend only on the variogram and the relative positions of all the vectors $\mathbf{x}_0$, $\mathbf{x}_i$ $(i=1,\ldots,n)$. It is therefore possible, given a variogram from a reconnaissance survey, to find the most cost-effective sample grid for kriging such that the kriging variance is acceptably small (McBratney et al., 1981; Oliver and Webster, 1987).

Fig. 1 shows how the kriging variance at the centre of a unit interval square grid depends on the distance parameter and ratio of spatial dependence for a

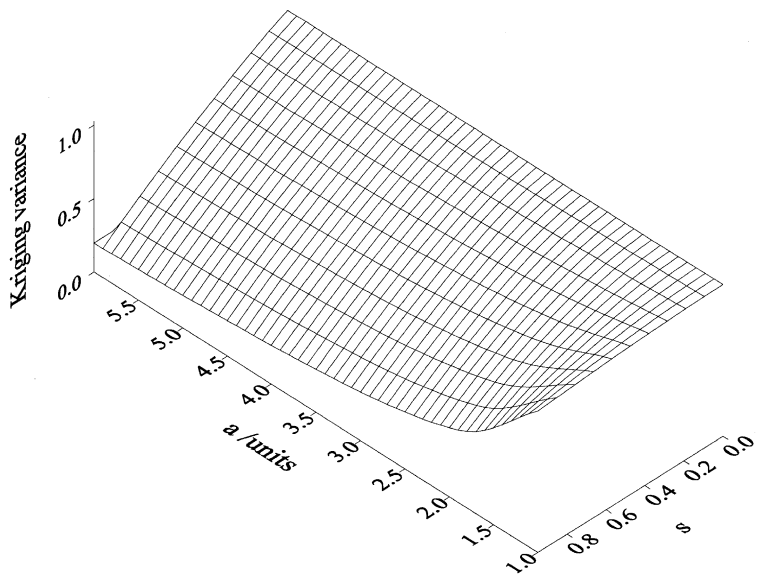

Fig. 1. Kriging variance at the centre of a unit-interval square grid as the function of distance parameter and ratio of spatial dependence for a spherical variogram of unit a priori variance.

process with a spherical variogram and unit a priori variance. It is clear from this figure that errors in the variogram parameters will propagate through Eq. (8) to errors in the estimate of the kriging variance.

Two reasons are given for using the precision of the estimate of the kriging variance as the objective function for optimizing sampling for the variogram.

First, when the objective of a reconnaissance survey to estimate the variogram is to allow the design of the final sample grid so as to achieve a target precision, using the techniques of McBratney et al. (1981), maximising the precision of the estimates of the kriging variance will be of direct practical significance. Fig. 2 shows the result of a simulation. A realization of a random process, with a spherical variogram with unit a priori variance, a distance parameter of 4 units and a ratio of spatial dependence of 0.7, was sampled without replacement 100 times, each sample comprising 120 sites on randomly located regular transects of unit spacing. The variogram was then estimated from each sample by the method of maximum likelihood, then used to calculate the spacing of a square grid required such that the kriging variance at the centre of a grid cell is 0.7. The figure shows the histogram of grid spacings. The variation reflects sample error of the variogram parameters. The range of grid spacings is from less than 0.5 to more than 3.5 units of distance. The first and third quartiles of these spacings correspond to sampling intensities of 0.44 and 0.17 samples per square unit of distance, respectively. Sampling variation of this order, a factor of more than 2, is not insignificant when the costs of sample collection and analysis over a large area are considered, so minimizing this variation is a worthwhile objective.

The second reason for selecting this objective function is that the kriging variance is known to be sensitive to errors in the variogram, as shown in Fig. 1.

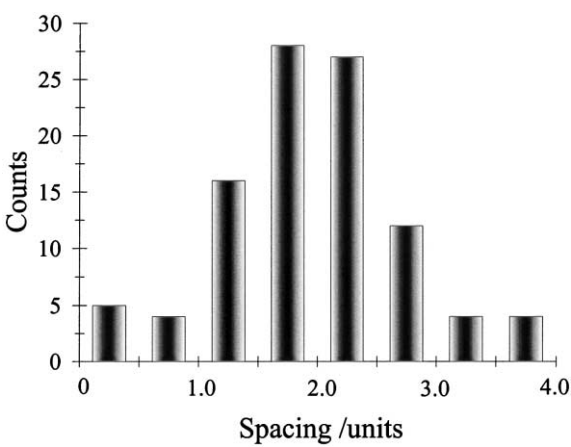

Fig. 2. Histogram of inferred intervals for a square grid to achieve a kriging variance of 0.7. The intervals are inferred from 100 sample variograms each estimated by maximum likehood from 120 observations on transects.

This makes the kriging variance a useful overall measure of the correctness of the variogram, which is why it is used for this purpose when evaluating variograms by cross-validation or jacknifing (e.g. Lark 2000b,c; Geiler et al.,1997; Voltz and Webster, 1990).

In this paper I consider the point-kriging variance at the centre of a square grid of specified interval, which is obtained from an estimate of the variogram and denoted σ_K^2. I seek to minimize the error in this quantity due to sample error in the variogram parameters in $\boldsymbol{\theta}$. The error variance of σ_K^2 may be written, using a first-order Taylor series approximation as

$$V_e\left(\sigma_K^2\right) = \sum_{i=1}^{p}\sum_{j=1}^{p}\rho_{i,j}\sigma_i\sigma_j\frac{\partial\sigma_K^2}{\partial\theta_i}\frac{\partial\sigma_K^2}{\partial\theta_j}, \tag{9}$$

where $\rho_{i,j}$, σ_i and σ_j are the correlation coefficient and standard deviations, respectively, of variogram parameters θ_i and θ_j, which can be calculated directly from $\mathbf{C}(\theta)$, and so which depend, for given variogram parameters, on the position of the sample sites. The two partial derivatives in Eq. (9) must be obtained numerically. More detail on how this was done is given in the Section 3.1.1.

2.2. Optimization

Having defined an objective function, the challenge is to find the configuration of sample points that will minimize it. The technique that was used here is simulated spatial annealing as developed by van Groenigen (1999).

Simulated annealing is a method of combinatorial optimization. Its name derives from an analogy with the annealing process by which a molten metal, on slow cooling, tends to a global energetic minimum—a perfect crystalline solid. Simulated annealing starts with an objective function $\phi(s)$ of some state of the system, s. Let the state of the system be s_i at time i on a discrete time scale. A proposed state for time $i+1$ is obtained by a random perturbation of the system. If this improves the objective function (reduces it in a minimization problem such as this), then the perturbation is accepted. If the objective function deteriorates under the proposed change (increases), then the acceptance or rejection of the proposed state s_{i+1} is decided by a random process. This is the basis of the Metropolis algorithm (see Aarts and Korst, 1989) and may be written as

$$P_\tau(s_i \rightarrow s_{i+1}) = \begin{cases} 1 & \phi(s_{i+1}) \leq \phi(s_i) \\ \exp\left\{\dfrac{\phi(s_i) - \phi(s_{i+1})}{\tau}\right\} & \phi(s_{i+1}) > \phi(s_i). \end{cases} \tag{10}$$

$P_\tau(s_i \rightarrow s_{i+1})$ denotes the probability that the proposed transition is accepted. A decision in any instance is then made by obtaining a realization of a random

binary variable $d \in \{0,1\}$ with expectation $E[d] = P_\tau(s_i \rightarrow s_{i+1})$. If $d = 1$ the proposed perturbation is adopted, otherwise the system remains unchanged over this time interval.

The constant τ in Eq. (10) is of particular importance. Making τ smaller reduces the probability, ceteris paribus, that a given transition $s_i \rightarrow s_{i+1}$ will be accepted. If the second line of Eq. (10) is compared to the Boltzmann distribution function, then it is seen that τ is analogous to the temperature of a system. Reducing τ is therefore a 'cooling' step. Note that the symbol c is usually used for τ in the literature on simulated annealing. It is not used here to avoid confusion with the variance components of variograms.

The basic principle of simulated annealing is to let a system undergo a chain of transitions with random perturbations that are accepted or rejected according to Eq. (10). At fixed interval τ is reduced according to a cooling schedule.

If the cooling schedule is well-chosen, then the system is expected to converge to a value of $\phi(s)$ close to the global minimum. Selecting a good schedule for a particular type of problem relies to some degree on trial and error. Cooling the system too rapidly will cause convergence to a poor local optimum (like 'quenching' in the analogous case of a cooling metal). Too slow a schedule prevents convergence within an acceptable time period.

Spatial Simulated Annealing (SSA) is a special case of simulated annealing developed by van Groenigen (1999) and shown to be useful for optimizing various geostatistical functions whose arguments are the co-ordinates of a set of points. It should not be confused with other instances in geostatistics where the simulated annealing principle is used for optimization (e.g. in fitting variogram models, (Pardo-Iguzquiza, 1997) or simulating random fields (Deutsch and Journel, 1992)). The key features of SSA are as follows.

(i) The cooling factor τ is reduced by a factor α_τ at the end of each fixed-length chain of transitions. Thus:

$$\tau_{k+1} = \alpha_\tau \tau_k, \tag{11}$$

where $0 < \alpha_\tau < 1$. The value of α_τ is set to achieve a specified final value τ from a specified initial value τ_0. The value of τ_0 is chosen so that 95% or more of the initial perturbations of the system are accepted.

(ii) In SSA a state is defined by the Cartesian co-ordinates of M sample points in space constrained to lie within a region $\mathscr{R}$. A single transition $s_i \rightarrow s_{i+1}$ is achieved by a random perturbation of the co-ordinates of one of these points, and each point is perturbed in turn. In two-dimensions, this is done by selecting a random direction, then moving the point h units in this direction where h is a random variable uniformly distributed on the closed interval $[0, h_{\max}]$. If this perturbation takes the point to a location outside the region to be sampled (or to an inaccessible location within the region such as a pond or a building), then the point is returned to its original location and a new perturbation is generated at random. The initial value of $h_{\max}$ was set to half the length

of the region to be sampled. The value of $h_{\max}$ is then reduced at each cooling step by a factor so that

$$h_{\max_{k+1}} = \alpha_h h_{\max_k}, \tag{12}$$

where $0 < \alpha_h < 1$. Van Groenigen (1999) found that this procedure increases the efficiency of SSA, since as we approach a global optimum it becomes less and less likely that a large perturbation will improve the objective function.

2.3. Questions to be addressed empirically

A method has been proposed for optimizing the sampling configuration for estimating the variogram by maximum likelihood. In the remainder of this paper I use simulation to address some questions about its usefulness. These are as follows.

(i) How does the performance of the optimized sampling design compare to that of a regular sample arrays for realizations of a random function with a known variogram?

(ii) How do the properties of an optimized sampling configuration vary with the parameters of the specified variogram?

(iii) How sensitive is the performance of a configuration of sampling points to the parameters of the specified variogram?

(iv) Does optimizing the average of the objective function computed for several different variograms produce a sample configuration with advantages over systematic sampling for a random function with a variogram similar to those used in the optimization? This might be one way of avoiding the difficulty that the optimum sample array can only be found when we know the values of the parameters that we wish to estimate.

(v) Does a two-phased sampling scheme in which an initial systematic sample is supplemented with additional points, whose location is optimized in accordance with an initial estimate of the variogram, have advantages over putting the same overall sampling effort into a single systematic sample?

3. Materials and methods

3.1. Basic methodology

Five experiments were carried out, using a basic set of methods which are described first.

3.1.1. Optimization

The SSA was implemented as described in Section 2.2 above to generate a sample array. A number, N, of labelled sample points were dispersed at random

within a square region $\mathscr{R}$ with sides of length 50 units. The possible co-ordinates for a point within this space are effectively continuous and there were no absolute constraints on the location of points within the area. To avoid degenerate solutions in which pairs of points become vanishingly close (van Groenigen, 1999), a minimum distance of 1 unit between any two points was enforced.

The initial value of the temperature variable τ was adjusted by trial and error so that approximately 95% of the initial perturbations of the system were accepted before the first cooling step. The factor α_τ was set so that this acceptance rate would be reduced to somewhere below 0.1% at the final cooling step. The initial value of $h_{\max}$ was set to 25 units and α_τ set so that the final value of $h_{\max}$ was around 0.5 units. Each point was then perturbed at random but subject to the constraint that it remained within the region $\mathscr{R}$ and did not approach any other point by an amount less than the specified threshold. The optimization then proceeded as described in Section 2.2.

Two basic objective functions were used in this study. The first was the Taylor series approximation to the error variance of the kriging variance in Eq. (9)—kriging at the centre of a square grid of five-unit interval from the 36 nearest neighbours. The second was the standard error of the distance parameter of the variogram a. These were based on a specified variogram.

The objective functions were obtained as follows. For a given set of locations of the sampling points, the correlation matrix **A** was computed from the specified variogram parameters. Eq. (7) was then used to calculate the information matrix of the variogram parameters, **I**, element by element. When complete **I** was inverted to find the covariance matrix of the parameters, $\mathbf{C}(\theta)$, for the given set of sample points. The elements on the principal diagonal of $\mathbf{C}(\theta)$ were extracted. The square root of each element is the standard error of the estimate of the corresponding variogram parameter. The standard error of a is one of the objective functions that was used. The correlation of the variogram parameters can similarly be extracted from $\mathbf{C}(\theta)$. The only terms of Eq. (9) unaccounted for are the partial derivatives of the kriging variance with respect to the variogram parameters. These were obtained numerically using the function DERIV in the IMSL numerical library (IMSL, 1994). This calls on a subroutine that was written to return the kriging variance for the specified variogram, and the target site and neighbours described above.

3.1.2. Simulation and sampling

The efficiency of a sample array was then tested by using it to sample from random fields. One hundred independent unconditioned realizations were simulated for a process with a specified variogram, each realization simulated at the nodes of a 51×51 unit grid. The simulations were obtained with the SASIM procedure in GSLIB (Deutsch and Journel, 1992). This uses simulated annealing to maximize the conformity of the exhaustive variogram of the simulated array to a specified variogram. The procedure SASIM was chosen in preference to

turning bands, LU decomposition or sequential Gaussian simulation. Exhaustive variograms of data produced by these latter methods fluctuate about the specified variogram and variation of the variogram parameters estimated from samples of these data will always have a substantial component which is due to this fluctuation (Webster and Oliver, 1992; Journel and Huijbregts, 1978). In this study we are interested in the sampling error. If a carefully chosen cooling schedule is used, then SASIM will simulate data with an exhaustive variogram which conforms closely to the specified variogram in each realization. Thus the variation in the estimates of variogram parameters from samples of a series of these realizations can be attributed almost entirely to sampling error, the fluctuation variance is negligible.

The simulated data are on a grid of discrete nodes. For this reason the co-ordinates generated by the SSA must be rounded to identify the simulated values to be extracted in the sample. This will result in some loss of optimality. Simulating a finer grid (i.e. with an interval set at some fraction of the distance unit) would reduce this effect. However, the chosen grid interval was a good compromise between resolution and the requirement for computing time.

3.1.3. Estimation

The procedures described so far generate 100 samples from independent realizations of a random function. The variogram was then estimated from each sample by maximum likelihood. It is known that ML estimates of the variance of a process are biased, but this can be corrected by applying the multiplier $(N)/(N-1)$, where N is the number of data in a single sample. Having generated 100 independent estimates of the variogram, one from each sample, then different variables were extracted from these—either particular variogram parameters or estimates of the kriging variance.

3.2. Specific experiments

For convenience, all the optimized sample arrays described below are summarized in Table 1.

3.2.1. Experiment 1

This objective of this experiment was to compare the performance of optimized sample arrays with systematic grids. This specifically addressed question (i) in Section 2.3.

The locations of $N = 49$ sample sites were optimized to estimate an exponential variogram with $a = 3$ units and $s = 0.9$ for a variable with a priori variance of 1.0 within a square region of side length 50 units. The objective function was the estimation variance of the kriging variance. This is array A in Table 1. This sample size is small compared to the usual recommendations for estimating the variogram. However, the focus of this study is on the variability of the estimates

Table 1
Details of sample arrays generated by spatial simulated annealing

Array	Number of sites	Objective function[b]	Specified variogram[a]: $\gamma(h) = (1.0 - s) + s\{1 - \exp(-h/a)\}$	
			a	s
A	49	evkv	3	0.9
B	49	sea	3	0.9
C	49	evkv	1	0.1
D	49	evkv	3	0.1
E	49	evkv	5	0.1
F	49	evkv	1	0.5
G	49	evkv	3	0.5
H	49	evkv	5	0.5
I	49	evkv	1	0.9
J	49	evkv	3	0.9
K	49	evkv	5	0.9
L	49	$\overline{\text{evkv}}$	1, 3, 5	0.5
M	49	$\overline{\text{evkv}}$	3	0.1, 0.5, 0.9
N	49	$\overline{\text{sea}}$	1, 3, 5	0.1
O	49	$\overline{\text{sea}}$	1, 3, 5	0.5
P	49	$\overline{\text{sea}}$	1, 3, 5	0.9
Q	80 (49 fixed)	evkv	4.8	0.6

[a]Where more than one parameter value is given then the objective function is an average computed over all variograms.

[b]evkv — estimation variance of the kriging variance (5 unit square grid). sea — standard error of the spatial parameter. An overbar indicates the average value of the estimation variance over all specified variograms.

of variogram parameters and not on estimating variograms for use. Increasing N increases the computing time needed to optimize the sample lay out since each evaluation of the objective function requires two inversions of an $N \times N$ matrix. In the context of this study where many sample arrays were generated for investigation computing time for larger sample sizes would become prohibitive.

The optimized sampling array was then used to sample each of 100 realizations of the specified random function. Each realization was also sampled with two randomly located square grids each of 49 samples, one with an interval of 2 units and the other with an interval of 6 units. The variograms were estimated by maximum likelihood from each of the three samples drawn from each of the 100 realizations and used to compute the kriging variances for different square grids.

3.2.2. Experiment 2

The objective of this experiment was to compare the performance of optimized and systematic sampling arrays where the objective function of interest in the standard error of the distance parameter, a. This experiment also addressed

question (i). The same procedure was followed as in Experiment 1, except that the objective function for optimizing the sampling array was the standard deviation of a (minimized) (array B in Table 1).

3.2.3. Experiment 3

The objective of this experiment was to compare the form of sampling arrays optimized for different variograms (addresses question (ii)), and also to compare their performance for sampling simulated fields (addressing questions (i) and (iii)).

An optimum configuration of 49 sample sites within the same square region was obtained for specified variograms with a priori variance of 1.0 and all nine combinations of the following parameter values: $a = 1$, 3, 5 grid units and $s = 0.1$, 0.5, 0.9. In all cases the objective function was the estimation variance of the kriging variance, (arrays C–K in Table 1). Comparing the resulting arrays addresses question (ii). The nine sample configurations were then used to sample 100 realizations of a random function with an exponential variogram, a priori variance of 1.0, $a = 3$ grid units and $s = 0.5$. Each realization was also sampled on transects. Seven transects each with seven observations at regularly spaced sites were randomly located along the rows or columns of the data grid. This was done four times for each realization using transects with spacing 1, 3, 5 and 7 grid units, respectively. Thirteen estimates of the variogram were therefore obtained from each realization of the random function from the nine optimized arrays and the four sets of transects. Kriging variances were computed from each variogram for specified kriging grids and compared to the values computed from the specified variogram. This allows the efficiency of optimized and systematic sampling to be compared (question (i)). It also allows us to investigate the efficiency of optimized arrays for sampling realizations of random functions other than the specified one (question (iii)).

Two sample configurations were then generated by optimizing 'hybrid' objective functions. In the first case (array L) the parameter s was fixed at 0.5, but the objective function (error variance of the kriging variance) was computed as the average estimation variance of the kriging variance for variograms with $a = 1$, 3, and 5 units. In the second case (array M) a was fixed at 3 units and the average objective function computed for $s = 0.1$, 0.5 and 0.9. These two sample arrays were then used to sample the same simulated data described in the previous paragraph and the kriging variances were computed from the estimated variograms. This addresses question (iv), whether uncertainty about the underlying parameters that we want to estimate can be dealt with by using these average objective functions.

3.2.4. Experiment 4

Experiment 3 suggested that the efficiency of an optimized sample array is particularly sensitive to the specified distance parameter a. One practical

approach to sampling would be to sample to estimate this parameter provisionally, then to optimize the remaining sampling points on the basis of this initial estimate. In this experiment two sampling strategies for estimating the distance parameter were compared.

The first sampling strategy was to use optimized arrays. Three optimized arrays of 49 sites (arrays N, O and P) were generated with hybrid objective functions, with $s = 0.1$, 0.5 and 0.9, respectively, and the objective function, the average error variance of the distance parameter when this is 1, 3 and 5 units. These arrays were applied to realizations of two random functions, both with a priori variance 1.0 and $s = 0.5$ and with $a = 3$ and 5 units, respectively.

The second strategy used systematic sampling. The simulated data were also sampled with four sets of random transects (seven transects of seven sites) with spacing 1, 3, 5 and 7 units, respectively (as in Experiment 3).

3.2.5. Experiment 5

This experiment compared simple systematic sampling with phased sampling (systematic sampling in the first phase, then a second phase where additional sites are sampled at optimized locations). This addressed question (v).

The simple systematic sample consisted of 10 transects, each of eight points at intervals of 3 grid units, laid down at random. This was applied to 100 realizations of a random function with $a = 5$ units and $s = 0.5$. Kriging variances were computed from each estimated set of variogram parameters.

The phased sampling started with one randomly located set of seven transects each with seven points at intervals of 3 grid units. This had been used in Experiment 4 to sample realizations of a random function with $a = 5$ units and $s = 0.5$. A further 31 points were added to one such set of transects by a modified version of the SSA algorithm in which the 49 points on the transects were not disturbed but the additional 31 points were placed so as to optimize an objective function (estimation variance of the kriging variance) derived from the covariance matrix for the variogram parameters computed from the correlation matrix for all 80 points. The specified variogram was the exponential with the average values of a and s, which had been obtained with the initial (transect) sampling design in Experiment 4. The resulting sample array (array Q) was applied to each of 100 realizations of a random function with $a = 5$ units and $s = 0.5$. Kriging variances were computed from each estimated set of variogram parameters.

4. Results

4.1. Experiment 1

Fig. 3 shows the distribution of sample points in optimized array A with the histogram of lag distances between these points inset. The progress of the

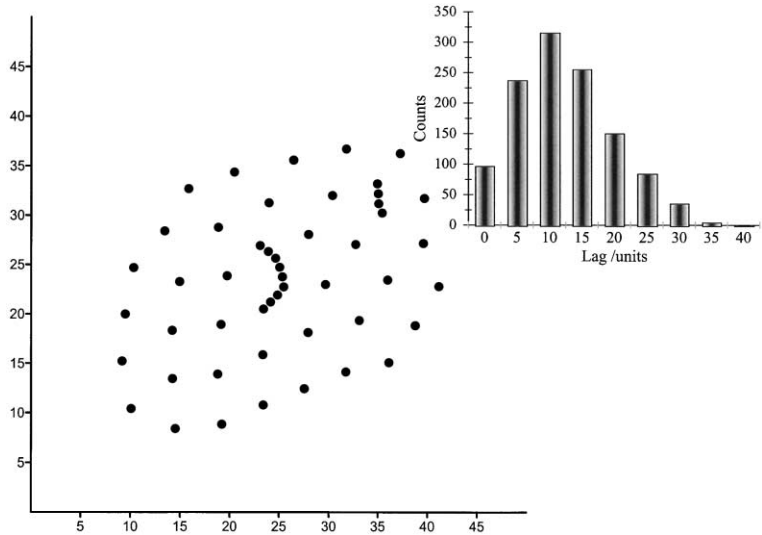

Fig. 3. Optimized sample array A with (inset) histogram of lag distances in array A.

optimization is shown in Fig. 4, a graph of the change in the objective function with successive perturbations of the sample configuration. The final array is more or less regular with each point about 5 units from its nearest neighbour, with the exception of two short chains of points that generate nearly 100 comparisons over lags less than 5 units. Contrary to the conjecture of Bresler and Green (1982), the histogram of lag distances does not resemble a uniform frequency distribution.

Fig. 5(a–c) shows the average kriging variance for different kriging grid intervals computed from variograms estimated with sample array A or from the two regular sample grids. The 5 and 95 percentiles of these estimates are shown, as well as the kriging variances computed for the specified variogram.

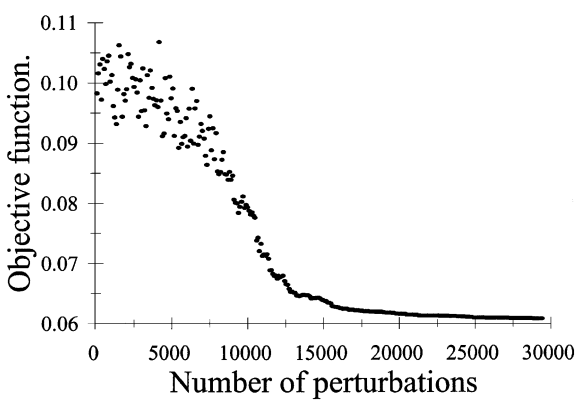

Fig. 4. Progress of the optimization of array A.

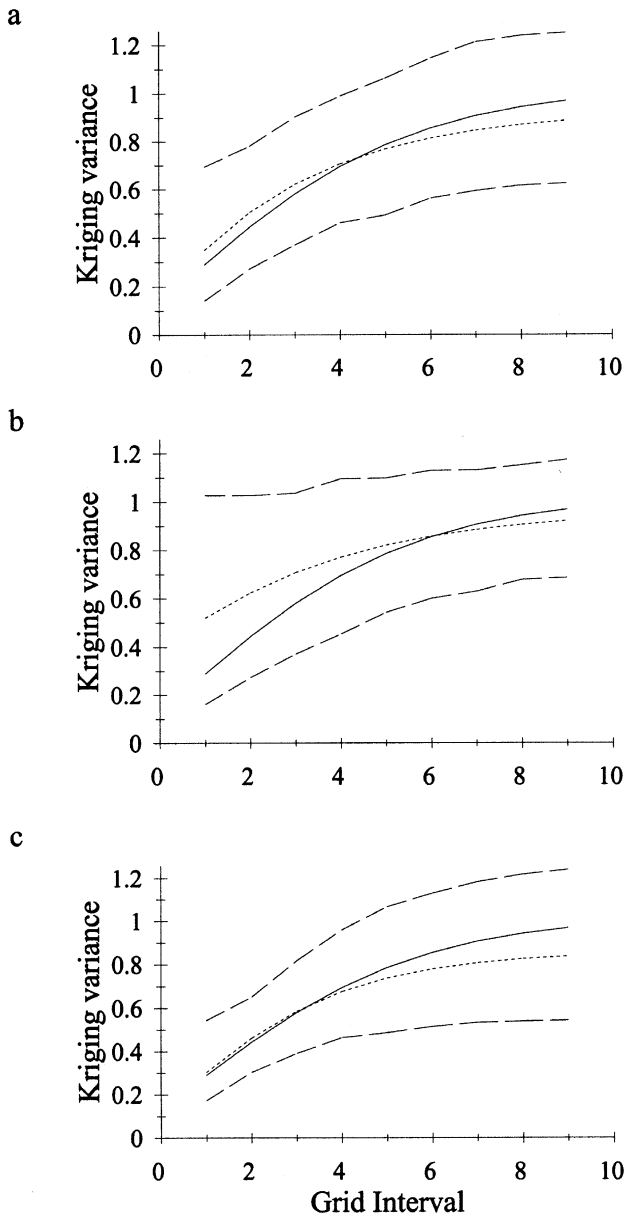

Fig. 5. Average kriging variance for different kriging grid intervals (dotted line) computed from variograms estimated with sample array A (a) or regular sample grids interval 6 units (b) or 2 units (c). Values for the specified variogram (solid line) and 0.05 and 0.95 quantiles of the estimates (broken line) are shown.

The variance of the estimated kriging variances from each set of variogram estimates is shown in Fig. 6(a) and the root mean-squared error (RMSE) from the values computed from the specified variogram are shown in Fig. 6(b). While

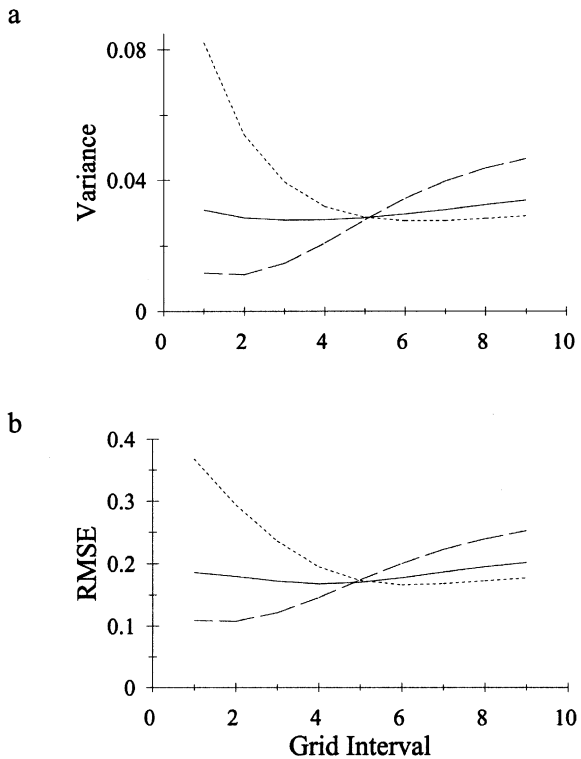

Fig. 6. Variance (a) and RMSE (b) of the estimated kriging variance from variograms estimated from sample array A (solid line), 6-unit grid (broken line) and 2-unit grid (dotted line).

the variance measures the precision of the different sets of estimated kriging variances, the RMSE also measures the bias.

It is interesting that the variograms estimates from the coarser sampling grid seem to overestimate the kriging variance of the finer kriging grids, while the variograms estimated from the finer sampling grids overestimate the kriging variance for the coarser kriging grids. The variance and RMSE of the estimated kriging variances are similarly sensitive to the kriging grid interval. The kriging variance for all kriging grids calculated from the variograms estimated from sample array A were close to the true values. The variance and RMSE of the kriging variances from the variograms obtained with sample array A were not sensitive to the size of the kriging grid.

4.2. Experiment 2

Fig. 7 shows the distribution of points in sample array B with the histogram of lag distances inset. Again the array is predominantly regular and with fewer

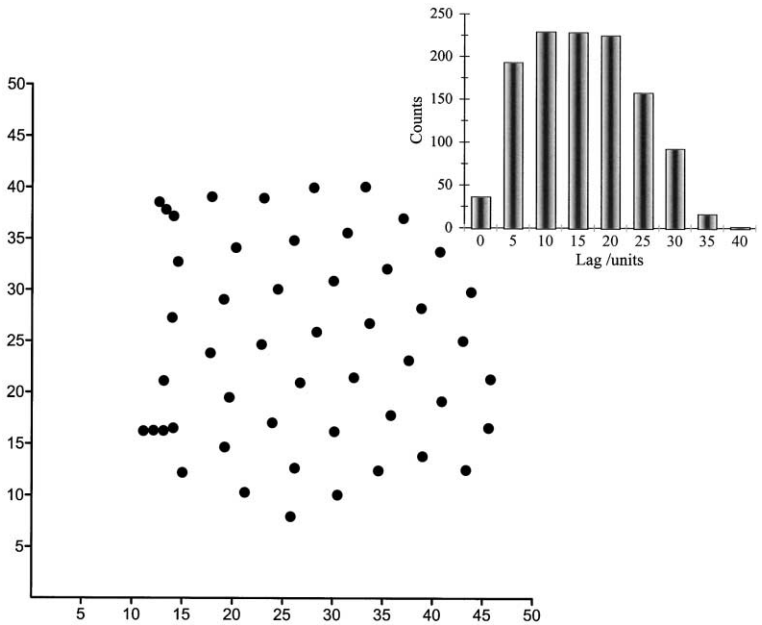

Fig. 7. Optimized sample array B with (inset) histogram of lag distances in array B.

points in short chains than in array A, so there are fewer short lag distances. The distribution of lag distances is closer to a uniform distribution than that for array A.

Table 2 shows the mean, variance and RMSE of the estimates of the distance parameter *a* obtained from sampling the simulated fields with array B. The corresponding results for array A and the two square sampling grids are also given. The variance and RMSE is smallest for the 2 unit grid, but the bias is

Table 2
Mean, variance and RMSE of estimates of the spatial parameter *a*
The true value of *a* is 3.0.

Sampling scheme	Mean estimate	Variance	RMSE
Array B	2.7	2.3	1.5
Grid, 2 unit interval	2.3	0.7	1.1
Grid, 6 unit interval	3.6	17.8	4.3
Array A	2.8	2.7	1.6

considerable. The variance and RMSE are smaller for sample array B than for sample array A or the 6 unit grid.

4.3. Experiment 3

The distribution of sample sites and the histograms of lag distances for arrays C – K are shown in Figs. 8 and 9. None of the histograms tend to a uniform frequency distribution. There are three broad types of sample array.

(i) Clusters of points, usually with a bimodal distribution of lags. These occur when the distance parameter and/or the spatial dependence is small.
(ii) A more or less regular distribution of points with some of the points supplemented by an additional observation at a short distance. This pattern arises with a larger distance parameter and intermediate spatial dependence.

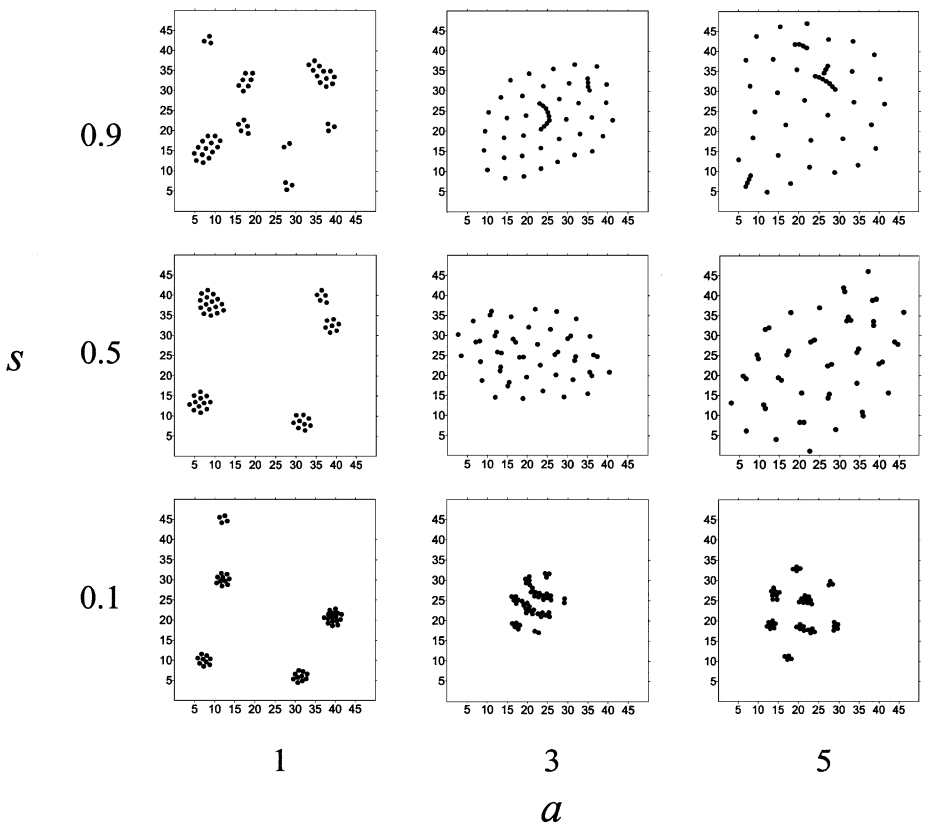

Fig. 8. Optimized sample arrays C–K. These are arranged according to the specified spatial parameter *a* (columns) and spatial dependence *s* (rows).

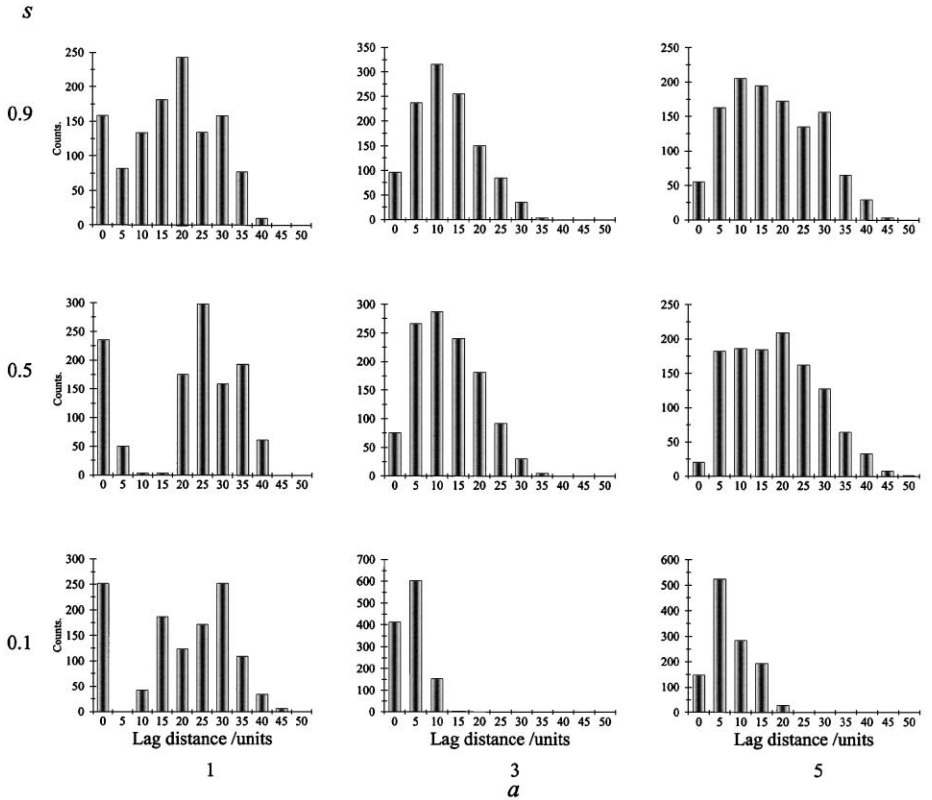

Fig. 9. Histograms of lag distances for arrays C–K, arranged as in Fig. 8.

(iii) A more or less regular array of points with some short ‘chains’ giving rise to comparisons over short lags. This arises from the largest distance parameter and large spatial dependence.

For reasons of space, the arrays obtained by optimizing ‘average’ objective functions are not shown, but array L followed pattern (ii) and array M pattern (iii).

Figs. 10–12 show, respectively, the average kriging variance (for a kriging grid of interval 5 units), the variance of this quantity and the RMSE from the 100 variograms obtained by sampling with arrays C–M and random transects of different spacing.

There is a tendency to underestimate the variance, but this bias is smallest for array G (which is optimized for the variogram specified for simulating the data). The variance of the kriging variance and the RMSE are both smallest for variograms obtained with array G.

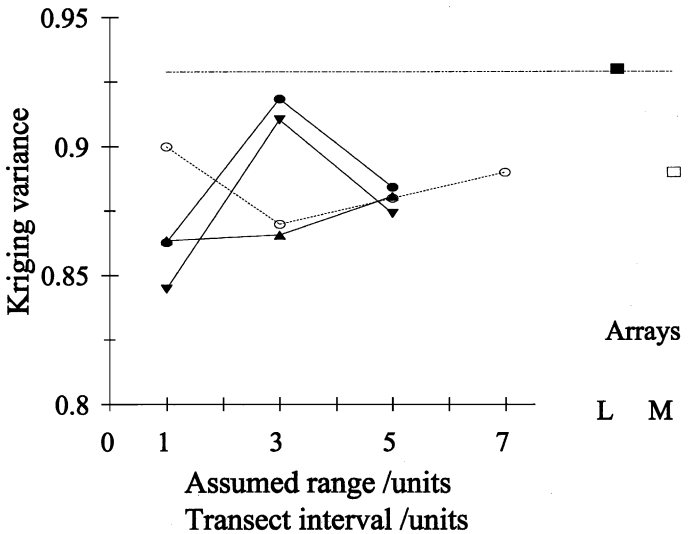

Fig. 10. Average kriging variance (square grid of interval 5 units) computed from variograms obtained by sampling with arrays C–M and randomly located transects of different spacing. Specified variogram had a spatial parameter 3 and spatial dependence 0.5. Note that the scale on the abscissa serves two purposes showing the spatial parameter assumed for optimizing arrays C–K, and the interval on the random transects. The horizontal broken line shows the actual kriging variance for the variogram used to simulate the data. Symbols are as follows: ○ Randomly located transects, ▲ Spatial dependence assumed 0.1, ● Spatial dependence assumed 0.5, ▼ Spatial dependence assumed 0.9.

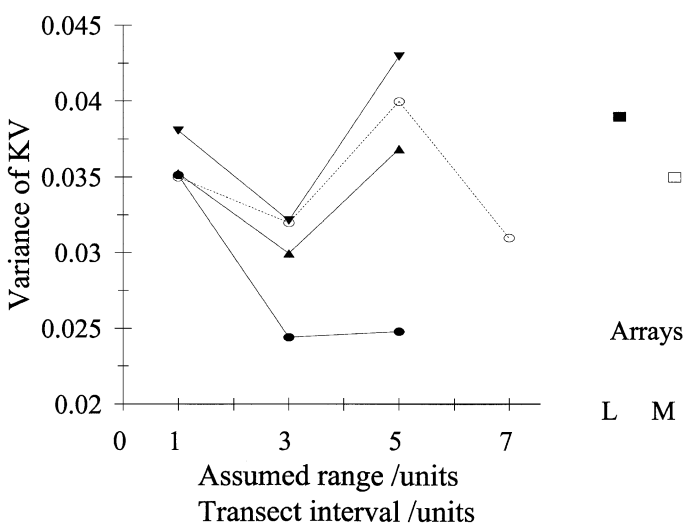

Fig. 11. Variance of the kriging variances computed from variograms obtained from different sampling schemes. The format and symbols are as in Fig. 10.

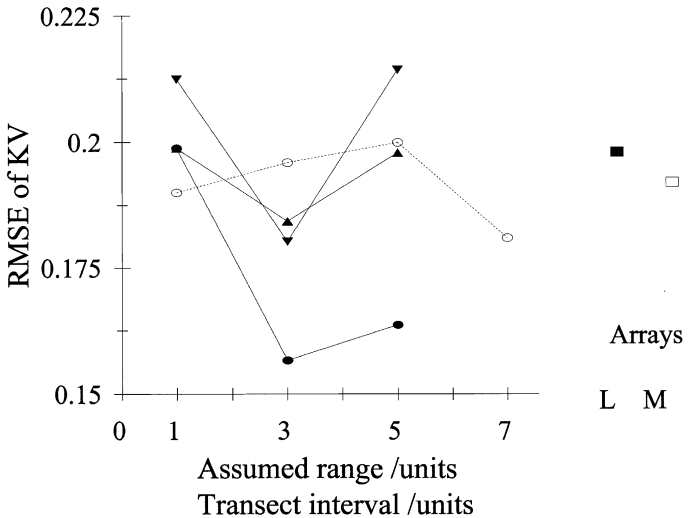

Fig. 12. RMSE of the kriging variances computed from variograms obtained from different sampling schemes. The format and symbols are as in Fig. 10.

In general, the bias and RMSE are of similar order for the transects and the arrays optimized for variograms other than the one from which the data were simulated. Nonetheless, among these arrays the RMSE is generally smaller for those optimized for a variogram with the correct distance parameter. This suggests that an initial sampling array which provides a reasonable initial

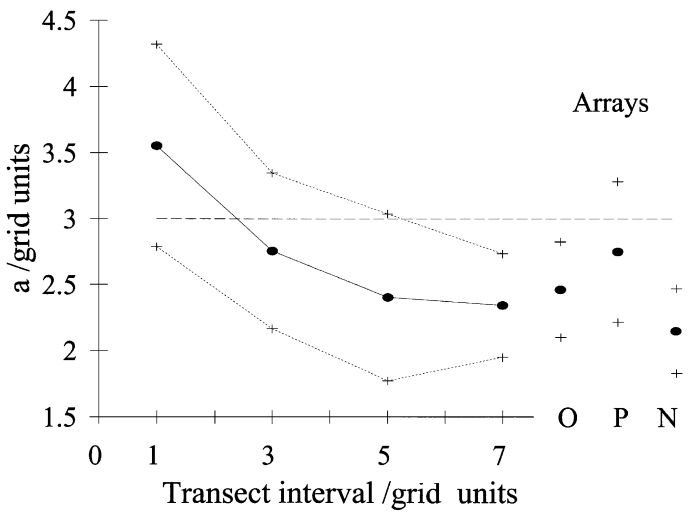

Fig. 13. Mean estimates and their 95% confidence limits for the spatial parameter for realizations of a random function with an exponential variogram ($a = 3$ units, $s = 0.5$). Estimated from randomly located transects or optimized arrays N–P.

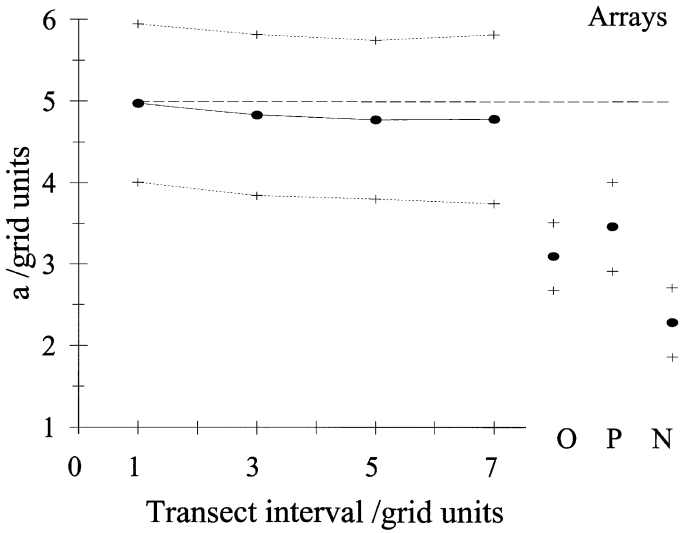

Fig. 14. Mean estimates and their 95% confidence limits for the spatial parameter for realizations of a random function with an exponential variogram ($a = 5$ units, $s = 0.5$). Estimated from randomly located transects or optimized arrays N–P.

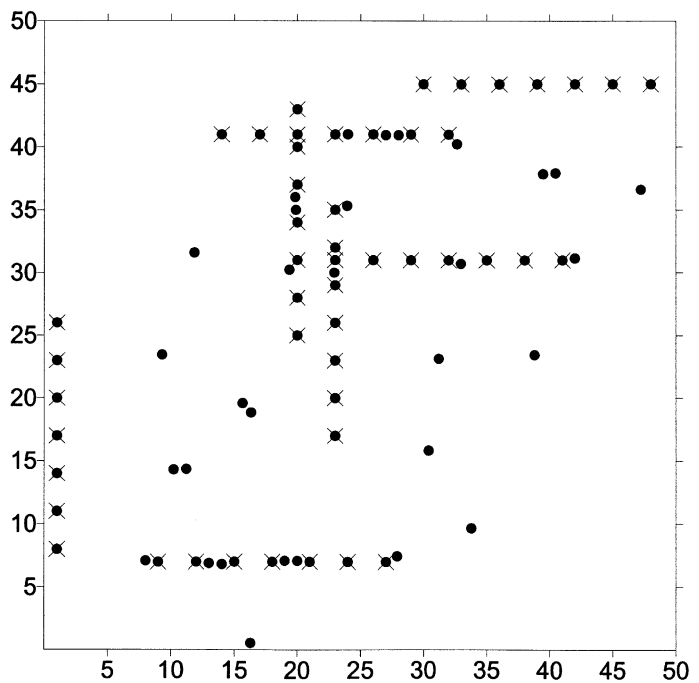

Fig. 15. Sample array Q. The crosses indicate points in the original transects.

estimate of the distance parameter could then be supplemented to provide an optimum sampling array for estimating the variogram.

4.4. Experiment 4

Fig. 13 shows the mean estimates of the distance parameter for realizations of a random function with an exponential variogram ($a = 3$ units, $s = 0.5$). The 95% confidence limits for these estimates are also shown. The estimates were obtained from transects of different spacing or from three arrays designed to minimize the average standard deviation of the parameter for variograms with $a = 1$, 3 and 5 and with spatial dependence 0.1 (array N), 0.5 (array O) or 0.9 (array P). Fig. 14 shows the same results when the sample configurations were used to sample realizations of another random function with an exponential variogram ($a = 5$ units, $s = 0.5$).

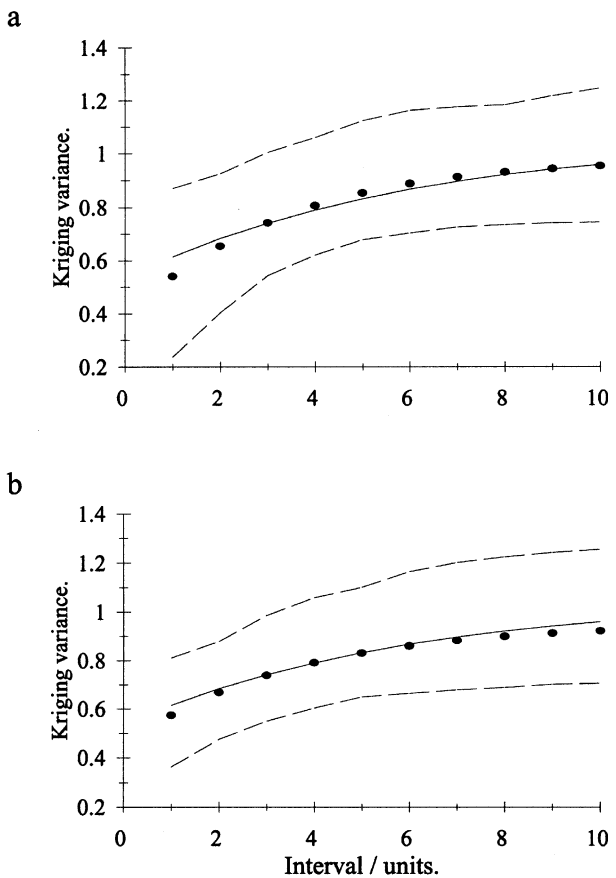

Fig. 16. The mean kriging variance with the 0.05 and 0.95 percentiles (broken line) for square kriging grids of different interval computed from variograms estimated either from transects (a) or array Q. The solid line shows the true values.

It is notable that the optimized arrays gave biased estimates of the distance parameter in most cases, tending to underestimate it. The smallest bias (not statistically significant for estimates of the process where $a = 3$) was obtained with array P, which has the largest spatial dependence (0.9). This suggests that

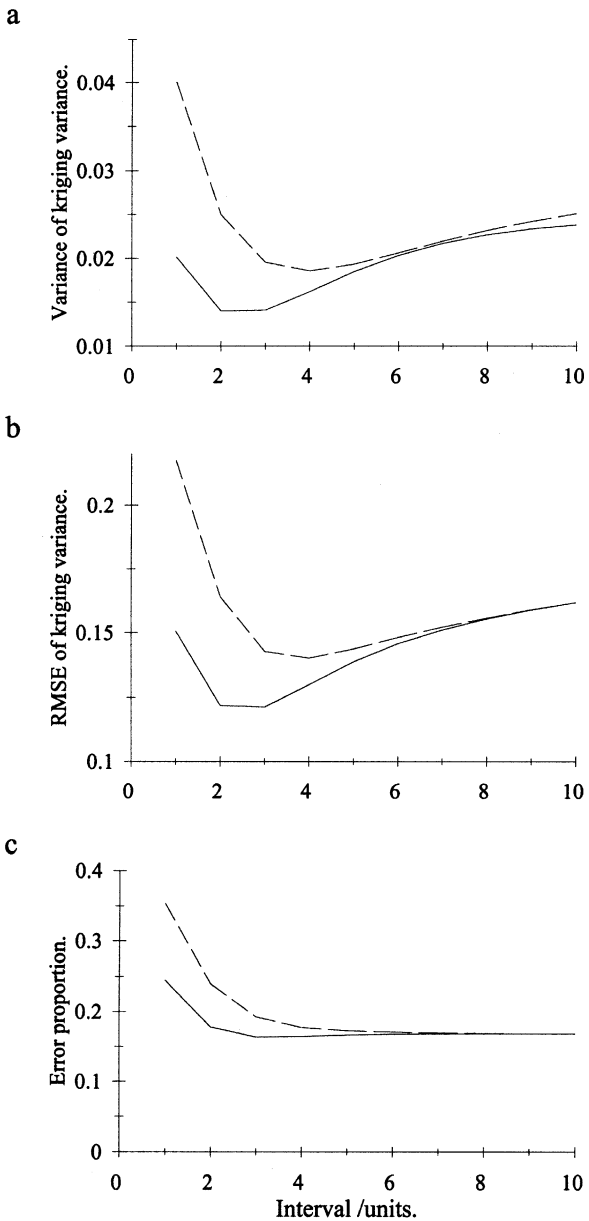

Fig. 17. Variance (a), RMSE (b) and RMSE as a proportion of the true kriging variance (c) for the estimates presented in Fig. 16. Solid line: array Q, broken line; transects.

the bias may be due to undue influence on the objective function of variograms with weak spatial structure. The mean estimates of the distance parameter obtained from the transects show more variation than the estimates from the optimized arrays, but in most instances show no bias.

4.5. Experiment 5

Fig. 15 shows the sample array generated by supplementing a set of transects optimally (array Q). It is interesting that about half the points added to the original transects fill in spaces between the transects while the remainder are inserted into spaces in the transects. This suggests that the original transects were too coarsely spaced.

The mean kriging variance for square kriging grids of different interval computed from variograms estimated either from array Q or from the same number of points on transects with the original spacing are shown in Fig. 16(a,b) with the 0.05 and 0.95 percentiles. Note the greater variability of the estimates based on the transect sampling for kriging grids with smaller spacings. Fig. 17 emphasises this showing the variance (a), RMSE (b) and RMSE as a proportion of the true kriging variance (c). There are clear advantages in using the sampling scheme obtained by optimal supplementation of transects. This is probably because the basic transect spacing is too coarse to characterize spatial dependence at short lags adequately.

5. Discussion and conclusions

I return to the questions posed in Section 2.3. The experiments addressed to these questions support the following conclusions.

(i) These simulations show that optimizing the configuration of sampling points with respect to objective functions derived from the maximum likelihood estimator of the parameters of the variogram can improve the efficiency of estimation. This is seen particularly in the stability of the error of estimated kriging variance (Fig. 5) and in the error of estimates from optimized arrays presented in Fig. 12. These effects are not large, though, and may not justify the extra time and effort needed by comparison to sampling on transects.

(ii) This study shows in a rigorous way how the optimum sampling strategy changes with the spatial dependence of the underlying variable. A random function with a small distance parameter and/or a small spatial dependence is best sampled by scattered clusters of points which generate many comparisons over short lags. By contrast a long-range process with strong spatial dependence is best sampled with a more or less regular array, usually supplemented with a few chains of points to generate comparisons over short lags. In intermediate

situations short-range comparisons are generated by distributing additional points more evenly over the array, each one associated with a point in the basic grid.

(iii) When we are entirely ignorant about the underlying variogram of a property, then the most robust approach is to sample on transects. When we have a prior idea of the size of the distance parameter of the variogram, then optimizing the sampling array based on this value may be advantageous even if we have to guess the ratio of spatial dependence.

(iv) None of the hybrid objective functions generated sampling arrays with useful properties, and in some cases estimates based on the resulting samples were biased. This may be because the objective functions were based on simple averages for strongly contrasting variograms. Weighted averaging might perform better as a way of incorporating prior knowledge about spatial variability into the sampling design.

(v) There are benefits in the staged sampling approach used to generate array Q by comparison to investing the same overall sampling effort entirely on transects. This may be the most appropriate way to use the optimization method proposed in this paper.

Given these answers to the original questions, what general conclusions may be drawn? This study has shown how spatial variability of a process controls the optimum sampling scheme. The transition from more or less regularly spaced sample sites, to regularly spaced pairs of sample sites to clusters of sample sites as the spatial continuity of the variable becomes weaker is consistent with the expectations of geostatisticians, but the form of the sample arrays and the precise way in which they change with variogram parameters has not previously been demonstrated. Even when a formal optimization of the sampling design is not attempted, planning an exploratory sampling exercise with one of these patterns in mind, depending in part on prior knowledge about the likely form of the variability of interest might be useful.

Second, it has been shown that there is scope to improve the efficiency of a sample by adjusting the distribution of sample points on the basis of some prior knowledge about the spatial variability of the target variable. Results suggest that information about the distance parameter may be most useful. For practical purposes a phased sampling scheme is suggested. The optimization method is used to identify sample sites for a second phase of sampling, supplementing an initial phase of sampling on a robust design, such as randomly located transects, which provides initial estimates of the variogram parameters. It might also be useful for supplementing a basic space-filling design such as a square or triangular grid, which provides the basis for the final kriging grid.

Third, it has been shown that minimizing a ‘hybrid’ objective function representing a range of variograms is to be treated with caution. Further research might be worthwhile here to define objective functions that can incorporate prior information about spatial variability in a more robust way than the simple

averaging used here. Lark (2000d) has shown that prior information about spatial variability might be expressed through the formalism of the fuzzy variogram. Using fuzzy memberships as weights might improve the performance of the objective functions explored in this study, which are averages of the results for different variograms. Any such development will need quicker computation of the objective function, perhaps using the approximate maximum likelihood proposed by Pardo-Iguzquiza and Dowd (1997), which can be calculated with less computational effort than the conventional maximum likelihood.

The principle of using an optimized phased sampling procedure may be applied practically in soil survey. The first step of this procedure would be a reconnaissance survey using either randomly located transects with regularly spaced sample sites—when the study area is not very large in relation to the anticipated spatial scales of interest, (e.g. in precision farming)—or using randomly distributed transects with exponential spacings (Pettitt and McBratney, 1993) or nested sampling (Webster and Oliver, 2000)—in large areas where the scales of importance may differ by orders of magnitude (e.g. in regional surveys).

The second phase of sampling would then be designed using SSA to supplement the existing sample sites optimally for estimating variograms from the data. Other considerations might also be incorporated into the SSA design where appropriate. For example, in a large regional survey limits could be set on the total distance that has to be travelled in order to visit all the new sample sites.

After phase 2, SSA as described by van Groenigen (1999) would then be used to design a third sampling phase from which a kriged map of the variable is produced, ensuring that the kriging variance at any location is kept below a threshold value.

Acknowledgements

This work was funded by the Biotechnology and Biological Sciences Research Council, UK through its Competitive Strategic Grant to Silsoe Research Institute. The author would like to thank the two referees who read the first version of this paper for comments that have improved its presentation.

References

Aarts, E., Korst, J., 1989. Simulated Annealing and Boltzmann Machines—A Stochastic Approach to Combinatorial Optimization and Neural Computing. Wiley, New York.

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

Bresler, E., Green, R.E., 1982. Soil parameters and sampling scheme for characterizing soil

hydraulic properties of a watershed. Technical 148. Water Resources Research Center, University of Hawaii, Honolulu (not available but cited by Warrick and Myers, 1987).

Brus, D.J., de Gruijter, J.J., 1994. Estimation of nonergodic variograms and their sampling variance by design-based sampling strategies. Mathematical Geology 26, 437–454.

Burgess, T.M., Webster, R., 1980. Optimal interpolation and isarithmic mapping of soil properties: I. The semivariogram and punctual kriging. Journal of Soil Science 31, 315–331.

Deutsch, C.V., Journel, A.G., 1992. GSLIB Geostatistical Software and User's Guide. Oxford University Press, New York.

Diggle, P.J., Ribeiro, P.J., 1999. Bayesian inference in Gaussian model based geostatistics. In: Atkinson, P.M., Riding, A.E., Tate, N.J. (Eds.), Proceedings of the Geostats-UK '99 Conference. University of Southampton, Southampton, pp. 29–43.

Geiler, H., Aschenbrenner, F., Dengel, H.S., Donsbach, A., Ostermann, R., Knoblich, K., Maurer, W., 1997. Räumliche variabilität und verteilung von schwermetallen in den böden einer 1 km^2 großen Modellfläche im Siegerland. Zeitschrift für Pflanzenernährung und Bodenkunde 160, 603–612.

IMSL, 1994. IMSL MATH/LIBRARY User's Manual, Version 3.0. Visual Numerics, Houston, TX.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, New York.

Kitanidis, P.K., 1987. Parametric estimation of covariances of regionalised variables. Water Resources Bulletin 23, 557–567.

Lark, R.M., 2000a. A geostatistical extension of the sectioning procedure for handling differences of scale between functional models of soil processes and information on one or more input variables. Geoderma 95, 89–112.

Lark, R.M., 2000b. A comparison of some robust estimators of the variogram for use in soil survey. European Journal of Soil Science 51, 137–157.

Lark, R.M., 2000c. Estimating variograms of soil properties by method-of-moments and maximum likelihood. European Journal of Soil Science 51, 717–728.

Lark, R.M., 2000d. Designing sampling grids from imprecise information on soil variability: an approach based on the fuzzy kriging variance. Geoderma 98, 35–59.

Matheron, G., 1962. Traité de Géostatistique Appliqué, Tome 1. Memoires du Bureau de Recherches Géologiques et Minières, Paris.

McBratney, A.B., Webster, R., Burgess, T.M., 1981. The design of optimal sampling schemes for local estimation and mapping of regionalised variables. Computers and Geosciences 7, 331–334.

Müller, W.G., Zimmerman, D.L., 1999. Optimal designs for variogram estimation. Environmetrics 10, 23–37.

Oliver, M.A., 1999. Exploring soil spatial variation geostatistically. In: Stafford, J.V. (Ed.), Precision Agriculture '99, Part 1. Sheffield Academic Press, Sheffield, pp. 3–17.

Oliver, M.A., Webster, R., 1987. The elucidation of soil pattern in the Wyre forest of the West-Midlands, England: 2. Spatial-distribution. Journal of Soil Science 38, 293–307.

Papritz, A., Webster, R., 1995. Estimating temporal change in soil monitoring: II. Sampling from simulated fields. European Journal of Soil Science 46, 13–27.

Pardo-Iguzquiza, E., 1997. MLREML: A computer program for the inference of spatial covariance parameters by maximum likelihood and restricted maximum likelihood. Computers and Geosciences 23, 153–162.

Pardo-Iguzquiza, E., Dowd, P.A., 1997. AMLE3D: a computer program for the inference of spatial covariance parameters by approximate maximum likelihood estimation. Computers and Geosciences 23, 793–805.

Pardo-Iguzquiza, E., Dowd, P.A., 1998. Maximum likelihood inference of spatial covariance parameters of soil properties. Soil Science 163, 212–219.

Pettitt, A.N., McBratney, A.B., 1993. Sampling designs for estimating spatial variance-components. Applied Statistics 42, 185–209.

Russo, D., 1984. Design of an optimal sampling network for estimating the variogram. Soil Science Society of America Journal 48, 708–716.

Stein, M.L., 1988. Asymptotically efficient prediction of a random field with a misspecified covariance function. Annals of Statistics 16, 55–63.

van Groenigen, J.W., 1999. Constrained optimisation of spatial sampling. ITC Publication Series, vol. 65, Enschede.

Viloria, J., 1989. The inter-relation between soil survey and soil information systems: optimization of the data gathering process. D.Phil. Thesis. University of Oxford.

Voltz, M., Webster, R., 1990. A comparison of kriging, cubic splines and classification for predicting soil properties from sample information. Journal of Soil Science 41, 473–490.

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

Webster, R., Oliver, M.A., 1992. Sample adequately to estimate variograms of soil properties. Journal of Soil Science 43, 177–192.

Webster, R., Oliver, M.A., 2000. Geostatistics for Environmental Scientists. Wiley, Chichester.