A Composite Likelihood Approach to Semivariogram Estimation

Frank C. CURRIERO and Subhash LELE

This article proposes the use of estimating functions based on composite likelihood for the estimation of isotropic as well as geometrically anisotropic semivariogram parameters. The composite likelihood approach is objective, eliminating the specification of distance lags and lag tolerances associated with the commonly used moment estimator. Extensions to the geometric anisotropy case include a parameterized transformed distance function, which eliminates the subjective estimation of the parameters of geometric anisotropy. The composite likelihood approach requires no matrix inversions and the estimators are shown to be consistent in a fashion similar to maximum likelihood and restricted maximum likelihood but without reliance on strong distributional assumptions. Predictions based on composite likelihood estimates performed very well using isotropic and geometric anisotropic simulated data and compared favorably to predictions based on the traditional approach in the isotropic case. Comparisons were also made using data collected on iron-ore measurements where previous analyses determined a geometric anisotropic semivariogram model to be appropriate.

Key Words: Estimating functions; Geometric anisotropy; Geostatistics; Kriging; Simulation; Spatial dependence; Variogram.

1. INTRODUCTION

Kriging, a geostatistical tool for spatial prediction, has proved extremely useful in various substantive areas of application such as mining, environmental remediation, and ecology. See Cressie (1991), Journel and Huijbregts (1978), and Haining (1990) for theoretical underpinnings of kriging and various applications. A key component in the geostatistical process is the estimation and modeling of spatial dependence which is usually accomplished with the semivariogram function, under the assumption of isotropic or anisotropic behavior (definitions provided in the next section). Common methods for estimation of semivariogram parameters are method of moments, maximum likelihood, and restricted maximum likelihood. Zimmerman and Zimmerman (1991) provided a review and comparison of these methods.

Recently Lele (1997) introduced composite likelihood methods for the estimation of semivariogram parameters. The use of composite likelihood, similar to maximum

Frank C. Curriero is a Research Scientist, Department of Biostatistics, The Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205 (E-mail: fcurrier@jhsph.edu). Subhash Lele is Associate Professor, Department of Biostatistics, The Johns Hopkins University, 615 N. Wolfe Street, Baltimore, MD 21205.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 4, Number 1, Pages 9–28

likelihood and restricted maximum likelihood methods, eliminates the need for subjective specification of distance lags and lag tolerances often required for the method of moments estimator, while at the same time retains the model robustness properties of method of moments. Lele (1997) compared performance of the method of moments estimation with the composite likelihood method using Godambe's information criterion (Godambe 1960) for isotropic semivariograms. His results indicate considerable gains in efficiency for the composite likelihood approach when compared to method of moments estimation based on least squares.

The objectives of this article are two-fold. First, we present the composite likelihood approach as an objective procedure for the estimation of semivariogram parameters with isotropic and geometric anisotropic data. This technique provides a less subjective approach than the currently used methods based on graphical descriptions. The composite likelihood estimators are also shown to be consistent in a fashion similar to maximum likelihood and restricted maximum likelihood but without reliance on strong distributional assumptions. Second, we evaluate *predictive* performance of the composite likelihood approach using both isotropic and geometric anisotropic simulated data. Data collected on iron-ore measurements are also re-analyzed to illustrate the composite likelihood approach as an objective and simultaneous procedure for geometric anisotropic semivariogram parameter estimation.

2. NOTATION AND PRELIMINARIES

Let $Z(\mathbf{s})$ represent a spatial process, where $\mathbf{s}$ denotes location coordinates, often taken to be of dimension two, and $Z(\cdot)$ the process value at location $\mathbf{s}$. Consider the ordinary kriging model (Cressie 1991) which assumes $E(Z(\mathbf{s})) = \mu$. The semivariogram for the $Z(\cdot)$ process, denoted by $\gamma(\mathbf{s}_i, \mathbf{s}_j)$, is given by

$$\gamma(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{2}\mathrm{var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)).$$

The variogram is defined to be $2\gamma(\mathbf{s}_i, \mathbf{s}_j)$. Further assume the $Z(\cdot)$ process to be stationary in the sense that $\gamma(\mathbf{s}_i, \mathbf{s}_j)$ reduces to a function of $\mathbf{s}_i - \mathbf{s}_j$ and $\mathrm{var}(Z(\mathbf{s}))$, if it exists, is constant. Covariances can also be modeled when $\mathrm{var}(Z(\mathbf{s})) < \infty$ through the covariogram function $C(\cdot)$, defined to be

$$C(\mathbf{s}_i - \mathbf{s}_j) = \mathrm{var}(Z(\mathbf{s})) - \gamma(\mathbf{s}_i - \mathbf{s}_j).$$

In geostatistics spatial dependence is commonly modeled with the semivariogram or variogram and subsequently used in kriging (prediction of unsampled locations).

Let $||\mathbf{s}_i - \mathbf{s}_j||$ denote the Euclidean distance between locations $\mathbf{s}_i$ and $\mathbf{s}_j$. At times we will use $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ and $d_{ij} = d(\mathbf{s}_i, \mathbf{s}_j) = ||\mathbf{s}_i - \mathbf{s}_j||$. Suppose $\gamma(\mathbf{s}_i, \mathbf{s}_j) = \gamma(||\mathbf{s}_i - \mathbf{s}_j||)$; that is, the semivariogram depends only on the distance between the locations. Such semivariograms are called isotropic semivariograms. One of the commonly used isotropic semivariogram models is the exponential model

$$\gamma(||\mathbf{s}_i - \mathbf{s}_j||; \phi) = c_0 + \sigma^2 \left(1 - \rho^{||\mathbf{s}_i - \mathbf{s}_j||}\right),$$

where $\phi = (c_0, \sigma^2, \rho)$. The parameters c_0 and σ^2 are called the nugget and the sill, respectively, and $c_0+\sigma^2$ represents the process variance. The parameter ρ measures spatial dependence. In the geostatistical literature, ρ is often parameterized as $\exp(-1/a)$, where a is called the range of the process. The value $3a$ is referred to as the effective range, denoting the approximate distance at which observations become spatially uncorrelated.

We say that the process Z is anisotropic when the dependence between $Z(\mathbf{s}_i)$ and $Z(\mathbf{s}_j)$ is a function of both distance and the direction between locations $\mathbf{s}_i$ and $\mathbf{s}_j$. Geometric anisotropy proposes

$$\gamma(\mathbf{s}_i, \mathbf{s}_j; \phi) = \gamma(||\mathbf{A}(\mathbf{s}_i - \mathbf{s}_j)||; \phi),$$

where $\mathbf{A}$ is a square matrix of appropriate dimension. That is, the process is isotropic after the space is transformed in an affine fashion. Such transformations can be interpreted as a rotation and stretching of the coordinate axes. See Journel and Huijbregts (1978, p. 177) for details on geometric anisotropy.

3. CLASSICAL ESTIMATION PROCEDURES

3.1 METHOD OF MOMENTS

Let $\{z(\mathbf{s}_1), \ldots, z(\mathbf{s}_n)\}$ represent an observed set of spatial data. Undoubtedly the most popular approach to semivariogram estimation is that based on method of moments. The classical method of moments estimator of the semivariogram (due to Matheron 1962) is given by

$$\widehat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (z(\mathbf{s}_i) - z(\mathbf{s}_j))^2, \tag{3.1}$$

where the set $N(\mathbf{h})$, assuming isotropy,

$$N(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j) : \|\mathbf{s}_i - \mathbf{s}_j\| = \|\mathbf{h}\|; i, j = 1, \ldots, n\},$$

is understood to contain all pairs of locations that are separated by distance $\|\mathbf{h}\|$ with $|N(\mathbf{h})|$ being the number of such pairs. In practice a smoothed version of (3.1) is usually employed because $|N(\mathbf{h})|$ is often too small to allow for any reliable averaging, especially when data are irregularly spaced. The common method is to group the distances $d(\mathbf{s}_i, \mathbf{s}_j)$ into bins according to chosen distance lags and lag tolerances, similar to histogram smoothing. The corresponding averaged $\frac{1}{2}(z(\mathbf{s}_i) - z(\mathbf{s}_j))^2$ in each bin is taken as the semivariogram estimate for that distance lag. Lag tolerances must be chosen so that adequate spatial resolution and stability in the smoothed estimator are retained. Journel and Huijbregts (1978) suggested choosing lag tolerances so that at least 30 location-to-location pairs fall within each bin. Further smoothing is accomplished by considering only those location pairs $(\mathbf{s}_i, \mathbf{s}_j)$ that are within half the maximum $d(\mathbf{s}_i, \mathbf{s}_j)$, since the variability in these squared differences is usually extreme for locations "too far" apart.

Once lag tolerances have been specified, a plot of the chosen distance lags, denoted by $\mathbf{h}(j), j = 1, \ldots, k$, versus the corresponding semivariogram estimate, $\widehat{\gamma}(\mathbf{h}(j))$, provides a graphical display of the empirical semivariogram. Let $\gamma(\mathbf{h}; \phi)$ be the parameteric model to be fitted. Then the parameters in ϕ can be estimated with least squares by minimizing

$$\sum_{j=1}^{k} w_j \left(\widehat{\gamma}(\mathbf{h}(j)) - \gamma(\mathbf{h}(j); \phi)\right)^2 \tag{3.2}$$

with respect to ϕ. In practice, the weights w_j are commonly defined so that (3.2) represents a weighted least squares (WLS) solution (Cressie 1985).

Directional semivariogram estimates can be calculated by further restricting the paired differences in (3.1) to also be within given directions defined by prespecified angles and angle tolerances (Isaaks and Srivastava 1989). The directional moment semivariogram estimator can be denoted by $\widehat{\gamma}(\mathbf{h}; \tau)$, where τ is a direction chosen as an angle from a fixed azimuth. The set $N(\mathbf{h}; \tau)$ would then contain the pairs of locations $(\mathbf{s}_i, \mathbf{s}_j)$ that are in each respective distance bin and within the specified angle tolerance of τ. In practice, anisotropic spatial dependence is usually investigated with graphical procedures such as directional semivariogram estimates and rose diagrams describing the ellipse of anisotropy (Isaaks and Srivastava 1989). Rotation and stretching parameters for the case of geometric anisotropy (Journel and Huijbregts 1978) are usually determined based on these procedures. In addition to having to specify the angles τ and angle tolerances, characterizing anisotropy in this fashion also reduces the amount of data (i.e., location-to-location pairs) available for each semivariogram lag estimate.

The method of moments approach to semivariogram estimation is subjective and can be sensitive to the chosen distance lags and lag tolerances (Myers, Begovich, Butz, and Kane 1982). Furthermore, Webster and Oliver (1992) suggested that it may take up to 150 or 200 sample observations for such a procedure to produce reliable semivariogram estimates. On the other hand, the method of moments procedure is robust, requiring no strong distributional assumptions about the process $Z(\mathbf{s})$, and the plotted empirical semivariogram provides a valuable graphical tool for exploring spatial dependence.

3.2 MAXIMUM LIKELIHOOD AND RESTRICTED MAXIMUM LIKELIHOOD APPROACH

If the underlying process follows a Gaussian distribution, one can also use maximum likelihood (ML) or restricted maximum likelihood (REML) methods to estimate the semivariogram parameters. Let $\mathbf{Z} = \left(Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)\right)$ and assume

$$\mathbf{Z} \sim N(\mu \mathbf{1}, \mathbf{C}(\phi)),$$

where μ is a constant, $\mathbf{1}$ is the vector of $1's$, and $\mathbf{C}(\phi)$ is defined through the covariogram function with $C_{ij}(\phi) = \mathrm{cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = \mathrm{var}(Z(\mathbf{s})) - \gamma(\mathbf{s}_i - \mathbf{s}_j; \phi)$. The ML estimate of ϕ is obtained by maximizing the log-likelihood

$$\mathrm{l}(\phi, \mu; \mathbf{Z}) = -\frac{1}{2} \log |\mathbf{C}(\phi)| - \frac{1}{2} \left(\mathbf{Z} - \mu\mathbf{1})'\mathbf{C}(\phi)^{-1}(\mathbf{Z} - \mu\mathbf{1}\right), \tag{3.3}$$

where $|\cdot|$ denotes the determinant.

It is advisable to eliminate the nuisance parameter μ before estimating ϕ. This is achieved by considering the likelihood of the contrasts. Consider a vector of contrasts

$\mathbf{Z}_c = \{Z(\mathbf{s}_i) - Z(\mathbf{s}_1), i = 2, \ldots, n\}$. It is easy to see that this vector corresponds to multiplying the original data vector $\mathbf{Z}$ by a matrix $\mathbf{Q}$ such that its first column consists of -1 and the ith column consists of zeros except in the ith place. Thus,

$$\mathbf{Z}_c = \mathbf{QZ} \sim N(\mathbf{0}, \mathbf{QC}(\phi)\mathbf{Q}').$$

The likelihood corresponding to $\mathbf{Z}_c$ is given by

$$\mathrm{L}(\phi; \mathbf{Z}_c) = \frac{1}{(2\pi)^{n/2} |\Psi(\phi)|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{Z}_c'\Psi^{-1}(\phi)\mathbf{Z}_c\right\}, \tag{3.4}$$

where

$$\begin{aligned} \Psi_{ii}(\phi) &= 2\gamma(\mathbf{s}_i, \mathbf{s}_1; \phi) \\ \Psi_{ij}(\phi) &= \gamma(\mathbf{s}_i, \mathbf{s}_1; \phi) + \gamma(\mathbf{s}_j, \mathbf{s}_1; \phi) - \gamma(\mathbf{s}_i, \mathbf{s}_j; \phi). \end{aligned}$$

Notice that this is a function of ϕ only and maximizing with respect to ϕ yields the restricted maximum likelihood (REML) estimator. Details on the ML and REML methods for semivariogram estimation can be found in Zimmerman and Zimmerman (1991).

The methods of ML or REML theoretically yield consistent estimators but require a full specification of the probabilistic model. Moreover, they involve inversion of large matrices which can be computationally prohibitive. Uniqueness of the maximum is also not always guaranteed (Warnes and Ripley 1987). In view of this, a natural question to ask would be: Can we approximate the likelihood function by something that behaves almost like a likelihood but is easy to deal with, both computationally and mathematically, and is less reliant on such strong distributional assumptions? The composite likelihood approach presented next addresses this issue. Recently, Barry, Crowder, and Diggle(1998) provided a parametric version of variogram estimation based on quasi-likelihood. Their technique is free of strong distributional assumptions, however, and still requires matrix inversions, although of a reduced size to that found in ML and REML, and is based on results obtained from the moment estimator (3.1).

4. COMPOSITE LIKELIHOOD APPROACH

The idea of composite likelihood, although discussed in various disguises such as pseudolikelihood (Besag 1975) or partial likelihood (Cox 1975), was developed in its own right by Lindsay (1988). To construct a composite likelihood, one starts with a set of conditional or marginal events for which one can write log-likelihoods. The log-composite likelihood is then formed by adding together these individual component log-likelihoods (Lindsay 1988). Therefore, log-composite likelihood simply refers to the pooling of log-likelihood contributions in an additive fashion in circumstances where the components do not necessarily represent independent replicates.

There are two motivations for constructing composite likelihoods. First, they provide a substitute method of estimation when maximum likelihood is very difficult to calculate. Second, they sometimes represent that portion of the model we are most comfortable with modeling, and the resultant estimators can be consistent even when full maximum likelihood estimators are not, a form of consistency robustness. Examples of composite likelihood applications include Heagerty and Lele (1998) for binary data in space; generalized Mantel–Haenszel analysis for 2×2 tables (Liang 1987); and working independence generalized estimating equations (Liang and Zeger 1986) for longitudinal data. In the following we illustrate the use of composite likelihood for semivariogram estimation in the isotropic and geometric anisotropic cases. Details regarding the validity of the composite likelihood approach are then discussed followed by a summary comparison.

4.1 COMPOSITE LIKELIHOOD ESTIMATION FOR ISOTROPY AND GEOMETRIC ANISOTROPY

Consider the ordinary kriging situation,

$$\begin{aligned} E(Z(\mathbf{s}_i)) &= \mu \\ \text{var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)) &= E(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)) \\ &= 2\gamma(d_{ij}; \phi). \end{aligned}$$

Let $\mathbf{V}$ be the vector of contrasts, $v_{ij} = Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$. Now consider the product of the marginal densities of these contrasts, namely,

$$\text{CL}(\phi; \mathbf{V}) = \prod_{i=1}^{n-1} \prod_{j>i} f(v_{ij}; \phi). \tag{4.1}$$

This is a "composite likelihood" because each of the components $f(v_{ij}; \phi)$ is a legitimate likelihood. Following this definition of composite likelihood (Lindsay 1988), selection of $f(\cdot)$ is restricted only to the pool of valid density functions. Now, as a matter of convenience, take $v_{ij} \sim N(0, 2\gamma(d_{ij}; \phi))$, an assumption that will be relaxed later, so that,

$$f(v_{ij}; \phi) = \frac{1}{\sqrt{2\pi}\sqrt{2\gamma(d_{ij}; \phi)}} \exp\left\{ -\frac{\left(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right)^2}{4\gamma(d_{ij}; \phi)} \right\}.$$

Hence, ignoring constant terms, this negative log-composite likelihood up to a constant can be written as

$$\sum_{i=1}^{n-1} \sum_{j>i} \left\{ \frac{\left(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right)^2}{2\gamma(d_{ij}; \phi)} + \log\left(\gamma(d_{ij}; \phi)\right) \right\}. \tag{4.2}$$

Composite likelihood semivariogram estimates can then be obtained by minimizing the above quantity with respect to ϕ.

At this point, note that we considered the contrasts $v_{ij} = Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$ to eliminate the nuisance parameter μ (as in spirit of REML). The pairwise composite likelihood formulated in (4.1), along with normality of the $v_{ij}'s$, were taken to arrive at the objective function (4.2) that is (a) computationally simpler than the ML or REML objective functions; and (b) as we demonstrate in the next section, corresponds to a zero unbiased estimating function irrespective of the Gaussian density assumed for the $v'_{ij}s$. Clearly, other composite likelihoods could have been constructed using different functions of the data with possibly different forms of compositing.

The ideas discussed previously can be extended when the process $Z(\mathbf{s})$ is anisotropic, that is dependence between $Z(\mathbf{s}_i)$ and $Z(\mathbf{s}_j)$ is a function of both the distance and the direction between locations $\mathbf{s}_i$ and $\mathbf{s}_j$. Here we consider the case of geometric anisotropy,

$$\gamma(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\phi}) = \gamma\left(\|\mathbf{A}(\mathbf{s}_i - \mathbf{s}_j)\|; \boldsymbol{\phi}\right),$$

where $\mathbf{s}_i, \mathbf{s}_j \in \Re^2$ so that $\mathbf{A}$ is a 2×2 matrix. Let $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_n)$ denote the $2 \times n$ matrix of coordinate locations at which observations are made. Let $\mathbf{S}^* = \mathbf{AS}$ denote the linear transformation of the original space $\mathbf{S}$. Geometric anisotropy contends that on this linearly transformed space, the semivariogram is isotropic. Now consider the spectral decomposition of the matrix $\mathbf{A}$,

$$\mathbf{A} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} p & 0 \\ 0 & q \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix},$$

for $\alpha, \theta \in \Re$, and $p, q > 0$. With $\mathbf{A}$ decomposed as above, the transformation $\mathbf{AS}$ first rotates the space $\mathbf{S}$ through the angle θ (positive θ in the counterclockwise direction), stretches each axes (the rows of $\mathbf{S}$) by p and q, respectively, and then rotates again through an angle α. This last rotation through the angle α is unnecessary, because of invariance with respect to rotation; however, it is sometimes included for graphical purposes taking $\alpha = -\theta$ so to rotate back to the original perspective.

Expressing $\mathbf{s}_i$ as $\mathbf{s}_i = (x_i, y_i)'$, the distances in the transformed space $\mathbf{S}^* = \mathbf{AS}$, with $\mathbf{S}^* = (\mathbf{s}_1^*, \ldots, \mathbf{s}_n^*)$, can be written as

$$\begin{aligned} \|\mathbf{s}_i^* - \mathbf{s}_j^*\| &= \{p^2\left((x_i - x_j)^2\cos^2\theta + (y_i - y_j)^2\sin^2\theta \right. \\ &\quad \left. -2(x_i - x_j)(y_i - y_j)\cos\theta\sin\theta\right) \\ &\quad +q^2\left((x_i - x_j)^2\cos^2\theta + (y_i - y_j)^2\sin^2\theta \right. \\ &\quad \left. +2(x_i - x_j)(y_i - y_j)\cos\theta\sin\theta\right)\}^{1/2} \\ &= d_{ij}^*(p, q, \theta). \end{aligned}$$

These do not involve the parameter α, the invariance mentioned previously. The negative log-composite likelihood for the geometrically anisotropic process based on $Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$ can be written as

$$\sum_{i=1}^{n-1}\sum_{j>i}\left\{\frac{\left(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right)^2}{2\gamma\left(d_{ij}^*(p, q, \theta); \boldsymbol{\phi}\right)} + \log\gamma\left(d_{ij}^*(p, q, \theta); \boldsymbol{\phi}\right)\right\}, \tag{4.3}$$

which can be minimized with respect to the parameters $(\boldsymbol{\phi}, p, q, \theta)$.

Estimation of geometric anisotropy parameters using the transformation $\|\mathbf{A}(\mathbf{s}_i - \mathbf{s}_j)\|$ is not unique to the composite likelihood approach. Similar techniques have been used in regards to ML and REML (Vecchia 1988; Hobert, Altman, and Schofield 1997).

Although, notice that

$$d_{ij}^*(p, q, \theta) = q d_{ij}^*(p/q, 1, \theta).$$

Hence if the geometric anisotropic exponential semivariogram model is used,

$$\gamma(d^*_{ij}(p,q,\theta);\phi) = c_0 + \sigma^2(1 - \rho^{d^*_{ij}(p,q,\theta)}),$$

with $\phi = (c_0, \sigma^2, \rho)$, then only $(c_0, \sigma^2, \widetilde{\rho}, \lambda, \theta)$, where $\widetilde{\rho} = \rho^q$ and $\lambda = p/q$ are identifiable. This identifiability problem is not unique to the method of composite likelihood. It also occurs with the methods of maximum likelihood and restricted maximum likelihood.

Method of moments estimation can provide a graphical check on the appropriateness of the model and parameter estimates in the presence of isotropy or geometric anisotropy. In the geometric anisotropic case this can be accomplished by first deforming the sampling space according to $\widehat{\theta}$ and $\widehat{\lambda}$. The distances in this deformed space, $d^*_{ij}(\widehat{\lambda}, 1, \widehat{\theta})$, can then be used with the moment semivariogram estimator (3.1) to estimate the transformed isotropic spatial dependence. The fitted model,

$$\gamma\left(d^*_{ij}\left(\widehat{\lambda}, 1, \widehat{\theta}\right); \widehat{\phi}\right),$$

can then be plotted through these estimates to gauge the fit. We illustrate this technique using the iron-ore residuals in Section 6.

4.2 COMPOSITE LIKELIHOOD ESTIMATING FUNCTIONS

In the following we detail some properties of the composite likelihood approach based on the concept of estimating functions (Godambe and Kale 1991). The development further clarifies and extends the results found in Lele (1997).

Consider the composite likelihood estimating procedure for the stationary isotropic case discussed in the previous section. Let $\gamma(d_{ij};\phi)$ denote the semivariogram model. To obtain the CL estimator of ϕ, one minimizes the negative log-composite likelihood given in (4.2)

$$\sum_{i=1}^{n-1}\sum_{j>i}\left\{\frac{\left(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right)^2}{2\gamma(d_{ij};\phi)} + \log\left(\gamma(d_{ij};\phi)\right)\right\}.$$

The corresponding estimating function is given by (Lele 1997, eq. 3)

$$\sum_{i=1}^{n-1}\sum_{j>i}\frac{\frac{d}{d\phi}\gamma(d_{ij};\phi)}{\gamma(d_{ij};\phi)}\left[\frac{\left(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right)^2}{2\gamma(d_{ij};\phi)} - 1\right] = 0. \tag{4.4}$$

Next we discuss several properties of the CL semivariogram estimator based on the above estimating function.

4.2.1 Consistency

The estimating function (4.4) is zero unbiased as long as $E(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)) = 2\gamma(d_{ij};\phi)$. We do not need to have correct specification of the marginal distribution of the contrasts v_{ij}. This is in contrast to the nonzero unbiasedness of the ML or REML estimating functions if the joint distribution is misspecified. Then, following the arguments in Crowder (1986) as modified by Heagerty and Lele (1998), it is easy to show that under increasing domain asymptotics there exists a consistent solution to the estimating functions (4.4); even if the marginal distribution of the contrasts v_{ij} are non-Gaussian, as long as $E(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)) = 2\gamma(d_{ij};\boldsymbol{\phi})$ holds along with some conditions on the rate of decay of the correlations between any two locations. Consistency of the composite likelihood estimators, therefore, depends only on correct specification of the first two moments of the process. This provides robustness properties similar to those described by Godambe and Thompson (1984). Similar consistency results for the method of moments is difficult to prove.

4.2.2 Obtaining the Consistent Estimator

In practice, unfortunately, there may be several solutions to the estimating function (4.4). The same holds true for semivariogram estimates based on ML and REML (Warnes and Ripley 1987) and method of moments. In the case of composite likelihood, we show that, in fact, the global minimum of the negative log-composite likelihood (4.2) is the consistent estimator. This is in the spirit of Wald (1949). Although similar results may be proved for ML and REML, the distinct feature of the results for CL is that it is robust against misspecification of the joint or bivariate marginal distributions. We will not produce all the mathematical details of the proof here, but only give the following key inequality.

Following Li (1997), the proof can be completed under proper rate of decay for the correlations. Assuming all expectations exist, the key idea behind Wald's proof is the fact that

$$E\left[-\text{log-likelihood}(\mathbf{Y},\beta)\right] > E\left[-\text{log-likelihood}(\mathbf{Y},\beta_0)\right]$$

if β_0 is the true parameter for data vector $\mathbf{Y}$. Li (1997) (see also Li 1996) showed that if a similar inequality holds true for any objective function, consistency follows. The objective function corresponding to the estimating function (4.4) is the negative log-composite likelihood given in (4.2).

Notice that

$$E_{\boldsymbol{\phi}_0}\left[\frac{\left(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\right)^2}{2\gamma(d_{ij};\boldsymbol{\phi})} + \log\left(\gamma(d_{ij};\boldsymbol{\phi})\right)\right] = \frac{\gamma(d_{ij};\boldsymbol{\phi}_0)}{\gamma(d_{ij};\boldsymbol{\phi})} + \log\left(\gamma(d_{ij};\boldsymbol{\phi})\right).$$

Thus, to prove $E\left[-\log \mathrm{CL}(\boldsymbol{\phi})\right] \geq E\left[-\log \mathrm{CL}(\boldsymbol{\phi}_0)\right]$, we need to prove that

$$\frac{\gamma(d_{ij};\boldsymbol{\phi}_0)}{\gamma(d_{ij};\boldsymbol{\phi})} + \log\left(\gamma(d_{ij};\boldsymbol{\phi})\right) \geq \frac{\gamma(d_{ij};\boldsymbol{\phi}_0)}{\gamma(d_{ij};\boldsymbol{\phi}_0)} + \log\left(\gamma(d_{ij};\boldsymbol{\phi}_0)\right)$$

or, equivalently,

$$\log\frac{\gamma(d_{ij};\boldsymbol{\phi})}{\gamma(d_{ij};\boldsymbol{\phi}_0)} \geq 1 - \frac{\gamma(d_{ij};\boldsymbol{\phi}_0)}{\gamma(d_{ij};\boldsymbol{\phi})}.$$

The proof follows by noticing that $\log(x) > 1 - \dfrac{1}{x}$, for all $x > 0$, with equality when $x = 1$, which occurs only when $\boldsymbol{\phi} = \boldsymbol{\phi}_0$, the true parameter.

We again emphasize here that this result holds under the less stringent condition on the model, $E(Z(\mathrm{s}_i) - Z(\mathrm{s}_j)) = 2\gamma(d_{ij};\phi)$. The marginal distribution of the contrasts v_{ij} need not be Gaussian. It is also easy to see that the result does not depend on the stationarity or isotropy assumption. We do, however, need ergodicity and law of large number to hold, as is needed with method of moments, maximum likelihood, and restricted maximum likelihood.

4.2.3 Distance Weighting

The estimating function (4.4) is written in the form of a weighted sum of zero unbiased component estimating functions with weights,

$$\frac{\frac{d}{d\phi}\gamma(d_{ij};\phi)}{\gamma(d_{ij};\phi)},$$

which represents the information content in each of these components (Godambe and Kale 1991). Note that if the distance between locations $\mathrm{s}_i, \mathrm{s}_j$ is large, $\frac{d}{d\phi}\gamma(d_{ij};\phi)$, and hence the corresponding weight, are close to zero provided $\gamma(d_{ij};\phi)$ has an asymptote as $d_{ij} \to \infty$. Thus, these estimating functions automatically down weight those pairs of observations that are far apart, or equivalently that are not informative about the semivariogram parameter (ML and REML estimating functions down weight similarly). This weighting scheme is not a simple function of the distance. It is a function of the information content of the component estimating function, in contrast to the method of moments weighting scheme.

Also note, the composite likelihood constructed in (4.1) considers the product of all pairwise contrasts. Clearly this formulation can be restricted to consider only those pairs that are within a given distance range, as is commonly suggested with method of moments estimation. It is not evident that the ML and REML objective functions share this flexibility. All our composite likelihood analyses that follow, however, are based on the full composite shown in (4.1).

Remark 1. *Note the objective function shown in (4.2) is different than the objective function described by Cressie (1991, p. 96, eq. 2.6.12) for approximating the WLS MM solution. It is easy to see that a straightforward application of Cressie's equation leads to a nonzero unbiased estimating function and hence to inconsistent estimators. See also Barry et al. (1998)*

4.3 SUMMARY COMPARISONS

In this section we review the merits of the composite likelihood approach to semivariogram estimation, stressing that it combines many positive features of existing techniques based on method of moments, maximum likelihood, and restricted maximum likelihood. For convenience, the following points are summarized in Table 1.

Table 1. Summary comparisons of the method of moments (MM), maximum likelihood (ML), restricted maximum likelihood (REML), and composite likelihood (CL) approaches to semivariogram estimation

	Semivariogram estimation			
Pros and cons	*MM*	*ML*	*REML*	*CL*
Method does not require strong dist. assumptions	√			√
Optimization does not require matrix inversion	√			√
Method does not require distance binning		√	√	√
Objective fitting of geometric anisotropy		√	√	√

Similar to method of moments and in contrast to maximum likelihood and restricted maximum likelihood:

- The composite likelihood approach does not require the correct specification of either the joint or marginal distribution of the process. Only the model for the semivariogram needs to be specified correctly, a form of consistency robustness.
- The composite likelihood approach is computationally substantially simpler, requiring no matrix inversions.

Similar to maximum likelihood and restricted maximum likelihood and in contrast to method of moments:

- The composite likelihood approach does not require specification of distance lags and lag tolerances.
- The composite likelihood approach provides an objective procedure for estimation in the presence of geometric anisotropy, eliminating the need to determine the rotation and stretching parameters, which is commonly based on graphical procedures.

In addition, based on the results in previous section, it can be seen that composite likelihood behaves in a manner similar to a likelihood function. One can use CL ratio to test for the presence of geometric anisotropy. Let $\gamma\left(\|\mathbf{A}(\mathbf{s}_i - \mathbf{s}_j)\|; \phi, \mathbf{A}(\theta, \lambda)\right)$ represent a chosen semivariogram model where the matrix of geometric anisotropy is parameterized as

$$\mathbf{A}(\theta, \lambda) = \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix},$$

θ and λ represent the rotation and stretching parameters, respectively. Note that isotropic spatial dependence corresponds to $\lambda = 1$ and is invariant with respect to rotation, so take $\theta = 0$ in such situations. Let H_1 hypothesize isotropic spatial dependence, characterized by $\gamma\left(\|\mathbf{A}(\mathbf{s}_i - \mathbf{s}_j)\|; \phi_1, \mathbf{A}(0, 1)\right)$, and let H_2 hypothesize geometric anisotropy, characterized by $\gamma\left(\|\mathbf{A}(\mathbf{s}_i - \mathbf{s}_j)\|; \phi_2, \mathbf{A}(\theta, \lambda)\right)$. The vectors ϕ_1 and ϕ_2 represent the same parameters as ϕ, but under the different models hypothesized by H_1 and H_2. Now consider the composite likelihood ratio

$$\Lambda = \frac{\mathrm{CL}_1\left(\widehat{\phi}_1, \mathbf{A}(0,1); \mathbf{V}\right)}{\mathrm{CL}_2\left(\widehat{\phi}_2, \mathbf{A}(\widehat{\theta},\widehat{\lambda}); \mathbf{V}\right)},$$

where $\mathbf{V}$ represents the vector of contrasts $v_{ij} = (Z(\mathbf{s}_i - Z(\mathbf{s}_j))$, $\widehat{\phi}_1$ and $\widehat{\phi}_2$ represent the CL estimates obtained by minimizing (4.2) and (4.3), respectively, and $\mathrm{CL}_1(\cdot)$ and $\mathrm{CL}_2(\cdot)$ denote the corresponding minimized values. The sampling distribution of $-2\log\Lambda$ under H_1 can be obtained with parametric bootstrap (Dennis and Taper 1994) and used to test for the presence of geometric anisotropy.

Furthermore, the composite likelihood approach can be made robust against outliers and influential observations using ideas presented in Lindsay (1994). It can also be adapted to account for biases incurred by preferential type sampling strategies, such as the size-biased designs discussed in Patil and Rao (1978). Extending the use of composite likelihood semivariogram estimation to these areas and other possible applications are currently being developed. It is also easy to observe that the composite likelihood procedure can be extended to discrete data situations (e.g., Heagerty and Lele 1998) provided $\mathrm{var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))$ can be properly specified.

5. PREDICTIVE PERFORMANCE COMPARISONS

Here we use the ordinary kriging model to study the predictive performance of composite likelihood semivariogram estimation using isotropic and geometric anisotropic simulated data. The exponential semivariogram model, parameterized as

$$\gamma(\mathbf{s}_i, \mathbf{s}_j; \mathbf{A}(\theta,\lambda), \sigma^2, \rho) = \sigma^2\left(1 - \rho^{\|\mathbf{A}(\mathbf{s}_i - \mathbf{s}_j)\|}\right), \tag{5.1}$$

was used as the true model of spatial dependence. The matrix of geometric anisotropy was parameterized as

$$\mathbf{A} = \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}. \tag{5.2}$$

Simulation experiments were conducted using an 8×8 regular grid with 1 unit interval spacing and two 15×15 grids obtained by halving the grid spacings to .5 (infill asymptotics) and doubling the grid spacing to 2 (increasing domain asymptotics). Let

$$\mathbf{S}_1 = \{\mathbf{s}_1, \ldots, \mathbf{s}_{n1}\}$$

represent these locations, where $n1 = 64$ for the 8×8 grid and $n1 = 225$ for the 15×15 grids. Data were simulated at these locations using the LU decomposition of the covariance matrix (Ripley 1987) under different variogram parameter combinations. We considered geometric anisotropy parameters $\theta = (0°,\ 30°,\ 60°)$ and $\lambda = (1,\ 2,\ 3)$. Note that isotropic spatial dependence corresponds to $\lambda = 1$ and is invariant with respect to rotation. The parameter ρ was varied to represent relatively weak, moderate, and strong levels of spatial dependence. Because ρ is spatial scale dependent, this was accomplished by setting the distance (effective range) at which values become approximately uncorrelated to be .20, .50, and .80 times the maximum distance over the domain $\mathbf{S}_1$. Thus, $\rho = (.22,\ .54,\ .68)$ for the 8×8 grid and 15×15 infill grid, and $\rho = (.47,\ .74,\ .83)$ for the 15×15 increasing domain grid. The variance parameter σ^2 was fixed at 1.

After specifying the set of sample locations $\mathbf{S}_1$, an additional set of 25 locations were randomly selected throughout the domain. These locations, denoted by the set

$$\mathbf{S}_2 = \{\mathbf{s}_{n1+1}, \ldots, \mathbf{s}_{n1+25}\},$$

represent locations for which kriged predictions were generated. The simulated values at these locations, however, played no role in estimating semivariogram parameters. They were taken to represent "true" values used for evaluating predictive performance.

Let $\mathbf{Z}_1 = \{z(\mathbf{s}_1), \ldots, z(\mathbf{s}_{n1})\}$ and $\mathbf{Z}_2 = \{z(\mathbf{s}_{n1+1}), \ldots, z(\mathbf{s}_{n1+25})\}$ represent the simulated data at the sampled $\mathbf{S}_1$ and unsampled $\mathbf{S}_2$ locations, respectively. Semivariogram parameters were estimated by minimizing the corresponding negative log-composite likelihood, omitting the θ and λ parameters for isotropic cases. In addition for isotropic data, parameters were also estimated via a more traditional approach by using the WLS procedure of Cressie (1985) to fit the exponential semivariogram model (5.1) by the method of moments estimator given in (3.1). No attempt was made to automate a procedure based on method of moments for the geometric anisotropic data. Also, note Zimmerman and Zimmerman (1991) concluded that, for purposes of kriging isotropic Gaussian data, little is sacrificed by using the simpler, more popular WLS MM approach to semivariogram estimation as opposed to ML or REML. Thus, in interest of brevity, we decided not to include predictions based on ML or REML estimates in our simulations.

Root mean squared error (RMSEP) in predictions

$$\text{RMSEP} = \left\{ \frac{1}{25} \sum_{i=n1+1}^{25} \left(z^*(\mathbf{s}_i) - z(\mathbf{s}_i)\right)^2 \right\}^{1/2},$$

where $z^*(\cdot)$ represents the ordinary kriged prediction based on an estimated variogram, were generated for each simulated analysis. We also considered a RMSEP value based on the specified variogram parameters used to simulate the data. Averaging these RMSEP's over many simulations provides a Monte Carlo estimate of the true prediction error at $\mathbf{S}_2$.

The following algorithm summarizes the simulations.

Step 1. Select a grid type $\mathbf{S}_1$ and prediction locations $\mathbf{S}_2$, geometric anisotropy parameters θ and λ, and spatial dependence level ρ (weak, moderate, strong).

Step 2. Simulate the data sets $\mathbf{Z}_1$ and $\mathbf{Z}_2$.

Step 3. Based on $\mathbf{Z}_1$ estimate the semivariogram parameters using the composite likelihood approach. For isotropic data ($\lambda = 1$), also estimate semivariogram parameters using the traditional WLS MM approach.

Step 4. Calculate RMSEP statistics based on the composite likelihood estimated semivariogram, on a semivariogram estimated using the traditional WLS approach for isotropic cases, and a semivariogram using the parameter values specified in Step 1.

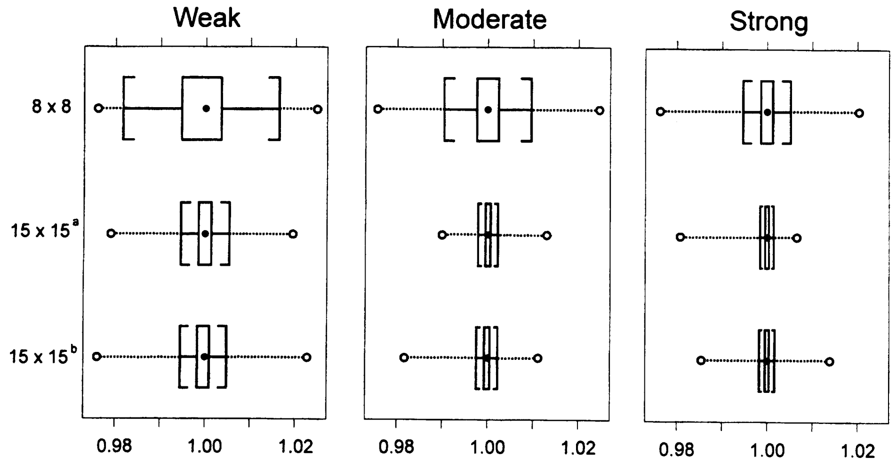

Figure 1. Boxplot distributional summaries for the isotropic simulations displayed according to level of spatial dependence (weak, moderate, strong) and grid type (8× 8, 15× 15^a infill, 15× 15^b increasing domain). The consistency of CL estimators is already evident from the reduction in variance for the increasing domain case.

For isotropic spatial dependence, generate a prediction efficiency ratio by dividing the RMSEP statistics from the traditional WLS MM approach to that from the composite likelihood approach.

Step 5. Repeat Steps 2 through 4 500 times for each grid type, anisotropy, and dependence level combination specified in Step 1.

All computations were performed with S-plus (MathSoft Inc. 1995). For the isotropic cases, functions from S-plus Spatial Stats (Kalunzy, Vega, Cardoso, and Shelly 1996) were used for method of moments semivariogram estimation combined with SAS procedure NLIN (SAS Institute 1990) for the WLS fit of (5.1). Composite likelihood estimates were based on the nonlinear multivariable minimization routine (nlminb) supplied by S-plus. Parameter starting values were set at $\rho = .1$, $\sigma^2 = \text{var}(\mathbf{Z}_l)$, $\theta = 0°$, and $\lambda = 1$. To avoid numerical difficulties, we restricted $\sigma^2 \geq .01$ and $.01 \leq \rho \leq .99$.

Figure 1 displays distributional summaries for the prediction efficiency ratios from the various isotropy simulations. All these distributions are concentrated about 1.00, indicating practically no difference in prediction between the composite likelihood and WLS MM semivariogram estimation. This conjecture is demonstrated analytically in Lele (1997) and, based on these simulations, appears to be supported more with the larger sample size and stronger levels of spatial dependence. The distributions for the 8×8 grid were trimmed up to 10% so the graphical comparisons across the different grids could be made on the same scale without loss of visual appeal.

Table 2 contains the results for the geometric anisotropy simulations. The values listed in the table are the estimated composite likelihood prediction errors divided by the estimated true prediction errors. These prediction errors were obtained by averaging the RMSEP values over each respective set of 500 simulations. The composite likelihood prediction error was always within 10% of the true prediction error for the 8×8 grid. There does not appear to be any noticeable effect due to the level of spatial dependence,

Table 2. Estimated composite likelihood prediction errors relative to the estimated true prediction errors for simulations from the 8× 8 grid, 15× 15[a] infill grid, and 15× 15[b] increasing domain grid. All values have been rounded to two decimal place accuracy.

Grid	Spatial	Geometric anisotropy					
		$\theta=0°$		$\theta=30°$		$\theta=60°$	
Type	Dependence	$\lambda=2$	$\lambda=3$	$\lambda=2$	$\lambda=3$	$\lambda=2$	$\lambda=3$
	Weak	1.05	1.06	1.06	1.06	1.05	1.05
8×8	Moderate	1.07	1.09	1.06	1.08	1.05	1.09
	Strong	1.07	1.10	1.05	1.07	1.06	1.09
	Weak	1.02	1.05	1.02	1.02	1.04	1.05
15× 15[a]	Moderate	1.03	1.03	1.03	1.03	1.04	1.12
	Strong	1.04	1.04	1.05	1.05	1.05	1.11
	Weak	1.03	1.05	1.02	1.02	1.04	1.08
15× 15[b]	Moderate	1.03	1.04	1.03	1.04	1.05	1.10
	Strong	1.04	1.04	1.04	1.04	1.05	1.12

and larger differences seem to occur with more severe anisotropy ($\lambda = 3$). The relative difference in prediction errors for the most part drops to within 5% for the 15×15 infill and 15×15 increasing domain grids. Some exceptions for the 15×15 grids were observed for the $\theta = 60°$, $\lambda = 3$ simulations.

6. A RE-ANALYSIS OF IRON-ORE RESIDUALS

Previously, Cressie (1985, 1986) analyzed % Fe_2O_3 measurements collected from an ore deposit in Australia. The goal of the analysis was prediction of % Fe_2O_3 levels at unsampled locations within the deposit. Two key components in applying kriging to spatial data are the characterizations of trend and dependence. For the iron-ore data, Cressie (1986) used median polish to estimate the trend, revealing spatially dependent residuals. The original data can be found in Cressie (1986). The detrended iron-ore residuals are shown here in Figure 2. Cressie (1986) determined that these iron-ore residuals exhibited a geometric anisotropy which was accounted for by doubling the scale in the north-south direction and modeled using a spherical semivariogram.

Let $\{z(\mathbf{s}_1), \ldots, z(\mathbf{s}_{112})\}$ represent the detrended iron ore residuals shown in Figure 2. We considered the following exponential semivariogram model

$$\gamma(\mathbf{s}_i, \mathbf{s}_i; \mathbf{A}(\theta, \lambda), c_0, \sigma^2, \rho) = c_0 + \sigma^2 \left(1 - \rho^{\|\mathbf{A}(\mathbf{s}_i - \mathbf{s}_j)\|}\right), \qquad (6.1)$$

which is consistent with that proposed in Zimmerman and Zimmerman (1991), who also used the residuals in Figure 2 to demonstrate common variogram estimation procedures available at that time. The matrix of geometric anisotropy was parameterized as in (5.2). For pragmatic reasons we rescaled the original data locations to a 1 meter × 1 meter grid so that possible strong levels of spatial dependence were within our working parameter space $\rho \in [.01, .99]$. It is unclear whether previous authors analyzing this same data performed similar rescaling. Consequently, since ρ is spatial scale dependent, the actual CL estimated value of ρ can not be accurately compared to estimates obtained

↑ N

		2.14	-3.90	1.29	0.73	-2.60	0.62	-0.62	-6.46							
		0.40	0.66	1.45	-1.01	-0.04	1.48	0.04	-1.50	-8.28	-7.45					
		0.58	5.04	-0.27	4.17	0.04	4.26	1.72	1.78	0.00	-2.47	-0.10	-0.80	-1.62	-3.02	-5.34
		-0.18	2.61	-1.42	2.52	1.99	-3.99	3.87	1.63	2.15	0.18	3.05	-2.15	-1.28	-0.18	
		0.18	-0.66	-0.67	-5.73	1.54	0.66	-4.28	-1.12	0.00	-4.57	1.40	0.00	1.28	0.18	
-1.43	-2.48	-0.78	-2.22	0.27	1.71	-3.42	-6.90	-0.04	1.12	2.04	0.07	2.64	1.24	2.82	-2.08	0.00
1.43	2.48	-1.52	2.84	-4.37	-0.73	0.74	-4.74	-7.38	1.78	-1.60	0.33	0.00	0.00	-5.02	1.88	5.26
		-0.90	2.66	1.05	-3.21	-0.04	-0.62	0.04	-1.40	-0.18	1.15	-1.98	4.32	5.90	5.60	
										2.60	-0.07	0.00				

Figure 2. *Iron-ore median polished residuals. Grid spacings are 50 meters by 50 meters.*

from the other studies. However, and more importantly, the fitted semivariograms and corresponding predictive performances can be accurately compared.

Table 3 contains composite likelihood parameter estimates obtained by minimizing (4.3) using $\gamma(\cdot)$ as defined above in (6.1). Also included are estimates using various other procedures obtained from Zimmerman and Zimmerman (1991) who, following the advice in Cressie (1986), predetermined that $\mathbf{A}(\widehat{\theta},\ \widehat{\lambda}) = \mathbf{A}(90°,\ 2)$. The composite likelihood approach yielded estimates of geometric anisotropy ($\widehat{\theta} = 89.994°$, $\widehat{\lambda} = 2.270$) similar to those found in Cressie (1986). To graphically gauge the fit of these semivariogram estimates, we display in Figure 3 the fitted semivariogram models based on CL, WLS MM, and REML. Each estimated isotropic semivariogram is plotted through the moment estimator (3.1) after first deforming the sampling space by $\widehat{\theta}$ and $\widehat{\lambda}$ to account for the geometric anisotropy. Note the composite likelihood fitted semivariogram model is shown separately since the plotted moment estimator differs slightly from that for the WLS MM and REML approachs due to the different estimates of geometric anisotropy.

Predictive performance of the various estimated semivariogram models were evalu-

Table 3. Composite likelihood (CL) semivariogram parameter estimates for the iron-ore residuals. Also included are estimates using other procedures obtained from Zimmerman and Zimmerman (1991); WLS with the moment semivariogram estimator, WLS with the robust semivariogram estimator of Cressie (1985), MLE, REML, and generalized MIVQ of Kitanidis (1983). The θ and λ parameters for these other procedures were estimated a priori and fixed at 90° and 2, respectively.

	Parameter estimates				
Method	ρ	σ^2	c_0	θ	λ
WLS MME	.703	3.618	4.888	90°	2
WLS Robust	.770	2.783	5.098	90°	2
MLE	.700	3.090	5.147	90°	2
REML	.895	6.382	5.336	90°	2
GMIVQ	.866	3.734	5.630	90°	2
CL $\mathbf{A}(\theta, \lambda)$	.571	3.801	4.193	89.994°	2.270

ated using the cross-validation statistics

$$\mathrm{CV}_1 = \frac{1}{112}\sum_{i=1}^{112}\{(z_i(\mathbf{s}_i) - z^*_{-i}(\mathbf{s}_i))/\sigma_{-i}(\mathbf{s}_i)\}$$

$$\mathrm{CV}_2 = \left\{\frac{1}{112}\sum_{i=1}^{112}\{(z_i(\mathbf{s}_i) - z^*_{-i}(\mathbf{s}_i))/\sigma_{-i}(\mathbf{s}_i)\}^2\right\}^{1/2}$$

$$\mathrm{CV}_3 = \left\{\frac{1}{112}\sum_{i=1}^{112}(z_i(\mathbf{s}_i) - z^*_{-i}(\mathbf{s}_i))^2\right\}^{1/2},$$

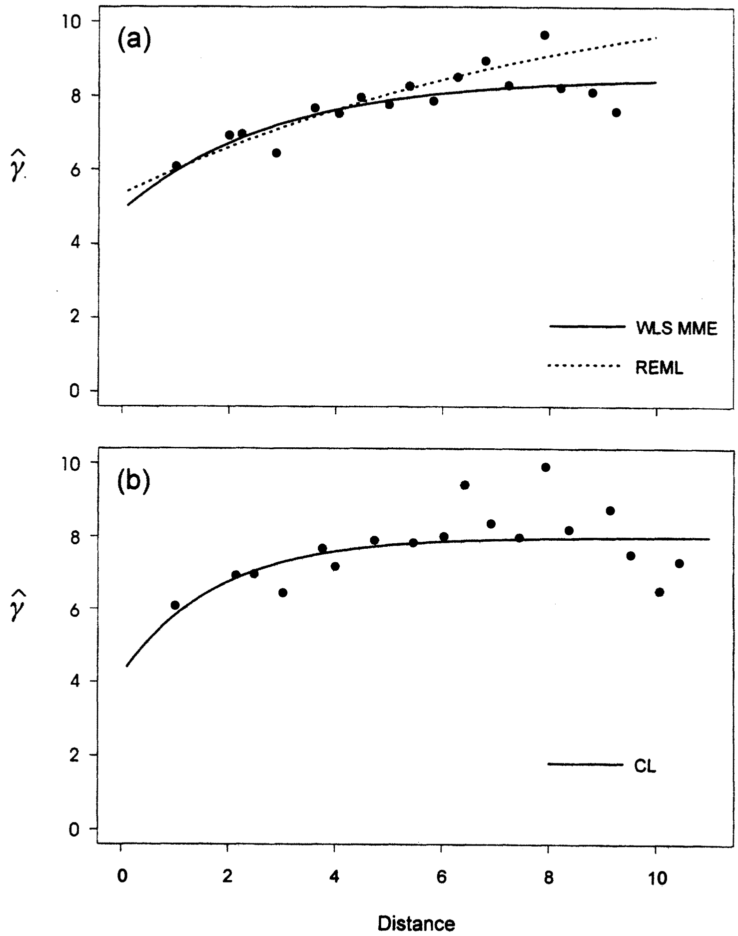

Figure 3. Fitted semivariogram models for the iron-ore residuals from (a) WLS with MME and REML, and (b) composite likelihood using $\mathbf{A}(\theta, \lambda)$. The semivariogram estimates were based on the moment estimator after first deforming the sampling space according to $\widehat{\theta}$ and $\widehat{\lambda}$ in Table 2.

Table 4. Cross-validation kriging results for the iron-ore residuals using the semivariogram estimates listed in Table 2. Included are the results from the estimated spherical model suggested in Cressie (1986). Also listed is the negative log-composite likelihood (CL) evaluated at the corresponding set of parameter estimates. A value using estimates from Cressie (1986) is not applicable (NA) for comparison since a different semivariogram model was used.

Source	*CL*	CV_1	CV_2	CV_3
Cressie	NA	−.000659	1.009542	2.596724
WLS MME	18979.17	−.000507	1.004833	2.632716
WLS Robust	18989.40	−.000499	1.033206	2.625841
MLE	18975.39	−.000448	1.001081	2.633959
REML	19045.87	−.000704	.999227	2.603348
GMIVQ	18987.39	−.000556	1.001254	2.615581
CL $\mathbf{A}(\theta, \lambda)$	18973.45	−.000348	1.015207	2.656975

where $z^*_{-i}(\mathbf{s}_i)$ and σ^2_{-i} denotes, respectively, the kriged prediction and kriged prediction variance at location $\mathbf{s}_i$ using the iron-ore residuals with observation $z(\mathbf{s}_i)$ deleted. The results are listed in Table 4. By construction CV_1 and CV_2 should be close to 0 and 1, respectively, and small values of CV_3 are preferred. Performance of the composite likelihood approach is similar to those based on the other methods. Also listed in Table 4 are the negative log-composite likelihood (CL) values obtained by evaluating (4.3) using the model in (6.1) and corresponding parameter estimates in Table 2. Not surprisingly, the composite likelihood parameter estimates obtained by minimizing (4.3) yielded the smallest CL value. A spherical semivariogram model was used in Cressie (1986) and thus is not applicable for comparison with the other CL values based on the exponential semivariogram.

Remark 2. *Similar to other semivariogram model fitting routines, the composite likelihood approach exhibited some sensitivity to starting values when used with the iron-ore residuals. When parameter estimates vary with starting values, one can use the moment semivariogram estimator (3.1), after proper deformation, along with the estimated model to gauge the fit. Cross-validation results can also be used to help evaluate performance. These techniques were applied with various sets of starting values when analyzing the iron-ore residuals. Although semivariogram parameter estimates varied with starting values, the cross-validation prediction statistics were reasonably stable. Alternatively, although not impelmented here, the minimax approach suggested in Li (1996, 1997) may provide a way to choose a particular solution.*

7. DISCUSSION

This article presented the composite likelihood approach to semivariogram estimation as an approximate likelihood based method combining many positive features of existing techniques based on method of moments, maximum likelihood, and restricted maximum likelihood. The main advantages of composite likelihood semivariogram estimation are that the procedure is objective, eliminating the distance or distance/direction binning often required for method of moments estimation in the isotropic and geometric anisotropic situations; it is computationally feasible, requiring no matrix inversions; and the procedure is statistically sensible, leading to consistent estimators without strong distributional assumptions. Predictions based on composite likelihood estimators worked very well under the isotropy and geometric anisotropy conditions we considered. The composite likelihood approach is also very flexible. Extensions to geometric anisotropy in three dimensions is straightforward, whereas the graphical procedures based on the moment estimator become even more subjective and questionable. In addition to the remarks in Section 4.3, the application of composite likelihood to universal and intrinsic random function kriging (Cressie 1991) and Cox point processes (Cox and Isham 1980) are currently being explored.

ACKNOWLEDGMENTS

Frank C. Curriero's work was partially supported by a grant from USDA CSREES Competitive Grants Program (95-37302-1905, PIs Michael E. Hohn and Andrew M. Liebhold) and by the U.S. Environmental Protection Agency's Climate Policy and Assessment Division, cooperative agreement CR823143 (Jonathan A. Patz, PI). Subhash Lele's work was partially supported by a grant from DOE (DE-FC07-94, Professor Daniel Goodman PI) and by the U.S. Environmental Protection Agency's Climate Policy and Assessment Division, cooperative agreement CR823143 (Jonathan A. Patz, PI). The authors thank the associate editor and referees for their many helpful comments.

[Received July 1997. Revised August 1998.]

REFERENCES

Barry, J., Crowder, M., and Diggle, P. (1998), "Parametric Estimation of the Variogram," *Mathematical Geology* (submitted).

Besag, J. E. (1975), "Statistical Analysis of Non-Lattice Data," *The Statistician*, 24, 179–195.

Cox, D. R. (1975), "Partial Likelihood," *Biometrika*, 62, 269–276.

Cox, D. R., and Isham, V. (1980), *Point Processes*, New York: Chapman and Hall.

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

——— (1986), "Kriging Nonstationary Data," *Journal of the American Statistical Association*, 81, 625–634.

——— (1991), *Statistics for Spatial Data*, New York: Wiley.

Crowder, M. J. (1986), "On Consistency and Inconsistency of Estimating Equations," *Econometric Theory*, 2, 305–330.

Dennis, B., and Taper, M. L. (1994), "Density Dependence in Time Series Observations of Natural Populations: Estimation and Testing," *Ecological Monographs*, 64, 205–224.

Godambe, V. P. (1960), "An Optimum Property of Regular Maximum Likelihood Estimation," *Annals of Mathematical Statistics*, 31, 1208–1212.

Godambe, V. P., and Kale, B. K. (1991), "Estimating Functions: An Overview," in *Estimating Functions*, ed. V. P. Godambe, London: Oxford University Press, pp. 3–20.

Godambe, V. P., and Thompson, M. E. (1984), "Robust Estimation Through Estimating Equations," *Biometrika*, 71, 115–125.

Haining, R. (1990), *Spatial Data Analysis in the Social and Environmental Sciences*, New York: Cambridge University Press.

Heagerty, P. J., and Lele, S. R. (1998), "A Composite Likelihood Approach to Binary Data in Space," *Journal of the American Statistical Association*, 93, 1099–1111.

Hobert, J. P., Altman, N. S., and Schofield, C. L. (1997), "Analyses of Fish Species Richness with Spatial Covariate," *Journal of the American Statistical Association*, 92, 846–854.

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

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

Kalunzy, S. P., Vega, S. C., Cardoso, T. P., and Shelly, A. A. (1996), *S+SPATIALSTATS, User's Manual*, Seattle, WA: MathSoft.

Lele, S. R. (1997), "Estimating Functions for Semivariogram Estimation," in *Selected Proceedings of the Conference on Estimating Functions*, eds. I. V. Basawa, V. P. Godambe, and R. L. Taylor, Haywood, CA: Institute of Mathematical Statistics, pp. 381–396.

Li, B. (1996), "A Minimax Approach to Consistency and Efficiency for Estimating Equations," *The Annals of Statistics*, 24, 1283–1297.

——— (1997), "On Consistency of Generalized estimating Equations," in *Selected Proceedings of the Conference on Estimating Functions*, eds. I. V. Basawa, V. P. Godambe, and R. L. Taylor, Haywood, CA: Institute of Mathematical Statistics, pp. 115–136.

Liang, K-Y. (1987), "Extended Mantel-Haenszel Estimating Procedure for Multivariate Logistic Regression Models," *Biometrics*, 43, 289–299.

Liang, K-Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.

Lindsay, B. G. (1988), "Composite Likelihood Methods," *Contemporary Mathematics*, 80, 221–239.

——— (1994), "Efficiency Versus Robustness: The case for Minimum Hellinger Distance and Related Methods," *The Annals of Statistics*, 22, 1081–1114.

Matheron, G. (1962), *Traite et Geostatistique Appliquee (Tome I), Memoires du Bureau de Recherches Geologiques et Minieres*, No. 14, Paris: Editions Technip.

MathSoft, Inc. (1995), *S-PLUS User's Manual* (ver. 3.3 for Windows), Seattle, WA: MathSoft.

Myers, D. E., Begovich, C. L., Butz, T. R., and Kane, V. E. (1982), "Variogram Models for Regional Groundwater Geochemical Data," *Mathematical Geology*, 14, 629–644.

Patil, G. P., and Rao, C. R. (1978), "Weighted Distributions and Size-Biased Sampling with Applications to Wildlife Populations and Human Families," *Biometrics*, 34, 179–189.

Ripley, B. D. (1987), *Stochastic Simulation*, New York: Wiley.

SAS Institute, Inc. (1990), *SAS/STAT User's Guide* (ver. 6), North Carolina: SAS Institute.

Vecchia, A. V. (1988), "Estimation and Model Identification for Continuous Spatial Processes," *Journal of the Royal Statistical Society*, Ser. B, 50, 297–312.

Wald, A. (1949), "Note on the Consistency of Maximum Likelihood Estimate," *The Annals of Mathematical Statistics*, 20, 595–601.

Warnes, J. J., and Ripley, B. D. (1987), "Problems with Likelihood Estimation of Covariance Functions of Spatial Gaussian Processes," *Biometrika*, 74, 640–642.

Webster, R., and Oliver, M. A. (1992), "Sample Adequately to Estimate Variograms of Soil Properties," *Journal of Soil Science*, 43, 177–192.

Zimmerman, D. L., and Zimmerman, M. B. (1991), "A Comparison of Spatial Semivariogram Estimators and Corresponding Ordinary Kriging Predictors," *Technometrics*, 33, 77–91.