Published in final edited form as:

Stat Sin. 2013 January ; 23(1): . doi:10.5705/ss.2011.172.

Multivariate spatial nonparametric modelling via kernel processes mixing*

Montserrat Fuentes and

Professor of Statistics at North Carolina State University (NCSU). Tel.: (919) 515-1921, Fax: (919) 515-1169

Brian Reich

Assistant Professor of Statistics at NCSU

Montserrat Fuentes: fuentes@stat.ncsu.edu

SUMMARY

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), which 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 non-stationary 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, which often hinders the spatial analysis of large data sets. 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.

Keywords

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 tail and extremes of multivariate distributions, and it is computationally efficient.

*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.

Our nonparametric model extends the stick-breaking (SB) prior of Sethuraman (1994), which 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 , $X_i \stackrel{iid}{\sim} F_o$, and F_o is a known distribution. A special case of this prior is the Dirichlet process prior with infinite M and $V_i \stackrel{iid}{\sim} \text{Beta}(1, v)$ (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 et al. (2005a), 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 kernels functions in our proposed methodology impose in a very natural form a 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 for the first time an extension of the stick-breaking prior to the multivariate spatial setting allowing for nonseparability and nonstationarity in the spatial cross-dependency between the outcomes of interest. The multivariate aspect as well as the nonstationary approach are the main contributions of the work presented here. We assign at each location a different, unknown distribution, and we smoothing the distributions in space with space-dependent kernel functions, allowing then for nonstationarity. The approach we introduce here is the first to allow both the probabilities p_i and the values X_i to depend on space, and that way 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 it avoids inverting large matrices and calculating determinants.

We apply the methods in this paper to model and characterize the complex spatial structure of air pollution, in particular to components of fine particulate matter. Fine particulate matter

(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. PM_{2.5} has over 40 components and two of the main components 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 PM_{2.5} chemistry changes with space and time so its association with the health endpoints could change across space and time. Speciated 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 an innovative multivariate spatial model for chemical components of PM_{2.5} across the entire United States. We implement this model to two of the PM_{2.5} greatest components, 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 new 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, which 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 the computing methods and MCMC algorithms. In Section 7, we illustrate the proposed methods with a simulation study. In Section 8, we present an application to air pollution. We conclude with Section 9 that has some remarks and final comments.

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, i.e., $Y(s) \sim F_s(Y)$. 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 will assume $d = 2$.

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

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

where X is a Gaussian process (GP) with covariance Σ_X , that has diagonal elements σ_i^2 . $p_1(s) = V_1(s)$, $p_i(s) = V_i(s) \prod_{j=1}^{i-1} (1 - V_j(s))$, and $V_i(s) = K_i(s) V_i$, $V_i \sim \text{Beta}(a, b)$ independent over i , we have $\sum_i p_i(s) = 1$ for all s (by Appendix A.1). The weight function, K_i , is a spatial kernel centered at knot $\varphi_i \in \mathbb{R}^2$ with bandwidth parameter ε_i . More generally, in our application we fit elliptical kernels K_i , with B_i being the 2×2 matrix that controls the shape of the ellipse. We represent $B_i = T(\varphi_i) T(\varphi_i)'$, where $T(\varphi_i)'$ denotes the transpose of $T(\varphi_i)$, and $T_{kk}(\varphi_i)$ is the (k, k) element of the matrix $T(\varphi_i)$. 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, $B_i = \varepsilon_i I$, 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 also 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),

$$\text{cov}(X(\phi_i), X(\phi_j)) = \sigma_i \sigma_j \rho(|\phi_i - \phi_j|), \quad (2)$$

where $\sigma_i = \sigma(\phi_i)$ is the variance of the process $X(\phi_i)$ and it is space-dependent, and ρ is a correlation function (e.g. a Matérn), $|\phi_i - \phi_j|$ 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. The spatial uniform is the prior chosen here for the knots. 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 et al. (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 to be space-dependent but only the angles, and the only type of kernels considered were squared exponential. In this work apart from generalizing the type of kernel convolution presented by Higdon et al. (1999) with our space-dependent kernel functions, the stick-breaking structure adds to the kernel convolution by allowing for lack of normality, and introducing the flexibility of having the data to determine the locations of the most relevant knots, i.e. the knots with large probabilities $p_i(s)$.

Model (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 (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 Appendix A.1 we prove that the representation in (1) introduces a properly defined process prior, using the Kolmogorov existence theorem. In Section 4 we study the marginal and conditional properties of model (1), for a stationary process X .

3 Multivariate spatial model

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

$$F_s(Y_k) = \sum_{i=1}^M V_{i,k}(s) \prod_{j < i} [1 - V_{j,k}(s)] \delta(X_k(\phi_i)), \quad (3)$$

where, $p_{1,k}(s) = V_{1,k}(s)$, $p_{i,k}(s) = V_{i,k}(s) \prod_{j=1}^{i-1} (1 - V_{j,k}(s))$, and $V_{i,k}(s) = K_{i,k}(s)V_i$, $V_i \sim \text{Beta}(a, b)$ 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 covariance $\Sigma^{(i)}$ that depends on space via the knot locations. By allowing $\Sigma^{(i)}$ to depend on i , and therefore changing with location, 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 model separately 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, thus, it could be negative in some areas and positive in others. We call this phenomenon “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, Reich, Fuentes, and Davis, 2009).

As an example, a simpler representation of the nonparametric multivariate spatial model in (3), would be obtained by sharing the univariate underlying process X across the p spatial processes, i.e. having the following representation for the distribution of each $Y_k(s)$,

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

where the V and K random weights are different for each process Y_k . This allows the spatial processes Y_k , with $k = 1, \dots, p$, 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 would not allow for non-stationarity in the sign. The assumption of sharing the underlying process X , in the context of speciated particulate matter, may be justified by having the different components of $\text{PM}_{2.5}$ sharing the same pollution sources.

4 Asymptotic dependence and weak convergence

Other nonparametric continuous spatial representations in the literature require repeated measurements. Then, we need to study the theoretical properties of the proposed representation, to justify the propriety of this prior, the nonstationarity of the covariance, and then to be able to understand the necessary conditions on the kernels and X process for all this to hold without replications.

In this section we first study the properties of the spatial dependence induced by our proposed mixture model. Initially, conditioning on the mixture weights, in which case as the kernel bandwidths become smaller, only the kernels with knots in the vicinity of the data points would be in the mixture, and one would expect to have a limiting process with similar properties as our base process X . So if X is stationary we would expect a stationary conditional dependence. The asymptotic results presented in this section should help us to understand better the role of the different components in our mixture representation.

Throughout this paper all proofs can be found in the Appendix.

4.1 Conditional univariate spatial covariance

Assume the spatial stick-breaking process prior $F_s(Y)$ in (1) for the data process Y . Throughout this section, without loss of generality, we assume $B_i = \varepsilon_i I$, where I is the 2×2 identity matrix. Conditional on the probability masses $p_i(s)$ in (1) but not on the latent process X , the covariance between two observations is as follows,

$$\begin{aligned} \text{cov}(Y(s), Y(s')) | p(s), p(s'), C &= \sum_i \sigma_i^2 p_i(s) p_i(s') + \sum_{i_1 \neq i_2} p_{i_1}(s) p_{i_2}(s') C(|\varphi_{i_1} - \varphi_{i_2}|) \\ &= \sum_i \sigma_i^2 \left[K_i(s) K_i(s') V_i^2 \prod_{j < i} (1 - ((K_j(s) + K_j(s')) V_j + K_j(s) K_j(s') V_j^2)) \right] \\ &+ \sum_{i_1 \neq i_2} [K_{i_1}(s) K_{i_2}(s') V_{i_1} V_{i_2} C(|\varphi_{i_1} - \varphi_{i_2}|) \prod_{j_1 < i_1, j_2 < i_2} (1 - (K_{j_1}(s) V_{j_1} + K_{j_2}(s') V_{j_2} + K_{j_1}(s) K_{j_2}(s') V_{j_1} V_{j_2}))], \end{aligned} \quad (4)$$

where $p(s) = (p_1(s), p_2(s), \dots)$ denotes the potentially infinite dimensional vector with all the probability masses $p_i(s)$ in the mixture defined in (1), C is the covariance function of X , and $\sigma_i^2 = \text{cov}(X(\varphi_i), X(\varphi_i))$.

The conditional covariance of the data process Y in (4) between any two locations s and s' , 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 interesting result is formally presented in Theorem 1.

Theorem 1—Assuming the spatial stick-breaking process prior $F_s(Y)$ presented in (1) for a data process Y , and conditioning on the probabilities $p_i(s)$ but not the underlying process X . The conditional covariance of the data process Y between any two locations s and s' , $\text{cov}(Y(s), Y(s'))$, approximates the covariance function C of the underlying process X , evaluated at $|s - s'|$, as the bandwidth parameter ε_i of each kernel function K_i in (1) goes uniformly to zero for all i , i.e. $\varepsilon_i < \varepsilon$, for all i , such that $\varepsilon \rightarrow 0$. Assuming that K_i is a kernel with compact support for all i , and that C has a first order derivative, C' , which is a bounded nonnegative function.

4.2 Conditional multivariate spatial covariance

Assuming the multivariate spatial stick-breaking process prior $(F_s(Y_1), \dots, F_s(Y_p))$ in (3) for the data processes $Y_1(s), \dots, Y_p(s)$, and conditioning 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 $\mathbf{X} = (X_1, X_2)$. Then, the conditional cross-covariance between any pair $Y_1(s)$ and $Y_2(s')$, is

$$\begin{aligned} \text{cov}(Y_1(s), Y_2(s')) | p_1(s), p_2(s'), C &= \sum_i p_{i,1}(s) p_{i,2}(s') C_{1,2}(\varphi_i, \varphi_i) \\ &+ \sum_{i_1 \neq i_2} p_{i_1,1}(s) p_{i_2,2}(s') C_{1,2}(\varphi_{i_1}, \varphi_{i_2}) \\ &= \sum_i [K_{i,1}(s) K_{i,2}(s') V_{i,1} V_{i,2} C_{1,2}(\varphi_i, \varphi_i) \prod_{j < i} (1 - K_{j,1}(s) V_{j,1} - K_{j,2}(s') V_{j,2} + K_{j,1}(s) K_{j,2}(s') V_{j,1} V_{j,2})] \\ &+ \sum_{i_1 \neq i_2} [K_{i_1,1}(s) K_{i_2,2}(s') V_{i_1,1} V_{i_2,2} C_{1,2}(\varphi_{i_1}, \varphi_{i_2}) \prod_{j_1 < i_1, j_2 < i_2} (1 - (K_{j_1,1}(s) V_{j_1,1} + K_{j_2,2}(s') V_{j_2,2} + K_{j_1,1}(s) K_{j_2,2}(s') V_{j_1,1} V_{j_2,2}))]. \end{aligned} \quad (5)$$

The conditional cross-covariance between any pair $Y_1(s)$ and $Y_2(s')$ presented in (5) 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 , as the bandwidth

parameters of the kernel functions become smaller. This result is formally presented in Theorem 2.

Theorem 2—Assuming the multivariate spatial stick-breaking process prior $(F_s(Y_1), \dots, F_s(Y_p))$ in (3) for the data processes $Y_1(s), \dots, Y_p(s)$, and conditioning 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 $X = (X_1, X_2)$. Assuming also that $C_{1,2}$, the cross-covariance function $\text{cov}(X_1(s), X_2(s')) = C_{1,2}(s, s')$ of the underlying process $\mathbf{X} = (X_1, X_2)$, has first order partial derivatives that are bounded nonnegative functions, and the kernels functions used in the definition of the probabilities $p_{i,1}(s), p_{i,2}(s')$ for each data process $Y_1(s)$ and $Y_2(s)$ have compact support. Then, the conditional cross-covariance between any pair $Y_1(s)$ and $Y_2(s')$, $\text{cov}(Y_1(s), Y_2(s'))$, approximates the cross-covariance function $C_{1,2}(s, s')$ of the underlying process $\mathbf{X}$, as the bandwidth parameters $\varepsilon_{i,1}$ and $\varepsilon_{i,2}$ of the kernel functions in $F_s(Y_1)$ and $F_s(Y_2)$ go uniformly to zero for all i , i.e. $\varepsilon_{i,k} < \varepsilon$, for all i and for $k = 1, 2$, with $\varepsilon \rightarrow 0$.

4.3 Marginal properties: univariate case

We study the marginal properties of the spatial stick prior representation in (1). We start by assuming that the V_i components in (1) are independent and they share the same prior $V_i \sim \text{Beta}(a, b)$. Then, $E(V_i) = E(V)$, and $E(V_i^2) = E(V^2)$, for all i . We assume $B_i = \varepsilon_i I$, to simplify the presentation of the integrals with respect to the kernel components. We also assume the covariance, C , of the underlying process X has a stationary covariance, such that $C(0) = \sigma^2$, and it is also an integrable function. We consider independent priors for the bandwidth and the knot parameters of the kernel functions K_i . In the following section we will 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 between two observations is

$$\begin{aligned} \text{cov}(Y(s), Y(s')) &= \sigma^2 c_2 E(V^2) \sum_i [1 - 2c_1 E(V) + c_2 E(V^2)]^{i-1} + \sum_{i_1 \neq i_2} c_{1,2} [1 - 2c_1 E(V) + c_2 E(V^2)]^{(i_1-1)(i_2-1)} \\ &= \sigma^2 \frac{\gamma(s, s')}{2(1+b/(a+1)) - \gamma(s, s')} + E(V)^2 \sum_{i_1 \neq i_2} c_{1,2} [1 - 2c_1 E(V) + c_2 E(V^2)]^{(i_1-1)(i_2-1)} \end{aligned} \quad (6)$$

where

$$\begin{aligned} \gamma(s, s') &= \frac{\int \int K_i(s) K_i(s') p(\varphi_i, \varepsilon_i) d\varphi_i d\varepsilon_i}{\int \int K_i(s) p(\varphi_i, \varepsilon_i) d\varphi_i d\varepsilon_i}, \\ c_1 &= \int \int K_i(s) p(\varphi_i, \varepsilon_i) d\varphi_i d\varepsilon_i, \\ c_2 &= \int \int K_i(s) K_i(s') p(\varphi_i, \varepsilon_i) d\varphi_i d\varepsilon_i, \\ c_{1,2} &= \int \int \int K_{i_1}(s) K_{i_2}(s') C(|\varphi_{i_1} - \varphi_{i_2}|) p(\varphi_{i_1}, \varepsilon_{i_1}) d\varphi_{i_1} d\varepsilon_{i_1} p(\varphi_{i_2}, \varepsilon_{i_2}) d\varphi_{i_2} d\varepsilon_{i_2}. \end{aligned}$$

In the expression for the marginal covariance between observations $Y(s)$ and $Y(s')$ in (6), the first term 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 in (6) is due to the spatial dependency of the process X . The first term in (6) is a function of s and s' through the function $\gamma(s, s')$. The function $\gamma(s, s')$ is a stationary function, i.e. $\gamma(s, s') = \gamma_0(s - s')$ when the kernel functions are the same across space, rather than being space-dependent functions.

For example, under an squared exponential kernel function with bandwidths equal to ε . Since $\gamma(s, s')$ is the familiar squared exponential correlation function with *range* parameter $\sqrt{2\varepsilon}$, we have $\gamma(s, s') = \gamma_0(h) = 0.5 \exp(-h\sqrt{h/2\varepsilon})$, for $s - s' = h$. To obtain these results we assume there is a buffer zone to avoid the edge effect problem. In expression (6), $c_{1,2}(s, s')$ denoted as $c_{1,2}$, is the only component that is a function of the covariance of the underlying spatial process X . When the covariance of X is stationary, then $c_{1,2}(s, s') = c_0(h)$. For instance, when X has a squared exponential covariance function with range ρ , under an squared exponential kernel function with fixed priors for the bandwidth, we obtain that $c_{1,2}(s, s')$ is stationary and for small values of the kernel bandwidth ε , $c_{1,2}(s, s')$ approximates the squared exponential correlation function. More specifically, we have

$$c_{1,2} = (2\pi)^2 \left(\frac{\varepsilon^2 \rho}{\rho + 2\varepsilon} \right) \exp \left\{ -h' h / 2 \left(\frac{\rho(\rho + 2\varepsilon)}{\rho + \varepsilon} \right)^{-1} \right\} \quad (7)$$

where $(\varepsilon_{i1}, \varepsilon_{i2}) = (1, \varepsilon)$ or $(\varepsilon_{i1}, \varepsilon_{i2}) = (\varepsilon, 1)$, i.e. for $i_1 = 1$, or $i_2 = 1$, which are the dominant terms in the summation presented in (6), and when ρ and ε are fixed. Therefore, the second term in (6), as the kernel bandwidths become smaller, is the familiar squared exponential correlation function with range ρ .

In the expression for the marginal covariance of the process Y as presented in (6), 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 , in general would be 1 degree higher than the degree of differentiability for the covariance of X , assuming the kernel functions are analytic functions. In (6), 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 would not be stationary even for an i.i.d. X underlying process (see eq. (8)).

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

$$\begin{aligned} \text{cov}(Y(s), Y(s')) = & \sum_i \sigma_i^2 E(V_i^2) [1 - c_1(s)E(V_i) - c_1(s')E(V_i) + c_2(s, s')E(V_i^2)]^{(i-1)} \\ & + \sum_{i_1 \neq i_2} c(s, s')E(V_{i_1} V_{i_2}) [1 - c_1(s)E(V_{i_1}) - c_1(s')E(V_{i_2}) + c_2(s, s')E(V_{i_1} V_{i_2})]^{(i_1-1)(i_2-1)} \end{aligned} \quad (8)$$

where

$$\begin{aligned} c_1(s) &= \int \int K_i(s) p(\varepsilon_i | \varphi_i) p(\varphi_i) d\varphi_i d\varepsilon_i \\ c_2(s, s') &= \int \int K_i(s) K_i(s') p(\varepsilon_i | \varphi_i) p(\varphi_i) d\varphi_i d\varepsilon_i \\ c(s, s') &= \int \int \int K_{i_1}(s) K_{i_2}(s') C(|\varphi_{i_1} - \varphi_{i_2}|) p(\varphi_{i_1}) p(\varepsilon_{i_1} | \varphi_{i_1}) d\varphi_{i_1} d\varepsilon_{i_1} p(\varphi_{i_2}) p(\varepsilon_{i_2} | \varphi_{i_2}) d\varphi_{i_2} d\varepsilon_{i_2} \end{aligned}$$

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 presented in (8), 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 , in general would be 1 degree higher than the degree of differentiability for the covariance of X , assuming the kernel functions are analytic functions.

4.5 Marginal properties: multivariate cross covariance

We present marginal properties of the multivariate spatial stick prior representation in (3). We study the marginal cross-covariance between any pair of data 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 space-dependent functions, so they are a function of the knots, i.e. for data process Y_k the kernel function $K_{i,k}$ has a bandwidth $\varepsilon_{i,k}(\phi_{i,k})$ that is a function of the kernel knot $\phi_{i,k}$. Then, integrating over the probability masses, the covariance between two observations is

$$\begin{aligned} \text{cov}(Y_1(s), Y_2(s')) \\ = \sum_{i_1, i_2} E(V_{i_1,1} V_{i_2,2}) c_{i_1, i_2}^*(s, s') [1 - c_{1,1}^*(s) E(V_{i_1,1}) - c_{2,2}^*(s') E(V_{i_2,2}) + c_{i_1, i_2}^*(s, s') E(V_{i_1,1} V_{i_2,2})]^{(i_1-1)(i_2-1)}, \end{aligned} \quad (9)$$

where,

$$\begin{aligned} c_{1,k}^*(s) &= \int \int K_{i,k}(s) p(\varphi_{i,k}) p(\varepsilon_{i,k} | \varphi_{i,k}) d\varphi_{i,k} d\varepsilon_{i,k} \\ c_{2,2}^*(s, s') &= \int \int \int K_{i,1}(s) K_{i,2}(s') p(\varphi_{i,1}) p(\varepsilon_{i,1} | \varphi_{i,1}) d\varphi_{i,1} d\varepsilon_{i,1} p(\varphi_{i,2}) p(\varepsilon_{i,2} | \varphi_{i,2}) d\varphi_{i,2} d\varepsilon_{i,2} \\ c_{i_1, i_2}^*(s, s') &= \int \int \int K_{i_1,1}(s) K_{i_2,2}(s') C_{1,2}(\varphi_{i_1}, \varphi_{i_2}) p(\varphi_{i_1}, 1) p(\varepsilon_{i_1,1} | \varphi_{i_1,1}) d\varphi_{i_1,1} d\varepsilon_{i_1,1} p(\varphi_{i_2,2}) p(\varepsilon_{i_2,2} | \varphi_{i_2,2}) d\varphi_{i_2,2} d\varepsilon_{i_2,2} \end{aligned}$$

The marginal covariance in (9) allows for lack of stationarity in the sign of the spatial cross-dependency structure. Thus, the covariance, $\text{cov}(Y_1(s), Y_2(s'))$, could have a different sign depending on location. This is induced by the change of sign in the cross-covariance $C_{1,2}$ of the underlying process X . The dependency structure in (9) will be in general nonseparable, in the sense that we can not separate the dependence between Y_1 and Y_2 and the spatial dependence between locations s and s' . This still would be the case even if the covariance of the multivariate process X is separable.

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 (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 (1). If the probability masses $p_i(s)$ in (1) are almost surely continuous in s (i.e., as $|s_1 - s_2| \rightarrow 0$, then $p_i(s_1) - p_i(s_2) \rightarrow 0$ with probability one), then $Y(s_1)$ converges weakly to $Y(s_2)$ ($F_{s_1}(Y)$ is close to $F_{s_2}(Y)$) with probability one as $|s_1 - s_2| \rightarrow 0$.

If the kernel functions $K_i(s)$ in (1) are continuous functions in s , then we will have that the probability masses $p_i(s)$ are almost surely continuous in s , and $F_{s_1}(Y)$ will approximate $F_{s_2}(Y)$ as $|s_1 - s_2| \rightarrow 0$ (by Theorem 3). From Theorem 3, we do not need X (underlying process) to be almost surely continuous to have weak convergence for Y . However, to have almost surely continuous realizations for Y , we need X to be almost surely continuous (see Theorem 4).

We present next one of the main features of our model, which is the fact that the limiting process has continuous realizations.

5 Continuous realizations for the limiting process

We say that a process Y has almost surely continuous realizations, if $Y(s)$ converges to $Y(s_0)$ with probability one as $|s - s_0| \rightarrow 0$, for every $s_0 \in \mathcal{R}^2$.

Unlike other nonparametric spatial models in the literature (Gelfand et al., 2005), our representation does not require repeated measures at the observation locations, because the latent process specification has continuous realization in the limit. If the underlying surface of interest is continuous and smooth, this continuity is not only a desirable theoretical property, but it also should offer 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 would only need one component with the model introduced here. Furthermore, a discontinuous SB prior would never correspond to a spatial Gaussian process.

Theorem 4—Let $Y(s)$ be a random field, with random distribution given by $F_s(Y)$ as in (1). If the underlying stationary process $X(s)$ is almost surely (a.s.) continuous in s for every $s \in \mathcal{R}^2$, (i.e., as $|s - s_0| \rightarrow 0$, then $X(s) - X(s_0) \rightarrow 0$ with probability one), then as the bandwidth parameter ε_i , assuming $B_i = \varepsilon_i I$, of each kernel function K_i in (1) goes uniformly to zero for all i , and the knots become more dense (i.e. $M \rightarrow \infty$), the process Y has a.s. continuous realizations.

In Theorem 4, the continuity of the realizations for the process Y is established based on having kernels with small bandwidths, the number of knots going to infinite within our fixed domain, and an underlying process X that is almost surely continuous. It is important to note that for a stationary spatial process X , the choice of the covariance function determines whether process realizations are a.s. continuous (Kent, 1989). Kent 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 the infinite stick-breaking process. We focus on the following truncation approximation to (1),

$$\sum_{i=1}^N p_i(s) \delta(X(\varphi_i)) + (1 - \sum_{i=1}^N p_i(s)) \delta(X(\varphi_0))$$

resulting in a distribution $F_{s,N}$ such that $F_{s,N} \rightarrow F_s$, with F_s the distribution of our stick-breaking process at location s . Letting $p_0(s)$ denote the probability mass on $X(\varphi_0)$, 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 work 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. bandwidth) are not (discrete) conditionally autoregressive spatial processes, but rather continuous spatial processes.

6.2 MCMC details

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

$$P(g(s)=m) = \frac{p_m(s) \Phi(X(\phi_m), \Sigma_e)}{\sum_{i=1}^M p_i(s) \Phi(X(\phi_i), \Sigma_e)}, \quad (10)$$

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

$$\begin{aligned} \text{Var}[U_k(\phi_i)|\text{rest}]^{-1} &= \{\Omega_k^{-1}\}_{ii} + \sum_{l=1}^n \sum_{k'=1}^p I(g(s_l)=i) \left(\frac{a_{kk'}}{\sigma_{k'}^2} \right)^2 \\ E[U_k(\phi_i)|\text{rest}] &= \text{Var}[U_k(\phi_i)|\text{rest}] \left[-\sum_{l \neq i} \{\Omega_k^{-1}\}_{li} U_k(\phi_l) + \sum_{l=1}^n \sum_{k'=1}^p I(g(s_l)=i) \frac{r_{k'}(s_l) a_{kk'}}{\sigma_{k'}^2} \right], \end{aligned} \quad (11)$$

and $r_k(s_l) = Y_k(s_l) - \sum_{k' \neq k} a_{kk'}(\phi_i) U_{k'}(\phi_i)$. The full conditional for $a_{kk}(\phi_i)$ is nearly identical to the full conditional of $U_k(\phi_i)$ and 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 $V_i^{(t-1)}$ otherwise. To compute the acceptance ratio R , we calculate the stick-breaking probabilities using all the current values of the parameters, denoted $p_m^{(t-1)}(s)$, and using all the current values with $V_i^{(t-1)}$ replaced by the candidate V_i^* , denoted $p_m^*(s)$. The acceptance ratio is then

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

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 were 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 standard 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 would be 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 by definition does not need to be an infinite mixture, and in many cases a small number of components can explain very well the complex underlining spatial structure.

7 Simulation Study

We generate bivariate spatial data on a 15×15 regular grid of spatial locations spanning $[0, 1]^2$. Let $(y_1(s), y_2(s))'$ be the observation at location $s \in [0, 1]^2$. We generate the spatial data by first drawing two independent bivariate Gaussian processes $(z_1^{(1)}(s), z_2^{(1)}(s))'$ and $(z_1^{(2)}(s), z_2^{(2)}(s))'$ with stationary and separable covariances

$$\text{Cov} \left[(z_1^{(j)}(s), z_2^{(j)}(s))', (z_1^{(j)}(t), z_2^{(j)}(t))' \right] = \exp(-\|s-t\|/\rho_j) \sum_j, \text{ where } \rho_1 = 0.05, \rho_2 = 0.25,$$

$$\sum_1 = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}, \text{ and } \sum_2 = \begin{pmatrix} 1 & -0.8 \\ -0.8 & 1 \end{pmatrix}.$$

These stationary Gaussian processes are used to generate non-stationary and/or non-Gaussian processes in the following four ways:

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

where $w(s) = \exp(-10 [(s_1 - 0.2)^2 + (s_2 - 0.2)^2])$, $f(c) = G^{-1}(\Phi(c))$, 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 function for the first component $\sqrt{w(s)}$ high in the lower left corner near $s = (0.2, 0.2)'$, and low in the upper right corner. Therefore, the spatial range is small (following ρ_1) and the cross-correlation is positive (following $\sum_1$) in the lower left; the spatial range is large (following ρ_2) and the cross-correlation is negative (following $\sum_2$) 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.

From each design we generate a $K = 25$ 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 are retained, while the second response y_2 is 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 without (“partial data”) an observation of y_2 .

We fit four models to each data set. The first model is the stationary, Gaussian model $y_j(s) \sim N(\mu_j(s), \sigma_j^2)$, where $\mu_j(s)$ is a bivariate Gaussian processes with separable covariance $\text{Cov}[(\mu_1(s), \mu_2(s))', (\mu_1(t), \mu_2(t))'] = \exp(-\|s - t\|/\rho_j)\Sigma_j$. For priors we assume an inverse gamma for $\sigma_j^2 \sim \text{InvGamma}(0.1, 0.1)$, $\Sigma \sim \text{Wishart}(2.1, 0.1I_2)$, and $\rho \sim \text{Unif}(0, 1)$. The prior distribution for the variance parameter σ_0^2 of the measurement error process is an Inverse Gamma $\text{InvGamma}(0.1, 0.1)$. The second model is also Gaussian with a separable covariance, but we allow for nonstationarity using the kernel convolution model for $\mu_j(s)$ (Higdon et al., 1999). That is,

$$\mu_j(s) = \sum_{k=1}^M w_{jk}(\|s - \phi_k\|)\theta_k \quad (12)$$

where $\theta_k \stackrel{iid}{\sim} N(0, 1)$, w_{jk} is a kernel function described below, and $\phi_k \in \mathcal{R}^2$ are fixed spatial knots. We pick the kernel function to approximate the Matérn covariance

$$\text{Cov}[(\mu_1(s), \mu_2(s))', (\mu_1(s+h), \mu_2(s+h))'] = \frac{\tau^2}{2\nu^{-1}\Gamma(\nu)} \left(\frac{2\nu^{1/2}\|h\|}{\rho} \right)^\nu \mathcal{K}_\nu \left(\frac{2\nu^{1/2}\|h\|}{\rho} \right), \quad (13)$$

where $\mathcal{K}$ is the modified Bessel function. The Matérn covariance has three parameters: $\tau^2 > 0$ controls the variance, $\rho > 0$ controls the spatial range of the correlation, and $\nu > 0$ controls the smoothness of the process. The kernel representation allows the spatial range to vary spatially, and across response type. The Matérn kernel function with spatially-varying range is

$$w_{jk}(u) = \tau_j^2 \frac{\gamma(\nu_j+1)^{1/2} \nu_j^{\nu_j/4+1/4} |u|^{\nu_j/2-1/2}}{\pi^{1/2}\Gamma(\nu_j/2+1/2)\Gamma(\nu_j)^{1/2} \rho_{jk}^{\nu_j/2+1/2}} \mathcal{K}_{\nu_j/2+1/2} \left(\frac{2\nu_j^{1/2}|u|}{\rho_{jk}} \right). \quad (14)$$

We take $\log(\rho_{jk})$ 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 assume $\nu_j \sim \text{Unif}(0, 10)$, $\tau_j^{-2} \sim \text{Gamma}(0.01, 0.01)$, $\bar{\rho}_j \sim N(-2, 1)$, and $\eta_j \sim \text{Unif}(0, 1)$. Note the prior 95% interval for $\exp(\rho_j)$ is roughly (0, 1), comparable to prior for the spatial range in Model 1. The knots have a spatial uniform prior.

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

Overall, the proposed model outperforms its competitors, and offers a significant reduction in the MSE. For complete data locations, in design 1, the stationary Gaussian case, all model perform similarly, except for the Reich and Fuentes (2007) SB prior that slightly

underperforms all other competitors. When modelling complex structures (design 4) using our proposed nonparametric model, there is a significant reduction of 41% 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, with design 2 a non-Gaussian but stationary model, 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 outperforms all other models at partial data locations. In particular, for design 4 there is reduction of 28% in the MSE when compared to the stationarity Gaussian model, a reduction of 26% when compared to the nonstationary Gaussian model, and a 9% reduction when compared to the Reich and Fuentes (2007) spatial SB prior. We conclude that 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 alternative standard modelling frameworks 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 is capable of reproducing the 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 proposed here, we analyze simultaneously monthly average values of ammonium and nitrate for January, 2007 at 209 monitoring stations in the US, the data are plotted in Figure 1.

These data were obtained from the US EPA <http://www.epa.gov/airexplorer/index.htm>. The spatial locations are projected to a flat surface using the Mercator projection and re-scaled to $[0, 1]^2$. The data are right-skewed, so we use log transformation for each pollutant. For each pollutant we remove 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 compare the four models described in Section 7 using five-fold cross validation. For the stick-breaking models we find that $M = 225$ terms is sufficient to approximate the infinite dimensional model. For all models, the priors are the same as those in Section 7, using different space-dependent uniform kernels for each pollutant, which share knots and bandwidth. Having different bandwidths for both pollutants did not seem necessary in this particular application. The observations are 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.

Figure 3 plots the estimated cross-correlation between ammonium and nitrate. More

specifically, we plot $\sum_{(12)}^{(i)} / \sqrt{\sum_{(11)}^{(i)} \sum_{(22)}^{(i)}}$ as a function of the knots ϕ_i , where $\Sigma^{(i)}$ is described in Section 3. The correlation varies dramatically across space, ranging from around 0.2 in the Rocky Mountains and Great Plains (RMGP) area, to 0.9 in the Northeast. In the RMGP area both pollutants show high spatial heterogeneity and rapid spatial changes that occur at different rates for both pollutants, responding to changes in elevation, weather and pollution sources. Thus, as a result there is less spatial cross-dependence in this area, in particular compared to the Northeast, where weather, elevation and sources remain more stable.

There is also considerable variation in the bandwidths in Figure 4, where we plot the posterior mean of the spatially-varying bandwidth. The largest bandwidths are in the Midwest and Southwest, where both pollutants show smoother spatial patterns.

Convergence is monitored using trace plots of the deviance and several parameters. The deviance converges after 1000 iterations, which is the burn-in used in these simulations.

Finally, we conduct a sensitivity analysis to various modeling assumptions. We refit the model increasing the number of terms in the mixture model to $M = 400$, and also with squared-exponential kernels. Figure 5 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 study sensitivity with respect to prior distributions for model parameters. We include here results for the 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 equal to 1, $\text{LogNormal}(0,1)$, rather than using 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, varies 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 the prior distribution for knots. In summary, we conclude that the results are robust to the choice of kernel and prior distributions, as well as increasing the number of mixture components and prior distribution for knot selection.

9 Discussion

In this paper we introduce a new modelling nonparametric framework for multivariate spatial processes, that is flexible enough to characterize complex nonstationary dependency and cross-dependency structures, while avoiding specifying a Gaussian distribution. One of the main advantages of the proposed multivariate spatial stick breaking approach is that it is very computationally efficient, and that should be widely useful in situations where several complex spatial processes need to be modelled simultaneously.

The univariate kernel stick breaking version also provides a useful alternative to recently developed spatial stick-breaking processes (Griffin and Steel, 2006; Dunson and Park, 2008; Gelfand et al., 2007; Reich and Fuentes, 2007). The univariate version has the attractive feature that the limiting process is continuous, rather than discrete as in the other kernel-type stick breaking approaches, and also allows for non-stationarity.

An advantage of the formulation presented in this paper is that many of the tools developed for Dirichlet processes can be applied with some modifications, and that has allowed us to study the statistical properties of the presented methods. The application presented in this paper is for a multivariate spatial process, but the presented framework can be applied to

multivariate spatial-temporal processes by using space-time kernels. In our future work, we will implement and extend the presented methods to spatial temporal processes.

References

- An Q, Wang C, Shterev I, Wang E, Carin L, Dunson D. Hierarchical kernel stick-breaking process for multi-task image analysis. International Conference on Machine Learning (ICML). 2008 to appear.
- Batalha JRF, Saldiva PHN, Clarke RW, Coull BA, Stearns RC, Lawrence J, Murthy GKG, Koutrakis P, Godleski JJ. Concentrated ambient air particles induce vasoconstriction of small pulmonary arteries in rats. *Environmental Health Perspectives*. 2002; 110:1191–1197.10.1289/ehp.5484 [PubMed: 12460797]
- Choi J, Reich BJ, Fuentes M, Davis JM. Multivariate spatial-temporal modeling and prediction of speciated fine particles. *Journal of Statistical Theory and Practice*. 2009; 3:407–418. [PubMed: 20407598]
- Clarke RW, Coull BA, Reinisch U, Catalano P, Killingsworth CR, Koutrakis P, Kavouras I, Murthy GKG, Lawrence J, Lovett E, Wolfson JM, Verrie RL, Godleski JJ. Inhaled concentrated ambient particles are associated with hematologic and bronchoalveolar lavage changes in canines. *Environmental Health Perspectives*. 2000; 108:1179–1187.10.2307/3434831 [PubMed: 11133399]
- Dominici F, Daniels M, Zeger SL, Samet JM. Air Pollution and Mortality: Estimating Regional and National Dose-Response Relationships. *Journal of the American Statistical Association*. 2002; 97:100–111.
- Dunson DB, Park JH. Kernel stick-breaking processes. *Biometrika*. 2008; 95:307–323. [PubMed: 18800173]
- Gelfand AE, Schmidt AM, Banerjee S, Sirmans CF. Nonstationary Multivariate Process Modelling through Spatially Varying Coregionalization (with discussion), *Test*. 2004; 13(2):1–50.
- Gelfand AE, Kottas A, MacEachern SN. Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*. 2005; 100:1021–1035.
- Gelfand, AE.; Guindani, M.; Petrone, S. Bayesian nonparametric modeling for spatial data analysis using Dirichlet processes. In: Bernardo, J., et al., editors. *Bayesian Statistics*. 2007. p. 8
- Griffin JE, Steel MFJ. Order-based dependent Dirichlet processes. *Journal of the American Statistical Association*. 2006; 101:179–194.
- Higdon, D.; Swall, J.; Kern, J. Non-stationary spatial modeling. In: Bernardo, JM.; Berger, JO.; David, AP.; Smith, AFM., editors. *Bayesian Statistics*. Vol. 6. Oxford University Press; 1999. p. 761–768.
- Laud P, Ibrahim J. Predictive model selection. *Journal of the Royal Statistical Society B*. 1995; 57:247–262.
- Hughes LS, Allen JO, Salmon LG, Mayo PR, Johnson RJ, Cass GR. Evolution of nitrogen species air pollutants along trajectories crossing the Los Angeles area. *Environmental Science and Technology*. 2002; 36:3928–3935.10.1021/es0110630
- Matérn, B. *Meddelanden från Statens Skogsforskningsinstitut*. Vol. 49. Almaenna Foerlaget; Stockholm: 1960. Spatial Variation.
- Palacios MB, Steel MFJ. Non-Gaussian Bayesian geostatistical modelling. *Journal of the American Statistical Association, Theory and Methods*. 2006; 101:604–618.
- Papaspiliopoulos O, Roberts G. Retrospective MCMC for Dirichlet process hierarchical models. *Biometrika*. 2008; 95:169–186.
- Petrone, S.; Guindani, M.; Gelfand, AE. Tech. report at the Statistics Department at Duke University. Durham, NC: 2008. Hybrid Dirichlet mixture models for functional data.
- Reich BJ, Fuentes M. A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *Annals of Applied Statistics*. 2007; 1:249–264.
- Sethuraman J. A constructive definition of Dirichlet priors. *Statistica Sinica*. 1994; 4:639–650.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society B*. 2002; 64:583–639.

- Song HR, Fuentes M, Ghosh S. A comparative study of Gaussian geostatistical models and Gaussian Markov random field models. *Journal of Multivariate analysis*. 2008; 99:1681–1697. [PubMed: 19337581]
- Rubin DB. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics*. 1984; 12:1151–1172.

(a) Monitoring data locations

(b) Ammonium

(c) nitrate

Figure 1. Maps of monitoring data locations and log-transformed ammonium and nitrate data.

(a) True cross-covariance

(b) Estimated cross-covariance

Figure 2. Comparison between the true cross-covariance structure and the posterior mean of the cross-covariance using our proposed method.

Figure 3.
Map of the correlation between ammonium and nitrate.

Figure 4.
Posterior mean of the bandwidth parameter.

Figure 5. Posterior mean (in log scale) for models with different kernels (Uniform or Gaussian) and different number of terms in the mixture ($M = 225$ or $M = 400$).

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)

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