*Mathematical Geology, Vol. 36, No. 3, April 2004 (© 2004)*

# Fitting the Linear Model of Coregionalization by Generalized Least Squares[1]

**Bernard Pelletier,[2,3] Pierre Dutilleul,[3] Guillaume Larocque,[2] and James W. Fyles[2]**

*In geostatistical studies, the fitting of the linear model of coregionalization (LMC) to direct and cross experimental semivariograms is usually performed with a weighted least-squares (WLS) procedure based on the number of pairs of observations at each lag. So far, no study has investigated the efficiency of other least-squares procedures, such as ordinary least squares (OLS), generalized least squares (GLS), and WLS with other weighing functions, in the context of the LMC. In this article, we compare the statistical properties of the sill estimators obtained with eight least-squares procedures for fitting the LMC: OLS, four WLS, and three GLS. The WLS procedures are based on approximations of the variance of semivariogram estimates at each distance lag. The GLS procedures use a variance–covariance matrix of semivariogram estimates that is (i) estimated using the fourth-order moments with sill estimates ($GLS_1$), (ii) calculated using the fourth-order moments with the theoretical sills ($GLS_2$), and (iii) based on an approximation using the correlation between semivariogram estimates in the case of spatial independence of the observations ($GLS_3$). The current algorithm for fitting the LMC by WLS while ensuring the positive semidefiniteness of sill matrix estimates is modified to include any least-squares procedure. A Monte Carlo study is performed for 16 scenarios corresponding to different combinations of the number of variables, number of spatial structures, values of ranges, and scale dependence of the correlations among variables. Simulation results show that the mean square error is accounted for mostly by the variance of the sill estimators instead of their squared bias. Overall, the estimated $GLS_1$ and theoretical $GLS_2$ are the most efficient, followed by the WLS procedure that is based on the number of pairs of observations and the average distance at each lag. On that basis, $GLS_1$ can be recommended for future studies using the LMC.*

**KEY WORDS:** direct and cross semivariograms, empirical variance and bias, fourth-order moments, multivariate nested semivariogram model, positive semidefiniteness, sill estimators.

## INTRODUCTION

In geostatistical studies, the linear model of coregionalization (LMC) is fitted to direct and cross experimental semivariograms evaluated from multivariate

---

[1]Received 6 March 2003; accepted 24 November 2003.

[2]Department of Natural Resource Sciences, McGill University, Macdonald Campus, Ste-Anne-de-Bellevue, Québec, Canada H9X 3V9; e-mail: bernard.pelletier@mail.mcgill.ca

[3]Department of Plant Science, McGill University, Macdonald Campus, Ste-Anne-de-Bellevue, Québec, Canada H9X 3V9.



0882-8121/04/0400-0323/1 

spatial data collected for $p$ random variables. The particularity of the LMC is that all $p(p+1)/2$ semivariograms are modeled as linear combinations of the same set of $S$ basic semivariogram functions (Journel and Huijbregts, 1978, p. 172). The LMC can be written as a multivariate nested semivariogram model:

$$\boldsymbol{\Gamma}(\mathbf{h}) = \sum_{s=1}^{S} \mathbf{B}_s g_s(\mathbf{h}) \quad (1)$$

where $\boldsymbol{\Gamma}(\mathbf{h})$ is the $p \times p$ matrix of semivariogram values at lag $\mathbf{h}$ and $\mathbf{B}_s$ is the $p \times p$ matrix of sills of the basic semivariogram function $g_s(\mathbf{h})$, which denotes any permissible semivariogram model. The $S$ basic semivariogram functions correspond to spatial structures or processes with different ranges, including the nugget effect. For the LMC to be permissible, each matrix $\mathbf{B}_s$ is constrained to be positive semidefinite. This is a necessary condition for the positive semidefiniteness of the variance–covariance matrix in the system of cokriging equations (Goovaerts, 1994; Journel and Huijbregts, 1978) and for the valid utilization of sill matrices as variance–covariance matrices in multivariate analyses (Wackernagel, 1995, chap. 27; Wackernagel, Petitgas, and Touffait, 1989).

An iterative algorithm was proposed by Goulard (1989) for fitting the LMC by weighted least squares (WLS), while ensuring the positive semidefiniteness of sill matrix estimates. In the application of the LMC algorithm described in Goulard and Voltz (1992), the weighing function involved in the WLS fitting is $N(\mathbf{h})$, the number of pairs of observations used to estimate the semivariogram values at each lag. Accordingly, most studies based on the LMC have used $N(\mathbf{h})$ as the WLS weighing function (Castrignanò and others, 2000a,b; Dobermann, Goovaerts, and George, 1995; Dobermann, Goovaerts, and Neue, 1997; Goovaerts, Sonnet, and Navarre, 1993; Goovaerts and Webster, 1994; Voltz and Goulard, 1994; Webster, Atteia, and Dubois, 1994). To our knowledge, however, the statistical properties of that WLS estimator of sills have not been thoughtfully investigated in the context of the LMC, and have not been compared to those of other estimators obtained by ordinary least squares (OLS), generalized least squares (GLS), or WLS based on other weighing functions in this context.

Studies comparing different least-squares estimation procedures have been conducted primarily for the fitting of a single direct semivariogram model (Cressie, 1985; Genton, 1998b; Müller, 1999; Pardo-Igúzquiza, 1999; Zhang, Van Eijkeren, and Heemink, 1995; Zimmerman and Zimmerman, 1991). In these studies, it is generally acknowledged that GLS should be better than OLS because the former takes into account the heteroscedasticity and correlation of semivariogram estimates at different lags (Cressie, 1985; Genton, 1998b; Jian, Olea, and Yu, 1996; McBratney and Webster, 1986; Müller, 1999; Zimmerman and Zimmerman, 1991). In the common situation where the number of model parameters to be estimated is smaller than the number of lags, Lahiri, Lee, and Cressie (2002) have demonstrated that the GLS estimator is asymptotically more efficient than the OLS and WLS estimators. In practice, however, many authors have preferred to work with WLS because GLS requires the estimation of the variance–covariance matrix of semivariogram estimates, which has been considered to be too computationally intensive until recently (Cressie, 1985; Jian, Olea, and Yu, 1996; McBratney and Webster, 1986). With the increase in computer power, the processing time has become less prohibitive and a number of recent studies have proposed GLS estimation procedures for fitting a single direct semivariogram model (Genton, 1998b; Müller, 1999; Pardo-Igúzquiza and Dowd, 2001).

The main objective of our study was to test the expectation that the GLS estimator of sills possesses better statistical properties than the OLS and WLS estimators in the context of the LMC. We proceeded by simulation and conducted a Monte Carlo study because of the effect that ensuring the positive semidefiniteness of sill matrix estimates was likely to have, and in fact had on the final individual sill estimates. This effect prevented the application of theoretical properties known for sill estimators in the case of the fitting of a single direct semivariogram model. In a preliminary step, we present a simple procedure for the estimation of the variance–covariance matrix of semivariogram estimates for Matheron's (1962) classical estimator. We have used the matrix notation extensively, which greatly simplifies the theoretical development. Thereafter, we modify the LMC fitting algorithm proposed by Goulard (1989) to make it more general and applicable to GLS or any other least-squares estimation procedure such as OLS and WLS. Recall here that in the framework of semivariogram model fitting, OLS assumes the homoscedasticity and independence of the semivariogram estimates, whereas WLS relaxes the homoscedasticity assumption while maintaining the independence condition on the semivariogram estimates.

## GENERALIZED LEAST SQUARES FOR THE LINEAR MODEL OF COREGIONALIZATION

Consider a second-order stationary, $p$-variate spatial process $\{\mathbf{Z}(\mathbf{u}) \mid \mathbf{u} \in D\}$ with $\mathbf{Z}(\mathbf{u}) = (Z_1(\mathbf{u}), \ldots, Z_p(\mathbf{u}))^{\mathrm{T}}$ and $D \subset \mathrm{R}^d$, $d \geq 1$. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a set of $n$ sampling locations and $\gamma_{ij}^*(\mathbf{h})$ be the experimental semivariogram between the random fields $Z_i$ and $Z_j$ at lag $\mathbf{h}$; $\gamma_{ij}^*(\mathbf{h})$ is a direct semivariogram if $i = j$ and a cross semivariogram otherwise.

So far, the use of the GLS estimation method in semivariogram model fitting has been limited to a single direct semivariogram (i.e., $p = 1$ and $i = j$), where the vector of sills and the vector of ranges are the parameters to be estimated (Cressie, 1985; Genton, 1998b; Müller, 1999; Pardo-Igúzquiza and Dowd, 2001). In this case, the estimation of model parameters can be viewed as a nonlinear regression problem (Lahiri, Lee, and Cressie, 2002, p. 77), based on the following model:

$$\boldsymbol{\gamma}_{ii}^{*} = \boldsymbol{\gamma}(\boldsymbol{\beta}_{ii}, \boldsymbol{\varphi}_{ii}) + \boldsymbol{\varepsilon}_{ii} \tag{2}$$

where $\boldsymbol{\gamma}_{ii}^{*} = (\gamma_{ii}^{*}(h_1), \ldots, \gamma_{ii}^{*}(h_K))^{\mathrm{T}}$ is the $K \times 1$ observable random vector of the direct experimental semivariogram at distance lags $\{h_1, \ldots, h_K\}$, $\boldsymbol{\varphi}(\boldsymbol{\beta}_{ii}, \boldsymbol{\varphi}_{ii}) = (\gamma(h_1, \boldsymbol{\beta}_{ii}, \boldsymbol{\varphi}_{ii}), \ldots, \gamma(h_K, \boldsymbol{\beta}_{ii}, \boldsymbol{\varphi}_{ii}))^{\mathrm{T}}$ is the $K \times 1$ vector of values of the semivariogram model at distance lags $\{h_1, \ldots, h_K\}$ for the $S \times 1$ vector of ranges $\boldsymbol{\varphi}_{ii}$ and the corresponding $S \times 1$ vector of sills $\boldsymbol{\beta}_{ii}$, and $\boldsymbol{\varepsilon}_{ii}$ is the $K \times 1$ unobservable random vector of errors. The semivariogram model is defined by

$$\boldsymbol{\gamma}(\boldsymbol{\beta}_{ii}, \boldsymbol{\varphi}_{ii}) = \mathbf{G}(\boldsymbol{\varphi}_{ii})\boldsymbol{\beta}_{ii} \tag{3}$$

where $\mathbf{G}(\boldsymbol{\varphi}_{ii})$ is the $K \times S$ matrix of values of the basic semivariogram functions for the vector of ranges $\boldsymbol{\varphi}_{ii}$. The variance–covariance matrix of the random vector of errors $\boldsymbol{\varepsilon}_{ii}$ is then a function of both the sill and the range parameters:

$$\mathrm{Var}(\boldsymbol{\varepsilon}_{ii}) = \boldsymbol{\Sigma}(\boldsymbol{\beta}_{ii}, \boldsymbol{\varphi}_{ii}) \tag{4}$$

and the GLS fitting procedure consists in minimizing with respect to $\boldsymbol{\beta}_{ii}$ and $\boldsymbol{\varphi}_{ii}$ the function

$$\mathrm{WSS}\,(\boldsymbol{\beta}_{ii}, \boldsymbol{\varphi}_{ii}) = (\boldsymbol{\gamma}_{ii}^{*} - \mathbf{G}(\boldsymbol{\varphi}_{ii})\boldsymbol{\beta}_{ii})^{\mathrm{T}}\, \boldsymbol{\Sigma}(\boldsymbol{\beta}_{ii}, \boldsymbol{\varphi}_{ii})^{-1}(\boldsymbol{\gamma}_{ii}^{*} - \mathbf{G}(\boldsymbol{\varphi}_{ii})\boldsymbol{\beta}_{ii}) \tag{5}$$

In the LMC, the $p$ direct semivariograms and the $p(p-1)/2$ cross semivariograms need to be modeled simultaneously, using the same $S$ basic semivariogram functions. It follows that the vector of ranges $\boldsymbol{\varphi}$ is the same for all semivariograms and that the sills become the only model parameters to be estimated.

The vector of ranges $\boldsymbol{\varphi}$ can be identified in different ways. It can be fixed by the experimenter on the basis of expert knowledge, ancillary information, or research hypotheses about the system under study (Goovaerts, 1997, p. 106). Different combinations of the number of spatial structures, type of basic semivariogram functions (e.g., spherical, exponential, Gaussian), and value of ranges can be compared on the basis of a weighted sum of squares (Goulard and Voltz, 1992). Semivariogram model fitting can also be viewed as a least-squares estimation problem for nonlinear models whose parameters separate (Golub and Pereyra, 1973; Guttman, Pereyra, and Scornik, 1973; Osborne, 1970). Accordingly, the ranges (i.e., the nonlinear parameters) can be estimated by least squares prior to the sills (i.e., the linear parameters).

For a given vector of ranges $\varphi$, the problem of estimating the sills for any semivariogram in the LMC becomes a problem of linear regression:

$$\gamma_{ij}^{*} = \mathbf{G}(\varphi)\boldsymbol{\beta}_{ij} + \varepsilon_{ij} \tag{6}$$

where $\gamma_{ij}^{*}$ can represent a direct or a cross experimental semivariogram depending on the values of $i$ and $j$, and $\mathbf{G}(\varphi)$ is the $K \times S$ matrix of values of the basic semivariogram functions. In this case, the variance–covariance matrix of the random errors is a function of the sills to be estimated and $\varphi$:

$$\text{Var}(\varepsilon_{ij}) = \boldsymbol{\Sigma}(\boldsymbol{\beta}_{ij}; \varphi) \tag{7}$$

The GLS fitting procedure for the LMC consists in minimizing with respect to $\boldsymbol{\beta}_{ij}$ the function

$$\text{WSS}(\boldsymbol{\beta}_{ij}) = (\gamma_{ij}^{*} - \mathbf{G}(\varphi)\boldsymbol{\beta}_{ij})^{\text{T}}\, \boldsymbol{\Sigma}(\boldsymbol{\beta}_{ij}; \varphi)^{-1}(\gamma_{ij}^{*} - \mathbf{G}(\varphi)\boldsymbol{\beta}_{ij}) \tag{8}$$

for $i,\ j = 1, \ldots, p$. This is equivalent to minimizing

$$\text{WSS}(\mathbf{B}) = \sum_{i=1}^{p}\sum_{j=1}^{p} \text{WSS}(\boldsymbol{\beta}_{ij}) \tag{9}$$

with respect to $\mathbf{B}$ whose entry $(i, j)$ is $\boldsymbol{\beta}_{ij}$ $(i,\ j = 1, \ldots, p)$. Note that each cross semivariogram is counted twice in Equation (9) in order to have a criterion for a Euclidean-like distance. Since the rest of this article is about GLS for the LMC where the ranges are assumed to be known, from now on the symbol $\varphi$ will be dropped from formulas for notational simplicity.

Because the variance–covariance matrix of the experimental semivariogram in Equation (8), now denoted by $\boldsymbol{\Sigma}(\boldsymbol{\beta}_{ij})$, is itself a function of the sill parameters to be estimated, the GLS fitting procedure is iterative (Cressie, 1985; Müller, 1999). Accordingly, the sill estimates at step $\tau + 1$ in this procedure are calculated as follows:

$$\hat{\boldsymbol{\beta}}_{ij}^{\tau+1} = \left(\mathbf{G}^{\text{T}}\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\beta}}_{ij}^{\tau}\right)^{-1}\mathbf{G}\right)^{-1}\mathbf{G}^{\text{T}}\,\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\beta}}_{ij}^{\tau}\right)^{-1}\gamma_{ij}^{*} \tag{10}$$

until successive sill estimates provide criterion values that differ by less than a predetermined (usually very small) quantity, which means that a minimum of $\text{WSS}(\boldsymbol{\beta}_{ij})$ is reached. The OLS estimates or any of the WLS estimates of sills provide a suitable initial solution (i.e., $\tau = 0$). The estimation of $\boldsymbol{\Sigma}(\boldsymbol{\beta}_{ij})$ is a key point in the iterative GLS procedure. The development below is for Matheron's classical semivariogram estimator. We discuss the cases of other, more robust estimators later in this article.

In its most general form, Matheron's estimator can be expressed as a bilinear form

$$\gamma_{ij}^{*}(\mathbf{h}) = \mathbf{z}_i^{\mathrm{T}}\mathbf{A}(\mathbf{h})\mathbf{z}_j \tag{11}$$

where $\mathbf{z} = (Z(\mathbf{u}_1), \ldots, Z(\mathbf{u}_n))^{\mathrm{T}}$ is the data vector for the random field $Z$ and $\mathbf{A}(\mathbf{h})$ is the spatial design matrix described in the Appendix (Eq. (A1)). Under the Gaussian assumption, the entry $(h_k, h_{k'})$ of $\boldsymbol{\Sigma}(\boldsymbol{\beta}_{ij})$ can be calculated as the covariance between two bilinear forms (Searle, 1971, p. 66):

$$\begin{aligned}\mathrm{Cov}(\gamma_{ij}^{*}(h_k), \gamma_{ij}^{*}(h_{k'})) &= \mathrm{tr}(\mathbf{A}(h_k)\mathbf{C}(\boldsymbol{\beta}_{ij})\mathbf{A}(h_{k'})\mathbf{C}(\boldsymbol{\beta}_{ij})) \\ &\quad + \mathrm{tr}(\mathbf{A}(h_k)\mathbf{C}(\boldsymbol{\beta}_{ii})\mathbf{A}(h_{k'})\mathbf{C}(\boldsymbol{\beta}_{jj}))\end{aligned} \tag{12}$$

where $\mathbf{C}(\boldsymbol{\beta}_{ij})$ is the $n \times n$ matrix of covariances between $\mathbf{z}_i$ and $\mathbf{z}_j$, while $\mathbf{C}(\boldsymbol{\beta}_{ii})$ and $\mathbf{C}(\boldsymbol{\beta}_{jj})$ are the $n \times n$ variance–covariance matrices of $\mathbf{z}_i$ and $\mathbf{z}_j$, respectively. The variance of the semivariogram estimator at distance lag $h$ follows from $h_k = h_{k'} = h$ in (11). If $i = j$, the experimental semivariogram is direct and its estimator is expressed as a quadratic form, so that Equation (12) becomes:

$$\mathrm{Cov}(\gamma_{ii}^{*}(h_k), \gamma_{ii}^{*}(h_{k'})) = 2\,\mathrm{tr}(\mathbf{A}(h_k)\mathbf{C}(\boldsymbol{\beta}_{ii})\mathbf{A}(h_{k'})\mathbf{C}(\boldsymbol{\beta}_{ii})) \tag{13}$$

which provides a much simpler expression for the evaluation of the fourth-order moments described in Pardo-Igúzquiza and Dowd (2001) and Ortiz and Deutsch (2002). A modification of Equation (12) that accelerates the process of evaluation is given in the Appendix. In practice (see Eq. (10)), an estimate $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}_{ij})$ of the variance–covariance matrix of the semivariogram estimator is obtained by substituting sill estimates $\hat{\boldsymbol{\beta}}_{ij}$ for $\boldsymbol{\beta}_{ij}$ in Equation (12).

## GLS FITTING ALGORITHM FOR THE LMC

The separate estimation of sill vectors $\boldsymbol{\beta}_{ij}$ $(i, j = 1, \ldots, p)$ by iterative GLS (see above) does not ensure the positive semidefiniteness of the $p \times p$ sill matrix estimates $\hat{\mathbf{B}}_s = (\hat{\beta}_{ij,s})$ where $s = 1, \ldots, S$ and $\hat{\beta}_{ij,s}$ denotes the GLS estimate of the sill for structure $s$ in semivariogram $(i, j)$. The algorithm proposed by Goulard (1989) to fit the LMC while ensuring that the sill matrix estimates are positive semidefinite was described for WLS using model-free weighing functions such as $N(h)$, the number of pairs of observations used to evaluate the semivariogram estimates at distance lag $h$. In this section, the algorithm is modified to permit the use of any least-squares procedure, including those (e.g., Cressie's

WLS, GLS) for which a model-based metric is used in the minimization of the objective function. The GLS fitting algorithm for the LMC follows the same basic steps as Goulard's (1989) algorithm and hence, some details about these steps are omitted below, while the differences between the algorithms are highlighted.

*Step 0*—Initialize $\tau$ to 0 and evaluate $\mathrm{WSS}(\hat{\mathbf{B}}^{\tau})$. Sills estimated by OLS or GLS are examples of initial values for $\hat{\mathbf{B}}_s$ $(s = 1, \ldots, S)$; the latter are obtained at the end of the iterative procedure defined by Equation (10). Eventually, they both provide the same results (not reported here).

*Step 1*—Among the $S$ spatial structures, select structure $s_0$ and subtract from the experimental semivariograms the part that is modeled by the other $S - 1$ structures:

$$\mathbf{\Gamma}^*_{s_0}(h_k) = \mathbf{\Gamma}^*(h_k) - \sum_{s \neq s_0} \hat{\mathbf{B}}^{\tau}_s g_s(h_k), \quad k = 1, \ldots, K \tag{14}$$

where $\mathbf{\Gamma}^*(h_k)$ is the matrix of total experimental semivariograms $\{\gamma^*_{ij}(h_k);\ i, j = 1, \ldots, p\}$ at distance lag $h_k$.

*Step 2*—Fit a model to each $\gamma^*_{ij,s_0}$ individually:

$$\hat{\beta}^{\tau+1}_{ij,s_0} = \left(\mathbf{g}^{\mathrm{T}}_{s_0} \mathbf{\Sigma}\left(\hat{\boldsymbol{\beta}}^{\tau}_{ij}\right)^{-1} \mathbf{g}_{s_0}\right)^{-1} \mathbf{g}^{\mathrm{T}}_{s_0} \mathbf{\Sigma}\left(\hat{\boldsymbol{\beta}}^{\tau}_{ij}\right)^{-1} \gamma^*_{ij,s_0}, \quad i, j = 1, \ldots, p \tag{15}$$

where $\mathbf{g}_{s_0}$ is the $K \times 1$ vector of values of the basic semivariogram function $g_{s_0}(h_k)$ for spatial structure $s_0$ at $h_k$ $(k = 1, \ldots, K)$.

*Step 3*—Perform the spectral decomposition of the $p \times p$ matrix $(\hat{\beta}^{\tau+1}_{ij,s_0})$ to obtain $\mathbf{Q}\,\mathbf{\Lambda}\,\mathbf{Q}^{\mathrm{T}}$, where $\mathbf{Q}$ contains the eigenvectors and $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues. Any element of $\mathbf{\Lambda}$ that is negative is then set at zero to obtain $\mathbf{\Lambda}^+$ and the matrix of sill estimates $\hat{\mathbf{B}}^{\tau}_{s_0}$ is replaced by $\mathbf{Q}\,\mathbf{\Lambda}^+\mathbf{Q}^{\mathrm{T}}$. The vector $\hat{\boldsymbol{\beta}}^{\tau}_{ij}$ in $\Sigma(\hat{\boldsymbol{\beta}}^{\tau}_{ij})$ is updated by replacing its $s_0$th component by the entry $(i, j)$ of $\mathbf{Q}\,\mathbf{\Lambda}^+\mathbf{Q}^{\mathrm{T}}$.

Repeat Steps 1–3 until the $S$ structures have been completed once.

*Step 4*—Compute $\mathrm{WSS}(\hat{\mathbf{B}}^{\tau+1})$. If WSS $(\hat{\mathbf{B}}^{\tau+1})$ differs from $\mathrm{WSS}(\hat{\mathbf{B}}^{\tau})$ by less than a predetermined, very small quantity, then stop. Otherwise, increment $\tau$ by 1 (i.e., $\tau = \tau + 1$) and go to Step 1.

The check for positive semidefiniteness in Step 3 above differs from the one in Goulard (1989) in that it is performed on $(\hat{\beta}^{\tau+1}_{ij,s_0})$ instead of the $p \times p$ matrix with entry $(i, j) = \mathbf{g}^{\mathrm{T}}_{s_0} \mathbf{\Sigma}(\hat{\boldsymbol{\beta}}^{\tau}_{ij})^{-1} \gamma^*_{ij,s_0}$. This difference is due to the fact that Goulard (1989) uses only one vector of weights $\{w(h_k); k = 1, \ldots, K\}$ for the WLS fitting of all semivariograms instead of $p(p + 1)/2$ matrices $\mathbf{\Sigma}(\hat{\boldsymbol{\beta}}^{\tau}_{ij})^{-1}$. Consequently, the equation in Goulard (1989) that corresponds to Equation (15) here can be rewritten as

$$\hat{\beta}_{ij,s_0}^{\tau+1} = \left(\sum_{k=1}^{K} w(h_k)\, g_{s_0}(h_k)^2\right)^{-1} \sum_{k=1}^{K} w(h_k) g_{s_0}(h_k) \gamma_{ij,s_0}^{*}(h_k) \tag{16}$$

Because the first multiplicative factor on the right-hand side of Equation (16) is constant for all semivariograms, the check for positive semidefiniteness needs to be performed only on the second factor, as proposed by Goulard (1989). In the cases of model-based WLS and GLS, however, the check for positive semidefiniteness must be performed on the $p \times p$ matrix $(\hat{\beta}_{ij,s_0}^{\tau+1})$ because both factors on the right-hand side of Equation (15) are then expected to differ among semivariograms.

The algorithm presented in this section can also be used for OLS and WLS procedures. It suffices to replace matrices $\boldsymbol{\Sigma}(\boldsymbol{\beta}_{ij})$ $(i, j = 1, \ldots, p)$ by the identity matrix $\mathbf{I}_K$ for OLS, by the diagonal matrix $\mathbf{V} = \operatorname{diag}(v(h_1), \ldots, v(h_K))$ for model-free WLS procedures (e.g., $v(h_k) = N(h_k)^{-1}$ $(k = 1, \ldots, K)$) and by diagonal matrices $\mathbf{V}(\boldsymbol{\beta}_{ij}) = \operatorname{diag}(v(h_1, \boldsymbol{\beta}_{ij}), \ldots, v(h_K, \boldsymbol{\beta}_{ij}))(i, j = 1, \ldots, p)$ for iterative model-based WLS procedures (e.g., Cressie, 1985). The diagonal entries of matrices $\mathbf{V}$ and $\mathbf{V}(\boldsymbol{\beta}_{ij})$ are measures of the uncertainty of semivariogram estimates at each lag. Note that the vector of weights $(w(h_1), \ldots, w(h_K))$ described for WLS by Pardo-Igúzquiza (1999) and Goulard and Voltz (1992), among others, corresponds to the diagonal of the inverse of $\mathbf{V}$ or $\mathbf{V}(\boldsymbol{\beta}_{ij})$ above.

## MONTE CARLO STUDY

A Monte Carlo study was performed to compare the statistical properties of eight least-squares estimators of sills in the context of the LMC: the OLS estimator plus four WLS and three GLS estimators. The weighing functions of the four WLS estimators are based on those presented in Pardo-Igúzquiza (1999) for fitting a single direct semivariogram model. The WLS procedures are defined by the following matrices whose diagonal entries follow from four approximations of the variance of direct semivariogram estimates at $K$ lags:

WLS$_1$: $\mathbf{V} = \operatorname{diag}(N(h_1)^{-1}, \ldots, N(h_K)^{-1})$, where $N(h_k)$ denotes the number of pairs of observations at distance lag $h_k$ (Goulard and Voltz, 1992);

WLS$_2$: $\mathbf{V} = \operatorname{diag}(\delta(h_1)^2/N(h_1), \ldots, \delta(h_K)^2/N(h_K))$, where $\delta(h_k)$ denotes the average distance for the $k$th lag (Zhang, Van Eijkeren, and Heemink, 1995);

WLS$_3$: $\mathbf{V}(\boldsymbol{\beta}_{ij}) = \operatorname{diag}(\gamma(h_1, \boldsymbol{\beta}_{ij})^2/N(h_1), \ldots, \gamma(h_K, \boldsymbol{\beta}_{ij})^2/N(h_K))(i, j = 1, \ldots, p)$, using Cressie's (1985) approximation for the variance of semivariogram estimates;

$\text{WLS}_4$: $\mathbf{V}(\boldsymbol{\beta}_{ij}) = \text{diag}(\{\gamma(h_1, \boldsymbol{\beta}_{ij})\}^2, \ldots, \{\gamma(h_K, \boldsymbol{\beta}_{ij})\}^2)(i, j = 1, \ldots, p)$, using the square of the modeled semivariogram ordinate as an approximation of the variance (Pardo-Igúzquiza, 1999).

The $\text{WLS}_3$ and $\text{WLS}_4$ procedures are model-based and iterative. Since the corresponding weighing functions were only discussed for direct semivariograms (Cressie, 1985; Pardo-Igúzquiza, 1999), it was necessary to adapt them for cross semivariograms (see the Appendix).

We studied the following three GLS procedures:

$\text{GLS}_1$: Iterative procedure based on the variance–covariance matrix estimates $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}_{ij})(i, j = 1, \ldots, p)$ obtained by using Equation (12) with sill estimates;

$\text{GLS}_2$: Procedure based on the true variance–covariance matrices $\boldsymbol{\Sigma}(\boldsymbol{\beta}_{ij})(i, j = 1, \ldots, p)$ calculated by using Equation (12) with the theoretical sills, for purposes of comparison with $\text{GLS}_1$ in which the $\boldsymbol{\beta}_{ij}$'s are estimated—$\text{GLS}_2$ is iterative only if any of the sill matrix estimates is not positive semidefinite;

$\text{GLS}_3$: Iterative procedure based on an approximation of variance–covariance matrices $\boldsymbol{\Sigma}(\boldsymbol{\beta}_{ij})(i, j = 1, \ldots, p)$ that uses the formula for the correlation between semivariogram estimates in the case of spatial independence of the observations (Genton, 1998b). The $\text{GLS}_3$ procedure follows from the suggestion made by Genton (1998b) and Furrer and Genton (1999) of using only the lags corresponding to the vertical and horizontal directions in a regular sampling grid. Technical details about $\text{GLS}_3$ are given in the Appendix.

All simulated data were generated on a $12 \times 12$ regular grid (i.e., $n = 144$). We considered 16 scenarios for our Monte Carlo study, corresponding to different combinations of four simulation parameters (Table 1):

(1) the number of spatial structures $S$ was 2 or 3, that is, a nugget effect representing the measurement error and spatial sources of variation at distances smaller than the shortest sampling distance (i.e., 1 unit) and one or two spherical semivariogram models;
(2) for scenarios with $S = 2$, three increasing values of range were considered for the spherical model: 2, 3.5, and 5; for $S = 3$, all scenarios were based on ranges of 2 and 5 for short- and long-range spherical models, respectively;
(3) the number of variables $p$ was 2 or 6;
(4) the correlations between variables were of same sign, or not, for all spatial structures. This latter parameter reflects various scale dependencies in the correlations among variables. Scale dependence here is defined as a change in the sign or in the magnitude, or both, of the correlation between

**Table 1.** Simulation Parameter Values used in the 16 Scenarios of the Monte Carlo Study, With the Location of Each Scenario in Figures 1–4

| Scenario | Number of spatial structures $S$ | Ranges for spherical models $\varphi_2$ | $\varphi_3$ | Scale dependence | Correlation sign $s_1$ | $s_2$ | $s_3$ | Number of variables $p$ | Panel in Figures 1–4 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 2 | | No | + | + | | 2 | A |
| 2 | 2 | 2 | | No | + | + | | 6 | A |
| 3 | 2 | 2 | | Yes | − | + | | 2 | A |
| 4 | 2 | 2 | | Yes | − | + | | 6 | A |
| 5 | 2 | 3.5 | | No | + | + | | 2 | B |
| 6 | 2 | 3.5 | | No | + | + | | 6 | B |
| 7 | 2 | 3.5 | | Yes | − | + | | 2 | B |
| 8 | 2 | 3.5 | | Yes | − | + | | 6 | B |
| 9 | 2 | 5 | | No | + | + | | 2 | C |
| 10 | 2 | 5 | | No | + | + | | 6 | C |
| 11 | 2 | 5 | | Yes | − | + | | 2 | C |
| 12 | 2 | 5 | | Yes | − | + | | 6 | C |
| 13 | 3 | 2 | 5 | No | + | + | + | 2 | D |
| 14 | 3 | 2 | 5 | No | + | + | + | 6 | D |
| 15 | 3 | 2 | 5 | Yes | + | − | + | 2 | D |
| 16 | 3 | 2 | 5 | Yes | + | − | + | 6 | D |

variables from one spatial structure to another. For each spatial structure, the magnitude of the (multiple) correlation was $\sqrt{0.5}$ for all scenarios. For $p = 6$, only some of the pairs of variables were correlated, namely $Z_1$ with $Z_4$, $Z_2$ with $Z_5$, and $Z_3$ with $Z_6$.

The matrix of theoretical sill values $\mathbf{B}_s$ was calculated as the product of the $p \times p$ matrix of correlations between variables at spatial structure $s$ and the proportion of the total variation allocated to that structure, to insure that $\sum_{s=1}^{S} \mathbf{B}_s = \mathrm{Cov}(\mathbf{Z})$ under the assumption of second-order stationarity. For $S = 2$, the nugget effect and the spherical semivariogram model were given respectively 1/3 and 2/3 of the total variation, while for $S = 3$, each spatial structure was allocated 1/3 of the total variation.

Each $p$-variate spatial data set was simulated using

$$\mathbf{z} = \boldsymbol{\Psi}\boldsymbol{\varepsilon} \qquad (17)$$

where $\mathbf{z}$ is the $np \times 1$ vector of simulated data, $\boldsymbol{\varepsilon}$ is an $np \times 1$ vector of $N(0, 1)$ pseudo-random numbers, and $\boldsymbol{\Psi}$ is the $np \times np$ lower triangular matrix resulting from the Cholesky decomposition of the $np \times np$ variance–covariance matrix $\mathbf{C}$, such that $\mathbf{C} = \boldsymbol{\Psi}\boldsymbol{\Psi}^{\mathrm{T}}$. The matrix $\mathbf{C}$ was calculated as $\sum_{s=1}^{S} (\mathbf{B}_s \otimes \boldsymbol{\rho}_s)$, with $\boldsymbol{\rho}_s$ the

$n \times n$ matrix of values of the basic auto- or cross-correlation function for spatial structure $s$ and $\otimes$ the Kronecker product. Sample data for each random variable were standardized to zero mean and unit variance.

The experimental semivariograms were evaluated up to the 6-unit distance lag. Distance lags $\{h_1, \ldots, h_K\}$ were chosen to correspond to exact distances between grid nodes, so that all pairs of observations used to evaluate the semivariogram at a given distance lag $h_k$ were located exactly at that distance and the average distance at the $k$th lag was equal to the exact distance. The number of lags $K$ was thus 18 for all procedures, except $GLS_3$ for which $K = 6$ (Genton, 1998b). An LMC was fitted to the $p(p+1)/2$ experimental semivariograms, using the algorithm described in the previous section.

For each of the eight least-squares estimators of sills, the empirical bias, the empirical variance, and the mean square error were calculated from 2500 simulation runs for each of the 16 scenarios. We used our own simulation program written in MATLAB language (The MathWorks, 2002) for this purpose.

## SIMULATION RESULTS

Results on the empirical variance of sill estimators in the 16 scenarios are reported separately for direct and cross semivariograms (Figs. 1 and 2), and so are results on the empirical bias (Figs. 3 and 4). The values reported are average empirical variances or biases calculated over the total number of direct or cross semivariograms for a given spatial structure in each of the scenarios. The statistical properties of the eight least-squares estimators of sills were compared for 12 scenarios with $S = 2$ and four scenarios with $S = 3$, that is, in 36 situations. By definition, the mean square error (MSE) of an estimator is equal to its empirical variance plus its squared empirical bias; the empirical bias is obtained by subtracting the theoretical value of the parameter from the sample mean of estimates. In our study, the MSE of all sill estimators was accounted for mostly by the empirical variance instead of the squared empirical bias whose values were relatively small—outside the context of the LMC, this could be expected on a theoretical basis—and this is why we do not report results on the MSE. We present results on the empirical bias instead of the squared empirical bias for the following reason. Although a low value of average empirical bias can result from the combination of positive and negative empirical biases, we observed that the sign of empirical biases for a given spatial structure in a given scenario was generally constant and that the patterns were basically the same. We also chose to report results on nonsquared empirical biases because of the additional information that their sign provided in terms of direction (upwards vs. downwards). Below, we give a general overview of our simulation results before reporting on specific effects of the simulation parameters.

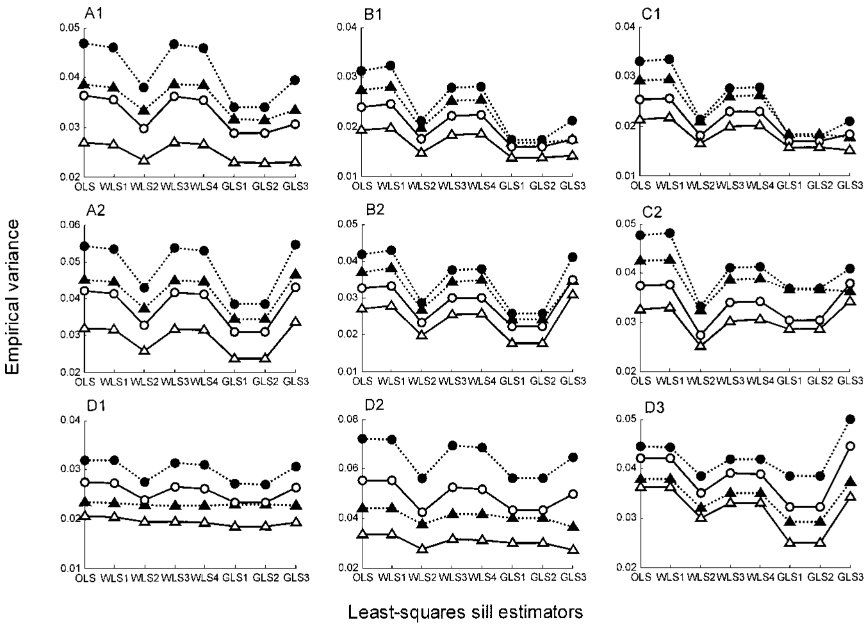


**Figure 1.** Empirical variance of sill estimators for direct semivariograms. See Table 1 for the description of the scenarios found in panels A–D. The number following the panel letter identifies the spatial structure $s$, with 1 for the nugget effect. Circles (●, ○) are for $p = 2$ variables, and triangles (△, ▲) for $p = 6$. Filled symbols (●, ▲) represent the absence of scale dependence of the correlations, and empty symbols (○, △), the presence of such scale dependence. See text for the definition of the least-squares estimation procedures.

## General Overview

Differences in the empirical variance of the eight least-squares estimators of sills are relatively constant among spatial structures and scenarios (Figs. 1 and 2). For direct semivariograms, the $GLS_1$ and $GLS_2$ estimators have the two smallest variances in 26 situations, and are among the best three in 32 situations (Fig. 1). For cross semivariograms, $GLS_1$ and $GLS_2$ have the two smallest variances in 32 situations, and are among the best three in all 36 situations (Fig. 2). Results for the $GLS_3$ procedure proposed by Genton (1998b) are more dependent on the combination of simulation parameter values considered in a given scenario. $GLS_3$ is also the only least-squares procedure for which the algorithm did not converge for some of the 2500 simulation runs. Since the major part of the MSE comes from the empirical variance of sill estimators, the results above already support the conclusion that the $GLS_1$ estimator be recommended in practice in the context of the LMC because of its greater efficiency overall, compared to the OLS, $WLS_{1–4}$ and $GLS_3$ estimators.

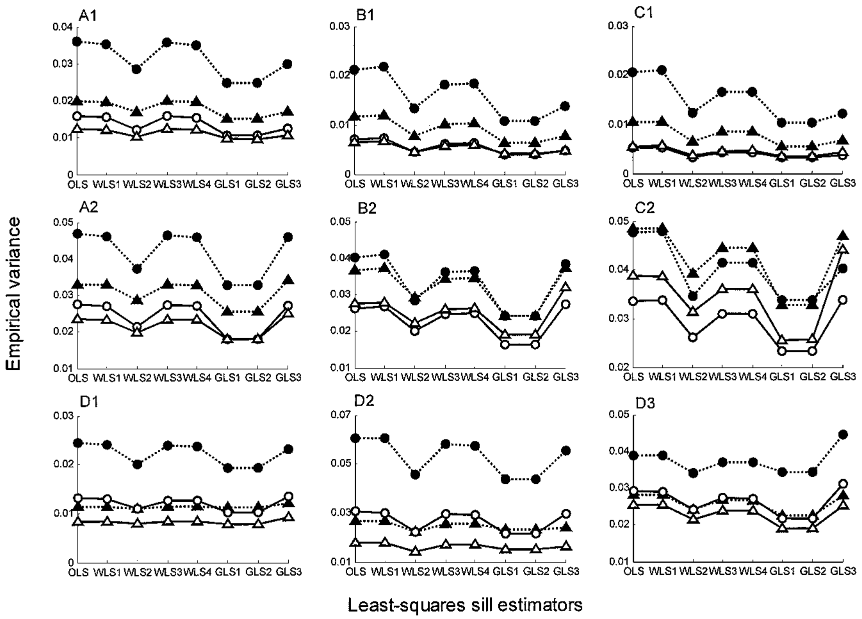


**Figure 2.** Empirical variance of sill estimators for cross semivariograms. See Table 1 and caption of Figure 1 for the description of scenarios represented in each panel and the meaning of symbols.

The other least-squares estimation procedure that performs well in terms of empirical variance is the model-free $WLS_2$, which is among the best three in 24 situations for direct semivariograms (Fig. 1) and 35 situations for cross semivariograms (Fig. 2). The model-based $WLS_3$ and $WLS_4$ generally possess a smaller empirical variance than OLS and $WLS_1$, but rarely perform better than $GLS_1$, $GLS_2$, or $WLS_2$.

The patterns observed for the empirical bias (Figs. 3 and 4) vary with the combinations of simulation parameter values in the scenarios. Squared empirical biases tend to be of one or two orders of magnitude smaller than empirical variances.

## Effect of Simulation Parameters

The four simulation parameters used to define the 16 scenarios in the Monte Carlo study—number of spatial structures, value of ranges, number of variables, and scale dependence of the correlations—have less complex effects on the empirical variance than on the empirical bias of sill estimators. For direct and cross semivariograms (Figs. 1 and 2), the following effects on the empirical variance are observed: (i) all other things being equal, variances are greater when correlations

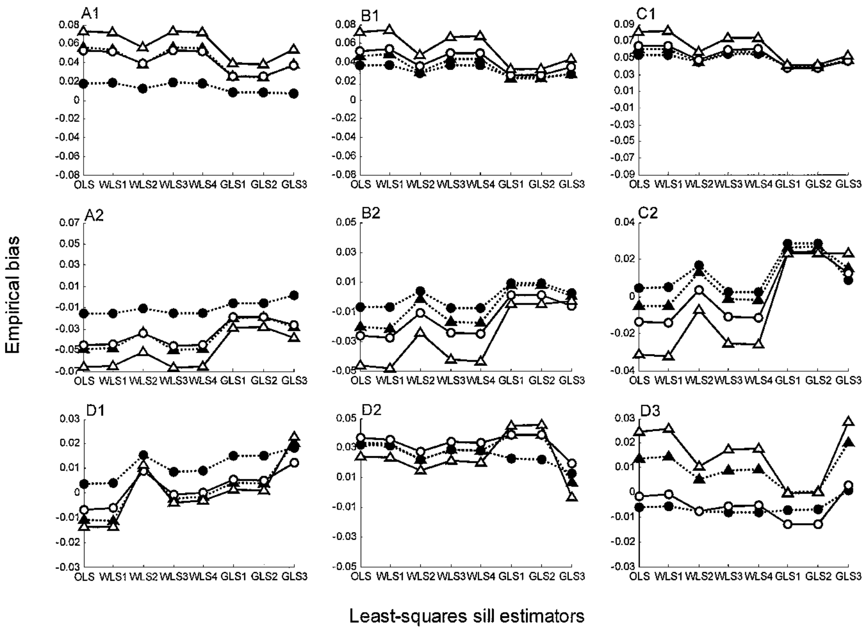


**Figure 3.** Empirical bias of sill estimators for direct semivariograms. See Table 1 and caption of Figure 1 for the description of scenarios represented in each panel and the meaning of symbols.

are not scale-dependent; (ii) $WLS_3$ and $WLS_4$ have smaller variances than OLS as the range increases; (iii) $GLS_3$ has relatively higher variances at spatial structures that are more autocorrelated.

The sign or direction of the empirical bias is associated primarily with the spatial structure for which the sills are estimated. For $S = 2$, there is a positive bias for the nugget effect and, for most of the least-squares estimation procedures, a negative bias for the spherical structure (Fig. 3, panels A and B; and Fig. 4, panels A and B). For $S = 3$, the sill estimates for the first spherical structure are biased upwards for direct and cross semivariograms (Fig. 3, panel D2 and Fig. 4, panel D2). Differences in the sign or direction of the empirical bias among least-squares estimation procedures exist, however (Fig. 3, panels C2 and D1; and Fig. 4, panel C2). Another noticeable effect is that the empirical bias of sills for cross semivariograms is inflated when correlations are scale-dependent (Fig. 4, with the exceptions of panels B1, C1, and C2). This notwithstanding, results on the empirical bias suggest complex interactions among the four simulation parameters.

The differences in the simulation parameter values among scenarios affected the ability of the LMC algorithm to converge for one of the least-squares procedures, $GLS_3$. For this procedure, the algorithm converged in all 2500 simulation runs for only four scenarios (1, 3, 4, and 13), but did not converge in more than

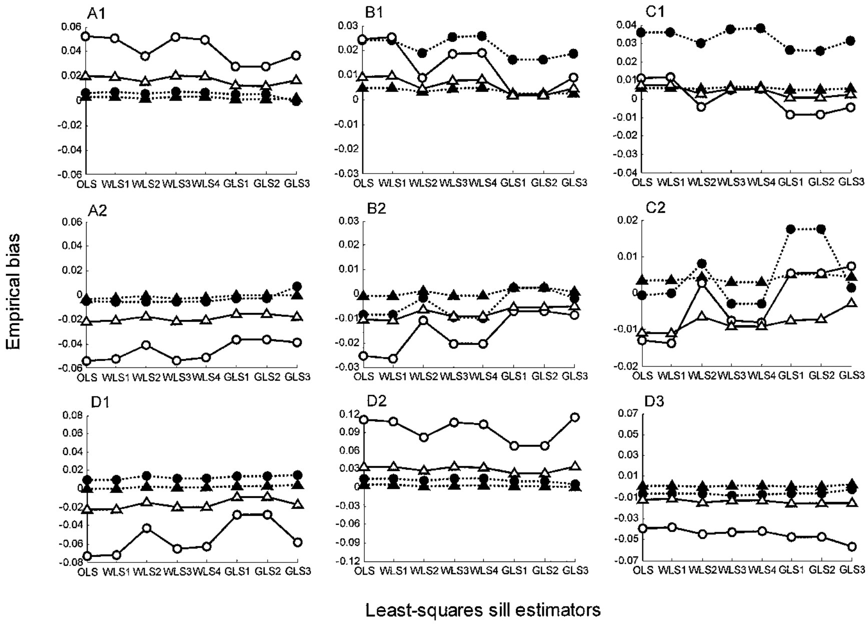


**Figure 4.** Empirical bias of sill estimators for cross semivariograms. See Table 1 and caption of Figure 1 for the description of scenarios represented in each panel and the meaning of symbols.

100 simulation runs for six scenarios, all characterized by $p = 6$ (14, 8, 16, 12, 6, and 10).

## DISCUSSION

To the best of our knowledge, all studies comparing the statistical properties of different least-squares estimators have been conducted within the context of fitting a single direct semivariogram model (see our literature review in the Introduction). A main question is, therefore, to determine whether the statistical properties of least-squares estimators reported in these studies are maintained within the context of the LMC.

The general overview of our simulation results suggests that some of the patterns observed in the studies focusing on the fitting of a single direct semivariogram model are, in fact, maintained in the context of the LMC. On the one hand, OLS is always among the worst, when not the worst of all least-squares estimators. On the other hand, the greater efficiency of the GLS estimator compared to the OLS and WLS estimators, which was demonstrated in past studies (Genton, 1998b; Lahiri, Lee, and Cressie, 2002; Müller, 1999; Pardo-Igúzquiza and Dowd, 2001), is also observed in the context of the LMC. In particular, it must be noted that $WLS_1$, which has been used for many studies based on the LMC in the past, performs relatively poorly. Furthermore, we observed that the better performance of the GLS estimators of sills is more pronounced for cross than for direct semivariograms, which strongly supports their use in studies investigating scale-dependent relationships among variables. To our knowledge, our study is the first that addressed the issues of estimating the vector of weights (WLS) and the variance–covariance matrix of cross semivariogram estimates (GLS) for fitting a cross semivariogram model and assessed the statistical properties of the corresponding least-squares estimators.

Results for $GLS_1$ and $GLS_2$ are very similar. There are at least three possible explanations for this similarity. First, the use of $\hat{\boldsymbol{\beta}}_{ij}$ in $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}_{ij})$ is adequate in $GLS_1$, so that $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}_{ij})$ is close to the theoretical $\boldsymbol{\Sigma}(\boldsymbol{\beta}_{ij})$ that is used in $GLS_2$ but is unknown in practice. Second, the knowledge of the correct autocorrelation structure through the ranges in the semivariogram model is more important than the sills themselves. Third, the sill matrix estimates of $GLS_1$ and $GLS_2$ require similar correction for positive semidefiniteness. These three explanations may apply altogether or not, depending on the situation. For instance, the constraint of positive semidefiniteness may be more a problem for $p = 6$ than for $p = 2$.

The $GLS_3$ procedure proposed by Genton (1998b) did not perform as well as the $GLS_1$ procedure based on the direct computation of the variance–covariance matrix of semivariogram estimates that we propose. In fact, simulation results presented in Genton (1998b, Table 2, p. 336) for regularly spaced data in $R^1$ indicated that, although his GLS procedure was more efficient than $WLS_3$ (Cressie, 1985) for estimating the range of the model, there were no clear differences between procedures in the empirical bias and variance of sill estimators for the nugget effect and the spherical structure. Note that in the GLS procedure of Genton (1998b), a direct semivariogram model can be fitted to robust semivariogram estimators (Cressie and Hawkins, 1980; Genton, 1998a; Lark, 2000) which, contrary to Matheron's classical estimator, are not expressed as a quadratic form. In the same procedure, however, the correlation matrix of Matheron's semivariogram estimates is used as an approximation to the correlation matrix of robust semivariogram estimates. Incorporating robust semivariogram estimators in the context of the LMC seems possible to us, but will require further investigation with regard to cross semivariograms (Lark, 2002).

The good performance of $WLS_2$ (Zhang, Van Eijkeren, and Heemink, 1995) observed in our study is in agreement with results of studies focusing on the fitting of a single direct semivariogram model. Indeed, those results demonstrated that $WLS_2$ had better finite sample properties than Cressie's (1985) $WLS_3$ procedure (Zhang, Van Eijkeren, and Heemink, 1995) and that $WLS_2$ performed better than other WLS procedures in the presence of a nugget effect (Pardo-Igúzquiza, 1999). An interesting aspect of $WLS_2$ is that it is model-free and hence, depends only on the spatial design of the sampling scheme. Although it is not clear why $WLS_2$ performs better than other WLS procedures, Zhang, Van Eijkeren, and Heemink (1995) highlighted the fact that the model-based $WLS_3$ was not using the same weight when positive and negative deviations of the empirical semivariogram ordinate from the fitted model were of same magnitude. Note that this is also true for $WLS_4$ and $GLS_3$ (Genton, 1998b), since both use values of the experimental semivariogram for their weighing procedure (see Equation (A7) in the Appendix).

The contribution of the squared empirical bias to the mean square error of sill estimators was found to be relatively small in our study. However, the fact that there is a bias is indicative of some effect of the LMC algorithm on the statistical properties of sill estimators because in theory, least-squares estimators of slopes in linear models are known to be unbiased (Searle, 1971, p. 89). There may be two reasons for such an effect. First, some of the matrices of initial sill estimates had to be transformed to ensure their positive semidefiniteness. Second, since all direct and cross semivariogram models were fitted using the same set of basic semivariogram functions, the range(s) imposed by the LMC might not correspond exactly to the best range(s) possible in the least-squares sense for some of the empirical semivariograms. This problem might have been exacerbated when a certain degree of nonstationarity was induced in the simulated data, especially in scenarios with longer theoretical range. The complex patterns of the empirical bias observed in our Monte Carlo study may also reflect interactions between the LMC algorithm and the simulation parameter values used in the scenarios. The difference in the effect of scale dependence of the correlations on the results for direct and cross semivariograms suggests that the estimators of their respective sills present different characteristics.

## CONCLUDING REMARKS

Our main objective was to test the expectation that the GLS estimator of sills possesses better statistical properties than the OLS and WLS estimators in the context of the LMC, when the ranges are assumed to be known and only the sills are unknown. The results of our Monte Carlo study support the conclusion that this is the case because the better properties in terms of variance insured in theory appear to be maintained with minimum bias after application of the LMC algorithm. In addition, the algorithm for fitting the LMC presented in this article extends the algorithm proposed by Goulard (1989), by including any type of least-squares procedure. Thus, the use of GLS can improve the performance of future studies using the LMC. It must be noted that the use of distance classes instead of exact distances in the computation of direct and cross empirical semivariograms has an effect at various degrees on the performance of all least-squares procedures, but when this is the case, the GLS procedure appears to keep its better relative efficiency overall (results not reported here). Another point that is worth mentioning for future applications to real data sets is that the standardization of sample data for each variable that is commonly performed prior to any quantitative analysis has an effect on the distribution of sill estimates because it results in a different plateau in direct empirical semivariograms. Guidelines for future work in the context of the LMC include the question of robust estimation of cross semivariograms, the assessment of GLS sill estimators to investigate scale-dependent relationships among variables, and the development of a theoretical framework to further explain the statistical properties of sill estimators.

## ACKNOWLEDGMENTS

The second and fourth authors acknowledge individual research grants from the Natural Sciences and Engineering Research Council of Canada. For the Monte Carlo study, we benefited from computing facilities made available through a Canada Foundation for Innovation grant. We are grateful to two reviewers for their comments on an earlier draft of this article.

## REFERENCES

Castrignanò, A., Giugliarini, L., Risaliti, R., and Martinelli, N., 2000a, Study of spatial relationships among some soil physico-chemical properties of a field in central Italy using multivariate geostatistics: Geoderma, v. 97, no. 1–2, p. 39–60.

Castrignanò, A., Goovaerts, P., Lulli, L., and Bragato, G., 2000b, A geostatistical approach to estimate probability of occurrence of Tuber melanosporum in relation to some soil properties: Geoderma, v. 98, no. 3–4, p. 95–113.

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

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

Dobermann, A., Goovaerts, P., and George, T., 1995, Sources of soil variation in an acid Ultisol of the Philippines: Geoderma, v. 68, no. 3, p. 173–191.

Dobermann, A., Goovaerts, P., and Neue, H. U., 1997, Scale-dependent correlations among soil properties in two tropical lowland rice fields: Soil Sci. Soc. Am. J., v. 61, no. 5, p. 1483–1496.

Furrer, R., and Genton, M. G., 1999, Robust spatial data analysis of Lake Geneva sediments with S+SpatialStats: Syst. Res. Inform. Sci., v. 8, no. 4, p. 257–272.

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

Genton, M. G., 1998b, Variogram fitting by generalized least squares using an explicit formula for the covariance structure: Math. Geol., v. 30, no. 4, p. 323–345.

Golub, G. H., and Pereyra, V., 1973, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate: SIAM J. Numer. Anal., v. 10, no. 2, p. 413–432.

Goovaerts, P., 1994, On a controversial method for modeling a coregionalization: Math. Geol., v. 26, no. 2, p. 197–204.

Goovaerts, P., 1997, Geostatistics for natural resources evaluation: Oxford University Press, New York, 483 p.

Goovaerts, P., Sonnet, P., and Navarre, A., 1993, Factorial kriging analysis of springwater contents in the Dyle River basin, Belgium: Water Resour. Res., v. 29, no. 7, p. 2115–2125.

Goovaerts, P., and Webster, R., 1994, Scale-dependent correlation between topsoil copper and cobalt concentrations in Scotland: Eur. J. Soil Sci., v. 45, no. 1, p. 79–95.

Gorsich, D. J., Genton, M. G., and Strang, G., 2002, Eigenstructures of spatial design matrices: J. Multivariate Anal., v. 80, no. 1, p. 138–165.

Goulard, M., 1989, Inference in a coregionalization model, *in* Armstrong, M., ed., Geostatistics, vol. 1: Kluwer Academic, Dordrecht, p. 397–408.

Goulard, M., and Voltz, M., 1992, Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix: Math. Geol., v. 24, no. 3, p. 269–286.

Guttman, I., Pereyra, V., and Scornik, H. D., 1973. Least-squares estimation for a class of nonlinear models: Technometrics, v. 15, no. 2, p. 209–218.

Jian, X. D., Olea, R. A., and Yu, Y. S., 1996, Semivariogram modeling by weighted least squares: Comput. Geosci., v. 22, no. 4, p. 387–397.

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

Lahiri, S. N., Lee, Y. D., and Cressie, N., 2002, On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters: J. Stat. Plan. Infer., v. 103, no. 1–2, p. 65–85.

Lark, R. M., 2000, A comparison of some robust estimators of the variogram for use in soil survey: Eur. J. Soil Sci., v. 51, no. 1, p. 137–157.

Lark, R. M., 2002, Robust estimation of the pseudo cross-variogram for cokriging soil properties: Eur. J. Soil Sci., v. 53, no. 2, p. 253–270.

Matheron, G., 1962, Traité de géostatistique appliquée, Tome 1: Mémoires du Bureau de Recherches Géologiques et Minières, no. 14: Éditions Technip, Paris, 333 p.

McBratney, A. B., and Webster, R., 1986, Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates: J. Soil Sci., v. 37, no. 4, p. 617–639.

Müller, W. G., 1999, Least-squares fitting from the variogram cloud: Stat. Probabil. Lett., v. 43, no. 1, p. 93–98.

Ortiz, C. J., and Deutsch, C. V., 2002, Calculation of uncertainty in the variogram: Math. Geol., v. 34, no. 2, p. 169–183.

Osborne, M. R., 1970, A class of nonlinear regression problems, *in* Andersen, R. S., and Osborne, M. R., eds., Data representation: University of Queensland Press, St. Lucia, p. 94–101.

Pardo-Igúzquiza, E., 1999, VARFIT: A fortran-77 program for fitting variogram models by weighted least squares: Comput. Geosci., v. 25, no. 3, p. 251–261.

Pardo-Igúzquiza, E., and Dowd, P., 2001, Variance–covariance matrix of the experimental variogram: Assessing variogram uncertainty: Math. Geol., v. 33, no. 4, p. 397–419.

Searle, S. R., 1971, Linear models: Wiley, New York, 532 p.

The MathWorks, Inc., 2002, MATLAB Version 6.5: The MathWorks, Inc., Natick, MA.

Voltz, M., and Goulard, M., 1994, Spatial interpolation of soil moisture retention curves: Geoderma, v. 62, no. 1–3, p. 109–123.

Wackernagel, H., 1995, Multivariate geostatistics: An introduction with applications: Springer-Verlag, Berlin, 256 p.

Wackernagel, H., Petitgas, P., and Touffait, Y., 1989, Overview of methods for coregionalization analysis, *in* Armstrong, M., ed., Geostatistics, vol. 1: Kluwer Academic, Dordrecht, p. 409–420.

Webster, R., Atteia, O., and Dubois, J. P., 1994, Coregionalization of trace metals in the soil in the Swiss Jura: Eur. J. Soil Sci., v. 45, no. 2, p. 205–218.

Zhang, X. F., Van Eijkeren, J. C. H., and Heemink, A. W., 1995, On the weighted least-squares method for fitting a semivariogram model: Comput. Geosci., v. 21, no. 4, p. 605–608.

Zimmerman, D. L., and Zimmerman, M. B., 1991, A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors: Technometrics, v. 33, no. 1, p. 77–91.

## APPENDIX

1. The spatial design matrix $\mathbf{A}(\mathbf{h})$ is calculated as

$$\mathbf{A}(\mathbf{h}) = \frac{\boldsymbol{\eta}(\mathbf{h}) - \mathbf{M}(\mathbf{h})}{2N(\mathbf{h})} \qquad \text{(A1)}$$

where $\mathbf{M}(\mathbf{h})$ is the $n \times n$ binary matrix $(m_{uv})$ with $m_{uv} = 1$ if the vector linking sampling locations $u$ and $v$ corresponds to lag $\mathbf{h}$ and $m_{uv} = 0$ otherwise, $\boldsymbol{\eta}(\mathbf{h})$ is the $n \times n$ diagonal matrix $(\eta_{uv})$ with $\eta_{uu} = \sum_{v=1}^{N} m_{uv}$ (i.e., the number of neighbors for sampling location $u$ at lag $\mathbf{h}$), and $N(\mathbf{h})$ is the number of pairs of observations at lag $\mathbf{h}$.

2. The $n \times n$ covariance matrix $\mathbf{C}(\boldsymbol{\beta}_{ij})$ can be calculated as

$$\mathbf{C}(\boldsymbol{\beta}_{ij}) = \sum_{s=1}^{S} \beta_{ij,s} \boldsymbol{\rho}_s \qquad \text{(A2)}$$

where $\boldsymbol{\rho}_s$ is the $n \times n$ matrix of values of the basic auto- or cross-correlation function for spatial structure $s$. Instead of evaluating Equation (12), $p(p + 1)/2$ times one entry at a time, it is possible to take advantage of the fact that $\boldsymbol{\rho}_s$ is the same for all semivariograms, by using the following equation:

$$\begin{aligned} &\mathrm{Cov}(\gamma_{ij}^*(h_k), \gamma_{ij}^*(h_{k'})) \\ &= \sum_{r=1}^{S} \sum_{q=1}^{S} [(\beta_{ij,r}\beta_{ij,q} + \beta_{ii,r}\beta_{jj,q})\, \mathrm{tr}\,(\boldsymbol{\rho}_r \mathbf{A}(h_k) \boldsymbol{\rho}_q \mathbf{A}(h_{k'}))] \end{aligned} \qquad \text{(A3)}$$

where each term $\mathrm{tr}(\boldsymbol{\rho}_r \mathbf{A}(h_k) \boldsymbol{\rho}_q \mathbf{A}(h_{k'}))$ needs to be calculated only once.

3. We adapted Cressie's (1985) approximation for the variance of a direct semivariogram to use it with a cross semivariogram, by applying the general form of Equation (12) with $i \neq j$ and $h_k = h_{k'}$. This adaptation provided us with the following approximation for the variance of a cross semivariogram

$$\begin{aligned} &\mathrm{var}(\gamma_{ij}^*(h_k)) \\ &\cong \frac{\gamma_{ij}(h_k, \boldsymbol{\beta}_{ij})^2 + \gamma_{ii}(h_k, \boldsymbol{\beta}_{ii})\gamma_{jj}(h_k, \boldsymbol{\beta}_{jj})}{N(h_k)}, \quad k = 1, \ldots, K \end{aligned} \qquad \text{(A4)}$$

4. In Genton (1998b, p. 332), the covariance matrix of semivariogram estimates is given by

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}_{ij}) = \mathbf{R} \circ \boldsymbol{\xi}(\boldsymbol{\beta}_{ij}) \tag{A5}$$

where the symbol $\circ$ denotes the Hadamard product of two matrices. The entry $(h_k, h_{k'})$ of the $\mathbf{R}$ matrix is

$$\mathrm{Corr}(\gamma_{ij}^*(h_k), \gamma_{ij}^*(h_{k'})) = \frac{\mathrm{tr}[\mathbf{A}(h_k)\mathbf{A}(h_{k'})]}{\sqrt{\mathrm{tr}[\mathbf{A}(h_k)\mathbf{A}(h_k)]\mathrm{tr}[\mathbf{A}(h_{k'})\mathbf{A}(h_{k'})]}} \tag{A6}$$

which is calculated under the assumption of spatial independence of the observations and hence, depends only on the spatial design matrix $\mathbf{A}(\mathbf{h})$. The entry $(h_k, h_{k'})$ of the matrix $\boldsymbol{\xi}(\boldsymbol{\beta}_{ij})$ in Equation (A5) is calculated by using an extension of Equation (A4) without the constraint $h_k = h_{k'}$:

$$\frac{0.5(\gamma_{ij}(h_k, \boldsymbol{\beta}_{ij})\gamma_{ij}(h_{k'}, \boldsymbol{\beta}_{ij}) + \gamma_{ii}(h_k, \boldsymbol{\beta}_{ii})\gamma_{jj}(h_{k'}, \boldsymbol{\beta}_{jj}))}{\sqrt{N(h_k)N(h_{k'})},} \quad i, j = 1, \ldots, p \tag{A7}$$

GLS$_3$ utilizes only the distance lags $h_k$ that are found on the vertical and horizontal directions of a regular $12 \times 12$ grid and that correspond to integer values. This is why for $h_k = 5$, a different spatial design matrix $\mathbf{A}(h_k)$ had to be calculated for GLS$_3$ in our Monte Carlo study because the one used for the other seven least-squares procedures included diagonal directions. In this case, we followed Genton (1998b) and Gorsich, Genton, and Strang (2002), who used the Kronecker product of the spatial design matrices calculated separately for each direction to compute spatial design matrices in 2D.