Journal of Applied Statistics, Vol. 23, No. 4, 1996, 435–449

CARFAX

An evaluation of a non-parametric method of estimating semi-variograms of isotropic spatial processes

S. CHERRY, J. BANFIELD & W. F. QUIMBY, *Department of Mathematical Sciences, Montana State University, USA*

SUMMARY *Semi-variograms are useful for describing the correlation strucure of spatial random variables. Valid semi-variograms must be conditionally negative definite. To ensure this restriction when estimating these functions, a valid parametric model is typically fitted to a sample semi-variogram. Recently, a method of fitting valid semi-variograms without having to choose a parametric family has been described in the literature. The method is based on the spectral representation of positive definite functions. In this paper, the method is evaluated using simulated data. The fits obtained using the non-parametric method are compared with fits obtained by fitting four parametric models (exponential, Gaussian, rational quadratic and power) to simulated data using non-linear least squares. The comparisons are based on the integrated squared errors of the resulting fits. The non-parametric estimator always resulted in fits that were as good as those obtained using the parametric models. The non-parametric method is faster, easier to use and more objective than the parametric methods. Some examples are presented.*

1 Introduction

Semi-variograms are used by geostatisticians to describe correlation structure among spatial random variables (Cressie, 1991). They are useful as a description of spatial dependence and also play a key role in spatial prediction (i.e. kriging).

Following Cressie (1991), let

$$\{Z(s): s \in D\}$$

be a spatial stochastic process, where $D \subset \mathrm{R}^d$. $Z(s)$ represents a random variable for each location s, where s is assumed to vary continuously over D.

Correspondence: S. Cherry, Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717, USA.

0266-4763/96/040435-15 $6.00

The semi-variogram is defined to be

$$\gamma(s_i, s_j) = \frac{\mathrm{Var}[Z(s_i) - Z(s_j)]}{2}$$

For second-order stationary isotropic random fields, there is a simple relationship between semi-variograms and covariance functions. Denoting a covariance function by $C(h)$ and a semi-variogram by $\gamma(h)$, where h denotes the Euclidean distance between two points, it is not difficult to show that

$$\gamma(h) = C(0) - C(h)$$

where $C(0) = \mathrm{Var}[Z(s)]$. The quantity $C(0)$ is referred to as the 'sill' in geostatistics.

Valid covariance functions must be positive definite. Similarly, valid semi-variogram functions must satisfy a mathematical property known as conditional negative definiteness. Letting h_{ij} denote the Euclidean distance between two points s_i and s_j in R^d, these two properties are defined as follows.

A function $C(h_{ij})$ is said to be positive definite in R^d if

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j C(h_{ij}) \geqslant 0$$

for all λ_i, h_{ij} and n. A function $\gamma(h_{ij})$ is said to be conditionally negative definite in R^d *if*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma(h_{ij}) \leqslant 0$$

for all h_{ij} and n, and for all λ_i, with the condition that $\Sigma\ \lambda_i = 0$.

It is easy to see that, if $C(h)$ is positive definite in R^d, then $k - C(h)$ is conditionally negative definite in R^d for any $k \in \mathrm{R}^1$. Also, it should be noted that functions that are positive definite (conditionally negative definite) in R^d are not necessarily positive definite (conditionally negative definite) in R^p where $p > d$, but are positive definite (conditionally negative definite) in R^p where $p < d$.

To guarantee conditional negative definiteness, geostatisticians typically fit known valid models to sample semi-variograms. A variety of methods have been used to fit semi-variograms, including maximum-likelihood, restricted maximum-likelihood and non-linear least-squares methods (Cressie, 1991). The most commonly used method has been to fit a model by inspection. The choice of a model is subjective and implies assumptions about the underlying spatial process. This subjectivity can lead to differences in conclusions about the correlation structure of the process (Englund, 1990). It has been recognised that a non-parametric method of estimating semi-variogram functions would be desirable (Cressie, 1991), but satisfying the property of conditional negative definiteness has been an obstacle.

Recently, several papers have addressed this issue. Shapiro and Botha (1991) and Sampson and Guttorp (1992) estimated semi-variograms using a linear combination of a more general class of known valid functions. Hall *et al.* (1994) used a modified kernel regression technique. Lele (1995) has developed a method of constructing conditionally negative definite matrices for use in kriging. His method does not require the specification of a parametric family of semi-variograms and performs well in practice (Lele, 1995). It does not actually produce an explicit semi-variogram, and does not provide estimates of parameters such as the sill.

In what follows, the Shapiro–Botha (SB) method of fitting valid semi-variograms

will be examined further. Section 2 describes the SB method of estimating semi-variograms. Section 3 describes how the SB method was implemented, and Section 4 discusses the results of comparing the SB method with a more traditional parametric method. Two examples are presented in Section 5, and conclusions and recommendations for implementation are discussed in Section 6.

2 The SB method

The SB non-parametric estimation procedure relies on the spectral representation of isotropic positive definite functions. Schoenberg (1938) established the following results.

A function $C(h)$ is positive definite in R^d if and only if it has a spectral representation

$$C(h) = \int_0^\infty \Omega_d(ht)\, \mathrm{d}F(t)$$

where

$$\Omega_d(x) = \left(\frac{2}{x}\right)^{(d-2)/2} \Gamma\left(\frac{d}{2}\right) \mathcal{J}_{(d-2)/2}(x)$$

Here, $\mathcal{J}_v$ is the Bessel function of the first kind of order v, and F is a non-decreasing bounded function on $t \geqslant 0$.

Fortunately, the somewhat complicated $\Omega_d(x)$ reduces as follows:

- $\Omega_1(x) = \cos(x)$ for one dimension;
- $\Omega_2(x) = \mathcal{J}_0(x)$ for two dimensions;
- $\Omega_3(x) = \sin(x)/x$ for three dimensions.

Shapiro and Botha (1991) formulated the problem of non-parametric estimation of semi-variograms as follows. Using the spectral properties of positive definite functions, and assuming second-order stationarity and isotropy, they wrote the semi-variogram as

$$\begin{aligned}\gamma(h) &= C(0) - C(h) \\ &= C(0) - \int_0^\infty \Omega_d(ht)\, \mathrm{d}F(t)\end{aligned}$$

Letting F be a step function with non-negative jumps p_j at nodes t_j, $j = 1, \ldots, m$, they considered the problem of finding an $(m+1)$-dimensional vector $\mathbf{p} = (p_1, \ldots, p_m, c_0)'$ that minimizes

$$\mathbf{Q}(\mathbf{p}) = \sum_{i=1}^{r} w_i \left[\hat{\gamma}(h_i) - c_0 + \sum_{j=1}^{m} \Omega_d(h_i t_j) p_j\right]^2$$

where $\hat{\gamma}(h)$ is a sample semi-variogram and $c_0 = C(0)$. The w_i terms are weights and r is the number of lags at which a semi-variogram estimate exists. The minimization is carried out under the constraints that

$$p_j \geqslant 0, \qquad j = 1, \ldots, m$$

and

$$c_0 - \sum_{j=1}^{m} p_j \geqslant 0$$

The problem was modified for the comparisons in this paper. On recognizing that

$$C(0) = \int_0^{\infty} \mathrm{d}F(t) = \sum_{j=1}^{m} p_j$$

the above problem can be recast as one of finding an m-dimensional vector $\mathbf{p} = (p_1, \ldots, p_m)'$ to minimize

$$\mathbf{Q}(\mathbf{p}) = \sum_{i=1}^{r} w_i \left\{ \hat{\gamma}(h_i) - \sum_{j=1}^{m} [1 - \boldsymbol{\Omega}_d(h_i t_j)] p_j \right\}^2 \tag{1}$$

subject to the constraint that $p_j \geqslant 0$, $j = 1, \ldots, m$.

Let $\mathbf{A}$ be an $r \times m$ matrix with elements $\{a_{ij}\} = \{1 - \boldsymbol{\Omega}_d(h_i t_j)\}$, let p be an $m \times 1$ solution vector and let $\mathbf{W}$ be an $r \times r$ weight matrix. Then, equation (1) can be written as

$$\mathbf{Q}(\mathbf{p}) = (\hat{\gamma} - \mathbf{A}\mathbf{p})' \mathbf{W} (\hat{\gamma} - \mathbf{A}\mathbf{p}) \tag{2}$$

For convenience, $\mathbf{W}$ is assumed to be an identity matrix in this paper. The resulting non-parametric estimate (with solution $\tilde{\mathbf{p}}$) of the semi-variogram has an explicit representation as

$$\hat{\gamma}(h) = \sum_{j=1}^{m} [1 - \boldsymbol{\Omega}_d(h t_j)] \tilde{p}_j \tag{3}$$

Note that $\Sigma_{j=1}^{m} \tilde{p}_j$ is an estimate of the sill.

Shapiro and Botha (1991) did not compare their method with more traditional parametric methods. They did not discuss how to choose the nodes, choosing them in their examples in what they describe as an *ad hoc* manner. They did discuss how to choose appropriate weights based on the work of Cressie (1985, 1991), but noted that, for their examples, the weights made little difference.

3 Fitting non-parametric semi-variograms

Let $Z(s_i)$, $i = 1, \ldots, n$, be n observations from a spatial random process. Generally, estimation of the semi-variogram function starts with the semi-variogram cloud (Cressie, 1991); a plot of the $[n(n-2)]/2$ distinct values of

$$Y_{ij} = \frac{[Z(s_i) - Z(s_j)]^2}{2}$$

against h_{ij}. A preliminary smoothing of this cloud yields a sample semi-variogram.

The most widely used smoothing method is the classical method of moments estimator of Matheron (Cressie, 1991). This has the form

$$\hat{\gamma}(h) \equiv \frac{1}{2|N(h)|} \sum_{|N(h)|} [Z(s_i) - Z(s_j)]^2 \tag{4}$$

where s_i and s_j are in R^d, $N(h) \equiv \{(s_i, s_j): h + \| s_i - s_j \|; i, j = 1, \ldots, n\}$, and $|N(h)|$ is

the number of pairs in $N(h)$. If the data are irregularly spaced, then this sample semi-variogram estimator is modified (Cressie, 1991). Although $\hat{\gamma}(h)$ is unbiased, it is not resistant to outliers.

The topic of the best way to determine sample semi-variograms has been an active area of research in itself. In particular, Cressie (1991) argues strongly that the semi-variogram cloud and the classical method of moments sample semi-variogram are not appropriate, because of their sensitivity to outliers, and discusses alternatives. However, the semi-variogram cloud and method of moments estimator remain the methods of choice for the majority of practitioners. Most available software depends heavily on these (see, for example, Deutsch & Journel, 1992). All sample semi-variograms used in this paper were determined in this manner.

Gaussian random fields were simulated for 50 locations spaced one unit apart on a one-dimensional transect. This yielded a total of 1225 distinct pairs of

$$Y_{ij} = \frac{[Z(s_i) - Z(s_i)]^2}{2}$$

over 49 lags. It is typical to consider only lags with an adequate number of observations—usually 30 or more (Cressie, 1991). Thus, the number of lags considered for the comparisons was taken to be 20, and the resulting semi-variogram cloud is a scatterplot of 790 points of Y_{ij} versus 20 lags with $50 - h$ observations at lag h. The nugget effect was set to zero for all the simulations.

Evaluating the SB method using simulated one-dimensional data can be criticized, because most spatial data occur in two or three dimensions. However, the form of sample semi-variograms, i.e. the 'data' that are being used for the fitting, is not affected much by the spatial dimension in which the data were collected. At most, the points in the sample semi-variogram might not occur at equally spaced lags. Thus, the results of an evaluation based on one-dimensional data should be applicable to data collected from higher dimensions.

Shapiro and Botha (1991) motivated their discussion with the one-dimensional version of $\Omega_d(ht)$, although they briefly discussed the two- and three-dimensional versions. The three-dimensional version of the SB estimator was chosen for all non-parametric fits in this paper, i.e.

$$\Omega_3(ht) = \frac{\sin(ht)}{ht}$$

This version will yield non-parametric estimates that are guaranteed to be conditionally negative definite for spatial data from one, two or three dimensions.

A solution to equation (2) was found using the program NNLS described in Lawson and Hanson (1974). A solution requires that the nodes (i.e. the t_j terms) be chosen before minimization. The selection of a set of nodes involves the selection of a set of functions of the form $[1 - \Omega_d(ht_j)]$ that will be used to construct the estimated semi-variogram. The set of nodes selected will affect the fit. However, it is inefficient to try to customize the selection of nodes to a given fitting problem. Such a method of choosing nodes is also open to the criticism that one can achieve any fit desired. One of the strengths of the non-parametric method is that it has the potential to be less subjective than the parametric fitting procedures currently used. The choice of nodes should be made in some systematic, objective way that produces a collection of functions that is rich enough (in some sense) to capture the behavior of the sample semi-variogram. Sampson and Guttorp (1992) fitted

TABLE 1. The minimum, 25th percentile, median, 75th percentile and maximum of the residual norms from 100 fits, using four different selections of nodes

	200 nodes	500 nodes (first set)	1000 nodes	500 nodes (second set)
Minimum	0.60	0.60	0.60	0.60
25th percentile	1.79	1.78	1.78	1.79
Median	2.55	2.52	2.52	2.55
75th Percentile	4.17	4.15	4.14	4.14
Maximum	12.37	12.32	12.31	12.33

Note: The simulated data are from an exponential model with a sill and range of 10.

their non-parametric semi-variograms using an algorithm originally designed for use in estimating mixtures. A different approach is described in the following.

There is no mathematical reason to restrict the number of nodes to be less than the number of observations. One could simply saturate the node space by choosing a large number of nodes and letting the NNLS procedure, along witht data, pick the nodes that are important. The rest will be assigned jumps of zero.

There are three questions that need to be resolved with the saturation approach.

(1) How should the nodes be selected?
(2) How many nodes are enough?
(3) Does the saturation approach lead to overfitting?

These questions are addressed next. In each instance, the comments refer to fits obtained to sample semi-variograms determined from simulated realizations of an isotropic Gaussian random field with a true semi-variogram function that was exponential with a sill of 10 and range of 10. (The range is defined as the spatial lag h at which $\gamma(h) = 0.95C(0)$.)

3.1 Node selection and number

Initially, fits based on collections of several hundred nodes were evaluated. The nodes were all restricted to the interval [0, 20]. Achieving a good fit requires a suitable number of nodes near zero. These are associated with lower frequency curves that allow the behavior of the data at larger lags to be captured. However, the function

$$1 - \frac{\sin(ht)}{ht}$$

rises so quickly to an asymptote with increasing t that relatively few nodes larger than four are necessary to capture the behavior of the data at smaller lags.

Theoretically, there is no upper limit to the number of nodes. Practically, however, the number chosen should be kept small enough to make the fitting computationally feasible.

Table 1 shows the minimum, median, maximum and the interquartile range of the Euclidean norms of the residual vectors from fits to 100 simulated data sets.There were four different node selections. The selections were as follows:

(1) 200 nodes, with 100 equispaced in (0, 4] and 100 equispaced in [4.16, 20];
(2) 500 nodes, with 250 equispaced in (0, 4] and 250 equispaced in [4.064, 20];

(3) 1000 nodes, with 500 equispaced in (0, 4] and 500 equispaced in [4.032, 20];
(4) 500 modes equispaced in (0, 20].

Increasing the number of nodes did not increase the accuracy of the fit, but did increase computational time. It appears that the first selection of 200 nodes works well and this selection was used in all subsequent fits. Although results are not presented here, some experimental fitting with the two-dimensional version of the non-parametric estimator also worked well with that selection of nodes.

3.2 Saturation and overfitting

Table 1 provides evidence that increasing the number of nodes does not result in overfitting. One reason is that the non-negativity constraints on the parameter estimates have a smoothing effect. In effect, constrained least-squares fitting is a form of regularization analagous to penalized least-squares routines used in the solution of ill-posed inverse problems.

The main reason for the lack of overfitting is that the three-dimensional version of the SB estimator was chosen. This automatically imposes smoothness constraints. A comparison of $1-\Omega_d(x)$ for increasing values of d will show that valid non-parametric semi-variograms constructed from the higher-dimensional versions will be smoother than those constructed from the lower-dimensional versions. This heuristic argument for increasing smoothness with increasing d is made rigorous in Schoenberg (1938, p. 822).

Even if interpolation is not a problem, some may still feel that the estimated functions are not sufficiently smooth. Shapiro and Botha (1991) show how it is possible to impose monotonicity and convexity constraints on the fits, by imposing appropriate constraints on first and second derivatives. Also, choosing $d=\infty$ will result in a monotonic increasing fit. Sampson and Guttorp (1992) chose this version ($\Omega_\infty = \exp(-ht)^2$) for their fits. It is possible to obtain smoother sample semi-variograms by smoothing the semi-variogram cloud by splines, kernel regression or isotonic regression, for example. The SB method could be applied to these smoother sample semi-variograms to yield valid fits. The SB method can, in fact, be thought of as a filter that transforms any sample semi-variogram into a conditionally negative definite function. Cherry (1994) discusses a penalized fitting procedure within the context of non-parametric sill estimation that also has the benefit of imposing additional smoothness constraints on the estimated semi-variograms.

4 Comparison of non-parametric and parametric methods

The comparisons are based on fitting a non-parametric model and various parametric models to sample semi-variograms based on simulated realizations of five random fields, each with a true exponential semi-variogram function with a sill of 10 and ranges that varied from two to 18 in increments of four. The exponential model was chosen because it is easy to simulate and is a popular choice for parametric fitting. A total of 100 realizations were simulated with 50 locations spaced one unit apart on a one-dimensional transect for each of the five random fields. The nugget effect was set to zero for each model.

Four parametric models (exponential, Gaussian, rational quadratic and power) were fitted to the sample semi-variogram using a non-linear least-squares program

(UMSOLVE) in Matlab. Cressie (1985) recommends the use of weighted least-squares fits and Shapiro and Botha (1991) discuss how to incorporate Cressie's suggested weights into their method. A brief examination of results from both weighted least-squares fits and non-weighted least-squares fits revealed no appreciable differences for the parametric and non-parametric methods, and only unweighted fits were used for the comparisons presented below.

With the nuggest effect equal to zero, the models have the following forms:

- exponential: $\gamma(h) = s[1 - \exp(-3h/r)]$;
- Gaussian: $\gamma(h) = s[1 - \exp(-3h/r)^2]$;
- rational quadratic: $\gamma(h) = s\{h^2/[1 + (h^2/r)]\}$;
- power: $\gamma(h) = bh^{\lambda}$.

The parameters s and r correspond to the sill and range respectively. The power model has no sill or range.

The power model was included despite the fact that it cannot be the semi-variogram for a second-order stationary process, because it does not have a sill. However, the model can give good fits to data from a second-order stationary process in which all the Y_{ij} terms in the semi-variogram cloud occur at lags less than the range.

Initial starting values for the parameters in the parametric fits were the true values of the underlying exponential model, except for the power model, where starting values consistent with a straight line with a slope of 1 were chosen. The fits were compared by computing the integrated squared error (ISE) over the interval [0, 20], where

$$\mathrm{ISE} = \int_0^{20} [\hat{\gamma}(h) - \gamma(h)]^2 \, \mathrm{d}h$$

was evaluated using the trapezoid rule ($\hat{\gamma}(h)$ is the estimated semi-variogram and $\gamma(h)$ is the true semi-variogram function).

Figure 1 shows four box plots of the resulting ISEs. The figure only gives results for data from realizations of the four random fields with true ranges of 6–18. For the simulated data from a random field with a range of two, the rational quadratic and power models fit so poorly that their ISEs swamped the results from the other fits if they were included in the box plot. The points in the figure that lie above the upper whisker are outliers. Table 2 gives the minimum, maximum, and the 25th, 50th and 75th percentiles for the ISE values calculated from the simulations.

The NNLS procedure always converges (Lawson & Hanson, 1974), but the non-linear least-squares program occasionally failed to converge. This was a problem for the rational quadratic model when the simulated data came from the exponential model with a range of two (10 of the 100 attempted fits failed to converge), and for the exponential model when the simulated data came from the exponential model with a range of 18 (six of the 100 attempted fits failed to converge).

The failure of the rational quadratic model to give an adequate fit of the data from the simulation with a range of two is puzzling. It has the same general shape as the true underlying exponential model. However, the failure of the power model (Table 2) is not surprising at all. The exponential model with a sill of 10 and a range of two rises quickly to its asymptote. The power model simply cannot track this behavior.

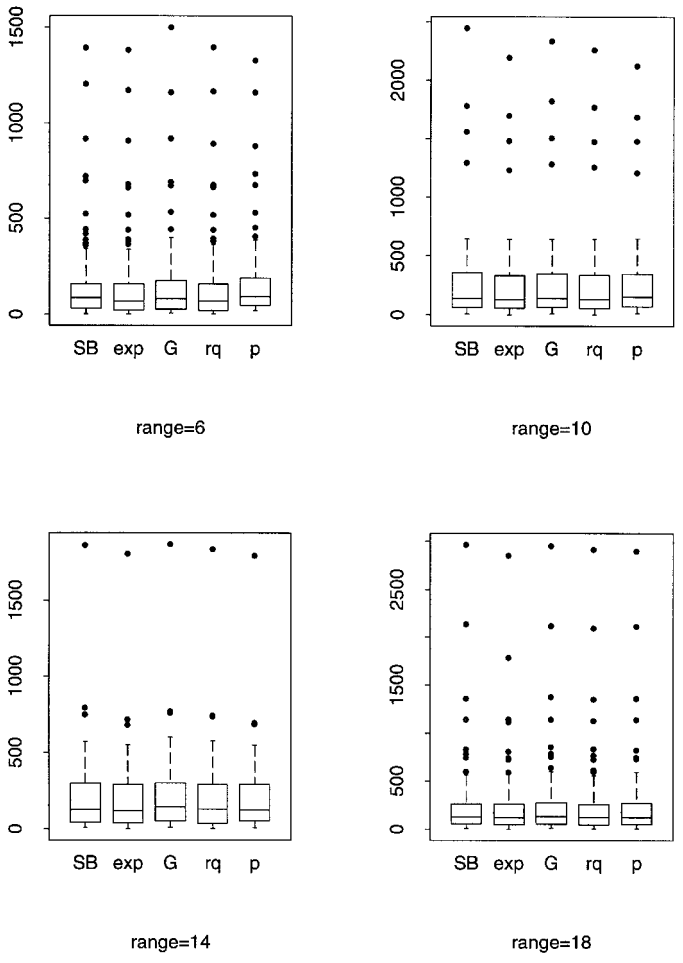

FIG. 1. Box plots of integrated squared errors from non-parametric (SB) and parametric models, fitted to simulated data sets based on exponential models with sills of 10 and the indicated ranges.

Data were also generated for two other isotropic random processes. One had a true semi-variogram function that was a mixture of a rational quadratic model and a hole-effect model valid for three-dimensional spatial data. With the nugget effect set to zero, the semi-variogram function is

$$\gamma(h) = 10\left[1 - \frac{sin(h)}{h}\right] + \frac{2h^2}{[1 + (h^2/2)]}$$

The sill for this semi-variogram is 14.

The minimum, 25th percentile, median, 75th percentile and maximum ISE values for the five comparisons are shown in Table 3. The non-parametric fits again compare favorably with the parametric fits. Data for the other process (a mixture of spherical semi-variogram models with a sill of four) are not presented, in the interest of space, but the SB method again performed well.

Also not presented here are results of simulations in which the one- and two-dimensional versions of Ω_d were used. The two-dimensional version produced

TABLE 2. The minimum, 25th percentile, median, 75th percentile and maximum ISE values from non-parametric and parametric semi-variogram models fitted to 100 simulated data sets

Range	Percentile	SB model	Exponential model	Gaussian model	Rational quadratic	Power model
2	Minimum	3.97	2.29	2.29	9.71	79797.97
	25th	25.84	23.66	23.66	534.16	226248.16
	Median	68.13	64.77	64.77	1857.10	356117.07
	75th	138.24	133.73	133.73	5477.22	536841.67
	Maximum	762.90	813.44	813.44	66672.13	2639251.15
6	Minimum	3.222	2.03	7.12	2.61	20.10
	25th	32.09	22.10	27.36	21.13	48.01
	Median	86.63	68.38	80.82	69.36	93.70
	75th	159.09	157.20	174.98	157.65	184.75
	Maximum	1392.48	1380.91	1500.23	1397.24	1327.76
10	Minimum	5.94	1.79	9.92	3.53	15.00
	25th	65.50	59.87	68.68	60.04	72.24
	Median	140.75	131.86	139.93	134.36	152.92
	75th	357.36	327.91	345.53	330.64	342.24
	Maximum	2441.96	2190.03	2331.54	2257.03	2121.19
14	Minimum	7.55	0.94	10.86	3.56	9.21
	25th	42.27	40.69	50.05	39.23	53.73
	Median	125.52	119.48	143.13	130.51	123.44
	75th	296.72	286.54	299.19	289.034	287.39
	Maximum	1859.67	1804.29	1866.11	1833.58	1791.90
18	Minimum	4.94	0.38	10.83	3.48	5.51
	25th	55.47	44.95	53.10	46.89	48.73
	Median	125.30	119.09	131.37	121.09	123.23
	75th	253.44	252.53	272.82	251.47	267.96
	Maximum	2962.94	2850.68	2947.62	2910.38	2895.87

Note: In each case, the true semi-variogram model was exponential with no nugget effect and a sill of 10. The ranges varied as indicated.

results comparable with those reported above. The one-dimensional version performed poorly. For example, the minimum, median and maximum ISE values from one-dimensional fits to simulated data from the process with an exponential semi-variogram with a sill and range of 10 were 147.54, 358.81 and 63842.17 respectively. The corresponding ISE values for the three-dimensional version were 5.94, 140.76 and 2441.96 (Table 2).

5 Examples

In this section, two examples of the use of the SB method will be presented. These are examples of the non-parametric estimation of semi-variogram functions and the non-parametric estimation of the sill for three actual data sets taken from the literature. The non-parametric and parametric estimates are compared.

5.1 Example 1

The first data set comes from Clark (1979). The data are in the form of silver concentrations sampled from an ore body. Only data from the first 75 sample

TABLE 3. The minimum, 25th percentile, median, 75th percentile and maximum ISE values from non-parametric and parametric semi-variogram models fitted to 100 simulated data sets

Percentile	SB model	Exponential model	Gaussian model	Rational quadratic	Power model
Minimum	3.69	39.06	17.63	30.42	105.20
25th	89.71	107.16	81.39	93.49	169.36
Median	273.34	235.61	201.06	204.49	305.20
75th	535.76	490.06	479.80	480.11	547.42
Maximum	2513.58	2729.25	2751.94	2740.22	2753.77

Note: The true semi-variogram model was a mixture of rational quadratic and hole effect models with no nugget effect and a sill of 14.

locations are used. This is the data set that Shapiro and Botha (1991) used to illustrate their method. Clark (1979) fitted a spherical semi-variogram to these data by eye. Her estimated semi-variogram took the form

$$\gamma(h)=\begin{cases}0, & h=0\\ 11[(3/2)(h/50)-(1/2)(h/50)^3], & 0<h\leqslant 50\\ 11 & h\geqslant 50\end{cases}$$

The non-parametric fit was achieved using the same collection of nodes as was used above with $d=3$. Figure 2 shows the actual data, Clark's fit and the non-parametric fit. Clark's sill estimate is 11. The sill estimate from the non-parametric fit is 11.33 (see Cherry (1994) for a discussion of the estimation of the sill).

5.2 Example 2

The second data set is also from Clark (1979). The data are logged nickel concentrations from an ore body. The experimental semi-variogram is shown in Fig. 3. Clark estimated a nugget effect of 0.40 and the data shown have been corrected to correspond to a nugget effect of zero.

Clark fitted a complicated mixture of spherical models to these data. Her model (absent of the nugget effect) has the form

$$\gamma(h)=\begin{cases}0, & h=0\\ 1.15[(3/2)(h/12)-(1/2)(h/12)^3]+[(3/2)(h/60)-(1/2)(h/60)^3], & 0<h\leqslant 12\\ 1.15+((3/2)(h/60)-(1/2)(h/60)^3), & 12<h\leqslant 60\\ 2.15, & h\geqslant 60\end{cases}$$

Figure 3 shows the actual data, Clark's fit and the non-parametric fit. Clark's sill estimate is 2.15. The sill estimate for the penalized non-parametric fit was 2.15. Once again, the parametric model was fitted by eye.

6 Conclusions

The results presented here show that the non-parametric method proposed by Shapiro and Botha fits as well as the parametric models currently used (based on the comparison of integrated squared errors). There is rarely a well-defined dividing line between objectivity and subjectivity, but the SB method generally seems to be more objective than the parametric methods. It is easier to implement, even if the parametric models are being fitted by inspection. The NNLS algorithm is fast and

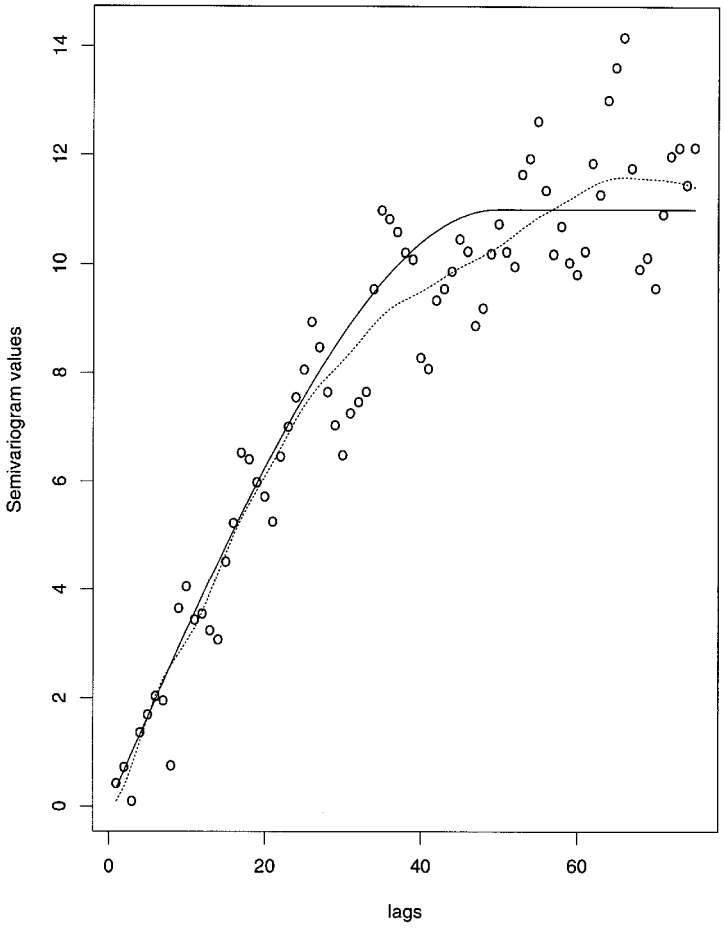

FIG. 2. Sample semi-variogram (0) and two semi-variograms fits to Clark's (1979) silver data: ——, spherical model with sill= 11; – – –, non-parametric fit with sill= 11.33.

always converges, and there are many similar algorithms available in the quadratic programming literature that would also work. The method is robust to the selection of nodes, provided that the selection results in a collection of functions of the form

$$1 - \Omega_d(ht)$$

that is sufficiently rich to capture the behavior of the data. For $d = 3$, it is important to have a good collection of nodes near the origin. As a general recommendation, the selection of 200 nodes used above worked well.

The observation that the SB method produced 'comparable' results could be criticised, in that it sometimes produced inferior results. However, if a method is simple and easy to apply to a wide class of problems, and produces results that are not generally inferior to other commonly accepted methods, then it is worthy of consideration. The SB method is easier, because parametric fitting of semi-variograms is rarely as straightforward as has been presented here. It is common for geostatisticians to have to fit mixtures of parametric models to achieve good fits. This is one reason why fitting by inspection is so common. Issaks and Srivastava (1989) present a detailed discussion of fitting semi-variograms, and non-linear

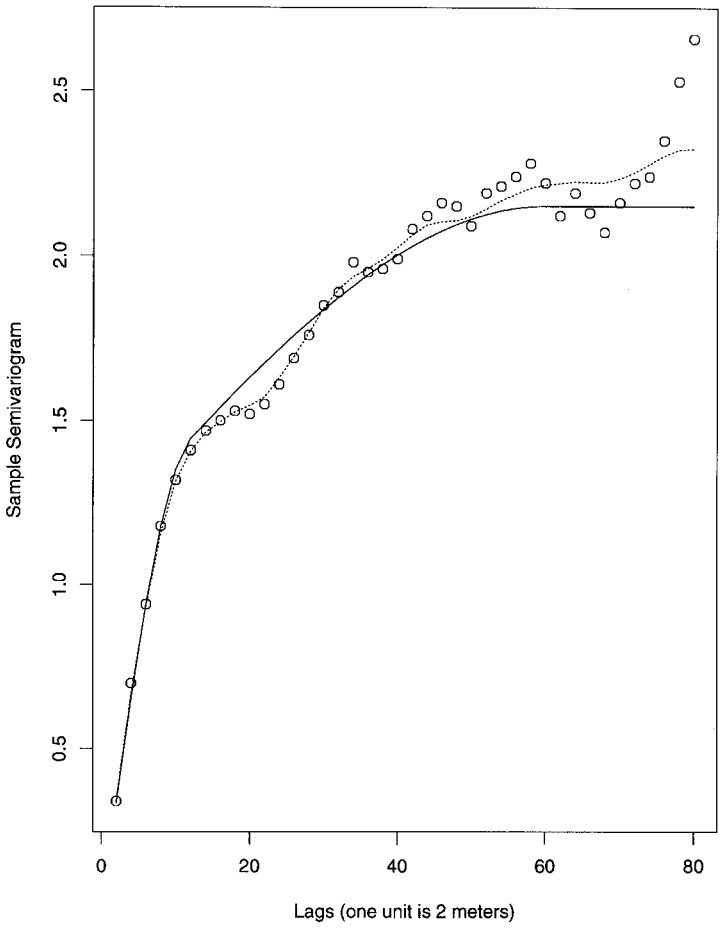

FIG. 3. Sample semi-variogram (0) and two semi-variogram fits to Clark's (1979) nickel data: ——, spherical model with sill= 2.15; – – –, penalized non-parametric model with sill= 2.15. One unit of lag is 2 m.

least-squares methods (and other parametric methods) are noticeably absent from their discussion. Fitting the semi-variogram models, they discuss using non-linear least-squares methods would be time-consuming and challenging.

The rather complicated model that Clark (1979) fitted in example 2 also illustrates this point (see Fig. 3). No single semi-variogram model is going to fit well enough by itself, and trying to fit such a mixture of models by non-linear least-squares methods would be difficult.

An anonymous reviewer expressed concern with the failure of the estimation procedure to deal with positive correlations between neighboring estimates of $\gamma(h)$. While these correlations are also routinely ignored when fitting parametric models, the non-parametric method might be more likely to overfit spurious structure in the sample semi-variogram. This is a valid criticism. The extent of the problem is unknown, but there should be less of a difference when the parametric model that is fitted is a complicated mixture. This is another argument for using higher-dimensional versions of Ω_d. Other posible methods of dealing with the problem are to estimate the correlation structure and take it into account in the weight matrix

W; and to use other methods of estimating the sample semi-variogram. In addition to being less sensitive to outliers, the robust estimator of Cressie (Cressie, 1991, pp. 74-76) is less sensitive to correlations among the estimates of $\gamma(h)$.

If one is also interested in a non-parametric estimate of the sill, then it may be necessary to implement the penalized fitting procedure described in Cherry (1994). However, the method is still faster and easier to use than non-linear least-squares methods. This penalized fitting procedure has the added advantage of imposing additional smoothness constraints. The details have been omitted here, owing to space considerations.

The most obvious disadvantage to the SB method as implemented here is the lack of a nugget effect. The estimated semi-variograms necessarily pass through the origin. Simply estimating the nugget effect by considering it as an additional parameter will not work, because, with no data at the origin, the NNLS algorithm will assign that parameter a value of zero. One possibility is to find n_0 and p_j, for $j = 1, \ldots, m$, to minimize

$$\mathbf{Q}(\mathbf{p} = \sum_{i=1}^{r} w_i \left\{ \hat{\gamma}(h_i - n_0 - \sum_{j=1}^{m} [1 - \Omega_d(h_i t_j) p_j] \right\}^2$$

with n_0 (the estimated nugget effect) and the p_j terms constrained to be non-negative. The minimization could be carried out by first fixing n_0 with an initial guess and solving for the p_j terms using the NNLS algorithm, and then fixing the p_j terms and solving for n_0. This procedure could be continued in an iterative manner until convergence—which is guaranteed, because each step is a minimization and $\mathbf{Q}(\mathbf{p})$ is bounded below by zero. This is a topic for further work.

The non-parametric method of Shapiro and Botha (1991) has been implemented as a FORTRAN program and easy-to-use extended S functions (Becker *et al.*, 1988). The package is available as a SHAR archive in STATLIB under the name npvar.sh.

Acknowledgements

This research was supported in part by the Office of Naval Research under Contract N-00014-89-J-1114. SC completed part of the work while a postdoctoral student with the Geophysical Statistics Project (NSF Grant DMS-9312686) at the National Center for Atmospheric Research in Boulder, CO, USA.

REFERENCES

BECKER, R. A., CHAMBERS, J. M. & WILKS, A. R. (1988) *The New S Language* (Pacific Grove, CA, Wadsworth and Brooks).

CHERRY, S. (1994) Nonparametric estimation of semivariogram functions, Unpublished PhD Dissertation, Montana State University, Bozeman, MT.

CLARK, I. (1979) *Practical Geostatistics* (London, Elsevier).

CRESSIE, N. A. (1985) Fitting variogram models by weighted least squares, *Journal of the International Association for Mathematical Geology*, 17, pp. 693–702.

CRESSIE, N. A. (1991) *Statistics for Spatial Data* (New York, Wiley).

DEUTSCH, C. V. & JOURNEL, A. G. (1992) *GSLIB: Geostatistical Software Library and User's Guide* (New York, Oxford University Press).

ENGLUND, E. J. (1990) A variance of geostatisticians, *Mathematical Geology*, 22, pp. 417–455.

HALL, P., FISHER, N. I. & HOFFMAN, B. (1994) On the nonparametric estimation of covariance functions, *The Annals of Statistics*, 22, pp. 2115–2134.

ISSAKS, E. H. & SRIVASTAVA, R. M. (1989) *An Introduction to Applied Geostatistics* (New York, Oxford University Press).

LAWSON, C. L. & HANSON, R. J. (1974) *Solving Least Squares Problems* (Englewood Cliffs, NJ, Prentice-Hall).

LELE, S. (1995) Inner product matrices, kriging, and nonparametric estimation of the variogram, *Mathematical Geology*, 27, pp. 673–692.

LELE, S. (1995) Personal communication.

SAMPSON, P. D. & GUTTORP, P. (1992) Nonparametric estimation of nonstationary spatial covariance structure, *Journal of the American Statistical Association*, 87, pp. 108–119.

SCHOENBERG, I. J. (1938) Metric spaces and completely monotone functions, *Annals of Mathematics*, 39, 811-841.

SHAPIRO, A. & BOTHA, J. D. (1991) Variogram fitting with a general class of conditionally nonnegative definite functions, *Computational Statistics and Data Analysis*, 11, pp. 87-96.