Optimal Design for Variogram Estimation

Müller, Werner; Zimmerman, Dale L.

DOI:
10.57938/d73aea87-0c8a-4420-a886-242386d1f265

Published: 01/01/1997

Document Version:
Publisher's PDF, also known as Version of record

Document License:
Unspecified

Optimal Design for Variogram Estimation

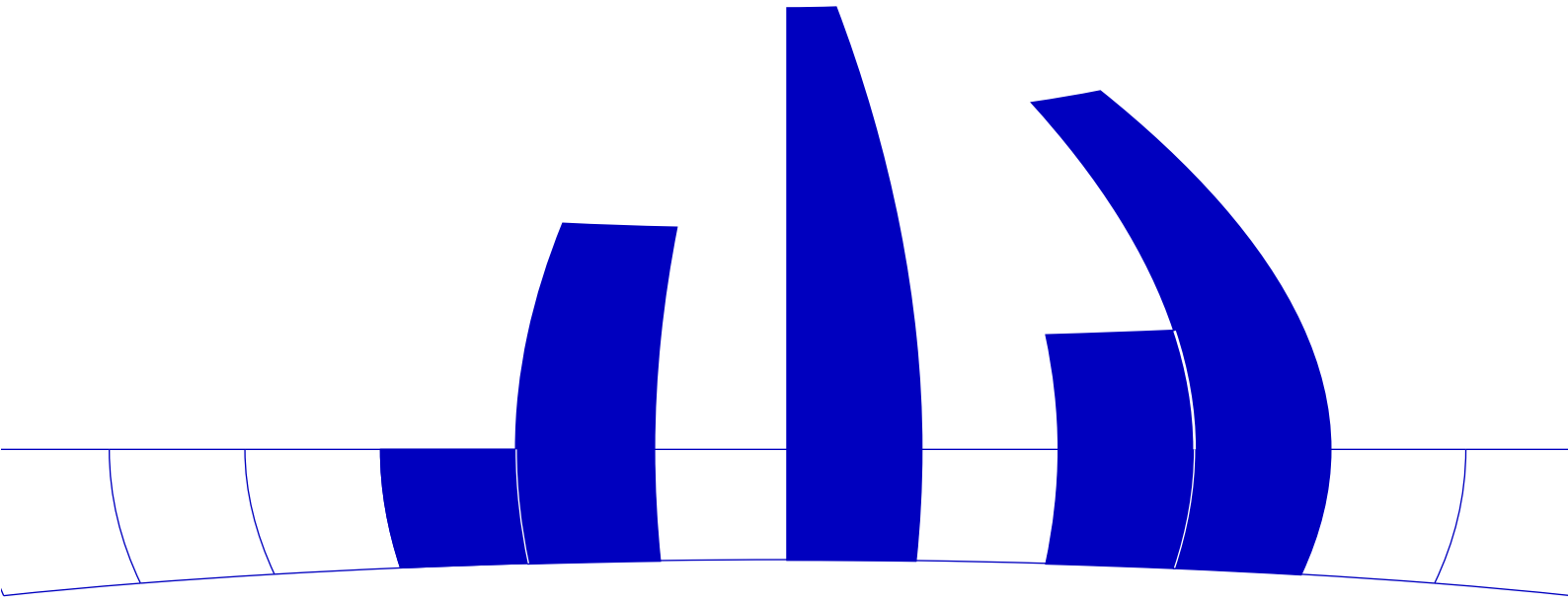
Werner Müller, Dale L. Zimmerman

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 51
November 1997

http://statmath.wu-wien.ac.at/



Optimal Design for Variogram Estimation

Werner G. Müller, Dale L. Zimmerman

No. 51, November 1997

OPTIMAL DESIGNS FOR VARIOGRAM ESTIMATION

Werner G. Müller[1] and Dale L. Zimmerman
Department of Statistics, University of Economics, Vienna,
and
Department of Statistics and Actuarial Science, University of Iowa, Iowa City

Abstract

The variogram plays a central role in the analysis of geostatistical data. A valid variogram model is selected and the parameters of that model are estimated before kriging (spatial prediction) is performed. These inference procedures are generally based upon examination of the empirical variogram, which consists of average squared differences of data taken at sites lagged the same distance apart in the same direction. The ability of the analyst to estimate variogram parameters efficiently is affected significantly by the sampling design, i.e., the spatial configuration of sites where measurements are taken.

In this paper, we propose design criteria that, in contrast to some previously proposed criteria oriented towards kriging with a known variogram, emphasize the accurate estimation of the variogram. These criteria are modifications of design criteria that are popular in the context of (nonlinear) regression models. The two main distinguishing features of the present context are that the addition of a single site to the design produces as many new lags as there are existing sites and hence also produces that many new squared differences from which the variogram is estimated. Secondly, those squared differences are generally correlated, which inhibits the use of many standard design methods that rest upon the assumption of uncorrelated errors.

Several approaches to design construction which account for these features are described and illustrated with two examples. We compare their efficiency to simple random sampling and regular and space-filling designs and find considerable improvements.

KEY WORDS: D-Optimality, Geostatistics, Kriging, Spatial dependence.

1 Introduction

The geostatistical method has proven to be a useful approach to the analysis of spatially correlated data in diverse applications including mining engineering, soil

[1]This joint research was initiated when Werner G. Müller spent a semester as a visiting assistant professor at the University of Iowa. His stay was partially supported by the Fulbright Commission Mutual Educational Exchange Grant, the Exportakademiestipendium der Bundeswirtschaftskammer Österreich and the OeNB-WU-Förderungspreis.

and crop science, hydrology, and ecology. Briefly, it uses data collected at locations in some spatial domain $\mathcal{S} \subset \mathbb{R}^d$ to build models of spatial dependence and to make statistical inferences about the parameters of those models and/or unobserved values of the variate under study. Classically, this is accomplished by carrying out the following sequence of steps:

1. Posit a model of (possibly nonstationary) mean plus intrinsically stationary error for the process (random field) $\{Z(s) : s \in \mathcal{S}\}$ that generated the data. This model has the form

$$Z(s) = \mu(s) + \varepsilon(s), \quad s \in \mathcal{S}, \tag{1}$$

   where the mean, $E\{Z(s)\} \equiv \mu(s)$, is a linear function of one or more unknown large-scale variation (trend) parameters β [that is, $\mu(s) = x(s)'\beta$ for all $s \in \mathcal{S}$] and the error process $\varepsilon(\cdot)$ has mean zero and is intrinsically stationary, i.e., $\mathrm{var}\{\varepsilon(s) - \varepsilon(t)\}$ depends only on $s - t$.

2. Estimate nonparametrically the variogram $\gamma(h)$ of the error process, defined by

$$\gamma(h) = \frac{1}{2}\mathrm{Var}\{\varepsilon(s+h) - \varepsilon(s)\}, \quad s, s+h \in \mathcal{S}. \tag{2}$$

3. Select a valid (conditionally negative semidefinite), generally nonlinear, parametric model $\gamma(h;\theta)$ that is compatible with a plot of the estimated variogram. Here, θ is a $p \times 1$ vector of unknown parameters belonging to a specified parameter space $\Theta \subset \mathbb{R}^p$.

4. Fit the chosen model to the estimated variogram to estimate the model's parameters.

In some instances [e.g. Hoeksema and Kitanidis (1985)] the objective of the analysis is merely to characterize the spatial dependence structure of the process, in which case the analysis terminates with Step 4. Usually, however, the ultimate goal of the analysis is the prediction of unobserved values of the process, in which case there is a fifth step:

5. "Krige," (i.e., predict) the unobserved value(s) and estimate the corresponding variance(s) of prediction error.

The "standard" prediction procedure (there are several variations) is universal kriging, which is merely best linear unbiased prediction under the assumptions that the chosen variogram model is correct and that the estimates of its parameters are the true parameter values. A complete treatment of the geostatistical method can be found, for instance, in the books by Isaaks and Srivastava (1989) and Cressie (1993).

The focus of this paper is on Step 4, the variogram parameter estimation component of a geostatistical analysis. The ability of an investigator to estimate the parameters of a variogram model efficiently is affected significantly by the sampling design, i.e., the locations of the sites $s_1, \ldots, s_n \in \mathcal{S}$ where data are taken. For instance, if $Z(\cdot)$ is isotropic (i.e., if the variogram depends only on the separation distance $r = (h'h)^{1/2}$ between sites) and no two sites are within a distance $\epsilon > 0$ of each other, then it will be difficult, if not impossible, to estimate precisely those parameters whose primary influence is on the behavior of the variogram at distances less than ϵ. Furthermore, Stein (1988) has shown that, for efficient prediction, proper identification of the variogram is much more important at "small" distances than at "large" distances.

Because kriging is usually the ultimate objective of a geostatistical analysis, most authors who have considered design aspects in this context have emphasized the utility of designs for prediction under an assumption that both the functional form of the variogram and the values of its parameters are known; see Bras and Rodriguez-Iturbe (1976), McBratney, Webster, and Burgess (1981), Yfantis, Flatman, and Behar (1987), Barnes (1989), Cressie, Gotway, and Grondona (1990), and Benedetti and Palma (1995). Inasmuch as the variogram must in practice be estimated, however, a total emphasis on prediction with no regard to the utility of the design for variogram estimation seems, to the authors, to "put the cart before the horse." In this paper, we propose reasonable criteria for choosing, and compare several practical approaches for constructing, sampling designs that emphasize the accurate estimation of the variogram. These designs could be adopted at the early stages of a sampling program until the variogram is sufficiently well-estimated, after which one could shift to an existing approach that emphasizes prediction.

Of course, there is a large literature on designs for estimating parameters in the (linear and nonlinear) regression context [e.g. the books by Fedorov (1972), Pázman (1986), and Pukelsheim (1993)]. Many aspects of the design approach proposed herein are adaptations and/or modifications of existing approaches to the peculiarities of the present context. A fundamental difference between the two contexts is the "order" of the random process's properties which are the objects of attention. In the regression context attention focusses on the expectation function, i.e., the first-order properties of the process. In the present context primary interest lies in the variogram, which characterizes the second-order properties. When the process is second-order stationary, an alternative characterization of its second-order properties is the covariance function $C(h)$, which is related to the variogram through the equation $\gamma(h) = C(0) - C(h)$. Modelling and estimating the variogram, rather than the covariance function, has several advantages, however; see Robinson (1990) and Cressie and Zimmerman (1992). Design for estimating second-order properties has attracted considerable interest in other areas of statistics as well, e.g. computer simulation experiments [Sacks et al. (1989)].

Russo (1984), Warrick and Myers (1987), and Morris (1991) have also considered design criteria for variogram estimation. Russo considered a design to be optimal if

it minimized the dispersion of lags (i.e., displacements between the sites $s_1, ..., s_n$) within lag classes, while Warrick and Myers proposed a criterion that measures how well the lag distribution corresponding to a design conforms to a prespecified distribution. The approaches presented here deal with the more fundamental issue of what that prespecified distribution should be. Morris proposed refinements of the Russo and Warrick-Myers approaches using an index called the maximum equivalent uncorrelated pairs, which accounts for correlations among estimates of $\gamma(h)$ at different lags. Our approach also attempts to account for these correlations.

Pettitt and McBratney (1993) have suggested designs for variogram estimation in the absence of any knowledge about the scales of variation. Such designs could be used as starting designs for the algorithms presented in this paper, which require some prior knowledge about the form and magnitude of local variation.

2 Background

2.1 Variogram Models

A function that satisfies the following important properties is said to be a valid variogram model:

1. $\gamma(\cdot)$ is "conditionally" negative semidefinite, i.e., $\sum_i \sum_j \lambda_i \lambda_j \gamma(s_i - s_j) \leq 0$ for all $s_i \in S$ and all $\lambda_1, \lambda_2, \ldots$ such that $\sum_i \lambda_i = 0$.

2. $\gamma(0) = 0$.

3. $\gamma(-h) = \gamma(h)$.

4. $\lim\{\gamma(h)/\|h\|^2\} = 0$ as $\|h\| \to \infty$.

Valid variogram models possess a number of attributes that are related to the spatial dependence structure of the process and that can therefore have a bearing on design. A brief description of a few of these follows; more complete descriptions can be found in Journel and Huijbregts (1978, pp. 36-40, 163-183).

Value near origin. Although $\gamma(0) = 0$, it is not necessary that $\gamma(h) \to 0$ as $\|h\| \downarrow 0$. In practice there is usually observed a discontinuity at the origin, known as the nugget effect, which classically [Matheron (1963)] has been ascribed to micro-scale variation.

Value at infinity. The fourth property of a valid variogram listed above says that $\gamma(h) = o(\|h\|^2)$ as $\|h\| \to \infty$, so it is not necessary that $\gamma(h)$ be bounded. Often, in practice, however, $\gamma(h)$ is bounded and $\lim \gamma(h)$ as $\|h\| \to \infty$ exists; in such a case this limit is called the sill. When a sill exists, $\varepsilon(\cdot)$ is second-order stationary and, provided that the spatial correlation vanishes as $\|h\| \to \infty$, the sill is equal to the variance of the process. Furthermore, if the sill is attained by $\gamma(h)$ at a finite distance a, then two values of $\delta(\cdot)$ located a distance greater than a apart are

uncorrelated and $\gamma(h)$ is said to have range a. It is theoretically possible for $\gamma(h)$ to be bounded but lack a sill; for instance, $\gamma(h)$ can be periodic, but this is rarely seen in practice. In fact, typically $\gamma(h)$ is a nondecreasing function of $\|h\|$.
Dependence on direction. If $\gamma(h)$ depends only on $\|h\|$, then $\gamma(h)$ [and $\varepsilon(\cdot)$] is said to be isotropic and we write it as $\gamma(r)$, where $r = \|h\|$; otherwise it is said to be anisotropic. d-dimensional To facilitate the exposition we will concentrate on the isotropic case in this paper. Generalizations to the anisotropic case are possible, however.

Four nontrivial variogram models in common use today are the exponential, Gaussian, spherical, and power models. Isotropic versions of these models, with nugget effects, are as follows.

- Exponential: $\gamma_E(r;\theta) = \theta_1 + \theta_2\{1 - \exp(-\theta_3 r)\}$, for $r > 0$
- Gaussian: $\gamma_G(r;\theta) = \theta_1 + \theta_2\{1 - \exp(-\theta_3 r^2)\}$, for $r > 0$
- Spherical: $\gamma_S(r;\theta) = \begin{cases} \theta_1 + \theta_2(\frac{3r}{2\theta_3} - \frac{r^3}{2\theta_3^3}), & \text{for } 0 < r \leq \theta_3 \\ \theta_1 + \theta_2, & \text{for } r > \theta_3 \end{cases}$
- Power: $\gamma_P(r;\theta) = \theta_1 + \theta_2 r^{\theta_3}$, for $r > 0$

The parameter space Θ for the exponential, Gaussian, and spherical models is $\Theta = \{\theta : \theta_1 \geq 0, \theta_2 \geq 0, \theta_1 + \theta_2 \neq 0, \theta_3 > 0\}$, whereas that for the power model is $\Theta = \{\theta : \theta_1 \geq 0, \theta_2 \geq 0, \theta + \theta_2 \neq 0, 0 < \theta_3 < 2\}$. Observe that all models but $\gamma_P(r;\theta)$ possess sills which are equal to $\theta_1 + \theta_2$ and that $\gamma_S(r;\theta)$ has range θ_3. All four models above, like nearly all commonly used variogram models, are nonlinear in their parameters. The sole exception is the so-called linear model, which is the special case of $\gamma_P(r;\theta)$ obtained when $\theta_3 = 1$. Other valid variogram models exist besides the ones mentioned here, such as the logarithmic model and the Matérn class of models (of which the exponential is a special case), but they are used less frequently. Somewhat more frequently, linear combinations of the four models mentioned here are used; in geostatistical jargon such models are called "nested structures." For further descriptions of these models, see Matérn (1960), Journel and Huijbregts (1978), and Isaaks and Srivastava (1989).

2.2 Variogram Estimation by Generalized Least Squares

Suppose that data $Z(s_1), \ldots, Z(s_n)$ are observed at sites $s_1, \ldots, s_n \in S \subset \mathbb{R}^2$. Suppose further that, by exploratory data analysis or any other means available, a model of the form (1) is posited for $Z(\cdot)$. For convenience of exposition, assume that the process has constant mean, i.e., that $\mu(s) \equiv \beta$ in model (1). The constant-mean assumption implies that $Z(\cdot)$, a partial realization of which is observable, and $\varepsilon(\cdot)$, which is completely unobservable, have identical variograms. For $r > 0$, define $S(r) = \{(i,j) : i,j = 1,\ldots, n \text{ and } \|s_i - s_j\| = r\}$; in words, $S(r)$ is the set of all pairs of points (s_i, s_j), henceforth called couples, located r units of distance apart

and hence (under the isotropy assumption) $\gamma(s_i - s_j)$ is constant on this set. Also let $N(r)$ denote the number of couples belonging to $S(r)$ and let $L = \{r : r > 0 \text{ and } N(r) > 0\}$. The elements of L are called lags, and the space $\mathcal{L}$ spanned by L the lag-space.

It is natural to estimate the variogram by

$$\hat{\gamma}(r) = \frac{1}{N_{H_r}} \sum_{s_i, s_j \in H_r} [Z(s_i) - Z(s_j)]^2, \tag{3}$$

where for the classical variogram estimator of Matheron (1963), we take $H_r = S(r)$ and $N_{H_r} = N(r)$. If the data are irregularly spaced, H_r is usually enlarged to contain a certain neighbourhood of r, but always such that $\sum_L N_{H_r} = \binom{n}{2} \equiv N$. ¿From a visual inspection of $\hat{\gamma}(r)$ for various lags, a valid parametric variogram model, $\gamma(r, \theta)$, is selected.

If the number of observation sites is not very large we can restrict our attention to a particular case of (3), namely to choose each H_r to contain only one element $r_{ij} = \|s_i - s_j\|$, i.e. $N_{H_r} \equiv 1$, $r = 1, \ldots, N$. The collection of pairs $\{r_{ij}, \hat{\gamma}(r_{ij})\}$ is then called the variogram cloud. Estimation from the variogram cloud clearly utilizes the most information and is preferable when feasible.

The next step is to estimate θ by fitting the chosen model to the nonparametric estimator. A parametric fit $\gamma(r, \hat{\theta})$ of the variogram can be obtained by least squares regression through the variogram cloud. It is obvious that the "observations" $\hat{\gamma}(r_{ij})$ are generally correlated and thus ordinary least squares yields inefficient estimates. The situation calls for generalized least squares estimation, i.e.

$$\hat{\theta}_{GLS} = \arg\min_{\theta} [\hat{\gamma} - \gamma(\theta)]^T \Sigma^{-1}(\theta) [\hat{\gamma} - \gamma(\theta)], \tag{4}$$

where $\hat{\gamma} = [\hat{\gamma}(r_{11}), \hat{\gamma}(r_{12}), \ldots, \hat{\gamma}(r_{n-1,n})]^T$, $\gamma(\theta) = [\gamma(r_{11}, \theta), \gamma(r_{12}, \theta), \ldots, \gamma(r_{n-1,n}, \theta)]^T$, and $\Sigma(\theta)$ denotes the parameterized form of the covariance matrix of $\hat{\gamma}$. If we now assume that $Z(\cdot)$ is a Gaussian random field, Cressie (1985) showed that

$$E(\hat{\gamma}(r_{ij})) = \gamma(r_{ij})$$

and

$$\mathrm{Cov}(\hat{\gamma}(r_{ij}), \hat{\gamma}(r_{kl})) = [\gamma(\|s_i - s_k\|) + \gamma(\|s_j - s_l\|) - \gamma(\|s_i - s_l\|) - \gamma(\|s_j - s_k\|)]^2. \tag{5}$$

The (direct) minimizer in (4) is generally inconsistent [Müller (1997)] and it is necessary to adopt an iterative algorithm. This algorithm is computationally quite intensive and requires the inversion of an $N \times N$ matrix at each step:

$$\hat{\theta}_{m+1} = \arg\min_{\theta} [\hat{\gamma} - \gamma(\theta)]^T \Sigma^{-1}(\hat{\theta}_m) [\hat{\gamma} - \gamma(\theta)],$$

$$\bar{\theta} = \lim_{m \to \infty} \hat{\theta}_m,$$

where the entries of $\Sigma(\hat{\theta}_m)$ can be estimated from the parametric version of (5) evaluated at $\theta = \hat{\theta}_m$.

In practice the iterations have to be started at an initial guess (say $\hat{\theta}_0 = \hat{\theta}_{OLS} = \arg\min_\theta[\hat{\gamma}-\gamma(\theta)]^T[\hat{\gamma}-\gamma(\theta)]$). The procedure then yields asymptotically efficient and consistent estimates. To make the algorithm computationally more feasible, $\Sigma(\hat{\theta}_m)$ could be approximated by a corresponding diagonal matrix, a suggestion that goes back to Cressie (1985) which we will take up later in the design context.

3 Design for variogram estimation

Let $\xi_n = (s_1, \ldots, s_n)$, where $s_i \in \mathcal{S}$. Now we consider the choice of ξ_n that will result in estimates of θ of relatively high precision. The two key features that distinguish this design problem from similar design problems in nonlinear regression settings are:

- The addition of a new site to an existing design of k sites produces not one but k new lags and hence k additional squared differences from which the variogram is estimated. Any sensible design criterion should take account of all of these new lags.
- The standard assumption of uncorrelated errors is not valid.

3.1 Preliminary design considerations

Starting designs

We assume throughout that data have been (or will be) observed at a starting design of $n_0 > 1$ sites in $\mathcal{S}$. There are several good reasons for using a starting design. First, data from a starting design can be used to carry out the initial stage of a geostatistical analysis, that is, to determine an appropriate mean function $\mu(s)$ in (1) and to check that the error process $\varepsilon(\cdot)$ satisfies the intrinsic stationarity assumption. Second, the data provide initial estimates of $\mu(s)$ and θ; the former is useful for the estimation of $\gamma(h)$ when $\mu(s)$ is not constant and the latter is useful because the optimal choice of the remaining $n - n_0$ sites generally depends on θ. Third, an appropriate choice of starting design will ensure a reasonable degree of overall coverage of the study area. This is important because designs selected with only parameter estimation in mind can leave large unsampled "holes" in $\mathcal{S}$. Fourth, if chosen wisely, the initial data can be used to check for anisotropy. Fifth, a properly chosen starting design ensures the positive definiteness of the covariance matrices involved in the computations.

Of course, in some situations the statistician plays no role in the selection of the starting design: some data have already been collected, perhaps where it was most convenient to do so, and the design problem is that of determining how best to augment the existing design. In other situations (e.g., soil surveys), however, one

may be able to build the design "from scratch." In the latter case, a starting design that satisfactorily accomplishes the goals described above is a regular grid, such as a square or triangular grid. Such a design is well suited for determining an appropriate mean function and for checking whether the intrinsic stationarity assumption is reasonably well satisfied; it achieves good overall coverage of $\mathcal{S}$; and it yields replicated lags in numerous directions for checking for anisotropy. Other possibilities for the starting design include the minimax or maximin distance designs of Johnson et al. (1990) and the "transect" designs of Pettitt and McBratney (1993). The latter of these were developed specifically for the purpose of variogram estimation.

Perhaps more important than the choice among these possible starting designs is the choice of n_0, i.e., the number of sites to be allocated to the starting design. Warrick and Myers (1987) suggested taking $n_0 = 0.5n$, but to us this seems excessive and, because of the relative importance of small lags for variogram estimation, it can cause the remaining $n - n_0$ optimally chosen sites to be concentrated very closely to one another. We suggest $n_0 \leq 0.3n$ instead, a rough justification of which is the following.

Take n to be fixed and suppose a regular grid of n_0 sites (where $n_0 < n$ is to be determined) are to be used in a starting design. Let us classify a lag as "large" or "small" according to whether it is greater than or less than η, where η is a positive number smaller than half the grid spacing of a regular grid of n sites. Then whatever the value of n_0, all $\binom{n_0}{2}$ lags in the starting design are large and there is a dire need for small lags. The most efficient way to produce small lags is to take all of the $n - n_0$ remaining sites to be within η of each other and within η of an arbitrary site in the starting design. This results in $\binom{n-n_0+1}{2}$ small lags, but it also produces an additional $(n_0 - 1)(n - n_0)$ large lags. The value of n_0 for which the number of small lags equals the number of large lags is

$$n_0 = n + \frac{1}{2} - \frac{1}{2}\sqrt{2n^2 - 2n + 1} \doteq 0.3n.$$

There are more small lags than large lags when $n_0 << 0.3n$ and vice versa when $n_0 >> 0.3n$. To allow for the possibility of a larger proportion of small lags in the final design, we recommend a value of n_0 smaller than $0.3n$, say, $n_0 = 0.2n$, provided that an acceptable level of overall coverage can still be achieved.

Discretization

A simplification made in the examples and experiments described in this paper is that of restricting the set from which each site in the design is chosen to a finite subset of $\mathcal{S}$. In some cases this restriction may come about naturally, as only a few sites in $\mathcal{S}$ are suitable for the kind of measurements being made. A classical example of this is meterological station data. In other situations, such as many soil studies, there is in principle no reason why attention must be restricted to a finite subset. In these cases, however, it simplifies matters greatly to maximize the design criterion by a grid search over a discretization of $\mathcal{S}$. Provided that the grid is not too fine, this procedure is computationally feasible; it is also more reliable than more

sophisticated optimization algorithms (e.g, Nelder-Mead simplex) since the criterion function generally is not unimodal.

In situations where replicate measurements at sites are not possible, such as when taking the measurement exhausts the experimental material, the set of available sites will get smaller as new sites are added to the design. To allow for this possibility, $\mathcal{S}_k$ shall denote the set of sites that can be added at the k-th stage of a sequential procedure.

Locally optimal designs

As noted in Section 2.1, variogram models are usually nonlinear in the parameters and thus an additional (though widely known) problem in optimal design theory arises, namely the dependence of the information matrix upon the values of the unknown parameters [see a recent discussion of the problem in Atkinson and Haines (1996)].

In particular, the information matrix that corresponds to $\hat{\theta}_{GLS}$ from (4) is given by

$$M(\xi_n;\theta) = G^T(\xi_n;\theta)\Sigma^{-1}(\theta;\xi_n)G(\xi_n;\theta), \tag{6}$$

where

$$G(\xi_n;\theta) = \frac{\partial\gamma(r_i;\theta)}{\partial\theta_j}|_\theta \quad (i = 1,\ldots,N; \quad j = 1,\ldots,p),$$

and $\Sigma(\theta;\xi_n)$ is a modified notation to reflect the dependence of the covariance matrix of $\hat{\gamma}$ with elements given by (5) upon the respective design ξ_n. It is evident that an attempt to optimize (6) for estimation of the unknown θ leads to a circular problem. If data are available from the starting design (or if any other reliable prior information exists), a reasonable way out is to replace θ by a preliminary estimate $\hat{\theta}_0$ and thereby find a "local" solution. All the expressions in the next section will thus be dependent upon $\hat{\theta}_0$ and we will omit this argument where it leads to no ambiguity.

3.2 Augmenting an existing design by one site

Assuming that data from a design of n sites are available, we might face the problem of deciding where to put one additional site in order to retrieve the maximum amount of information about θ. This problem has been treated by Zimmerman and Homer (1991). They suggested the use of the so-called D-criterion, which is the maximization of the determinant of the information matrix, i.e.

$$s^* = \arg\max_{s\in\mathcal{S}_{n+1}} |M(\xi_{n+1})|, \tag{7}$$

where $\xi_{n+1}^T = [\xi_n^T, s]$. This criterion was in the nonlinear regression context firstly proposed by Box and Lucas (1959). Straightforward optimization of (7) leads to an optimization problem that requires the inversion of the $\binom{n+1}{2} \times \binom{n+1}{2}$ dimensional matrix $\Sigma(\hat{\theta}_0;\xi_{n+1})$ for each candidate point.

However, it follows from well-known matrix inversion and determinant formulae [e.g. Rao (1973)] that we can rewrite

$$\begin{aligned} |M(\xi_{n+1})| &= |M(\xi_n) + \Gamma(s,\xi_n)V(s,\xi_n)\Gamma^T(s,\xi_n)| \\ &= |M(\xi_n)||V(s,\xi_n)||V^{-1}(s,\xi_n) + \Gamma^T(s,\xi_n)M^{-1}(\xi_n)\Gamma(s,\xi_n)|, \end{aligned}$$

where

$$\Gamma(s,\xi_n) = G^T(\xi_n)\Sigma^{-1}(\xi_n)\Sigma(s,\xi_n) - G(s),$$

and

$$V(s,\xi_n) = [\Sigma(s) - \Sigma^T(s,\xi_n)\Sigma^{-1}(\xi_n)\Sigma(s,\xi_n)]^{-1}$$

from the block decomposition

$$\Sigma(\xi_{n+1}) = \begin{pmatrix} \Sigma(\xi_n) & \Sigma(s,\xi_n) \\ \Sigma^T(s,\xi_n) & \Sigma(s) \end{pmatrix}. \tag{8}$$

Therefore an equivalent optimization problem to (7) is

$$s^* = \arg\max_{s\in\mathcal{S}_{n+1}} |V(s,\xi_n)||V^{-1}(s,\xi_n) + \Gamma^T(s,\xi_n)M^{-1}(\xi_n)\Gamma(s,\xi_n)|, \tag{9}$$

which requires the inversion of the $\binom{n}{2} \times \binom{n}{2}$ matrices $\Sigma(\xi_n)$ and $M(\xi_n)$ only once, and for each candidate point just the inversion of the $n \times n$ matrix $V(s,\xi_n)$ is needed.

3.3 Augmenting several sites

In principle the design method of the previous section can be extended to choose the "optimal" locations for some additional $q > 1$ sites. Thus, we may consider the problem of choosing additional sites $s_{n+1}, s_{n+2}, \ldots, s_{n+q}$ so as to maximize

$$|V(\mathbf{s}_{n+1},\xi_n)||V^{-1}(\mathbf{s}_{n+1},\xi_n) + \Gamma^T(\mathbf{s}_{n+1},\xi_n)M^{-1}(\xi_n)\Gamma(\mathbf{s}_{n+1},\xi_n)|, \tag{10}$$

where $\mathbf{s}_{n+1} = (s_{n+1},\ldots,s_{n+q})'$ and $V(\mathbf{s}_{n+1},\xi_n)$ and $\Gamma(\mathbf{s}_{n+1},\xi_n)$ are defined by a block decomposition analogous to (8). However, the computational burden of this problem is prohibitive for typical applications. An alternative, computationally simpler approach is to apply rule (9) sequentially to each of the q points to be added. Thus, the method for augmenting one site is extended to the following:

$$s^*_{i+1} = \arg\max_{s\in\mathcal{S}_{i+1}} |V(s,\xi_i)||V^{-1}(s,\xi_i) + \Gamma^T(s,\xi_i)[G(\xi_i)\Sigma^{-1}(\xi_i)G^T(\xi_i)]^{-1}\Gamma(s,\xi_i)|, \tag{11}$$

where $i = n, n+1, \ldots, n+q$. This sequential optimization approach, when combined with sequential data collection, also permits the continual updating of the GLS estimate of θ, which is important for reasons mentioned previously. Sequential data collection is feasible in some applications [see Graham and McLaughlin (1989a),

(1989b); Loaiciga et al. (1992)]; in many other applications, however, it is not feasible and a one-shot design is required.

The single point correction algorithm (11) is an adaptation of a well-known algorithm of Brimkulov, Krug and Savanov (1980) for regression designs with correlated observations [see also Näther (1985)]. It is simple to extend (11) to a Fedorov exchange-type algorithm, as follows: at every step i add the point s_i^* to the current design $\xi_{(i)n}$ and delete the point

$$s_i^- = \arg\min_{s \in \xi_{(i)n}} \quad |V(s, \xi_{(i,s)n})| \times \\ |V^{-1}(s, \xi_{(i,s)n}) + \Gamma^T(s, \xi_{(i,s)n}) M^{-1}(\xi_{(i,s)n}) \Gamma(s, \xi_{(i,s)n})|,$$

where $\xi_{(i,s)n} = \{\xi_{(i)n} - s\}$.

Unfortunately there is currently no proof (as there is for the classical uncorrelated case - diagonal Σ) that such a procedure converges and though we have empirical evidence that the improvements by it are considerable, eventually we might be trapped at a local optimum.

There have been several other attempts to solve the design problem in the correlated errors setting. The classical one is by Sacks and Ylvisaker (1966) who optimize the rate of convergence to an asymptotic design. Recently, Pázman and Müller (1996) have provided a differentiable extension of information matrices and based a gradient algorithm on it. Fedorov and Flanagan (1996) suggest using an eigenvector expansion of the covariance kernel. All three methods seem unsuitable for the current problem. The first does not provide information on the finite sample properties of its resulting designs, the second requires nonsingular covariance matrices defined on the whole collection of candidate points (which might hardly be the case in geostatistical practice), whereas the third requires knowledge of the eigenvectors of the covariance kernel and depends highly on the order of the employed expansion.

3.4 Alternative methods which ignore correlations

A major disadvantage of the algorithms given in the previous section is that they can be computationally very demanding. Even the simplified version (11) might be a computational burden in some settings. It might therefore be helpful in the design context to take up Cressie's (1985) suggestion to approximate Σ by a diagonal matrix, which speeds up computations considerably [as was done by Zimmerman and Homer (1991)]. Note, however, that the effect of ignoring large correlations (due to small lag distances) might be substantial.

Nevertheless, the restriction to the framework of uncorrelated observations allows us to develop alternatives to (11) that avoid some of its disadvantages. Müller and Zimmerman (1995) have suggested a two-stage strategy:

(a) Find the optimal configuration of $n/2$ distances $\xi_{\mathcal{L}}^*$ in the lag space $\mathcal{L}$ (we assume n is even). These will serve as the lags between sites within couples.

(b) Map this configuration into the original site space $\mathcal{S}$ while simultaneously ensuring that that sites from different couples are sufficiently far apart that the corresponding observations are uncorrelated (or as nearly so as possible).

This procedure has the advantage that (a) is in fact the solution of a classical design problem, thus many algorithms with guaranteed convergence are available; see Pukelsheim (1993).

Finding the solution of (b) was also the goal of Warrick and Myers (1987). They suggest finding an optimal site space design by minimizing

$$(\xi_{\mathcal{L}}^* - \xi_{\mathcal{L}})' W (\xi_{\mathcal{L}}^* - \xi_{\mathcal{L}}),$$

where $\xi_{\mathcal{L}}^*$ denotes a target distribution of lags, $\xi_{\mathcal{L}}$ denotes the distribution of lags resulting from a prospective design ξ, and $W = \text{diag}(w_i)$ is a matrix whose diagonal elements are prespecified weights reflecting the importance of certain lags. This rule is basically what is known as *least squares multidimensional scaling* [see e.g. Cox and Cox (1995)]. In fact we may employ a number of techniques developed in that field for the solution of (b). Warrick and Myers (1987) suggest the use of a uniform target distribution $\xi_{\mathcal{L}}^*$, which is not a particularly good choice as we shall see later. There is still one open question related to (b). Since (a) is based on the assumption of uncorrelated observations it would, strictly, be necessary to sufficiently separate the couples yielding the $\frac{n}{2}$ optimal lags so that the remaining $n(n-2)/2$ lags are distant enough not to produce correlations. That is, we should minimize

$$\sum_{L_{\frac{n}{2}}} w(\xi_{\mathcal{L}}^* - \xi_{\mathcal{L}})^2, \quad \text{s.t. } \xi_{\mathcal{L}j} \notin L_{\frac{n}{2}} > \text{range},$$

where $L_{\frac{n}{2}}$ denotes the set of points corresponding to the $\frac{n}{2}$ optimal lags. This can be accomplished using techniques from Lee (1984) or the algorithm given in Müller and Zimmerman (1995). It would guarantee an optimal (or nearly optimal) allocation. However, it is not clear, *a priori*, if this advantage outweighs the deliberate sacrifice of a very large portion of the observations. In the examples we will compare this approach to an alternative that utilizes the information from all observations. We will refer to these approaches as Müller and Zimmerman '95 and Müller and Zimmerman '97, respectively.

3.5 Example and Experiments

To illustrate the properties of the approaches described in this section, we have considered several examples, two of which are reported here. It so happens that the form of the variogram has very little impact on the optimal design, though the range does have considerable impact. Therefore we concentrated on the spherical variogram, $\gamma_S(r;\theta)$, in our examples and varied only its range, θ_3.

The first example considered is taken directly from Warrick and Myers (1987). They seek an optimal allocation of $n = 30$ points on a $400m \times 400m$ square grid with

smallest sampling element of size $2m \times 2m$. Their aim was a uniform distribution of lags over the lag classes $0 - 20m$, $20 - 40m$, ..., $180 - 200m$. For our calculations we chose the prior estimates $\hat{\theta}_1 = \hat{\theta}_2 = 1$ and $\hat{\theta}_3 = 150$, the magnitude of variation of which seems in accordance with the situation in Warrick and Myers (1987).

The resulting designs for the methods presented in this paper are given in Figure 1. Their relative performance can be assessed by looking at Table 1, which gives the values of the design criterion (the determinant of Σ) and the respective calculation times on a Pentium 200Mhz Pc. The value for *eqobs* in the tables denotes the number of "optimal" (according to Brimkulov et al. exchange type algorithm) observations that are sufficient to yield the same amount of information as the respective design.

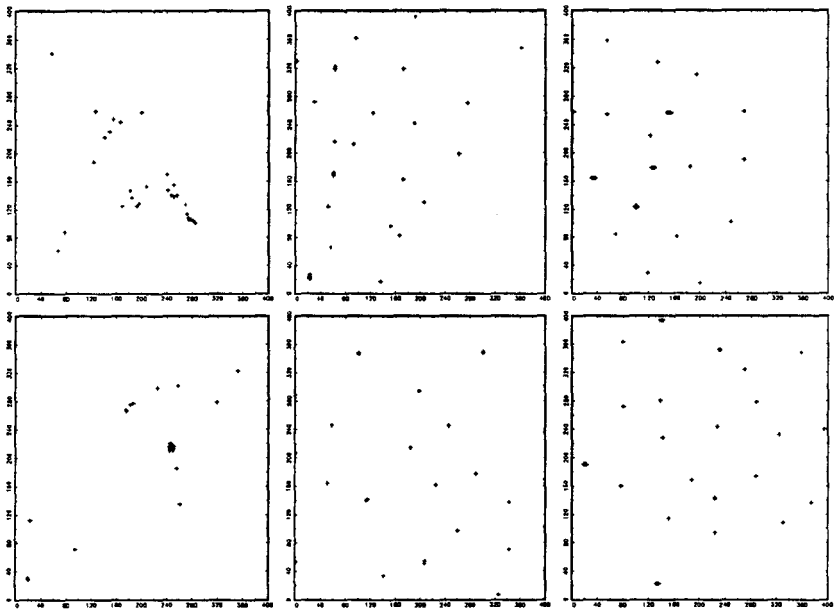


Figure 1: Resulting designs for various methods: Warrick & Myers, Brimkulov et al., Brimkulov et al. (exchange), Müller & Zimmerman '97, Müller & Zimmerman '95, Zimmerman & Homer (from upper left to lower right, by rows).

It can be seen that the best design is achieved by the computationally most demanding algorithm, the exchange version of the Brimkulov et al. type. In fact, all methods (except for Warrick and Myers) give reasonable preformance. However, it seems that the two-stage methods (Müller and Zimmerman '95 and '97) have the speed advantage, which will be of even greater importance for larger samples. It must be noted that there is effectively no difference between the '95 version and the '97 one, which is remarkable considering the amount of information that was neglected by the former. The '97 version also yields a large cluster of about half the observations (producing many small lags), whereas the other designs spread out much more over the design region (with the '95 spreading out the most).

Note that even the best from the simple random sampling designs is much worse

than all other methods that were compared. Specifically it is about three times less efficient than an optimal (Brimkulov et al.) design. Designs from a regular rectangular grid (6×5) or space filling designs are unsuitable for the purpose of variogram estimation, since they fail to produce small lags, which is reflected by an *eqobs*= 4 in both cases.

Design method	criterion values	calculation time (min)	*eqobs*
random sampling	101.72	0	10
Warrick & Myers	16.10	17	18
Müller & Zimmerman '95	4.77	8	27
Müller & Zimmerman '97	4.87	14	27
Brimkulov et al.	2.41	270	29
Brimkulov et al. (exchange)	1.58	4380	30
Zimmerman & Homer	3.09	65	28

Table 1: Best performance out of 5 runs, criterion values $\times 10^6$.

The second example was considered to assess the robustness of the various methods with respect to misspecifying the parameters in the prior guess that leads to the local solution. We again seek an optimal allocation of now $n = 12$ points on a $400m \times 400m$ square grid with smallest sampling element of size 2×2. The assumed correct $\theta_1 = \theta_2 = 1$ and $\theta_3 = 150$ was selected and we varied $\hat{\theta}_3 = 100, 150, 200$. Variations in the other parameters had only neglible effects on the designs.

Here, again the Brimkulov et al. exchange type algorithm gives the best result and for the moderate sample size the calculation times are acceptable. It turns out that the '95 version of the two-stage algorithm is very sensitive to specifying a too low a range, whereas the robustness of the other methods is quite remarkable.

Design method	criterion values for $\hat{\theta} =$ (1,1,100)	(1,1,150)	(1,1,200)	calculation time (min)	*eqobs*
random sampling	534	733	1150	0	6
Warrick & Myers	852	798	919	6	6
Müller & Zimmerman '95	1530	588	117	4	7
Müller & Zimmerman '97	166	123	157	6	10
Brimkulov et al.	248	103	141	3	10
Brimkulov et al. (exchange)	123	42	89	45	12
Zimmerman & Homer	383	136	164	2	10

Table 2: Best performance out of 15 runs, criterion values $\times 10^6$.

4 Discussion

We have adapted several design criteria from the nonlinear regression context for the peculiarities of the variogram estimation context. Moreover, we have described the strengths and weaknesses of these approaches and compared them to simple random sampling, regular and space-filling grids, and previously proposed design criteria for variogram estimation. Finally, we have demonstrated the superiority of our design criteria over these others on two examples.

Our final comments pertain to the possibility of broadening the focus of geostatistical design criteria in two ways. First, our criteria were developed under the assumption that the variogram belongs to a known parametric family (e.g. spherical). In practice we must use the data first to make the choice of a parametric family and then to estimate parameters. Conventional practice for the model selection component is to choose a model that is compatible with a plot of Matheron's classical variogram estimator given by (5). Design principles for this component could possibly be developed by adapting existing model selection design criteria for the context of regression with uncorrelated errors, such as those proposed by Hunter and Reiner (1965) and Atkinson and Fedorov (1975a, 1975b).

Second, a focus on design criteria related only to variogram estimation, with no account taken of the efficacy of the design for prediction (kriging), is inappropriate when prediction is the ultimate objective of the analysis. Suitable design criteria for this case should combine the objectives of variogram estimation and prediction. The development of such criteria requires further investigation.

References

Atkinson, A.C., and Fedorov, V.V. (1975), "The Design of Experiments for Discriminating Between Two Rival Models", *Biometrika*, 62, 57-70.

Atkinson, A.C., and Fedorov, V.V. (1975), "Optimal Design: Experiments for Discriminating Between Several Models", *Biometrika*, 62, 289-303.

Atkinson, A.C., and Haines L.M. (1996), "Designs for Nonlinear and Generalized Models", in *Handbook of Statistics Vol. 13*, S.Gosh and C.R. Rao, eds., Elsevier.

Barnes, R. J. (1989), "Sample Design for Geologic Site Characterization", in *Geostatistics*, M. Armstrong, ed., 809-822, Dordrecht: Kluwer.

Benedetti, R., and Palma, D. (1995), "Optimal Sampling Designs for Dependent Spatial Units", *Environmetrics*, 6, 101-114.

Box, G. E. P., and Lucas, H. L. (1959), "Design of Experiments in Nonlinear Situations", *Biometrika*, 46, 77-90.

Bras, R. L., and Rodriguez-Iturbe, I. (1976), "Network Design for the Estimation of Areal Mean of Rainfall Events", *Water Resources Research* , 12, 1185-1195.

Brimkulov, U.N., Krug, G.K., and Savanov, V.L. (1980), "Cislennoje postrojenije tocnych planov eksperimenta pri korrelirovannych izmerenijach", *Zavodskaja Laboratorija (Industrial Laboratory)*, 46, 435-442.

Cox, T.F., and Cox, M.A.A. (1994), *Multidimensional Scaling*, Chapman & Hall, London.

Cressie, N. (1985), "Fitting Variogram Models by Weighted Least Squares", *Mathematical Geology*, 17, 563-586.

Cressie, N. (1993), *Statistics for Spatial Data* (revised ed.), New York: John Wiley.

Cressie, N., Gotway, C. A., and Grondona, M. O. (1990), "Spatial Prediction from Networks", *Chemometrics and Intelligent Laboratory Systems*, 7, 251-271.

Cressie, N., and Zimmerman, D. L. (1992), "On the Stability of the Geostatistical Method", *Mathematical Geology*, 24, 45-59.

Fedorov, V. V. (1972), *Theory of Optimal Experiments*, Academic Press, New York.

Fedorov, V. V., and Flanagan, D. (1997), "Optimal Monitoring Network Design based on Mercer's Expansion of Covariance Kernel", ...

Graham, W., and McLaughlin, D. (1989a), "Stochastic Analysis of Nonstationary Subsurface Solute Transport, 1. Unconditional Moments." *Water Resources Research*, 25(2), 215-232.

Graham, W., and McLaughlin, D. (1989b), "Stochastic Analysis of Nonstationary Subsurface Solute Transport, 2. Conditional Moments." *Water Resources Research*, 25(11), 2331-2356.

Hoeksema, R. J., and Kitanidis, P. K. (1985), "Analysis of the Spatial Structure of Properties of Selected Aquifers", *Water Resources Research*, 21, 563-572.

Hunter, W.G., and Reiner, A.M. (1965), "Designs for Discrimination Between Two Rival Models", *Technometrics*, 7, 307-323.

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

Journel, A. G., and Huijbregts, C. J. (1978), *Mining Geostatistics*, London: Academic Press.

Johnson, M. E., Moore, L.M., and Ylvisaker, D. (1990), "Minimax and maximin distance designs", *Journal of Statistical Planning and Inference*, 26, 131-148.

Lee, S. (1984), "Multidimensional Scaling Models with Inequality and Equality Constraints", *Commun. Statist.-Simula. Computa.*, 13(1), 127-140.

Loaiciga, H. A., Charbeneau, R. J., Everett, L. G., Fogg, G. E., Hobbs, B. F., and Rouhani, S. (1992), "Review of Ground-Water Quality Monitoring Network Design", *Journal of Hydraulic Engineering*, 118(1), 11-37.

Matérn, B. (1960), "Spatial Variation", *Meddelanden fran Statens Skogforskninginstitut*, 49, 1-144.

Matheron, G. (1963), "Principles of Geostatistics", *Economic Geology*, 58, 1246-1266.

McBratney, A. B., Webster, R., and Burgess, T. M. (1981), "The Design of Optimal Sampling Schemes for Local Estimation and Mapping of Regionalized Variables-I", *Computers and Geosciences*, 7, 331-334.

Müller, W.G. (1997), "On least squares variogram fitting", submitted.

Müller, W.G., and Zimmerman, D.L. (1995), "An Algorithm for Sampling Optimization for Semivariogram Estimation", in MODA4 - Advances in Model-Oriented Data Analysis, C.P. Kitsos, and W.G. Müller, eds., Physica, Heidelberg.

Näther, W. (1985), *Effective Observations of Random Fields*, Teubner Texte zur Mathematik, 72, Teubner Verlag, Leipzig.

Pázman, A. (1986), *Foundations of Optimum Experimental Design*, Reidel, Dordrecht.

Pázman, A., and Müller, W. G. (1996), "Design Measures and Extended Information Matrices for Optimal Designs when the Observations are Correlated", TR No. 48, Dept. of Statistics, University of Economics Vienna.

Pettitt, A.N., and McBratney, A.B. (1993), "Sampling Designs for Estimating Spatial Variance Components", *Applied Statistics*, 42(1), 185-209.

Pukelsheim, F. (1993), *Optimal Design of Experiments*, Wiley, New York.

Ripley, B. D. (1981), *Spatial Statistics*, New York: John Wiley.

Rao, C.R. (1973), *Linear Statistical Inference and its Applications*, 2nd Edition, Wiley, New York.

Robinson, G. K. (1990), "A Role for Variograms", *Australian Journal of Statistics*, 32, 327-335.

Russo, D. (1984), "Design of an Optimal Sampling Network for Estimating the Variogram", *Soil Science Society of America Journal*, 48, 708-716.

Sacks, J., Welch, W.J., Mitchell, T.J., and H.P. Wynn (1989), "Design and Analysis for Computer Experiments", *Statistical Science*, 4, 409-423.

Sacks, J., and Ylvisaker, D. (1966), "Designs for regression problems with correlated errors", *Ann. Math. Statist.*, 37, 66-89.

Stein, M. L. (1988), "Asymptotically Efficient Prediction of a Random Field with a Misspecified Covariance Function", *Annals of Statistics*, 16, 55-63.

Warrick, A. W., and Myers, D. E. (1987), "Optimization of Sampling Locations for Variogram Calculations", *Water Resources Research*, 23, 496-500.

Yfantis, E. A., Flatman, G. T., and Behar, J. V. (1987), "Efficiency of Kriging Estimation for Square, Triangular, and Hexagonal Grids", *Mathematical Geology*, 19, 183-205.

Zimmerman, D.L., and Homer, K.E. (1991), "A Network Design Criterion for Estimating Selected Attributes of the Semivariogram", *Environmetrics*, 2(4), 425-441.