Math Geosci (2011) 43: 243–259
DOI 10.1007/s11004-010-9314-5

Robust Resampling Confidence Intervals for Empirical Variograms

Robert Graham Clark · Samuel Allingham

Received: 22 February 2010 / Accepted: 4 August 2010 / Published online: 3 December 2010

Abstract The variogram function is an important measure of the spatial dependencies of a geostatistical or other spatial dataset. It plays a central role in kriging, designing spatial studies, and in understanding the spatial properties of geological and environmental phenomena. It is therefore important to understand the variability attached to estimates of the variogram. Existing methods for constructing confidence intervals around the empirical variogram either rely on strong assumptions, such as normality or known variogram function, or are based on resampling blocks and subject to edge effect biases. This paper proposes two new procedures for addressing these concerns: a quasi-block-bootstrap and a quasi-block-jackknife. The new methods are based on transforming the data to decorrelate it based on a fitted variogram model, resampling blocks from the decorrelated data, and then recorrelating. The coverage properties of the new confidence intervals are compared by simulation to a number of existing resampling-based intervals. The proposed quasi-block-jackknife confidence interval is found to have the best properties of all of the methods considered across a range of scenarios, including normally and lognormally distributed data and misspecification of the variogram function used to decorrelate the data.

Keywords Spatial analysis · Variograms · Bootstrap · Jackknife · Block bootstrap · Block jackknife

R.G. Clark (✉)
Centre for Statistical and Survey Methodology, University of Wollongong, Wollongong, NSW 2522, Australia
e-mail: Robert_Clark@uow.edu.au

S. Allingham
Centre for Health Services Development, University of Wollongong, Wollongong, NSW 2522, Australia
e-mail: Sam_Allingham@uow.edu.au

1 Introduction

Spatial datasets consist of one or more variables measured at a number of locations defined by a coordinate system. One of the aims of many spatial analyses is to understand how the relationship between the variables measured at two points depends on their positions. Typically nearby points tend to be similar while distant points are less so, but the way in which the relationship depends on distance can be very different in different examples, and a good understanding of this is crucial in the design and analysis of spatial studies.

The sample variogram is a widely used tool for this purpose (Cressie and Hawkins 1980). Variograms can be generalized to the cross-variogram for multivariate data (Chiles and Delfiner 1999), but we will consider situations where a single variable Z is measured on a set of points defined in two dimensions. We write $Z(s)$ for the value of Z at a point s. First and second order stationarity is assumed; that is, the first two moments of $Z(s)$ are assumed to not depend on s. It will also be assumed that $E[Z(s)] = 0$; in practice, this would be achieved approximately by subtracting a grand mean or the fitted value from a regression model from each observation. Let h be the distance between two points s_1 and s_2, defined using the Cartesian or some other metric. The variogram

$$\gamma(h) = \mathrm{Var}\big[Z(s_1) - Z(s_2)\big] = E\big[Z(s_1) - Z(s_2)\big]^2 \tag{1}$$

is assumed to depend only on h and not on s_1 and s_2. Anisotropic models, in which the variogram also depends on the direction of the vector $s_1 - s_2$, are sometimes used but will not be considered here. The covariogram and correlogram are alternative, related representations of spatial dependency; however, the variogram has the advantage that it may exist in situations where covariances and variances do not (Cressie 1993).

The value of $\gamma(0)$ is equal to 0. The value of $C_0 = \lim_{h \downarrow 0} \gamma(h)$ (called the nugget parameter) may be non-zero, due to measurement error or discontinuities in the data (e.g., nuggets or seams in geological data). The variogram generally increases as h increases, with more distant points becoming less correlated. The least upper bound of the variogram is called the sill and will be denoted as $C_0 + C$. In some models, the variogram will be equal to the sill for $h \geq a$, in which case a is referred to as the range of influence. In other cases, the variogram will approach its sill asymptotically, in which case the practical range (or effective range) is defined to be the distance at which the variogram reaches a value of 0.95 times the sill (Oliver and Webster 2001). The marginal variance of Z (if it exists) is equal to $(C_0 + C)/2$.

The integral range is another measure of the strength of spatial correlations over a region V with volume $|V|$. Let $\rho(h)$ be the correlation between values at points separated by distance h. The integral range is defined by $A = \int \rho(h)\, dh$ (Chiles and Delfiner 1999). If A is large compared to $|V|$, this means that spatial correlations are very pervasive. To see this, let $\bar{Z}_V$ be the mean of the variable Z over all points in the region. Then the variance of $\bar{Z}_V$ is equal to the variance that would be obtained by an uncorrelated sample of M observations from V, where $M = |V|/A$ (Chiles and Delfiner 1999). Thus, full observation of the spatial field V is in some sense equivalent to M independent observations, at least for the purpose of estimating the

mean of Z. The quantity M can be considered to be the effective sample size of the region V.

The variogram is commonly estimated by the binned empirical variogram. Suppose that distances h are partitioned into bins (usually non-overlapping) B_k, $k = 1, \ldots, K$, with midpoints h_k. Let $h_{ij} = \|s_i - s_j\|$ be the distance between points i and j, and let P_k contain all pairs of points such that $h_{ij} \in B_k$. The binned empirical variogram at h_k is defined as

$$\hat{\gamma}(h_k) = \frac{1}{|P_k|} \sum_{(ij)\in P_k} \big(Z(s_i) - Z(s_j)\big)^2 \tag{2}$$

(Matheron 1963). If the bins B_k are of width 0 (i.e., if P_k only contains points exactly h_k apart), then (2) is exactly unbiased for the true variogram. There is a bias when bins are of width greater than 0 (Emery 2007), which can be minimized by making the bins narrow enough so that the variogram does not change substantially over $h \in B_k$. The binned empirical variogram can be sensitive to outliers and outlier-robust versions have been proposed; Cressie and Hawkins (1980), Emery (2007), Cressie (1993), Dowd (1984), Lark (2000) evaluated a number of these and recommended the method of Dowd (1984).

The variogram is usually modelled as a smooth curve with several parameters to be estimated. One benefit of modelling is that most commonly used models guarantee that the modelled variance–covariance matrix of any set of points is positive semi-definite. This is necessary for most kriging methods, and also allows for simulation of data from the model. Four commonly used variogram models are

$$\gamma(h) = C_0 + C\big(1 - e^{-h/a}\big), \quad \text{(the exponential model)} \tag{3}$$

$$\gamma(h) = C_0 + C\big(1 - e^{-h^2/a^2}\big), \quad \text{(the Gaussian model)} \tag{4}$$

$$\gamma(h) = \left\{ \begin{array}{ll} C_0 + C(7(h/a)^2 - 8.75(h/a)^3 + 3.5(h/a)^5 - 0.75(h/a)^7) & \\ \quad \text{if } h \le a & \\ C_0 + C \quad \text{if } h > a & \end{array} \right\}, \tag{5}$$

(the cubic model)

$$\gamma(h) = \left\{ \begin{array}{ll} C_0 + C(1.5h/a - 0.5(h/a)^3) & \text{if } h \le a \\ C_0 + C & \text{if } h > a \end{array} \right\}, \tag{6}$$

(the spherical model)

where h is greater than 0 and a is a parameter controlling the range. The first two variogram models approach their sills asymptotically, with the Gaussian model appearing more parabolic for short distances than the exponential model. The cubic and spherical variogram models both reach their sills at distance a; the cubic is less smooth than the spherical variogram.

Estimated variograms are important in the prediction of unknown points (referred to as kriging, Chiles and Delfiner 1999). The variogram is also often of scientific interest in its own right. For example, Wang et al. (2009) used empirical variograms to compare the spatial structure of stable carbon isotope and organic carbon content

at two sites in southern Africa. Cosh et al. (2003) compared empirical variograms representing different time points to observe the effect of dry periods on vegetation density. Shafer and Varljen (1990) noted that it is useful to compare empirical variograms over time to determine if a spatially correlated variable is time invariant. They also proposed that, when conducting a repeated spatial survey, variograms from previous and current iterations should be compared to identify significant changes in the spatial structure of the variable of interest and ensure that the survey design remains optimal. To make such comparisons meaningful, the uncertainty surrounding the empirical variogram needs to be taken into consideration.

It is therefore of interest to calculate confidence intervals for empirical variograms. Standard jackknife and bootstrap confidence intervals assume that data are independent. Section 2 reviews a number of resampling techniques that have been proposed for spatial data and variograms. One of these approaches is the quasi-bootstrap, where the data is transformed to create an approximately independent vector of observations, using the estimated covariance matrix of the observed data. This approach works well provided that the covariance matrix can be accurately estimated, but this is usually based on a model for the variogram or correlogram, and results can be sensitive to misspecification of this model. Another approach is to divide the region into blocks and resample these blocks. This works well provided data from different blocks is approximately uncorrelated, but this will not be the case unless the effective range is less than the block size. Section 3 proposes two new methods for computing confidence intervals for empirical variograms, designed to overcome these shortfalls by combining aspects of the quasi-bootstrap and block resampling approaches. The use of Box–Cox transformations to deal with non-normal data is also discussed. Section 4 summarizes a simulation study comparing the new approaches to existing methods. Data was initially simulated from a normal distribution with an exponential variogram. To assess robustness to modelling assumptions, datasets were also generated from a lognormal distribution and a Gaussian variogram. Section 5 contains conclusions.

2 Review of Resampling-Based Confidence Intervals for Variograms

In resampling methods for variance estimation or confidence intervals, data is resampled from the observed data or from a model estimated using the observed data, and the statistic of interest (e.g., the empirical variogram for a given bin) is calculated for each of these replicate datasets. The variance over the replications of the statistic is used to estimate the variance of the statistic. Commonly used replication techniques include the following: the nonparametric bootstrap, where observations are selected by simple random sampling with replacement from the original data; the parametric bootstrap where replicate datasets are generated from a model fitted to the population; and the jackknife, where one observation or a group of observations is dropped from each replicate (Davison and Hinkley 1997). Confidence intervals can be constructed by assuming that the statistic is normally distributed. Alternatively, if a bootstrap method is used, confidence intervals can be constructed using the percentiles of the replicate statistics.

Standard nonparametric jackknife and bootstrap methods assume that data is independent and identically distributed. Sebastyén (2004) applied the standard bootstrap procedure in conjunction with the percentile method to construct confidence intervals for variograms. This approach is flawed because spatial correlations are not allowed for; therefore, these confidence intervals would perform poorly unless these correlations are negligible (Tang et al. 2006).

Parametric bootstraps have been applied to time series data (Davison and Hinkley 1997), but less so to spatial data. The parametric bootstrap handles correlated data naturally, provided the multivariate distribution of the data can be correctly specified and estimated (Cressie 1993). Lahiri (2003) described a range of block resampling procedures for spatial and other dependent data. One example of this is a method of bootstrapping blocks or tiles in two or more dimensions, also discussed by Hall (1985). The approach is to partition up the space into non-overlapping blocks. Replicates are then constructed by replacing each block's data with the data from another block selected at random from all blocks. Let T_i refer to block i, where $i = 1, \ldots, m$, $\mathbf{Y}_i$ refer to the vector of variables observed from block i, and $\hat{\theta}(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)$ refer to the statistic of interest calculated using the observed data from all blocks. Let $\mathbf{Y}_i^*$ refer to the bootstrapped observations for block i, equal to $\mathbf{Y}_{j(i)}$ where $j(i)$ is selected at random from $\{1, \ldots, m\}$. The bootstrap replicate is then given by $\hat{\theta}^* = \hat{\theta}(\mathbf{Y}_1^*, \ldots, \mathbf{Y}_m^*)$. The bias of the bootstrap variance estimator was shown by Hall (1985) to be asymptotically negligible under two assumptions:

(i) The statistic of interest should be expressible as $\hat{\theta}(\mathbf{Y}_1, \ldots, \mathbf{Y}_m) = m^{-1} \sum_{i=1}^{m} \hat{\theta}(\mathbf{Y}_i)$, where $\hat{\theta}(\mathbf{Y}_i)$ is an estimator of the parameter of interest calculated using only the data in block i.
(ii) The covariances between $\hat{\theta}(\mathbf{Y}_i)$ and $\hat{\theta}(\mathbf{Y}_j)$ are asymptotically negligible when blocks i and j are adjacent and are zero otherwise.

The first assumption would not be satisfied for many nonlinear statistics, including empirical variograms, although it may be approximately satisfied provided that blocks are large. In the case of estimating variograms, both assumptions probably require that the blocks are large relative to the lags of interest as well as relative to the effective range.

Kunsch (1989) specified a block jackknife method for spatial statistics. The region was divided into B blocks, as in Hall (1985). B replicate datasets $\{\mathbf{Z}_b^* : b = 1, \ldots, B\}$ were created by dropping one block b at a time. Applying this method to empirical variograms, we would let $\hat{\gamma}_{k(b)}$ be the estimate for bin k calculated from replicate b. The jackknife variance estimator is then

$$\hat{\text{var}}\big[\hat{\gamma}(h_k)\big] = \frac{B-1}{B} \sum_{b=1}^{B} \big\{\hat{\gamma}_b(h_k) - \bar{\hat{\gamma}}_{\cdot}(h_k)\big\}^2. \tag{7}$$

Confidence intervals can be calculated using the normal approximation for either $\hat{\gamma}(h_k)$ or $\log(\hat{\gamma}(h_k))$.

The block bootstrap just described creates edge effects where the correlations between adjacent points which lie on the edges of two different blocks is much less in

the bootstrap samples than in the original data. The effect is negligible as the block size tends to infinity, but it can be important in practice. The same is true of the block jackknife, although the effects are probably less severe as only one block is dropped at a time with the remaining blocks staying in position. Different variants of block resampling have been proposed to better deal with edge effects. These include so-called post-blackening of time series data, where the residuals of each observation given the preceding observation are resampled, rather than resampling the observations themselves (Davison and Hinkley 1997).

Solow (1985) developed a quasi-bootstrap approach based on transformation for one-dimensional time series data. A second order stationary process $Z(t)$ was assumed where t is time. The first step was to estimate a variance–covariance matrix $\mathbf{V}$ of the data $\mathbf{Z}$ by fitting an exponential variogram model. A Cholesky decomposition $\hat{\mathbf{V}} = \hat{\mathbf{L}}\hat{\mathbf{L}}^{\mathrm{T}}$ was then calculated where $\hat{\mathbf{L}}$ is a lower triangular matrix. The transformed data $\mathbf{U} = \hat{\mathbf{L}}^{-1}\mathbf{Z}$ should then be approximately uncorrelated with unit variances. Bootstrap samples $\mathbf{U}^*$ were selected by simple random sampling with replacement from $\mathbf{U}$, and were then recorrelated using $\mathbf{Z}^* = \hat{\mathbf{L}}\mathbf{U}^*$. Statistics of interest were calculated from the so-called quasi-bootstrap sample $\mathbf{Z}^*$ and the process repeated many times. The variance over the replicate empirical variograms was then used to estimate the variance of the statistics of interest. The approach was used to estimate variances and calculate confidence intervals for functions of the one-dimensional variogram, and worked reasonably well in a small simulation study. Solow (1985) commented that robustness to misspecification of the variogram model was an open question. The method could similarly be applied to stationary two-dimensional data by using a fitted model for the variogram to obtain $\hat{\mathbf{V}}$.

Tang et al. (2006) modified the approach of Solow (1985) and extended it to two-dimensional data. The values of $\mathbf{U}^*$ were generated from independent standard normal distributions rather than by resampling from $\mathbf{U}$. These were then recorrelated using $\mathbf{Z}^* = \hat{\mathbf{L}}\mathbf{U}^*$ and the method then proceeded as in Solow (1985). A simulation study found that it gave appropriate coverage levels for confidence intervals for a grand mean parameter, even when the variogram model was misspecified. The method would be expected to depend to some extent on the data being normally distributed, and this was not evaluated. Shafer and Varljen (1990) proposed a somewhat ad hoc resampling method for statistics based on spatial data. Resamples were constructed by firstly selecting $n/3$ points by simple random sampling without replacement from all n points. For each of these points, three associated points were selected, each of which were random in orientation and separation distance up to the correlation length. The method was trialled on a dataset of groundwater nitrogen/nitrate concentrations, but has not been validated by a simulation study. Gascuel-Odoux and Boivin (1994), Huisman et al. (2003), and Wingle and Poeter (1993) also made use of this approach.

A number of methods for estimating variogram uncertainty were evaluated in a simulation study in Marchant and Lark (2004). Several theoretical expressions for the variances of the empirical variogram were compared to simulation estimates of the variances with both being based on an assumption of normally distributed data.

3 Two Quasi-Block-Resampling Methods

3.1 Quasi-Block-Jackknife

The first step in both methods is to calculate the binned empirical variogram using (2). A variogram model (e.g., the exponential model (3)) is then fitted to the data. This can be done by maximum likelihood, but this requires a full specification of the joint distribution of the spatial field. Instead, we suggest using a least squares fit to the binned empirical variogram to estimate the parameters C_0, C, and a. The Newton–Raphson algorithm is the usual method for nonlinear least squares; however, we found that this often failed to converge in the simulation study in the next section. A grid search procedure can be used instead, where a grid over possible values for the variance, nugget, and scale parameters is constructed. A shortfall of the grid search is that the grid must be reasonably coarse to make the computation feasible, resulting in less precise parameter estimates. This would be a problem if the variogram model fit was of interest in its own right, but we will be using it only as a tool to construct confidence intervals, so a grid search may be adequate. In our simulation study, we used a grid of 80 possible values for each of the three parameters, and this seemed to be adequate (for details, see Sect. 4.1).

The variance–covariance matrix $\hat{\mathbf{V}}$ of the data $\mathbf{Z}$ is then calculated according to this fitted model, and a Cholesky decomposition $\hat{\mathbf{V}} = \hat{\mathbf{L}}\hat{\mathbf{L}}^{\mathrm{T}}$ is calculated. The data is then transformed using $\mathbf{U} = \hat{\mathbf{L}}^{-1}\mathbf{Z}$. So far, this is identical to the quasi-bootstrap method. However, the quasi-bootstrap method used a simple bootstrap which assumed that the elements of $\mathbf{U}$ are independent. If the variogram model is misspecified, this may not be justified, so a block jackknife can be used instead. This relies on the correlations between blocks having a negligible effect, which is a weaker assumption. The method proceeds as follows:

(i) Divide the region into B blocks. Block jackknife replicates $\mathbf{U}_b^*$ are constructed by dropping all of the observations in $\mathbf{U}$ from block b, for $b = 1, \ldots, B$.
(ii) Let $\hat{\mathbf{V}}_b$ be the submatrix of $\hat{\mathbf{V}}$ obtained by dropping the covariances involving one or two points in block b. Let $\hat{\mathbf{V}}_b = \hat{\mathbf{L}}_b\hat{\mathbf{L}}_b^{\mathrm{T}}$ be its Cholesky decomposition. Recorrelate $\mathbf{U}_b^*$ by calculating $\mathbf{Z}_b^* = \hat{\mathbf{L}}_b\mathbf{U}_b^*$.
(iii) Calculate the empirical variogram for $\mathbf{Z}_b^*$; let be $\hat{\gamma}_{k(b)}$ the empirical variogram for bin k.
(iv) Estimate $\mathrm{var}[\hat{\gamma}_k]$ using the usual jackknife approach

$$\hat{\mathrm{var}}[\hat{\gamma}_k] = \frac{B-1}{B}\sum_{b=1}^{B}\{\hat{\gamma}_{k(b)} - \bar{\hat{\gamma}}_{k(.)}\}^2.$$

Confidence intervals can be calculated using a normal approximation. Alternatively, $\mathrm{var}[\log(\hat{\gamma}_k)]$ can be estimated and confidence intervals can be constructed based on $\hat{\mathrm{var}}[\log(\hat{\gamma}_k)]$.

A disadvantage of the block-jackknife approach is that the data from each block are not independent, leading to biased variance estimators. This is sometimes called an edge effect since the dependencies are greatest between observations on either side

of the edge between two blocks. The approach described here reduces this problem because the blocks are dropped from the transformed data $\mathbf{U}$ rather than from the original data $\mathbf{Z}$. The values in $\mathbf{U}$ from different blocks will be much closer to being independent because if the assumed variogram model is correct, then every observation in $\mathbf{U}$ will be approximately independent.

3.2 Quasi-Block-Bootstrap

Again, the first step is to fit a variogram model, calculate the Cholesky decomposition, and transform the data using $\mathbf{U} = \hat{\mathbf{L}}^{-1}\mathbf{Z}$. The block bootstrap is then calculated as follows:

(i) Divide the region into B blocks. The observations in each block are then replaced by the observations from a donor block selected at random from all other blocks. This gives a block bootstrap replicate $\mathbf{U}_j^*$, for $j = 1, \ldots, J$ where J is the number of replicates.
(ii) Recorrelate $\mathbf{U}_j^*$ by calculating $\mathbf{Z}_j^* = \hat{\mathbf{L}}\mathbf{U}_j^*$.
(iii) Calculate the empirical variogram for $\mathbf{Z}_j^*$; let be $\hat{\gamma}_{k(j)}$ be the empirical variogram for bin k and replicate j.
(iv) Estimate $\mathrm{var}[\hat{\gamma}_k]$ using the variance over the J bootstrap replicates

$$\hat{\mathrm{var}}[\hat{\gamma}_k] = (J-1)^{-1} \sum_{j=1}^{J} \left\{ \hat{\gamma}_{k(j)} - \bar{\hat{\gamma}}_{k(.)} \right\}^2.$$

The variance of $\log(\hat{\gamma}_k)$ can be similarly estimated. Confidence intervals can be calculated using a normal approximation for $\hat{\gamma}_k$ or $\log(\hat{\gamma}_k)$, or using a percentile method.

As for the quasi-block-jackknife, this method should be less sensitive to misspecification of the variogram model than the quasi-bootstrap, and less subject to edge effects than the block bootstrap.

3.3 Using a Box–Cox Transformation to Deal with Non-normal Data

The methods described in Sects. 3.1 and 3.2 did not require any assumption that data is normally distributed. However, the empirical variogram is sensitive to outliers, and this could result in poor confidence intervals if the data is very skewed or heavy tailed. A Box–Cox transformation of the observed data could be used to make the observations more normally distributed. This would mean replacing each observation Z by $g(Z)$ where g belongs to the Box–Cox family

$$g(z_i) = \begin{cases} \frac{z^{\lambda}-1}{\lambda(GM(z))^{\lambda-1}} & \text{if } \lambda \neq 0, \\ GM(z)\log(z_i) & \text{if } \lambda = 0, \end{cases}$$

where $GM(z) = (z_1 \cdots z_n)^{1/n}$ is the geometric mean of the data, and λ is a parameter to be determined. The method can only be used for data which is strictly positive. The value of λ is typically estimated by maximum likelihood, under the assumption that

there exists a λ such that $g(Z)$ is normally distributed. The original data Z would then be transformed to $g(Z)$ using the estimated λ, and the methods of Sects. 3.1 and 3.2 would be applied to the transformed data, to give the empirical binned variogram and confidence intervals for $g(Z)$.

Another approach that could be used for non-normal data is to transform using the empirical distribution function, or a smoothed estimate of the distribution function.

4 Simulation Study

4.1 Methodology

A simulation study was conducted in the R statistical environment (R Development Core Team 2007). For each model considered, 500 datasets of 2500 correlated normally distributed values were generated on a 50 by 50 regular square grid. Distances between distinct points ranged from 1 to 69.3. Variables on this spatial field were generated as follows:

(i) Normally distributed values were generated using the exponential variogram model (3) and the Gaussian variogram model (4). In each case, the nugget parameter C_0 was set to 0.5, and C was set to 2. This implies that the variables have a marginal variance of 1.25. The range parameter, a, was chosen so that the effective range was 0 (i.e., independent data), 2, 5, 10 and 15. The data were generated using the GaussRF function in the RandomFields package (Schlather 2006), using the direct matrix decomposition method.

(ii) In addition, lognormally distributed spatially correlated variables were generated by exponentiating the variables from (i) at each point, after multiplying by $\sqrt{0.25}/\sqrt{1.25}$. This meant that these variables had marginal $LN(0, 0.25)$ distributions. This is a moderately right-skewed distribution, with median 1, mean 1.13, and skewness 1.75.

We found that when the effective range was equal to 15, all of the confidence intervals considered performed poorly in some respect. When this parameter was much greater than 15, none of the confidence intervals were adequate. This is perhaps not surprising. When the effective range is 15, the integral range of the exponential variogram model is equal to $A = 2\pi 15^2/9 = 157.1$ (Garrigues et al. 2006, Table 3), so that the effective sample size is given by the area of the region divided by A, giving an effective sample size of 2500/157.1, or just 15.9. In other words, the complete region, even if fully observed, would in a sense contain information equivalent to just 16 independent observations of Z.

The binned empirical variogram (2) was calculated for each simulation and variable. The first three bins were defined to be of width zero and contain the values 1, $\sqrt{2}$, and 2. This was done because the data are on a regular grid with unit spacing, so that there were many pairs of points whose distances are exactly equal to these values, but no pairs whose distances lie between these values. Subsequent bins were intervals of width 1 up to a distance of 10, and then width 2 up to a distance of 40. Thus the midpoints of the bins were $\{1, \sqrt{2}, 2, 2.5, 3.5, \ldots, 9.5, 11, 13, 15, \ldots, 39\}$.

Variance estimates for the log of the binned empirical variogram at each midpoint were calculated using five methods:

(i) A block jackknife. The 50 by 50 grid was partitioned into 25 square blocks each of size 10 by 10. For each simulated dataset, 25 jackknife variograms were estimated by dropping one square at a time. The variance of $\log(\hat{\gamma}(h_k))$ was estimated as described in Sect. 2.

(ii) A block bootstrap. The 50 by 50 grid was partitioned into 25 square blocks each of size 10 by 10. The variance of $\log(\hat{\gamma}(h_k))$ was estimated as described in Sect. 1.

(iii) A quasi-bootstrap, using the method of Solow (1985) described in Sect. 1. The variance $\mathbf{V}$ was estimated by fitting an exponential variogram model. The parameters of the variogram model were estimated using a grid of 80 possible values for each of the sill $(C + C_0)$, nugget (C_0) and scale parameters (a). The grid values for the variance ranged from 0 to 3 times the maximum value of the empirical variogram; the values for the nugget ranged from 0 to 3 times the value of the empirical variogram for the first non-zero bin; and the values of the scale ranged from 0.01 times the mean of the bin midpoints to 2 times the largest bin midpoint. In each case, the grid was spaced more finely for the first half of the range for each parameter, as this was considered more likely to contain the optimal parameter values. The parameter values which minimised the sum of the squared differences between the modelled and empirical binned variogram were selected.

(iv) A quasi-block-jackknife, as described in Sect. 3.1, using the same fitted exponential variogram model.

(v) A quasi-block-bootstrap, as described in Sect. 3.2, using the same fitted exponential variogram model.

The bootstrap methods were all calculated using 100 replicates. 90% confidence intervals for the log of the variogram were calculated using the normal approximation, and then exponentiated to give confidence intervals for the variogram.

This method for confidence interval calculation was used for the normally distributed and lognormally distributed variables. For the latter variables, a Box–Cox transformation approach was also evaluated. The boxcox function in the MASS package in the R statistical environment (R Development Core Team 2007) was used with default settings to maximize the likelihood over $\lambda \in \{-2, -1.9, -1.8, \ldots, 2\}$ assuming independent data. (The assumption of independence is not satisfied for spatial data, but the resulting estimate of λ is still thought to be reliable enough for our purposes.) The original data Z was transformed to $g(Z)$ using the estimated λ, and the above five methods were used to give confidence intervals for the variogram of $g(Z)$.

The non-coverage rates of the intervals were estimated for each bin for each simulation by taking the proportion of the 500 cases where the confidence interval did not cover the true value of the variogram. For convenience, in calculating coverage, the true variogram for each bin was approximated by the mean over the 500 replicates of the empirical variogram for the bin. We found that this was a very good approximation for the true variogram provided the bin is sufficiently narrow so that $\gamma(h)$ does not vary significantly over the bin, which was the case for our bin ranges.

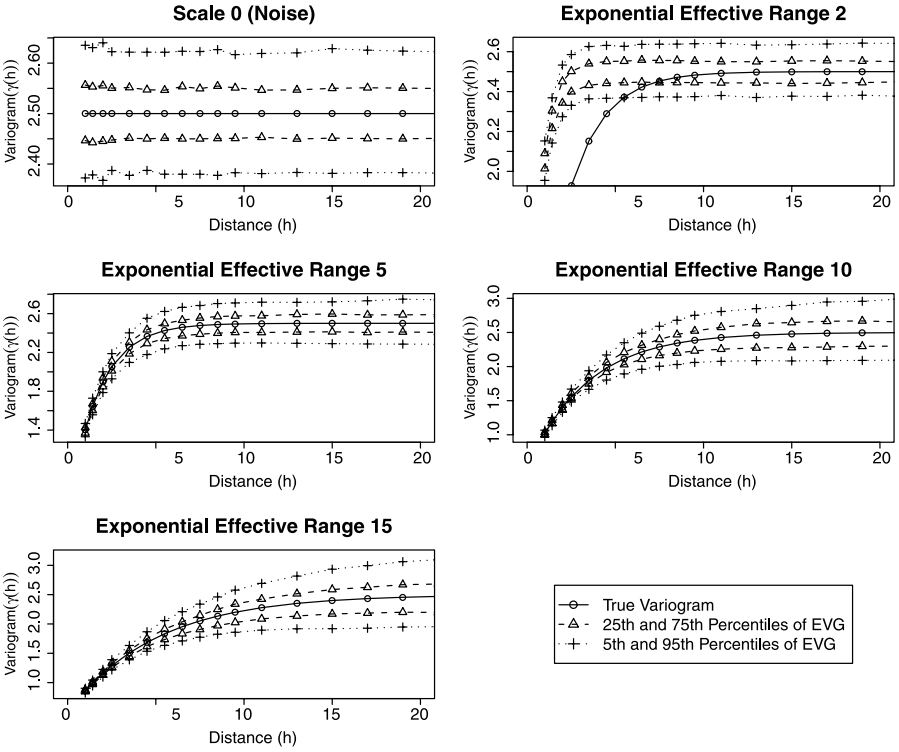

Fig. 1 True variograms and distribution of empirical variogram for exponential variogram models and normally distributed data

4.2 Results

4.2.1 True Variograms and Variability of Empirical Variograms

Figure 1 shows the true variogram for exponential models with effective range 0 (uncorrelated data), 2, 5, 10, and 15. The figure also shows the upper and lower fifth and 25th percentiles of the empirical variograms over the 500 simulations. It can be seen that for uncorrelated data, the variability of the empirical variogram is about the same for all bins. As the effective range increases so that there are higher correlations in the dataset, a fanning out pattern is evident with greater variability associated with longer distances.

4.2.2 Coverage when Exponential Variogram Model Is Correct and Data Are Normal

Figure 2 shows the non-coverage rates of the various confidence intervals over the 500 simulations. All of the confidence intervals did well for uncorrelated data, as would be expected, but had higher non-coverage as the effective range increased. The quasi-block-jackknife consistently had the coverage closest to the nominal 10% rate,

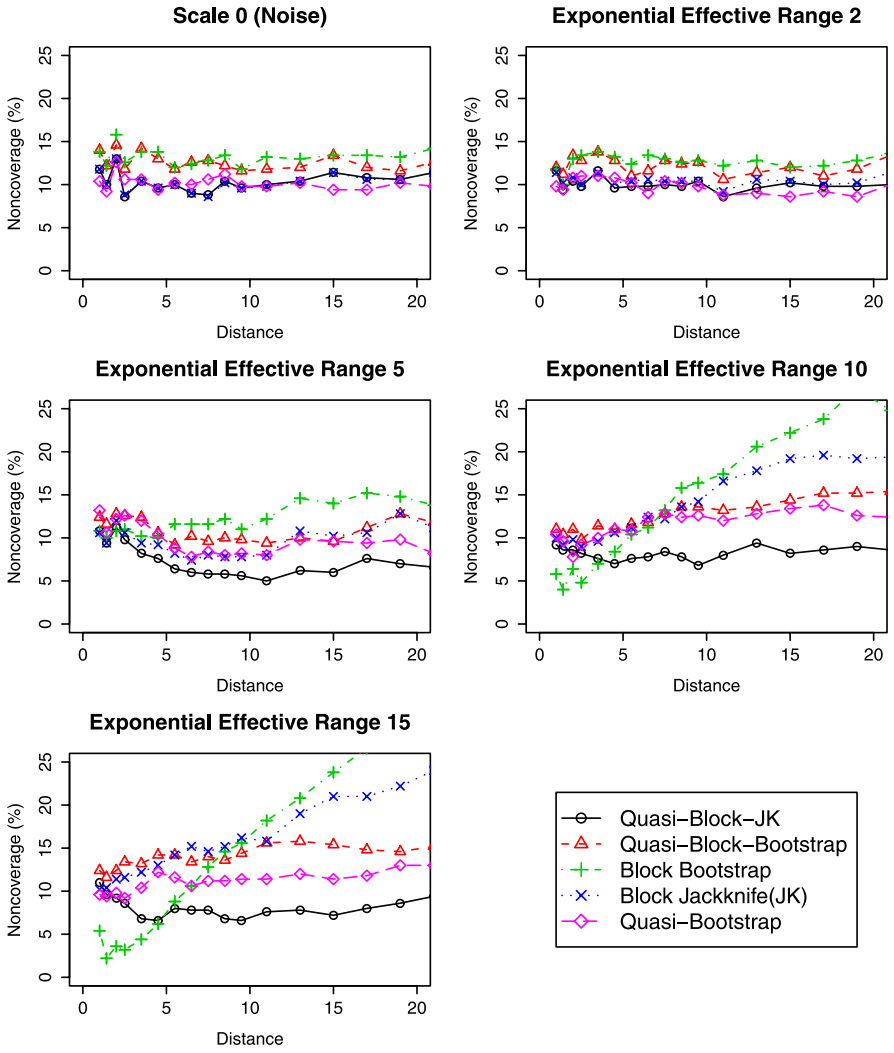

Fig. 2 Noncoverage rates of confidence intervals for VGs, and for normal data simulated from exponential variogram models with different effective ranges (90% nominal coverage)

followed by the quasi-bootstrap, the quasi-block-bootstrap, the block jackknife, and the block bootstrap. The block bootstrap was clearly the worst performer, presumably because edge effects were substantial when the effective range was 5 or more. It should be noted that this is not the situation for which the block bootstrap is designed as correlations across blocks are clearly not negligible.

4.2.3 Exponential Variogram Model Is Incorrect and Data Are Normal

The quasi-resampling methods work by fitting an exponential variogram model to the observed data. In practice, the true variogram is likely to differ from the exponential

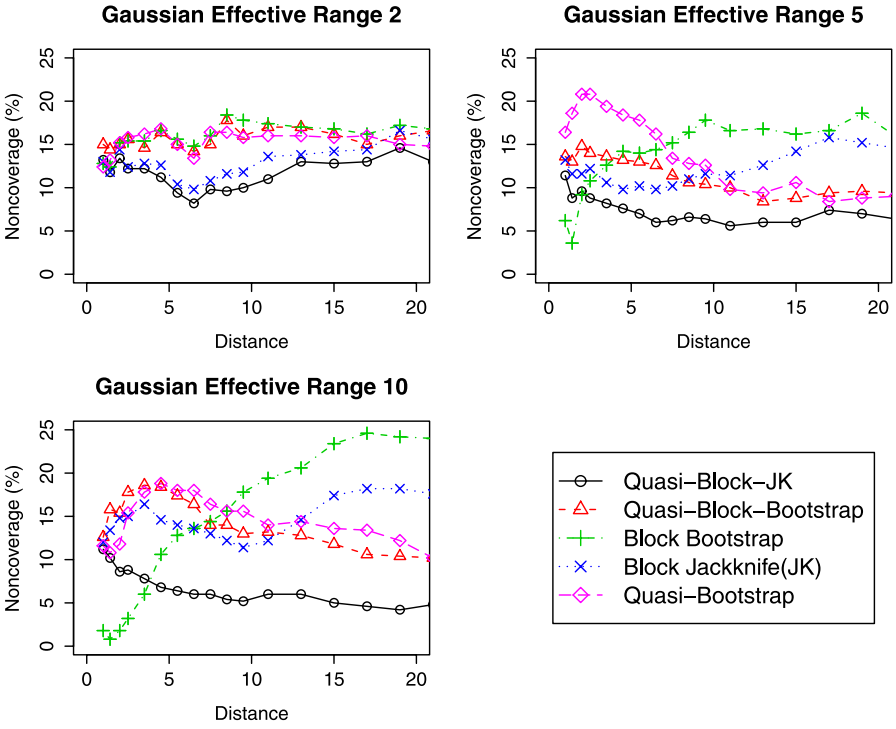

Fig. 3 Noncoverage rates of confidence intervals for VGs, and for normal data simulated from Gaussian variogram models (90% nominal coverage)

model to some extent. Figure 3 shows the non-coverage of the various confidence intervals when the data was generated by a Gaussian variogram model with effective range of 2, 5, or 10. All five methods were adequate when the effective range was 2. For longer effective ranges, only the quasi-block-jackknife gave acceptable coverage levels.

4.2.4 Lognormal Data

Figure 4 shows non-coverage rates for the lognormally distributed data, and for exponential and Gaussian variogram models with effective ranges 0, 2, and 10. All of the methods were adequate when the range was 0, but became dramatically worse for longer ranges. The scale of the y-axis in Fig. 4 differs from Figs. 2, 3, and 5 to accommodate this. The quasi-block-jackknife was again clearly the best method, but still had unacceptably high non-coverage when the effective range was 10.

4.2.5 Lognormal Data when a Box–Cox Transformation is Applied

Figure 5 shows the results for the same lognormal data as Fig. 4 when a Box–Cox transformation is applied to the data before implementing the confidence interval methods. The true value of λ for this data was 0 and the estimator $\hat{\lambda}$ was close to

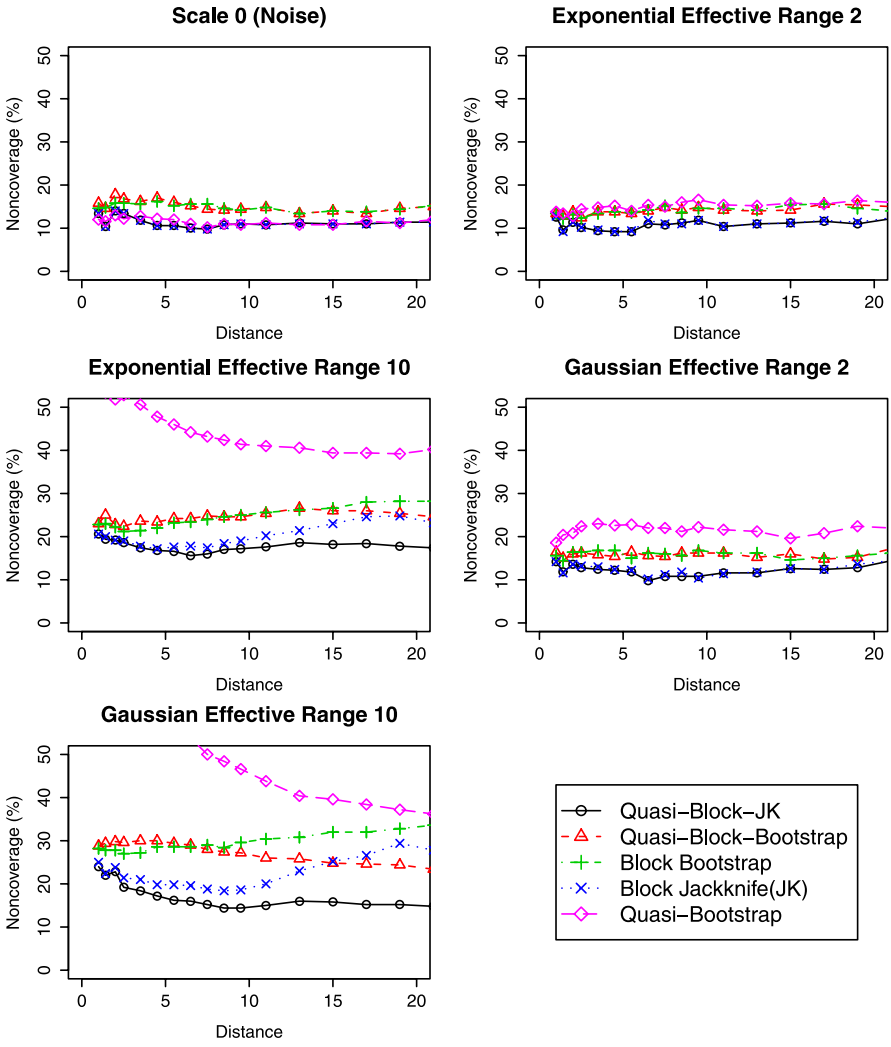

Fig. 4 Noncoverage rates of confidence intervals for VGs, and for lognormal data simulated from several variogram models (90% nominal coverage)

unbiased with standard deviation of 0.16 or less over the 500 simulations. Comparing to Fig. 4 shows that making the transformation substantially improves the coverage for all of the confidence interval methods. The quasi-block-jackknife remains the best of the methods, with non-coverage fairly close to 10% in all cases.

4.2.6 Other Methods

A number of other evaluations were also conducted. The performance of the five confidence interval methods under cubic and spherical variogram models was evaluated in Clark and Allingham (2010) with similar conclusions.

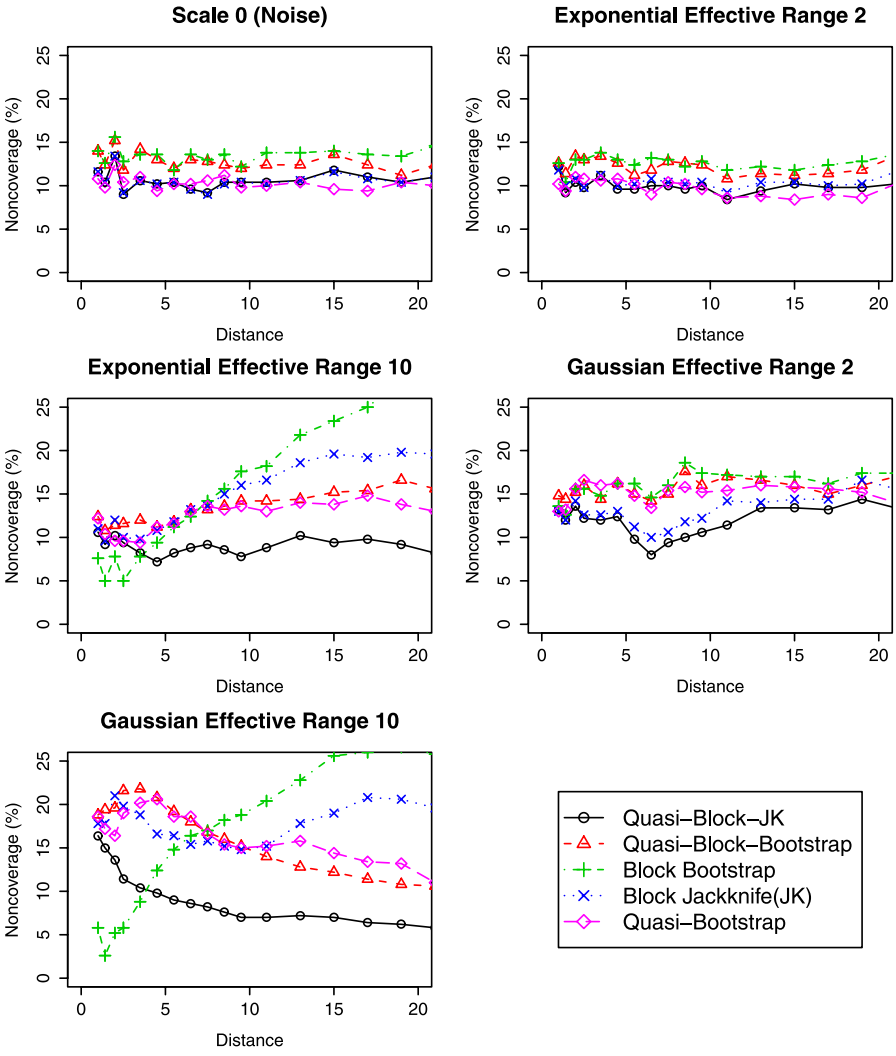

Fig. 5 Noncoverage rates of confidence intervals for VGs of Box–Cox-transformed data, and for lognormal data simulated from several variogram models (90% nominal coverage)

A number of other confidence interval methods have also been evaluated. A standard bootstrap, where individual observations were resampled ignoring spatial location, had very high non-coverage except when the effective range was 0. A parametric bootstrap was implemented with multivariate normally distributed data simulated using the fitted exponential variogram model. This performed very similarly to the quasi-bootstrap when the data was normal, but substantially worse when the data was lognormal. The block bootstrap and quasi-block-bootstrap shown were calculated using fixed blocks. The data was divided up into 25 blocks, and these blocks were effectively interchanged at random. Moving block bootstraps were also calculated, where the donor blocks could be any 10 by 10 block in the region. The results

were very similar to the fixed block bootstraps. Percentile confidence intervals were also calculated for the three bootstrap methods, giving a slight improvement over the normal approximation confidence intervals shown in the figures.

5 Discussion

A range of methods have been proposed for calculating confidence intervals for statistics calculated from spatial data, but their applicability to empirical variograms has been given relatively little attention. These methods include the block bootstrap, the block jackknife and the quasi-bootstrap. The simulations described in this paper show that all of these methods, as well as new quasi-block-bootstrap and quasi-block-jackknife methods developed in this article, perform reasonably well when the effective range is short (1/25 of the width of a square region). For longer range dependency, the block jackknife performs less well while the block bootstrap performs very poorly. The quasi-bootstrap gives close to correct coverage when the assumed variogram model is correct, but is sensitive to misspecification of this model. The new quasi-block-jackknife was clearly the best performer out of the methods considered, and was the only method with good coverage properties in all of the scenarios considered except when the effective range was 20% or more of the region width (in this case all of the methods performed poorly). It was robust both to non-normality and to misspecification of the variogram function.

Acknowledgements The authors thank Professors Alan Welsh and Ray Chambers for their helpful comments and advice. A referee made detailed comments which improved the paper considerably.

References

Chiles J-P, Delfiner P (1999) Geostatistics: Modelling spatial uncertainty. Wiley, New York

Clark RG, Allingham S (2010) Further simulation results on the robustness of resampling confidence intervals for empirical variograms. University of Wollongong Centre for Statistical and Survey Methodology Working Paper 02-10. http://www.cssm.uow.edu.au and http://ro.uow.edu.au/

Cosh MH, Stedinger J, Brutsaert W (2003) Time changes in spatial structure of surface variability in the southern great plains. Adv Water Resour 26:407–415

Cressie N (1993) Statistics for spatial data. Wiley, New York

Cressie N, Hawkins D (1980) Robust estimation of the variogram: I. Math Geol 12(2):115–125

Davison A, Hinkley D (1997) Bootstrap methods and their application. Cambridge University Press, Cambridge

Dowd P (1984) The variogram and kriging: robust and resistant estimators. In: Geostatistics for natural resource characterization. Reidel, Dordrecht, pp 91–106

Emery X (2007) Reducing fluctuations in the sample variogram. Stoch Environ Res Risk Assess 21(4):391–403

Garrigues S, Allard D, Baret F, Weiss M (2006) Quantifying spatial heterogeneity at the landscape scale using variogram models. Remote Sens Environ 103:81–96

Gascuel-Odoux C, Boivin P (1994) Variability of variograms and spatial estimates due to soil sampling: a case study. Geoderma 62:165–182

Hall P (1985) Resampling a coverage pattern. Stoch Process Appl 20:231–246

Huisman J, Snepvangers J, Bouten W, Heuvelink G (2003) Monitoring temporal development of spatial soil water content variation: comparison of ground penetrating radar and time domain reflectometry. Vadose Zone J 2(4):519–529

Kunsch HR (1989) The jackknife and the bootstrap for general stationary observations. Ann Math Stat 17(3):1217–1241
Lahiri S (2003) Resampling methods for dependent data. Springer, New York
Lark RM (2000) A comparison of some robust estimators of the variogram for use in soil survey. Eur J Soil Sci 51:137–157
Marchant B, Lark R (2004) Estimating variogram uncertainty. Math Geol 36(8):867–898
Matheron G (1963) Geostatistics. Econ Geol 58(2):1246–1266
Oliver M, Webster R (2001) Geostatistics for environmental scientists. Wiley, Chichester
R Development Core Team (2007) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0
Schlather M (2006) RandomFields: Simulation and analysis of random fields. R package version 1.3.29
Sebastyén Z (2004) Uncertainty investigation of the semivariogram by means of the bootstrap method. Acta Geol Hung 47(1):83–91
Shafer J, Varljen M (1990) Approximation of confidence limits on sample semivariograms from single realizations of spatially correlated random fields. Water Resour Res 26(8):1787–1802
Solow A (1985) Bootstrapping correlated data. Math Geol 17(7):769–775
Tang L, Schucany W, Woodward W, Gunst R (2006) A parametric spatial bootstrap. Technical Report SMU-TR-337, Southern Methodist University, Dallas, Texas
Wang L, Okin GS, Caylor KK, Macko SA (2009) Spatial heterogeneity and sources of soil carbon in southern African savannas. Geoderma 149:402–408
Wingle W, Poeter E (1993) Uncertainty associated with semivariograms used for site simulation. Ground Water 31(5):365–376