Mathematical Geology, Vol. 30, No. 4, 1998

Variogram Fitting by Generalized Least Squares Using an Explicit Formula for the Covariance Structure[1]

Marc G. Genton[2]

In the context of spatial statistics, the classical variogram estimator proposed by Matheron can be written as a quadratic form of the observations. If data are Gaussian with constant mean, then the correlation between the classical variogram estimator at two different lags is a function of the spatial design matrix and the variance matrix. When data are independent with unidimensional and regular support, an explicit formula for this correlation is available. The same is true for a multidimensional and regular support as can be shown by using Kronecker products of matrices. As variogram fitting is a crucial stage for correct spatial prediction, it is proposed to use a generalized least squares method with an explicit formula for the covariance structure (GLSE). A good approximation of the covariance structure is achieved by taking account of the explicit formula for the correlation in the independent situation. Simulations are carried out with several types of underlying variograms, as well as with outliers in the data. Results show that this technique (GLSE), combined with a robust estimator of the variogram, improves the fit significantly.

KEY WORDS: spatial statistics, scale estimation, robustness, dependent data, correlation structure, ordinary kriging.

INTRODUCTION

Variogram estimation and variogram fitting are two crucial stages of spatial prediction. Because they determine the kriging weights, they must be carried out carefully, otherwise kriging can produce noninformative maps. Careful fitting implies on one hand the use of a highly robust variogram estimator (see Genton, 1996, 1998). On the other hand, variogram estimates at different spatial lags are correlated, for the same observation is used for different lags. As a consequence, variogram fitting by ordinary least squares is not satisfactory. This problem is addressed in this paper by focusing on a generalized least squares

[1]Received 6 February 1997; revised 26 June 1997.
[2]Department of Mathematics, Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland. e-mail: genton@dma.epfl.ch

0882-8121/98/0500-0323$15.00/1

method which takes account of the variance–covariance structure of the variogram estimates.

Consider a spatial stochastic process

$$\{Z(\mathbf{x}) : \mathbf{x} \in D\} \tag{1}$$

where $D \subset \mathbb{R}^d$, $d \geq 1$, is a fixed region. Assume that this process is ergodic and satisfies the hypothesis of intrinsic stationarity:

(a) $E(Z(\mathbf{x})) = \mu = \text{constant}, \quad \forall \mathbf{x} \in D$

(b) $\text{Var}(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})) = 2\gamma(\mathbf{h}), \quad \forall \mathbf{x}, \mathbf{x} + \mathbf{h} \in D$

where $2\gamma(\mathbf{h})$ is the variogram. Suppose that $2\hat{\gamma}(\mathbf{h})$ is a variogram estimator for a given lag $\mathbf{h}$, based on a sample $\{Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n)\}$ of the spatial process, and let $\mathbf{h}_1, \ldots, \mathbf{h}_k$ be the vector lags defined by $\mathbf{h}_i = i\mathbf{h}/\|\mathbf{h}\|$, $i = 1, \ldots, k$, where $1 \leq k \leq K$ and K is the maximal possible distance between data in the direction $\mathbf{h}$. Denote further by $2\hat{\boldsymbol{\gamma}} = (2\hat{\gamma}(\mathbf{h}_1), \ldots, 2\hat{\gamma}(\mathbf{h}_k))^T \in \mathbb{R}^k$ the random vector with variance–covariance matrix $\text{Var}(2\hat{\boldsymbol{\gamma}}) = \tau^2\Omega$, where τ^2 is a real positive constant. The method of generalized least squares consists to determine the estimator $\hat{\boldsymbol{\theta}}$ which minimizes

$$G(\boldsymbol{\theta}) = (2\hat{\boldsymbol{\gamma}} - 2\boldsymbol{\gamma}(\boldsymbol{\theta}))^T\Omega^{-1}(2\hat{\boldsymbol{\gamma}} - 2\boldsymbol{\gamma}(\boldsymbol{\theta})) \tag{2}$$

where $2\boldsymbol{\gamma}(\boldsymbol{\theta}) = (2\gamma(\mathbf{h}_1, \boldsymbol{\theta}), \ldots, 2\gamma(\mathbf{h}_k, \boldsymbol{\theta}))^T \in \mathbb{R}^k$ is the vector of a valid parametric variogram, and $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. Note that $2\gamma(\mathbf{h}, \boldsymbol{\theta})$ is generally a nonlinear function of the parameter $\boldsymbol{\theta}$. Journel and Huijbregts (1978) suggest using only lag vectors $\mathbf{h}_i$ such that $N_{\mathbf{h}_i} > 30$ and $0 < i \leq K/2$. This empirical rule is often met in practice.

The method of generalized least squares (GLS) for variogram fitting was studied by Cressie (1985) in the case where the variance–covariance matrix Ω is diagonal, which leads to the method of weighted least squares (WLS). This approach was at the time dictated by two requirements. On one hand, it led to a computationally simple procedure. On the other hand, the exact computation of Ω was impossible.

THE CORRELATION STRUCTURE OF VARIOGRAM ESTIMATORS

We now turn to the study of the correlation structure of Matheron's classical variogram estimator, as well as some other—more robust—estimators.

Matheron's Classical Variogram Estimator

The classical variogram estimator proposed by Matheron (1962), based on the method-of-moments, is

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{N_{\mathbf{h}}} \sum_{N(\mathbf{h})} (Z(\mathbf{x}_i) - Z(\mathbf{x}_j))^2, \qquad \mathbf{h} \in \mathbb{R}^d \tag{3}$$

where $N(\mathbf{h}) = \{(\mathbf{x}_i, \mathbf{x}_j): \mathbf{x}_i - \mathbf{x}_j = \mathbf{h}\}$ and $N_\mathbf{h}$ is the cardinality of $N(\mathbf{h})$. The simple form of this estimator allows us to write (3) as a quadratic form. In fact, if $\mathbf{z} = (Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_n))^T$ is the data vector and $A(\mathbf{h})$ is the spatial design matrix of the data at lag $\mathbf{h}$, then

$$2\hat{\gamma}(\mathbf{h}) = \mathbf{z}^T A(\mathbf{h})\mathbf{z} \tag{4}$$

It is then straightforward to compute the first and second moments of the above expression, as shown in the following theorem.

Theorem 1. Let $\mathbf{z}$ be a random vector with expectation $E(\mathbf{z}) = \mu \mathbf{1}_n$ and variance–covariance matrix $\mathrm{Var}(\mathbf{z}) = \Sigma$. Then, the expectation of Matheron's classical variogram estimator is

(a) $E(2\hat{\gamma}(\mathbf{h})) = \mathrm{tr}[A(\mathbf{h})\Sigma]$

Moreover, if $\mathbf{z}$ is Gaussian, then

(b) $\mathrm{Var}(2\hat{\gamma}(\mathbf{h})) = 2\ \mathrm{tr}[A(\mathbf{h})\Sigma A(\mathbf{h})\Sigma]$

(c) $\mathrm{Cov}(2\hat{\gamma}(\mathbf{h}_1), 2\hat{\gamma}(\mathbf{h}_2)) = 2\ \mathrm{tr}[A(\mathbf{h}_1)\Sigma A(\mathbf{h}_2)\Sigma]$

(d) $\mathrm{Corr}(2\hat{\gamma}(\mathbf{h}_1), 2\hat{\gamma}(\mathbf{h}_2)) = \dfrac{\mathrm{tr}[A(\mathbf{h}_1)\Sigma A(\mathbf{h}_2)\Sigma]}{\sqrt{\mathrm{tr}[A(\mathbf{h}_1)\Sigma A(\mathbf{h}_1)\Sigma]\mathrm{tr}[A(\mathbf{h}_2)\Sigma A(\mathbf{h}_2)\Sigma]}}$

where $\mathrm{tr}[\cdot]$ is the trace operator.

Proof. Relations (a) and (b) are well known (Cressie, 1991). The third relation is shown as follows:

$$\begin{aligned}\mathrm{Cov}(2\hat{\gamma}(\mathbf{h}_1), 2\hat{\gamma}(\mathbf{h}_2)) &= \mathrm{Cov}(\mathbf{z}^T A(\mathbf{h}_1)\mathbf{z}, \mathbf{z}^T A(\mathbf{h}_2)\mathbf{z}) \\ &= \tfrac{1}{2}[\mathrm{Var}(\mathbf{z}^T(A(\mathbf{h}_1) + A(\mathbf{h}_2))\mathbf{z}) - \mathrm{Var}(\mathbf{z}^T A(\mathbf{h}_1)\mathbf{z}) \\ &\quad - \mathrm{Var}(\mathbf{z}^T A(\mathbf{h}_2)\mathbf{z})] \\ &= \mathrm{tr}[(A(\mathbf{h}_1) + A(\mathbf{h}_2))\Sigma(A(\mathbf{h}_1) + A(\mathbf{h}_2))\Sigma] \\ &\quad - \mathrm{tr}[A(\mathbf{h}_1)\Sigma A(\mathbf{h}_1)\Sigma] - \mathrm{tr}[A(\mathbf{h}_2)\Sigma A(\mathbf{h}_2)\Sigma] \\ &= 2\ \mathrm{tr}[A(\mathbf{h}_1)\Sigma A(\mathbf{h}_2)\Sigma]\end{aligned}$$

Finally, relation (d) is an automatic by-product of relations (b) and (c). □

From now on, we consider data on a unidimensional support of n points, regularly spaced. In addition, we suppose that data are independent or only slightly correlated, in such a way that the variance–covariance matrix can be written as

$$\Sigma = \sigma^2 I_n \tag{5}$$

where I_n is the identity matrix of size $n \times n$, and σ^2 is a real positive factor. Formula (d) for the correlation reduces to

$$\mathrm{Corr}(2\hat{\gamma}(h_1), 2\hat{\gamma}(h_2)) = \frac{\mathrm{tr}[A(h_1)A(h_2)]}{\sqrt{\mathrm{tr}[A(h_1)A(h_1)]\ \mathrm{tr}[A(h_2)A(h_2)]}}, \tag{6}$$

which depends only on the spatial design matrix $A(h)$.

Definitions (3) and (4) of Matheron's classical variogram estimator give the expression of the spatial design matrix $A(h)$, of size $n \times n$, in the unidimensional case. It can be built by superposing identity matrices I_{n-h}, of size $(n - h) \times (n - h)$, in the following way:

$$A(h) = \frac{1}{n-h} \begin{pmatrix} I_{n-h} & -I_{n-h} \\ -I_{n-h} & I_{n-h} \end{pmatrix} \tag{7}$$

If $h < n/2$, the word "superposition" means that two elements located at the same place are added. There are three possible forms of the matrix $A(h)$, depending on h, and they are:

$$A(h) = \frac{1}{n-h} \left[\begin{array}{ccc|ccc|ccc} 1 & & & -1 & & & & & \\ & \ddots & & & \ddots & & & O & \\ & & 1 & & & -1 & & & \\ \hline -1 & & & 2 & & & -1 & & \\ & \ddots & & & \ddots & & & \ddots & \\ & & -1 & & & 2 & & & -1 \\ \hline & & & -1 & & & 1 & & \\ & O & & & \ddots & & & \ddots & \\ & & & & & -1 & & & 1 \end{array} \right], \quad \text{if } h < \frac{n}{2}$$

$$A(h) = \frac{1}{n-h} \left[\begin{array}{ccc|ccc} 1 & & & -1 & & \\ & \ddots & & & \ddots & \\ & & 1 & & & -1 \\ \hline -1 & & & 1 & & \\ & \ddots & & & \ddots & \\ & & -1 & & & 1 \end{array} \right], \quad \text{if } h = \frac{n}{2}$$

$$A(h) = \frac{1}{n-h} \left[\begin{array}{ccc|c|ccc} 1 & & & & -1 & & \\ & \ddots & & O & & \ddots & \\ & & 1 & & & & -1 \\ \hline & O & & O & & O & \\ \hline -1 & & & & 1 & & \\ & \ddots & & O & & \ddots & \\ & & -1 & & & & 1 \end{array} \right], \quad \text{if } h > \frac{n}{2}$$

When $h > n/2$, the four identity matrices are disjunct, whereas for $h = n/2$, they adjoin. As soon as $h < n/2$, they are superposed and consequently, a part of the diagonal of $A(h)$ takes the value $2(n - h)$. In this situation, the matrix $A(h)$ has a particular form, which is called *tri-ridged*. We present now a result which allows us to compute the trace of the product of tri-ridged matrices.

Lemma 1. Let U and V be two symmetric, tri-ridged matrices, of size $n \times n$, given by

$$U = \begin{pmatrix} a_1 & & b_1 & & \\ & & & \ddots & \\ b_1 & & \ddots & & b_p \\ & \ddots & & & \\ & & b_p & & a_n \end{pmatrix}$$

and

$$V = \begin{pmatrix} c_1 & & d_1 & & \\ & & & \ddots & \\ d_1 & & \ddots & & d_q \\ & \ddots & & & \\ & & d_q & & c_n \end{pmatrix}$$

where $1 \leq p \leq n$, $1 \leq q \leq n$. We note $\mathbf{a} = (a_1, \ldots, a_n)^T$, $\mathbf{b} = (b_1, \ldots, b_p)^T$, $\mathbf{c} = (c_1, \ldots, c_n)^T$ and $\mathbf{d} = (d_1, \ldots, d_q)^T$ the ridge values. Then

$$\mathrm{tr}[UV] = \begin{cases} \mathbf{a} \cdot \mathbf{c} & \text{if } p \neq q \\ \mathbf{a} \cdot \mathbf{c} + 2\mathbf{b} \cdot \mathbf{d} & \text{if } p = q \end{cases}$$

where $\cdot$ is the usual scalar product.

Proof. By direct computation, we have

$$\mathrm{tr}[UV] = \sum_{i=1}^{n} (UV)_{ii} = \sum_{i=1}^{n} \sum_{k=1}^{n} u_{ik}v_{ki} = \sum_{i=1}^{n} \sum_{k=1}^{n} u_{ik}v_{ik}$$

$$= \begin{cases} \mathbf{a} \cdot \mathbf{c} & \text{if } p \neq q \\ \mathbf{a} \cdot \mathbf{c} + 2\mathbf{b} \cdot \mathbf{d} & \text{if } p = q \end{cases} \qquad \square$$

Corollary 1.1. If the ridge values **a**, **b**, **c**, and **d** are constant, that is to say $\mathbf{a} = (a, \ldots, a)^T$, $\mathbf{b} = (b, \ldots, b)^T$, $\mathbf{c} = (c, \ldots, c)^T$, and $\mathbf{d} = (d, \ldots, d)^T$, then

$$\mathrm{tr}[UV] = \begin{cases} n\, ac & \text{if } p \neq q \\ n\, ac + 2p\, bd & \text{if } p = q \end{cases}$$

The matrix $A(h)$ belongs to that class of symmetric, tri-ridged matrices. With the previous lemma and corollary, we have the following:

Theorem 2. Let $2\hat{\gamma}(h) = \mathbf{z}^T A(h)\mathbf{z}$ be Matheron's classical variogram estimator. Let $\mathbf{z} \in \mathbb{R}^n$ be the data vector with unidimensional and regular support of n points in $\mathbb{R}^1$. Suppose that $E(\mathbf{z}) = \mu \mathbf{1}_n$ and $\text{Var}(\mathbf{z}) = \sigma^2 I_n$. Then, the expectation of Matheron's classical variogram estimator is

(a) $E(2\hat{\gamma}(h)) = 2\sigma^2$

Moreover, if $\mathbf{z}$ is Gaussian, then

$$\text{(b) } \text{Var}(2\hat{\gamma}(h)) = \begin{cases} \dfrac{12\sigma^4}{(n-h)} - \dfrac{4\sigma^4 h}{(n-h)^2} & \text{if } h < \dfrac{n}{2} \\ \dfrac{8\sigma^4}{n-h} & \text{otherwise,} \end{cases}$$

and for $h_1 < h_2$

$$\text{(c) } \text{Cov}(2\hat{\gamma}(h_1), 2\hat{\gamma}(h_2)) = \begin{cases} \dfrac{4\sigma^4(2n - h_1 - 2h_2)}{(n-h_1)(n-h_2)} & \text{if } h_1 + h_2 < n \\ \dfrac{4\sigma^4}{n-h_1} & \text{otherwise,} \end{cases}$$

(d) $\text{Corr}(2\hat{\gamma}(h_1), 2\hat{\gamma}(h_2)) =$

$$\begin{cases} \dfrac{2n - h_1 - 2h_2}{\sqrt{(3n-4h_1)(3n-4h_2)}} & \text{if } h_2 < \dfrac{n}{2} \\ \dfrac{2n - h_1 - 2h_2}{\sqrt{2(3n-4h_1)(n-h_2)}} & \text{if } h_2 \geq \dfrac{n}{2} \text{ and } h_1 + h_2 < n \\ \dfrac{n - h_2}{\sqrt{2(3n-4h_1)(n-h_2)}} & \text{if } h_1 < \dfrac{n}{2} \text{ and } h_1 + h_2 \geq n \\ \dfrac{1}{2}\sqrt{\dfrac{n-h_2}{n-h_1}} & \text{if } h_1 \geq \dfrac{n}{2} \end{cases}$$

These explicit formulas will be useful for computing the variance-covariance matrix Ω of the generalized least squares method discussed in the next section.

Other Variogram Estimators

Consider two other variogram estimators, one proposed by Cressie and Hawkins (1980), and Q_{N_h}, an explicit and highly robust estimator, proposed by Rousseeuw and Croux (1992, 1993) in the context of scale estimation. Genton

(1995, 1996, 1998) discusses the use of this latter for variogram estimation. The form of these estimators does not allow explicit computing of their correlation structure, even in the independent case. Nevertheless, one can estimate these structures by simulation. Note that the correlation structure does not change with the variance σ^2 of the data. 10000 unidimensional samples of size $n = 100$ were simulated. Each observation was independently and identically distributed, according to a standard Gaussian distribution. The correlation structure of each estimator was estimated with the samples. Results are shown in Figure 1, in the form of contour plots of the correlation structure. The first plot is obtained with the explicit formula given in Theorem 2, whereas the others are estimated by simulations. The structure for Matheron's classical variogram estimator is in close agreement with the theory. The remaining two plots show that the general shape of the correlation structure for Cressie and Hawkins'

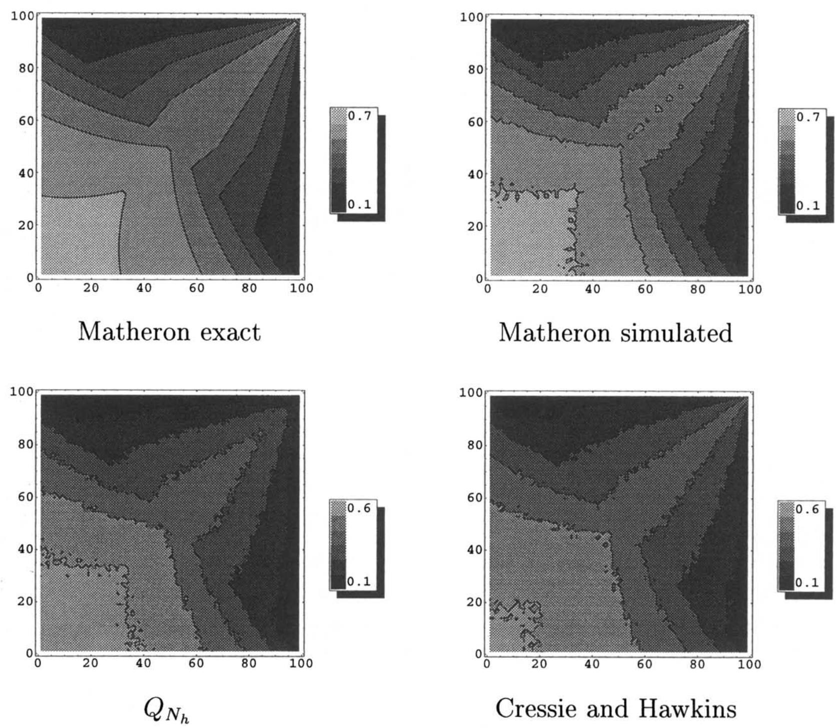


Figure 1. These plots show dependence of Corr($2\hat{\gamma}(h_1)$, $2\hat{\gamma}(h_2)$) on lags h_1 (horizontal axis) and h_2 (vertical axis), for independently and identically distributed observations from a standard Gaussian distribution. This changes with the estimator $2\hat{\gamma}(\cdot)$ used. For Matheron's classical variogram estimator, explicit computation is possible (Theorem 2). In other situations, one has to perform simulations (I used 10000 replications of size $n = 100$). Plots show contour lines for correlations between 0 and 0.1, between 0.1 and 0.2, etc.

estimator and for Q_{N_h} are quite similar, although the maximal correlation becomes smaller. In fact, Q_{N_h} is the estimator whose correlation structure is closest to Matheron's.

When data are dependent, the correlation between the observations themselves modifies the correlation of the variogram estimators. Simulations with data having various degrees of dependency show that differences can happen. Nonetheless, in practical situations, the correlation structure obtained from independent data is remarkably close to the true correlation structure.

THE COVARIANCE STRUCTURE OF GENERALIZED LEAST SQUARES ESTIMATES FOR DATA IN $\mathbb{R}^1$

Cressie (1985) shows that the approximation

$$\mathrm{Var}(2\hat{\gamma}(h)) \cong 2\,\frac{(2\gamma(h,\,\theta))^2}{N_h}, \qquad h = 1, \ldots, k \tag{8}$$

is satisfactory for Matheron's classical variogram estimator, and

$$\mathrm{Var}(2\hat{\gamma}(h)) \cong 2.885\,\frac{(2\gamma(h,\,\theta))^2}{N_h}, \qquad h = 1, \ldots, k \tag{9}$$

is satisfactory for Cressie and Hawkins' estimator. I present in the following lemma a generalization of these two results to M-estimators of scale.

Variogram estimation can be tackled via scale estimation based on the process of differences $V(h) = Z(x + h) - Z(x)$, which has zero expectation and a variance of $2\gamma(h)$ under the hypothesis of intrinsic stationarity of the process Z. M-estimators of scale (Hampel and others, 1986) form a wide class of estimators, which are asymptotically defined as the solution S of the implicit equation

$$\int \chi(x/S(F_V))dF_V(x) = 0 \tag{10}$$

where χ is a real, symmetric, and sufficiently regular function, and F_V is the marginal distribution of the possibly dependent process V. The special choice $\chi(x) = |x|^q - \int |x|^q\, dF_V(x)$, $q > 0$, leads to the so-called L^q estimators of scale (Genton and Rousseeuw, 1995), where L^2 corresponds to Matheron's classical variogram estimator, whereas $L^{1/2}$ corresponds to Cressie and Hawkins' estimator (Genton, 1998).

Lemma 2. For every M-estimator of scale $S(F_V)$ applied to a process of differences $V(h)$, dependent, with expectation zero and variance $2\gamma(h)$, the asymptotic variance under dependence of $S^2(F_V)$ is equal to

$$V^*(S^2, F_V) = K_h^*\gamma(h)^2$$

where K_h^* is a real positive constant, which depends on S and on h through F_V.

Proof. From the implicit Equation (10), the influence function of S^2 is

$$IF(u, S^2, F_V) = 2\frac{\chi(u/S(F_V))}{\int x/S(F_V)\chi'(x/S(F_V))dF_V(x)} S^2(F_V)$$

and the asymptotic variance under dependence (Portnoy, 1977) equals

$$\begin{aligned} V^*(S^2, F_V) &= 4\frac{A(\chi, F_V) + 2\sum_{k=1}^{\infty} A^*(\chi, F_V^{(k)})}{B^2(\chi, F_V)} S^4(F_V) \\ &= 16\frac{A(\chi, F_V) + 2\sum_{k=1}^{\infty} A^*(\chi, F_V^{(k)})}{B^2(\chi, F_V)} \gamma(h)^2 \\ &= K_h^*\gamma(h)^2 \end{aligned}$$

where

$$\begin{aligned} A(\chi, F_V) &= \int \chi^2(x/S(F_V))dF_V(x) \\ B(\chi, F_V) &= \int (x/S(F_V))\chi'(x/S(F_V))dF_V(x) \\ A^*(\chi, F_V^{(k)}) &= \int\int \chi(x_1/S(F_V))\chi(x_2/S(F_V))dF_V^{(k)}(x_1, x_2) \end{aligned}$$

and $F_V^{(k)}$ is the bivariate distribution of the pair $(V_i(h), V_{i+k}(h))$. □

The constant K_h^* depends on the type of estimator, as well as the underlying dependence of the process V. If V is independent and Gaussian, the asymptotic variance of Matheron's estimator is

$$V^*((L^2)^2, F_V) = 8\gamma(h)^2$$

whereas Cressie and Hawkins' estimator satisfies

$$V^*((L^{1/2})^2, F_V) = 11.54\gamma(h)^2$$

These are exactly the approximations (8) and (9) given by Cressie (1985). When the spatial stochastic process Z is independent and Gaussian, the asymptotic variance of Matheron's estimator is

$$V^*((L^2)^2, F_V) = 12\gamma(h)^2$$

which corresponds to the explicit result given in Theorem 2. The distinction between the case $h < n/2$ and $h \geq n/2$ does not appear in Lemma 2, because the result is asymptotic. We are now prepared to compute the form of an element of the matrix $\Omega = \Omega(\theta)$ by writing

$$\begin{aligned} \Omega_{ij} &= \text{Cov}(2\hat{\gamma}(i), 2\hat{\gamma}(j)) \\ &= \text{Corr}(2\hat{\gamma}(i), 2\hat{\gamma}(j)) \sqrt{\text{Var}(2\hat{\gamma}(i))\text{Var}(2\hat{\gamma}(j))} \\ &\cong \text{Corr}(2\hat{\gamma}(i), 2\hat{\gamma}(j))K_i^*\gamma(i, \theta)K_j^*\gamma(j, \theta)/\sqrt{N_iN_j} \end{aligned}$$

where K_i^* and K_j^* are the constants for $h = i$ and $h = j$, respectively. When the spatial stochastic process Z is independent, K_h^* is the same constant for each h. So, the approximation

$$\Omega_{ij} = \text{Corr}(2\hat{\gamma}(i), 2\hat{\gamma}(j))\gamma(i, \boldsymbol{\theta})\gamma(j, \boldsymbol{\theta})/\sqrt{N_i N_j}$$

is used, where the correlation $\text{Corr}(2\hat{\gamma}(i), 2\hat{\gamma}(j))$ can be approximately computed by Matheron's classical estimator under independence and gaussianity given explicitly in part (d) of Theorem 2. The factor τ^2 in the variance of $2\hat{\gamma}$ allows one to include this approximation, as well as the factor K_h^* and the asymptotic treatment of the variance of $2\hat{\gamma}$. The matrix Ω can be written

$$\Omega = R \circ S(\boldsymbol{\theta}) \tag{11}$$

where $R \in \mathbb{R}^{k \times k}$ is the correlation matrix given in Theorem 2, $S(\boldsymbol{\theta}) = \mathbf{s}(\boldsymbol{\theta})\mathbf{s}(\boldsymbol{\theta})^T \in \mathbb{R}^{k \times k}$ with $\mathbf{s}(\boldsymbol{\theta}) = (\gamma(1, \boldsymbol{\theta})/\sqrt{N_1}, \ldots, \gamma(k, \boldsymbol{\theta})/\sqrt{N_k})^T \in \mathbb{R}^k$, and $\circ$ is the Hadamard product of two matrices (Horn and Johnson, 1991). We show in the next lemma that the matrix Ω in (11) is a variance–covariance matrix, that is to say positive definite, which guarantees the existence of the inverse matrix Ω^{-1}.

Lemma 3. The matrix $\Omega = R \circ S(\boldsymbol{\theta})$ is positive definite $\forall \boldsymbol{\theta} \in \Theta$.

Proof. The matrix R is positive definite, because it is a correlation matrix, the elements of which are defined in Theorem 2. Moreover, the matrix $S(\boldsymbol{\theta})$ is positive semi-definite $\forall \boldsymbol{\theta} \in \Theta$, because it possesses only one nonzero eigenvalue, which is $\lambda = \|\mathbf{s}(\boldsymbol{\theta})\|^2 = \Sigma_{i=1}^k \gamma(i, \boldsymbol{\theta})^2/N_i$. In addition, the diagonal elements of $S(\boldsymbol{\theta})$ are strictly positive and equal to $S_{ii}(\boldsymbol{\theta}) = \gamma(i, \boldsymbol{\theta})^2/N_i$, $\forall i = 1, \ldots, k$. Applying Schur's Theorem (Horn and Johnson, 1991), which states that the matrix resulting of the Hadamard product between a positive definite matrix and a positive semidefinite matrix with nonzero diagonal elements, is positive definite, we conclude that $\Omega = R \circ S(\boldsymbol{\theta})$ is positive definite $\forall \boldsymbol{\theta} \in \Theta$. □

The asymptotic distribution of the generalized least squares estimator $\hat{\boldsymbol{\theta}} \in \mathbb{R}^p$ is given by Fedorov (1974) and Seber and Wild (1989). Under some regularity conditions, $\sqrt{k}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically distributed as a Gaussian law $N_p(\mathbf{0}, v^2\Upsilon)$. The parameters Υ and v^2 can be estimated consistently by

$$\hat{\Upsilon} = k[D(\hat{\boldsymbol{\theta}})^T\Omega(\hat{\boldsymbol{\theta}})^{-1}D(\hat{\boldsymbol{\theta}})]^{-1}$$

$$\hat{v}^2 = \frac{1}{k - p} G(\hat{\boldsymbol{\theta}})$$

where $D(\boldsymbol{\theta}) = \partial 2\gamma(\boldsymbol{\theta})/\partial\boldsymbol{\theta} \in \mathbb{R}^{k \times p}$ is the matrix of partial derivatives of $2\gamma(\boldsymbol{\theta})$.

The method of generalized least squares with an explicit formula for the covariance structure (GLSE) that I propose for variogram fitting is summarized in the following algorithm:

(1) Determine the matrix $\Omega = \Omega(\boldsymbol{\theta})$ with element Ω_{ij} given by

$$\text{Corr}(2\hat{\gamma}(i), 2\hat{\gamma}(j))\gamma(i, \boldsymbol{\theta})\gamma(j, \boldsymbol{\theta})/\sqrt{N_i N_j}$$

and $\text{Corr}(2\hat{\gamma}(i), 2\hat{\gamma}(j))$ equal to the formula (d) of Theorem 2.

(2) Choose $\theta^{(0)}$ and let $i = 0$.
(3) Compute the matrix $\Omega(\theta^{(i)})$ and determine $\theta^{(i+1)}$ which minimizes

$$G(\theta) = (2\hat{\gamma} - 2\gamma(\theta))^T \Omega(\theta^{(i)})^{-1}(2\hat{\gamma} - 2\gamma(\theta))$$

(4) Repeat (3) until convergence to obtain $\hat{\theta}$.

In step (2), the choice of $\theta^{(0)}$ can be carried out randomly, or with the result of a fit by ordinary least squares (OLS) or weighted least squares (WLS).

SIMULATIONS

In order to analyze the performances of our method, simulations of spatial data in $\mathbb{R}^1$ were carried out for different kind of typical dependence situations. I have chosen the following models:

1. An exponential variogram

$$\gamma(h, a, b, c) = \begin{cases} 0 & \text{if } h = 0 \\ a + b\left(1 - \exp\left(\frac{-h}{c}\right)\right) & \text{otherwise} \end{cases}$$

with parameters

(i) $a = 1, b = 2, c = 1$
(ii) $a = 1, b = 2, c = 5$
(iii) $a = 1, b = 2, c = 15$

2. A spherical variogram

$$\gamma(h, a, b, c) = \begin{cases} 0 & \text{if } h = 0 \\ a + b\left(\frac{3}{2}\left(\frac{h}{c}\right) - \frac{1}{2}\left(\frac{h}{c}\right)^3\right) & \text{if } 0 < h \leq c \\ a + b & \text{if } h > c \end{cases}$$

with parameters

(i) $a = 1, b = 2, c = 3$
(ii) $a = 1, b = 2, c = 15$
(iii) $a = 1, b = 2, c = 45$

3. A power variogram

$$\gamma(h, a, b, c) = \begin{cases} 0 & \text{if } h = 0 \\ a + bh^c & \text{otherwise} \end{cases}$$

with parameters

(i) $a = 0$, $b = 2$, $c = 0.5$

(ii) $a = 0$, $b = 2$, $c = 1.5$

4. A spherical variogram $\gamma(h, 1, 2, 15)$, with a contamination of $\epsilon\%$ of data by Gaussian data from $N(0, 25)$ with

(i) $\epsilon = 5\%$

(ii) $\epsilon = 10\%$

These variograms represent three typical dependencies. The first two correspond to spatial stochastic processes which are second order stationary, whereas the third one comes from an intrinsically stationary process. The exponential variogram reaches its sill only asymptotically, but one generally assigns it a range of $3c$, which corresponds roughly to the ranges of the spherical variogram given in 2. Finally, a perturbed variogram is used in order to test the robustness properties of the method.

The following lemma shows that the various parameters of these variograms have different practical significance. Parameter c, characterizing shape and sometimes range of the variogram, is the most important one. Parameters a and b have no influence on weights of ordinary kriging, but parameter b influences the kriging variance.

Lemma 4. Weights of ordinary kriging are invariant by linear transformation of the variogram.

Proof. Let $\gamma^*(\mathbf{h}) = \alpha + \beta\gamma(\mathbf{h})$ be the variogram obtained by a linear transformation, where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ are constants. The linear system of ordinary kriging is

$$\begin{pmatrix} \Gamma & \mathbf{1}_n \\ \mathbf{1}_n^T & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \eta \end{pmatrix} = \begin{pmatrix} \gamma \\ 1 \end{pmatrix}$$

For $\gamma^*(\mathbf{h})$, the linear system of ordinary kriging is

$$\begin{pmatrix} \beta\Gamma + \alpha\mathbf{1}_n\mathbf{1}_n^T & \mathbf{1}_n \\ \mathbf{1}_n^T & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \eta \end{pmatrix} = \begin{pmatrix} \beta\gamma + \alpha\mathbf{1}_n \\ 1 \end{pmatrix}$$

which is equivalent to

$$\begin{pmatrix} \Gamma & \mathbf{1}_n \\ \mathbf{1}_n^T & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \eta/\beta \end{pmatrix} = \begin{pmatrix} \gamma \\ 1 \end{pmatrix}$$

Therefore, weights of ordinary kriging with $\gamma^*(\mathbf{h})$ are the same as those with $\gamma(\mathbf{h})$. Kriging variance, however, changes if $\beta \neq 1$. □

I now briefly describe the simulation setup I used. For each of the above ten situations, I generated 100 Gaussian samples of size $n = 200$. As the exponential and spherical variograms represent second order stationary processes, the algorithm of Durbin–Levinson (Brockwell and Davis, 1991) is used. For the power variogram, which represents intrinsically stationary processes, the innovation algorithm (Brockwell and Davis, 1991) is used. Finally, for the perturbed spherical variogram, simply replace ϵ% of data by outliers drawn from a Gaussian distribution $N(0, 25)$. On each sample, the variogram is estimated by Matheron's classical estimator, noted L^2, and the highly robust estimator Q_{N_h}. Next, the fit of the variogram on half of the points is done via weighted least squares (WLS) of Cressie (1985) and generalized least squares (GLSE). The initialization of this last method is done with the result of WLS, as well as with several random choices. No convergence problem has been detected and the minimum is generally reached in three or four iterations.

Results of the simulations are shown in Tables 1, 2, 3, and 4. For each situation, the mean of every parameter, with associated standard deviation, is

Table 1. Results of Simulations for the Exponential Variogram[a]

		L^2 and WLS	Q_{N_h} and WLS	L^2 and GLSE	Q_{N_h} and GLSE
expo(1,2,1)	$\bar{a}$	1.038	0.976	0.959	0.885
	$\hat{\sigma}_{\bar{a}}$	0.087	0.086	0.087	0.087
	$\bar{b}$	2.136	2.104	2.070	2.198
	$\hat{\sigma}_{\bar{b}}$	0.083	0.083	0.085	0.087
	$\bar{c}$	5.286	2.799	2.289	**1.410**
	$\hat{\sigma}_{\bar{c}}$	2.393	0.962	0.784	**0.122**
expo(1,2,5)	$\bar{a}$	1.037	0.998	1.006	1.021
	$\hat{\sigma}_{\bar{a}}$	0.042	0.042	0.034	0.039
	$\bar{b}$	2.157	2.221	2.208	2.296
	$\hat{\sigma}_{\bar{b}}$	0.077	0.078	0.084	0.086
	$\bar{c}$	9.337	8.763	7.593	**6.469**
	$\hat{\sigma}_{\bar{c}}$	1.385	1.634	0.809	**0.598**
expo(1,2,15)	$\bar{a}$	0.965	0.909	0.982	1.007
	$\hat{\sigma}_{\bar{a}}$	0.027	0.019	0.019	0.019
	$\bar{b}$	2.774	2.512	2.613	2.429
	$\hat{\sigma}_{\bar{b}}$	0.280	0.230	0.195	0.149
	$\bar{c}$	28.442	19.313	21.774	**16.188**
	$\hat{\sigma}_{\bar{c}}$	5.102	3.605	2.775	**1.899**

[a] For each situation, variograms are estimated by L^2 or Q_{N_h}, and fitted by WLS or GLSE. The mean over 100 replications of the fitted parameter $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{c})^T$, as well as an associated standard deviation, are computed. The best fit of parameter c, the most important for ordinary kriging, is written in boldface.

Table 2. Results of Simulations for the Spherical Variogram[a]

		L^2 and WLS	Q_{N_h} and WLS	L^2 and GLSE	Q_{N_h} and GLSE
sph(1,2,3)	$\bar{a}$	1.022	0.995	0.990	0.963
	$\hat{\sigma}_{\bar{a}}$	0.051	0.058	0.051	0.058
	$\bar{b}$	1.993	2.073	2.009	2.104
	$\hat{\sigma}_{\bar{b}}$	0.055	0.062	0.056	0.064
	$\bar{c}$	3.847	3.897	3.566	**3.525**
	$\hat{\sigma}_{\bar{c}}$	0.355	0.330	0.216	**0.172**
sph(1,2,15)	$\bar{a}$	1.071	1.061	1.031	1.081
	$\hat{\sigma}_{\bar{a}}$	0.027	0.028	0.023	0.028
	$\bar{b}$	2.117	2.157	2.171	2.279
	$\hat{\sigma}_{\bar{b}}$	0.086	0.084	0.086	0.095
	$\bar{c}$	21.100	20.111	19.685	**17.627**
	$\hat{\sigma}_{\bar{c}}$	1.646	1.531	1.576	**0.994**
sph(1,2,45)	$\bar{a}$	1.004	0.976	1.034	1.082
	$\hat{\sigma}_{\bar{a}}$	0.016	0.014	0.019	0.019
	$\bar{b}$	8.558	2.368	4.793	2.458
	$\hat{\sigma}_{\bar{b}}$	4.684	0.199	1.741	0.215
	$\bar{c}$	294.657	**47.579**	139.287	48.564
	$\hat{\sigma}_{\bar{c}}$	143.604	**4.631**	70.621	6.985

[a]For each situation, variograms are estimated by L^2 or Q_{N_h}, and fitted by WLS or GLSE. The mean over 100 replications of the fitted parameters $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{c})^T$, as well as an associated standard deviation, are computed. The best fit of parameter c, the most important for ordinary kriging, is written in boldface.

Table 3. Results of Simulations for the Power Variogram[a]

		L^2 and WLS	Q_{N_h} and WLS	L^2 and GLSE	Q_{N_h} and GLSE
pow(0,2,0.5)	$\bar{a}$	0.168	0.097	0.132	0.040
	$\hat{\sigma}_{\bar{a}}$	0.035	0.056	0.026	0.014
	$\bar{b}$	1.325	1.772	1.263	1.644
	$\hat{\sigma}_{\bar{b}}$	0.074	0.094	0.058	0.053
	$\bar{c}$	0.451	0.325	**0.491**	0.382
	$\hat{\sigma}_{\bar{c}}$	0.026	0.025	**0.026**	0.014
pow(0,2,1.5)	$\bar{a}$	0.873	0.010	0.067	0.000
	$\hat{\sigma}_{\bar{a}}$	0.742	0.007	0.013	0.000
	$\bar{b}$	2.588	9.291	1.621	3.556
	$\hat{\sigma}_{\bar{b}}$	0.528	1.541	0.102	0.155
	$\bar{c}$	1.290	0.742	**1.357**	0.956
	$\hat{\sigma}_{\bar{c}}$	0.036	0.037	**0.025**	0.019

[a]For each situation, variograms are estimated by L^2 or Q_{N_h}, and fitted by WLS or GLSE. The mean over 100 replications of the fitted parameter $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{c})^T$, as well as an associated standard deviation, are computed. The best fit of parameter c, the most important for ordinary kriging, is written in boldface.

Table 4. Results of Simulations for the Perturbed Spherical Variogram[a]

		L^2 and WLS	Q_{N_h} and WLS	L^2 and GLSE	Q_{N_h} and GLSE
sph(1,2,15)	$\bar{a}$	1.071	1.061	1.031	1.081
$\epsilon = 0\%$	$\hat{\sigma}_{\bar{a}}$	0.027	0.028	0.023	0.028
	$\bar{b}$	2.117	2.157	2.171	2.279
	$\hat{\sigma}_{\bar{b}}$	0.086	0.084	0.086	0.095
	$\bar{c}$	21.100	20.111	19.685	**17.627**
	$\hat{\sigma}_{\bar{c}}$	1.646	1.531	1.576	**0.994**
sph(1,2,15)	$\bar{a}$	2.379	1.364	2.431	1.449
$\epsilon = 5\%$	$\hat{\sigma}_{\bar{a}}$	0.077	0.034	0.079	0.038
	$\bar{b}$	1.923	2.390	2.001	2.524
	$\hat{\sigma}_{\bar{b}}$	0.075	0.094	0.080	0.104
	$\bar{c}$	22.693	20.402	22.618	**19.668**
	$\hat{\sigma}_{\bar{c}}$	2.007	1.901	1.959	**1.486**
sph(1,2,15)	$\bar{a}$	3.543	1.804	3.561	1.872
$\epsilon = 10\%$	$\hat{\sigma}_{\bar{a}}$	0.090	0.044	0.094	0.045
	$\bar{b}$	1.834	2.561	1.871	2.696
	$\hat{\sigma}_{\bar{b}}$	0.087	0.092	0.088	0.102
	$\bar{c}$	27.356	21.519	26.877	**19.841**
	$\hat{\sigma}_{\bar{c}}$	3.446	2.044	2.989	**1.576**

[a] For each situation, variograms are estimated by L^2 or Q_{N_h}, and fitted by WLS or GLSE. The mean over 100 replications of the fitted parameters $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{c})^T$, as well as an associated standard deviation, are computed. The best fit of parameter c, the most important for ordinary kriging, is written in boldface.

computed over the 100 simulations. The best fit of parameter c, the most important one for ordinary kriging, is written in boldface. It is straightforward to see that use of the generalized least squares method (GLSE), in conjunction with a robust estimator of the variogram, like Q_{N_h}, improves the fit significantly. Nevertheless, note in Table 3 that the variogram estimation by L^2 seems best, at least when $c = 1.5$. In fact, when $1 < c < 2$, the spatial stochastic process Z possesses long-range dependencies and is usually called a long memory process (Beran, 1994). In this situation, and more particularly for large lags h and samples of small size, estimators L^2 and Q_{N_h} behave quite differently for L^2 makes explicit use of the assumption that $E(V(h)) = 0$, whereas Q_{N_h} does not. However, processes with a power variogram are not common in practice. Finally, Table 4 shows that even with outliers in the data, our method picks the true underlying variogram.

Figures 2 and 3 present effects of outliers on estimation by L^2 or Q_{N_h} and fitting by WLS or GLSE. Here, the data is simulated with a spherical variogram sph(1, 2, 15), perturbed by $\epsilon = 5\%$ and $\epsilon = 10\%$ observations from $N(0, 25)$.

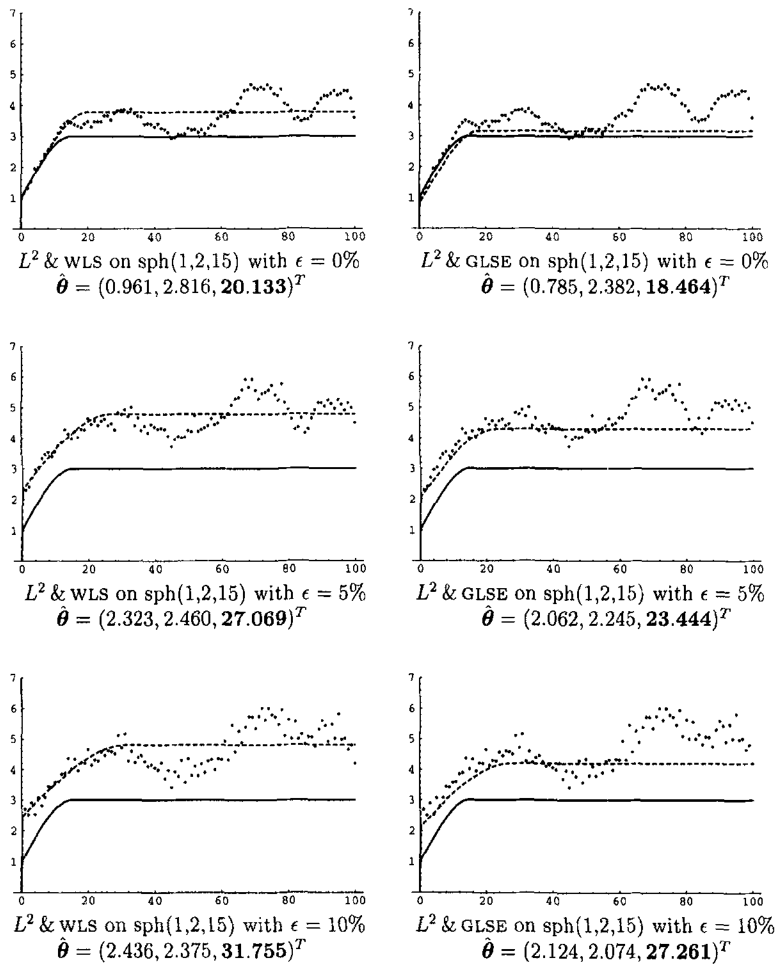
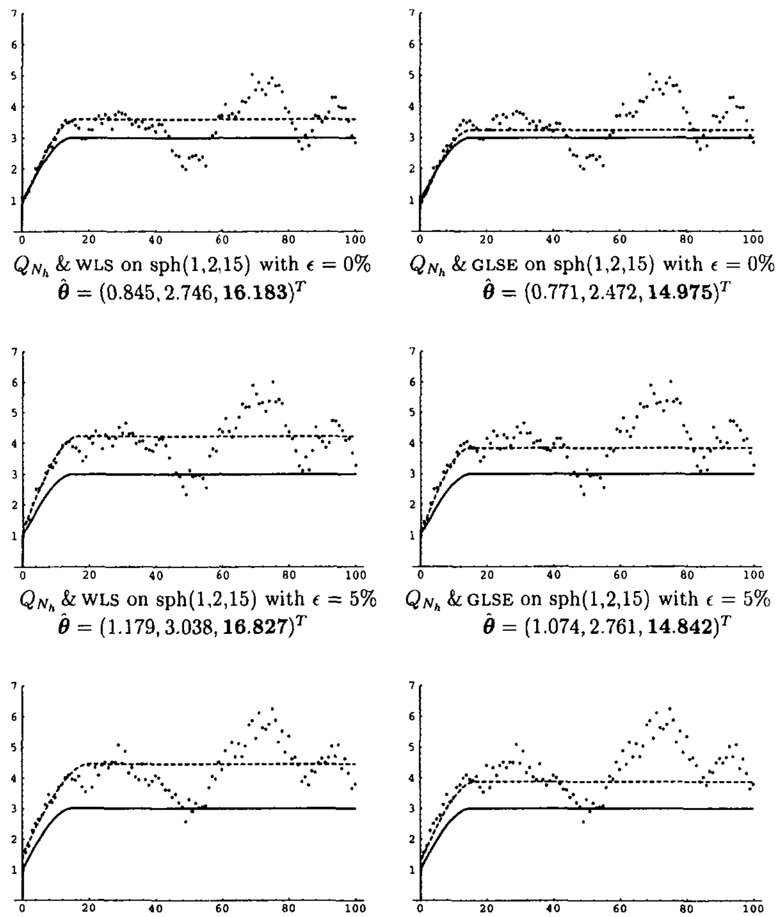


Figure 2. Example of estimation by L^2 of a perturbed spherical variogram and fit with WLS or GLSE (underlying variogram is solid line and fitted variogram is dashed line).

On each graph, the underlying variogram is represented by a solid line and the fitted variogram by a dashed line. The effect of perturbations is noticeable by a greater vertical variability of the variogram estimations. For L^2, a horizontal deformation is added, which leads to an increase of the range, expressed through the parameter c. This phenomena occurs to a much lesser extent for the Q_{N_h} estimator. When fitting, the method GLSE tends to reduce this effect. There-



Q_{N_h} & WLS on sph(1,2,15) with $\epsilon = 0\%$
$\hat{\boldsymbol{\theta}} = (0.845, 2.746, \mathbf{16.183})^T$

Q_{N_h} & GLSE on sph(1,2,15) with $\epsilon = 0\%$
$\hat{\boldsymbol{\theta}} = (0.771, 2.472, \mathbf{14.975})^T$

Q_{N_h} & WLS on sph(1,2,15) with $\epsilon = 5\%$
$\hat{\boldsymbol{\theta}} = (1.179, 3.038, \mathbf{16.827})^T$

Q_{N_h} & GLSE on sph(1,2,15) with $\epsilon = 5\%$
$\hat{\boldsymbol{\theta}} = (1.074, 2.761, \mathbf{14.842})^T$

Q_{N_h} & WLS on sph(1,2,15) with $\epsilon = 10\%$
$\hat{\boldsymbol{\theta}} = (1.481, 2.970, \mathbf{20.285})^T$

Q_{N_h} & GLSE on sph(1,2,15) with $\epsilon = 10\%$
$\hat{\boldsymbol{\theta}} = (1.247, 2.632, \mathbf{16.581})^T$

Figure 3. Example of estimation by Q_{N_h} of a perturbed spherical variogram and fit with WLS or GLSE (underlying variogram is solid line and fitted variogram is dashed line).

fore, the combination of Q_{N_h} and of GLSE gives the best estimation of the parameter c.

Note that there exist other ways of computing the matrix Ω arising in the method of generalized least squares. Here are some possibilities:

- $\Omega_{ij} = \mathrm{Corr}(2\hat{\gamma}(i),\ 2\hat{\gamma}(j))\hat{\gamma}(i)\hat{\gamma}(j)/\sqrt{N_i N_j}$, with $\mathrm{Corr}(2\hat{\gamma}(i),\ 2\hat{\gamma}(j))$ given by formula (d) of Theorem 2.

- $\Omega_{ij} = \text{Cov}(2\hat{\gamma}(i), 2\hat{\gamma}(j))$, with $\text{Cov}(2\hat{\gamma}(i), 2\hat{\gamma}(j))$ given by formula (c) of Theorem 2.
- $\Omega_{ii} = \hat{\gamma}(i)^2/N_i$.

Use of these three propositions on simulations leads to poor results, the quality of which are located between WLS of Cressie (1985) and our method GLSE.

GENERALIZATION TO DATA IN $\mathbb{R}^d$

A generalization of the previous method of variogram fitting to data in $\mathbb{R}^d$, with $d \geq 2$, is now described. It is based once again on the correlation structure of Matheron's classical variogram estimator. Consider a spatial stochastic process $\{Z(\mathbf{x}):\mathbf{x} \in D\}$, $D \subset \mathbb{R}^d$, which is ergodic, intrinsically stationary and isotropic. This means that the variogram depends only on the norm of the lag vector $\mathbf{h}$, and not on direction. Suppose there are $n = \Pi_{i=1}^{d} n_i$ data points, located on a multidimensional and regular grid, where n_i is the number of points along the ith unidimensional axis. The variogram $2\gamma(h)$ of this process is estimated on the axes of the grid by Matheron's classical variogram estimator $2\hat{\gamma}(h) = \mathbf{z}^T A(h)\mathbf{z}$, where $\mathbf{z} \in \mathbb{R}^n$ is the data vector. This means that one neglects a few directions, which are diagonally oriented. Therefore, the spatial design matrix $A(h)$ can be split along each axis of the grid. Using the notion of Kronecker product between matrices (Fang and Zhang, 1990), it can be written as

$$A(h) = \frac{1}{N_h} \sum_{i=1}^{d} I_{n_d \cdot \cdots \cdot n_{i+1}} \otimes A^{(i)}(h) \otimes I_{n_{i-1} \cdot \cdots \cdot n_1} \tag{12}$$

where $A^{(i)}(h) \in \mathbb{R}^{n_i \times n_i}$ is the spatial design matrix along axis i, given at the beginning of the paper, that is to say

$$A^{(i)}(h) = \begin{pmatrix} I_{(n_i - h)+} & -I_{(n_i - h)+} \\ -I_{(n_i - h)+} & I_{(n_i - h)+} \end{pmatrix}$$

and $N_h = \Sigma_{i=1}^{d} [(n_i - h)_+(\Pi_{j \neq i} n_j)]$ is the number of differences of length h on the grid. The notation $(n_i - h)_+$ means

$$(n_i - h)_+ = \begin{cases} n_i - h & \text{if } h < n_i \\ 0 & \text{otherwise} \end{cases}$$

Behavior of the correlation structure of Matheron's classical variogram estimator is shown in the next theorem in a particular context.

Theorem 3. Let $2\hat{\gamma}(h) = \mathbf{z}^T A(h)\mathbf{z}$ be Matheron's classical variogram estimator, applied on a multidimensional and regular grid $D \subset \mathbb{R}^d$. Let $\mathbf{z} \in \mathbb{R}^n$ be the data vector, supported on $n = \Pi_{i=1}^{d} n_i$ points, where n_i is the number of points on each unidimensional axis $i = 1, \ldots, d$. Suppose that $E(\mathbf{z}) = \mu \mathbf{1}_n$ and $\text{Var}(\mathbf{z}) = \sigma^2 I_n$. Note $N_h = \Sigma_{i=1}^{d} [(n_i - h)_+ (\Pi_{j \neq i} n_j)]$ the number of differences of length h on the grid. Then, the expectation of Matheron's classical variogram estimator is

(a) $E(2\hat{\gamma}(h)) = 2\sigma^2$

Moreover, if $\mathbf{z}$ is Gaussian, then

$$\text{(b) } \text{Var}(2\hat{\gamma}(h)) = \frac{1}{N_h^2} \left[\sum_{i=1}^{d} \frac{n(n_i - h)_+^2}{n_i} \text{Var}^{(i)}(2\hat{\gamma}(h)) + 8\sigma^4 \sum_{i \neq j} \frac{n(n_i - h)_+ (n_j - h)_+}{n_i n_j} \right]$$

and for $h_1 < h_2$

$$\text{(c) } \text{Cov}(2\hat{\gamma}(h_1), 2\hat{\gamma}(h_2)) = \frac{1}{N_{h_1} N_{h_2}} \left[\sum_{i=1}^{d} \frac{n(n_i - h_1)_+ (n_i - h_2)_+}{n_i} \cdot \text{Cov}^{(i)}(2\hat{\gamma}(h_1), 2\hat{\gamma}(h_2)) + 8\sigma^4 \sum_{i \neq j} \frac{n(n_i - h_1)_+ (n_j - h_2)_+}{n_i n_j} \right]$$

$$\text{(d) } \text{Corr}(2\hat{\gamma}(h_1), 2\hat{\gamma}(h_2)) = \frac{\text{Cov}(2\hat{\gamma}(h_1), 2\hat{\gamma}(h_2))}{\sqrt{\text{Var}(2\hat{\gamma}(h_1)) \, \text{Var}(2\hat{\gamma}(h_2))}}$$

where $\text{Var}^{(i)}$ and $\text{Cov}^{(i)}$ are, respectively, the unidimensional variance and covariance along each axis i, given in Theorem 2.

Proof. Recall that the Kronecker product of two matrices $A \in \mathbb{R}^{n \times p}$ and $B \in \mathbb{R}^{m \times q}$ is defined by

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1p}B \\ \vdots & & \vdots \\ a_{n1}B & \cdots & a_{np}B \end{pmatrix} \in \mathbb{R}^{nm \times pq}$$

and that it has some simple properties such as $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ and $\text{tr}[A \otimes B] = \text{tr}[A] \, \text{tr}[B]$. Using the special form of $A(h)$ given by

$$A(h) = \frac{1}{N_h} \sum_{i=1}^{d} I_{n_d \cdot \ldots \cdot n_{i+1}} \otimes A^{(i)}(h) \otimes I_{n_{i-1} \cdot \ldots \cdot n_1}$$

one can write for the covariance

$$\mathrm{Cov}(2\hat{\gamma}(h_1), 2\hat{\gamma}(h_2))$$

$$= 2\sigma^4 \mathrm{tr}[A(h_1)A(h_2)]$$

$$= \frac{2\sigma^4}{N_{h_1}N_{h_2}} \mathrm{tr}\left[\left(\sum_{i=1}^{d} I_{n_d \cdot \ldots \cdot n_{i+1}} \otimes A^{(i)}(h_1) \otimes I_{n_{i-1} \cdot \ldots \cdot n_1}\right)\right.$$

$$\left. \cdot \left(\sum_{j=1}^{d} I_{n_d \cdot \ldots \cdot n_{j+1}} \otimes A^{(j)}(h_2) \otimes I_{n_{j-1} \cdot \ldots \cdot n_1}\right)\right]$$

$$= \frac{2\sigma^4}{N_{h_1}N_{h_2}} \mathrm{tr}\left[\sum_{i=1}^{d}\sum_{j=1}^{d} I_{n_d \cdot \ldots \cdot n_{i+1}}\right.$$

$$\left. \otimes A^{(i)}(h_1) \otimes I_{n_{i-1} \cdot \ldots \cdot n_{j+1}} \otimes A^{(j)}(h_2) \otimes I_{n_{j-1} \cdot \ldots \cdot n_1}\right]$$

$$= \frac{2\sigma^4}{N_{h_1}N_{h_2}} \left[\sum_{i=1}^{d} (n_d \cdot \ldots \cdot n_{i+1}) \mathrm{tr}[A^{(i)}(h_1)A^{(i)}(h_2)](n_{i-1} \cdot \ldots \cdot n_1)\right.$$

$$+ \sum_{i \neq j} (n_d \cdot \ldots \cdot n_{i+1}) \mathrm{tr}[A^{(i)}(h_1)](n_{i-1} \cdot \ldots \cdot n_{j+1})$$

$$\left. \cdot \mathrm{tr}[A^{(j)}(h_2)](n_{j-1} \cdot \ldots \cdot n_1)\right]$$

$$= \frac{1}{N_{h_1}N_{h_2}} \left[\sum_{i=1}^{d} \frac{n(n_i - h_1)_+(n_i - h_2)^+}{n_i} \mathrm{Cov}^{(i)}(2\hat{\gamma}(h_1), 2\hat{\gamma}(h_2))\right.$$

$$\left. + 8\sigma^4 \sum_{i \neq j} \frac{n(n_i - h_1)_+(n_j - h_2)_+}{n_i n_j}\right]$$

The expression for the variance is obtained by letting $h_1 = h_2 = h$ in the previous equation. □

Theorem 3 shows that variance and covariance of Matheron's classical variogram estimator are linear combinations of unidimensional variances and covariances. Nevertheless, this is not the case for the correlation. Figure 4 visualizes a contour plot of the correlation structure of Matheron's classical variogram estimator for different spatial dimensions d. The first graph corresponds to a unidimensional support of size 100, and one finds exactly the same picture as in Figure 1. Thereafter, the second and third graph correspond to bidimensional supports, and the following to tridimensional supports. Hence, these graphs show that correlation increases with spatial dimension d, and that shape changes as well. Therefore, it is necessary to take account of this new correlation when the variogram is fitted. Consequently, the generalized least

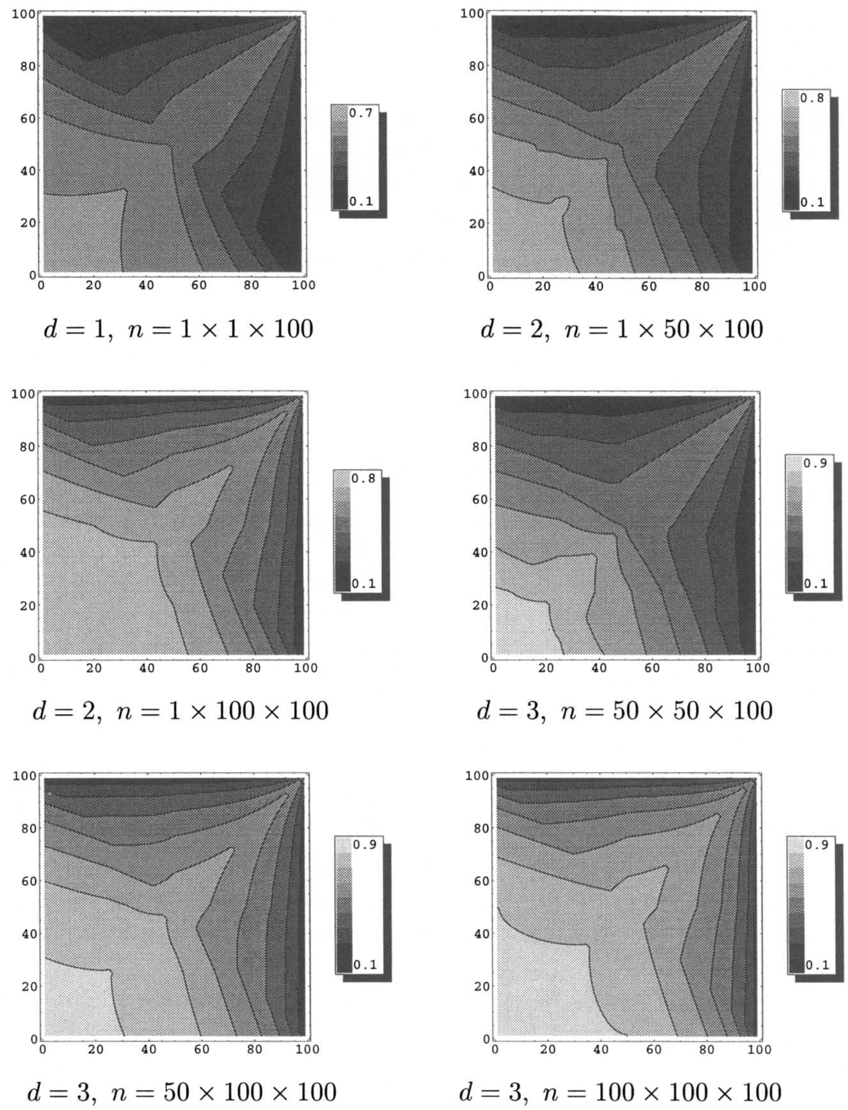


Figure 4. Dependence of Corr($2\hat{\gamma}(h_1)$, $2\hat{\gamma}(h_2)$) on lags h_1 (horizontal axis) and h_2 (vertical axis), for independently and identically distributed observations from a standard Gaussian distribution. This changes with spatial dimension d and the number of data $n = n_1 \times n_2 \times n_3$. For Matheron's classical variogram estimator, explicit computation is possible (Theorem 3). Plots show contour lines for correlations between 0 and 0.1, between 0.1 and 0.2, etc.

squares method is the same as before, except that the correlation is given by formula (d) of Theorem 3.

Note further that if one wishes to estimate the variogram in a unique direction of the grid, given by the axis i, one only needs to let $A^{(j)}(h) = O$ for $j \neq i$ in the expression (12) of $A(h)$. Other situations which can arise in practice are irregular grids, unidimensional or multidimensional. In that case, one can adapt our method of fitting in the following manner:

- Neglect irregularity of the grid, if not too big, and use the correlation given by Theorem 3.
- Take partially account of the irregularity of the grid by modifying the number N_h of differences of length h in the formula for the correlation given by Theorem 3.

CONCLUSIONS

In this paper, I derived the correlation structure of Matheron's classical variogram estimator in the case of independent Gaussian observations on a multidimensional and regular grid $D \subset \mathbb{R}^d$, with $d \geq 1$. This structure has been used to approximate the variance-covariance matrix occurring in the generalized least squares method for variogram fitting. Statistical properties of the variogram estimator are thereby taken into account. Several simulations show that our technique (GLSE), combined with a robust estimator of the variogram, improves the fit significantly, even if outliers are present in the data.

ACKNOWLEDGMENTS

This work contains parts of my PhD dissertation, which was written under the generous guidance of Prof. Stephan Morgenthaler. I wish to thank the Swiss National Science Foundation for its financial support.

REFERENCES

Beran, J., 1994, Statistics for long-memory processes: Chapman & Hall, New York, 315 p.

Cressie, N., 1985, Fitting variogram models by weighted least squares: Math. Geology, v. 17, no. 5, p. 563–586.

Cressie, N., 1991, Statistics for spatial data: John Wiley & Sons, New York, 900 p.

Cressie, N., and Hawkins, D. M., 1980, Robust estimation of the variogram, I: Math. Geology, v. 12, no. 2, p. 115–125.

Fang, K., and Zhang, Y., 1990, Generalized multivariate analysis: Springer, Beijing, 220 p.

Fedorov, V. V., 1974, Regression problems with controllable variables subject to error: Biometrika, v. 61, no. 1, p. 49–56.

Genton, M. G., 1995, Robustesse dans l'estimation du variogramme: Bulletin de l'Institut International de Statistique, Beijing, China, v. 1, p. 400–401.

Genton, M. G., 1996, Robustness in variogram estimation and fitting: unpubl. doctoral dissertation, no. 1595, Swiss Federal Inst. Technology, Lausanne, 132 p.

Genton, M. G., 1998, Highly robust variogram estimation: Math. Geology, v. 30, no. 2, p. 213–221.

Genton, M. G., and Rousseeuw, P. J., 1995, The change-of-variance function of *M*-estimators of scale under general contamination: Jour. Computational and Applied Mathematics, v. 64, no. 1, p. 69–80.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A., 1986, Robust statistics, the approach based on influence functions: John Wiley & Sons, New York, 502 p.

Horn, R. A., and Johnson, C. R., 1991, Topics in matrix analysis: Cambridge Univ. Press, Cambridge, 607 p.

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

Matheron, G., 1962, Traité de géostatistique appliquée, Tome I: Mémoires du Bureau de Recherches Géologiques et Minières, no. 14, Editions Technip, Paris, 333 p.

Portnoy, S., 1977, Robust estimation in dependent situations: The Annals of Statistics, v. 1, p. 22–43.

Rousseeuw, P. J., and Croux, C., 1992, Explicit scale estimators with high breakdown point: L_1 Statistical Analyses and Related Methods, p. 77–92.

Rousseeuw, P. J., and Croux, C., 1993, Alternatives to the median absolute deviation: Jour. Am. Stat. Assoc., v. 88, no. 424, p. 1273–1283.

Seber, G. A. F., and Wild, C. J., 1989, Nonlinear regression: John Wiley & Sons, New York, 768 p.