Computational Statistics & Data Analysis 11 (1991) 87–96
North-Holland

Variogram fitting with a general class of conditionally nonnegative definite functions

A. Shapiro and J.D. Botha
University of South Africa, Pretoria, 0001 Republic of South Africa

Received August 1989
Revised December 1989

Abstract: In this paper we propose an easily implementable and computationally flexible approach to fitting of a permissible variogram function. It is shown that if the weighted least squares criterion is chosen, then the optimization problem thus obtained is reducible to a quadratic programming problem. Various requirements such as smoothness, monotonicity or convexity of the fitted function, can be formulated as additional linear constraints. One-, two- and three-dimensional cases are discussed and isotropic variograms are considered.

Keywords: Conditional nonnegative definiteness, Geostatistics, Fourier transform, Hankel transform, Quadratic programming, Variogram, Weighted least squares

1. Introduction

Estimation of the variogram and subsequent fitting of an appropriate variogram model is an important problem in geostatistics, hydrosciences and other geophysical sciences (see, e.g., Journel and Huijbregts (1978), Christakos (1984) and references therein). However, it often happens in practice that known variogram models do not provide a satisfactory approximation of the experimental variogram values. Therefore the practitioner may be tempted to search for a new model giving a more adequate explanation of the data. This raises the question of testing the validity of the proposed model. Unfortunately, the standard techniques involved are complicated and difficult to apply (Christakos (1984), Armstrong and Diamond (1984)).

In this paper we propose an easily implementable approach to fitting of (isotropic) variogram functions. Instead of choosing a variogram model depending on a small number of parameters, we consider a broad class of permissible variograms, and fit the data by an element from this class which minimizes a given criterion. Therefore the proposed techniques can be considered as *nonparametric*. Flexibility of the method is then demonstrated by showing how various requirements, such as variogram smoothness, monotonicity or convexity, can easily be incorporated into the fitting process. It should be understood that we are not concerned with the problem of calculation (estimation) of the variogram values from the data available. For a thorough discussion of this problem see Cressie and Hawkins (1980) and Cressie (1985). It will be assumed that the experimental variogram values are given, and we then concentrate on the problem of variogram fitting.

2. Preliminary discussion and problem formulation

In this section we will formulate the basic problem of variogram fitting. Let $Z(\boldsymbol{x})$ be a real-valued, second-order stationary stochastic process, defined over a domain D of $\mathbb{R}^d$, and consider the function

$$\gamma(\boldsymbol{h}) = 1/2 \operatorname{var}\{Z(\boldsymbol{x}+\boldsymbol{h}) - Z(\boldsymbol{x})\}, \qquad \boldsymbol{x} \text{ and } \boldsymbol{x}+\boldsymbol{h} \in D.$$

The functions $\gamma(\boldsymbol{h})$ and $2\gamma(\boldsymbol{h})$ are known as the *semivariogram* and *variogram* functions, respectively (Matheron (1963)). Notice that second-order stationarity of $Z(\boldsymbol{x})$ implies that the absolute value of $\gamma(\boldsymbol{h})$ is bounded. Therefore in this paper we deal with *bounded* variogram functions. Since we will be concerned with estimation of $\gamma(\boldsymbol{h})$ on a *finite* interval, there is no really loss of generality in considering bounded variograms as compared with the so-called intrinsic hypothesis which allows for unbounded variogram models.

Since $2\gamma(\boldsymbol{h})$ is defined as the variance of increments of $Z(\boldsymbol{x})$, it must satisfy certain conditions. Following Christakos (1984) we say that a function $g(\boldsymbol{h})$, $\boldsymbol{h} \in \mathbb{R}^d$, is a *permissible* semivariogram function if it is continuous (except possibly at the origin), $g(\boldsymbol{h}) = g(-\boldsymbol{h})$, $g(\boldsymbol{h}) \geq 0$ for all $\boldsymbol{h}$, and $-g(\boldsymbol{h})$ is *conditionally nonnegative definite*, that is

$$-\sum_{i,\, j=1}^{m} \lambda_i \lambda_j g(\boldsymbol{x}_i - \boldsymbol{x}_j) \geq 0, \tag{1}$$

for any $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^d$ and any $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ such that $\sum_{i=1}^{m} \lambda_i = 0$. The requirement of conditional nonnegative (or rather positive) definiteness for $-\gamma(\boldsymbol{h})$ is especially important in general linear-model estimation when deriving the best linear unbiased estimator (Matheron (1973)).

We will restrict our attention to *isotropic* functions $g(\boldsymbol{h})$, which depend on $\boldsymbol{h}$ only through its Euclidean norm

$$r = |\boldsymbol{h}| = \left(h_1^2 + \cdots + h_d^2\right)^{1/2}.$$

That is

$$g(\boldsymbol{h}) = g(|\boldsymbol{h}|) = g(r).$$

The class of all permissible isotropic functions will be denoted by $\mathcal{G}_d$. Notice that $\mathcal{G}_d$ depends on the dimensionality d, and if $k > d$, then $\mathcal{G}_k \subset \mathcal{G}_d$.

Let $\hat{\gamma}_i = \hat{\gamma}(r_i)$, $i = 1, \ldots, n$, be estimated semivariogram values. For example, one can use the classical estimator

$$\hat{\gamma}(r) = \tfrac{1}{2}N^{-1} \sum_{i=1}^{N} [Z(\boldsymbol{x}_i + \boldsymbol{h}) - Z(\boldsymbol{x}_i)]^2,$$

where $|\boldsymbol{h}| = r$ and N is the number of lag-h differences (Matheron (1963)). For a discussion of other, more robust and seemingly better, estimators $\hat{\gamma}$ see Cressie and Hawkins (1980). Choose an objective function $\rho(t)$ and positive numbers (weights) $w_1, \ldots, w_n$. Then we formulate the basic fitting problem as follows: Find a function $\bar{g} \in \mathscr{G}_d$ such that:

$$\sum_{i=1}^{n} w_i \rho(\hat{\gamma}_i - \bar{g}(r_i)) = \inf_{g \in \mathscr{G}_d} \sum_{i=1}^{n} w_i \rho(\hat{\gamma}_i - g(r_i)). \tag{2}$$

The standard choice of the objective function is $\rho(t) = t^2$, in which case problem (2) becomes a weighted least squares estimation, although some other choices (such as $\rho(t) = |t|$) are possible. It was shown by Cressie (1985) that approximately optimal (in a certain statistical sense) weights are given by $w_i = n_i g(r_i)^{-2}$, where n_i is the number of pairs used to estimate the semivariogram at the ith lag. These weights depend on the fitted semivariogram function $g(r)$ and are not known *a priori*. We discuss how to deal with this problem later. We note that the class $\mathscr{G}_d$ is too large to give a meaningful variogram estimator $\bar{g}$. Even the verification of existence of such a minimizer $\bar{g}$ is a non-trivial matter. Subsequently we restrict the estimation problem (2) to a manageable one which is still sufficiently flexible for practical purposes.

Some simple properties of the class $\mathscr{G}_d$ are now in order. Clearly if g_1 and g_2 are in $\mathscr{G}_d$, then $\alpha g_1 + \beta g_2 \in \mathscr{G}_d$ for any nonnegative numbers α and β. If $g \in \mathscr{G}_d$ and c_0 is a constant, then the function $g(r) + c_0$ is in $\mathscr{G}_d$ provided it has nonnegative values for all $r \geq 0$. Consider an isotropic, nonnegative definite function $f(r)$, i.e. for every $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^d$ and $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$,

$$\sum_{i,j=1}^{m} \lambda_i \lambda_j f(|\boldsymbol{x}_i - \boldsymbol{x}_j|) \geq 0.$$

Then for any constant c_0 the function $f(r) - c_0$ is conditionally nonnegative definite. If in addition $f(r)$ is continuous and $-f(r) + c_0 \geq 0$ for all $r \geq 0$, then $g(r) = -f(r) + c_0 \in \mathscr{G}_d$. Therefore we restrict ourselves to a subclass of $\mathscr{G}_d$ consisting of functions g representable in the form $g(r) = c_0 - f(r)$, where $f(r)$ is an isotropic nonnegative definite function. Since the absolute value of a nonnegative definite function is bounded by its value at the origin, it follows that if $g(0) = c_0 - f(0) \geq 0$, then $g(r)$ is greater than or equal to zero for all $r \geq 0$. It also follows that the functions $g(r)$ thus obtained are *bounded*.

Since it was assumed that the process $Z(\boldsymbol{x})$ is second-order stationary we have that

$$\gamma(\boldsymbol{h}) = c(\boldsymbol{0}) - c(\boldsymbol{h}),$$

where $c(\boldsymbol{h})$ is the covariance of $Z(\boldsymbol{x}+\boldsymbol{h})$ and $Z(\boldsymbol{x})$. Therefore in the representation $g=c_0-f$ one can interpret f as the corresponding covariance function.

3. The one-dimensional case

In this section we outline an approach to solving the basic fitting problem in the one-dimensional case, i.e. $d=1$. From Bochner's theorem we know that a function f is continuous and nonnegative definite if and only if it is the Fourier transform of a nonnegative bounded Borel measure μ. That is, f is representable in the form

$$f(r)=\int_0^\infty \cos(rt)\,\mathrm{d}F(t), \qquad r\geq 0,$$

where $F(t)$ is a bounded function which is monotonically increasing on the interval $[0,\infty)$. Consider the associated functions $g(r)=c_0-f(r)$. The weighted least squares fitting problem will now be formulated as follows: Find a bounded monotonically increasing function $F(t)$ and a constant c_0 which minimize

$$\sum_{i=1}^n w_i\left[\hat{\gamma}_i-c_0+\int_0^\infty \cos(r_i t)\,\mathrm{d}F(t)\right]^2 \tag{3}$$

subject to

$$c_0-\int_0^\infty \mathrm{d}F(t)\geq 0. \tag{4}$$

The above estimation problem is still too general to be solved numerically. Therefore we restrict it further by considering only atomic measures μ. That is, let $F(t)$ be a step function with a finite number of positive jumps $y_1,\ldots,y_m$ at points $t_1,\ldots,t_m$. This reduces the problem to the following finite-dimensional version: Find the $(m+1)$-dimensional vector $z=(y_1,\ldots,y_m,c_0)^T$ which minimizes the function

$$Q(z)=\sum_{i=1}^n w_i\left[\hat{\gamma}_i-c_0+\sum_{j=1}^m \cos(r_i t_j)y_j\right]^2 \tag{5}$$

subject to

$$y_j\geq 0, \quad j=1,\ldots,m, \quad \text{and} \tag{6}$$

$$c_0-\sum_{j=1}^m y_j\geq 0. \tag{7}$$

The objective function $Q(z)$ is quadratic and can be written in the following form. Consider the data vector $\hat{\gamma}=(\hat{\gamma}_1,\ldots,\hat{\gamma}_n)^{\mathrm{T}}$, let $W=\mathrm{diag}(w_i)$ be the $n\times n$ diagonal matrix of weights and let $A=(a_{ij})$ be the $n\times(m+1)$ matrix given by

$a_{ij} = -\cos(r_i t_j)$ for $i = 1,\ldots, n$; $j = 1,\ldots, m$, and $a_{i(m+1)} = 1$ for $i = 1,\ldots, n$. Then

$$Q(z) = (\hat{\gamma} - Az)^T W(\hat{\gamma} - Az). \tag{8}$$

Also, one may note that the constraints (6) and (7) are linear. The minimization problem is then a quadratic programming problem, which can be solved by standard techniques (see, e.g., Gill, Murray and Wright (1981)). After an optimal solution, $\bar{z} = (\bar{y}_1,\ldots, \bar{y}_m, \bar{c}_0)^T$, is calculated the fitted semivariogram $\bar{g}(r)$ is obtained by the formula

$$\bar{g}(r) = \bar{c}_0 - \sum_{j=1}^{m} \cos(rt_j)\bar{y}_j. \tag{9}$$

Notice that $g(0)$ is given by the left-hand side of the inequality (7). Therefore the inequality constraint (7) can be replaced by the (linear) equality constraint

$$c_0 - \sum_{j=1}^{m} y_j = b, \tag{10}$$

where b is a prescribed nonnegative number. A choice of b strictly greater than zero will correspond to the so-called *nugget effect* (Journel and Huijbregts (1978, p. 39)).

The quadratic term of the function $Q(z)$ is given by $z^T A^T W A z$. Therefore, in order to ensure nonsingularity (and hence positive definiteness) of the matrix $A^T WA$, m must satisfy $m + 1 \le n$. This motivated us to choose $m = n - 1$. The points $t_1,\ldots, t_m$ were taken with a fixed step-length δ, that is, $t_j = \delta j$, $j = 1,\ldots, m$, where the positive number δ was chosen on an ad hoc basis.

The optimal solution of the optimization problem (5)–(7) may follow the data values $\hat{\gamma}_i$ too closely. Consequently, especially when the $\hat{\gamma}_i$ are scattered, the fitted function $\bar{g}(r)$ may change rapidly. In order to eliminate such noisy behavior, further constraints are required. For example, one could impose a smoothness condition by requiring the derivative

$$g'(r) = \sum_{j=1}^{m} t_j \sin(rt_j) y_j \tag{11}$$

to be bounded. It follows from (11) that

$$|g'(r)| \le \sum_{j=1}^{m} t_j |y_j|.$$

Therefore we introduce the (linear) constraint

$$\sum_{j=1}^{m} t_j y_j \le K, \tag{12}$$

where K is some chosen positive constant. We shall refer to the inequality constraint (12) as the "gradient bound".

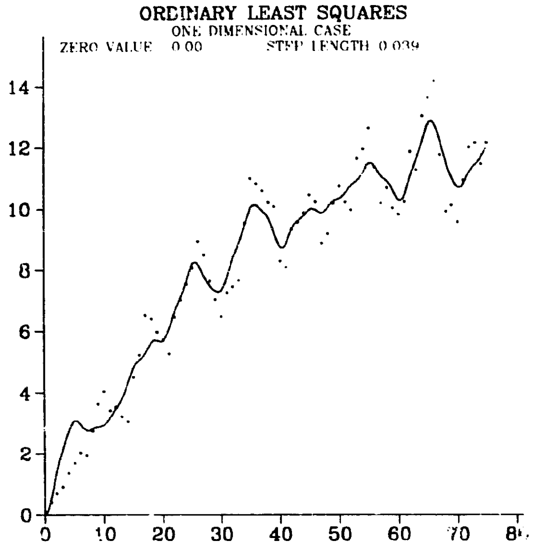

Fig. 1.

When the weights are taken as suggested by Cressie (1985), $w_i = n_i g(r_i)^{-2}$, the obtained optimization problem is not quadratic. We propose a simple iterative procedure (reiterated least squares) as a way of dealing with this problem. Fix the weights w_i and solve the corresponding quadratic programming problem. Use the calculated semivariogram $\bar{g}(r)$ to update the weights by taking $w_i = n_i \bar{g}(r_i)^{-2}$. Substitute the new set of weights into the quadratic programming problem, and so forth. Make a few iterations untill the calculated weights stabilize. In our experience, the proposed procedure always converged in a few (two-three) iterations and was insensitive to starting values of the weights w_i. The estimated semivariogram thus obtained usually did not differ significantly from the semivariogram fitted by taking all weights equal one (the ordinary least squares).

As an example, we analysed an experimental semivariogram given in Clark (1979, p. 21) with the number n of data points equal to 75. The following values of the parameters were used. All weights w_i are equal to one, the step $\delta = 0.039$, the number of frequencies $m = n - 1 = 74$, the constant b in the constraint (10) is zero and the constant K in the gradient bound inequality (12) is taken to be the maximal slope of the experimental semivariogram. The quadratic programming problem was solved by applying subroutine QPROG from the IMSL software library. The experimental values and the fitted semivariogram are shown in Figure 1.

It can be seen that the semivariogram values are fairly scattered, and the fitted semivariogram $\bar{g}(r)$ behaves quite irregularly. One can force smoother behavior of $\bar{g}(r)$ by reducing the constant K in the gradient bound constraint (12). There are also two other alternatives. We could require that the fitted semivariogram function be monotonically increasing or that it be convex.

The monotonicity condition can be ensured by requiring that $g'(r)$ be nonnegative for all $r > 0$. In order to solve this numerically, the following discretiza-

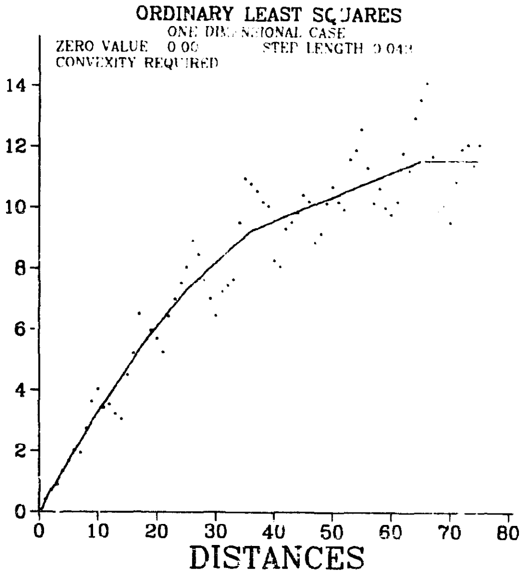

Fig. 2.

tion will be employed

$$g'(r_i) = \sum_{j=1}^{m} t_j \sin(r_i t_j) y_j \geq \epsilon, \qquad i = 1, \ldots, n, \tag{13}$$

where ϵ is a small positive number and the r_i are some chosen points. We take r_i to be the lag corresponding to the experimental semivariogram values. Notice that the inequality constraints (13) are linear in y_j. Similarly the convexity (or rather, mathematically speaking, the concavity) of the fitted semivariogram function can be ensured by the constraints

$$g''(r_i) = \sum_{j=1}^{m} t_j^2 \cos(r_i t_j) y_j \leq -\epsilon, \qquad i = 1, \ldots, n. \tag{14}$$

The same experimental semivariogram is now analysed by replacing the gradient bound constraint (12) by the convexity constraints (14). The semivariogram thus obtained is shown in Figure 2.

It can be seen that the convexity requirement automatically enforces smooth behavior of the fitted function.

4. Higher-dimensional cases

The approach outlined in the previous section can easily be extended to higher dimensions by replacing the Fourier transform by the Hankel transform (see, e.g., Sneddon (1951) for the definition and basic properties of the Hankel transform). In this section we discuss this in some detail for the cases $d = 2$ and $d = 3$.

Consider Bessel functions $J_k(x)$ of the first kind. A two-dimensional isotropic, nonnegative definite function $f(r)$ can be represented as the zero-order Hankel transform

$$f(r)=\int_0^{\infty} tJ_0(rt)\,\mathrm{d}F(t) \tag{15}$$

of the monotonically increasing function $F(t)$. Consequently the two-dimensional analogue of the problem (5)–(7) is: Find the $(m+1)$-dimensional vector $z=(y_1,\ldots,y_m,c_0)^T$ which minimizes the function

$$Q^{(2)}(z)=\sum_{i=1}^{n} w_i\left[\hat{\gamma}_i-c_0+\sum_{j=1}^{m} J_0(r_it_j)y_j\right]^2 \tag{16}$$

subject to constraints (6) and (7).

If $\bar{z}=(\bar{y}_1,\ldots,\bar{y}_m,\bar{c}_0)^T$ is the solution of the above problem, then the fitted (two dimensional) semivariogram is given by

$$\bar{g}(r)=\bar{c}_0-\sum_{j=1}^{m} J_0(rt_j)\bar{y}_j. \tag{17}$$

As in the one-dimensional case, the inequality constraint (7) can be replaced by the equality constraint (10). An analogue for the smoothness condition also can be given. Since $J_0'(x)=-J_1(x)$ we have

$$g'(r)=\sum_{j=1}^{m} t_jJ_1(rt_j)y_j.$$

Therefore one can impose the constraints

$$a_i\le\sum_{j=1}^{m} t_jJ_1(r_it_j)y_j\le b_i, \tag{18}$$

which correspond to the smoothness requirement (for example $a_i=-K$ and $b_i=K$, $i=1,\ldots,n$). The choice of $a_i=\epsilon$ and $b_i=\infty$, $i=1,\ldots,n$, where ϵ is a small positive number, will ensure the monotonicity condition.

Since $xJ_1'(x)=xJ_0(x)-J_1(x)$ we can formulate the convexity condition in the form of constraints

$$g''(r_i)=\sum_{j=1}^{m}\left[t_j^2J_0(r_it_j)-r_i^{-1}t_jJ_1(r_it_j)\right]y_j\le-\epsilon,\qquad i=1,\ldots,n.$$

Finally, for $d=3$, the corresponding Hankel transform, of order 1/2, provides three-dimensional isotropic, nonnegative definite functions in the form

$$f(r)=r^{-1/2}\int_0^{\infty} t^{3/2}J_{1/2}(rt)\,\mathrm{d}F(t),$$

where $F(t)$ is a monotonically increasing function. Since

$$J_{1/2}(x)=(2/\pi)^{1/2}x^{-1/2}\sin x$$

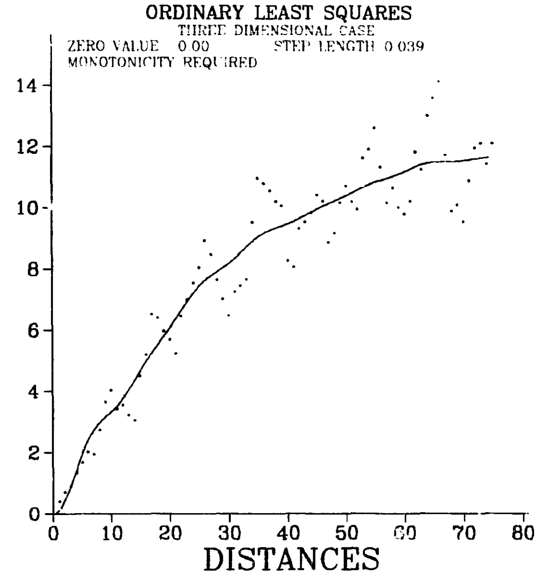

Fig. 3.

we obtain

$$f(r)=\int_0^\infty t^2 S(rt)\,\mathrm{d}F(t),$$

where $S(x)=x^{-1}\sin x$ (the constant $(2/\pi)^{1/2}$ is omitted). The corresponding objective function becomes

$$Q^{(3)}(z)=\sum_{i=1}^{n} w_i\left[\hat{\gamma}_i-c_0+\sum_{j=1}^{m} S(r_i t_j)y_j\right]^2.$$

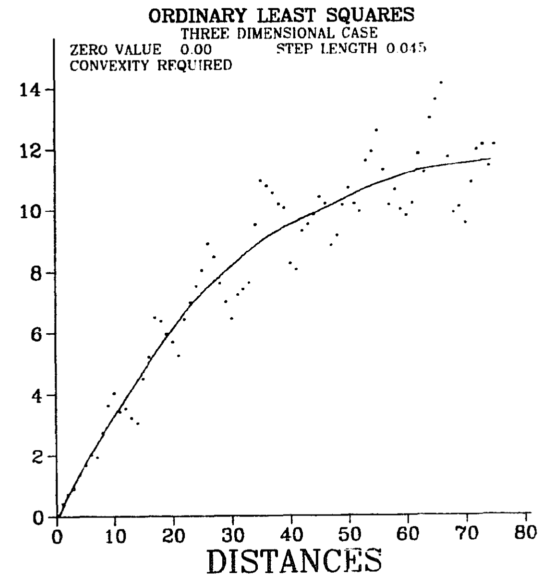

Fig. 4.

The constraints (6) and (7) remain the same and the fitted semivariogram is

$$\bar{g}(r) = \bar{c}_0 - \sum_{j=1}^{m} S(rt_j)\bar{y}_j.$$

The smoothness, monotonicity and convexity conditions can also be written in the form of linear constraints.

In Figures 3 and 4, the experimental semivariogram is approximated by three-dimensional semivariogram functions subject to the monotonicity and convexity conditions, respectively.

Again, smooth behavior of the fitted function is enforced here by either of the monotonicity or convexity conditions.

Acknowledgement

The authors are indebted to Prof. F.E. Steffens for bringing the problem to their attention, and for many helpful discussions. Without his kind collaboration and encouragement this research would never have been completed.

References

Armstrong, M. and Diamond, P., 1984, Testing variograms for positive-definiteness. *Math. Geol.*, **16**, pp. 407–421.

Christakos, G., 1984, On the problem of permissible covariance and variogram models. *Water Resources Res.*, **20**, pp. 251–265.

Clark, I., 1979, *Practical Geostatistics.* Applied Science Publishers, London.

Cressie, N. and Hawkins, D.M., 1980, Robust estimation of the variogram, I. *Math. Geol.* **12**, pp. 115–125.

Cressie, N., 1985, Fitting variogram models by weighted least squares. *Math. Geol.*, **17**, pp. 563–586.

Gill, P.E., Murray, W. and Wright, M.H., 1981, *Practical Optimization.* Academic Press, London.

Journel, A.G. and Huijbregts, Ch.J., 1978, *Mining Geostatistics.* Academic Press, London.

Matheron, G., 1963, Principles of geostatistics. *Economic Geology*, **58**, pp. 1246–1266.

Matheron, G., 1973, The intrinsic random functions and their applications. *Adv. in Applied Probability*, **5**, pp. 439–468.

Sneddon, I.N., 1951, *Fourier Transforms.* McGraw Hill, New York.