

Sensitivity analysis and model selection for a generalized convolution model for spatial processes

Anandamayee Majumdar*, Debashis Paul† and Jason Kaye‡

Abstract. We examine a flexible class of nonstationary stochastic models for multivariate spatial data proposed by [Majumdar et al. \(2010\)](#). This covariance model is based on convolutions of spatially varying covariance kernels with centers corresponding to the centers of “local stationarity”. A Bayesian method for estimation of the parameters in the model based on a Gibbs sampler is applied to simulated data to check for model sensitivity. The effect of a perturbation of the model in terms of kernel centers is also examined. Finally, the method is applied to a bivariate soil chemistry data from the the Central Arizona Phoenix Long Term Ecological Project (CAP LTER). Prediction bias, prediction standard deviation and predictive coverage are examined for different candidate models. In addition, a comparison with the bivariate stationary coregionalization model introduced by [Wackernagel \(2003\)](#) is carried out. A variant of the model proposed in [Majumdar et al. \(2010\)](#), with *random* kernel centers, is also examined. The latter model is seen to work much better than the stationary coregionalization model, and to perform comparably with the model with random kernel centers. Simulations indicate that the model is sensitive to under- or over-specification of kernel centers. On the other hand, application to real data seems to indicate that centroids of the regions that are homogeneous can be used as means of the random kernel centers. Cross validation can be used as a way of finding the best model with an appropriate number of kernels.

Keywords: convolution; multivariate spatial data; nonstationary covariance; posterior inference; model robustness

1 Introduction

Spatial modeling with flexible classes of covariance functions is one of the central topics of spatial statistics. A traditional approach is to model the spatial process as a parametric stationary process. However, the stationarity assumption is often violated in practice, particularly when the data come from large, heterogeneous, geographic regions. In soil science, environmental science, etc., it is often reasonable to view the data as realizations of processes that only in a small neighborhood of a location behave like stationary processes. In addition, there is a need to model two or more processes simul-

*Department of Mathematics and Statistics, Arizona State University, Tempe, AZ, <mailto:ananda@math.asu.edu>

†Department of Statistics, University of California, Davis, CA, <mailto:debashis@wald.ucdavis.edu>

‡Department of Crop Sciences, Pennsylvania State University, University Park, PA, <mailto:jpk12@psu.edu>

taneously and account for the possible correlation among various coordinate processes. For example, Majumdar and Gelfand (2007) considered an atmospheric pollution data consisting of 3 pollutants : CO , NO and NO_2 , whose concentrations in the atmosphere are correlated. Recently, Majumdar et al. (2010) proposed a flexible model for multivariate nonstationary spatial processes that allows for incorporation of extraneous information on possible spatial inhomogeneity. In the current paper, we study the aforementioned model in detail for sensitivity analysis and model selection. A key question studied here is the impact of possible mis-specification of the number and the locations of the centers of local stationarity for multivariate spatial processes.

Before going into the issues we are primarily concerned with in this paper, we give a brief overview of the existing literature on modeling nonstationary spatial processes. A considerable amount of work in spatial statistics has focussed on modeling *locally stationary* processes (Higdon 1997; Fuentes 2002; Nychka et al. 2002; Gelfand et al. 2004; Fuentes et al. 2005; Paciorek and Schervish 2006). Higdon et al. (1999) and Higdon (2002) modeled the process as a convolution of a stationary process with a kernel of varying bandwidth. Fuentes and Smith (2001) and Fuentes (2002) considered a convolution model in which the kernel has a fixed bandwidth, while the process has a spatially varying parameter. Nychka et al. (2002) considered a multiresolution analysis-based approach to model the spatial inhomogeneity that utilizes the smoothness of the process and its effect on the covariances of the basis coefficients, when the process is represented in a suitable wavelet-type basis. One of the central themes of the various modeling schemes described above is that a process may be represented in the spectral domain *locally* as a superposition of Fourier frequencies with suitable (possibly spatially varying) weight functions. Paciorek and Schervish (2006) derived an explicit representation for the covariance function for Higdon's model when the kernel is multivariate Gaussian and used it to define a nonstationary version of the Matérn covariance function by utilizing the *Gaussian scale mixture* representation of positive definite functions.

The modeling approaches mentioned so far focus primarily on univariate processes. Existing approaches to modeling the multivariate processes include the work by Gelfand et al. (2004) that utilizes the idea of *coregionalization* to model the covariance of $\mathbf{Y}(s)$ (taking values in \mathbb{R}^N) as

$$\text{Cov}(\mathbf{Y}(s), \mathbf{Y}(s')) = \sum_{j=1}^N \rho_j(s - s') \mathbf{T}_j,$$

where $\rho_j(\cdot)$ are stationary covariance functions, and \mathbf{T}_j are positive semidefinite matrices of rank 1. Christensen and Amemiya (2002) considered a different class of multivariate processes that depend on a latent shifted-factor model structure.

Majumdar et al. (2010) deal with the question of modeling nonstationary multivariate spatial processes. The model proposed there can be viewed as a generalization of the convolution model for correlated Gaussian processes proposed by Majumdar and Gelfand (2007). The authors extend their aforementioned model to nonstationary settings by incorporating information on local inhomogeneity through specification of kernel weights. A key motivation is the assertion that when spatial inhomogene-

ity in the process is well-understood in terms of dependence on certain features of the geographical locations (e.g. local values of some covariates), it makes sense to use that information directly in the specification of the covariance kernel. For example, soil concentrations of Nitrogen, Carbon, and other nutrients and/or pollutants, that are spatially distributed, are relatively homogeneous across similar land use types (e.g. agricultural, urban, desert, transportation - and so on), but are non-homogeneous across spatial locations with different land-use types. Usually the land use types and their boundaries are clearly known (typically from satellite imagery). Thus in this instance nonstationary models will be clearly advantageous compared to stationary models. Another example concerns the land values and different economic indicators in a region. Usually land values are higher around (possibly multiple) business centers, and such information may be incorporated as the known centers of the kernels in the model (1) for the covariance kernel of the spatial process. The models (1) proposed by Majumdar et al. (2010) behave locally like stationary processes, but are *globally* nonstationary. Moreover, this model allows for the degree of correlations among the coordinate processes to vary across different spatial scales.

In this paper, model performance is examined by varying the underlying parameters of the model, when the number and/or locations of the centers of local stationarity are misspecified. We then concentrate on an application to the soil chemistry data discussed earlier, where the different land use zones are well-known. Satellite maps provide the necessary geographic information about the land use zones that can be used to specify the kernel centers in this application. But it may well happen that some of the land use zones are redundant in terms of model performance, or that a land use subtype not taken into account previously plays a significant role in terms of determining the spatial inhomogeneity. Hence an appropriate model selection procedure for choosing the centers of local stationarity is important for the specification of an “optimal” model from the point of view of predictive performance. We propose and examine one such method by utilizing the principle of cross-validation. In addition to avoiding redundancy and over-parametrization, these results may also throw light on the hitherto unknown ecological/environmental aspects of the data.

The main contributions of this paper are: (i) testing for model sensitivity when the number and centers of local stationarity are misspecified; (ii) an application to a bivariate soil chemistry data with focus on model selection; and (iii) data analysis extended to the case when kernel centers are considered random, also as a part of model selection. We present a Bayesian estimation procedure based on *Gibbs sampling* for estimating the parametric covariance functions and predicting the outcomes, and apply it to simulated and real data.

This paper is organized as follows. In Section 2, we present a special parametric subclass of the bivariate spatial model specified in Majumdar et al. (2010) to illustrate some of the key features. In Section 3, we give a summary of the simulation studies based on this model that illustrates the sensitivity to mis-specifications of the number of kernels in terms of under- or over-representation of the centers of local stationarity. We measure the performance of the mis-specified model using prediction coverage of 50 hold out points. In Section 4 we discuss the application of the generalized convolution

model to a bivariate soil-chemistry data coming from the CAP LTER project in the Phoenix metropolitan area. In Section 6, we discuss some practical aspects of modeling using the current framework and the related research directions. Some technical details can be found in the *supplementary materials* of Majumdar et al. (2010).

2 Specification of the nonstationary covariance model

In this section we give a short description of the model proposed in Majumdar et al. (2010). This model has a natural appeal from the perspective of modeling spatially inhomogeneous multivariate processes, and renders the problem of estimating the nonstationary covariance kernel computationally quite tractable. The general form of the covariance kernel for an N -dimensional nonstationary process in \mathbb{R}^d is described below.

Let $\{t_l : l = 1, \dots, L\}$ be a sequence of points in \mathbb{R}^d ; for each l , let $K_l(\cdot)$ be a nonnegative kernel with $\int K_l(x)dx = 1$; and let $\{\Sigma_l : l = 1, \dots, L\}$ be a sequence of $d \times d$ positive definite matrices. Define, for $\omega \in \mathbb{R}^d$,

$$f_j(s, \omega) = \sum_{l=1}^L |\Sigma_l|^{-1/2} K_l(\Sigma_l^{-1/2}(s - t_l)) f_j(\omega; \theta_{jl}), \quad j = 1, \dots, N, \quad (1)$$

where for every fixed (j, l) , $f_j(\cdot; \theta_{jl})$ is a spectral density function belonging to a parametric family parameterized by θ_{jl} . Let, $\rho_{jj'}(\omega) = \rho_0(\omega; \nu_{jj'}, \kappa)$, for parameters $\{\nu_{jj'}\}_{j,j'=1}^N$ and κ be such that the matrix $\mathbb{R}(\omega) = ((\rho_{jj'}(\omega)))_{j,j'=1}^N$ is positive definite with diagonal entries equal to 1. Also, let $\rho_1(s - t; \tau)$, for some parameter τ , be the covariance kernel of a stationary process in \mathbb{R}^d . Then, the covariance kernel $C^*(s, t)$ of the N -dimensional process, is determined through

$$\begin{aligned} C_{jj'}^*(s, t) &= \rho_1(s - t; \tau) \sum_{l,l'=1}^L |\Sigma_l|^{-1/2} K_l(\Sigma_l^{-1/2}(s - t_l)) |\Sigma_{l'}|^{-1/2} K_{l'}(\Sigma_{l'}^{-1/2}(t - t_{l'})) \\ &\quad \cdot \int_{\mathbb{R}^d} e^{i\omega^T(s-t)} f_j(\omega; \theta_{jl}) f_{j'}(\omega; \theta_{j'l'}) \rho_{jj'}(\omega) d\omega, \quad 1 \leq j, j' \leq N. \end{aligned} \quad (2)$$

Thus, defining $G_{jj'}(s; \theta_{jl}, \theta_{j'l'}, \nu_{jj'}, \kappa) = \int_{\mathbb{R}^d} e^{i\omega^T s} f_j(\omega; \theta_{jl}) f_{j'}(\omega; \theta_{j'l'}) \rho_0(\omega; \nu_{jj'}, \kappa) d\omega$,

$$\begin{aligned} C_{jj'}^*(s, t) &= \rho_1(s - t; \tau) \sum_{l,l'=1}^L |\Sigma_l|^{-1/2} K_l(\Sigma_l^{-1/2}(s - t_l)) |\Sigma_{l'}|^{-1/2} K_{l'}(\Sigma_{l'}^{-1/2}(t - t_{l'})) \\ &\quad \cdot G_{jj'}(s - t; \theta_{jl}, \theta_{j'l'}, \nu_{jj'}, \kappa). \end{aligned} \quad (3)$$

Typically, the sequence $\{t_l\}_{l=1}^L$ may be assumed given.

2.1 A bivariate process through specification of a parametric spectral density

Majumdar et al. (2010) considered a special case of the process with covariance kernel given by (3) to maintain a balance among flexibility, computational cost and interpretability. By assuming that $f_j(\omega; \theta_{jl})$ is of the form $c_{jl}\gamma(\omega; \tilde{\theta}_{jl})$ for some scale parameter $c_{jl} > 0$, for a parametric class of spectral densities $\gamma(\cdot; \tilde{\theta})$ that is closed under product, and choosing $\rho_0(\omega; \nu_{jj'}, \kappa)$ to be of the form $\nu_{jj'}\alpha(\omega; \kappa)$, where $\alpha(\omega; \kappa) \equiv \alpha(\omega)$ is a real-valued function satisfying $-\frac{1}{N-1} \leq \alpha(\omega) \leq 1$, and the $N \times N$ matrix $\mathbf{N} = ((\nu_{jj'})_{1 \leq j, j' \leq N})$ is a positive definite correlation matrix, they obtained a closed form expression for the functions $G_{jj'}(s - t; \theta_{jl}, \theta_{j'l'}, \nu_{jj'}, \kappa)$. The details can be found in Section 3 of Majumdar et al. (2010).

For the applications we focus on in this paper, the process is two-dimensional (that is, $N = 2$) on a two-dimensional domain (that is, $d = 2$). For example, the two variables may be soil salinity and soil moisture content, or, temperature and pressure fields, etc. We focus on a special case of the aforementioned model. The model specification in the bivariate case simplifies considerably by making the additional assumption that the spectral density γ is Gaussian. For this reason, we give a detailed description of the model and the estimation procedure for this special case only. We model $\gamma(\cdot; \tilde{\theta})$ as a Gaussian spectral density with scale parameter $\tilde{\theta}$, so that $\gamma(\omega; \tilde{\theta}) = \frac{1}{2}(\pi\tilde{\theta})^{-\frac{1}{2}}e^{-\omega^2/4\tilde{\theta}}$; and that $\rho_1(s - t; \tau) = e^{-\tau\|s-t\|}$ (exponential correlation kernel). Next, we parametrize the scaling matrices Σ_l by the Cholesky decomposition (with $\Sigma_l^{-1} = \Sigma_l^{-1/2}(\Sigma_l^{-1/2})^T$):

$$\Sigma_l^{-1/2} = \begin{pmatrix} \sigma_{11l} & 0 \\ \sigma_{21l} & \sigma_{22l} \end{pmatrix}. \tag{4}$$

With these, we obtain a simplified expression for the covariance kernels $C_{jj'}^*(s, t)$. For ease of expressions, we consider the case when $\tilde{\theta}_{jl} = \tilde{\theta}_l$ for $j = 1, 2$. Then,

$$C_{jj'}^*(s, t) = e^{-\tau\|s-t\|} \frac{1}{2\pi} \sum_{l, l'=1}^L \sigma_{11l}\sigma_{22l}\sigma_{11l'}\sigma_{22l'}c_{jl}c_{j'l'}\nu_{jj'}\Gamma_{jj'}(s - t; \tilde{\theta}_l, \tilde{\theta}_{l'}, \kappa) \cdot \exp\left(-\frac{1}{2} \|\Sigma_l^{-1/2}(s - t_l)\|^2 - \frac{1}{2} \|\Sigma_{l'}^{-1/2}(t - t_{l'})\|^2\right), \tag{5}$$

where $\kappa = (\alpha_1, \alpha_2, \beta)$, $\Sigma_l^{-1/2}$ has the form (4), and the $\Gamma_{jj'}(s; \tilde{\theta}_l, \tilde{\theta}_{l'}, \kappa)$ are given as follows.

$$\begin{aligned} \Gamma_{12}(s; \tilde{\theta}_l, \tilde{\theta}_{l'}, \kappa) &= \Gamma_{21}(s; \tilde{\theta}_{l'}, \tilde{\theta}_l, \kappa) \\ &= \frac{1}{2\sqrt{\pi}} \left[g_1(\alpha_1, \tilde{\theta}_l, \tilde{\theta}_{l'})e^{-g_2(\alpha_1, \tilde{\theta}_l, \tilde{\theta}_{l'})\|s\|^2} - \beta g_1(\alpha_2, \tilde{\theta}_l, \tilde{\theta}_{l'})e^{-g_2(\alpha_2, \tilde{\theta}_l, \tilde{\theta}_{l'})\|s\|^2} \right], \end{aligned} \tag{6}$$

$$\Gamma_{jj}(s; \tilde{\theta}_l, \tilde{\theta}_{l'}; \kappa) = \frac{1}{2\sqrt{\pi}} \frac{1}{\sqrt{\tilde{\theta}_l + \tilde{\theta}_{l'}}} \exp\left(-\frac{\tilde{\theta}_l\tilde{\theta}_{l'}}{\tilde{\theta}_l + \tilde{\theta}_{l'}} \|s\|^2\right), \quad j = 1, 2 \tag{7}$$

where, for $k = 1, 2, 1 \leq l, l' \leq L$,

$$g_1(\alpha_k, \tilde{\theta}_l, \tilde{\theta}_{l'}) = \sqrt{\frac{\alpha_k}{\alpha_k(\tilde{\theta}_l + \tilde{\theta}_{l'}) + \tilde{\theta}_l\tilde{\theta}_{l'}}}, \quad g_2(\alpha_k, \tilde{\theta}_l, \tilde{\theta}_{l'}) = \frac{\alpha_k\tilde{\theta}_l\tilde{\theta}_{l'}}{\alpha_k(\tilde{\theta}_l + \tilde{\theta}_{l'}) + \tilde{\theta}_l\tilde{\theta}_{l'}}. \tag{8}$$

3 Simulation results

We discuss some simulation results for the special case of the bivariate model specified in Section 2.1. We fixed $\sigma_{11l} = \sigma_{11}$, $\sigma_{22l} = \sigma_{22}$, $\sigma_{21l} = \sigma_{21}$, $c_{jl} = c$, and $\tilde{\theta}_{jl} = \theta$ for all $l = 1, \dots, L$; and $\alpha_1 = \alpha_2 = \alpha$. Note that, with this simplification, we can reparametrize by defining $\gamma := \theta^2/\alpha$ so that the functions g_1 and g_2 defined in (8) can be expressed as $g_1(\gamma, \theta) = 1/\sqrt{2\theta + \gamma}$ and $g_2(\gamma, \theta) = \theta^2/(2\theta + \gamma)$. Since β and ν_{12} are not identifiable together in this bivariate model, we set $\beta = 0$. Further, we set $\sigma_{11} = \sigma_{22} = 1$, $\tau = 0.1$. The parameters c , θ , α , σ_{21} , and ν_{21} are chosen to be random. We generated bivariate Gaussian data with mean $\mathbf{0}$.

For estimation, we treated β , σ_{11} , σ_{22} and τ as *known*, and the other five parameters, namely c , θ , α , σ_{21} , and ν_{21} , as *unknown* and estimated them from the data using a Gibbs sampling procedure. From equations (5), (6) and (7), c^2 is a scale parameter, we used an *InvGamma*(2, 1) prior for c^2 . For the (positive) parameter α , we assumed a *Gamma*(0.01, 10) prior. For θ , we assumed a *Gamma*(0.1, 10) prior. Since ν_{21} is restricted to the interval $(-1, 1)$, and is a measure of global association between processes, we assumed a positive association through a *Uniform*(0, 1) prior. Finally, we chose a *N*(0, 10) prior for σ_{21} .

The posterior distribution of c^2 is an Inverse Gamma. The posterior distributions of the rest of the parameters do not have closed forms. Hence we employed Gibbs sampling within a Metropolis Hastings algorithm to obtain posterior samples of the parameters. Burn-in was obtained with 2000 iterations, and we thinned the samples by 20 iterations to obtain 1000 uncorrelated samples from the joint posterior distribution of $(c, \theta, \alpha, \sigma_{21}, \nu_{21})$ given the data. Sensitivity analysis of the priors was carried out by varying the means and variances. The priors prove to be fairly robust with respect to the posterior inference results. For data simulated using $n = 50$ locations, we present further results in the subsections below.

3.1 Sensitivity to under-specification of kernels

To understand the sensitivity of the model to under-specification of number of kernel centers, or equivalently the number of centers of “local stationarity”, we performed some simulation studies for the bivariate ($N = 2$) case, in which we specified the number of kernels, $L = 4$; $\sigma_{1l} = \sigma_{22l} = \sigma_{21l} = 1$, $c_{jl} = 2$, $\theta_{jl} = 0.1$ for all $j = 1, \dots, N$, $l = 1, \dots, 4$; $\tau = 0.5$, $\nu_{12} = \nu_{21} = 0.8$, $\alpha_1 = \alpha_2 = \alpha$, $\beta = 0.1$. We generated 100 realizations (on the unit square $[0, 1] \times [0, 1]$) of a bivariate spatial process with centers of the four kernels $t_1 = (0.1, 0.7)$, $t_2 = (0.6, 0.1)$, $t_3 = (0.9, 0.6)$, $t_4 = (0.6, 0.9)$. We then used three of the kernel centers in lieu of all four, and carried out prediction at 50 holdout points. The prediction coverage of the 95% prediction intervals was noted in 100 simulations. The results are summarized in the table of figures Table 1. The locations of the kernel centers are denoted in figures (a)-(d) in Table 1. The blue square corresponds to the kernel that was left out during estimation, to distinguish the kernel from the holdout points, even though the true model included this kernel. The red triangles correspond to spatial points where the prediction coverage was below 85%. In Table 1, we note

that out of 50, there are at most 3 points that have prediction coverage below 85%, and about 94% of the remaining data yields more than 95% coverage. Similarly, in Table 3, for the second spatial process, there is one case for which all prediction intervals resulted in 95% or more coverage. In the three remaining combinations, the maximum number of spatial points where the prediction coverage was below 85% was 4. The remainder of the 92% of the holdout points yield at least 95% coverage. We observe that the points where the prediction is less than satisfactory are spatial points outside the convex hull of the kernel centers used in the model.

3.2 Sensitivity to over-specification of kernels

We next perform simulation studies to understand the sensitivity of the model to over-specification of number of kernel centers or equivalently the over-specifications of the number of center of “local stationarity”. We again performed simulation studies for the bivariate ($N = 2$) case, in which we set the number of kernels, $L = 3$, $\sigma_{1l} = \sigma_{22ll} = \sigma_{21l} = 1$, $c_{jl} = 2$, $\theta_{jl} = 0.1$ for all $j = 1, \dots, N$, $l = 1, \dots, 4$; $\tau = 0.5$, $\nu_{12} = \nu_{21} = 0.8$ and $\alpha_1 = \alpha_2 = \beta = 0.1$. We generated 100 realizations (on the unit square $[0, 1] \times [0, 1]$) of a bivariate spatial process with centers of the three out of four kernel centers: $t_1 = (0.1, 0.7)$, $t_2 = (0.6, 0.1)$, $t_3 = (0.9, 0.6)$, $t_4 = (0.6, 0.9)$. We then used all four of the kernel centers in lieu of the three and carried out prediction at 50 holdout points. Again, the prediction coverage of the 95% prediction intervals were noted in 100 simulations. The results are summarized in Table 4. The locations of the kernel centers used in the MCMC are displayed in figures (a)-(b). The blue spatial point corresponds to the kernel center that was redundant, since the true model did not include this. The red triangles correspond to spatial points where the prediction coverage was below 80% and the violet star corresponds to a point that has below 85% but above 80% coverage. In Table 4, focusing on the first spatial process, we note that out of 50, there are three points that yield prediction coverage below 80%, and one point that is close to 85%, about 92% of the remaining hold-out points yield more than 95% coverage. In Table 4, for the second spatial process, we considered one case for which all but one prediction intervals resulted in 95% or more coverage. There was one spatial point for which prediction coverage was only 56%. We again note that the points where the prediction result is less than satisfactory are spatial points fall outside or just outside the convex hull of the kernel centers used in the model.

3.3 Sensitivity to mis-specification : impact of parameter values

We next study the effect of the parameters on the performance of the mis-specified models. We performed simulation studies for the bivariate case in which we specified the number of kernels, $L = 4$; $\sigma_{1l} = \sigma_{22ll} = \sigma_{21l} = 1$, $c_{jl} = 2$, $\theta_{jl} = 0.1$ for all $j = 1, \dots, N$, $l = 1, \dots, 4$; $\tau = 0.5$, $\nu_{12} = \nu_{21} = 0.8$, and $\alpha_1 = \alpha_2 = \beta = 0.1$. We then kept all the parameters fixed and varied one parameter at a time to realize the effect of the parameter on the under-specified model. As before, we generated 100 realizations (on the unit square $[0, 1] \times [0, 1]$) of a bivariate spatial process with four kernel centers:

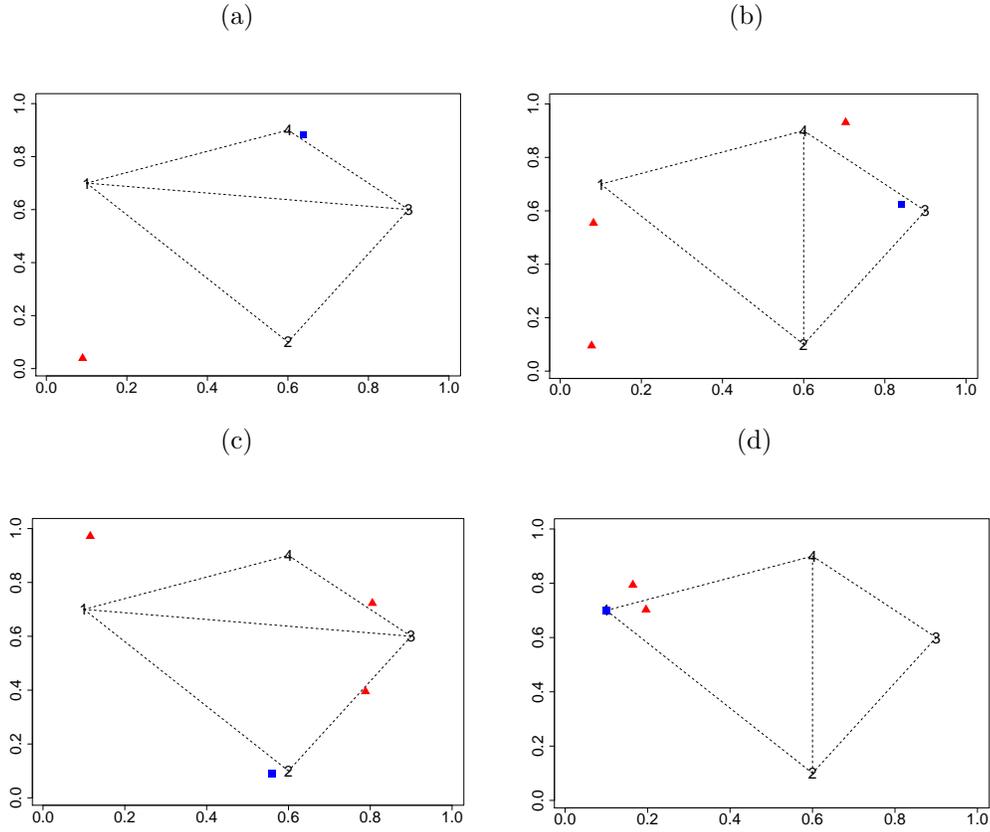


Table 1: Spatial plot identifying low prediction coverage points for the 1st process when kernel centers are under-specified. Red triangles identify locations of low ($< 80\%$) posterior predictive coverage and blue squares identify the kernel center that was erroneously left out of the model during estimation.

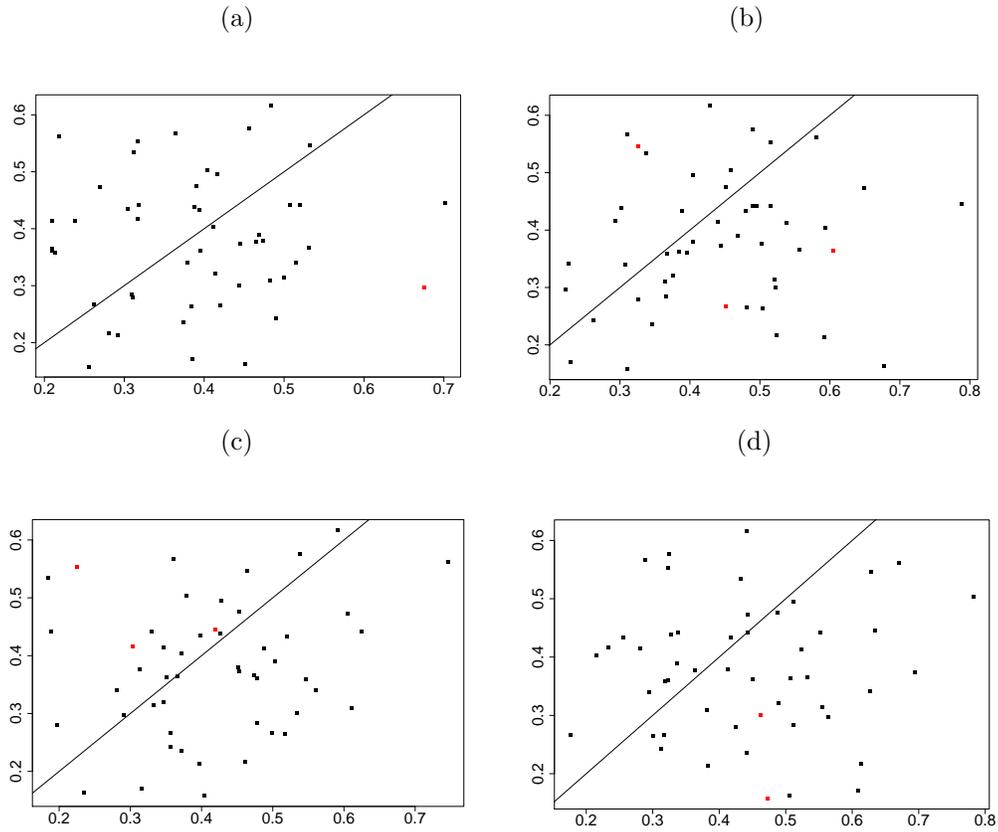


Table 2: Comparison of the length of prediction intervals for correct and under-specified (in terms of kernel centers) models for the 1st process. The length of the prediction interval under the model represents the Y-axis and the same under the under-specified model represents the X-axis. Red triangles identify points of low posterior predictive coverage. The solid line represents the $X = Y$ line.

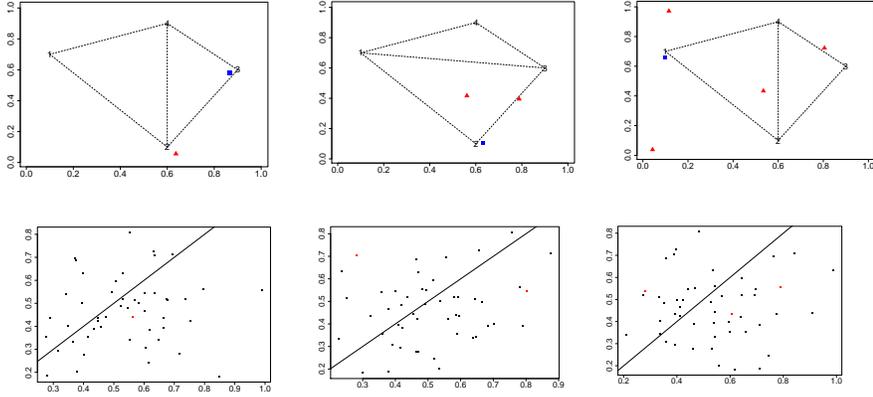


Table 3: Top panel: Spatial plot identifying low prediction coverage points for the 2nd process when kernel centers are under-specified (blue squares identifying the kernel center erroneously left out). Red triangles identify points with low prediction coverage. Bottom panel: Comparison of the length of prediction intervals for correct and under-specified (in terms of kernel centers) for the 2nd process. The length of the prediction interval under the correct model represents the Y-axis and the same for the under-specified model represents the X-axis (the solid line represents the $X = Y$ line). Red triangles identify locations of low posterior predictive coverage.

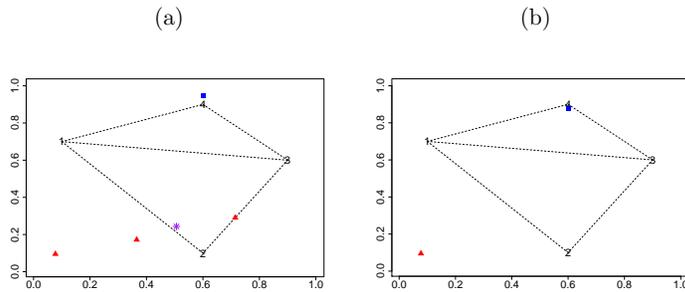


Table 4: Spatial plot identifying low prediction coverage points when kernel centers are over-specified (blue squares identifying the kernel center that was erroneously included). Red triangles identify locations of low (< 80%) posterior predictive coverage. A violet star identify locations of medium (80%-85%) posterior predictive coverage. Left panel represents the first, and right panel represents the second spatial process

$t_1 = (0.1, 0.7)$, $t_2 = (0.6, 0.1)$, $t_3 = (0.9, 0.6)$, $t_4 = (0.6, 0.9)$. We then used the first three kernel centers in lieu of the four and carried out prediction at 50 holdout points. Again, the prediction coverage of the 95% prediction intervals were noted in 100 simulations. The results are summarized in the following table of figures. The locations of the kernel centers used in the MCMC are denoted from 1-4 in the figures. The blue spatial point corresponds to the kernel center that was not used in prediction, even if the true model did include it. The red points correspond to spatial points where the prediction coverage was below 80%.

In Table 5, we focus on the effect of changing the “cross-correlation parameter” $\nu_{12} = \nu_{21}$. Observe that ν_{12} controls the degree of association between the two coordinate processes. When we changed this parameter from 0.8 (top panel) to 0.2 (bottom panel), the number of points that yield prediction coverage below 80% went up from one (2%) to three (6%) for the first process and from none (0%) to three (6%) in the second process. This is to be expected as a higher absolute value of the parameter ν_{12} involves more sharing of information between process I and process II, and thereby is expected to predict better. We also study the change regarding the parameter $\sigma_{21l} = \sigma_{21}$. By changing this parameter from 1 to 0.5, the number of points that yield prediction coverage below 80% went up from one (2%) to six (12%) for the first process and from none (0%) to two (4%) for the second process. Again, this is to be expected as a higher absolute value of the parameter σ_{12} involves higher spatial covariance within each process, and thereby by sharing of spatial information of each is expected to predict better. The same direction of impact is observed on over-specified models when these parameters are changed.

Next, we studied the impact of assuming misspecified values of the parameter $\sigma_{1l} = \sigma_{22ll}$ which had been considered fixed ($= 1$) throughout in this exercise. In table 6, we a) under-specified this value to 0.5 (top panel) and also b) over-specified the same value to 2 (lower panel), and thus, compared average prediction coverage (left panels) and average prediction interval lengths (right panels), between correct and misspecified models. For both cases the misspecified model yielded poor performance regarding both prediction coverage and length of prediction intervals, for both Y_1 and Y_2 processes. The lesson to take home from this exercise is that the diagonal elements of Σ_l need to be treated as random to yield better prediction performance.

We also studied the impact of assuming misspecified values of the “global association” parameter ν_{21} . When the true value of $\nu_{21} = 0.8$, for the estimation procedure we chose a fixed value of $\nu_{21} = 0.0$, comparing it with the “correct” approach where ν_{21} was assumed random. The results are illustrated in Table 7. The misspecified model yielded worse performance regarding prediction coverage. However, the length of the prediction interval on an average was slightly larger for the first process and a bit smaller for the second process. Hence we can conclude that, disregarding the association parameter (i.e., assuming independence of processes) tends to decrease the prediction coverage, but may not adversely affect the prediction accuracy as determined by the length of the prediction interval.

Next, we compare performance of the generalized convolution model performance in

terms of parameter estimation, when using only one process realization (for example, Y_1) versus the “full” bivariate process realization (Y_1 and Y_2). The comparison of prediction performance is demonstrated in table 8. For this comparison, we use the median of the MSE (left panel) and standard deviation of the MSE (right panel) of the relevant random parameters – namely, c , σ_{21} , θ and a . It is clear from the table, that while the parameters a and c have comparable values of median MSE and sd MSE, the bivariate model does a better job at estimation for the parameters θ and σ_{21} .

Finally, we compare performance of the generalized convolution model performance in terms of prediction coverage and prediction accuracy, when using only one process realization (for example, Y_1) versus the “full” bivariate process realization (Y_1 and Y_2). The comparison of estimation performance between bivariate and univariate models is demonstrated in table 9. We observe that using the bivariate model improved the prediction coverage. The length of prediction intervals had an overall reduction, even though the length of the prediction intervals for the bivariate cases were more tight. Overall, we expect some decline in prediction coverage when making use of only one process realization, compared to the situation where we make use of the bivariate process realization. In our case, points where the prediction coverage was below 80% went from 0 (0%) to 5 (10%) by ignoring the bivariate process realization and using only one process realization.

3.4 Sensitivity to jitter in the kernel centers

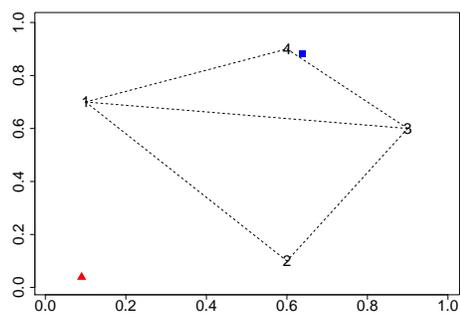
Our next step was to examine the performance of the model when the kernel centers had a jitter added to them. In the simulations, instead of the true kernel centers, the model used had slight random, independent shifts for each of the centers. Prediction coverage for 50 hold out points indicated that some prediction intervals had low coverage. An increase in the magnitude of the jitter worsened the model performance. A way to address the problem was to use kernel centers at the corners of the unit square and one kernel center at the center of the unit square, thus spacing out the centers, and this improved the model performance quite a bit in terms of the prediction coverage. Note that points too far from all the kernel centers could have a low impact in the model, and thereby points far away from any kernel center might yield bad prediction. Thus, when the model is known to be nonstationary, yet the centers of local stationarity can not be identified, this idea of spacing out the kernel centers may yield reasonable results. We note that model choice is still required by varying the number of the kernel centers.

4 Data Application: The CAP LTER soil Nitrogen and Carbon content

Two of the most basic metrics of ecosystem structure are the quantities of carbon (C) and nutrients such as nitrogen (N) stored in soil. Soil C is the largest store of carbon in the terrestrial biosphere, so understanding spatial variation in this property is essential for improving understanding of soil-carbon-climate change linkages ([Jobbagy and Jackson](#)

Process I

Process II



Process I

Process II

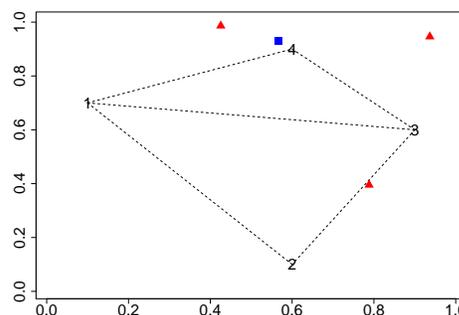
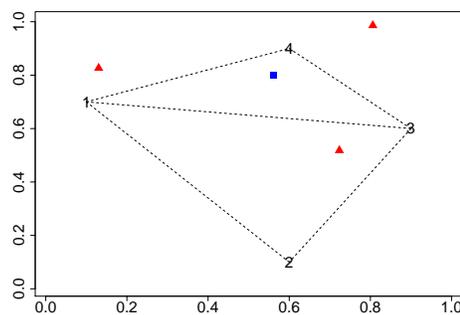
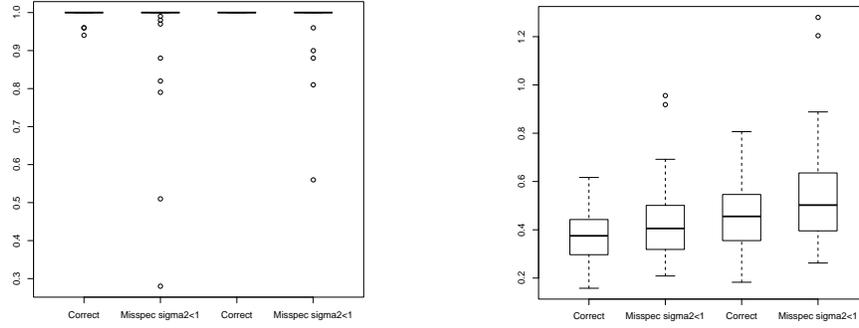


Table 5: Comparison of prediction coverage for different “true” parameter values. Red triangles identifying spatial points with low posterior predictive coverage for the first (left) and second (right) spatial process when we have $\nu_{12} = 0.8$ (top panel) and $\nu_{12} = 0.2$ (bottom panel). Note that for process II, there are no spatial points with low post. pred. coverage when we have $\nu_{12} = 0.8$ (top right panel)

(a)



(b)

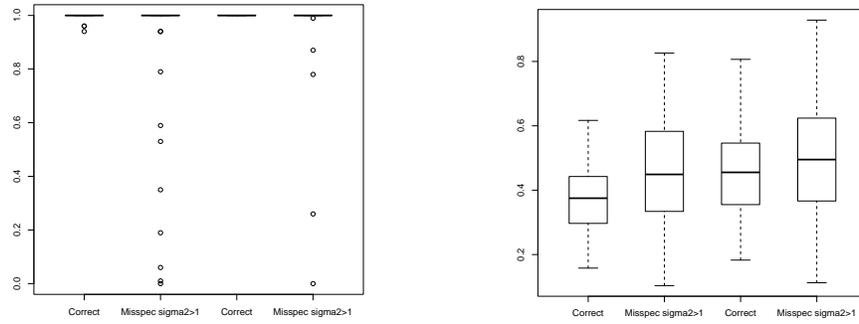


Table 6: Comparison between correct and misspecified models when (a) $\sigma_{1l} = \sigma_{22ll} = 0.5$ is erroneously used (half the correct value) and when (b) $\sigma_{1l} = \sigma_{22ll} = 2$ is erroneously used (twice the correct value). The left panel compares prediction coverage, the right panel compares the length of the intervals. In each figure the left two box plots correspond to Y_1 and the two right box plots correspond to Y_2 .

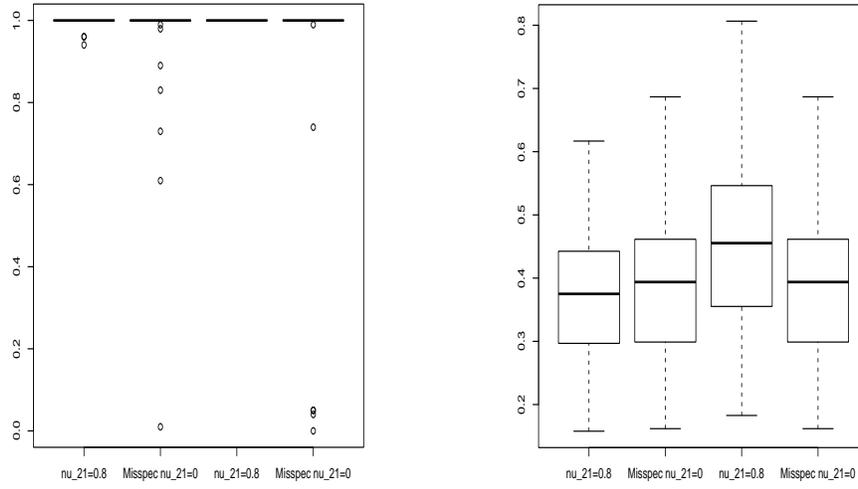
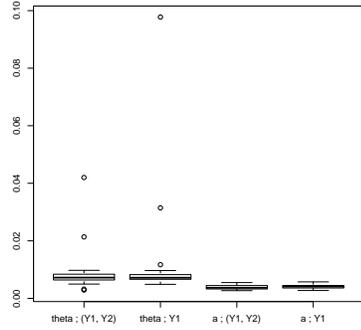


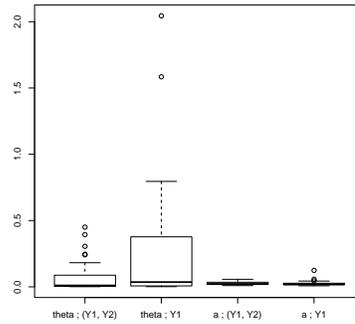
Table 7: Comparison between correct and misspecified models when $\nu_{21} = 0$ corresponding to no association is erroneously assumed (correct value is $\nu_{21} = 0.8$). Left panel compares prediction coverage and right panel compares the length of the intervals. In each figure the two leftmost box plots correspond to Y_1 and the two rightmost box plots correspond to Y_2 .

(a) Parameters θ and a

median of MSE

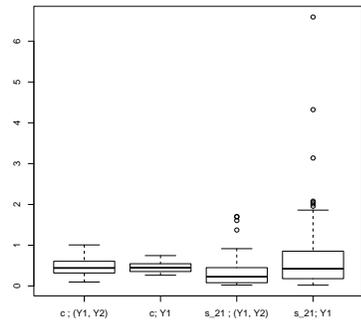


sd of MSE



(b) Parameters c and σ_{21}

median of MSE



sd of MSE

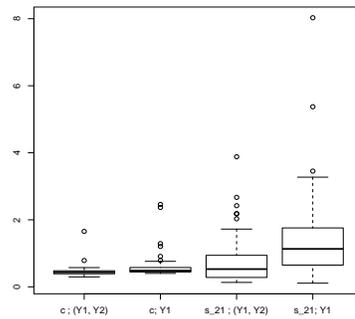


Table 8: Comparison of model estimation performance between bivariate and univariate processes for parameters (a) θ and a and (b) c and σ_{21} . Left panel compares median (med) MSE and right panel compares standard deviation (sd) of the MSE's.

Prediction coverage

Length of prediction intervals

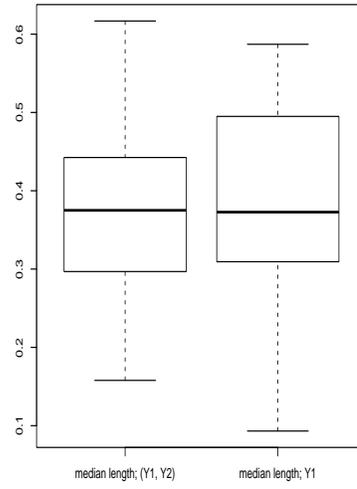
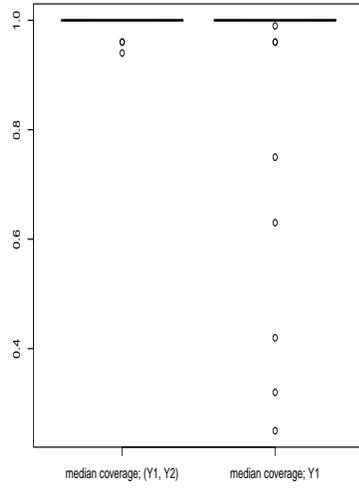


Table 9: Comparison between the bivariate and the univariate model. Left panel compares prediction coverage and the right panel compares the lengths of the intervals. In each figure the box plot on the left corresponds to (Y_1, Y_2) and the box plot on the right corresponds to Y_1 .

2000). Soil N is a major source of nutrients to plants and is also an important sink for atmospheric nitrogen pollution, thus, again, understanding variation in this property will increase our ability to predict environmental change (Vitousek et al. 1997). In previous studies, a common approach to modeling variability in these chemicals has been to assemble soil carbon and nutrient data from many sites and use traditional multiple regression techniques to determine which independent variables are significant predictors of nutrient and carbon pools (Burke et al. 1989). While this approach provides estimates of variability in regional soil carbon storage, it has several weaknesses. Large scale studies simplify the region by restricting analyses to a few landscape components (e.g. just natural or agricultural portions) and these components are generally described by separate models. By avoiding urban ecosystems, these models also simplify the potential drivers of soil chemical variability by excluding socioeconomic measures from the list of independent variables. Oleson et al. (2006) have introduced a spatial Bayesian approach to explain the variability arising from different land use patterns, but they have treated the data across land use types as independent. This approach fails to attribute spatial association in the boundary regions of the land use specifications.

One of our goals here is to improve upon these modeling approaches by increasing the number of land use types being represented in the model, and by incorporating the local stationarity structure in the model. Soil C and N are significantly impacted by human management practices. Specifically, fertilization, irrigation, and plant manipulations (crops, lawn plantings, etc.) on urban and agricultural lands lead to large changes in soil C and N (Kaye et al. 2005, 2008). These management practices may lead to spatial structures that are distinct among land use types. In other words, land use patterns may impart local stationarity to soil C and N storage. Since management practices that alter soil C and N are strongly correlated with mapped land use categories, satellite-derived maps of land use could be coupled with models of local stationarity to greatly improve our predictions of regional scale soil C and N storage. In our prior work describing the soil C and N from the mixed land-use area around Phoenix, AZ (Kaye et al. 2008; Majumdar et al. 2008), we observed large differences in soil C and N among the satellite-mapped land use types, but we determined the spatial structure globally. Thus, the major advancement here is to explicitly model the local stationarity in this data set.

Another shortcoming of traditional regression approaches is that multiple soil processes are not modeled simultaneously, even when we know they are associated. For example, in trying to understand how C and N vary across a landscape separate models are constructed for each element, despite the fact that C:N ratios are known to be quite predictable. Most of the N in soil (> 90%) is covalently bound to organic (i.e. C containing) molecules. Thus, as C storage in the soil increases the N storage also increases because these covalent bonds stabilize the N in soil (Kaye et al. 2002). It would be advantageous, and more realistic, to include C and N (and possibly other elements) in the same model, so that variability in one soil pool could be used to improve prediction of other soil pools of interest.

Keeping these in perspective, in this section we handle the problem of modeling and analyzing the spatial distribution of C and N in soil using a spatial regression approach in which the residual variation is modeled through the generalized convolution model

studied in earlier section. Effectively, we apply a Bayesian solution to the problem of modeling the spatial distribution of multi-element soil chemistry across a landscape of heterogeneous land use.

4.1 The data set

The research was conducted within and around the Phoenix Metropolitan area of 3.5 million people (US Census Bureau, 2000). The study region was a roughly rectangular area of 6400 km² that includes the city and surrounding agricultural lands and desert. A randomized, tessellation-stratified design was achieved by superimposing a 4 km × 4 km grid on the study area, giving 462 potential sampling units. It was expected that landscape heterogeneity would be greater in the urban core, so a random sample point was assigned within every square inside the urban core and in every third square outside that area, giving a total sample size of 206.

4.2 Modeling

Various modeling schemes have been developed and applied to explain soil-nutrient and carbon concentrations. These include deterministic and stochastic models, phenomenological and mechanistic models, spatial and non-spatial models. In the context of explaining soil nutrient characteristics, the data layers are inherently spatial and highly correlated. The data arise from different land use types within which the observed process can be assumed to be fairly homogeneous. Since some land use categories mentioned in Table 10 (the land use map) do not have enough points needed for making valid predictions, we clumped some of the land use categories together leaving only those that had a significant representation in terms of the proportion of areal coverage. Thus, the main land use categories used in our analysis are: 1) urban 2) open desert 3) agriculture and (remaining groups put together) 4) a mixture of multiple land-use types. For these land use zones, the kernel centers are evaluated using empirical estimates, namely, the averages of spatial coordinates (i.e., the centroids) of the available data points. We use 50% of the data to build the model, and the other 50% to make predictions. Thus the application of the generalized convolution model for the covariance, as discussed in Section 2, for modeling the aforementioned data seems reasonable. Some spatial features can be described through the mean structure of the model (or in the case of categorical or count data, the mean on a transformed scale) and the mean of the spatial distribution of log(Organic C) and log(Total N) shall be modeled using the spatial regressor variables.

We model the joint distribution of the natural logarithms of Organic Carbon and Total Nitrogen concentrations (gm/m²) (a log transform of the bivariate data revealed that a Gaussian model of the log transformed data was viable and that the residuals of the regression analysis were close to zero). Let $Y_1(s_j)$, $Y_2(s_j)$ denote the corresponding concentrations in the j -th spatial location s_j . We also observe at each plot elevation

Land cover map of CAPLTER

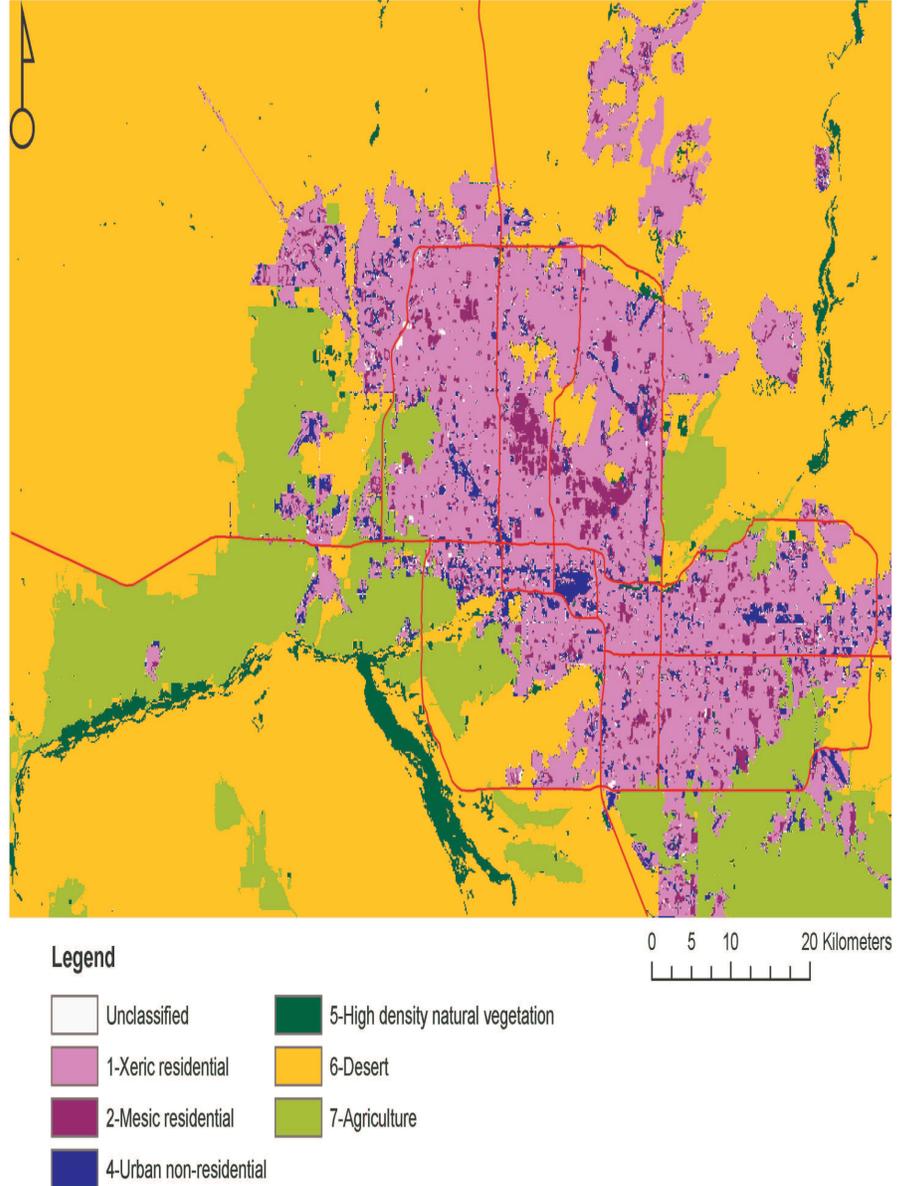


Table 10: Land use map of the Central Phoenix Arizona Long Term Ecological Project

(meters) and annual per capita income. So the model is

$$\mathbf{Y}(s_j) = \mathbf{X}_j\boldsymbol{\beta} + \boldsymbol{\epsilon}_j \quad (9)$$

where $\mathbf{Y}(s_j) = (Y_1(s_j), Y_2(s_j))^T$, \mathbf{X}_j is design matrix of order 2×6 . This design matrix is block diagonal in our case, each block containing three columns of regressor vectors. In this case these three regressors are the intercept term, percent of area in the soil that is impervious, (for instance, covered by concrete), and slope (the angle of the ground surface from the vertical in degrees from 0 to 90). These regressors were all significant in conventional regression analysis, and thus incorporated into the model.

$\boldsymbol{\beta}$ is a 6×1 vector, the first three corresponding to the regression coefficients for $\log(\text{Organic } C)$ and the last three corresponding to the regression coefficients for $\log(\text{Total } N)$. Finally the error component is $\boldsymbol{\epsilon}_j = (\epsilon_1(s_j), \epsilon_2(s_j))^T$. We assume that the latter has a cross covariance function given by (5) with $L =$ number of land use zones used in the model (3 or 4), $n = 98$, $N = 2$, $\sigma_{jil} = 1$ $c_{jl} = c$, $\alpha_1 = \alpha_2 = \alpha$, $\theta_{jl} = \theta$ and for identifiability purposes, as $\alpha_1 = \alpha_2 = \alpha$ so β is set to 0. So the covariance function involves a 6-parameter vector $(c, \theta, \sigma_{21}, \tau, \nu_{12}, \alpha)$. c is assigned an inverse Gamma prior with infinite variance, α , τ and θ are assigned Gamma priors with mean 1 and variance 10, ν_{12} is assigned the Uniform(0, 1) prior as the association between the two variables is known to be positive (Kaye et al. 2002). Finally the parameter σ_{21} is assigned a Normal(0, 10) prior. The regression coefficient vector $\boldsymbol{\beta}$ is assigned a Multivariate Normal prior with zero mean and large variance. We do not use any hyperpriors. As explained above, we use the centroids of the land use zones, based on the available data, to determine the t_l 's, i.e., the kernel centers.

Further, to compare the prediction performance of our model with that of a known bivariate stationary spatial model, we used the coregionalization model (stationary) of Wackernagel (2003) as implemented by the `spBayes` package (Carlin et al. 2007) in R, which uses one extra parameter than the model of generalized convolution, and compared predictive distributions of 95 holdout points (about 50% of the data available). The `spBayes` package uses priors with large variances (infinite variance for scale or variance parameters) for all but one of the parameters used in the model, and that is also the case in our model. One parameter, namely the decay parameter, was assigned a mean of 0.18 and variance of 0.54 in the model implemented by `spBayes`. The generalized convolution model, on the other hand, assigned a prior to the parameter α with mean 0.1 and variance 1. No hyper-prior was used in either model.

We then carry out model comparison using prediction coverage of the 95 hold out points as indicated in Table 11. Tables 12 and 13 show respectively the prediction bias and prediction standard deviations corresponding to each generalized convolution approach, with several choices of kernel centers. Cross-validation and model comparison were carried out using the performance of (i) predictive coverage of 95% predictive intervals of the 95 hold out points, (ii) prediction bias and (iii) prediction standard deviation. By cross-validation we mean the following. We are using 50% of the original sample (randomly selected from the data) for model fitting. The other 50% of the data are being used as hold-out points for checking (i) predictive coverage, (ii) prediction

Model	Prediction coverage of 95% pred. int.s log(Org C)	Prediction coverage of 95% pred. int.s log(Total N)
Correg.	47%	49%
(1, 2, 3)	80%	81%
(1, 2, 4)	81%	80%
(1, 3, 4)	82%	82%
(2, 3, 4)	81%	80%
(1, 2, 3, 4)	80%	80%
random centers (3)	83%	81%

Table 11: Percent of prediction coverage of 95 holdout points for the first (left) and second (right) spatial process

bias and prediction standard deviation. So, this procedure is similar in spirit to the v -fold cross validation procedure commonly used for model validation (cf. Breiman et al. (1984)). Here model (i, j, k) will denote a generalized convolution model with three kernels, with centers at land use zones i, j, k , respectively. Model (1, 2, 3, 4) will denote the “full” model, i.e. the model using all the four land use zones to determine the kernel centers (four). We employ another model (“random”) with random kernel centers, in which case we have 6 extra parameters for the coordinates of these centers, which are assigned independent Gaussian priors with mean set to the kernel centers of the “best model” (best in terms of prediction coverage) and variances equal to 0.5. Altering the variance does not improve the results we get. Finally, for model comparison, we have also used the stationary model of coregionalization. Table 11 shows that in terms of predictive coverage the choice of a generalized convolution approach which uses one less parameter than the stationary model of coregionalization does much better, even when only three kernel centers are used. The model (1, 3, 4) does a little better among the rest, but in terms of bias and standard deviations, all the models using the generalized convolution approach perform similarly. Overall the predictions are negatively biased as seen from Table 12. Using random kernel centers improves this situation a little, but does not improve the prediction coverage.

Scientific/practical merit of the findings can be summarized as follows. This study shows: (i) applicability to a large set of multivariate or univariate spatial data sets when the stationarity assumption is violated, (ii) effectiveness of a cross-validation technique for choosing the optimal model with minimal number of centers of local stationarity when the focus is to choose (a) high posterior predictive coverage (b) low posterior predictive bias and (c) low posterior predictive variance, (iii) a practical way to choose the kernel centers or centers of local stationarity by obtaining centroids of the “homogeneous” areas (such as the land uses) under consideration, (iv) when local stationarity is assumed, but the centers are not specified or known, then a choice of centers that span the spatial grid can be chosen, and the minimal grid which yields the best prediction inference over a random set of hold-out points can be selected.

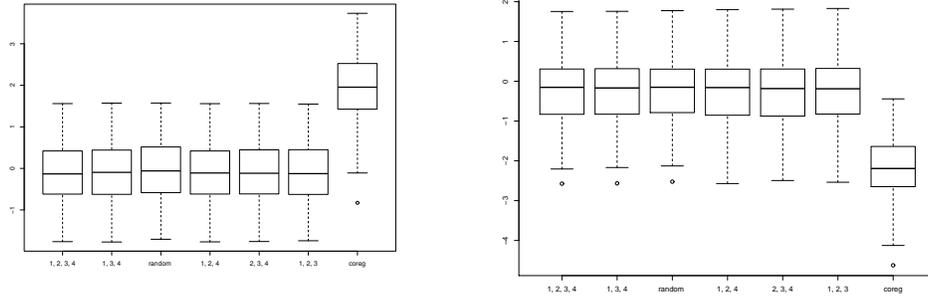


Table 12: Prediction bias of 95 holdout points for log(Org C) (left) and log(Total N) (right)

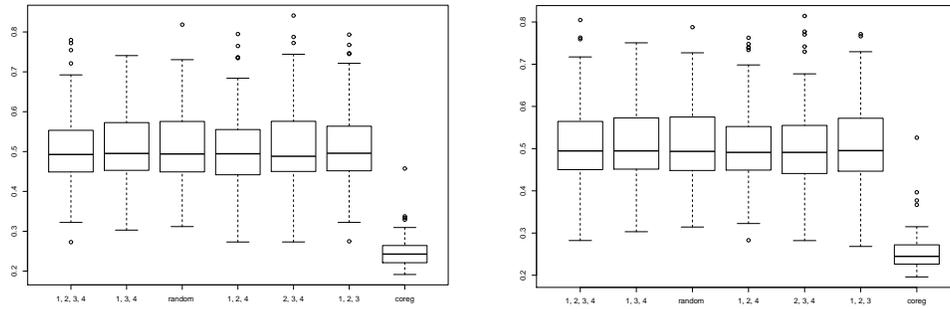


Table 13: Prediction standard deviation of 95 holdout points for log(Org C) (left) and log(Total N) (right)

5 Discussion

In this section we summarize some practical rules for choosing the kernel centers. When the area of spatial homogeneity are specified (such as census tracts, land use types etc) and when a large number of data points span these areas, then the centroids of these areas arising from the average of the x and y spatial locations can be chosen as the centers of local stationarity. Suppose K is the maximum number of centers. One can then examine among several submodels by comparing the full model – with all K centers with models that have less than or equal to $(K - 1)$ centers. So theoretically, one can consider $2^K - 1$ such models, but in practice a greedy procedure may be employed instead of running MCMC for all of these choices. By relaxing the definition a bit, we may declare a model with k kernel centers, where $1 < k < K$, to be optimal if this is the smallest model such that all the models with $(k - 1)$ centers, and all models with $m > k$ centers are less favorable according to some specific criteria. Our recommendation for the criteria of choosing the “optimal” model is to focus on predictive inference based on a large set of hold-out points. We can choose either the average posterior prediction coverage error, or posterior prediction mean squared error. The theoretical and practical performance of such procedures need to be examined.

In spite of the versatility of the proposed procedure, there are a few methodological limitations. Firstly, in spatial statistics, there is a potential lack of identifiability of the effects while estimating both the mean function and covariance kernel from the observed data. Indeed, from observed data it is difficult to ascribe the true source of heterogeneity in the process, that is, whether it is from the mean (representing the fixed effect) or from the covariance (representing the stochastic effect). In our modeling scheme, prior knowledge of the spatial field (in our application, the land use pattern) mitigates the problem by allowing us to choose the kernel centers with some degree of assurance. Nevertheless, the estimation of the bandwidth and directionality of the kernel can have an offsetting effect with the estimation of the mean, including the covariate effects. Secondly, there are a few settings in which the proposed modeling scheme may be inadequate for capturing the inhomogeneity in a nonstationary spatial process. One such scenario is when the zones of spatial homogeneities are either too fragmented, or occur along some thin or fibrous structures which cannot be represented very well by a limited number of kernels. Another potential difficulty with this procedure is that, if the number of kernels is over-specified and/or the kernel bandwidths are badly mis-specified, it may lead to a near-singularity of the covariance kernel of the process under the model. This will have negative implications in terms of model fitting as well as prediction. However, this issue can be addressed effectively by adopting an appropriate model selection procedure.

References

- Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Wadsworth, California. 514
- Burke, I. C., Yonker, C. M., Parton, W. J., Cole, C. V., Flach, K., and Schimel, D. S.

- (1989). “Texture, climate, and cultivation effects on soil organic matter content in U.S. grassland soils.” *Soil Science Society of America Journal*, 53: 800–805. 510
- Carlin, B. P., Banerjee, S., and Finley, A. O. (2007). “spBayes: An R package for univariate and multivariate hierarchical point-referenced spatial models.” *Journal of Statistical Software*, 19: 1–24. 513
- Christensen, W. F. and Amemiya, Y. (2002). “Latent variable analysis of multivariate spatial data.” *Journal of the American Statistical Association*, 97: 302–317. 494
- Fuentes, M. (2002). “Spectral methods for nonstationary spatial processes.” *Biometrika*, 89: 197–210. 494
- Fuentes, M., Chen, L., Davis, J., and Lackmann, G. (2005). “Modeling and predicting complex space-time structures and patterns of coastal wind fields.” *Environmetrics*, 16: 449–464. 494
- Fuentes, M. and Smith, R. (2001). “A new class of nonstationary models.” Technical report, North Carolina State University, Institute of Statistics. 494
- Gelfand, A. E., Schmidt, A. M., Banerjee, S., and Sirmans, C. (2004). “Nonstationary multivariate process modelling through spatially varying coregionalization.” *Test*, 13: 263–312. 494
- Higdon, D. M. (1997). “A process-convolution approach for spatial modeling.” *Computer Science and Statistics: Proceedings of the 29th Symposium Interface*. 494
- (2002). “Space and space-time modeling using process convolutions.” *Quantitative Methods for Current Environmental Issues*, 37–54. 494
- Higdon, D. M., Swall, J., and Kern., J. (1999). “Non-stationary spatial modeling.” *Bayesian Statistics*, 6: 761–768. 494
- Jobbagy, E. and Jackson, R. B. (2000). “The vertical distribution of soil organic carbon and its relation to climate and vegetation.” *Journal of the American Statistical Association*, 10: 423–436. 504
- Kaye, J. P., Barrett, J. E., , and Burke, I. C. (2002). “Stable carbon and nitrogen pools in grassland soils of variable texture and carbon content.” *Ecosystems*, 5: 461–471. 510, 513
- Kaye, J. P., Majumdar, A., Gries, C., Hope, D., Grimm, N., Zhu, W., and Baker, L. (2008). “The spatial distribution of C, N, and P in a region of mixed urban, desert, and agricultural land use.” *Ecological Applications*, 18: 132–145. 510
- Kaye, J. P., McCulley, R., and Burke, I. C. (2005). “Carbon fluxes, nitrogen cycling and soil microorganisms in adjacent urban, native and agricultural ecosystems.” *Global Change Biology*, 11: 575–587. 510

- Majumdar, A. and Gelfand, A. E. (2007). “Multivariate spatial modeling for geostatistical data using convolved covariance functions.” *Mathematical Geology*, 39: 225–245. 494
- Majumdar, A., Kaye, J. P., Gries, C., Hope, D., Burdick, R., and Grimm, N. (2008). “Hierarchical spatial modeling of multivariate soil nutrient concentrations in heterogeneous land-use patches of the Phoenix metropolitan area.” *Communications in Statistics : Simulation and Computation*, 37: 434–453. 510
- Majumdar, A., Paul, D., and Bautista, D. (2010). “A generalized convolution approach for modeling nonstationary multivariate spatial processes.” *Statistica Sinica*, 20: 675–695. 493, 494, 495, 496, 497
- Nychka, D., Wikle, C., and Royle, A. (2002). “Multiresolution models for nonstationary spatial covariance functions.” *Statistical Modeling*, 2: 299–314. 494
- Oleson, J., Hope, D., Gries, C., and Kaye, J. P. (2006). “A Bayesian approach to estimating regression coefficients for soil properties in land-use patches with varying degrees of spatial variation.” *Environmetrics*, 17: 517–525. 510
- Paciorek, C. J. and Schervish, M. (2006). “Spatial modeling using a new class of nonstationary covariance function.” *Environmetrics*, 17: 483–506. 494
- Vitousek, P. M., Aber, J. D., Howarth, R. W., Likens, G. E., Matson, P. A., Schindler, D. W., Schlesinger, W. H., and Tilman, D. G. (1997). “Latent variable analysis of multivariate spatial data.” *Ecological Applications*, 7: 737–750. 510
- Wackernagel, H. (2003). *Multivariate Geostatistics : An Introduction with Applications*. Springer, Berlin, 3 edition. 493

Acknowledgments

The authors express thanks to Alexander Buyantuyev for providing the landuse map. They are also thankful to the reviewers and the associate editor for their valuable comments. Majumdar’s research was supported by grants from NSF. The data and related research material is based upon work supported by the National Science Foundation under Grant No. DEB-0423704, Central Arizona - Phoenix Long-Term Ecological Research (CAP LTER). Paul’s research was partially supported by the NSF grant DMS-0806128.