Mathematical Geology, Vol. 39, No. 1, January 2007 (© 2007)
DOI: 10.1007/s11004-006-9069-1

Optimized Sample Schemes for Geostatistical Surveys[1]

B. P. Marchant[2] and R. M. Lark[2]

Sample schemes used in geostatistical surveys must be suitable for both variogram estimation and kriging. Previously schemes have been optimized for one of these steps in isolation. Ordinary kriging generally requires the sampling locations to be evenly dispersed over the region. Variogram estimation requires a more irregular pattern of sampling locations since comparisons must be made between measurements separated by all lags up to and beyond the range of spatial correlation. Previous studies have not considered how to combine these optimized schemes into a single survey and how to decide what proportion of sampling effort should be devoted to variogram estimation and what proportion devoted to kriging.

An expression for the total error in a geostatistical survey accounting for uncertainty due to both ordinary kriging and variogram uncertainty is derived. In the same manner as the kriging variance, this expression is a function of the variogram but not of the sampled response data. If a particular variogram is assumed the total error in a geostatistical survey may be estimated prior to sampling. We can therefore design an optimal sample scheme for the combined processes of variogram estimation and ordinary kriging by minimizing this expression. The minimization is achieved by spatial simulated annealing. The resulting sample schemes ensure that the region is fairly evenly covered but include some close pairs to analyse the spatial correlation over short distances. The form of these optimal sample schemes is sensitive to the assumed variogram. Therefore a Bayesian approach is adopted where, rather than assuming a single variogram, we minimize the expected total error over a distribution of plausible variograms. This is computationally expensive so a strategy is suggested to reduce the number of computations required.

KEY WORDS: variogram, ordinary kriging, maximum likelihood.

INTRODUCTION

The collection and analysis of samples for geostatistical surveys may be expensive and time-consuming. Significant savings may be made if sample schemes are optimized to maximize the information attained from a relatively small sample size. Several previous studies have considered how the location of sample locations may be optimized for either variogram estimation (Bogaert and Russo, 1999;

[1]Received 8 March 2005; accepted 10 April 2006; Published online: 28 February 2007.
[2]Bioinformatics and Biomathematics Division, Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, UK; e-mail: ben.marchant@bbsrc.ac.uk.

Müller and Zimmerman, 1999; Lark, 2002) or ordinary kriging (van Groenigen, Siderius, and Stein, 1999).

Often these two steps require quite different sample schemes. Ordinary kriging requires that the sampling locations are dispersed evenly over the region whereas variogram estimation requires comparisons between measured values separated by small, intermediate and large distances relative to the range of spatial correlation. Little consideration has been given to how much sampling effort should be devoted to each step or how to design a sample scheme which is suitable for both.

The optimal configuration of sampling locations depends upon the variogram and in particular the range of spatial correlation. Abt, Welch, and Sacks (1999) explored whether the addition of close pairs of observations in an otherwise regular survey of dioxin contamination would aid variogram estimation. The additional observations made little improvement to empirical prediction accuracy. In this case the regular portion of the survey, where locations were closely spaced relative to the range of spatial correlation, was sufficient for variogram estimation. This may not be the case for less intensive regular surveys.

Furthermore, van Groenigen (2000) and Lark (2002) found that the optimal sample schemes for both variogram estimation and kriging depend on the variogram of the variable being measured. The variogram is unknown prior to sampling and therefore the variogram must be selected arbitrarily if the sample scheme is to be optimized in this manner. An inaccurate choice of variogram will lead to an unsuitable sample scheme. In the case of variogram estimation, Bogaert and Russo (1999) rectified this problem using a Bayesian framework within which they integrated their objective function over all plausible variogram functions.

In this paper we discuss methods to design sample schemes which are optimal for an entire ordinary kriging survey including variogram estimation. A single expression for the total error in such surveys is derived and a spatial simulated annealing (SSA) algorithm is used to design optimal sample schemes which minimize this expression. The pattern of locations within these schemes is seen to vary according to the assumed variogram parameters and therefore a Bayesian framework similar to that used by Bogaert and Russo (1999) is adopted.

THEORY

Estimation of the Variogram

In geostatistics the value of a spatially correlated variable at a location $\mathbf{x}$ is regarded as a realization of a random function $Z(\mathbf{x})$. If the random function is intrinsically stationary then $\mathrm{E}[Z(\mathbf{x})]$ is constant for all $\mathbf{x}$ and the spatial correlation may be described by the variogram

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{E}[(Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h}))^2]. \tag{1}$$

Here $\mathbf{h}$ is the lag vector. In general, the variogram may be a function of both the length and direction of the lag vector. In this exploratory study however, only isotropic variograms are considered which are purely a function of the length of the vector which we denote by the scalar h. Furthermore we concentrate upon the spherical variogram model

$$\begin{aligned}
&\gamma(h) = c_0 + c_1\left(\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right) \quad \text{for } 0 < h \leq a,\\
&\gamma(h) = c_0 + c_1 \text{ for } h > a,\\
&\gamma(0) = 0.
\end{aligned} \tag{2}$$

The spherical model has three parameters, these are the nugget variance c_0, the sill variance c_1 and the range of spatial correlation a. We denote the vector of q variogram parameters by $\boldsymbol{\theta}$ so in the case of the spherical model $\boldsymbol{\theta} = (c_0, c_1, a)$.

Conventionally, the variogram is estimated from n sampled values of $Z(\mathbf{x})$, denoted $\mathbf{z}$, by the method of moments (Journel and Huijbregts, 1978). Pardo-Igúzquiza (1998) suggested that variograms are fitted by maximum likelihood rather than the method of moments. This method fits a variogram model such as Eq. (2) directly to the sampled response data under the assumption that it follows a multivariate Gaussian distribution. The maximum likelihood estimates of the variogram parameters, $\hat{\boldsymbol{\theta}}$, are the values of $\boldsymbol{\theta}$ which minimize the negative log-likelihood function

$$L(\boldsymbol{\theta};\mathbf{z}) = \frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln(\det[\mathbf{V}]) + \frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{z}-\boldsymbol{\mu}), \tag{3}$$

where $\mathbf{V}$ is the $n \times n$ covariance matrix of the sampled values and $\boldsymbol{\mu}$ is the length n vector with each entry equal to μ, the mean value of $Z(\mathbf{x})$. Eq. (3) is minimized numerically. In this paper we use simulated annealing (Pardo-Igúzquiza, 1997). For each $\boldsymbol{\theta}$, μ, is replaced by an estimate $\hat{\mu}$ by differentiating Eq. (3) with respect to μ and equating the resulting function to zero. This yields

$$\hat{\mu} = \left(\mathbf{1}_n^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{1}_n\right)^{-1}\mathbf{1}_n^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{z}, \tag{4}$$

where $\mathbf{1}_n$ is a $n \times 1$ vector with each element equal to 1 and $\mathbf{Q}$ is the correlation matrix (i.e. $\mathbf{V} = \sigma^2\mathbf{Q}$ for a process with total variance $\sigma^2 \equiv c_0 + c_1$).

The relative merits of the method of moments and maximum likelihood for estimating variograms are discussed by Lark (2000). One advantage of maximum likelihood is there is no necessity to bin the pair comparisons into lag classes and therefore no risk that spatial information may be lost by excessive smoothing of the experimental variogram. Variograms fitted by maximum likelihood are consistent with a Bayesian framework to account for uncertainty of geostatistical surveys. For both methods the uncertainty of variogram parameter estimates may be assessed via the inverse Fisher information matrix which we discuss below. However the calculation of the Fisher information matrix for the method of moments requires many more computations. Therefore in this paper we favour the maximum likelihood approach.

Variogram Uncertainty

The inverse of the Fisher information matrix is commonly used as an approximation to the variance-covariance matrix of fitted variogram parameters (e.g. Pardo-Igúquiza and Dowd, 2001b). The Fisher information matrix $\mathbf{F}(\boldsymbol{\theta})$ contains the expected values of the negative second order partial derivatives of the log-likelihood function evaluated at the maximum likelihood estimate $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, i.e.,

$$[F(\hat{\boldsymbol{\theta}})]_{ij} = -\mathrm{E}\left[\frac{\partial^2 L(\hat{\boldsymbol{\theta}}; \mathbf{z})}{\partial\theta_i \partial\theta_j^{\mathrm{T}}}\right]. \tag{5}$$

It is shown in standard texts (e.g. Dobson, 1990) that if the variogram parameter estimators are normally distributed then their covariance matrix may be approximated by $\mathbf{F}^{-1}(\boldsymbol{\theta})$.

Kitanidis (1987) showed that the Fisher information matrix can be expressed as

$$[\mathbf{F}(\boldsymbol{\theta})]_{ij} = \frac{1}{2}\mathrm{Tr}[\mathbf{Q}^{-1}\mathbf{Q}_i\mathbf{Q}^{-1}\mathbf{Q}_j], \tag{6}$$

where $\mathbf{Q}_i = \partial\mathbf{Q}/\partial\theta_i$ and Tr[ ] denotes the trace of a matrix. Note that the Fisher Information matrix is a function of the sampling configuration and variogram but not directly of the sampled response data. Therefore if the correlation structure is known, it is possible to calculate the Fisher Information matrix prior to sampling. Pardo-Igúquiza and Dowd (2001a) suggested an analogous method to approximate the variance-covariance matrix of variogram parameters fitted by the method of moments.

The Fisher Information matrix has been used in a number of previous studies considering variogram uncertainty. Pardo-Igúquiza and Dowd (2001b) and Pardo-Igúquiza (1997) provide FORTRAN code to compute the inverse Fisher Information matrix for method of moments variogram estimates and maximum likelihood variogram estimates respectively. Todini (2001) and Zimmerman and Cressie (1992) have combined the inverse Fisher Information matrix with a further Taylor series to demonstrate how ordinary kriging errors propagate from variogram errors. Müller and Zimmerman (1999), Bogaert and Russo (1999) and Lark (2002) minimize functions of the Fisher Information matrix to optimize sample schemes for variogram estimation.

Marchant and Lark (2004) explored the accuracy of the inverse Fisher Information matrix as a measure of variogram uncertainty via a simulation study. They found that, whilst the uncertainty of nugget and sill parameters of spherical and exponential variograms was accurately estimated, the uncertainty of range estimates was often under-estimated. Thus variogram uncertainty should be assessed via the inverse Fisher Information matrix with some caution. This is particularly true for parameters such as the variogram range a which have a highly nonlinear relationship with the negative log-likelihood function that is poorly approximated by a first order Taylor series. For this reason Zimmerman and Cressie (1992) assumed the variogram range was known when they considered the propagation of variogram errors and only allowed the nugget and sill to be fitted.

Ordinary Kriging

Ordinary kriging estimates the value of a random variable Z at location $\mathbf{x}_0$ by a weighted average of the value of Z measured at n_k sample locations $\mathbf{x}_i,\ i = 1, \ldots n_\mathrm{k}$ in the vicinity of $\mathbf{x}_0$. Thus

$$\hat{Z}\left(\mathbf{x}_0; \boldsymbol{\theta}\right) = \sum_{i=1}^{n_\mathrm{k}} \lambda_i z\left(\mathbf{x}_i\right), \tag{7}$$

where $\hat{Z}(\mathbf{x}_0)$ denotes the estimated value of Z at $\mathbf{x}_0$, $z(\mathbf{x}_i)$ the sampled value at $\mathbf{x}_i$ and λ_i the weights. The weights must sum to unity to ensure that the kriged estimate is unbiased and their chosen values ensure that the kriging variance $\sigma_\mathrm{k}^2(\mathbf{x}_0; \boldsymbol{\theta})$ defined as $E[(\hat{Z}(\mathbf{x}_0; \boldsymbol{\theta}) - z(\mathbf{x}_0; \boldsymbol{\theta}))^2]$ at $\mathbf{x}_0$ is minimized. The vector of weights $\boldsymbol{\lambda}$ may be determined by

$$\boldsymbol{\lambda} = \mathbf{A}^{-1}\mathbf{b}, \tag{8}$$

where

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{n_k} \\ \psi(\mathbf{x_0}) \end{bmatrix}, \tag{9}$$

$$\mathbf{A} = \begin{bmatrix} \gamma(|\mathbf{x}_1 - \mathbf{x}_1|) & \gamma(|\mathbf{x}_1 - \mathbf{x}_2|) & \dots & \gamma(|\mathbf{x}_1 - \mathbf{x}_{n_\mathrm{k}}|) & 1 \\ \gamma(|\mathbf{x}_2 - \mathbf{x}_1|) & \gamma(|\mathbf{x}_2 - \mathbf{x}_2|) & \dots & \gamma(|\mathbf{x}_2 - \mathbf{x}_{n_\mathrm{k}}|) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(|\mathbf{x}_{n_\mathrm{k}} - \mathbf{x}_1|) & \gamma(|\mathbf{x}_{n_\mathrm{k}} - \mathbf{x}_2|) & \dots & \gamma(|\mathbf{x}_{n_\mathrm{k}} - \mathbf{x}_{n_\mathrm{k}}|) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}, \tag{10}$$

and

$$\mathbf{b} = \begin{bmatrix} \gamma(|\mathbf{x}_1 - \mathbf{x}_0|) \\ \gamma(|\mathbf{x}_2 - \mathbf{x}_0|) \\ \vdots \\ \gamma(|\mathbf{x}_{n_\mathrm{k}} - \mathbf{x}_0|) \\ 1 \end{bmatrix}, \tag{11}$$

see, for example, Webster and Oliver (2001). Here, $\gamma(|\mathbf{x}_i - \mathbf{x}_j|)$ denotes the semivariance of Z between the locations $\mathbf{x}_i$ and $\mathbf{x}_j$ and $\psi(\mathbf{x}_0)$ is a Lagrange multiplier. The kriging variance, denoted $\sigma^2_\mathrm{k}(\mathbf{x}_0; \boldsymbol{\theta})$, may be determined from

$$\sigma^2_\mathrm{k}(\mathbf{x}_0; \boldsymbol{\theta}) = \mathbf{b}^\mathrm{T}\boldsymbol{\lambda}. \tag{12}$$

Propagation of Variogram Uncertainty into Kriging

If the variogram parameter vector $\boldsymbol{\theta}$ is known then $z(\mathbf{x_0})$ is a realization of a Gaussian process with mean $\hat{Z}(\mathbf{x}_0; \boldsymbol{\theta})$ and variance $\sigma^2_\mathrm{k}(\mathbf{x}_0; \boldsymbol{\theta})$. When an estimated variogram parameter vector $\hat{\boldsymbol{\theta}}$ is used however the uncertainty associated with this estimate propagates into the kriged estimate. The additional component of uncertainty in the kriged estimate due to variogram uncertainty was approximated by Zimmerman and Cressie (1992) using a first order Taylor series (Heuvelink, 1998). They found that when uncertainty in the estimate of $\boldsymbol{\theta}$ was included

$$\zeta = \mathrm{E}\left[\hat{Z}(\mathbf{x}_0;\boldsymbol{\theta})\right] \approx \hat{Z}(\mathbf{x}_0;\hat{\boldsymbol{\theta}}), \tag{13}$$

$$\tau^2 = \mathrm{E}[\big(\hat{Z}(\mathbf{x}_0;\boldsymbol{\theta}) - \mathrm{E}\left[z(\mathbf{x}_0;\boldsymbol{\theta})\right]\big)^2] = \sum_{i=1}^{q}\sum_{j=1}^{q}\left(\mathrm{Cov}\left(\theta_i,\theta_j\right)\frac{\partial \hat{Z}}{\partial \theta_i}\frac{\partial \hat{Z}}{\partial \theta_j}\right), \tag{14}$$

and that

$$\mathrm{E}\left[\sigma_{\mathrm{k}}^2(\mathbf{x}_0;\boldsymbol{\theta})\right] = \sigma_{\mathrm{k}}^2(\mathbf{x}_0;\hat{\boldsymbol{\theta}}). \tag{15}$$

Abt (1999) considered extensions of this expression but through simulation studies found that Eq. (15) was sufficient. Therefore when variogram uncertainty is included, the distribution of $Z(\mathbf{x}_0)$ may be approximated by a Gaussian process with mean $\hat{Z}(\mathbf{x}_0;\hat{\boldsymbol{\theta}})$ and variance $\sigma_{\mathrm{k}}^2(\mathbf{x}_0;\hat{\boldsymbol{\theta}}) + \tau^2$ where τ^2 is approximated by Eq. (14).

The $\mathrm{Cov}(\theta_i, \theta_j)$ terms in Eq. (14) may be approximated by entries of the inverse Fisher information matrix for $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Once the response data $\mathbf{z}$ have been collected then the derivatives in Eq. (14) may be calculated numerically. However, we wish to assess the effects of variogram uncertainty prior to sampling, i.e we wish to calculate

$$\mathrm{E}\left[\tau^2\right] \approx \sum_{i=1}^{q}\sum_{j=1}^{q}\left(\mathrm{Cov}\left(\theta_i,\theta_j\right)\mathrm{E}\left[\frac{\partial \hat{Z}}{\partial \theta_i}\frac{\partial \hat{Z}}{\partial \theta_j}\right]\right). \tag{16}$$

Todini and Ferraresi (1996) note that the derivatives in Eq. (14) may be calculated exactly by

$$\frac{\partial \hat{Z}}{\partial \theta_i} = \mathbf{z}^{\mathrm{T}}\frac{\partial \boldsymbol{\lambda}}{\partial \theta_i} = \mathbf{z}^{\mathrm{T}}\mathbf{A}^{-1}\left(\frac{\partial \mathbf{b}}{\partial \theta_i} - \frac{\partial \mathbf{A}}{\partial \theta_i}\mathbf{A}^{-1}\mathbf{b}\right). \tag{17}$$

Equation (17) requires that the sampled response data vector $\mathbf{z}$ is known but we note that

$$\mathrm{E}\left[\frac{\partial \hat{Z}}{\partial \theta_i}\frac{\partial \hat{Z}}{\partial \theta_j}\right] = \mathrm{E}\left[\mathbf{z}^{\mathrm{T}}\frac{\partial \boldsymbol{\lambda}}{\partial \theta_i}\mathbf{z}^{\mathrm{T}}\frac{\partial \boldsymbol{\lambda}}{\partial \theta_j}\right], \tag{18}$$

$$= \frac{\partial \boldsymbol{\lambda}^{\mathrm{T}}}{\partial \theta_i}\mathrm{E}[\mathbf{z}\mathbf{z}^{\mathrm{T}}]\frac{\partial \boldsymbol{\lambda}^{\mathrm{T}}}{\partial \theta_j}. \tag{19}$$

If we assume that the mean of $\mathbf{z}$ is zero, or that the mean value of $\mathbf{z}$ has been subtracted from $\mathbf{z}$, then

$$\mathrm{E}\left[\frac{\partial \hat{Z}}{\partial \theta_i}\frac{\partial \hat{Z}}{\partial \theta_j}\right] = \frac{\partial \boldsymbol{\lambda}^{\mathrm{T}}}{\partial \theta_i}\mathbf{V}\frac{\partial \boldsymbol{\lambda}^{\mathrm{T}}}{\partial \theta_j}, \tag{20}$$

where $\mathbf{V}$ is the covariance matrix of $\mathbf{z}$ with all the entries of the $(n_{\mathrm{k}}+1)$th row and column equal to zero. Thus $\mathrm{E}[\tau^2]$ may be approximated by substituting Eq. (20) and entries of the inverse Fisher information matrix into Eq. (16). This requires a variogram parameter vector $\boldsymbol{\theta}$ to be specified but does not require the sampled response data. Therefore the mean squared error (MSE) of the kriged estimate accounting for variogram uncertainty may be approximated prior to sampling by

$$\begin{aligned}\sigma_{\mathrm{T}}^2 = \mathbf{b}^{\mathrm{T}}\boldsymbol{\lambda} + \sum_{i=1}^{q}\sum_{j=1}^{q}\Bigg(\left[\mathbf{F}(\hat{\boldsymbol{\theta}})\right]_{ij}\mathbf{A}^{-1}\left(\frac{\partial \mathbf{b}}{\partial \theta_i} - \frac{\partial \mathbf{A}}{\partial \theta_i}\mathbf{A}^{-1}\mathbf{b}\right)\\ \times \mathbf{V}\mathbf{A}^{-1}\left(\frac{\partial \mathbf{b}}{\partial \theta_j} - \frac{\partial \mathbf{A}}{\partial \theta_j}\mathbf{A}^{-1}\mathbf{b}\right)\Bigg).\end{aligned} \tag{21}$$

Objective Functions for Optimized Sample Schemes

Generally, optimized sample schemes are designed by selecting the scheme $\mathbf{S}$ which minimizes some objective function $\phi(\mathbf{S})$ which is a measure of error in a survey. The earliest approaches to the design of optimal sample schemes for variogram estimation were based upon practical considerations of applying the method of moments. Russo (1984) suggested that the smoothing effect of using lag classes could be reduced if the variation of lag distances within a class was minimized. Warrick and Myers (1987) proposed that the distribution of pairs of sample locations between the lag classes should be controlled to maximize the conformity to a pre-specified function. However, van Groenigen, Siderius, and Stein (1999) found that sample configurations which optimized the Warwick and Myers (1987) criterion for a uniform distribution of lag distances gave less precise variogram estimates than regular square grids.

A number of authors have suggested that sample schemes may be optimized for variogram estimation by selecting a function of the inverse Fisher information matrix as the objective function. Bogaert and Russo (1999) and Müller and Zimmerman (1999) selected the determinant of this matrix for variogram estimation by the method of moments. Lark (2002) considered reconnaissance surveys where the variogram is estimated by maximum likelihood. Following the reconnaissance survey the variogram estimate is used to select a regular survey which is dense enough to ensure that a threshold on the kriging variance is satisfied everywhere. Lark (2002) minimized the precision of the kriging variance at the centre of a regular grid cell. This is a function of the inverse Fisher information matrix. Van Groenigen, Siderius, and Stein (1999) minimized the mean or maximum kriging variance over the study area to optimize sample schemes for ordinary kriging. Heuvelink, Brus, and De Gruijter (2004) optimized sample schemes for universal kriging.

All of the objective functions described above are functions of the variogram. Therefore a particular variogram must be assumed before the sample schemes can be optimized. Lark (2002) observed that the optimized schemes were very sensitive to changes in the assumed variogram parameters. Van Groenigen (2000) found that the sampling configurations which minimized the kriging variance were less sensitive to changes in variogram parameters. Generally the schemes which minimized the mean kriging variance across a region consisted of regularly spaced sampling locations although sampling locations did get shifted towards the boundary if the nugget to sill ratio became large. Since the variogram is unknown prior to sampling the influence of variogram parameters on optimal schemes creates a problem. To counter this Bogaert and Russo (1999) suggested that sampling should be optimized within a Bayesian framework. Rather than assuming a single variogram they assumed a particular variogram model and that prior to sampling the parameters of this model were characterized by a multivariate probability distribution $f(\boldsymbol{\theta})$ which was pre-specified over a set of plausible variogram parameters $\boldsymbol{\Theta}$. Once the response data had been collected the prior distribution $f(\boldsymbol{\theta})$ was updated by Bayes Rule. Their objective function became the expectation of the determinant of $\mathbf{F}^{-1}$ over this distribution of $\boldsymbol{\theta}$ values *i.e.*

$$\phi(\mathbf{S}) = \int_{\boldsymbol{\Theta}} \det[\mathbf{F}^{-1}(\boldsymbol{\theta})] f(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \tag{22}$$

If no data is available the choice of $f(\boldsymbol{\theta})$ is somewhat arbitrary. If some data is available Pardo-Igúquiza and Dowd (2003) describe how a probability distribution function based upon this data may be derived.

METHODS

Testing the Expression for σ_T^2

In this paper we have proposed Eq. (21) as an approximate expression for σ_T^2, the total variance of estimates from ordinary kriging accounting for variogram uncertainty. According to our approximation, variogram uncertainty does not introduce bias to ordinary kriging so σ_T^2 is also an approximation to the mean squared error (MSE). In deriving the expression for σ_T^2 we made a number of assumptions. First we assumed that the relationship between the derivative of the negative log-likelihood function for the fitted variogram and $\boldsymbol{\theta}$, and the relationship between the kriged estimate $\hat{Z}(\mathbf{x}; \boldsymbol{\theta})$ and $\boldsymbol{\theta}$ could be accurately approximated by first order Taylor series for all probable values of $\boldsymbol{\theta}$. Also we assumed that the variogram parameters and the kriged estimates had Gaussian distributions. These assumptions may not always be valid. For example, the Taylor series assumption may be invalid for highly nonlinear relationships or when the variances of the $\boldsymbol{\theta}$ values are large. The later case would require the Taylor series approximation to apply for a large portion of the $\boldsymbol{\theta}$ domain. Pardo-Igúzquiza and Dowd (2003) present examples where the distributions of $\boldsymbol{\theta}$ and $\hat{Z}$ are non-Gaussian. Therefore we carry out tests on simulated variables for different variogram parameters to see when Eq. (21) is a valid approximation.

The test region $\mathbf{X}$ consists of 42.3 hectares area of land in Silsoe, Bedfordshire, UK. The boundary of this region is shown in Figure 1. The approximation for σ_{T}^2 is tested by simulating variables over $\mathbf{X}$. All simulated realizations of $\mathbf{z}$ within this paper are generated from an LU decomposition of the joint covariance matrix $\mathbf{V}$ (Deutsch and Journel, 1998). In each test 10, 000 realizations of a variable

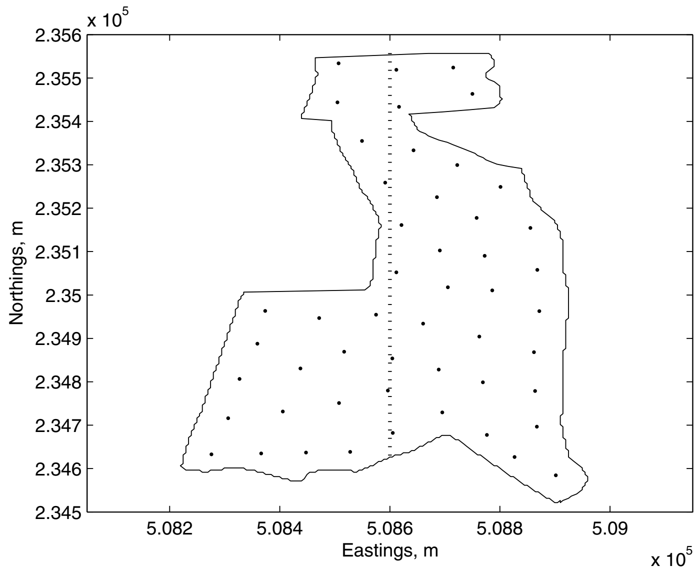

Figure 1. The boundary (continuous line) of the set of fields at Silsoe, Bedfordshire, England, considered in this study. The dispersed dots mark the position of the evenly spaced locations at which sample response data is simulated and the dotted line marks the position of the transect on which the kriged estimate of the variable is compared with the simulated value. Coordinates (m) on UK National Grid; expressed relative to 508 200, 234 500 on subsequent figures.

Table 1. Variogram Parameters for the Simulation Tests of the Expression for the Total Variance in an Ordinary Kriging Survey σ_{T}^2

Test	c_0	c_1	a/m
1	0.25	0.75	240.0
2	0.45	0.75	240.0
3	0.05	0.75	240.0
4	0.25	1.25	240.0
5	0.25	0.50	240.0
6	0.25	0.75	180.0
7	0.25	0.25	300.0

with a specified spherical variogram ($\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$) are simulated at the 50 locations marked by dots in Figure 1 and at a further 90 locations separated by 20.0m along the transect marked in Figure 1. The set of 50 locations represent the sampling locations in a survey. This sample response data is used to krige estimates of the variable at each of the 90 locations along the transect with estimated variogram parameter vectors that are subject to uncertainty. Motivated by Pardo-Igúzquiza and Dowd (2001a) we assume that the estimates of $\boldsymbol{\theta}$ have a covariance matrix which may be approximated by $\mathbf{F}^{-1}(\tilde{\boldsymbol{\theta}})$. For each realization, a different set of estimated variogram parameters to be used for ordinary kriging is sampled from the Gaussian distribution with the mean value given by $\tilde{\boldsymbol{\theta}}$ and covariance matrix given by $\mathbf{F}^{-1}(\tilde{\boldsymbol{\theta}})$. The estimated variogram parameter values are censored to ensure all of the parameters are non-negative. Thus the simulation tests only validate the expression for the propagation of variogram errors and not the expression for the variogram errors themselves. The mean squared difference between the kriged values and the simulated values over the 10,000 realizations is calculated for each of the 90 locations. These values are compared with the values of σ_{T}^2 calculated from Eq. (21). The tests are repeated for the seven sets of variogram parameters listed in Table 1.

Optimizing Sample Schemes for Specified Variograms

In this paper we wish to design sample schemes which are optimal for the whole kriging process including variogram estimation. Therefore as an objective function we select the mean value over the study region $\mathbf{X}$ of the approximate expression for the total variance σ_{T}^2 given by Eq. (21). We denote this objective function as $\phi_{\mathrm{T}}(\mathbf{S})$ and initially assume that the variogram is spherical with known parameter vector $\boldsymbol{\theta}$ such that

$$\phi_{\mathrm{T}}(\mathbf{S}; \boldsymbol{\theta}) = \mathrm{E}\left[\sigma_{\mathrm{T}}^2(\mathbf{x}|\boldsymbol{\theta})\right]. \tag{23}$$

Table 2. Spherical Variogram Parameters c_0, c_1 and a and Number of Sample Locations N for eachSample Scheme Optimization

Case	c_0	c_1	a /m	N
1	0.127	1.000	240.0	50
2	0.127	1.500	240.0	50
3	0.127	0.500	240.0	50
4	0.127	1.000	120.0	50
5	0.127	1.000	90.0	50
6	0.127	1.000	90.0	125

This objective function is approximated by the average value of $\sigma_{\mathrm{T}}^2(\mathbf{X}; \boldsymbol{\theta})$ on 1071 points on a regular grid of interval 20.0m within the boundaries of $\mathbf{X}$. Spatial simulated annealing (SSA), a stochastic optimization algorithm devised by van Groenigen, Siderius, and Stein (1999), is used to find N point sample schemes that minimize Eq. (23) for different assumed spherical variogram parameter vectors. To reduce the required computational load the sample locations may only be on the regular 20.0m grid. The values of N and the spherical variogram parameter vectors $\boldsymbol{\theta}$ used in the optimizations are listed in Table 2. The resulting sample schemes are compared to explore the sensitivity of the pattern of optimal sampling configurations to variogram parameters. For each set of variogram parameters SSA is used to design a second sample scheme which minimizes the mean kriging variance σ_{k}^2 over $\mathbf{X}$. The variance of variogram parameters predicted by the inverse Fisher information matrix, the kriging variance and the total variance are compared for the schemes optimized with each objective function.

Optimizing Sample Schemes for Distributions of Variogram Parameters

Since the objective function described by Eq. (23) varies according to the variogram parameters we follow Bogaert and Russo (1999) and consider a Bayesian framework where $\boldsymbol{\theta}$ is described by a distribution function $f(\boldsymbol{\theta})$. In this paper we assume that $f(\boldsymbol{\theta})$ has uniform density over a space of plausible variogram parameters $\boldsymbol{\Theta}$ and that $f(\boldsymbol{\theta})$ is zero outside $\boldsymbol{\Theta}$. Then the objective function becomes

$$\phi_{\mathrm{BT}}(\mathbf{S}) = \int_{\boldsymbol{\Theta}} \phi_{\mathrm{T}}(\mathbf{S}; \boldsymbol{\theta})\, f(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \tag{24}$$

The integral in Eq. (24) is approximated by calculating the values of $\phi_{\mathrm{T}}(\mathbf{S}; \boldsymbol{\theta})$ and $f(\boldsymbol{\theta})$ on a mesh of $\boldsymbol{\theta}$ values in q dimensional space. The minimization of Eq. (24) is therefore very costly to compute since each perturbation of a location in the SSA algorithm requires σ_{T}^2 to be calculated across $\boldsymbol{\Theta}$ and $\mathbf{X}$. Therefore we present

Table 3. Bounds on $\boldsymbol{\Theta}$ the Set of Plausible Variogram Parameters and Values of N Used to Optimize Sample Schemes within a Bayesian Framework

Case	Min c_0	Max c_0	Min c_1	Max c_1	Min a/m	Max a/m	N
7	0.0	0.8	0.1	1.3	3.0	27.0	30
8	0.0	0.3	0.4	1.5	90.0	290.0	100
9	0.0	0.3	0.4	1.5	50.0	250.0	100

only one direct minimization of Eq. (24) and rather than using the field shown in Fig. 1 we optimize $N = 30$ locations over a square region of length 30.0 m. The bounds on $\boldsymbol{\Theta}$ are listed as Case 7 in Table 3.

The number of computations required to minimize Eq. (24) may be reduced if the optimization is divided into two parts. Initially a N_{ok} location scheme which minimizes the mean kriging variance across $\mathbf{X}$ is designed. This requires that the variogram parameters are selected prior to the optimization but since optimal kriging variance schemes are relatively insensitive to the variogram parameters the choice should have little effect. The second part of the optimization selects the position of N_{v} sample locations which are optimal for estimating the variogram. The corresponding objective function is minimized at a single point in $\mathbf{X}$, denoted $\mathbf{X}_{\mathrm{opt}}$. The point $\mathbf{X}_{\mathrm{opt}}$ is the position more than 50m from the boundary of $\mathbf{X}$ which would have the largest kriging variance if the N_{v} location survey was applied. The objective function for the second part of the optimization is τ^2 at $\mathbf{X}_{\mathrm{opt}}$, that is the variance component of a kriged estimate due to variogram uncertainty at $\mathbf{X}_{\mathrm{opt}}$. This is approximated by Eq. (14). The covariance matrix in Eq. (14) is calculated using all of the $N_{\mathrm{ok}} + N_{\mathrm{v}}$ locations. However, the derivatives of $\hat{Z}$ are calculated using only the N_{ok} locations from the kriging variance phase of the optimization. Otherwise the objective function would be made zero if one of the N_{v} locations was positioned at $\mathbf{X}_{\mathrm{opt}}$. At the end of the second optimization phase the full objective function described by Eq. (24) is calculated. The choice of N_{ok} and N_{v} are arbitrary so the process is repeated with different combinations all summing to N and then scheme with the lowest value of $\phi_{\mathrm{BT}}(\mathbf{S})$ is selected. We present two 100 location examples of schemes which minimize Eq. (24) by this approach. The bounds on $\boldsymbol{\Theta}$ are listed as Cases 8 and 9 in Table 3.

RESULTS

Propagation of Variogram Uncertainty

Figure 2 shows the approximate MSE from Eq. (21), the kriging variance and the simulated MSE at each of the 90 test sites for a spherical variogram function

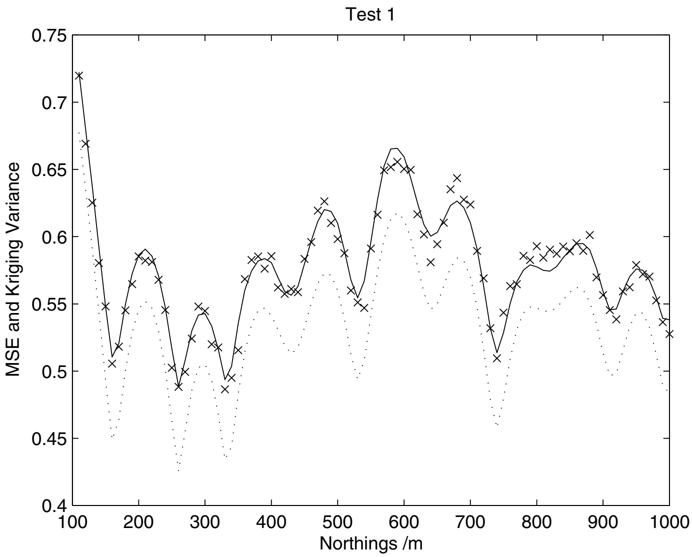

Figure 2. Predicted MSE (continuous line), kriging variance (dotted line) and simulated MSE (crosses) along the transect for a spherical variogram with $c_0 = 0.25$, $c_1 = 0.75$ and $a = 240.0\,\text{m}$.

with $c_0 = 0.25$, $c_1 = 0.75$ and $a = 240.0\,\text{m}$. The interval between sample points is approximately 100m. The simulated MSE is noticeably larger than the kriging variance at each of the locations. Thus for this sample scheme and variogram combination, variogram uncertainty has an observable effect on the precision of kriged estimates. This effect is well approximated by the theoretical expression for τ^2 given by Eq. (14).

Figure 3 shows the results of further simulated tests using the Test 2–5 variogram parameter vectors listed in Table 1. In these tests the nugget c_0 and sill c_1 parameters are varied. The difference between the MSE and kriging variance increases with both the nugget and sill parameters and the simulated MSE is reasonably approximated by Eq. (21) in each case. However the theoretical expression is less effective for Test 6 shown in Figure 4. In this example the variogram range has been reduced to 180.0 m. The sample scheme shown in Figure 1 is less suitable for estimating the variogram in this case since a lower proportion of pairs of locations are separated by distance less than the range of spatial correlation. Therefore the inverse Fisher information matrix predicts large variances for each parameter. The mean values plus or minus standard errors for c_0, c_1 and a are 0.25 ± 0.64, 0.75 ± 0.68 and $180.0\,\text{m} \pm 85.9\,\text{m}$ respectively. Therefore the distribution of variogram parameters must be significantly censored to ensure that none of the parameters are negative. Thus the distribution of variogram parameters that

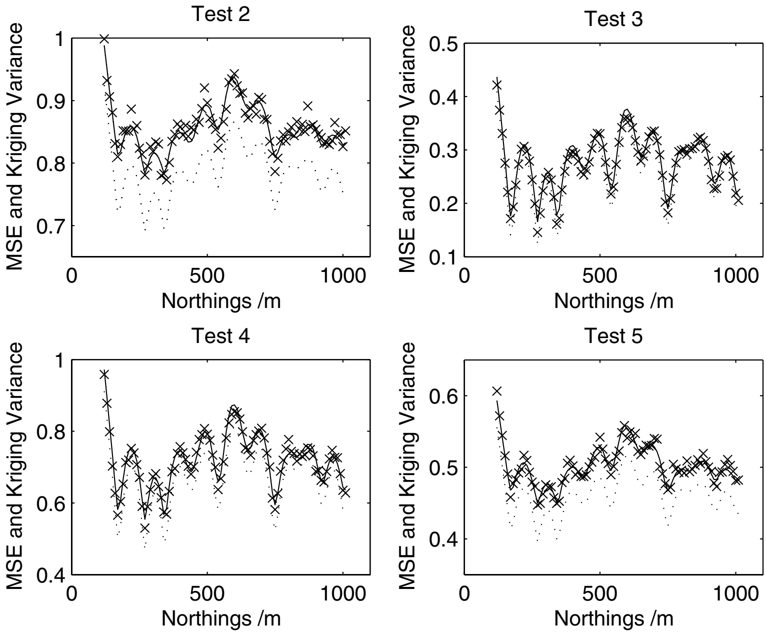

Figure 3. Predicted MSE (continuous line), kriging variance (dotted line) and simulated MSE (crosses) along the transect for the Test 2–5 spherical variograms with parameters as listed in Table 1.

was assumed to estimate the propagated error differs from the actual distribution. In addition the Taylor series that was assumed in estimating the propagated errors is only accurate if the errors in estimating the variogram parameters is small. Thus for situations such as Test 6 where the variogram is not precisely estimated the estimate of the propagated errors is also imprecise. In Test 7 which is also shown on Figure 4 the range is 320.0 m. The mean values plus or minus standard errors are 0.25 ± 0.17, 0.75 ± 0.37 and $320.0\,\text{m} \pm 89.5\,\text{m}$ respectively. Since the variogram errors are smaller and less censoring is required the propagated kriging errors are accurately approximated by (21).

Optimal Sample Schemes for Known Variograms

Figure 5 (KV) shows a 50 location sample scheme optimized to minimize the mean kriging variance across the region $\mathbf{X}$ for a spherical variogram with $c_0 = 0.127$, $c_1 = 1.000$ and $a = 240.0\,\text{m}$. The sample locations in Figure 5 (KV) are evenly spread across the field. A similar pattern of locations is seen when the

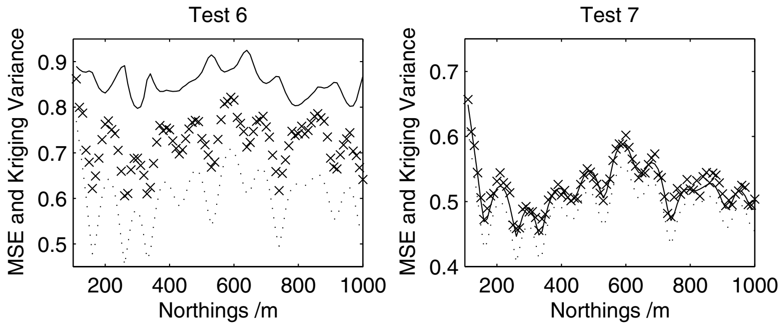

Figure 4. Predicted MSE (continuous line), kriging variance (dotted line) and simulated MSE (crosses) along the transect for the Test 6–7 spherical variograms with parameters as listed in Table 1.

kriging variance is minimized for the other variogram parameter vectors listed in Table 2.

Figure 5 (Cases 1–6) show 50 location sample schemes where the mean total variance σ_T^2 [Eq. (21)] has been minimized for different sets of spherical variogram parameters. Table 4 shows the standard deviation on the variogram parameters predicted by the Fisher information matrix for each of these schemes and sets of variogram parameters, the standard deviation for the same parameters where the kriging variance has been minimized, the kriging variance for each scheme and the mean of the total variance expression for each scheme. For each set of variogram parameters the optimized mean total variance schemes contain locations which are dispersed evenly over the region with a number of closer pairs. The closer pairs cause the standard deviation of the variogram parameters, particularly the nugget effect, to be reduced. These schemes have a larger mean kriging variance than the corresponding schemes where the kriging variance has been minimized, but the mean total variance is lower.

There are clear differences between the optimized mean total variance schemes for different variogram parameters. Case 1 shows the minimized total variance scheme for a spherical variogram with parameter values $c_0 = 0.127$, $c_1 = 1.000$ and $a = 240.0$ m. The majority of locations are spread evenly over the region but there are are three close pairs of locations separated by 20.0 m, the minimum separation distance allowed by the optimization algorithm.

When the sill parameter is increased so that the parameters are $c_0 = 0.127$, $c_1 = 1.5$ and $a = 240.0$ m has little effect of the optimal sampling array in Case 2 but when the nugget is increased in Case 3 there are more sets of close pairs. When the range is reduced to 120.0 m and 90.0 m in Cases 4 and 5 the number of close pairs increases and small transects form. When $a = 90.0$ m the 50 sample locations appear to be insufficient to cover the region. A large number of

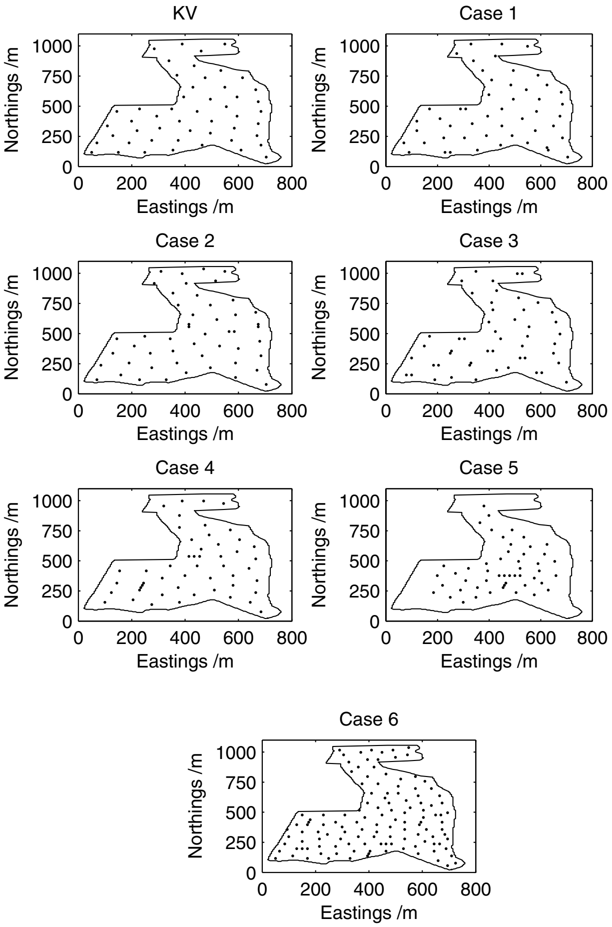

Figure 5. Sample scheme optimized to minimize the mean kriging variance across the trial region for a spherical variogram with parameters $c_0 = 0.127$, $c_1 = 1.000$ and $a = 240.0\,\mathrm{m}$ (KV) and sample schemes optimized to minimize the expected MSE across the trial region for spherical variograms with parameters listed in Table 2 (Cases 2–6).

close pairs are required to adequately describe the variogram which means some portions of the field are unsampled. Case 6 shows 125 locations optimized to minimize the mean total MSE for the same variogram parameters. There are still a number of close pairs arranged in a transect but the whole of the region is now covered.

Table 4. The Predicted Standard Error for each Variogram Parameter, the Mean Kriging Variance $E[\sigma_k^2(\mathbf{X}|\boldsymbol{\theta})]$ and the Mean Total Variance $E[\sigma_T^2(\mathbf{x}|\boldsymbol{\theta})]$ for Each of the Optimized Sample Schemes

Case	$\sigma(c_0)$	$\sigma(c_1)$	$\sigma(a)$/m	$E[\sigma_k^2(\mathbf{x}|\boldsymbol{\theta})]$	$E[\sigma_T^2(\mathbf{x}|\boldsymbol{\theta})]$
1 σ_T^2	0.18	0.37	50.85	0.46	0.48
1 σ_k^2	0.24	0.42	53.79	0.45	0.49
2 σ_T^2	0.31	0.42	33.42	0.78	0.81
2 σ_k^2	4.61	4.62	96.90	0.76	5.78
3 σ_T^2	0.85	0.89	121.94	0.94	0.95
3 σ_k^2	77.84	77.84	374.36	0.91	971.23
4 σ_T^2	0.17	0.59	48.95	0.60	0.63
4 σ_k^2	0.33	0.61	50.30	0.60	0.64
5 σ_T^2	0.28	0.46	71.80	0.94	0.97
5 σ_k^2	0.42	0.56	81.47	0.92	0.99
6 σ_T^2	0.26	0.32	15.00	0.65	0.66
6 σ_k^2	0.86	0.88	26.44	0.63	0.81

Optimal Sample Schemes for Distributions of Variogram Parameters

Figure 6 shows a 30 location sample which minimizes the mean value of ϕ_{BT} in a Bayesian framework for the Case 7 distribution of variogram parameters. As was the case when the variogram was known the optimized scheme has a number

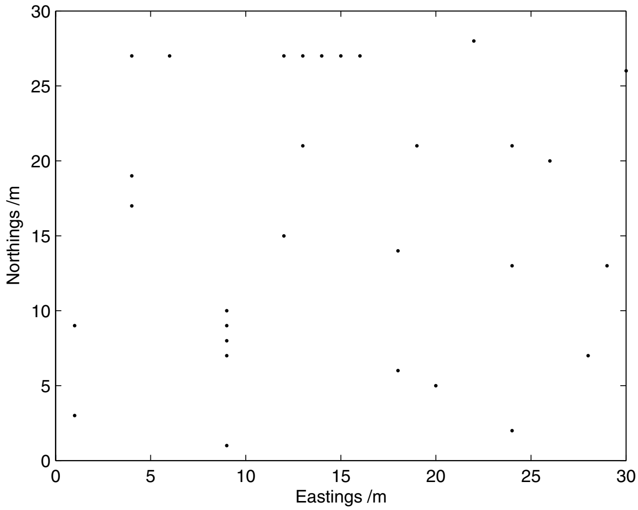

Figure 6. Sample scheme consisting of 30 locations designed by minimization of $\phi_{BT}(\mathbf{S})$ for the Case 7 distribution of variogram parameters.

of locations spread relatively evenly over the region although a few close pairs are included to provide information about the degree of spatial correlation over short distances.

This optimization requires a large number of computations. Table 5 shows the results from minimizing ϕ_{BT} by the computationally cheaper two stage method. A total of 100 locations are divided into two sets. One set is used to minimize the kriging variance, σ_{k}^2; the other set to minimize the additional variance due to variogram uncertainty, τ^2. Table 5 shows how the mean values of σ_{k}^2, τ^2 and σ_{T}^2 vary with N_{ok}, the number of locations devoted to minimizing σ_{k}^2 for the Case 8 and Case 9 distributions of variogram parameters. Generally σ_{k}^2 decreases and τ^2 increases with N_{ok}. There is some fluctuation but this may merely be an artefact due to the stochastic nature of the SSA algorithm. The rate of decrease of σ_{k}^2 is relatively slow whereas τ^2 increases rapidly as N_{ok} approaches 100. For the Case 8 distribution where the minimum value of a in $\boldsymbol{\Theta}$ was 50.0 m the smallest value of σ_{T}^2 occurred for $N_{\mathrm{ok}} = 80$. In Case 9, when the minimum value of a in $\boldsymbol{\Theta}$ was increased to 90.0 m, the smallest value of σ_{T}^2 occurred for $N_{\mathrm{ok}} = 95$.

DISCUSSION

For the relatively small sample size of 50 locations, estimation errors in the variogram have an observable effect on the accuracy of the kriged estimates. The magnitude of this effect is accurately approximated by Eq. (21) except when the variogram range is small relative to the interval between sampling locations. This is not surprising since when the range is small, few comparisons are made

Table 5. The Mean Kriging Variance $\mathrm{E}[\sigma_{\mathrm{k}}^2(\mathbf{X};\boldsymbol{\theta})]$, Mean Variance Contribution from Variogram Uncertainty $\mathrm{E}[\tau^2(\mathbf{X};\boldsymbol{\theta})]$ and the Mean Total Variance $\mathrm{E}[\sigma_{\mathrm{T}}^2(\mathbf{X}|\boldsymbol{\theta})]$ Across $\mathbf{X}$ for the Schemes Optimized in Case 8 and Case 9 with Different Number of Locations N_{ok} Devoted to Minimize the Kriging Variance

	Case 8			Case 9		
N_{ok}	$\mathrm{E}[\sigma_{\mathrm{k}}^2(\mathbf{x}\|\boldsymbol{\theta})]$	$\mathrm{E}[\tau^2(\mathbf{x}\|\boldsymbol{\theta})]$	$\mathrm{E}[\sigma_{\mathrm{T}}^2(\mathbf{x}\|\boldsymbol{\theta})]$	$\mathrm{E}[\sigma_{\mathrm{k}}^2(\mathbf{x}\|\boldsymbol{\theta})]$	$\mathrm{E}[\tau(\mathbf{x}\|\boldsymbol{\theta})]$	$\mathrm{E}[\sigma_{\mathrm{T}}^2(\mathbf{x}\|\boldsymbol{\theta})]$
55	5.55×10^{-1}	1.65×10^{-2}	5.71×10^{-1}	4.74×10^{-1}	8.67×10^{-3}	4.83×10^{-1}
60	5.53×10^{-1}	1.57×10^{-2}	5.68×10^{-1}	4.62×10^{-1}	1.01×10^{-2}	4.73×10^{-1}
65	5.47×10^{-1}	2.11×10^{-2}	5.68×10^{-1}	4.59×10^{-1}	9.21×10^{-3}	4.69×10^{-1}
70	5.38×10^{-1}	2.46×10^{-2}	5.63×10^{-1}	4.49×10^{-1}	1.14×10^{-2}	4.61×10^{-1}
75	5.51×10^{-1}	1.13×10^{-2}	5.62×10^{-1}	4.46×10^{-1}	1.02×10^{-2}	4.57×10^{-1}
80	5.38×10^{-1}	1.80×10^{-2}	5.56×10^{-1}	4.39×10^{-1}	1.18×10^{-2}	4.50×10^{-1}
85	5.33×10^{-1}	3.89×10^{-2}	5.72×10^{-1}	4.37×10^{-1}	1.15×10^{-2}	4.48×10^{-1}
90	5.45×10^{-1}	4.52×10^{-2}	5.80×10^{-1}	4.38×10^{-1}	1.31×10^{-2}	4.47×10^{-1}
95	5.55×10^{-1}	6.01×10^{-2}	6.07×10^{-1}	4.29×10^{-1}	1.76×10^{-2}	4.47×10^{-1}
100	5.32×10^{-1}	74.9	75.4	4.24×10^{-1}	1.05×10^{-1}	5.29×10^{-1}

between measurements separated by less than the range. Therefore the variogram is poorly approximated and in turn the Taylor series approximation is inaccurate. In this paper optimized sample schemes are considered. These schemes contain sufficient close pairs to estimate the spatial correlation over short distances so the approximation of the effects of variogram uncertainty on kriging should always be reasonable.

The simulation tests only explored the propagation of a known variogram uncertainty into kriging. They assumed that this uncertainty could be accurately approximated by the inverse Fisher information matrix. Although Marchant and Lark (2004) showed that this approximation is generally reasonable, Pardo-Igúzquiza and Dowd (2003) demonstrated that the assumption of variogram parameter estimates of having a multivariate Gaussian distribution is an over-simplification.

The expression for the total variance in an ordinary kriging survey was successfully used to optimize sample schemes when the variogram was known. These schemes ensured a good coverage of the region for kriging and a sufficient number of close locations for variogram estimation. In each case the uncertainty of variogram parameters predicted by the inverse Fisher information matrix was less than that for a scheme of the same size designed by minimization of the kriging variance. The reduction in variogram uncertainty was largest for variograms with a small range. The schemes which minimized the total variance had a marginally larger kriging variance than the schemes which minimized the kriging variance. The pattern of these schemes varied according to the variogram parameters. As the range of spatial correlation decreases the required number of close locations increases. Variables with a small range have to be sampled at a large number of locations to ensure that there is sufficient data for variogram estimation and ordinary kriging across the field. If the relative nugget effect is large the close locations are distributed as pairs whereas for small relative nugget transects of four or five close locations may be seen.

Since the optimal sample scheme does vary according to the variogram parameters it is prudent to adopt an optimization algorithm which considers a number of different variogram parameters. Therefore the optimization was implemented within a Bayesian framework. Rather than assuming a single variogram this method requires the practitioner to select a probability distribution for the variogram parameters. The relative sampling effort devoted to the variogram estimation and kriging processes varies according to the lower bound on the distribution of the range parameter. The selection of the parameter distributions will still be somewhat arbitrary but they may be based upon the intuition of the practitioner combined with previous results from sampling the same variable under similar conditions. Thus the Bayesian framework is more likely to lead to suitable sample schemes, but the computational load in the design of these schemes is a lot larger.

CONCLUSIONS

Abt, Welch, and Sacks (1999) observed that when a variable is intensely sampled relative to its range of spatial correlation a regular survey may be sufficient for variogram estimation. However, we see that when separation of locations in a regular sample scheme is around half the range of spatial correlation, variogram uncertainty may make an observable contribution to the errors in an ordinary kriging survey. For known variograms this error contribution may be accurately approximated prior to sampling via a Taylor series although this approximation does fail if few sample locations are separated by less than the variogram range.

Optimal sample schemes for both variogram estimation and ordinary kriging may be designed if an expression for the total sampling error in an ordinary kriging survey is minimized. This expression includes variogram parameter uncertainty and estimation uncertainty upon kriging. A particular variogram function must be assumed to calculate this total error expression. The resulting schemes are sensitive to the assumed variogram function. Since the variogram function is not known prior to sampling the total error expression should be minimized in a Bayesian framework which allows for variogram uncertainty. The minimization of the total error expression within the Bayesian framework is computationally expensive since each calculation of the total error function must be averaged over the study region and over plausible variogram parameters. The computational load may be reduced if the optimization is split into two parts; one to minimize the ordinary kriging errors and the other to minimize variogram uncertainty.

ACKNOWLEDGMENTS

This work was supported by the Biotechnology and Biological Sciences Research Council of the U.K. through Grant 204/D15335 and by the Home-Grown Cereals Authority of the U.K. through grant 2453.

REFERENCES

Abt, M., 1999, Estimating the prediction mean squared error in Gaussian stochastic processes with exponential correlation structure: Scand. Jour. Statist., v. 26, no. 4, p. 563–578.

Abt, M., Welch, W. J., and Sacks, J., 1999, Design and analysis for modelling and predicting spatial contamination: Math. Geol., v. 31, no. 1, p. 1–22.

Bogaert, P., and Russo, D., 1999, Optimal spatial sampling design for the estimation of the variogram based upon a least squares approach: Water Resour. Res., v. 35, no. 4, p. 1275–1289.

Deutsch, C. V., and Journel, A. G., 1998, GSLIB: Geostatistical software library and users guide, 2nd ed.: Oxford University Press, New York, 369 p.

Dobson, A. J., 1990, An introduction to generalized linear models, 2nd ed.: Chapman and Hall, London, 174 p.

Heuvelink, G. B. M., 1998, Error propagation in environmental modelling with GIS: Taylor, Francis, London, 127 p.

Heuvelink, G. B. M., Brus, D. J., and De Gruijter, J. J., 2004, Optimization of sample configurations for digital soil mapping with universal kriging: Proceedings of Global Workshop on Digital Soil Mapping, INRA, Montpellier.

Journel, A. G., and Huijbregts, C. J., 1978, Mining geostatistics: Academic Press, London, 600 p.

Kitanidis, P. K., 1987, Parametric estimation of covariances of regionalised variables: Water Resources Bulletin, v. 23, no. 4, p. 557–567.

Lark, R. M., 2000, Estimating variograms of soil properties by the method-of-moments and maximum likelihood: Eur. Jour. Soil Sci., v. 51, no. 4, p. 717–728.

Lark, R. M., 2002, Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood: Geoderma, v. 105, no. 1–2, p. 49–80.

Marchant, B. P., and Lark, R. M., 2004, Estimating variogram uncertainty: Math. Geol., v. 36, no. 8, p. 867–898.

Müller, W. G., and Zimmerman, D. L., 1999, Optimal designs for variogram estimation: Environmetrics, v. 10, no. 1, p. 23–37.

Pardo-Igúzquiza, E., 1997, MLREML: A computer program for the inference of spatial covariance parameters by maximum likelihood and restricted maximum likelihood: Comput. & Geosciences, v. 23, no. 2, p. 153–162.

Pardo-Igúzquiza, E., 1998, Maximum likelihood estimation of spatial covariance parameters: Math. Geol., v. 30, no. 1, p. 95–108.

Pardo-Igúzquiza, E., and Dowd, P. A., 2001a, Variance-covariance matrix of the experimental variogram: Assessing variogram uncertainty: Math. Geol., v. 33, no. 4, p. 397–419.

Pardo-Igúzquiza, E., and Dowd, P. A., 2001b, VARIOG2D: A computer program for estimating the semi-variogram and its uncertainty: Comput. & Geosciences, v. 27, no. 5, p. 549–561.

Pardo-Igúzquiza, E., and Dowd, P. A., 2003, Assessment of the uncertainty of spatial covariance parameters of soil properties and its use in applications: Soil Sci., v. 168, no. 11, p. 769–782.

Russo, D., 1984, Design of an optimal sampling network for estimating the variogram: Soil Sc. Soc. Am. J., v. 48, no. 4, p. 708–716.

Todini, E., 2001, Influence of parameter estimation uncertainty in Kriging: Part 1-Theoretical development: Hydrol. Earth Sci. Syst., v. 5, no. 2, p. 215–223.

Todini, E., and Ferraresi, M., 1996, Influence of parameter estimation uncertainty in Kriging: J. Hydrol., v. 175, no. 4, p. 555–566.

van Groenigen, J. W., Siderius, W., and Stein, A., 1999, Constrained optimisation of soil sampling for minimisation of the kriging variance: Geoderma v. 87, no. 3–4, p. 239–259.

van Groenigen, J. W., 2000, The influence of variogram parameters on optimal sampling schemes for mapping by kriging: Geoderma v. 97, no. 3–4, p. 223–236.

Warrick, A. W., and Myers, D. E., 1987, Optimization of sampling locations for variogram calculations: Water Resour. Res., v. 23, no. 3, p. 496–500.

Webster, R., and Oliver, M. A., 2001 Geostatistics for Environmental Scientists: John Wiley, Chichester, 271 p.

Zimmerman, D. L., and Cressie, N., 1992, Mean squared prediction error in the spatial linear model with estimated covariance parameters: Ann. Inst. Statist. Math. v. 44, no. 1, p. 27–43.