

MULTIVARIATE SPATIAL NONPARAMETRIC MODELLING VIA KERNEL PROCESSES MIXING

Montserrat Fuentes and Brian Reich

North Carolina State University, Raleigh

Abstract: In this paper we develop a nonparametric multivariate spatial model that avoids specifying a Gaussian distribution for spatial random effects. Our nonparametric model extends the stick-breaking (SB) prior of Sethuraman (1994), that is frequently used in Bayesian modelling to capture uncertainty in the parametric form of an outcome. The stick-breaking prior is extended here to the spatial setting by assigning each location a different, unknown distribution, and smoothing the distributions in space with a series of space-dependent kernel functions that have a space-varying bandwidth parameter. This results in a flexible nonstationary spatial model, as different kernel functions lead to different relationships between the distributions at nearby locations. This approach is the first to allow both the probabilities and the point mass values of the SB prior to depend on space. Thus, there is no need for replications and we obtain a continuous process in the limit. We extend the model to the multivariate setting by having, for each process, a different kernel function, but sharing the location of the kernel knots across the different processes. The resulting covariance for the multivariate process is in general nonstationary and nonseparable. The modelling framework proposed here is also computationally efficient because it avoids inverting large matrices and calculating determinants. We study the theoretical properties of the proposed multivariate spatial process. The methods are illustrated using simulated examples and an air pollution application to model components of fine particulate matter.

Key words and phrases: Dirichlet processes, nonseparability, nonstationarity, spatial models.

1. Introduction

This paper focuses on the problem of modelling the unknown distribution of a *multivariate spatial* process. We introduce a nonparametric model that avoids specifying a Gaussian distribution for the spatial random effects. This model is flexible enough to characterize the potentially complex spatial structures of the tails and extremes of multivariate distributions, and it is computationally efficient.

Our nonparametric model extends the stick-breaking (SB) prior of Sethuraman (1994) that is frequently used in Bayesian modelling to capture uncertainty

in the parametric form of a distribution. For general (non-spatial) Bayesian modelling, the stick-breaking prior offers a way to model a distribution of a parameter as an unknown quantity to be estimated from the data. The stick-breaking prior for the unknown distribution F , is

$$F \stackrel{d}{=} \sum_{i=1}^M p_i \delta(X_i),$$

where the number of mixture components M may be infinite, $p_i = V_i \prod_{j=1}^{i-1} (1 - V_j)$, $V_i \sim \text{Beta}(a, b)$ independent across i , $\delta(X_i)$ is the Dirac distribution with point mass at X_i , $X_i \stackrel{i.i.d.}{\sim} F_o$, and F_o is a known distribution. A special case of this prior is the Dirichlet process prior with infinite M and $V_i \stackrel{i.i.d.}{\sim} \text{Beta}(1, \nu)$ (Ferguson (1973)). The stick-breaking prior has been extended to the univariate spatial setting by incorporating spatial information into either the model for the values of X_i or the model for the masses p_i . Gelfand, Kottas, and MacEachern (2005), Gelfand, Guindani, and Petrone (2007), and Petrone, Guindani, and Gelfand (2008) model the X_i as vectors drawn from a spatial distribution, in particular Petrone, Guindani, and Gelfand (2008) extend this type of Dirichlet mixture model for functional data analysis. However, their model requires replication. Griffin and Steel (2006) propose a spatial Dirichlet model that permutes the V_i based on spatial location, allowing the occurrence of X_i to be more or less likely in different regions of the spatial domain. The kernel functions in our proposed methodology impose a natural ranking for the different mixture components based on distances of locations to knots, which is the role the permutations and ranking in Griffin and Steel's approach would play. Reich and Fuentes (2007) introduce a SB prior allowing the probabilities p_i to be space-dependent by using kernel functions that have independent and identically distributed (i.i.d.) bandwidths (not space-dependent), but this spatial SB prior has a limiting process that is not continuous. This model is similar to that of Dunson and Park (2008), who use kernels to smooth the weights in the non-spatial setting. An et al. (2008) extend the kernel SB prior for use in image segmentation. Reich and Fuentes (2007) use the kernel SB prior in a multivariate setting, but it has a separable cross-covariance.

Here we introduce an extension of the stick-breaking prior to the multivariate spatial setting that allows for nonseparability and nonstationarity in the spatial cross-dependency between the outcomes of interest. We assign at each location a different, unknown distribution, and we smooth the distributions in space with space-dependent kernel functions, allowing then for nonstationarity. This is the first approach that allows both the probabilities p_i and the values X_i to depend

on space, consequently there is no need for replications, and we obtain a continuous process in the limit. One of the main challenges when analyzing continuous spatial processes and making Bayesian spatial inference is calculating the likelihood function of the covariance parameters. For large datasets, calculating the matrix inverses and determinants in the likelihood might not be feasible; the modelling framework proposed here is computationally efficient because this is avoided.

We apply our methods to model and characterize the complex spatial structure of air pollution, in particular to components of fine particulate matter. Fine particulate matter ($\text{PM}_{2.5}$) is the general term used for a mixture of solid particles and liquid droplets in the air that are 2.5 microns in diameter and less. It includes aerosols, smoke, fumes, dust, ash, and pollen. $\text{PM}_{2.5}$ has over 40 components and two of the main ones are nitrate and ammonium. The study of the association between ambient particulate matter (PM) and human health has received much attention in epidemiological studies over the past few years, e.g., Dominici et al. (2002). Their results showed the importance of considering particle size, composition, and source information when modeling particle pollution health effects. The $\text{PM}_{2.5}$ chemistry changes with space and time so its association with the health endpoints could change across space and time. Speciated $\text{PM}_{2.5}$ is measured sparsely, and spatial interpolation is needed to conduct epidemiological studies of the spatial association of these pollutants and adverse health effects. We introduce a multivariate spatial model for chemical components of $\text{PM}_{2.5}$ across the United States, and focus on nitrate and ammonium. This is the first study with chemical components of particulate matter that allows the cross-dependency between components to vary spatially.

The paper is organized as follows. In Section 2, we introduce a univariate spatial model using an extension of the stick-breaking prior, which directly models nonstationarity. In Section 3, we present a multivariate extension of this spatial SB model that allows nonseparability and a cross dependency between spatial processes that is space-dependent. In Section 4, we present the conditional and marginal properties of the spatial SB process prior, and we study some asymptotic properties. In Section 5, we show the continuity of the limiting process. In Section 6 we discuss computing methods and MCMC algorithms. In Section 7, we illustrate the proposed methods with a simulation study. In Section 8, we present the application to air pollution. We conclude with remarks and comments in Section 9.

2. Univariate Nonstationary Spatial Model

The spatial distribution of a stochastic process $Y(s)$ is modeled using an extension of the stick-breaking prior that directly models nonstationarity. The

nonparametric spatial model assigns a different prior distribution to the stochastic process at each location, $Y(s) \sim F_s(Y)$. The distributions $F_s(Y)$ are unknown and smoothed spatially. The coordinate s is in $D \in \mathbb{R}^d$. To simplify notation throughout this section we take $d = 2$.

The prior for $F_s(Y)$ is the potentially infinite mixture

$$F_s(Y) = \sum_{i=1}^M p_i(s) \delta(X(\phi_i)) = \sum_{i=1}^M V_i(s) \prod_{j<i} [1 - V_j(s)] \delta(X(\phi_i)), \quad (2.1)$$

where X is a Gaussian process (GP) with covariance Σ_X that has diagonal elements σ_i^2 . $p_1(s) = V_1(s)$, $p_i(s) = V_i(s) \prod_{j=1}^{i-1} (1 - V_j(s))$, $V_i(s) = K_i(s) V_i$, with $V_i \sim \text{Beta}(a, b)$ independent over i , and $\sum_i p_i(s) = 1$ for all s (by supplemental material A.1). The weight function, K_i , is a spatial kernel centered at knot $\phi_i \in \mathbb{R}^2$ with bandwidth parameter ϵ_i . More generally we fit elliptical kernels K_i , with B_i the 2×2 matrix that controls the shape of the ellipse. We write $B_i = T(\phi_i) T(\phi_i)'$, where $T(\phi_i)'$ denotes the transpose of $T(\phi_i)$, and $T_{kk'}(\phi_i)$ is the (k, k') element of the matrix $T(\phi_i)$. We normalize the kernel functions to avoid lack of identifiability problems when estimating the V_i components.

In this representation, it is important to notice that the underlying GP X is defined on the *knot* space.

The spatial correlation of the process $Y(s)$ is controlled by the bandwidth parameters associated with knots in the vicinity of s (Section 4). To allow the correlation to vary from region to region, the bandwidth parameters are modeled as spatially varying parameters. Thus, we assign to $T_{ij}(\phi_i)$, for $i, j \in \{1, 2\}$, spatial Gaussian priors with non-zero mean and Matérn covariance functions (Matérn (1960)). For identification purposes we restrict the mean of the diagonal elements of T to be positive. We also consider isotropic kernels, $B_i = \epsilon_i I$, with a log-Gaussian prior for ϵ_i .

In addition to allowing the correlation (via the bandwidth parameters) to be a function of space, we allow the variance to be a spatial process. To do this, the spatial process $X(\phi_i)$ has a zero mean-Gaussian process prior with covariance (Palacios and Steel (2006)),

$$\text{cov}(X(\phi_i), X(\phi_j)) = \sigma_i \sigma_j \rho(|\phi_i - \phi_j|), \quad (2.2)$$

where $\sigma_i = \sigma(\phi_i)$ is the the variance of the process $X(\phi_i)$, space-dependent, ρ is a correlation function, and $|\phi_i - \phi_j|$ denotes the Euclidean distance between ϕ_i and ϕ_j . We assign to $\log(\sigma(\phi_i))$ a spatial Gaussian prior with non-zero mean and a Matérn covariance function. The knots ϕ are random, with spatially uniform prior. In practice we work with a discrete spatial uniform prior to facilitate

sampling of the Gaussian Process (GP) priors that have distances defined in the knot space.

Higdon, Swall, and Kern (1999) introduced a nonstationarity representation of a Gaussian process using a kernel convolution. In that representation the kernels did not have the widths of the elliptical bandwidths space-dependent, only the angles, and the only kernels considered were squared exponential. Here, apart from generalizing the type of kernel convolution in Higdon, Swall, and Kern (1999), the stick-breaking structure adds to the kernel convolution by allowing for lack of normality, and it introduces the flexibility of having the data determine the locations of the most relevant knots.

Model (2.1) is a Dirichlet Process (DP) mixture model with spatially varying weights. The process of interest that we model in practice is $Z(s) = Y(s) + e(s)$, where Y arrives from the DP model in (2.1), and $e(s) \sim N(0, \sigma_0^2)$ is the measurement error process. Thus, in our hierarchical Bayesian framework the response Z given the Y process (SB prior) is independent over space.

In supplemental material A.1 we show that the representation in (2.1) is a properly defined process prior. In Section 4 we study the marginal and conditional properties of (2.1) for a stationary process X .

3. Multivariate Spatial Model

We present a nonparametric multivariate spatial model, that is a multivariate extension of model (2.1). We explain the cross spatial dependency between p stochastic processes, $Y_1(s), \dots, Y_p(s)$, by introducing a model for the distribution of each $Y_k(s)$, for $k = 1, \dots, p$,

$$F_s(Y_k) = \sum_{i=1}^M V_{i,k}(s) \prod_{j<i} [1 - V_{j,k}(s)] \delta(X_k(\phi_i)), \quad (3.1)$$

where, $p_{1,k}(s) = V_{1,k}(s)$, $p_{i,k}(s) = V_{i,k}(s) \prod_{j=1}^{i-1} (1 - V_{j,k}(s))$, with $V_{i,k}(s) = K_{i,k}(s)V_i$, $V_i \sim \text{Beta}(a, b)$ independent over i . The kernel functions $K_{i,k}$ are space-dependent and modelled as in the univariate case, but having the knots of each kernel function shared across the p spatial processes. The p -dimensional process $\mathbf{X} = (X_1, \dots, X_p)$ has a multivariate normal prior. The cross-covariance for the multivariate process \mathbf{X} at each knot i is $\Sigma^{(i)} = A(\phi_i)A'(\phi_i)$, where A is a full rank lower triangular, $A(\phi_i) = \{a_{kk'}(\phi_i)\}_{kk'}$, and for each k and k' in $\{1, \dots, p\}$, $a_{kk'}$ are independent spatial Gaussian non-zero mean processes evaluated at location ϕ_i . For identification purposes we restrict the mean of the diagonal elements of A to be positive. Thus, the process $(X_1(\phi_i), \dots, X_p(\phi_i))$ has a multivariate normal prior with a covariance $\Sigma^{(i)}$ that depends on space via the knot locations. By allowing $\Sigma^{(i)}$ to depend on i , we obtain a cross covariance between the Y_k

processes that varies with space (nonstationary cross-dependence), and it is in general nonseparable in the sense that we do not separately model the cross-dependency between the p processes and the spatial dependency structure. We allow not only the magnitude of the cross-dependency structure to vary across space but also its sign; we call this “nonstationarity for the sign” of the cross-dependency. Most multivariate models, in particular separable models for the covariance, constrain the cross-dependency to be stationary with respect to its sign (e.g., Reich and Fuentes (2007)); Choi et al. (2009)).

As an example, a simpler representation of the nonparametric multivariate spatial model in (3.1) is

$$F_s(Y_k) = \sum_{i=1}^M V_{i,k}(s) \prod_{j<i} [1 - V_{j,k}(s)] \delta(X(\phi_i)),$$

where the V and K random weights are different for each process Y_k . This allows the spatial processes Y_k , to have different spatial structure: the V and K components explain the different spatial structure of the Y_k processes, as well as the strength of the cross-dependence between the Y_k 's. This simpler version can offer computational benefits but it does not allow for nonstationarity in the sign. In the context of speciated particulate matter, the sharing of the underlying process X , may be justified by the components of $\text{PM}_{2.5}$ sharing pollution sources.

4. Asymptotic Dependence and Weak Convergence

In this section we study the properties of the spatial dependence induced by our mixture model. Initially, conditioning on the mixture weights, in which case as the kernel bandwidths become smaller and only the kernels with knots in the vicinity of the data points would be in the mixture, one would expect to have a limiting process with similar properties as the base process X . So if X is stationary, we expect a stationary conditional dependence. The asymptotic results here aid the understanding of the different components in our mixture representation. Proofs are in the supplemental material .

4.1. Conditional univariate spatial covariance

Consider the prior in (2.1) for the data process Y . Without loss of generality, we take $B_i = \epsilon_i I$, where I is the 2×2 identity matrix. Conditional on the probability masses $p_i(s)$ in (2.1), but not on X , the covariance between two observations is,

$$\begin{aligned} & \text{cov}(Y(s), Y(s') | p(s), p(s'), C) \\ &= \sum_i \sigma_i^2 p_i(s) p_i(s') + \sum_{i_1 \neq i_2} p_{i_1}(s) p_{i_2}(s') C(|\phi_{i_1} - \phi_{i_2}|) \end{aligned}$$

$$\begin{aligned}
&= \sum_i \sigma_i^2 \left[K_i(s) K_i(s') V_i^2 \prod_{j < i} (1 - ((K_j(s) + K_j(s')) V_j + K_j(s) K_j(s') V_j^2)) \right] \\
&+ \sum_{i_1 \neq i_2} [K_{i_1}(s) K_{i_2}(s') V_{i_1} V_{i_2} C(|\phi_{i_1} - \phi_{i_2}|) \\
&\quad \prod_{j_1 < i_1} \prod_{j_2 < i_2} (1 - (K_{j_1}(s) V_{j_1} + K_{j_2}(s') V_{j_2}) + K_{j_1}(s) K_{j_2}(s') V_{j_1} V_{j_2})], \quad (4.1)
\end{aligned}$$

where $p(s) = (p_1(s), p_2(s), \dots)$ denotes the potentially infinite dimensional vector with the $p_i(s)$ in the mixture defined as (2.1), C is the covariance function of X , and $\sigma_i^2 = \text{cov}(X(\phi_i), X(\phi_i))$.

The conditional covariance of Y as (4.1), as the bandwidths of the kernel functions become smaller, is stationary and approximates the covariance function C of the underlying process X , $\text{cov}(X(s), X(s')) = C(|s - s'|)$. This result is stated in Theorem 1.

Theorem 1. *Consider the prior as (2.1) for the process Y , and assume that K_i is a kernel with compact support for all i , and that C has a bounded and nonnegative first derivative. Conditioned on the probabilities $p_i(s)$ but not X , the covariance of Y tends to the covariance function C of an isotropic process X , as the bandwidths of the kernel functions K_i as (2.1) go uniformly to zero.*

4.2. Conditional multivariate spatial covariance

Consider the multivariate prior $(F_s(Y_1), \dots, F_s(Y_p))$ as (3.1) for the processes $Y_1(s), \dots, Y_p(s)$, conditioned on the probabilities $p_{i,1}(s)$ and $p_{i,2}(s')$ for each pair of data processes $Y_1(s)$ and $Y_2(s')$, but not on the corresponding underlying multivariate process $\mathbf{X} = (X_1, X_2)$, the cross-covariance between any pair $Y_1(s)$ and $Y_2(s')$, is

$$\begin{aligned}
&\text{cov}(Y_1(s), Y_2(s') | p_1(s), p_2(s'), C) \\
&= \sum_i p_{i,1}(s) p_{i,2}(s') C_{1,2}(\phi_i, \phi_i) + \sum_{i_1 \neq i_2} p_{i_1,1}(s) p_{i_2,2}(s') C_{1,2}(\phi_{i_1}, \phi_{i_2}) \\
&= \sum_i [K_{i,1}(s) K_{i,2}(s') V_{i,1} V_{i,2} C_{1,2}(\phi_i, \phi_i) \\
&\quad \prod_{j < i} (1 - K_{j,1}(s) V_{j,1} - K_{j,2}(s') V_{j,2} + K_{j,1}(s) K_{j,2}(s') V_{j,1} V_{j,2})] \\
&\quad + \sum_{i_1 \neq i_2} [K_{i_1,1}(s) K_{i_2,2}(s') V_{i_1,1} V_{i_2,2} C_{1,2}(\phi_{i_1}, \phi_{i_2}) \\
&\quad \prod_{j_1 < i_1} \prod_{j_2 < i_2} (1 - (K_{j_1,1}(s) V_{j_1,1} + K_{j_2,2}(s') V_{j_2,2}) + K_{j_1,1}(s) K_{j_2,2}(s') V_{j_1,1} V_{j_2,2})]. \quad (4.2)
\end{aligned}$$

This cross-covariance approximates the cross-covariance function $\text{cov}(X_1(s), X_2(s')) = C_{1,2}(s, s')$ of the underlying process $\mathbf{X} = (X_1, X_2)$ used to define the process prior for Y_1 and Y_2 , for small bandwidths of the kernel functions. The formal result is in Theorem 2.

Theorem 2. *Consider the multivariate prior $(F_s(Y_1), \dots, F_s(Y_p))$ for the data processes $Y_1(s), \dots, Y_p(s)$ as (3.1), and suppose the cross-covariance function of the process $\mathbf{X} = (X_1, X_2)$ as (3.1) has first order partial derivatives that are bounded nonnegative functions, and that the kernels functions associated with the processes $Y_1(s)$ and $Y_2(s)$ have compact support. Conditioned on the probabilities $p_{i,1}(s)$ and $p_{i,2}(s')$ for each pair of data processes $Y_1(s)$ and $Y_2(s')$, but not on the corresponding process $X = (X_1, X_2)$, the cross-covariance $\text{cov}(Y_1(s), Y_2(s'))$ tends to the cross-covariance function $C_{1,2}(s, s')$ of the process \mathbf{X} as the bandwidths of the kernel functions for $F_s(Y_1)$ and $F_s(Y_2)$ go uniformly to zero.*

4.3. Marginal properties: univariate case

We study the marginal properties of the prior as (2.1). We start by assuming that the V_i as (2.1) share the Beta(a, b) prior. Then, $E(V_i) = E(V)$, and $E(V_i^2) = E(V^2)$, for all i . We take $B_i = \epsilon_i I$ to simplify the presentation. We suppose the covariance, C , of the process X is stationary with $C(0) = \sigma^2$, and is an integrable function. We consider independent priors for the bandwidth and the knot parameters of the kernel functions K_i , and we study the marginal properties when the V_i components are space dependent functions and X has a nonstationary variance.

Integrating over the probability masses, the marginal covariance is

$$\begin{aligned} \text{cov}(Y(s), Y(s')) &= \sigma^2 c_2 E(V^2) \sum_i [1 - 2c_1 E(V) + c_2 E(V^2)]^{i-1} \\ &\quad + \sum_{i_1 \neq i_2} c_{1,2} [1 - 2c_1 E(V) + c_2 E(V^2)]^{(i_1-1)(i_2-1)} \\ &= \sigma^2 \frac{\gamma(s, s')}{2(1 + b/(a+1)) - \gamma(s, s')} \\ &\quad + E(V)^2 \sum_{i_1 \neq i_2} c_{1,2} [1 - 2c_1 E(V) + c_2 E(V^2)]^{(i_1-1)(i_2-1)}, \quad (4.3) \end{aligned}$$

where

$$\begin{aligned} \gamma(s, s') &= \frac{\int \int K_i(s) K_i(s') p(\phi_i, \epsilon_i) d\phi_i d\epsilon_i}{\int \int K_i(s) p(\phi_i, \epsilon_i) d\phi_i d\epsilon_i}, \\ c_1 &= \int \int K_i(s) p(\phi_i, \epsilon_i) d\phi_i d\epsilon_i, \end{aligned}$$

$$c_2 = \iint K_i(s)K_i(s')p(\phi_i, \epsilon_i)d\phi_id\epsilon_i,$$

$$c_{1,2} = \iiint K_{i_1}(s)K_{i_2}(s')C(|\phi_{i_1} - \phi_{i_2}|)p(\phi_{i_1}, \epsilon_{i_1})d\phi_{i_1}d\epsilon_{i_1}p(\phi_{i_2}, \epsilon_{i_2})d\phi_{i_2}d\epsilon_{i_2}.$$

The first term in (4.3) corresponds to the marginal covariance if the underlying process X is i.i.d. across space rather than a spatial Gaussian process. The second term is due to the spatial dependency of the process X . The first term in (4.3) is a function of s and s' through the function $\gamma(s, s')$, and this can be written as $\gamma_0(s - s')$ when the kernel functions are the same across space, rather than being space-dependent.

Supposing there is a buffer zone to avoid edge effects, $c_{1,2}(s, s')$ as (4.3), denoted by $c_{1,2}$, is the only component that is a function of the covariance of X . When the covariance of X is stationary, $c_{1,2}(s, s') = c_0(h)$. For instance, if X has a squared exponential covariance with range ρ , under squared exponential kernel function with a fixed prior for the bandwidth, we obtain that $c_{1,2}(s, s')$ is stationary and, for small values of the kernel bandwidth ϵ ,

$$c_{1,2} = (2\pi)^2 \left(\frac{\epsilon^2 \rho}{\rho + 2\epsilon} \right) \exp \left\{ -\frac{h'h}{2} \left(\frac{\rho(\rho + 2\epsilon)}{\rho + \epsilon} \right)^{-1} \right\} \quad (4.4)$$

when $(\epsilon_{i_1}, \epsilon_{i_2}) = (1, \epsilon)$ or $(\epsilon_{i_1}, \epsilon_{i_2}) = (\epsilon, 1)$. The second term in (4.3), as the kernel bandwidths tend to zero, is the squared exponential correlation function with range ρ .

From (4.3), one sees that the degree of differentiability of the marginal covariance function for Y is, in general, one degree higher than that of the covariance of X . In (4.3), the moments of V are not space-dependent, but when V is a function of space and the kernel bandwidths are also space-dependent, the resulting marginal covariance is not stationary even for an i.i.d. X .

4.4. Marginal properties: univariate with space-dependent kernels

We calculate the marginal covariance, but we allow the functions V_i to be space dependent, $V_i \sim \text{Beta}(1, \tau_i)$, $E(V_i) = 1/(1 + \tau_i)$. We also allow the variance of the process X to be space-dependent, $\text{cov}(X(\phi_i), X(\phi_i)) = \sigma_i^2$, and the bandwidth parameters ϵ_i of the kernel functions to be space-dependent. Then, we have

$$\begin{aligned} \text{cov}(Y(s), Y(s')) &= \sum_i \sigma_i^2 E(V_i^2) [1 - c_1(s)E(V_i) - c_1(s')E(V_i) + c_2(s, s')E(V_i^2)]^{(i-1)} \\ &\quad + \sum_{i_1 \neq i_2} c(s, s') E(V_{i_1} V_{i_2}) [1 - c_1(s)E(V_{i_1}) - c_1(s')E(V_{i_2}) \\ &\quad + c_2(s, s') E(V_{i_1} V_{i_2})]^{(i_1-1)(i_2-1)}, \end{aligned} \quad (4.5)$$

where

$$\begin{aligned} c_1(s) &= \iint K_i(s)p(\epsilon_i|\phi_i)p(\phi_i)d\phi_id\epsilon_i, \\ c_2(s, s') &= \iint K_i(s)K_i(s')p(\epsilon_i|\phi_i)p(\phi_i)d\phi_id\epsilon_i, \\ c(s, s') &= \iiint K_{i_1}(s)K_{i_2}(s')C(|\phi_{i_1} - \phi_{i_2}|)p(\phi_{i_1})p(\epsilon_{i_1}|\phi_{i_1})d\phi_{i_1}d\epsilon_{i_1} \\ &\quad p(\phi_{i_2})p(\epsilon_{i_2}|\phi_{i_2})d\phi_{i_2}d\epsilon_{i_2}. \end{aligned}$$

The nonstationarity in the variance is due to the space-dependent variance σ_i , the nonstationarity in the correlation (range of dependence) is induced by the space-dependent bandwidths. In the expression for the marginal covariance of the process Y as (4.5), we have the integral of the covariance function for the underlying process X . Thus, the degree of differentiability of the marginal covariance function for Y , is one degree higher than the degree of differentiability for the covariance of X .

4.5. Marginal properties: multivariate cross covariance

We present marginal properties of the multivariate spatial representation as (3.1). We study the marginal cross-covariance between any pair of processes $Y_1(s)$ and $Y_2(s')$, assuming that the cross-covariance $C_{1,2}$ between the underlying processes X_1 and X_2 is an integrable function. We allow the bandwidth parameters to be functions of the knots. Then, integrating over the probability masses, the covariance between two observations is

$$\begin{aligned} \text{cov}(Y_1(s), Y_2(s')) &= \sum_{i_1, i_2} E(V_{i_1,1}V_{i_2,2})c_{i_1, i_2}^*(s, s')[1 - c_{1,1}^*(s)E(V_{i_1,1}) - c_2^*(s')E(V_{i_2,2}) \\ &\quad + c_{i_1, i_2}^*(s, s')E(V_{i_1,1}V_{i_2,2})]^{(i_1-1)(i_2-1)}, \end{aligned} \quad (4.6)$$

where

$$\begin{aligned} c_{1,k}^*(s) &= \iint K_{i,k}(s)p(\phi_{i,k})p(\epsilon_{i,k}|\phi_{i,k})d\phi_{i,k}d\epsilon_{i,k}, \\ c_2^*(s, s') &= \iiint K_{i,1}(s)K_{i,2}(s')p(\phi_{i,1})p(\epsilon_{i,1}|\phi_{i,1})d\phi_{i,1}d\epsilon_{i,1} \\ &\quad p(\phi_{i,2})p(\epsilon_{i,2}|\phi_{i,2})d\phi_{i,2}d\epsilon_{i,2}, \\ c_{i_1, i_2}^*(s, s') &= \iiint K_{i_1,1}(s)K_{i_2,2}(s')C_{1,2}(\phi_{i_1}, \phi_{i_2})p(\phi_{i_1,1})p(\epsilon_{i_1,1}|\phi_{i_1,1})d\phi_{i_1,1}d\epsilon_{i_1,1} \\ &\quad p(\phi_{i_2,2})p(\epsilon_{i_2,2}|\phi_{i_2,2})d\phi_{i_2,2}d\epsilon_{i_2,2}. \end{aligned}$$

The marginal covariance in (4.6) allows for a lack of stationarity in the sign of the spatial cross-dependency structure; this is induced by the change of

sign in the cross-covariance $C_{1,2}$ of the underlying process X . The dependency structure in (4.6) is in general nonseparable, in the sense that we cannot separate the dependence between Y_1 and Y_2 and the spatial dependence between locations s and s' .

4.6. Weak convergence

In this section we study the weak convergence for a spatial process $Y(s)$ with process prior $F_s(Y)$ as in (2.1), with the purpose of understanding if $F_{s_1}(Y)$ approximates $F_{s_2}(Y)$ when s_1 is close to s_2 .

Theorem 3. *Let $Y(s)$ be a random field, with random distribution given by $F_s(Y)$ as in (2.1). If the probability masses $p_i(s)$ in (2.1) are almost surely continuous, then $Y(s_1)$ converges weakly to $Y(s_2)$ with probability one as $|s_1 - s_2| \rightarrow 0$.*

If the kernel functions $K_i(s)$ as (2.1) are continuous functions, then the probability masses $p_i(s)$ are almost surely continuous and $F_{s_1}(Y)$ tends to $F_{s_2}(Y)$ as $|s_1 - s_2| \rightarrow 0$, Theorem 3. From Theorem 3, we do not need X to be almost surely continuous to have weak convergence for Y . However, to have almost sure continuity for Y , we need X to be almost surely continuous, Theorem 4.

5. Continuous Realizations for the Limiting Process

Unlike other nonparametric spatial models in the literature (Gelfand et al. (2004)), our representation does not require repeated measures at the observation locations, because the latent process specification has a continuous realization in the limit. If the underlying surface of interest is continuous and smooth, this continuity is not only desirable but it also offers computational benefits in relation to spatially discontinuous SB priors (e.g., Reich and Fuentes (2007)). For instance, a discontinuous spatial SB would need several mixture components to characterize a spatial Gaussian process, while we only need one component with the model introduced here.

Theorem 4. *Let $Y(s)$ be a random field given through $F_s(Y)$ as (2.1). If the underlying stationary process is almost surely (a.s.) continuous in s for every $s \in \mathcal{R}^2$, then as the bandwidth parameters go uniformly to zero and the knots become more dense, the process Y has a.s. continuous realizations.*

Kent (1989) shows that if the covariance of X admits a second order Taylor-series expansion with remainder that goes to zero at a rate of $2 + \delta$ for some $\delta > 0$, then X is a.s. continuous.

6. Computing Methods

6.1. Truncation

It is useful in practice to consider finite approximations to an infinite stick-breaking process. We focus on a truncation approximation to (2.1),

$$\sum_{i=1}^N p_i(s) \delta(X(\phi_i)) + (1 - \sum_{i=1}^N p_i(s)) \delta(X(\phi_0)),$$

that results in a distribution $F_{s,N}$ such that $F_{s,N} \rightarrow F_s$, with F_s the distribution of the stick-breaking process at location s . If $p_0(s)$ denotes the probability mass on $X(\phi_0)$, we have, a.s., $\sum_{i=0}^N p_i(s) = 1$ for all s . The proof of this result is a straightforward extension of Theorem 3 in Dunson and Park (2008).

Papaspiliopoulos and Roberts (2008) introduce an elegant computational approach to working with an infinite mixture for Dirichlet processes mixing. However, their approach would not be efficient in our setting, because it relies on Markovian properties of all parameters, and the spatial varying parameters in our model (e.g. bandwidths) are not (discrete) conditionally autoregressive spatial processes, but rather continuous spatial processes.

6.2. MCMC details

For computational purposes, let $g(\mathbf{s}) \sim \text{Categorical}(p_1(s), \dots, p_M(s))$ to indicate $Z(\mathbf{s})$'s mixture component, so that $Z(\mathbf{s})|g(\mathbf{s})$ is normal with mean $X(\phi_{g(\mathbf{s})})$ and diagonal covariance Σ_e , the covariance of the nugget effect. We write $X_j(\phi_i) = \sum_k a_{kj}(\phi_i) U_k(\phi_i)$, where U_1, \dots, U_p are independent spatial processes with mean zero, $\text{cov}(U_j(\phi_1), \dots, U_j(\phi_M)) = \Omega_j$, with the (u, v) element of Ω_j as $\rho_j(\phi_u, \phi_v)$. Using this parameterization, $g(\mathbf{s}_l)$, $U_k(\phi_i)$, and $a_{kk'}(\phi_i)$ have conjugate full conditional posteriors and are updated using Gibbs sampling. The full conditionals for $g(\mathbf{s})$ and $U_k(\phi_i)$ are

$$P(g(\mathbf{s}) = m) = \frac{p_m(\mathbf{s}) \Phi(X(\phi_m), \Sigma_e)}{\sum_{i=1}^M p_i(\mathbf{s}) \Phi(X(\phi_i), \Sigma_e)}, \quad (6.1)$$

where $\Phi(X(\phi_m), \Sigma_e)$ is the multivariate normal density with mean $X(\phi_m)$ and covariance Σ_e , $U_k(\phi_i)$'s full conditional is normal with

$$\begin{aligned} \text{Var}[U_k(\phi_i)|\text{rest}]^{-1} &= \{\Omega_k^{-1}\}_{ii} + \sum_{l=1}^n \sum_{k'=1}^p I(g(\mathbf{s}_l) = i) \left(\frac{a_{kk'}}{\sigma_{k'}} \right)^2 \\ E[U_k(\phi_i)|\text{rest}] &= \text{Var}[U_k(\phi_i)|\text{rest}] \\ &\quad \left[- \sum_{l \neq i} \{\Omega_k^{-1}\}_{li} U_k(\phi_l) + \sum_{l=1}^n \sum_{k'=1}^p I(g(\mathbf{s}_l) = i) \frac{r_{k'}(\mathbf{s}_l) a_{kk'}}{\sigma_{k'}^2} \right], \end{aligned} \quad (6.2)$$

and $r_{k'}(\mathbf{s}_l) = Y_{k'}(\mathbf{s}_l) - \sum_{k \neq k'} a_{kk'}(\phi_i) U_k(\phi_i)$. The full conditional for $a_{kk'}(\phi_i)$ is nearly identical to the full conditional of $U_k(\phi_i)$ and is not given here.

We sampled the nugget variances from their conjugate inverse gamma full conditionals. Conjugacy does not hold in general for the stick-breaking parameters V_i , B_i , or ϕ_i , or the spatial range parameters; these parameters are updated using Metropolis sampling with Gaussian candidate draws. For the stick-breaking parameters V_i at iteration t , we generate the candidate value $V_i^* \sim N(V_i^{(t-1)}, c)$, where $V_i^{(t-1)}$ is the value from the previous iteration and c is a tuning parameter. Then $V_i^{(t)}$ is set to V_i^* with probability $\min\{1, R\}$, and to $V_i^{(t-1)}$ otherwise. To compute the acceptance ratio R , we calculate the stick-breaking probabilities using all the current values of the parameters, denoted $p_m^{(t-1)}(\mathbf{s})$, and using all the current values with $V_i^{(t-1)}$ replaced by the candidate V_i^* , denoted $p_m^*(\mathbf{s})$. The acceptance ratio is then

$$R = \left[\prod_{l=1}^n \frac{p_{g(\mathbf{s}_l)}^*(\mathbf{s}_l)}{p_{g(\mathbf{s}_l)}^{(t-1)}(\mathbf{s}_l)} \right] \frac{Be(V_i^*; a, b)}{Be(V_i^{(t-1)}; a, b)},$$

where $Be(V; a, b)$ is the beta prior density. The remaining stick-breaking parameters B_i and ϕ_i are updated similarly. Candidates with no posterior mass are rejected and the candidate standard deviations are tuned to give acceptance rate near 0.4. We draw 25,000 MCMC samples and discard the first 5,000 as burn-in. Convergence is monitored using trace plots of the deviance as well as several representative parameters.

The number of sample points, n , is often used in spatial methods to describe the computational burden of the model because evaluating the Gaussian likelihood requires inverting an $n \times n$ covariance matrix, which is challenging for large n . In the proposed mixture model, most of the computational time is spent evaluating spatial processes defined at the M knots. Therefore, M may be a more relevant summary of computational complexity. For smooth processes, $M < n$ knots may be sufficient, resulting in a dimension reduction and improved computing for large n . However, for processes with small spatial range, the model may require $M = n$ or even $M > n$ knots, which is infeasible for large data sets. A good starting value for M is the expected effective number of observations. We easily determine if M is large enough by evaluating the posterior distribution of the mixture probabilities and determining if p_M , the residual mass not accounted by the other $M - 1$ mixture probabilities, is small enough. The stick-breaking prior process does not need to be an infinite mixture and, in many cases, a small number of components can explain well the underlining spatial structure.

7. Simulation Study

We generated bivariate spatial data on a 15×15 regular grid of spatial locations spanning $[0, 1]^2$. With $(y_1(\mathbf{s}), y_2(\mathbf{s}))'$ the observation at location $\mathbf{s} \in [0, 1]^2$, generated the spatial data by first drawing two independent bivariate Gaussian processes $(z_1^{(1)}(\mathbf{s}), z_2^{(1)}(\mathbf{s}))'$ and $(z_1^{(2)}(\mathbf{s}), z_2^{(2)}(\mathbf{s}))'$ with stationary and separable covariances $\text{Cov}[(z_1^{(j)}(\mathbf{s}), z_2^{(j)}(\mathbf{s}))', (z_1^{(j)}(t), z_2^{(j)}(t)')] = \exp(-\|\mathbf{s} - t\|/\rho_j) \Sigma_j$, where $\rho_1 = 0.05$, $\rho_2 = 0.25$, $\Sigma_1 = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$, and $\Sigma_2 = \begin{pmatrix} 1 & -0.8 \\ -0.8 & 1 \end{pmatrix}$. These stationary Gaussian processes were used to generate non-stationary and/or non-Gaussian processes in the following ways.

1. Stationary, Gaussian: $\mu_k(\mathbf{s}) = z_k^{(1)}(\mathbf{s})$.
2. Non-stationary, Gaussian: $\mu_k(\mathbf{s}) = \sqrt{w(\mathbf{s})}z_k^{(1)}(\mathbf{s}) + [1 - \sqrt{w(\mathbf{s})}]z_k^{(2)}(\mathbf{s})$.
3. Stationary, Non-Gaussian: $\mu_k(\mathbf{s}) = f[z_k^{(1)}(\mathbf{s})]$.
4. Non-stationary, Non-Gaussian: $\mu_k(\mathbf{s}) = f[\sqrt{w(\mathbf{s})}z_k^{(1)}(\mathbf{s}) + [1 - \sqrt{w(\mathbf{s})}]z_k^{(2)}(\mathbf{s})]$.

Here $w(\mathbf{s}) = \exp(-10 [(s_1 - 0.2)^2 + (s_2 - 0.2)^2])$, $f(c) = G^{-1}(\Phi(c))$, Φ is the standard normal CDF, and G is the Gamma(2,2) CDF. The second design is a linear combination of two Gaussian processes, with weight for the first component high in the lower left corner near $\mathbf{s} = (0.2, 0.2)'$, and low in the upper right corner. Therefore, the spatial range is small (following ρ_1) and the cross-correlation is positive (following Σ_1) in the lower left; the spatial range is large (following ρ_2) and the cross-correlation is negative (following Σ_2) in the lower left. The third design is stationary, but the function f transforms the marginal distribution from standard normal to the right-skewed Gamma(2,2). The final design is non-stationary and non-Gaussian.

For each design we generated $K = 25$ data sets by first drawing μ as described above, and then adding error with mean 0 and standard deviation 0.5 at each location and each response type. All observations for the first response y_1 were retained, while the second response y_2 was eliminated for a randomly selected subset of 20% of the spatial locations. Table 1 reports the mean squared error between the true and posterior mean μ for the second response for each model, separately for locations with “complete data” and with “partial data”.

We fit four models to each data set. The first model was the stationary, Gaussian model $y_j(s) \sim N(\mu_j(s), \sigma_j^2)$, where $\mu(s)$ was a bivariate Gaussian process with separable covariance $\text{Cov}[(\mu_1(\mathbf{s}), \mu_2(\mathbf{s}))', (\mu_1(t), \mu_2(t)')] = \exp(-\|\mathbf{s} - t\|/\rho) \Sigma$. For priors we took $\sigma_j^2 \sim \text{InvGamma}(0.1, 0.1)$, $\Sigma \sim \text{Wishart}(2.1, 0.1I_2)$, and $\rho \sim \text{Unif}(0, 1)$. The prior distribution for the variance parameter σ_0^2 of the measurement error process was $\text{InvGamma}(0.1, 0.1)$. The second model was also Gaussian

with a separable covariance, but we allowed for nonstationarity using the kernel convolution model for $\mu_j(\mathbf{s})$ (Higdon, Swall, and Kern (1999)):

$$\mu_j(\mathbf{s}) = \sum_{k=1}^M w_{jk}(\|\mathbf{s} - \phi_k\|)\theta_k, \quad (6.3)$$

where $\theta_k \stackrel{i.i.d.}{\sim} N(0, 1)$, w_{jk} as the kernel function described below, and the $\phi_k \in \mathcal{R}^2$ were fixed spatial knots. We chose the kernel function to approximate the Matérn covariance

$$\begin{aligned} & \text{Cov} [(\mu_1(\mathbf{s}), \mu_2(\mathbf{s}))', (\mu_1(\mathbf{s} + h), \mu_2(\mathbf{s} + h))'] \\ &= \frac{\tau^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\nu^{1/2}\|h\|}{\rho} \right)^\nu \mathcal{K}_\nu \left(\frac{2\nu^{1/2}\|h\|}{\rho} \right), \end{aligned} \quad (6.4)$$

where \mathcal{K} is the modified Bessel function, $\tau^2 > 0$ controls the variance, $\rho > 0$ controls the spatial range of the correlation, and $\nu > 0$ controls the smoothness of the process. The Matérn kernel function with spatially-varying range is

$$w_{jk}(u) = \tau_j^2 \frac{\gamma(\nu_j + 1)^{1/2} \nu_j^{\nu_j/4+1/4} |u|^{\nu_j/2-1/2}}{\pi^{1/2} \Gamma(\nu_j/2 + 1/2) \Gamma(\nu_j)^{1/2} \rho_{jk}^{\nu_j/2+1/2}} \mathcal{K}_{\nu_j/2+1/2} \left(\frac{2\nu_j^{1/2}|u|}{\rho_{jk}} \right). \quad (6.5)$$

We took $\log(\rho_{jk})$ to be Gaussian with mean $\bar{\rho}_j$ and covariance $\text{Cov}(\log(\rho_{jk}), \log(\rho_{jl})) = \delta_j^2 \exp(-\|\phi_k - \phi_l\|/\eta_j)$, independent across response type j . For priors we took $\nu_j \sim \text{Unif}(0,10)$, $\tau_j^{-2}, \delta_j \sim \text{Gamma}(0.01, 0.01)$, $\bar{\rho}_j \sim N(-2, 1)$, and $\eta_j \sim \text{Unif}(0, 1)$. Note the prior 95% interval for $\exp(\bar{\rho}_j)$ was roughly $(0, 1)$, comparable to that of for the spatial range in Model 1. The knots had a spatial uniform prior.

In addition to the Gaussian model, we fit two versions of the spatial stick-breaking model. Both models used uniform kernels. The first was the stationary model of Reich and Fuentes (2007): the model in Section 2 with constant bandwidths and cross-covariance, and the with i.i.d. X . The second was the full model in Section 3 that allows not only X , but also the bandwidths and cross-correlations to be spatial continuous processes. The knots had a spatial discrete uniform prior. The number of mixture components was fixed at $M = 225$ for all mixture models.

Overall, the proposed model outperformed its competitors, bringing a significant reduction in the MSE. For complete data locations, in the stationary Gaussian case, all model performed similarly, except for the Reich and Fuentes (2007) SB prior that slightly underperformed other competitors. In the fourth setting, using our proposed nonparametric model, there was a reduction of 41%

Table 1. Mean squared error (standard error) for the simulation study.

(a) Complete data locations

Simulation Design		Model			
Stationary	Gaussian	Stationary Gaussian	Non-stationary Gaussian	Stationary non-Gaussian	Non-stationary non-Gaussian
Yes	Yes	0.46 (0.01)	0.46 (0.01)	0.47 (0.01)	0.46 (0.01)
No	Yes	0.53 (0.03)	0.78 (0.05)	0.45 (0.01)	0.45 (0.01)
Yes	No	0.70 (0.01)	0.39 (0.01)	0.46 (0.02)	0.39 (0.01)
No	No	0.61 (0.02)	0.49 (0.03)	0.50 (0.02)	0.36 (0.01)

(b) Partial data locations

Simulation Design		Model			
Stationary	Gaussian	Stationary Gaussian	Non-stationary Gaussian	Stationary non-Gaussian	Non-stationary non-Gaussian
Yes	Yes	0.70 (0.02)	0.75 (0.03)	0.72 (0.02)	0.62 (0.03)
No	Yes	0.71 (0.02)	0.83 (0.03)	0.57 (0.02)	0.56 (0.03)
Yes	No	0.71 (0.02)	0.71 (0.02)	0.57 (0.02)	0.57 (0.03)
No	No	0.72 (0.02)	0.70 (0.02)	0.57 (0.02)	0.52 (0.02)

in the MSE for complete data locations with respect to the standard stationary Gaussian process, a reduction of 27% when compared to the kernel convolution nonstationary Gaussian model, and a reduction of 28% when compared to the spatial SB prior of Reich and Fuentes (2007). Similarly, in the second setting, there is a reduction of 15% in the MSE when compared to the spatial SB prior of Reich and Fuentes (2007).

The proposed model also outperformed all other models at partial data locations. Our nonparametric model is capable of characterizing complex nonstationary, and/or non-Gaussian data structures, as well as simpler stationary surfaces, and it seems to outperform the alternatives in the literature.

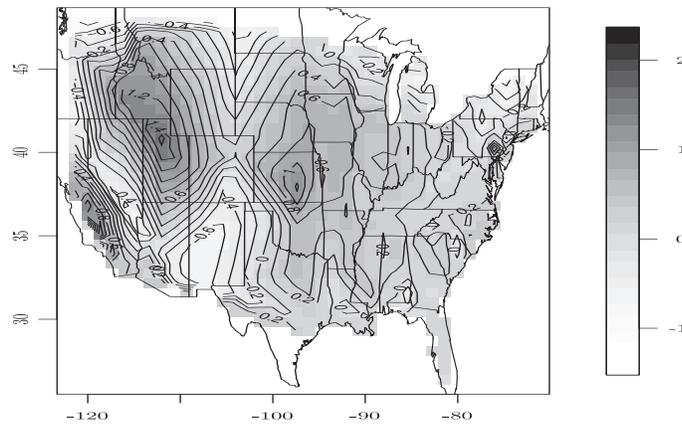
To evaluate how the proposed method captures the cross-dependence between data processes, in Figure 2 we compare the true cross covariance for the simulation design with Gaussian nonstationary data, with the average of the posterior mean of the cross-covariance for the full model. The similar patterns in Figure 2 indicate that our method can reproduce cross-dependence structure; similar results were obtained for other designs.

8. Analysis of Nitrate and Ammonium

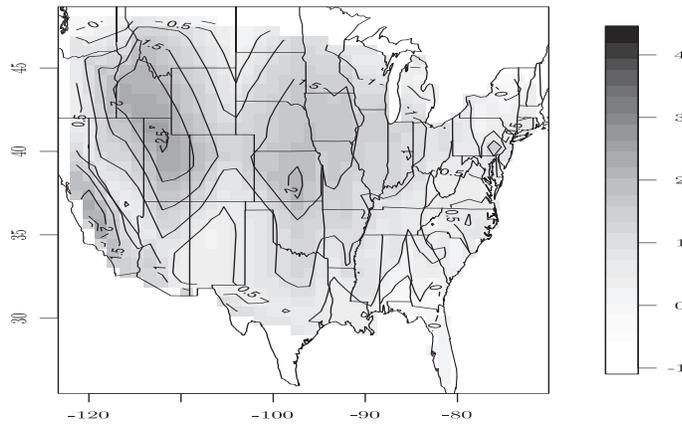
Nitrate and ammonium are two of the greatest contributors to the $PM_{2.5}$ mass, and studies conducted separately for each pollutant have found significant association with adverse health outcomes (e.g., Hughes et al. (2002), Batalha et al. (2002), and Clarke et al. (2000)). To illustrate the multivariate model



(a) Monitoring data locations



(b) Ammonium



(c) nitrate

Figure 1. Comparison between the true cross-covariance structure and the posterior mean of the cross-covariance using the proposed method.

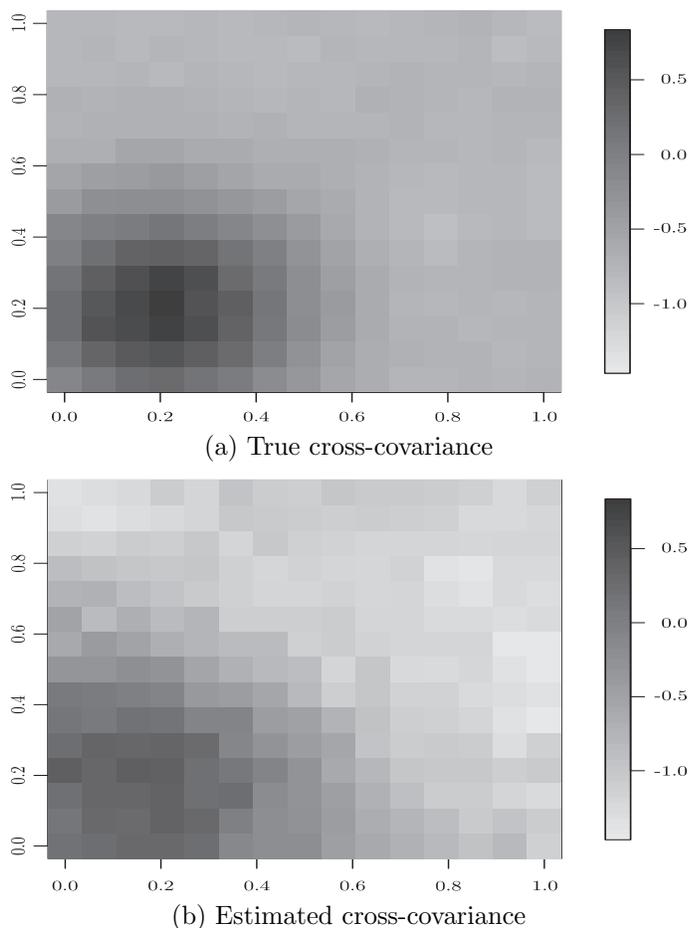


Figure 2. Maps of monitoring data locations and log-transformed ammonium and nitrate data.

proposed here, we analyze monthly average values of ammonium and nitrate for January, 2007 at 209 monitoring stations in the US. These data were obtained from the US EPA <http://www.epa.gov/airexplorer/index.htm>. The data are plotted in Figure 1.

Spatial locations were projected to a flat surface using the Mercator projection and re-scaled to $[0, 1]^2$. The data are right-skewed, so we used a log transformation for each pollutant and, for each pollutant, we removed a second-order mean trend with linear and quadratic effects for latitude and longitude, and the interaction between latitude and longitude, to account for large-scale spatial variation.

We compared the four models described in Section 7 using five-fold cross validation. For the stick-breaking models we found $M = 225$ terms sufficient to

Table 2. Five-fold cross validation root mean squared error and coverage probability for ammonium and nitrate.

(a) Root mean squared error

Stationary	Gaussian	Ammonium	nitrate
Yes	Yes	0.253	0.268
No	Yes	0.302	0.316
Yes	No	0.239	0.263
No	No	0.227	0.259

(b) Coverage probability of 90% predictive intervals

Stationary	Gaussian	Ammonium	nitrate
Yes	Yes	0.920	0.938
No	Yes	0.951	0.966
Yes	No	0.928	0.956
No	No	0.941	0.938

approximate the infinite dimensional model. For all models, the priors were those in Section 7, using different space-dependent uniform kernels for each pollutant by sharing knots and bandwidth. Observations were allocated to the five training data sets randomly over space and species. The mean root squared prediction (RMSE) error and coverage probabilities are given in Table 2. Our proposed multivariate nonstationary nonGaussian model reduces the RMSE by 24% for ammonium and 18% for nitrate when compared to the kernel convolution nonstationary Gaussian approach. It also outperforms the Gaussian model, and the nonparametric approach of Reich and Fuentes (2007), though, the gain is more moderate in that case. Our model reaches the nominal coverage probability, this did not seem to be a problem for any of the models.

There is also considerable variation in the bandwidths in Figure 3, where we plot their posterior mean. The largest bandwidths are in the Midwest and Southwest, where both pollutants show smoother spatial patterns.

Convergence was monitored using trace plots of the deviance and several parameters. The deviance converged after 1,000 iterations, which was the burn-in used in these simulations.

Finally, we conducted a sensitivity analysis to various modeling assumptions. We refit the model, increasing the number of terms in the mixture model to $M = 400$, and with squared-exponential kernels. Figure 4 plots the fitted values in the logarithmic scale (mean of the posterior predictive distribution). The fitted values are similar for the different kernel types and when we increase the number of terms in the mixture model. We also looked at sensitivity with respect to prior distributions for model parameters. We include here results for the

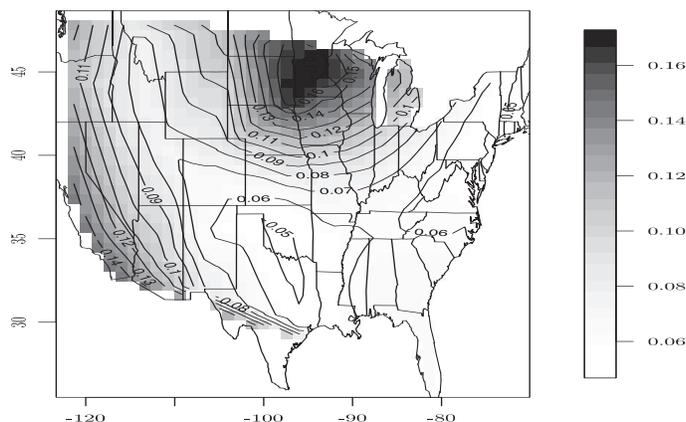


Figure 3. Posterior mean of the bandwidth parameter.

range parameter (one of the most difficult parameters to estimate). We refit the entire model using a log-normal distribution prior for the range parameters with location parameter 0 and squared scale $\text{LogNormal}(0,1)$, rather than uniform priors. The correlation between the fitted values from both models with the different prior distributions, and between the residuals after removing the spatial trend, varied between 0.997 and 1 for both pollutants. We also studied sensitivity to the distributions of knots; results not shown here indicate that the results were robust to their prior distribution. In summary, the results appear to be robust to the choice of kernel and prior distributions, as well to varying the number of mixture components and the prior distribution for knot selection.

9. Discussion

In this paper we introduce a nonparametric framework for modelling multivariate spatial processes that is flexible enough to admit complex nonstationary dependency and cross-dependency structures without specifying a Gaussian distribution. One advantage of our approach is that is computationally efficient and therefore useful in situations where several complex spatial processes need to be modelled simultaneously.

The univariate kernel stick breaking version provides an alternative to recently developed spatial stick-breaking processes (Griffin and Steel (2006); Dunson and Park (2008); Gelfand, Guindani, and Petrone (2007); Reich and Fuentes (2007)). The univariate version has a continuous limiting process, rather than discrete as in the other approaches, and it allows for non-stationarity.

Here many of the tools developed for Dirichlet processes can be applied with some modifications, and that allows us to study statistical properties. The application in the paper is to a multivariate spatial process, but our framework

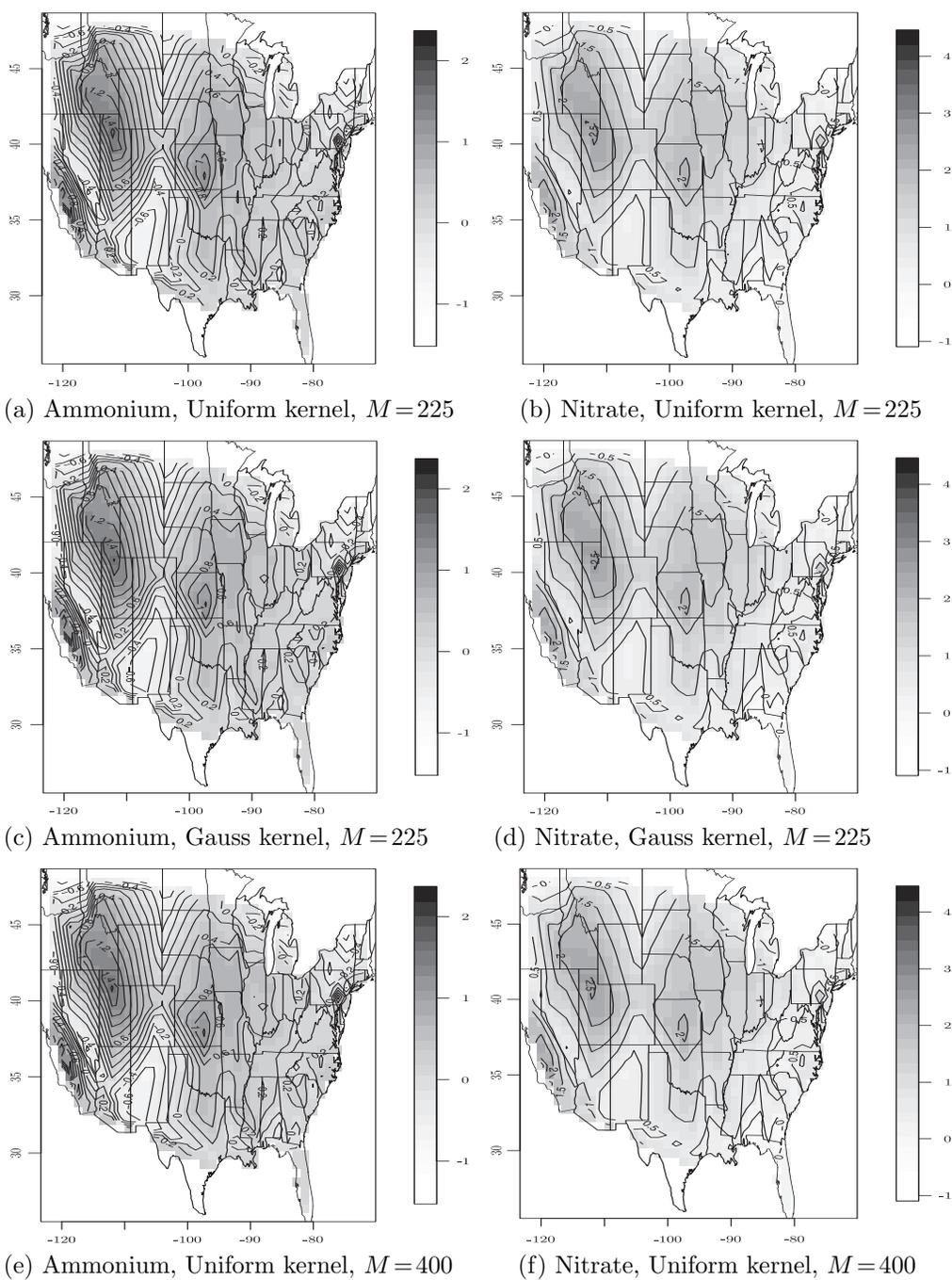


Figure 4. Posterior mean (in log scale) for models with different kernels (Uniform or Gaussian) and different number of terms in the mixture ($M = 225$ or $M = 400$).

can be applied to multivariate spatial-temporal processes by using space-time kernels. We plan to carry out this extension.

Acknowledgement

The authors thank the National Science Foundation (Reich, DMS-0354189; Fuentes DMS-0706731, DMS-0353029), the Environmental Protection Agency (Fuentes, R833863), and National Institutes of Health (Fuentes, 5R01ES014843-02) for partial support of this work.

References

- An, Q., Wang, C., Shterev, I., Wang, E., Carin, L. and Dunson, D. (2008). Hierarchical kernel stick-breaking process for multi-task image analysis. *International Conference on Machine Learning (ICML)*, to appear.
- Batalha, J. R. F., Saldiva, P. H. N., Clarke, R. W., Coull, B. A., Stearns, R. C., Lawrence, J., Murthy G. G. K., Koutrakis, P. and Godleski, J. J. (2002). Concentrated ambient air particles induce vasoconstriction of small pulmonary arteries in rats. *Environmental Health Perspectives* **110**, 1191-1197. doi: 10.1289/ehp.5484.
- Choi, J., Reich, B. J., Fuentes, M. and Davis, J. M. (2009). Multivariate spatial-temporal modeling and prediction of speciated fine particles. *Journal of Statistical Theory and Practice* **3**, 407-418.
- Clarke, R. W., Coull, B. A., Reinisch, U., Catalano, P., Killingsworth, C. R., Koutrakis, P., Kavouras, I., Murthy, G. G. K., Lawrence, J., Lovett, E., Wolfson, J. M., Verrie, R. L. and Godleski, J. J. (2000). Inhaled concentrated ambient particles are associated with hematologic and bronchoalveolar lavage changes in canines. *Environmental Health Perspectives* **108**, 1179-1187.
- Dominici, F., Daniels, M., Zeger, S. L and Samet, J. M. (2002) Air pollution and mortality: estimating regional and national dose-response relationships. *J. Amer. Statist. Assoc.* **97**, 100-111.
- Dunson, D. B. and Park, J. H. (2008). Kernel stick-breaking processes. *Biometrika* **95**, 307-323.
- Ferguson, T. (1973). Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230.
- Gelfand, A. E., Guindani, M. and Petrone, S. (2007). Bayesian nonparametric modeling for spatial data analysis using Dirichlet processes. In *Bayesian Statistics 8* (Edited by J. Bernardo et al.).
- Gelfand, A. E., Kottas, A. and MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *J. Amer. Statist. Assoc.* **100**, 1021-1035.
- Gelfand, A. E., Schmidt, A. M., Banerjee, S. and Sirmans, C.F. (2004). Nonstationary multivariate process modelling through spatially varying coregionalization (with discussion). *Test* **13**, 2, 1-50.
- Griffin, J. E. and Steel, M. F. J. (2006). Order-based dependent Dirichlet processes. *J. Amer. Statist. Assoc.* **101**, 179-194.
- Higdon, D., Swall, J. and Kern, J. (1999). Non-stationary spatial modeling. In *Bayesian Statistics 6* (Edited by Bernardo, Berger, David, and Smith), 761-768. Oxford University Press.

- Hughes, L. S., Allen, J. O., Salmon, L. G., Mayo, P. R., Johnson, R. J. and Cass, G. R. (2002). Evolution of nitrogen species air pollutants along trajectories crossing the Los Angeles area. *Environmental Science and Technology* **36**, 3928-3935.
- Kent, J. T. (1989). Continuity properties for random fields, *Ann. Probab.* **17**, 1432-1440.
- Laud, P. and Ibrahim, J. (1995). Predictive model selection. *J. Roy. Statist. Soc. Ser. B* **57**, 247-262.
- Matérn, B. (1960). *Spatial Variation*. Almaenna Foerlaget, Stockholm.
- Palacios, M. B. and Steel, M. F. J. (2006). Non-Gaussian Bayesian geostatistical modelling. *J. Amer. Statist. Assoc.* **101**, 604-618.
- Papaspiliopoulos, O. and Roberts, G. (2008). Retrospective MCMC for Dirichlet process hierarchical models. *Biometrika* **95**, 169-186.
- Petrone, S., Guindani, M. and Gelfand, A. E. (2008). Hybrid Dirichlet mixture models for functional data. Tech. report, Statistics Department at Duke University.
- Reich, B. J. and Fuentes, M. (2007). A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *Ann. Appl. Statist.* **1**, 249-264.
- Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.* **12**, 1151-1172.
- Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4**, 639-650.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *J. Roy. Statist. Soc. Ser. B* **64**, 583-639.
- Song, H. R., Fuentes, M. and Ghosh, S. (2008). A comparative study of Gaussian geostatistical models and Gaussian Markov random field models. *J. Multivariate Anal.* **99**, 1681-1697.

Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, NC 27695-8203, USA.

E-mail: fuentes@stat.ncsu.edu

Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, NC 27695-8203, USA.

E-mail: brian_reich@ncsu.edu

(Received July 2011; accepted November 2011)