

Hierarchical Factor Models for Large Spatially Misaligned Data: A Low-Rank Predictive Process Approach

Qian Ren* and Sudipto Banerjee**

Division of Biostatistics University of Minnesota, U.S.A.

*email: renxx014@umn.edu

**email: sudiptob@biostat.umn.edu

SUMMARY. This article deals with jointly modeling a large number of geographically referenced outcomes observed over a very large number of locations. We seek to capture associations among the variables as well as the strength of spatial association for each variable. In addition, we reckon with the common setting where not all the variables have been observed over all locations, which leads to *spatial misalignment*. Dimension reduction is needed in two aspects: (i) the length of the vector of outcomes, and (ii) the very large number of spatial locations. Latent variable (factor) models are usually used to address the former, although low-rank spatial processes offer a rich and flexible modeling option for dealing with a large number of locations. We merge these two ideas to propose a class of hierarchical low-rank spatial factor models. Our framework pursues stochastic selection of the latent factors without resorting to complex computational strategies (such as reversible jump algorithms) by utilizing certain identifiability characterizations for the spatial factor model. A Markov chain Monte Carlo algorithm is developed for estimation that also deals with the spatial misalignment problem. We recover the full posterior distribution of the missing values (along with model parameters) in a Bayesian predictive framework. Various additional modeling and implementation issues are discussed as well. We illustrate our methodology with simulation experiments and an environmental data set involving air pollutants in California.

KEY WORDS: Bayesian inference; Factor analysis; Gaussian predictive process; Linear model of coregionalization; Low-rank spatial modeling; Multivariate spatial processes; Spatial misalignment.

1. Introduction

With enhanced capabilities in collecting, storing, and accessing geographically referenced data sets, spatial analysts today frequently encounter data comprising a large number of variables across a very large number of locations. In several instances, inference focuses upon three major aspects: (i) estimate associations among the variables, (ii) estimate the strength of spatial association for each variable, and (iii) predict the outcomes at arbitrary locations.

Modeling multiple geographically referenced outcomes proceeds from two different premises. One approach (Royle and Berliner, 1999; Gelfand et al., 2004) considers a conditional regression-like approach where the marginal distribution of the first outcome is specified, followed by the conditional distribution of the second outcome given the first, and so on. This approach is more suitable when there is a natural “ordering” of the outcomes suggesting the sequence for constructing the conditional distributions.

Settings that lack such information prefer joint modeling for the set of outcomes to avoid the explosion in models emerging from alternate ordering schemes. Conditional inference can subsequently proceed from the joint distribution by conditioning on the relevant variables. The challenge in the joint modeling approach is to specify valid multivariate spatial processes using matrix-valued, cross-covariance functions

([e.g., Ch. 7]; Banerjee, Carlin, and Gelfand 2004). Gelfand et al. (2004) offer a detailed comparison of both approaches for multivariate spatial data. Here we focus upon the joint modeling approach.

The linear model of coregionalization (LMC) proposed by Matheron (1982) is among the most general models for multivariate spatial data analysis. Here, the spatial behavior of the outcomes is assumed to arise from a linear combination of the independent latent processes operating at different spatial scales (Chilés and Delfiner, 1999). The idea resembles latent factor analysis (FA) models for multivariate data analysis (e.g., Anderson, 2003) except that in the LMC the number of latent processes is usually taken to be the same as the number of outcomes. Then, an $m \times m$ covariance matrix has to be estimated for each spatial scale (see, e.g., Lark and Papritz, 2003; Castrignanó et al., 2005; Zhang, 2007; Finley et al., 2008), where m is the number of outcomes. When m is large (e.g., $m \geq 5$ and 300 spatial locations), obtaining such estimates is expensive. Schmidt and Gelfand (2003) and Gelfand et al. (2004) associate only an $m \times m$ triangular matrix with the latent processes. However, high dimensional outcomes are still computationally prohibitive for these models.

When the number of independent latent processes in an LMC is taken to be fewer than the number of outcomes (e.g., Zhang, 2007), we obtain a spatial factor model. Wang and

Wall (2003) studied multivariate indicators of cancer risk in Minnesota using one spatial factor. Liu, Wall, and Hodges (2005) deployed multiple factors with each underlying factor explaining its own unique set of observed/measured variables. Christensen and Amemiya (2002) developed semiparametric latent variable (shift-factor) models for rectangular grids. Hogan and Tchernis (2004) fitted a one-factor spatial model and compared the results using different forms of spatial dependence through the single factor. Minozzo and Frutini (2004) applied log-linear spatial FA to geo-referenced frequency counts adopting the classical proportional covariance model to the latent factors.

Our current work is similar to the aforementioned articles in its use of independent latent spatial processes as underlying factors. However, we propose three methodological innovations within this framework. First, we use a multivariate low-rank spatial process to achieve dimension reduction over space. Critically, we need to work with irregularly spaced locations that do not necessarily lie on a grid, nor can they be easily projected onto the same. Although there are several choices here, we deploy the multivariate Gaussian predictive process (Banerjee et al., 2008, 2010). The method is closely related to kernel-convolutions, splines and low-rank kriging, (see, e.g., Wikle and Cressie, 1999; Kamman and Wand, 2003; Ver Hoef et al., 2004; Paciorek, 2007; Cressie and Johannesson, 2008), all of which attempt to facilitate computation through lower dimensional process representations. The predictive process can be applied to any spatial correlation function and maintain the richness of desired hierarchical spatial modeling specifications using a set of locations (or knots).

Second, we do not fix the number of factors but model it stochastically. We do so differently from some existing approaches. For example, Lopes and West (2004) addressed this problem by constructing proposals using the results of a preliminary Markov chain Monte Carlo (MCMC) run under each model. Such an approach has high computational demands, becoming infeasible as the sample size and potential number of factors increase. Dunson (2006) introduced a model averaging method for FA, but the construction of factor selection is inconvenient for spatial models. Chen and Dunson (2003) proposed a Bayesian random effect selection method which is similar to what we propose but, instead of selecting random variables, we choose the underlying latent processes that capture spatial dependence. We avoid complex computational strategies such as reversible jump algorithms and build our adaptive models by utilizing some key identifiability results, hitherto largely unaddressed, to construct hierarchical models.

Third, we reckon with spatial misalignment in the context of spatial FA. Misalignment occurs frequently in spatial data when not all variables have been observed across all locations. Put differently, the sets of observed locations for the different outcomes are not identical (either because they are missing or have been collected by different monitoring sets). Assuming that all covariates are available at a location, we want to estimate the functional relationship between the covariates and the outcomes at that location—even if all the outcomes have not been observed there. We also seek to predict the outcomes at any arbitrary location in the domain, thereby estimating the response surfaces for each outcome.

Our motivating application pertains to ambient air quality assessment in California. The deleterious impact of air pollution upon health and quality of life is widely recognized as a major environmental issue (e.g., Dominici et al., 2006). Spatial interpolation of air pollutants plays a crucial role in assessing and monitoring ambient air quality and modeling multiple pollutants can capture associations within and across different locations, which can enhance predictive performance. Our data set includes five commonly encountered pollutants observed over 300 monitoring stations across California and is spatially misaligned in the aforementioned manner. Estimating fully specified joint models for such data will be exorbitant, which is why multivariate spatial modeling has rarely been undertaken in such frameworks.

The remainder of this article is organized as follows. Section 2 describes the features of the LMC and discusses the model construction, identifiability issues, stochastic selection, and prior specification. Section 3 outlines the proposed class of low-rank adaptive spatial factor models, how we handle misaligned data, and carry out inference. Section 4 illustrates the analysis for two simulated data sets and one air quality monitoring data. Finally, Section 5 concludes the paper with a summary and an eye toward future work.

2. Model Construction

2.1 LMC Model Structure and Specification

LMC consists of decomposing the set of original second-order random stationary outcomes into a set of reciprocally orthogonal regionalized factors. Suppose, for a study region $D \subseteq \mathbb{R}^d$, an $m \times 1$ process $\mathbf{Z}(\mathbf{s}) = (Z_1(\mathbf{s}), \dots, Z_m(\mathbf{s}))'$ is a second-order stationary process. Then, for all $\mathbf{s}, \mathbf{h} \in \mathbb{R}^d$ and $i, j = 1, \dots, m$, we have $E[Z_i(\mathbf{s})] = \mu_i$ and $\text{cov}\{Z_i(\mathbf{s}), Z_j(\mathbf{s} + \mathbf{h})\} = C_{ij}(\mathbf{h})$. The matrix-valued function $\mathbf{C}(\mathbf{h}) = \{C_{ij}(\mathbf{h})\}$ is called the multivariate cross-covariance. Generally for $i \neq j$, a change in the order of the variables or a change in the sign of the separation vector \mathbf{h} changes the values of the cross-covariances. If both the sequence and the sign are changed, we would have the same value $C_{ij}(\mathbf{h}) = C_{ji}(-\mathbf{h})$, which implies $\mathbf{C}(\mathbf{h}) = \mathbf{C}(-\mathbf{h})'$. $\mathbf{C}(\mathbf{h})$ must also be a positive definite function. That is, for any finite set of spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_n \in D$ and any vectors $\mathbf{a}_i \in \mathbb{R}^m, i = 1, \dots, n$, $\text{var}\{\sum_{i=1}^n \mathbf{a}_i' \mathbf{Z}(\mathbf{s}_i)\} = \sum_{i,j=1}^n \mathbf{a}_i' \mathbf{C}(\mathbf{s}_i - \mathbf{s}_j) \mathbf{a}_j \geq 0$. Specifying a valid cross-covariance function is less straightforward because of this constraint, but several spectral and constructive approaches (Ver Hoef and Barry, 1998; Chilés and Delfiner, 1999) have been proposed and used in multivariate spatial analysis.

A straightforward cross-covariance is the so-called intrinsic specification, $\mathbf{C}(\mathbf{h}) = \rho(\mathbf{h})\mathbf{T}$, where \mathbf{T} is an $m \times m$ positive definite matrix and $\rho(\mathbf{h})$ is a univariate correlation function. The limitation here is that each $Z_i(\mathbf{s})$ has the same spatial parameters, so each outcome would have the same strength in spatial association over the domain. An extension (Wackernagel, 2003) specifies $\mathbf{C}(\mathbf{h}) = \sum_{k=1}^r \rho_k(\mathbf{h}; \phi_k) \mathbf{T}_k$, where for each k , \mathbf{T}_k is a rank-one positive semidefinite matrix and $\rho_k(\mathbf{h}; \phi_k)$ is a correlation function that depends upon additional parameters ϕ_k . Here, r is the total number of different spatial correlation functions in the multivariate cross-covariance.

The spectral decomposition yields $\mathbf{T}_k = \xi_k \mathbf{u}_k \mathbf{u}_k'$, where \mathbf{u}_k is the normalized (i.e., $\|\mathbf{u}_k\| = 1$) eigenvector of \mathbf{T}_k

corresponding to the *only* positive eigenvalue ξ_k of \mathbf{T}_k . Because \mathbf{T}_k is symmetric with rank one, all its other eigenvalues are zero. This implies that there is an $m \times 1$ vector, $\boldsymbol{\lambda}_k = \sqrt{\xi_k} \mathbf{u}_k$, which is the square root of \mathbf{T}_k such that $\boldsymbol{\lambda}_k \boldsymbol{\lambda}_k' = \mathbf{T}_k$. Although $\|\mathbf{u}_k\| = 1$, it still has two possible directions. To build a one-to-one transformation between \mathbf{T}_k and $\boldsymbol{\lambda}_k$, we need to impose some further constraints on $\boldsymbol{\lambda}_k$ by, for instance, restricting the first element of $\boldsymbol{\lambda}_k$ to be positive.

Let $w_k(\mathbf{s})$, $k = 1, \dots, r$, be independently distributed univariate Gaussian processes, each with unit variance and a parametric correlation function. We write $w_k(\mathbf{s}) \sim GP(0, \rho_k(\cdot; \phi_k))$ with $\text{var}\{w_k(\mathbf{s})\} = 1$, $\text{cov}\{w_k(\mathbf{s}), w_k(\mathbf{t})\} = \rho_k(\mathbf{s}, \mathbf{t}; \phi_k)$, and $\text{cov}\{w_k(\mathbf{s}), w_l(\mathbf{t})\} = 0$ whenever $k \neq l$ for all \mathbf{s} and \mathbf{t} (even when $\mathbf{s} = \mathbf{t}$). Here $\rho_k(\cdot; \phi_k)$ is a correlation function associated with $w_k(\mathbf{s})$, and ϕ_k includes the spatial decay and smoothness parameters. We can easily see that the multivariate cross-covariance function for the process $w_k(\mathbf{s})\boldsymbol{\lambda}_k$ is $\rho_k(\mathbf{h}; \phi_k)\mathbf{T}_k$ and hence, for $\sum_{k=1}^r w_k(\mathbf{s})\boldsymbol{\lambda}_k$, the function is $\sum_{k=1}^r \rho_k(\mathbf{h}; \phi_k)\mathbf{T}_k$.

We often assume that the mean of the outcomes arises linearly in the predictors so that the mean of the j th outcome $Y_j(\mathbf{s})$ is modeled as $\mathbf{x}_j(\mathbf{s})'\boldsymbol{\beta}_j$, where $\mathbf{x}_{ij}(\mathbf{s})$ is a $p_j \times 1$ vector of predictors, assumed to be known at location \mathbf{s} , and $\boldsymbol{\beta}_j$ is the corresponding $p_j \times 1$ vector of slopes. Let $\mathbf{Y}(\mathbf{s})$ be the $m \times 1$ vector of outcomes with j th element $Y_j(\mathbf{s})$ and let $\mathbf{X}(\mathbf{s})'$ be an $m \times p$ block diagonal matrix, where $p = \sum_{j=1}^m p_j$ and the j th diagonal block is given by $\mathbf{x}_j(\mathbf{s})'$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \dots, \boldsymbol{\beta}_m')'$ is a $p \times 1$ vector of slopes. Then, the spatial factor model is

$$\begin{aligned} \mathbf{Y}(\mathbf{s}) &= \mathbf{X}(\mathbf{s})'\boldsymbol{\beta} + \sum_{k=1}^r w_k(\mathbf{s})\boldsymbol{\lambda}_k + \boldsymbol{\epsilon}(\mathbf{s}) \\ &= \mathbf{X}(\mathbf{s})'\boldsymbol{\beta} + \boldsymbol{\Lambda}\mathbf{w}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}), \end{aligned} \quad (1)$$

where $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_r)$ is a $m \times r$ matrix with k th column $\boldsymbol{\lambda}_k$, $\mathbf{w}(\mathbf{s}) = (w_1(\mathbf{s}), \dots, w_r(\mathbf{s}))'$, and $\boldsymbol{\epsilon}(\mathbf{s})$ is an $m \times 1$ vector of measurement errors distributed as $N(\mathbf{0}, \boldsymbol{\Psi})$. The measurement error variance $\boldsymbol{\Psi}$ can be any $m \times m$ positive definite matrix but is usually assumed diagonal with elements ψ_j^2 for $j = 1, \dots, m$ along the diagonal.

2.2 Identifiability in Spatial Factor Model

Model (1) is similar to FA models with the factor-loading matrix $\boldsymbol{\Lambda}$ and latent factor $\mathbf{w}(\mathbf{s})$, except that $\mathbf{w}(\mathbf{s})$ is now a multivariate stochastic process positing spatial dependence. As is well known (see, e.g., Anderson, 2003), orthogonal FA models must be further constrained to ensure identifiability. A widely used approach is to fix certain elements of $\boldsymbol{\Lambda}$ to constant values, usually to zeroes, such as restricting $\boldsymbol{\Lambda}$ to be an upper or lower triangular matrix with strictly positive diagonal elements (Lopes and West, 1999).

In LMCs, (see, e.g., Schmidt and Gelfand, 2003; Gelfand et al., 2004; Finley et al., 2008), a lower triangular $\boldsymbol{\Lambda}$ with positive diagonal elements identifies the covariances among the outcomes within a location because $\mathbf{C}(\mathbf{0}) = \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$. But what we really seek to model is $\mathbf{C}(\mathbf{h})$. Finding proper identifiability constraints on $\boldsymbol{\Lambda}$ is equivalent to finding transformations that retain the statistical properties of the model.

Let \mathbf{P} be an $r \times r$ orthogonal matrix, so that $\mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P}' = \mathbf{I}_r$. The random effect term in (1) can be written as $\boldsymbol{\Lambda}\mathbf{w}(\mathbf{s}) =$

$\boldsymbol{\Lambda}\mathbf{P}\mathbf{P}'\mathbf{w}(\mathbf{s}) = \bar{\boldsymbol{\Lambda}}\bar{\mathbf{w}}(\mathbf{s})$, where $\bar{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}\mathbf{P}$ and $\bar{\mathbf{w}}(\mathbf{s}) = \mathbf{P}'\mathbf{w}(\mathbf{s})$. For nonspatial or traditional FA, the elements of $\mathbf{w}(\mathbf{s})$ are uncorrelated and the model is invariant to any orthogonal transformation because $\mathbf{w}(\mathbf{s})$ and $\bar{\mathbf{w}}(\mathbf{s})$ have identical cross-covariances and $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \bar{\boldsymbol{\Lambda}}\bar{\boldsymbol{\Lambda}}'$. One can, therefore, obtain an infinite number of equivalent matrices of factor loadings by simply applying orthogonal transformations.

Matters are subtly different when $\mathbf{w}(\mathbf{s})$ is a spatial process. The cross-covariance matrix $\text{cov}\{\mathbf{w}(\mathbf{s}), \mathbf{w}(\mathbf{t})\} = \boldsymbol{\Gamma}(\mathbf{s}, \mathbf{t}; \phi)$, is now diagonal (and not an identity matrix as in traditional FA) with the k th diagonal element $\rho_k(\mathbf{s}, \mathbf{t}; \phi_k)$. Orthogonal transformations can, therefore, alter the distribution of $\bar{\mathbf{w}}(\mathbf{s})$. To be precise, now $\text{cov}\{\bar{\mathbf{w}}(\mathbf{s}), \bar{\mathbf{w}}(\mathbf{t})\} = \mathbf{P}'\boldsymbol{\Gamma}(\mathbf{s}, \mathbf{t}; \phi)\mathbf{P}$, which is neither necessarily diagonal nor equal to $\boldsymbol{\Gamma}(\mathbf{s}, \mathbf{t}; \phi)$. Therefore, spatial factor models are not necessarily invariant to *any* orthogonal transformation.

Our primary motivation for using a $\boldsymbol{\Lambda}$ without any pre-specified 0 element is stochastic modeling for the number of factors (see Section 2.3). Such a model would automatically identify the latent spatial processes and, hence, the corresponding columns of $\boldsymbol{\Lambda}$ that are retained in (1). Specifying $\boldsymbol{\Lambda}$ to be lower triangular is now problematic because it is unlikely that such a structure will be retained throughout the stochastic selection process. On the other hand, using a $\boldsymbol{\Lambda}$ that is identifiable but that does not restrict certain elements to be zero will avoid such awkwardness.

In fact, we argue (see Web Appendix A for details) that only two groups of orthogonal transformations lead to non-identifiability in spatial factor models. The first transformation \mathbf{P} satisfies $\mathbf{P}'\boldsymbol{\Gamma}(\mathbf{s}, \mathbf{t}; \phi)\mathbf{P} = \boldsymbol{\Gamma}(\mathbf{s}, \mathbf{t}; \phi)$. Such a \mathbf{P} must be diagonal with 1's and -1's only. It is a special *reflector* (i.e., $\mathbf{P}^2 = \mathbf{I}$) and is obtained as a product of elementary reflectors of the form $\mathbf{I} - 2\mathbf{e}_i\mathbf{e}_i'$, where \mathbf{e}_i is the i th Euclidean standard coordinate vector. The orthogonal matrix \mathbf{P} alters the sign of the columns in $\boldsymbol{\Lambda}$. To identify $\boldsymbol{\Lambda}$ (or have a one-to-one relationship between $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$), we need to specify one element in each column of $\boldsymbol{\Lambda}$ as positive or negative. Without losing generality, we could set the first row of $\boldsymbol{\Lambda}$ to be positive.

Permutation matrices constitute the second group of orthogonal transformations that lead to nonidentifiability in spatial factor models. A permutation matrix switches the elements of $\mathbf{w}(\mathbf{s})$ and the corresponding columns of $\boldsymbol{\Lambda}$ simultaneously, so the distributions of $\bar{\boldsymbol{\Lambda}}\bar{\mathbf{w}}(\mathbf{s})$ and $\boldsymbol{\Lambda}\mathbf{w}(\mathbf{s})$ are the same. To address such identifiability issues, we impose some constraints on $\rho_k(\mathbf{h}; \phi_k)$, more specifically on ϕ_k (Zhang, 2007). For simplicity, we consider the exponential correlation function $\rho_k(\mathbf{h}; \phi_k) = \exp(-\phi_k\|\mathbf{h}\|)$, which has a spatial decay parameter ϕ_k as the only unknown parameter. We require the range parameter ϕ_k , $k = 1, \dots, r$, to be ordered as $\phi_1 < \phi_2 \dots < \phi_r$ or $\phi_1 > \phi_2 \dots > \phi_r$. Simulation studies (see Sections 4.1 and 4.2) reveal that without this constraint, parameter estimation becomes problematic.

2.3 Adaptive Bayesian Factor Model

Determining the number of spatial factors (r) is challenging because of the lack of rigorous theoretical results that hint at the data's ability to inform about r . Generally, previous work Webster, Atteias, and Dubois (1994) employing the LMC assumes $r < m$ is fixed. The choice of the number of spatial processes and their respective scales is a critical point in

geostatistical models. In the applications of LMC to multivariate spatial analysis (see, e.g., Webster et al., 1994; Castrignanó et al., 2005; Buttafuoco et al., 2010), the spatial correlation functions $\rho_k(\mathbf{h}; \phi_k)$, the parameters ϕ_k , and r are obtained from empirical estimates of the auto- and cross-variograms before any modeling. This approach ignores the uncertainty in the estimates of the spatial parameters and may yield dubious inference.

We adapt earlier work by Kuo and Mallick (1998) and Chen and Dunson (2003) to propose an approach for selecting spatial processes corresponding to different spatial scales using a Bayesian hierarchical model. Indicator variables $\delta = \{\delta_1, \dots, \delta_r\}$, where each δ_k is supported at two points 1 and 0, are introduced in model (1) to yield:

$$\mathbf{Y}(\mathbf{s}) = \mathbf{X}(\mathbf{s})'\boldsymbol{\beta} + \sum_{k=1}^r \delta_k w_k(\mathbf{s}) \boldsymbol{\lambda}_k + \boldsymbol{\epsilon}(\mathbf{s}). \quad (2)$$

When $\delta_k = 1$, we include the k th spatial random process $w_k(\mathbf{s})$ and the corresponding $\boldsymbol{\lambda}_k$ in the model. Otherwise, we omit the k th spatial scale. This yields 2^r submodels. We call (2) the *adaptive Bayesian factor model*, in which the unknown parameters are estimated together with which factors will be retained.

2.4 Prior Specification

We place a multivariate normal prior on the slope parameter $\boldsymbol{\beta}$ with mean $\boldsymbol{\mu}_\beta$ and variance $\boldsymbol{\Sigma}_\beta$. Often, a flat prior $\boldsymbol{\Sigma}_\beta^{-1} = \mathbf{O}$ is used. Each diagonal element of $\boldsymbol{\Psi}$ is assigned an Inverse Gamma (IG) distribution. With $\psi_i^2 \sim \text{IG}(a, b)$, the prior for $\boldsymbol{\Psi}$ becomes $\pi(\boldsymbol{\Psi} | a, b) \propto \prod_{i=1}^m (\psi_i^2)^{-a-1} \exp\{-b/\psi_i^2\}$ with hyperparameters a and b . A customary choice is to use $a = 2$, which suggests a distribution with infinite variance and a mean of b (often gleaned from a semivariogram).

For simplicity, we assume independent priors for $\boldsymbol{\delta}$ and $\boldsymbol{\Lambda}$ (Kuo and Mallick, 1998). The indicator variables δ_k , $k = 1, \dots, r$, are taken a priori to be independent, with $p(\delta_k = 1 | \omega) = \omega$. We regard ω as unknown and assume that it has a uniform prior on $(0, 1)$. Imposing constraints on $\boldsymbol{\Lambda}$, only requires a minor modification in the derivation of the full conditional distribution. Here, we take independent priors such as $\Lambda_{jk} \sim N(0, c_0^2)$. The first row of $\boldsymbol{\Lambda}$ is restricted to be positive, that is, $\Lambda_{1k} \sim N(0, c_0^2) I(\Lambda_{1k} > 0)$ for $k = 1, \dots, r$, where $I(\cdot)$ is the indicator function. Berger and Pericchi (2001) suggest caution in the use of such priors in generic hierarchical models because the outcome of the model selection process can be quite sensitive to their vagueness. However, this seems to be less of an issue in dynamic factor models once the loading matrix is made identifiable (Lopes and West, 2004), whereupon inference and model selection were robust to the prior specifications. Similar results are obtained in random effect selection models (Chen and Dunson, 2003; Cai and Dunson, 2006). In our current context, restricting the first row of $\boldsymbol{\Lambda}$ to be positive, hence assigning a truncated normal to each element of the first row of $\boldsymbol{\Lambda}$, ensures robust model selection and related inference to c_0^2 .

One also needs to assign priors on $\boldsymbol{\phi} = \{\phi_k\}_{k=1}^r$. The prior for $\boldsymbol{\phi}$ depends upon the choice of correlation functions. Quite remarkably, the spatial process parameters are not consistently estimable and the effect of the prior does not disappear

with increasing amounts of data (Zhang, 2004). Hence, prior information becomes an even more delicate issue. Typically, we set prior distributions for the range parameters relative to the size of their domains. In this article, an exponential correlation function is used and the prior for the range parameter is specified on a support with upper and lower limits denoted as $\phi_u = -\log(0.01)/d_{\min}$ and $\phi_l = -\log(0.05)/d_{\max}$, where d_{\min} and d_{\max} are the minimum and maximum distances across all the locations (Wang and Wall, 2003). Because of identifiability issues discussed in Section 2.2, we construct a joint distribution for the ϕ_k 's to ensure ordering. In particular, we set

$$\pi(\boldsymbol{\phi}) = \pi(\phi_1)\pi(\phi_2 | \phi_1) \cdots \pi(\phi_k | \phi_{k-1}, \dots, \phi_1) \cdots \pi(\phi_r | \phi_{r-1}, \dots, \phi_1),$$

where $\pi(\phi_1)$ is a uniform density with support (ϕ_l, ϕ_u) and subsequently, for $k = 2, 3, \dots, r$,

$$\pi(\phi_k | \phi_{k-1}, \dots, \phi_1) \propto \exp\left\{-\frac{c_k}{\phi_k - \phi_{k-1}}\right\} \quad (3)$$

$$I(\phi_u > \phi_k > \phi_{k-1}).$$

Here the hyperparameters $c_k > 0$ controls the shape of the distribution and the separation of the ϕ_k 's. In all our subsequent analyses, we fix $c_k = 2k$ as a reasonable choice that delivers robust inference. For any finite domain this prior is proper (i.e., integrable). To offer some additional insight, let $\phi_k - \phi_{k-1} \in (0, hc_k)$ and $h = 4$. Numerical integration yields $p(\phi_k - \phi_{k-1} < c_k/2 | \phi_{k-1}) < 0.01$. This implies that ϕ_k is unlikely to appear in $(\phi_{k-1}, \phi_{k-1} + c_k/2)$ unless there is a strong mode from the likelihood. So (3) could efficiently separate the ϕ_k 's. On the other hand, $\frac{p(3c_k \leq \phi_k - \phi_{k-1} \leq 4c_k | \phi_{k-1})}{p(2c_k \leq \phi_k - \phi_{k-1} \leq 3c_k | \phi_{k-1})} = 1.1$, which indicates that $\pi(\phi_k | \phi_{k-1})$ becomes informative when ϕ_{k-1} and ϕ_k are away from each other.

An alternative specification is $\pi(\boldsymbol{\phi}) \propto I(\phi_l < \phi_1)I(\phi_1 < \phi_2) \cdots I(\phi_r < \phi_u)$. This is somewhat simpler than (3), but encounters problems in practical implementation. Here, the posteriors for ϕ_k and ϕ_{k+1} , although they theoretically obey $\phi_k < \phi_{k+1}$, can become very close to each other. In that case, one of the two latent processes becomes redundant, yet stochastic factor selection keeps both processes. The specification in (3) avoids this situation.

3. Predictive Process Factor Models

Although FA is a powerful tool for summarizing multivariate outcomes and conducting dimension reduction on the number of outcomes, the spatial FA is prohibitive with a large number of locations. A popular model-based approach for dimension reduction over space uses low- or fixed-rank representations for $\mathbf{w}(\mathbf{s})$. Their likelihood resembles linear-mixed models and can be estimated using standard algorithms with some minor adaptations.

Here we consider one such representation that projects the spatially associated latent factors $\mathbf{w}(\mathbf{s})$ onto a lower-dimensional subspace determined by a partial realization of the process $\mathbf{w}(\mathbf{s})$ over a manageable set of locations called “knots.” Unlike several other low-rank methods, the predictive process does not introduce additional parameters or kernels. Also, some approaches require empirical estimates of the data's covariance structure. This may be challenging here as

the variability in the “data” is assumed to be a sum of unobserved factors. Therefore, we cannot use variograms on the data to isolate empirical estimates of its spatial covariance structure from that of the individual factors.

3.1 Model Construction

A Gaussian predictive process (see, e.g., Banerjee et al., 2008; Finley et al., 2009; Banerjee et al., 2010) uses a set of fixed “knots” $\mathcal{S}^* = \{\mathbf{s}_1^*, \dots, \mathbf{s}_{n^*}^*\}$, $n^* \leq n$, which are usually fixed and may, but need not, form a subset of the observed locations $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$. The Gaussian process defined in Section 2.1 implies that $\mathbf{w}_k^* = (w_k(\mathbf{s}_1^*), \dots, w_k(\mathbf{s}_{n^*}^*))'$ follows $N_{n^*}(\mathbf{0}, \mathbf{D}_k^*(\phi_k))$, where $\mathbf{D}_k^*(\phi_k)$ is the $n^* \times n^*$ covariance matrix whose (i, j) th element is $\rho_k(\mathbf{s}_i^*, \mathbf{s}_j^*; \phi_k)$. The spatial interpolation or “kriging” function at a site \mathbf{s}_0 is given by $\tilde{w}_k(\mathbf{s}_0) = E[w_k(\mathbf{s}_0) | \mathbf{w}_k^*] = \mathbf{d}_k(\mathbf{s}_0; \phi_k)' \mathbf{D}_k^*(\phi_k)^{-1} \mathbf{w}_k^*$, where $\mathbf{d}_k(\mathbf{s}_0; \phi_k)$ is an $n^* \times 1$ vector whose i th element is $\rho_k(\mathbf{s}_0, \mathbf{s}_i^*; \phi_k)$. This defines the predictive process $\tilde{w}_k(\mathbf{s}) \sim GP(0, \tilde{\rho}_k(\cdot; \phi_k))$ derived from the parent process $w_k(\mathbf{s})$, where $\tilde{\rho}_k(\mathbf{s}_i, \mathbf{s}_j; \phi_k) = \mathbf{d}_k(\mathbf{s}_i; \phi_k)' \mathbf{D}_k^*(\phi_k)^{-1} \mathbf{d}_k(\mathbf{s}_j; \phi_k)$.

The predictive process underestimates the variance of the parent process $w_k(\mathbf{s}_0)$ at any location \mathbf{s}_0 because $\text{var}\{w_k(\mathbf{s}_0)\} - \text{var}\{\tilde{w}_k(\mathbf{s}_0)\} = \text{var}\{w_k(\mathbf{s}_0) - \tilde{w}_k(\mathbf{s}_0)\} \geq 0$. The veracity of this is an immediate consequence of the definition of $\tilde{w}_k(\mathbf{s})$ as a conditional expectation. This means that the estimated Ψ from the predictive process model roughly captures the same amount of variability as the estimated $\Psi + \mathbf{A}(\mathbf{I}_r - E[\tilde{\Gamma}(\mathbf{s}_0)])\mathbf{A}'$ from (1), where $\tilde{\Gamma}(\mathbf{s}_0)$ is an $r \times r$ diagonal matrix with k th diagonal element $\tilde{\rho}_k(\mathbf{s}_0, \mathbf{s}_0; \phi_k)$.

Defining $f_k(\mathbf{s}) | \tilde{w}_k(\mathbf{s}) \sim N(\tilde{w}_k(\mathbf{s}), \sigma_{f_k}^2(\mathbf{s}))$, where $\sigma_{f_k}^2(\mathbf{s}) = 1 - \tilde{\rho}_k(\mathbf{s}, \mathbf{s}; \phi_k)$, we easily obtain $\text{var}\{f_k(\mathbf{s})\} = \text{var}\{w_k(\mathbf{s})\} = 1$, as desired. Replace $\mathbf{w}(\mathbf{s})$ with $\mathbf{f}(\mathbf{s}) = (f_1(\mathbf{s}), \dots, f_r(\mathbf{s}))'$ in (2) and define $\mathbf{f}_k = (f_k(\mathbf{s}_1), \dots, f_k(\mathbf{s}_n))'$, $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_r)$ and $\mathbf{w}^* = (\mathbf{w}(\mathbf{s}_1^*), \dots, \mathbf{w}(\mathbf{s}_{n^*}^*))'$. This yields the posterior distribution $p(\mathbf{F}, \mathbf{w}^*, \beta, \Psi, \phi, \Lambda, \delta, \omega | \mathbf{Y})$ proportional to

$$\begin{aligned} & \pi(\omega) \times \pi(\phi) \times \prod_{k=1}^r \text{Ber}(\delta_k | \omega) \times N_m(\lambda_k | \mathbf{0}, c_0^2 \mathbf{I}_m) \\ & \times \mathbf{I}(\Lambda_{1k} > 0) \times N_p(\beta | \mu_\beta, \Sigma_\beta) \\ & \times \prod_{j=1}^m IG(\psi_j^2 | a, b) \times \prod_{k=1}^r N_n \\ & \times \left(\mathbf{f}_k | \mathbf{D}_k(\phi_k) \mathbf{D}_k^*(\phi_k)^{-1} \mathbf{w}_k^*, \Sigma_{\mathbf{f}_k}(\phi_k) \right) \\ & \times \prod_{k=1}^r N_{n^*}(\mathbf{w}_k^* | \mathbf{0}, \mathbf{D}_k^*(\phi_k)) \times \prod_{i=1}^n N_m \\ & \times \left(\mathbf{Y}(\mathbf{s}_i) | \mathbf{X}(\mathbf{s}_i)' \beta + \sum_{k=1}^r \delta_k f_k(\mathbf{s}_i) \lambda_k, \Psi \right), \end{aligned} \quad (4)$$

where $\mathbf{D}_k(\phi_k) = (\mathbf{d}_k(\mathbf{s}_1), \dots, \mathbf{d}_k(\mathbf{s}_n))'$ and $\Sigma_{\mathbf{f}_k}$ is an $n \times n$ diagonal matrix with i th diagonal element $\sigma_{f_k}^2(\mathbf{s}_i)$. $\pi(\omega)$ is uniformly distributed on $(0, 1)$ and $\pi(\phi)$ is defined in (3). Estimation of (4) proceeds using MCMC sampling (see Web Appendix B for details).

We offer some remarks on knot selection. With evenly distributed locations, knots on a uniform grid (see, e.g., Diggle and Lophaven, 2006) may suffice. With irregular locations, space-covering designs (e.g., Royle and Nychka, 1998) yield

a more representative set. In general, different knots selection methods still produce robust results in predictive process models (see, e.g., Finley et al., 2009; Banerjee et al., 2010). And the modified predictive process model $f_k(\mathbf{s})$ is even less sensitive than $\tilde{w}_k(\mathbf{s})$ to such choices. Here, we use the K-means clustering algorithm (Hartigan and Wong, 1979) to arrive at a set of knots.

3.2 Handling Missing Observations

Missing outcomes arise frequently in many environmental applications. The outcomes may be missing for a variety of unintended reasons: nonresponse, equipment failure, lack of collection, and so on. Bayesian analysis often proceeds from data augmentation (Tanner, 1993) or multiple imputation (Rubin, 1976) to handle the incomplete data problem. Both methods impute the missing values within a Gibbs step and then use the complete data set. Instead, here we condition on the latent factors, which simplifies the MCMC algorithm and reduces the computational burden. Conditional on the latent factor $\mathbf{f}(\mathbf{s})$, the outcomes are independent of each other, which yields a likelihood depending only upon observed data. The missing values can, then, be recovered from the posterior predictive distribution.

Let $\mathbf{Y}(\mathbf{s})$ be the $m \times 1$ vector of measured and unmeasured outcomes at site \mathbf{s} . Suppose the measured and unmeasured elements of $\mathbf{Y}(\mathbf{s})$ have indices $i_1, \dots, i_{d_{\mathbf{s}}}$ and $i_{d_{\mathbf{s}}+1}, \dots, i_m$, respectively, where $d_{\mathbf{s}}$ is the number of observed outcomes at \mathbf{s} . Let $\mathbf{v}_j, j = 1, \dots, m$, be an $m \times 1$ vector whose j th element is 1 and the rest are all 0. We can then construct matrices $\mathbf{R}_1(\mathbf{s})_{d_{\mathbf{s}} \times m}$ and $\mathbf{R}_2(\mathbf{s})_{(m-d_{\mathbf{s}}) \times m}$, which can extract the observed and unobserved elements when multiplied by $\mathbf{Y}(\mathbf{s})$. More precisely, $\mathbf{R}_1(\mathbf{s}) = (\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_{d_{\mathbf{s}}}})'$ and $\mathbf{R}_2(\mathbf{s}) = (\mathbf{v}_{i_{d_{\mathbf{s}}+1}}, \dots, \mathbf{v}_{i_m})'$.

Observe that $\mathbf{Y}_o(\mathbf{s}) = \mathbf{R}_1(\mathbf{s})\mathbf{Y}(\mathbf{s})$ and $\mathbf{Y}_u(\mathbf{s}) = \mathbf{R}_2(\mathbf{s})\mathbf{Y}(\mathbf{s})$ consist of the observed and unobserved elements, respectively. Multiplying both sides of model (2) by $\mathbf{R}_1(\mathbf{s})$ reveals that the likelihood $p(\mathbf{Y}_o(\mathbf{s}) | \mathbf{f}(\mathbf{s}), \gamma, \beta, \Lambda, \Psi)$ follows a multivariate normal distribution with variance $\mathbf{R}_1(\mathbf{s})\Psi\mathbf{R}_1(\mathbf{s})'$ and mean $\mathbf{R}_1(\mathbf{s})[\mathbf{X}(\mathbf{s})'\beta + \sum_{k=1}^r \delta_k f_k(\mathbf{s})\lambda_k]$. Note that $\mathbf{R}_1(\mathbf{s})\Psi\mathbf{R}_1(\mathbf{s})'$ is a $d_{\mathbf{s}} \times d_{\mathbf{s}}$ diagonal matrix with l th diagonal element $\psi_{i_l}^2$ for $l = 1, \dots, d_{\mathbf{s}}$. This ensures that $\mathbf{R}_1(\mathbf{s})\Psi\mathbf{R}_1(\mathbf{s})'$ is a valid covariance matrix, and so is $\mathbf{R}_2(\mathbf{s})\Psi\mathbf{R}_2(\mathbf{s})'$ for similar reasons. See Web Appendix B for implementation details.

3.3 Prediction and Predictive Model Comparison

Spatial analysis for data sets with missing values customarily entail three types of inference: (i) recover missing values, (ii) predict outcome $\mathbf{Y}(\mathbf{t}) = (Y_1(\mathbf{t}), \dots, Y_m(\mathbf{t}))'$ at any arbitrary location \mathbf{t} , and (iii) generate replicates for the observed data for use in model assessment and selection (Gelfand and Ghosh, 1998).

To recover the missing values, we use composition sampling to draw \mathbf{Y}_u from $p(\mathbf{Y}_u | \mathbf{Y}_o)$, where $\mathbf{Y}_o = (\mathbf{Y}_o(\mathbf{s}_1), \dots, \mathbf{Y}_o(\mathbf{s}_n))'$ and $\mathbf{Y}_u = (\mathbf{Y}_u(\mathbf{s}_1), \dots, \mathbf{Y}_u(\mathbf{s}_n))'$. Note that

$$\begin{aligned} p(\mathbf{Y}_u | \mathbf{Y}_o) &= \int \prod_{i=1}^n p(\mathbf{Y}_u(\mathbf{s}_i) | \mathbf{f}(\mathbf{s}_i), \theta) p(\mathbf{F}, \theta | \mathbf{Y}_o) d\mathbf{F} d\theta \\ &= \int \prod_{i=1}^n N(\mathbf{Y}_u(\mathbf{s}_i) | \mathbf{R}_2(\mathbf{s}_i)[\mathbf{X}(\mathbf{s}_i)'\beta + \tilde{\Lambda}\mathbf{f}(\mathbf{s}_i)], \\ &\quad \times \mathbf{R}_2(\mathbf{s}_i)\Psi\mathbf{R}_2(\mathbf{s}_i)') p(\mathbf{F}, \theta | \mathbf{Y}_o) d\mathbf{F} d\theta, \end{aligned}$$

where $\tilde{\Lambda} = (\delta_1 \lambda_1, \dots, \delta_r \lambda_r)$. Using the samples $\{\mathbf{F}^{(l)}, \boldsymbol{\theta}^{(l)}\}$ from $p(\mathbf{F}, \boldsymbol{\theta} | \mathbf{Y}_o)$, the missing value at \mathbf{s} is recovered by drawing $\mathbf{Y}_u^{(l)}(\mathbf{s})$ for each, $l = 1, \dots, L$, from an $(m - d_S) \times 1$ multivariate normal distribution with mean $\mathbf{R}_2(\mathbf{s})[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}^{(l)} + \tilde{\Lambda}^{(l)}\mathbf{f}^{(l)}(\mathbf{s})]$ and variance $\mathbf{R}_2(\mathbf{s})\boldsymbol{\Psi}^{(l)}\mathbf{R}_2(\mathbf{s})'$.

Prediction of the outcomes at unsampled or ungauged sites is often a major study objective. Using notations defined earlier, we have

$$\begin{aligned} p(\mathbf{Y}(\mathbf{t}) | \mathbf{Y}_o) &= \int p(\mathbf{Y}(\mathbf{t}) | \mathbf{f}(\mathbf{t}), \boldsymbol{\theta}) p(\mathbf{f}(\mathbf{t}) | \mathbf{F}, \boldsymbol{\phi}, \mathbf{Y}_o) \\ &\quad \times p(\mathbf{F}, \boldsymbol{\theta} | \mathbf{Y}_o) d\mathbf{F} d\boldsymbol{\theta} d\mathbf{f}(\mathbf{t}) \\ &= \int N(\mathbf{Y}(\mathbf{t}) | \mathbf{X}(\mathbf{t})'\boldsymbol{\beta} + \tilde{\Lambda}\mathbf{f}(\mathbf{t}), \boldsymbol{\Psi}) \prod_{k=1}^r \\ &\quad \times N(f_k(\mathbf{t}) | \mu_{f_k}(\mathbf{t}) | \mathbf{f}_k, \sigma_{f_k}^2(\mathbf{t}) | \mathbf{f}_k) \\ &\quad \times p(\mathbf{F}, \boldsymbol{\theta} | \mathbf{Y}_o) d\mathbf{F} d\boldsymbol{\theta} d\mathbf{f}(\mathbf{t}), \end{aligned}$$

where $\sigma_{f_k}^2(\mathbf{t}) | \mathbf{f}_k = 1 - \mathbf{d}_k(\mathbf{t}; \boldsymbol{\phi}_k)' \mathbf{D}_k^*(\boldsymbol{\phi}_k)^{-1} \mathbf{D}_k(\boldsymbol{\phi}_k)' \mathbf{G}_k(\boldsymbol{\phi}_k)^{-1} \mathbf{D}_k(\boldsymbol{\phi}_k) \mathbf{D}_k^*(\boldsymbol{\phi}_k)^{-1} \mathbf{d}_k(\mathbf{t}; \boldsymbol{\phi}_k)$ and $\mu_{f_k}(\mathbf{t}) | \mathbf{f}_k = \mathbf{d}_k(\mathbf{t}; \boldsymbol{\phi}_k)' \mathbf{D}_k^*(\boldsymbol{\phi}_k)^{-1} \mathbf{D}_k(\boldsymbol{\phi}_k)' \mathbf{G}_k(\boldsymbol{\phi}_k)^{-1} \mathbf{f}_k$. Here, $\mathbf{G}_k(\boldsymbol{\phi}_k)$ is the variance of \mathbf{f}_k , which equals $\mathbf{D}_k(\boldsymbol{\phi}_k) \mathbf{D}_k^*(\boldsymbol{\phi}_k)^{-1} \mathbf{D}_k(\boldsymbol{\phi}_k)' + \boldsymbol{\Sigma}_{\mathbf{f}_k}$. For prediction, we would first sample $f_k(\mathbf{t})^{(l)}$ from $N(f_k(\mathbf{t}) | \mu_{f_k}(\mathbf{t}) | \mathbf{f}_k, \sigma_{f_k}^2(\mathbf{t}) | \mathbf{f}_k)$ using posterior samples $\mathbf{f}_k^{(l)}$ and $\boldsymbol{\phi}_k^{(l)}$. Then, $\mathbf{Y}(\mathbf{t})^{(l)}$ is drawn from an $m \times 1$ multivariate normal distribution with mean $\mathbf{X}(\mathbf{t})'\boldsymbol{\beta}^{(l)} + \tilde{\Lambda}^{(l)}\mathbf{f}(\mathbf{t})^{(l)}$ and variance $\boldsymbol{\Psi}^{(l)}$.

Finally we turn to generating replicated observations for \mathbf{Y}_o . This is easily achieved by first sampling from $p(\mathbf{w}^*, \boldsymbol{\theta} | \mathbf{Y}_o)$. The replicate $\mathbf{Y}_o^{rep}(\mathbf{s})^{(l)}$ is generated from a $d_S \times 1$ multivariate normal distribution with mean $\mathbf{R}_1(\mathbf{s})[\mathbf{X}(\mathbf{s})'\boldsymbol{\beta}^{(l)} + \tilde{\Lambda}^{(l)}\mathbf{w}^*(\mathbf{s})]$ and variance $\mathbf{R}_1(\mathbf{s})[\boldsymbol{\Psi}^{(l)} + \tilde{\Lambda}^{(l)}\boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{s})(\boldsymbol{\phi}^{(l)})'\tilde{\Lambda}^{(l)}]\mathbf{R}_1(\mathbf{s})'$, where $\boldsymbol{\Sigma}_{\mathbf{f}}(\mathbf{s})$ is a diagonal matrix with k th diagonal element $\sigma_{f_k}^2(\mathbf{s})$. Then $\mu_i^{rep} = E[Y_{oi}^{rep} | \mathbf{Y}_o]$ and $\sigma_i^{2rep} = \text{var}\{Y_{oi}^{rep} | \mathbf{Y}_o\}$ can be calculated respectively, where i indexes the observed data. Gelfand and Ghosh (1998) present a posterior predictive criterion that balances goodness-of-fit and predictive variance under a squared error loss function. This assigns a score to each model that is the sum of two terms, P and G , where $G = \sum_i (\mu_i^{rep} - Y_{oi})^2$ is an error sum of squares and represents goodness-of-fit, whereas $P = \sum_i \sigma_i^{2rep}$ represents predictive variance and acts as a penalty term. Therefore, complexity is penalized and a parsimonious choice with lower $D = G + P$ indicates preferred models.

4. Illustrations

We executed our adaptive spatial factor models using the R programming language. The most demanding model (12×1 vector of outcomes across 1,000 locations) took approximately 48 hours to deliver its entire inferential output from 20,000 MCMC iterations, including 4,000 samples for burn-in, on a 3.10-GHz Intel i5-2400 processor with 4.0 Gbytes of RAM. The statistics of Gelman and Rubin (1992) was used

to assess chain convergence alongside visual inspection of the trace plots and empirical autocorrelation functions.

4.1 Simulation Study One

The objective of this simulation study is to explore different specifications for Λ and demonstrate model identifiability. Adaptive factor modeling and predictive processes are not employed here. The data generating model uses a 2×2 factor loading matrix Λ and a spatial range parameter $\boldsymbol{\phi}$ whose first element is smaller than the second. A synthetic data set comprising 250 locations within a $[0, 50] \times [0, 50]$ square was generated from (1) with $r = m = 2$. An isotropic exponential correlation function, $\rho_k(\mathbf{s} - \mathbf{t}; \boldsymbol{\phi}_k) = \exp(-\boldsymbol{\phi}_k \|\mathbf{s} - \mathbf{t}\|)$ for $k = 1, 2$, was used to produce a spatially dependent bivariate random field. The bivariate outcomes, $\mathbf{Y}(\mathbf{s})$, were simulated with the following parameters:

$$\boldsymbol{\beta} = \begin{pmatrix} 5 \\ 10 \end{pmatrix}, \quad \Lambda = \begin{pmatrix} 3 & 1 \\ -2 & 2 \end{pmatrix}, \quad \boldsymbol{\Psi} = \begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix}, \quad \text{and } \boldsymbol{\phi} = \begin{pmatrix} 0.1 \\ 0.6 \end{pmatrix},$$

where $\boldsymbol{\beta}$ is the mean vector for $\mathbf{Y}(\mathbf{s})$ and $\mathbf{C}(\mathbf{0}) = \Lambda\Lambda' = \begin{pmatrix} 10 & -4 \\ -4 & 8 \end{pmatrix}$. Note the strong negative cross-correlation between the two spatial processes. Also, the correlation decays six times faster in the second process than in the first, yielding spatial ranges $-\log(0.05)/\phi_1 \approx 30$ and $-\log(0.05)/\phi_2 \approx 5$ units for the first and second process, respectively.

Three models are estimated and compared. The first restricts Λ to be lower-triangular, as in the LMC setting, but does not order the ϕ_i 's a priori. The other two models use factor loadings with positive elements in the first row of Λ and use the prior in (3) for $\boldsymbol{\phi}$. One of these two models assumes $\phi_1 < \phi_2$, reflecting the true ordering, although the other reverses the inequality and results in a misspecified model a priori. However, because the factor loadings are now identifiable, the estimated Λ contains permuted columns according to the order of the ϕ_i 's. This is attractive in practice—one need not ascertain the correct ordering of the range parameters a priori. We arrange them in the same order as the true values although presenting estimates in Table 1.

In the first model $\Lambda_{21} = 0$, ϕ_i 's are assigned uniform priors over (ϕ_l, ϕ_u) without any further restrictions. Priors for the two other models are assigned as discussed in Section 2.4 with hyperparameters $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \mathbf{0}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} = \mathbf{O}$, $c_0^2 = 10$, $\phi_l = 0.047$, $\phi_u = 25$, $a = 2$, and $b = 5$. Posterior inferences for the latter two models are consistent. This indicates the importance of ordering constraints; without them, the posterior distributions for $\boldsymbol{\phi}$ and Λ can produce multiple modes as the MCMC chain could jump between the possible specifications (Lopes and West, 2004). The parameter Λ_{21} (boldface in Table 1), which was set as 0 in the lower triangular setting, is significantly different from 0 in both the models using general factor loadings. Thus, we can capture the underlying cross-covariance structure without restricting the loading matrix, but by imposing an ordering on the spatial decay parameters. In addition, our proposed models have a lower posterior predictive model comparison score ($D = 14,202$) than the lower triangular setting ($D = 14,941$). This, again, demonstrates a preference for using general factor loadings.

Table 1
Posterior percentiles (50%, 2.5%, and 97.5%) estimated for the parameters in different model specifications

True parameter	Lower triangular setting	General factor loadings	
		Increasing ϕ_k	Decreasing ϕ_k
$\beta_1 = 5$	4.77 (1.22, 7.49)	5.00 (3.29, 6.98)	5.58 (3.77, 7.85)
$\beta_2 = 10$	10.38 (8.65, 12.73)	10.15 (8.38, 11.60)	9.68 (7.73, 11.16)
$C(\mathbf{0})_{1,1} = 10$	8.59, (4.76, 15.08)	10.82 (6.53, 18.54)	10.65 (6.23, 17.66)
$C(\mathbf{0})_{2,1} = -4$	-5.64 (-9.99, -2.93)	-5.34 (-10.84, -1.82)	-5.11 (-9.88, -1.67)
$C(\mathbf{0})_{2,2} = 8$	12.17 (6.76, 17.54)	12.99 (7.15, 19.43)	12.83 (7.15, 18.42)
$\Psi_1 = 2$	3.65 (2.61, 4.88)	2.30 (1.19, 3.66)	2.35 (1.15, 3.55)
$\Psi_2 = 5$	3.27 (0.97, 8.38)	3.33 (1.06, 7.99)	3.21 (0.99, 7.99)
$\phi_1 = 0.1$	0.06 (0.05, 0.12)	0.07 (0.05, 0.13)	0.07 (0.05, 0.14)
$\phi_2 = 0.6$	0.95 (0.25, 2.15)	1.17 (0.52, 2.58)	1.15 (0.49, 2.44)
$\Lambda_{1,1} = 3$	2.93 (2.18, 3.88)	3.11 (2.32, 4.15)	3.07 (2.31, 4.06)
$\Lambda_{1,2} = -2$	-1.93, (-2.80, -1.17)	-2.53 (-3.42, -1.79)	-2.48 (-3.33, -1.68)
$\Lambda_{2,1} = 1$	Fixed at 0	1.02 (0.56, 1.68)	1.01 (0.55, 1.78)
$\Lambda_{2,2} = 2$	2.89, (1.84, 3.47)	2.57 (1.33, 3.22)	2.60 (1.34, 3.23)

4.2 Simulation Study Two

Now we demonstrate the adaptive spatial factor model using a simulated data set with missing values. We generated 1,000 locations in a $[0, 30] \times [0, 30]$ square. At each location, we simulated a 12×1 vector of outcomes, $\mathbf{Y}(\mathbf{s})$, from (1) with three factors and outcome-specific intercepts as the only regressors. Each spatial process, $w_k(\mathbf{s})$, was generated from an isotropic exponential correlation function. The parameters are defined as:

$$\boldsymbol{\mu} = \{9, 5, -7, -4, -15, 7, -5, -6, 13, -1, 0, -9\},$$

$$\boldsymbol{\phi} = \{0.1, 0.8, 1.5\},$$

$$\boldsymbol{\Psi} = \text{diag}\{0.75, 1.31, 1.37, 1.21, 1.44, 1.49, 1.47, 1.48, 0.50, 1.37, 0.65, 1.18\},$$

$$\boldsymbol{\Lambda}' = \begin{pmatrix} 2.2 & -2.6 & 1.8 & 0.4 & -2.8 & -1.1 & -3.0 & -1.0 & 1.9 & -2.0 & 1.4 & 1.7 \\ 1.6 & -1.4 & 2.1 & 2.6 & -2.6 & -2.4 & 2.9 & -1.2 & -0.1 & 2.8 & -1.4 & -2.8 \\ 2.0 & 0.9 & 2.8 & -0.8 & 2.7 & -1.7 & -2.9 & 0.9 & 0.1 & -2.9 & -2.2 & 3.0 \end{pmatrix}.$$

After the data was generated, we allowed a quarter of the locations to retain all the outcomes, although randomly omitting 2/3 of the outcomes from the remaining locations. This produced a data set with approximately 50% missingness.

We applied the Gibbs sampling algorithm described in Section 3 for estimation. For the predictive process, we selected 200 knots using a K-means clustering algorithm (Hartigan and Wong, 1979). The maximum number of latent factors, r , was taken to be 5. For comparison, we also fit spatial factor models using a fixed number of latent factors for the same data set. The prior distributions resemble those in Section 4.1, except that now $\phi_l = 0.07$ and $\phi_u = 310$ to better reflect the domain. Also, for the adaptive model we used independent Bernoulli priors for $\boldsymbol{\delta}$ with $p(\delta_k = 1 | \omega) = \omega$, $k = 1, \dots, r$ and $\omega \sim U(0, 1)$.

Stochastic selection produced the final model with three latent factors based upon the highest posterior probabilities. Figure 1 illustrates the posterior distributions for $\boldsymbol{\phi}$. The spatial parameters in the active processes are ϕ_1, ϕ_2, ϕ_3 (corresponding to $\delta_k = 1$) and their posteriors are well identified. The true values are all included in the central 95% credible

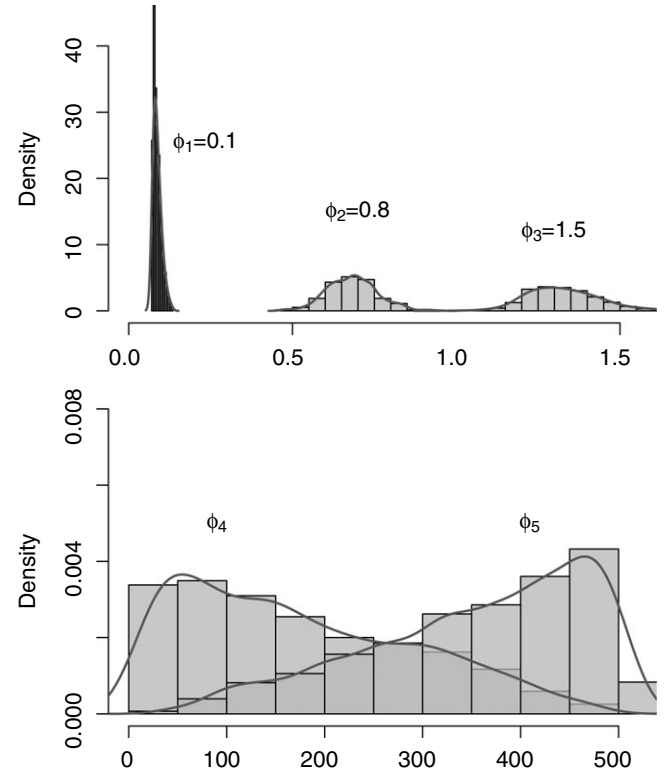


Figure 1. Histograms of the posterior distributions for $\boldsymbol{\phi}$.

intervals. The posteriors for ϕ_4 and ϕ_5 only reflect prior information. The wide range of their posteriors suggest that there is very little posterior learning and the MCMC sampler is sampling from the prior. The marginal posterior density for ω reveals substantial posterior learning. The credible intervals of $\boldsymbol{\beta}$ and $\boldsymbol{\Psi}$ contain the true values, as does most of the elements of $\boldsymbol{\Lambda}$, suggesting sound estimative performance.

Spatial factor models with fixed number of latent factors equaling 1, 2, 4, and 5 are also fitted. The posterior predictive criterion discussed in Section 3.3 is presented for model

Table 2
Simulation study two: Model comparison criteria

Criterion	Number of factors				
	$r = 1$	$r = 2$	Adaptive factor	$r = 4$	$r = 5$
G	69567	53893	45982	46011	46036
P	71698	60173	53758	53880	53939
D	141265	114066	99740	99891	99975

assessment and the scores are compared in Table 2. The adaptive spatial factor model with $D = 99740$ (boldface in Table 2) seems to be the winner here. There is overwhelming evidence that the regression model with random processes $f_1(\mathbf{s})$, $f_2(\mathbf{s})$, and $f_3(\mathbf{s})$ is optimal. This, pleasantly, agrees with the true specification.

4.3 Air Monitor Value Data

Our current application involves data compiled from the Environmental Protection Agency across 316 air monitors in California involving five pollutants: carbon monoxide (CO), nitrogen dioxide (NO₂), ozone (O₃), particulate matter with diameter < 2.5 micrometers (PM2.5), and particulate matter with diameter < 10 micrometers (PM10). Concentrations (ppm) measured by monitoring equipments are used for CO, NO₂, and O₃ and ($\mu\text{g}/\text{m}^3$) for PM2.5 and PM10. Here we use the annual average of the monitor values recorded in 2008. All the variables were standardized to mean 0 and variance 1.

For the modeling, each pollutant is assigned its own intercept term. Elevation (in kilometers) is the only predictor and is depicted as a contoured image in the top-left panel of Figure 2. The instruments only monitor some of the five pollutants at most of the sites, so the data set has about 53% of observations missing. The longitude and latitude of the monitors are transformed to Easting and Northing in kilometer units. For the predictive process, 50 knots were selected using a K-means clustering algorithm. Interpolation for the air pollutants and the locations of monitor sites (\cdot) appear in Figure 2.

Visually, the air pollutants exhibit spatially varying concentrations as well as the associations among them at a given location. We seek to model both kinds of dependence. It is not clear whether it is appropriate to assume a single factor underlying the different air pollutants or if additional factors need to be introduced. To address this question, we repeated the approach described in Section 3 using the exact model (4). The priors are the same as in Section 4.2 except for ϕ , which has distribution (3) with $\phi_l = 0.002$ and $\phi_u = 136$.

The model is implemented in the same manner as in Section 4.2. The maximum possible number of factors r is taken to be 3. We decide to use 4,000 iterations for a burn-in period, and then a further 16,000 iterations for posterior inference.

The probability assigned to the model with one latent factor is 1, suggesting that one factor is sufficient. The interpolation of the latent spatial processes are displayed in Figure 3. The panel at the top depicts the only active process ($\delta_1 = 1$), which presents some characteristics for the spatial concentrations of the air pollutants. The panels at the bottom do not show significant association with the data and

only demonstrate the prior information. For each air pollutant $Y_j(\mathbf{s})$, the percentage of the variance explained by the spatial factor $f_k(\mathbf{s})$, is simply $\delta_k \Lambda_{jk}^2 / (\sum_{k=1}^r \delta_k \Lambda_{jk}^2 + \psi_j^2)$, $j = 1, \dots, 5$. Overall, about 50% of the variation is explained by the first latent factor—NO₂ and CO are more closely related to this latent factor, although PM2.5 is weakly explained.

Table 3 presents the posterior inferences of β , Ψ , ϕ , Λ , and the correlation $\rho = \{\rho_{i,j}\}$, $i, j = 1, \dots, 5$, $j > i$, among air pollutants, which is estimated from the variance estimator $\Lambda\Lambda' + \Psi$. Only CO and NO₂ reveal significant (negative) intercept coefficients. The slope parameters seem to suggest that elevation has a significant impact only upon Ozone among the five pollutants. The posterior inference for ϕ_1 has median 0.018, which corresponds to the effective range (i.e., the distance at which the correlation drops to 0.05) roughly 160 kilometers. The posterior inferences for ϕ_2 and ϕ_3 indicate weak spatial dependence. So we only present the first column of Λ and omit the rest in Table 3. The Λ_{j1} 's reflect the influence of the common factor on the corresponding outcome variables. Because the residual variances, ψ_j^2 's, differ across j , the Λ_{j1} 's measure covariances on different scales and cannot be directly compared to assess strength of association between outcomes. The parameters $\psi_1^2, \dots, \psi_5^2$ can be interpreted as independent measurement error variances. The covariances of the outcomes collected within one set are $\lambda_1 \lambda_1' + \Psi$, where $\lambda_1 = (\Lambda_{11}, \dots, \Lambda_{51})'$. The corresponding correlations among the outcomes are also presented in Table 3.

5. Summary

We have addressed the problem of modeling large multivariate spatial data, where dimension reduction is sought both in the number of outcomes and in the number of spatial locations. The former is achieved with fewer number of factors, although the latter is achieved using a knot-based predictive process. Model-based strategies (Guhaniyogi et al., 2011) exist for choosing knots but recent findings, including our own explorations here, suggests that simple space-covering or clustering algorithms usually deliver robust inference. Our *adaptive* model lets the number of factors be stochastic and allows the data to drive the inference.

One can use either the predictive process $\tilde{w}_k(\mathbf{s})$ or its “modification” $f_k(\mathbf{s})$ in the adaptive spatial factor models but $f_k(\mathbf{s})$ adapts better to the data and, hence, is less sensitive to the knots. Although the substantive inference is quite robust to both these processes, some subtle differences are seen in the estimation algorithm with regard to transition among models. This is related to the strength of the spatial random field and is explained in Web Appendix C.

Given the widespread use of R as a statistical language, we presented computational benchmarks on R running on fairly standard architectures. In multivariate spatial analysis, “large” refers to the size of $m \times n$, where m is the number of outcomes and n is the number of locations. With R running on standard architectures, $mn > 1,000$ is usually deemed exorbitant without dimension reduction. Substantial computational gains accrue from using lower-level languages (C/C++) on shared memory systems.

Alternative approaches for modeling large spatial data sets include an SPDE/GMRF approach proposed by Lindgren, Rue, and Lindström (2011) that uses explicit Markov

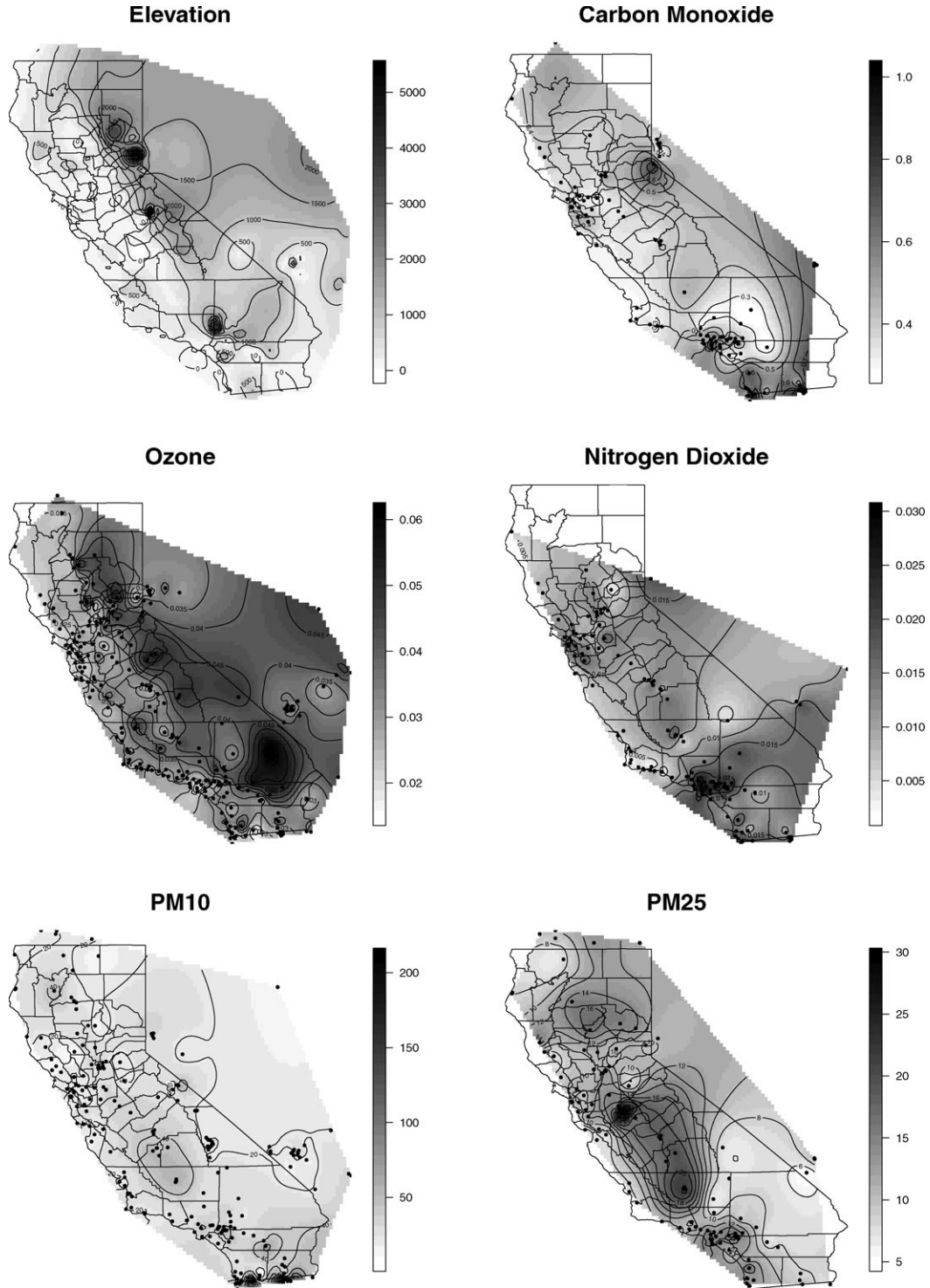


Figure 2. Interpolation of air pollutants measured on monitor sites across California.

representations of the Matérn covariance family using a class of stochastic partial differential equations. Rather than MCMC, they use a faster Integrated Nested Laplace Approximation algorithm for Bayesian inference. Another approach, termed covariance tapering (Furrer, Genton, and Nychka 2006), relies upon compactly supported correlation

functions to produce sparse covariance matrices containing only a moderate number of nonzero elements. How effective these alternative approaches will be with dynamic spatial factor models is yet to be ascertained.

Space-time or dynamic factor modeling using low-rank processes can also be envisioned. Misalignment can now occur

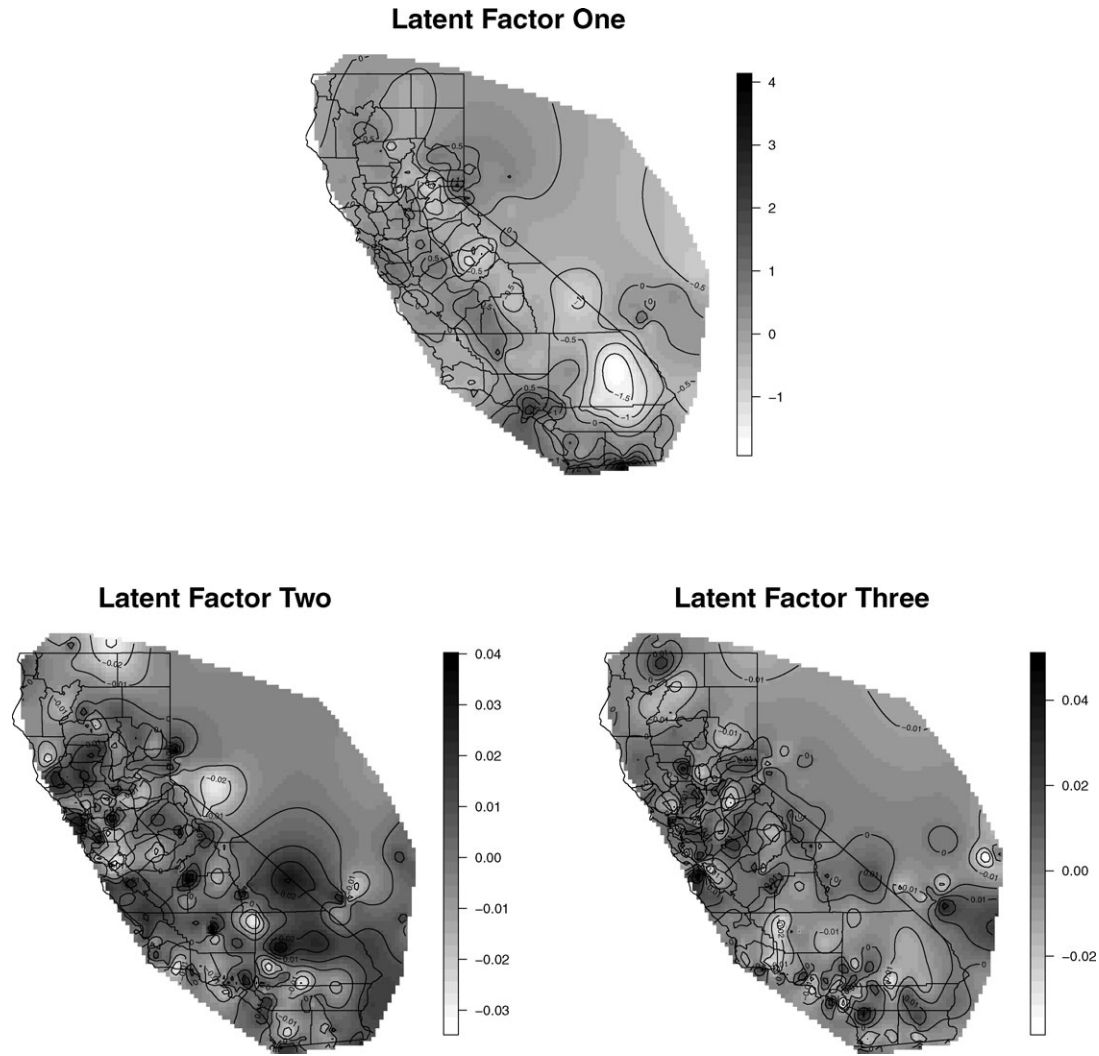


Figure 3. Interpolation of latent spatial factors across California.

Table 3

The posterior credible intervals estimated for the parameters in air pollutants data set. The subscripts 1–5 in β , Ψ , ρ , and the row index of λ_1 refer to CO , NO_2 , O_3 , PM_{10} , and PM_{25} , respectively. Subscripts on ϕ refer to the three spatial range parameters

Parameter	50% (2.5%, 97.5%)	Parameter	50% (2.5%, 97.5%)	Parameter	50% (2.5%, 97.5%)
$\beta_{1,0}$	−0.76 (−1.37, −0.29)	Λ_{11}	0.85, (0.58, 1.39)	$\rho_{1,2}$	0.60 (0.44, 0.80)
$\beta_{2,0}$	−0.61 (−1.23, −0.14)	Λ_{21}	0.88 (0.64, 1.47)	$\rho_{1,3}$	−0.45 (−0.70, −0.27)
$\beta_{3,0}$	0.04 (−0.24, 0.41)	Λ_{31}	−0.57 (−0.98, −0.36)	$\rho_{1,4}$	0.50 (0.33, 0.68)
$\beta_{4,0}$	−0.30 (−0.69, 0.05)	Λ_{41}	0.68, (0.49, 1.02)	$\rho_{1,5}$	0.39 (0.19, 0.64)
$\beta_{5,0}$	−0.27 (−0.72, 0.07)	Λ_{51}	0.55 (0.27, 1.00)	$\rho_{2,3}$	−0.48 (−0.73, −0.28)
$\beta_{1,Elev}$	−0.02 (−0.50, 0.45)	ψ_1^2	0.52 (0.35, 0.77)	$\rho_{2,4}$	0.52, (0.37, 0.70)
$\beta_{2,Elev}$	0.33 (−0.24, 0.91)	ψ_2^2	0.44 (0.30, 0.63)	$\rho_{2,5}$	0.41 (0.20, 0.66)
$\beta_{3,Elev}$	0.40 (0.21, 0.58)	ψ_3^2	0.59 (0.44, 0.79)	$\rho_{3,4}$	−0.39 (−0.60, −0.23)
$\beta_{4,Elev}$	0.02 (−0.28, 0.33)	ψ_4^2	0.63 (0.45, 0.85)	$\rho_{3,5}$	−0.30 (−0.55, −0.14)
$\beta_{5,Elev}$	−0.12 (−0.43, 0.16)	ψ_5^2	0.86 (0.62, 1.19)	$\rho_{4,5}$	0.33 (0.15, 0.55)
ϕ_1	0.018 (0.004, 0.039)				
ϕ_2	43 (5.1, 110)				
ϕ_3	101 (33, 136)				

both over space and over time to yield data matrices that are highly irregular. Our formulation can, nevertheless, be easily adapted to such settings. Also, although we attended only to point-referenced data here, adaptive space-time factor models could be used for multivariate regionally aggregated data (Jin, Banerjee, and Carlin 2007) as well. Here, usually the number of regions is not onerous, so dimension reduction is relevant over the number of outcomes (FA) and time (a temporal predictive process).

6. Supplementary Materials

The Web Appendices referenced in Sections 2.2, 3.1, 3.2, and 5, the ambient air quality data analyzed in Section 4.3, and the R code implementing our models are available with this paper at the *Biometrics* website on Wiley Online Library.

ACKNOWLEDGEMENTS

The authors thank the referees, associate editor, and editor for valuable comments resulting in substantial improvements to the manuscript. This work was supported by NIH grant 1-RC1-GM092400-01 and NSF grant DMS-1106609.

REFERENCES

- Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd edition. New York, NY, USA: Wiley.
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, Florida: CHAPMAN & HALL/CRC.
- Banerjee, S., Finley, A. O., Waldmann, P., and Ericsson, T. (2010). Hierarchical spatial process models for multiple traits in large genetic trials. *Journal of the American Statistical Association* **105**, 506–521.
- Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society Series B Statistical Methodology* **70**, 825–848.
- Berger, J. O. and Pericchi, L. R. (2001). Objective Bayesian methods for model selection: Introduction and comparison. *Lecture Notes-Monograph Series* **38**, 135–207.
- Buttafuoco, G., Castrignanò, A., Colecchia, A. S., and Ricca, N. (2010). Delineation of management zones using soil properties and a multivariate geostatistical approach. *The Italian Journal of Agronomy* **4**, 323–332.
- Cai, B. and Dunson, D. B. (2006). Bayesian covariance selection in generalized linear mixed models. *Biometrics* **62**, 446–457.
- Castrignanò, A., Cherubini, C., Giasi, C. I., Castore, M., Mucci, G. D., and Molinari, M. (2005). Using multivariate geostatistics for describing spatial relationships among some soil properties. In *ISTRO Conference Brno*, 383–390. Brno, Czech Republic: ISTRO.
- Chen, Z. and Dunson, D. B. (2003). Random effects selection in linear mixed models. *Biometrics* **59**, 762–769.
- Chilès, J. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. New York, USA: John Wiley.
- Christensen, W. F. and Amemiya, Y. (2002). Latent variable analysis of multivariate spatial data. *Journal of the American Statistical Association* **97**, 302–317.
- Cressie, N. A. C. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Mathematical Geology* **70**, 209–226.
- Diggle, P. and Lophaven, S. (2006). Bayesian geostatistical design. *Scandinavian Journal of Statistics* **33**, 53–64.
- Dominici, F., Peng, R., Bell, M., Pham, L., McDermott, A., Zeger, S., and Samet, J. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *Journal of the American Medical Association* **295**, 1127–1134.
- Dunson, D. B. (2006). Efficient Bayesian model averaging in factor analysis. *Isds discussion paper*, Duke University.
- Finley, A. O., Banerjee, S., Ek, A. R., and McRoberts, R. E. (2008). Bayesian multivariate process modeling for prediction of forest attributes. *Journal of Agricultural, Biological, and Environmental Statistics* **1**, 60–83.
- Finley, A. O., Sang, H., Banerjee, S., and Gelfand, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Computational Statistics & Data Analysis* **53**, 2873–2884.
- Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal Of Computational And Graphical Statistics* **15**, 502–523.
- Gelfand, A. E. and Ghosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika* **85**, 1–11.
- Gelfand, A. E., Schmidt, A. M., Banerjee, S., and Sirmans, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *TEST* **13**, 263–312.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–472.
- Guhaniyogi, R., Finley, A. O., Banerjee, S., and Gelfand, A. E. (2011). Adaptive gaussian predictive process models for large spatial datasets. *Environmetrics* **22**, 997–1007.
- Hartigan, J. A. and Wong, M. A. (1979). A k-means clustering algorithm. *Applied Statistics* **28**, 100–108.
- Hogan, J. W. and Tchernis, R. (2004). Bayesian factor analysis for spatially correlated data, with application to summarizing area-level material deprivation from census data. *Journal of the American Statistical Association* **99**, 314–324.
- Jin, X., Banerjee, S., and Carlin, B. P. (2007). Order free co-regionalized areal data models with application to multiple-disease mapping. *Journal Of The Royal Statistical Society Series B* **69**, 817–838.
- Kamman, E. and Wand, M. (2003). Geoadditive models. *Applied Statistics* **52**, 1–18.
- Kuo, L. and Mallick, B. (1998). Variable selection for regression models. *Sankhya, Series B, The Indian Journal of Statistics* **60**, 65–81.
- Lark, R. and Papritz, A. (2003). Fitting a linear model of coregionalization for soil properties using simulated annealing. *Geoderma* **115**, 245–260.
- Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society - Series B: Statistical Methodology* **73**, 423–498.
- Liu, X., Wall, M. M., and Hodges, J. S. (2005). Generalized spatial structural equation models. *Biostatistics* **6**, 539–557.
- Lopes, H. F. and West, M. (1999). Model uncertainty in factor analysis. Technical report, Institute of Statistics and Decision Sciences, Duke University.
- Lopes, H. F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica* **14**, 41–67.
- Matheron, G. (1982). Pour une analyse krigéante des données régionalisées. Report N-732, *Centre de Géostatistique*. Fontainebleau, France: E.N.S.M.P.
- Minozzo, M. and Fruttini, D. (2004). Loglinear spatial factor analysis: An application to diabetes mellitus complications. *Environmetrics* **15**, 423–434.
- Paciorek, C. J. (2007). Computational techniques for spatial logistic regression with large datasets. *Computational Statistics and Data Analysis* **51**, 3631–3653.
- Royle, J. A. and Berliner, L. M. (1999). A hierarchical approach to multivariate spatial modeling and prediction. *Journal of Agricultural Biological and Environmental Statistics* **4**, 29–56.

- Royle, J. A. and Nychka, D. (1998). An algorithm for the construction of spatial coverage designs with implementation in spls. *Computers & Geosciences* **24**, 479–488.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
- Schmidt, A. M. and Gelfand, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research* **108**, 8783–8788.
- Tanner, M. A. (1993). *Tools for Statistical Inference. Methods for the Exploration of Posterior Distributions and Likelihood Functions*. New York, USA: Springer-Verlag.
- Ver Hoef, J. M. and Barry, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference* **69**, 275–294.
- Ver Hoef, J. M., Cressie, N., and Barry, R. P. (2004). Flexible spatial models based on the fast fourier transform (fft) for cokriging. *Journal of Computational and Graphical Statistics* **13**, 265–282.
- Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction With Applications*. Berlin: Springer-Verlag Telos.
- Wang, F. and Wall, M. M. (2003). Generalized common spatial factor model. *Biostatistics* **4**, 569–582.
- Webster, R., Atteias, O., and Dubois, J. P. (1994). Coregionalization of trace metals in the soil in the swiss jura. *European Journal of Soil Science* **45**, 205–218.
- Wikle, C. and Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika* **86**, 815–829.
- Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association* **99**, 250–261.
- Zhang, H. (2007). Maximum-likelihood estimation for multivariate spatial linear coregionalization models. *Environmetrics* **18**, 125–139.

Received December 2011. Revised August, 2012.

Accepted August 2012.