ENVIRONMETRICS
Environmetrics 2007; **18**: 125–139
Published online 18 September 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/env.807

Maximum-likelihood estimation for multivariate spatial linear coregionalization models

Hao Zhang*,†

Department of Statistics, Washington State University, Pullman, WA 99164-3144, USA

SUMMARY

A multivariate spatial linear coregionalization model is considered that incorporates the Matérn class of covariograms. An EM algorithm is developed for maximum-likelihood estimation that has a few desirable properties and is capable of handling high-dimensional data. Most estimates in the EM algorithm are updated through closed form expressions and these estimates automatically satisfy necessary constraints. The model and algorithm are illustrated through a real example. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: cokriging; EM algorithm; linear coregionalization model; multivariate covariogram; spatial correlation

1. INTRODUCTION

Multiple spatial variables are often observed in many studies in environmental, agricultural, and ecological sciences. The observed values of these spatial variables are referred to as multivariate spatial data, which often possess two kinds of spatial correlation: spatial autocorrelation that exists between observations of an individual variable at different locations, and spatial cross-correlation that describes the correlation between two different variables measures at either the same or different locations. It is an important problem to model both kinds of spatial correlation. By appropriately accounting for and modeling the spatial correlation, efficient estimation, and better prediction can be achieved. For example, cokriging is a technique for linear prediction of one variable by making use of observed values of other variables, and can result in more precise prediction than the kriging methods that utilize only the spatial auto-correlation of this particular variable being predicted.

Cross correlation has been modeled through a multivariate covariogram and a cross-variogram (Wackernagel, 1998; Chilés and Delfiner, 1999), and also through pseudo cross variogram (Myers, 1991; Ver Hoef and Barry, 1998). If the objective of the analysis of multivariate spatial data is prediction, one can use either multivariate covariogram, multivariate variogram, or multivariate pseudo variogram although these methods may yield different prediction results. Wackernagel (1998)

*Correspondence to: H. Zhang, Department of Statistics, Washington State University, Pullman, WA 99164-3144, USA.
†E-mail: zhanghao@wsu.edu

Received 27 April 2006
Accepted 29 June 2006

discussed cokriging using either multivariate covariogram and variogram, while Ver Hoef and Barry (1998) discussed cokriging by employing the multivariate pseudo variogram. However, if the objective is on estimation—for example, estimation of some mean parameters, multivariate covariogram in general is employed. Therefore, modeling the multivariate covariogram has its own interest.

A number of models for the multivariate covariogram exists (see, e.g., Wackernagel, 1998; Chilés and Delfiner, 1999), among which the linear coregionalization model (LCM) is perhaps the most general one. For LCM, Goulard and Voltz (1992) studied least squares estimators for some of the model parameters using the empirical multivariate variogram, under the assumption that some other parameters in the model are known. Maximum likelihood estimates can be hard to find for two reasons, one being the potential high dimension of parameter space, and the other being the constraints on the parameters that are necessary for a valid covariogram. MLE must satisfy these constraints in order for the resulting multivariate covariogram to be a valid one. We will discuss more on this in Section 2.

The purpose of the paper is to develop an EM algorithm for the maximum-likelihood estimation for the parameters in the LCM. The EM algorithm (Dempster *et al*., 1977) is an iterative method for finding a maximum-likelihood estimate (MLE) when the data consist of missing or unobservable values. Our version of the EM algorithm for the LCM has a few desirable properties. First, in each iteration, most estimates are updated in closed form. Second, the estimates in each iteration automatically satisfy the necessary constraints for a valid multivariate covariogram. Finally, it leads to maximum likelihood estimates of the parameters. Due to their desirable properties, maximum-likelihood estimates are preferred when they are computable.

In this work, we also extend the LCM by incorporating the Matérn covariogram into the model. The Matérn class of covariograms has received much attention in recent years because it has a parameter that controls how smooth the process is. These smoothness parameters in general are estimated unsatisfactorily by least-squares based methods, as in the univariate case (Stein, 1999). The reason is that these parameters determine how smooth the processes are and hence have more to do with the covariogram or variogram at small lags than at large lags. Consequently, likelihood-based methods are better at estimating the behavior of covariogram near the origin. In the next two sections, we will introduce the LCM and the EM algorithm. We will provide explicit expressions for the EM algorithm in Section 3 and 4, we illustrate the model and the EM algorithm through a real example in an environmental study. Some discussion is provided in the final section.

2. LINEAR COREGIONALIZATION MODEL AND ESTIMATION

We start with a review of basic concepts about a multivariate second-order stationary process. Let $\boldsymbol{Y}(\boldsymbol{s}) = (Y_1(\boldsymbol{s})), \ldots, Y_p(\boldsymbol{s}))'$, $s \in R^d$ be a p-variate stochastic process, where $Y_i(\boldsymbol{s})$ represents the value of the ith variable at location $\boldsymbol{s}$. The process is said to be second-order stationary if for all $\boldsymbol{s}, \boldsymbol{h} \in R^d$ and $i, j = 1, \ldots, p$,

$$E[Y_i(\boldsymbol{s})] = m_i, \ \mathrm{Cov}[Y_i(\boldsymbol{s}), Y_j(\boldsymbol{s}+\boldsymbol{h})] = C_{ij}(\boldsymbol{h}) \tag{1}$$

The functions $C_{ij}(\cdot)$ are called the direct covariograms if $i = j$ and the cross-covariograms if $i \neq j$. The matrix-valued function $C(\boldsymbol{h}) = (C_{ij}(\boldsymbol{h}))$ is called the multivariate covariogram, which must be positive

DOI: 10.1002/env

definite in the sense that for any spatial locations $s_1, \ldots, s_n$ and any vectors $\boldsymbol{a}_i \in R^p, i = 1, \ldots, n$,

$$\mathrm{Var}\left(\sum_i \boldsymbol{a}_i' \boldsymbol{Y}(\boldsymbol{s}_i)\right) = \sum_{i,\, j=1}^{n} \boldsymbol{a}_i' C(s_i - s_j)\boldsymbol{a}_j \geq 0$$

Because of this constraint, it is a difficult problem to specify a valid multivariate covariogram that is not too complex to be estimated and yet capable of modeling a wide range of spatial correlations. Only a few multivariate covariogram models have been proposed and used in analyzing real multivariate spatial data.

The simplest model is the proportional correlation model (Chilés and Delfiner, 1999):

$$C(\boldsymbol{h}) = \mathbf{V}\rho(\boldsymbol{h}),\ \boldsymbol{h} \in R^d$$

where $\mathbf{V}$ is a $p \times p$ positive definite matrix and $\rho(\boldsymbol{h})$ is a correlation function (also called a correlogram). The proportional covariogram is also called the intrinsic covariogram (Wackernagel, 1998, Chapter 23). The proportional model can be used to build the multivariate nested covariogram (Wackernagel, 1998, Chapter 26):

$$C(\boldsymbol{h}) = \mathbf{V}_0 + \sum_{k=1}^{K} \mathbf{V}_k \rho_k(\boldsymbol{h}) \tag{2}$$

where for each k, $\mathbf{V}_k$ is a positive semi-definite matrix and $\rho_k(\boldsymbol{h})$ is a correlogram that depends on some additional one or two parameters. Wackernagel (1998) provided some examples of application of this model in which $\rho_k(\boldsymbol{h})$ are either exponential or spherical.

This covariogram corresponds to the linear coregionalization model (Chilés and Delfiner, 1999, Subsection 5.6.5; Wackernagel, 1998, Chapter 26):

$$\boldsymbol{Y}(s) = \mu(\boldsymbol{s}) + \sum_{k=0}^{K} \boldsymbol{X}_k(\boldsymbol{s}) \tag{3}$$

where $X_0(\boldsymbol{s})$ is a stationary but uncorrelated p-variate process with mean 0, that is,

$$EX_0(\boldsymbol{s}) = 0,\ \mathrm{Cov}(X_0(\boldsymbol{s}), X_0(\boldsymbol{s} + \boldsymbol{h})) = \mathbf{V}_0 1_{\{\boldsymbol{h} \neq 0\}}$$

$X_k(\boldsymbol{s})$ is a p-variate stationary process with mean 0 and a multivariate covariogram $\mathbf{V}_k\rho_k(\boldsymbol{h})$, $k = 1, \ldots, K$. In addition, the $(1 + K)$ processes are uncorrelated in the sense that for any $k \neq j$,

$$\mathrm{Cov}(\boldsymbol{X}_k(\boldsymbol{s}), \boldsymbol{X}_j(\tilde{\boldsymbol{s}})) = 0, \forall \boldsymbol{s}, \tilde{\boldsymbol{s}}$$

In this work, we consider $\rho_k(\boldsymbol{h}) = \rho(\boldsymbol{h};\psi_k)$ to be a Matérn correlation function that depends on the parameter $\psi_k = (\nu_k, \phi_k)'$, where

$$\rho(\boldsymbol{h};\psi) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\nu^{1/2}\|\boldsymbol{h}\|}{\phi}\right)^{\nu} K_\nu \left(\frac{2\nu^{1/2}\|\boldsymbol{h}\|}{\phi}\right) \tag{4}$$

and K_ν is the modified Bessel function of order ν as discussed by Abramowitz and Stegun (1967). The parameter ν_k controls the smoothness of the process $\boldsymbol{X}_k(\boldsymbol{s})$. The larger ν_k is, the smoother the process becomes. Therefore the components $\boldsymbol{X}_k(\boldsymbol{s})$, $K = 1, \ldots, k$, which have different smoothness parameters, are capable of representing different scales of variation.

Goulard and Voltz (1992) developed an algorithm for estimating $\mathbf{V}_k$ in the linear coregionalization model when the correlograms ρ_k are known. Hence, their method does not estimate the correlogram

DOI: 10.1002/env

parameters such as ν_k and ψ_k, $k = 1, \ldots, K$. That method is an extension of the least squares fitting of variogram in the univariate case. It first calls for non-parametric estimation of the direct variograms and cross-variograms γ_{ij} at some lags $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_N$ and then minimizes through an iterative procedure $\sum_{j=1}^{N} \mathrm{tr}((\hat{\Upsilon}(\boldsymbol{h}_j) - \Upsilon(\boldsymbol{h}_j))^2$, where $\Upsilon(\boldsymbol{h})$ is the variogram matrix whose (i, j)th element is

$$\gamma_{ij}(\boldsymbol{h}) = (1/2)\mathrm{Cov}[Y_i(\boldsymbol{s}+\boldsymbol{h}) - Y_i(\boldsymbol{s}), Y_j(\boldsymbol{s}+\boldsymbol{h}) - Y_j(\boldsymbol{s})] \tag{5}$$

and $\hat{\Upsilon}(h_j)$ is the empirical variogram matrix. The minimization is subject to the constraint that the estimates of the matrices $\mathbf{V}_k$ are all positive semi-definite.

There are some limitations of least-squares based method such as that in Goulard and Voltz (1992) for the LCM in this paper. First, since the parameter ν_k reflect the smoothness properties of the component $\boldsymbol{X}_k(\boldsymbol{s})$, the behavior of the variogram near the origin is more important to the estimation of ν_k and consequently variogram at large distance lags has little information about the ν_k. Therefore, estimation of ν_k based on empirical variogram is very inefficient. Second, this algorithm does not apply to the completely heterotopic case, that is, different variables are observed at completely different locations and there is no sampling location where more than one variable is observed. This is because the empirical cross-variograms defined in Equation (5) cannot be computed in this case. It may work in the partially heterotopic case though this algorithm only calculates the cross-variograms at isotopic locations where all variables are observed and hence reduces the efficiency of estimation. Definitions of heterotopy and isotopy can be found in Wackernagel (1998, p 159).

An alternative to the least-squares estimation is the maximum-likelihood estimation if the process is Gaussian. Mardia and Marshall (1984) described the Fisher scoring algorithm for the maximum-likelihood estimation for a univariate Gaussian stationary process, which can be extended straightforwardly to multivariate processes. However, the large number of parameters and the constraints on the parameters can make the implementation problematic. For example, each $\mathbf{V}_k$ has $p(p+1)/2$ parameters that are constrained to make $\mathbf{V}_k$ positive semi-definite. Estimates of these parameters ought to satisfy this constraint. In the next section, we introduce the EM algorithm for maximum-likelihood estimation, which overcomes these difficulties.

A version of EM algorithm has been developed by Zhu *et al*. (2005) for a multivariate spatio-temporal generalized linear mixed models, in which the random effects or latent variables follow an LCM with exponential covariograms. Our implementation of the EM algorithm in this work differs from that in Zhu *et al*. (2005) in that we provide explicit expression in closed form for the estimate of $\mathbf{V}_k$, $k = 0, \ldots, K$ in each iteration, while these estimates are given by constrained maximization in Zhu *et al*. (2005).

3. THE EM ALGORITHM FOR LCM

The EM algorithm is applied when there are missing values or latent variables. In the LCM (3), the processes $\{\boldsymbol{X}_k(\boldsymbol{s})\}, k = 1, \ldots, K$ are unobservable, and therefore the EM algorithm can be applied. We first introduce the following notations before introducing the EM algorithm. Recall that $\boldsymbol{Y}(\boldsymbol{s}) = (\mathrm{Y}_1(\boldsymbol{s}), \ldots, Y_p(\boldsymbol{s}))'$ is a p-variate process. Let $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ be the sampling locations where at least one of the p variables is observed and $\boldsymbol{Y}$ be the vector of all observations.

DOI: 10.1002/env

Hence $\boldsymbol{Y}$ consists of all those $Y_i(\boldsymbol{s}_j)$ that are observed. Let $\boldsymbol{X}_{ki}(\boldsymbol{s})$ be the ith element of $\boldsymbol{X}_k(\boldsymbol{s})$ and write

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)' = E(\boldsymbol{Y}(\boldsymbol{s}))$$

$$\boldsymbol{X}_{ki} = (\boldsymbol{X}_{ki}(s_1)), \ldots, \boldsymbol{X}_{ki}(s_n))',\ \boldsymbol{X}_k = (\boldsymbol{X}'_{k1}, \ldots, \boldsymbol{X}'_{kp})'$$

$$\boldsymbol{X} = (\boldsymbol{X}'_2, \ldots, \boldsymbol{X}'_p)'$$

$$\boldsymbol{Y}_i = (Y_i(s_n), \ldots, Y_i(s_n)), \boldsymbol{Y}^* = (\boldsymbol{Y}'_1, \ldots, \boldsymbol{Y}'_p)'$$

$$R_k(\psi_k) = (\rho_k(s_i - s_j, \psi_k))_{i,j=1}^n,\ \Sigma_k(\psi_k) = V_k \otimes R_k(\psi_k)$$

where $\otimes$ denotes the Kronecker product.

Note that we do not require that at each of the locations $\boldsymbol{s}_i$, all p variables $Y_1(\boldsymbol{s}_i), \ldots, Y_p(\boldsymbol{s}_i)$ are observable. Hence the observed vector $\boldsymbol{Y}$ may be a subset of $\boldsymbol{Y}^*$. The complete-data log likelihood is, apart from an additive constant,

$$\log L(\theta, \boldsymbol{Y}^*, \boldsymbol{X}) = -\frac{1}{2}\log(\Sigma_0) - \frac{1}{2}\left(\boldsymbol{Y}^* - \sum_{k=1}^K \boldsymbol{X}_k - \boldsymbol{\mu} \otimes 1\right)' \Sigma_0^{-1} \left(\boldsymbol{Y}^* - \sum_{k=1}^K \boldsymbol{X}_k - \mu \otimes 1\right)$$
$$-\frac{1}{2}\sum_{k=1}^K \left(\log\left(|\Sigma_k(\psi_k|) + \boldsymbol{X}'_k\, \Sigma_k^{-1}(\psi_k)\boldsymbol{X}_k\right)\right) \tag{6}$$

where $\theta = (\boldsymbol{\mu}, \mathbf{V}_0, \mathbf{V}_k, \psi_k, k = 1, \ldots, K)$ denotes all parameters in the LCM (3). The EM iterates as follows. At each iteration, there are two steps, the E-step and the M-step. The E-steps find the conditional expectation of the complete-data log likelihood. Specifically, given the estimate $\theta^{(m)}$ in the mth iteration, compute the conditional expectation of the complete-data log likelihood

$$Q(\theta \,|\, \theta^{(m)}, \boldsymbol{Y}) = E_{\theta(m)}[\log L(\theta, \boldsymbol{Y}^*, \boldsymbol{X}) \,|\, \boldsymbol{Y}]$$

where the conditional expectation is evaluated under the parameter $\theta = \theta(^{\mathrm{m}})$. At the M-step, $Q(\theta \,|\, \theta^{(m)}, \boldsymbol{Y})$ is maximized with respect to θ and the new estimate is

$$\theta^{(m+1)} = \text{ArgMax}\ Q(\theta \,|\, \theta^{(m)}, \boldsymbol{Y})$$

We now show that the maximization can be carried out mostly in closed-form. In view of Equation (6), $Q(\theta \,|\, \theta^{(m)}, \boldsymbol{Y})$ is a sum or $(1 + K)$ terms, each of which depends on a different subset of parameters. Hence maximizing $Q(\theta \,|\, \theta^{(m)}, \boldsymbol{Y})$ can be broken down into several separate small maximization problems. Specifically, $\mathbf{V}_k^{(m+1)}$ and $\psi_k^{(m+1)}, k \geq 1$ maximize

$$-E_{\theta(m)}\left[\log|\Sigma_k| + \boldsymbol{X}'_k\Sigma_k^{-1}\boldsymbol{X}_k\right] \tag{7}$$

which, because of $\log|\Sigma_k| = n\log|\mathbf{V}_k| + p\log|R_k(\psi_k)|$ and $\Sigma_k^{-1} = \mathbf{V}_k^{-1} \otimes R_k^{-1}(\psi_k)$ can be written as

$$-n\log|\mathbf{V}_k| - p\log|R_k(\psi_k)| - \sum_{i,j=1}^p \upsilon_{ij,k} E_{\theta(m)}[\boldsymbol{X}'_{ki} R_k^{-1}(\psi_k)\boldsymbol{X}_{kj} \,|\, \boldsymbol{Y}] \tag{8}$$

DOI: 10.1002/env

where $\upsilon_{ij,k}$ is the (i,j)th element of $\mathbf{V}_k^{-1}$. From the well-known property of matrix derivative (see, e.g., Schott, 1997, p333), we have

$$-\frac{\partial \log|\mathbf{V}_k|}{\partial \upsilon_{ij,k}} = \frac{\partial \log\left|\mathbf{V}_k^{-1}\right|}{\partial \upsilon_{ij,k}} = \mathrm{tr}\left(\mathbf{V}_k \frac{\partial \mathbf{V}_k^{-1}}{\partial \upsilon_{ij,k}}\right) = d_{ij}\sigma_{ij,k}$$

where d_{ij} is 2 if $i \neq j$ and 1 otherwise. It follows immediately that the derivative of (8) with respect to $\upsilon_{ij,k}$ is

$$d_{ij}(n\sigma_{ij,k} - E_{\theta(m)}[\boldsymbol{X}'_{ki}R_k^{-1}(\psi_k)\boldsymbol{X}_{kj} \mid \boldsymbol{Y}])$$

Hence, for any fixed ψ_k, (8) as a function of $\mathbf{V}_k$ (or equivalently of $\mathbf{V}_k^{-1}$) is maximized at $\mathbf{V}_k = \mathbf{V}_k^{(m)}(\psi_k)$ whose (i,j)th $(i,j = 1, \ldots, p)$ element is

$$\sigma_{ij,k}(\psi_k) = (1/n)E_{\theta(m)}[\boldsymbol{X}'_{ki}R_k^{-1}(\psi_k)\boldsymbol{X}_{kj} \mid \boldsymbol{Y}],\ k = 1, \ldots, K \tag{9}$$

Note that the double sum in (8) equals $n\mathrm{tr}(\mathbf{V}_k^{-1}\mathbf{V}_k^{(m)}(\psi_k))$. When $\mathbf{V}_k^{-1} = \mathbf{V}_k^{(m)}(\psi_k)$, this trace equals p and (8) becomes

$$-n\log\left|\mathbf{V}_k^{(m)}(\psi_k)\right| - p\log|R_k(\psi_k)| - np$$

Therefore, in the EM algorithm, estimates for ψ_k and $\mathbf{V}_k (k = 1, \ldots, K)$ are updated by

$$\psi_k^{(m+1)} = \mathrm{ArgMin}(n\log|\mathbf{V}_k^{(m)}(\psi_k)| + p\log|R_k(\psi_k)|) \tag{10}$$

$$\mathbf{V}_k^{(m+1)} = \mathbf{V}_k^{(m)}(\psi_k^{(m+1)}) \tag{11}$$

Next we give the closed-form solution for $\mathbf{V}_0^{(m+1)}$ and $\mu^{(\mathrm{m+1})}$, which minimizes

$$\log(|\Sigma_0|) + E_{\theta(m)}[(\boldsymbol{Y}^* - \sum_{k=1}^{K}\boldsymbol{X}_k - \boldsymbol{\mu}\otimes 1)'\Sigma_0^{-1}(\boldsymbol{Y}^* - \sum_{k=1}^{K}\boldsymbol{X}_k - \boldsymbol{\mu}\otimes 1)|\boldsymbol{Y}]$$

which equals, because $\boldsymbol{Y}^* - \sum_{k=1}^{K}\boldsymbol{X}_k = \boldsymbol{\mu}^{(m)}\otimes 1 + \boldsymbol{X}_0$ given $\theta = \theta^{(m)}$,

$$\log(|\Sigma_0|) + E_{\theta(m)}[(\boldsymbol{\mu}^{(m)}\otimes 1 + \boldsymbol{X}_0 - \boldsymbol{\mu}\otimes 1)'\Sigma_0^{-1}(\boldsymbol{\mu}^{(m)}\otimes 1 + \boldsymbol{X}_0 - \boldsymbol{\mu}\otimes 1)|\boldsymbol{Y}] \tag{12}$$

Let $\boldsymbol{Z} = \boldsymbol{\mu}^{(m)}\otimes 1 + \boldsymbol{X}_0$ and $h(\boldsymbol{\mu}) = (\boldsymbol{Z} - \boldsymbol{\mu}\otimes 1)'\Sigma_0^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}\otimes 1)$. Then

$$\begin{aligned}\frac{\partial h(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}'} &= 2(\boldsymbol{Z} - \boldsymbol{\mu}\otimes 1)'\Sigma_0^{-1}\tfrac{\partial}{\partial \boldsymbol{\mu}'}(\boldsymbol{Z} - \boldsymbol{\mu}\otimes 1)\\ &= 2(\boldsymbol{Z} - \boldsymbol{\mu}\otimes 1)'\Sigma_0^{-1}(I\otimes 1)\\ &= 2(\boldsymbol{Z} - \boldsymbol{\mu}\otimes 1)'(\mathbf{V}_0^{-1}\otimes 1)\end{aligned}$$

The new estimate $\boldsymbol{\mu}^{(m+1)}$ must satisfy

$$E_{\theta^{(m)}}\left(\frac{\partial h(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}'}|\boldsymbol{Y}\right) = 0$$

DOI: 10.1002/env

which is, after transpose,

$$(\mathbf{V}_0^{-1} \otimes 1)'[E_{\theta(m)}(\boldsymbol{Z} \mid \boldsymbol{Y}) - \boldsymbol{\mu} \otimes 1] = (\mathbf{V}_0^{-1} \otimes 1')E_{\theta(m)}(\boldsymbol{Z} \mid \boldsymbol{Y}) - (\mathbf{V}_0^{-1}\boldsymbol{\mu})n = 0$$

Hence the solution is

$$\boldsymbol{\mu}^{(m+1)} = (I \otimes 1')E_{\theta(m)}(\boldsymbol{Z} \mid \boldsymbol{Y})/n = \boldsymbol{\mu}^{(m)} + (I \otimes 1')E_{\theta(m)}(\boldsymbol{X}_0 \mid \boldsymbol{Y})/n \tag{13}$$

Componentwisely, Equation (13) is

$$\mu_i^{(m+1)} = \mu_i^{(m)} + E_{\theta(m)}(1'\boldsymbol{X}_{0i} \mid \boldsymbol{Y})/n, i = 1, \ldots, p$$

Now, write $\eta_i = E_{\theta(m)}(1'\boldsymbol{X}_{0i} \mid \boldsymbol{Y})/n$. Similar to the establishment of (9), we can show that

$$\begin{aligned} \sigma_{ij,0}^{(m+1)} &= (1/n)E_{\theta(m)}[(\boldsymbol{X}_{0i} - \eta_i 1)'(\boldsymbol{X}_{0j} - \eta_j 1)|\boldsymbol{Y}] \\ &= (1/n)E_{\theta(m)}(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0j} \mid \boldsymbol{Y}) - \eta_i\eta_j \end{aligned} \tag{14}$$

The EM algorithm therefore iterates according to Equations (10), (11), (13), and (14). It is stopped when the difference between the new estimates and old estimates is within a prespecified range. The estimates are updated by expressions of closed form except that for $\psi_k (k = 1, \ldots, K)$. However, ψ_k usually is scalar or a vector of dimension 2. Hence the maximization in (10) is not too cumbersome.

It is worthwhile to note that the matrix estimates $\mathbf{V}_k^{(m+1)} (k = 0, \ldots, K)$ are not only given in closed form but also semi-positive definite. It is one of the major advantages of the EM algorithm implemented in this way. We also note that because all variables are Gaussian, the conditional expectations in Equation (9), (13), and (14) can be calculated explicitly, as shown next.

We now provide explicit expressions for these conditional expectations to facilitate the implementation of the EM algorithm, and put the proof in the Appendix. For simplicity, we focus on the isotopic case here so that $\boldsymbol{Y}^* = \boldsymbol{Y}$—that is, all the p variables are observed at each of the n locations. We will briefly discuss the heterotopic case at the end of this section. Given the estimate $\theta^{(m)}$ at the mth iteration, let $\boldsymbol{\Gamma}^{(m)}$ denote the *inverse* of the covariance matrix

$$\mathrm{Var}(\boldsymbol{Y}) = \mathbf{V}_0^{(m)} \otimes I_n + \sum_{k=1}^{K} \mathbf{V}_k^{(m)} \otimes R_k(\psi_k^{(m)})$$

and partition $\boldsymbol{\Gamma}^{(m)}$ into block matrices $\Gamma^{(m)} = (\Gamma_{ij}^{(m)})_{i,j=1}^{p}$ where each $\Gamma_{ij}^{(m)}$ is $n \times n$. Define

$$\boldsymbol{Y}^{(m)} = \Gamma^{(m)}(\boldsymbol{Y} - \boldsymbol{\mu}^{(m)} \otimes 1_n) \tag{15}$$

$$\boldsymbol{Y}_i^{(m)} = \sum_{j=1}^{p} \Gamma_{ij}^{(m)}(\boldsymbol{Y}_j - \mu_j^{(m)} 1_n), i = 1, \ldots, p$$

Then $\boldsymbol{Y}^{(m)} = (\boldsymbol{Y}_{\boldsymbol{1}}^{(\boldsymbol{m})'}, \ldots, \boldsymbol{Y}_{\boldsymbol{p}}^{(\boldsymbol{m})'})'$. We arrange $\mathbf{Y}^{(m)}$ into an $n \times p$ matrix $\mathbf{Y}_{ma}^{(m)}$ whose ith column is $\mathbf{Y}_i^{(m)}, i = 1, \ldots, p$.

Define for $k = 1, \ldots, K$ the function of ψ_k

$$W_k^{(m)}(\psi_k) = R_k(\psi_k^{(m)})R_k^{-1}(\psi_k)R_k(\psi_k^{(m)})$$

DOI: 10.1002/env

and let $\mathbf{B}_m^{(m)}(\psi_k)$ be a $p \times p$ matrix whose (i,j)th element is $\mathrm{tr}(W_k^{(m)}(\psi_k)\Gamma_{ij}^{(m)}$. Then the matrix $\mathbf{V}_k^{(m)}$ whose (i,j)th element is defined by Equation (9) can be calculated as follows as will be shown in the Appendix.

$$\begin{aligned}\mathbf{V}_k^{(m)}(\psi_k) &= (1/n)\mathbf{V}_k^{(m)}\mathbf{Y}_{ma}^{(m)'}W_k^{(m)}(\psi_k)\mathbf{Y}_{ma}^{(m)}\mathbf{V}_k^{(m)} + (1/n)\mathbf{V}_k^{(m)}\mathrm{tr}\Big(R_k(\psi_k^{(m)})R_k^{-1}(\psi_k)\Big) \\ &\quad - (1/n)\mathbf{V}_k^{(m)}\mathbf{B}_k^{(m)}(\psi_k)\mathbf{V}_k^{(m)}, \quad k \geq 1\end{aligned} \tag{16}$$

We note that $\mathbf{V}_k^{(m)}$ and $\mathbf{V}_k^{(m)}(\psi_k)$ denote different matrices: The former is the estimate of $\mathbf{V}_k$ while the later is a function of both ψ_k and $\mathbf{V}_k^{(m)}$.

In summary, the EM can be implemented as follows:

Step 1. Start with some initial value $\theta^{(0)}$; set m = 0;

Step 2. Given $\theta^{(m)}$, calculate $\theta^{(m+1)} = (\mu^{(m+1)}, \mathbf{V}_0^{(m+1)}, \mathbf{V}_k^{(m+1)}, \psi_k^{(m+1)}, k \geq 1)'$ by

$$\boldsymbol{\mu}^{(m+1)} = \boldsymbol{\mu}^{(m)} + (1/n)\mathbf{V}_0^{(m)}\mathbf{Y}_{ma}^{(m)'}1 \tag{17}$$

$$\begin{aligned}\mathbf{V}_0^{(m+1)} &= \mathbf{V}_0^{(m)} + (1/n)\mathbf{V}_0^{(m)}\mathbf{Y}_{ma}^{(m)'}\mathbf{Y}_{ma}^{(m)}\mathbf{V}_0^{(m)} - (1/n)\mathbf{V}_0^{(m)}\mathbf{B}_0^{(m)}\mathbf{V}_0^{(m)} \\ &\quad - (1/n)^2\mathbf{V}_0^{(m)}\mathbf{Y}_{ma}^{(m)'}11'\mathbf{Y}_{ma}^{(m)}\mathbf{V}_0^{(m)}\end{aligned} \tag{18}$$

$$\psi_k^{(m+1)} = \mathrm{ArgMin}(n\log|\mathbf{V}_k^{(m)}(\psi_k)| + p\log|R_k(\psi_k)|),\ k \geq 1 \tag{19}$$

$$\mathbf{V}_k^{(m+1)} = \mathbf{V}_k^{(m)}(\psi_k^{(m+1)}),\ k \geq 1 \tag{20}$$

where $\mathbf{B}_0^{(m)}$ is a $p \times p$ matrix whose (i,j)th element is the trace of $\Gamma_{ij}^{(m)}$.

Step 3. Repeat Step 2 until $\left\|\theta^{(m+1)} - \theta^{(m)}\right\|^2 < \delta$ for some prespecified small number δ.

Explicit expressions for the conditional expectation can be still given in the heterotopic case. For example, consider the conditional expectation in Equation (9), which equals

$$E_{\theta(m)}\{E_{\theta(m)}[\boldsymbol{X}_{ki}'R_k^{-1}(\psi_k)\boldsymbol{X}_{kj} \mid \boldsymbol{Y}^*] \mid \boldsymbol{Y}\}$$

The last expression can be explicitly calculated by observing that $E_{\theta(m)}[\boldsymbol{X}_{ki}'R_k^{-1}(\psi_k)\boldsymbol{X}_{kj}|\boldsymbol{Y}^*]$ is the conditional expectation considered in the isotopic case and that the unobserved components of $\boldsymbol{Y}^*$ have a normal distribution conditional on $\boldsymbol{Y}$.

4. PREDICTION

Prediction of the variables at unsampled sites is often a major objective of a study. Following the notations in the previous section, we consider the minimum mean-square error (MMSE) prediction of $\boldsymbol{Y}(\boldsymbol{s}) = (Y_1(\boldsymbol{s}), \ldots, Y_p(\boldsymbol{s}))'$ at an unsampled location $\boldsymbol{s}$. The predictor is $E(\boldsymbol{Y}(\boldsymbol{s}) \mid \boldsymbol{Y})$, where the conditional expectation is evaluated at the parameter estimates found through the EM algorithm in the previous section. This leads to the so-called plug-in prediction. In this section, we assume model parameters are known and evaluate $E(\boldsymbol{Y}(\boldsymbol{s}) \mid \boldsymbol{Y})$.

DOI: 10.1002/env

Due to the additive structure of LCM, we have

$$E(\boldsymbol{Y}(\boldsymbol{s}) \mid \boldsymbol{Y}) = \boldsymbol{\mu} + \sum_{k=0}^{K} E(\boldsymbol{X}_k(s) \mid \boldsymbol{Y})$$

However, $E(\boldsymbol{X}_k(\boldsymbol{s})|\boldsymbol{Y}) = 0$ due to independence. Hence we only need to provide explicit expressions for $E(\boldsymbol{X}_k(\boldsymbol{s})|\boldsymbol{Y})$, $k = 1, \ldots, K$. We will first consider the isotopic case, that is, all the p variables $Y_i(\boldsymbol{s}_j)$ are all observed at each sampling site $\boldsymbol{s}_j, j = 1, \ldots, n$. The heterotopic case will be considered later. Because $\boldsymbol{X}_k(\boldsymbol{s})$ and $\boldsymbol{Y}$ are jointly normal, the conditional expectation equals

$$E(\boldsymbol{X}_k(\boldsymbol{s}) \mid \boldsymbol{Y}) = \mathrm{Cov}(\boldsymbol{X}_k(s), \boldsymbol{Y})\Gamma(\boldsymbol{Y} - \boldsymbol{\mu} \otimes 1)$$

where $\boldsymbol{\Gamma}$ denotes the inverse of the covariance matrix of $\boldsymbol{Y}$, as in the previous section. Because $\boldsymbol{Y} = \boldsymbol{\mu} \otimes 1 + \sum_{k=0}^{K} \boldsymbol{X}_k$ and $\boldsymbol{X}_k(\boldsymbol{s})$ and $\boldsymbol{X}_j$ are independent for any $j \neq k$, we have

$$\mathrm{Cov}(\boldsymbol{X}_k(\boldsymbol{s}), \boldsymbol{Y}) = \mathrm{Cov}(\boldsymbol{X}_k(\boldsymbol{s})\boldsymbol{X}_k) = \mathbf{V}_k \otimes r_k'$$

where $\boldsymbol{r}_k' = (\rho_k(\boldsymbol{s}_j - \boldsymbol{s}, \psi_k), j = 1, \ldots, p)$. The second equality of the above equation follows the obvious fact that $E(X_{ki}(\boldsymbol{s}), X_{kj}(\tilde{\boldsymbol{s}})) = \sigma_{ij,k}\rho_k(\tilde{\boldsymbol{s}} - \boldsymbol{s})$, where $\sigma_{ij,k}$ denotes the (i,j)th element of $\mathbf{V}_k$ as in the previous section. Hence, the MMSE prediction of $\boldsymbol{X}_k$ is

$$E(\boldsymbol{X}_k(\boldsymbol{s}) \mid \boldsymbol{Y}) = (\mathbf{V}_k \otimes \boldsymbol{r}_k')\Gamma(\boldsymbol{Y} - \boldsymbol{\mu} \otimes 1) = \mathbf{V}_k\mathbf{M}r_k \tag{21}$$

where $\mathbf{M}$ is a $p \times n$ matrix whose (i,j)th element is the $((i-1)n+j)$th element of the vector $\Gamma(\boldsymbol{Y} - \boldsymbol{\mu} \otimes 1)$. In other words, if we partition $\Gamma(\boldsymbol{Y} - \boldsymbol{\mu} \otimes 1)$ into p blocks of length n, the ith block is the ith row of $\mathbf{M}$.

The prediction variance is

$$\mathrm{Var}(\boldsymbol{X}_k(\boldsymbol{s}) \mid \boldsymbol{Y}) = \mathrm{Var}(\boldsymbol{X}_k(\boldsymbol{s})) - \mathrm{Cov}(\boldsymbol{X}_k(\boldsymbol{s}), \boldsymbol{Y})\Gamma\mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{X}_k(\boldsymbol{s})) \tag{22}$$

$$= \mathbf{V}_k - \left(\mathbf{V}_k \otimes r_k'\right)\Gamma(\mathbf{V}_k \otimes r_k) \tag{23}$$

Extension to the heterotopic case is straightforward by observing the following

$$E(\boldsymbol{X}_k(\boldsymbol{s}) \mid \boldsymbol{Y}) = E\{E(\boldsymbol{X}_k(\boldsymbol{s}) \mid \boldsymbol{Y}^*) \mid \boldsymbol{Y}\}$$

where, as defined in the previous section, $\boldsymbol{Y}^*$ denotes the vector of all variables, observable or not, at all sampling sites. We already know that $E(\boldsymbol{X}_k(\boldsymbol{s}) \mid \boldsymbol{Y}^*) = \mathbf{V}_k \otimes \boldsymbol{r}_k'\Gamma(\boldsymbol{Y}^* - \boldsymbol{\mu} \otimes 1)$. Hence

$$E(\boldsymbol{X}_k(\boldsymbol{s}) \mid \boldsymbol{Y}) = (\mathbf{V}_k \otimes \boldsymbol{r}_k')\Gamma(E(\boldsymbol{Y}^* \mid \boldsymbol{Y}) - \boldsymbol{\mu} \otimes 1) \tag{24}$$

There is a nice interpretation to Equation (24). If some $Y_i(\boldsymbol{s}_j)$ is not observed at a sampling site $\boldsymbol{s}_j$, then predict it using the observed $\boldsymbol{Y}$. We then have values for all $\boldsymbol{Y}_i(\boldsymbol{s}_j)$ $(i = 1, \ldots, p, j = 1, \ldots, n)$, which are either actually observed or estimated. We then use this value and Equation (21) to predict $\boldsymbol{X}_k(\boldsymbol{s})$.

Prediction variance in the heterotopic case can also be calculated by observing the following

$$\begin{aligned} \mathrm{Var}(\boldsymbol{X}_k(\boldsymbol{s}) \mid \boldsymbol{Y}) &= E\{\mathrm{Var}(\boldsymbol{X}_k(\boldsymbol{s}) \mid \boldsymbol{Y}^*) \mid \boldsymbol{Y}\} + \mathrm{Var}\{E(\boldsymbol{X}_k(s) \mid \boldsymbol{Y}^*) \mid \boldsymbol{Y}\} \\ &= \mathbf{V}_k - (\mathbf{V}_k \otimes \boldsymbol{r}_k')\Gamma(\mathbf{V}_k \otimes \boldsymbol{r}_k) + (\mathbf{V}_k \otimes \boldsymbol{r}_k')\Gamma\mathrm{Var}(\boldsymbol{Y}^* \mid \boldsymbol{Y})\Gamma(\mathbf{V}_k \otimes \boldsymbol{r}_k) \end{aligned} \tag{25}$$

The conditional variance $\mathrm{Var}(\boldsymbol{Y}^* \mid \boldsymbol{Y}))$ can be given because $\boldsymbol{Y}^*$ and $\boldsymbol{Y}$ are Gaussian. It will depend which and where the variables $Y_i(\boldsymbol{s}_j)$ are not observed.

DOI: 10.1002/env

5. EXAMPLE

In this section, we illustrate the EM algorithm for LCM by analyzing a bivariate spatial sample. The Ozark Highlands of the USA have large areas of intensive poultry production. Poultry litter is applied to permanent pastures of the region and often serves as the sole nutrient source of forage growth. The litter is routinely applied at a rate to meet forage N requirements. Sauer and Meek (2003) studied the spatial variation of plant-available phosphorus in pastures at a site located in Eastern Benton County, Arkansas, USA, where a few variables of soil properties, including phosphorus, N, and C, were measured on soil samples. This site was divided into 96 grids, each of which is 30 m long and 30 m wide. Soil samples were taken and measured from each of the grids. In this work, we analyze two variables N and C from the study, which are highly correlated. Both variables have skewed distributions. We apply the log transformation to both variables. Figure 1 shows the scatter plot of the transformed variables, and clearly reveals the high correlation between the two variables.

Empirical variograms for both variables $\log(N)$ and $\log(C)$ reveal a relatively strong nugget effect for each of the two variables. Therefore the model should include the measurement error term as in model (3). Another practical issue in formulating the model is the number of Matérn components in the model. It is a difficult problem to determine the number of components due to the lack of rigorous theoretical results. In this work, we use a model with one Matérn component, hereafter referred to as Model 1, and a model with 2 Matérn components, referred to as Model 2. We choose a preferable model by comparing the likelihoods and predictive performance of the two models.

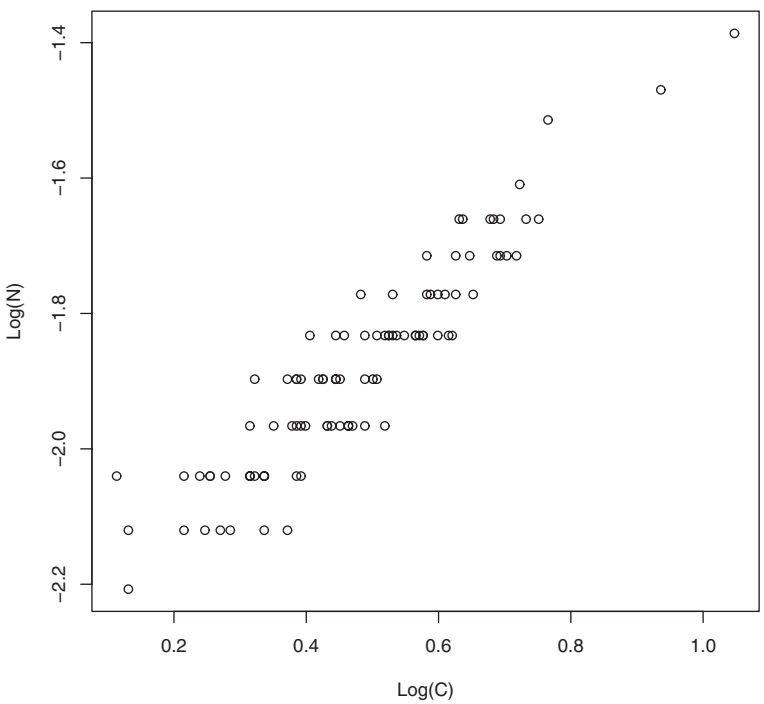

Figure 1. Scatter plot of log-transformed N against log-transformed C

DOI: 10.1002/env

Table 1. Summary of drop-one prediction for the two models

	Model 1		Model 2	
	log(*N*)	log(*C*)	log(*N*)	log(*C*)
MPE	−1.87848	0.48729	−1.87786	0.48642
MSPE	0.02373	0.02832	0.02374	0.02824

For Model 1, we discretize ν for $\nu = 0.25, 0.5, 1.0, 1.5, 2.0, 2.5$, and 3.0. There are two reasons for the discretization. First, precise estimation for ν may be hard to obtain as some numerical studies show (e.g., Stein, 1999, p220) that the data may not have sufficient information about the parameter ν. In general, data that collected at locations close to each other have more information about the smoothness parameter. Second, the discretization accelerates the convergence. For all practical purposes, the discretization should provide sufficient approximation. Applying the EM algorithm described in Section 3, we obtain the maximum likelihood estimates $\hat{\nu}_1 = 2.5$, $\hat{\psi}_1 = 31.06$ and

$$\hat{V}_0 = \begin{pmatrix} 0.13156 & 0.13150 \\ 0.13150 & 0.14531 \end{pmatrix}, \ \hat{V}_1 = \begin{pmatrix} 0.02228 & 0.01717 \\ 0.01717 & 0.02343 \end{pmatrix}$$

The estimates for the means of log(*N*) and log(*C*) are −1.8793 and 0.4966, respectively. The maximum log likelihood is 174.9.

For Model 2, we again discretize the parameters ν and take the values $\nu = (\nu_1, \nu_2)$ from the set $\{(\nu_1, \nu_2) : \nu_1 < \nu_2, \nu_1, \nu_2 = 0.25, 0.5, 1.0, 2.0, 2.5, 3.0\}$. The MLEs are $\hat{\nu}_1 = 1$, $\hat{\nu}_2 = 3$, $\hat{\psi}_1 = 6.01$, $\hat{\psi}_2 = 36.19$, and

$$\hat{V}_0 = \begin{pmatrix} 0.07168 & 0.07034 \\ 0.07034 & 0.07884 \end{pmatrix}, \ \hat{V}_1 = \begin{pmatrix} 0.06855 & 0.06677 \\ 0.06677 & 0.07486 \end{pmatrix}, \ \hat{V}_2 = \begin{pmatrix} 0.01276 & 0.00865 \\ 0.00865 & 0.01357 \end{pmatrix}$$

The estimates for the means of log(*N*) and log(*C*) are −1.8778 and 0.4962, respectively. The maximum log likelihood for this model is 175.7, which is slightly higher than that for Model 1. Considering that Model 2 has 5 more parameters than Model 1, the slightly higher log likelihood indicates that the Model 2 is perhaps not worthwhile.

Next we compare the predictive performance of the two models. We calculate the drop-one predictions so that at each of the 96 locations, say $\boldsymbol{s}_j$, we predict $Y(s_j) = (Y_1(s_j), Y_2(s_j))'$ using observations at all locations but $\boldsymbol{s}_j$. We calculate the plug-in predictions using both models and construct the mean prediction error (MPE) $\sum_{j=1}^{n}(Y_i(\boldsymbol{s}_j) - \hat{Y}_i(\boldsymbol{s}_j))/n$ and the mean-squared prediction error (MSPE). $\sum_{j=1}^{n}\{Y_i(\boldsymbol{s}_j) - \hat{Y}_i(\boldsymbol{s}_j)\}^2/n$, $i = 1$ for log(N) and $i = 2$ for log(C). Table 1 provides summaries of drop-one predictions.

The two models yield quite close predictive results while the second model provides a slightly smaller mean-squared prediction error. Hence model 1 is a parsimonious model with comparable performance to the second model, and is adopted in this work.

Next we generate a prediction surface for each variable, log(*N*) and log(*C*) by making predictions at 471 fine grid points. The predicted surfaces and contour plots are shown in Figure 2. These plots allow us to see the spatial variation of each of the two variables, log(*N*) and log(*C*), as well as how the two variables are spatially correlated with each other. For example, both variables have higher values near the center of the site while they both have lower values at the lower-right or the southeast corner.

DOI: 10.1002/env

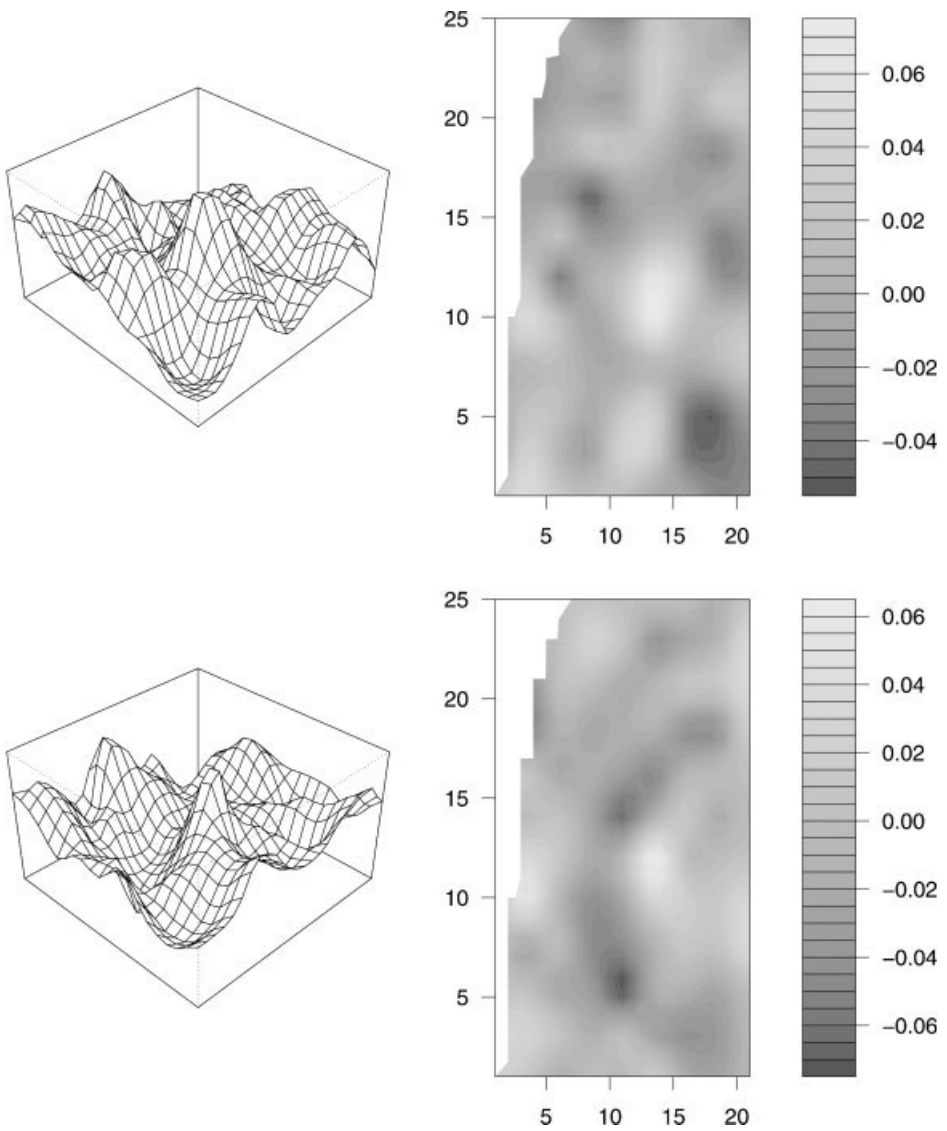

Figure 2. Perspective plot (left column) and contour plot for the prediction surface for variables log(*N*) (up row) and log(*C*) (lower row)

6. SUMMARY AND DISCUSSION

In this paper we developed an EM algorithm for the LCM, which can be implemented straightforwardly. Estimates for all parameters except those in the correlograms are updated in closed form. Therefore, this EM algorithm is capable of handling high-dimensional data. An increase in the dimension of the multivariate process does not cause a significant burden on computation.

DOI: 10.1002/env

In the LCM, we assumed that the components have Matérn covariograms. Zhang (2004) showed that for a univariate stationary process observed in a fixed spatial domain, not all parameters in the Matérn covariogram are consistently estimable. His techniques can be extended to the univariate LCM (i.e., $p = 1$) to show that not all parameters in the model are consistently estimable. In particular, the variances of the each component are not consistently estimable. There is no infill asymptotic results available for a multivariate spatial Gaussian process. However, we believe that inconsistency still exists in the multivariate case though we cannot justify that in this paper. For a given finite sample, the inconsistency usually translates into a large variance of estimator. More specifically, if a parameter cannot be estimated consistently under the infill asymptotic framework, its estimator usually has a large variance for a given finite sample.

This inconsistency problem only exists under the infill asymptotic framework. Under the increasing domain asymptotic framework in which the distance between any two spatial sampling locations is bounded from 0, all parameters are consistently estimable under regularity conditions. Recently, Zhang and Zimmerman (2005) compared the two asymptotic frameworks and concluded that for those parameters that can be estimated consistently, asymptotic results corresponding to the two frameworks approximate about equally well the finite sample distributions of the maximum-likelihood estimators of these parameters. However, for those parameters that cannot be estimated consistently under the infill asymptotic framework, the infill asymptotic approximation is preferable. Their results were established for univariate Gaussian process having an exponential covariogram. It will be a very interesting and harder problem to compare the two asymptotic frameworks for a multivariate process.

ACKNOWLEDGEMENTS

This work is partially supported by the National Science Foundation, DMS-0405782. The author thanks Dr Tom Sauer and Dr David Meek for providing data in the example, and a referee for helpful comments.

APPENDIX

We provide proofs for Equations (16)–(18). For any $k = 0, \ldots, K$, $\boldsymbol{X}_k, \boldsymbol{Y}$ have a joint multivariate normal distribution and $\mathrm{Cov}(\boldsymbol{X}_k, \boldsymbol{Y}) = \mathrm{Cov}(\boldsymbol{X}_k, \boldsymbol{X}_k) = \mathbf{V}_k^{(m)} \otimes R_k(\psi_k^{(m)})$, where hereinafter covariances and expectations are evaluated under $\theta = \theta^{(m)}$, and $R_0(\psi_0)$ is the constant $n \times n$ identify matrix. It follows the well-known properties of multivariate normal distribution that the conditional expectation and conditional covariance matrix of $\boldsymbol{X}_k$ given $\boldsymbol{Y}$ are

$$E(\boldsymbol{X}_k \mid \boldsymbol{Y}) = \mathbf{V}_k^{(m)} \otimes R_k(\psi_k^{(m)}) \boldsymbol{Y}^{(m)} \tag{26}$$

$$\mathrm{Var}(\boldsymbol{X}_k \mid \boldsymbol{Y}) = \mathbf{V}_k^{(m)} \otimes R_k(\psi_k^{(m)}) - \left(\mathbf{V}_k^{(m)} \otimes R_k(\psi_k^{(m)})\right) \Gamma^{(m)} \left(\mathbf{V}_k^{(m)} \otimes R_k(\psi_k^{(m)})\right) \tag{27}$$

where $\boldsymbol{Y}^{(m)}$ is given by Equation (15), and $\boldsymbol{\Gamma}^{(m)}$ is the inverse of $\mathrm{Var}(\boldsymbol{Y})$. To prove Equation (17), it suffices to show, from Equation (13), that

$$(I_p \otimes 1)' E(\boldsymbol{X}_0 \mid \boldsymbol{Y}) = \mathbf{V}_0^{(m)} \boldsymbol{Y}_{ma}^{(m)'} \mathbf{1}_n$$

DOI: 10.1002/env

From Equation (26), $E(\boldsymbol{X}_0 \mid \boldsymbol{Y}) = (\mathbf{V}_0^{(m)} \otimes I_n)\boldsymbol{Y}^{(m)}$. It follows that

$$(I_p \otimes 1)' E(\boldsymbol{X}_0 \mid \boldsymbol{Y}) = (I_p \otimes 1)'(\mathbf{V}_0^{(m)} \otimes I_n)\boldsymbol{Y}\boldsymbol{s}^{(m)} = (\mathbf{V}_0^{(m)} \otimes 1)' Y^{(m)}$$

It is straightforward to show

$$(\mathbf{V}_0^{(m)} \otimes 1')\boldsymbol{Y}^{(m)} = \mathbf{V}_0^{(m)} \boldsymbol{Y}_{ma}^{(m)'} 1_n$$

Indeed, the ith element of the left-hand side is $\sum_{j=1}^{p} \sigma_{ij,0}^{(m)} 1_n' \boldsymbol{Y}_j^{(m)}$, which equals to that of the right-hand side. Equation (17) is now proved.

To prove Equation (16), we apply the well-known fact that $E(X'AX) = E(X')AE(X) + \mathrm{tr}(A\mathrm{Var}(X))$. Then the conditional covariance in Equation (9) equals

$$(1/n)E(\boldsymbol{X}_{ki}' \mid \boldsymbol{Y})R_k^{-1}(\psi)E(\boldsymbol{X}_{ki} \mid \boldsymbol{Y}) + (1/n)\mathrm{tr}\{\mathrm{Cov}(\boldsymbol{X}_{ki}, \boldsymbol{X}_{ki} \mid \boldsymbol{Y})R_k^{-1}(\psi)\} \tag{28}$$

From Equation (26)

$$\begin{aligned} E_{\theta(m)}(\boldsymbol{X}_{ki} \mid \boldsymbol{Y}) &= \left(\mathbf{V}_k^{(m)}[i,] \otimes R_k(\psi_k^{(m)})\right)\boldsymbol{Y}^{(m)} \\ &= R_k(\psi_k^{(m)})\boldsymbol{Y}_{ma}^{(m)}\mathbf{V}_k^{(m)}[,i] \end{aligned} \tag{29}$$

It follows that the first term in Equation (28) equals the (i, j)th element of

$$(1/n)\mathbf{V}_k^{(m)}\boldsymbol{Y}_{ma}^{(m)'}W_k^{(m)}(\psi)\boldsymbol{Y}_{ma}^{(m)}\mathbf{V}_k^{(m)}$$

From Equation (27), the conditional covariance

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{X}_{ki}, \boldsymbol{X}_{kj} \mid \boldsymbol{Y}) &= \sigma_{ij,k}^{(m)} R_k(\psi_k^{(m)}) - \left(\mathbf{V}_k^{(m)}[i,] \otimes R_k(\psi_k^{(m)})\right)\Gamma^{(m)}\left(\mathbf{V}_k^{(m)}[,j] \otimes R_k(\psi_k^{(m)})\right) \\ &= \sigma_{ij,k}^{(m)} R_k(\psi_k^{(m)}) - \sum_{s=1}^{p}\sum_{t=1}^{p} \sigma_{is,k}^{(m)} R_k(\psi_k^{(m)}\Gamma_{st} R_k(\psi_k^{(m)}\sigma_{tj,k}^{(m)} \end{aligned} \tag{30}$$

Then the trace in Equation (28) equals

$$\sigma_{ij,k}^{(m)}\mathrm{tr}(R_k(\psi_k^{(m)})R_k^{-1}(\psi_k)) - \sum_{s=1}^{p}\sum_{t=1}^{p} \sigma_{is,k}^{(m)}\mathrm{tr}(R_k(\psi_k^{(m)})\Gamma_{st}R_k(\psi_k^{(m)}R_k^{-1}(\psi))\sigma_{tj,k}^{(m)}$$

Equation (17) now follows. Equation(18) can be established similarly.

REFERENCES

Abramowitz M, Stegun I. 1967. *Handbook of Mathematical Functions*. U.S. Government Printing Office: Washington, DC.

Chilés JP, Delfiner P. 1999. *Geostatistics: Modeling Spatial Uncertainty*. John Wiley: New York.

Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood estimation from incomplete data via the em algorithm (with discussion). *Journal of the Royal Statistical Society B* **39**: 1–38.

Goulard M, Voltz M. 1992. Linear coregionalization model: tools for estimation and choice of multivariate variograms. *Mathematical Geology* **24**: 269–286.

Mardia KV, Marshall RJ. 1984. Maximum likelihood estimation of models for residual covariance in spatial statistics. *Biometrika* **71**: 135–146.

Myers DE. 1991. Pseudo-cross variograms, positive-definiteness, and cokriging. *Mathematical Geology* **23**: 805–816.

DOI: 10.1002/env

Sauer T, Meek DW. 2003. Spatial variation of plant available phosphorus in pastures with contrasting management. *Soil Science and Social American Journal* **67**: 826–836.
Schott JR. 1997. *Matrix Analysis for Statistics*. John Wiley & Sons, Inc.: New York.
Stein ML. 1999. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer: New York.
Ver Hoef JM, Barry RP. 1998. Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference* **69**: 275–294.
Wackernagel H. 1998. *Multivariate Geostatistics: An Introduction with Applications*. Springer-Verlag Inc.: New York.
Zhang H. 2004. Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *Journal of the American Statistical Association* **99**: 250–261.
Zhang H, Zimmerman DL. 2005. Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92**: 921–936.
Zhu J, Eickhoff JE, Yan P. 2005. Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. *Biometrics* **61**: 674–683.

DOI: 10.1002/env