Multivariate spatial nonparametric modelling via kernel processes mixing *

Montserrat Fuentes and Brian Reich

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 nonstationary spatial model, as different kernel functions lead to different relationships between the

*M. Fuentes is a Professor of Statistics at North Carolina State University (NCSU). Tel.: (919) 515-1921, Fax: (919) 515-1169, E-mail: fuentes@stat.ncsu.edu. B. Reich is an Assistant Professor of Statistics at NCSU. 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. *Key words:* dirichlet processes, nonseparability, nonstationarity, spatial models.

distributions at nearby locations. This approach is the first to allow both the probabilities and the locations of the SB prior to depend on space, and that way 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 speciated particulate matter.

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, and it is flexible enough to characterize the potentially complex spatial structures of the tail and extremes of the multivariate distribution.

We develop a new multivariate model that overcomes the assumption of normality, stationarity and/or separability, and is computationally efficient. The assumption of separability offers a simplified representation of any variance-covariance matrix, and consequently, some remarkable computational benefits. The concept of separability refers a covariance matrix in d dimensions that is the kronecker product of less than d smaller covariance matrices, for example a multivariate spatial covariance is often modeled as the kronecker product of a spatial covariance matrix and a cross-covariance matrix. Suppose that $\{Y(s) : s \equiv (s_1, s_2, \cdots, s_d)' \in D \subset \mathbb{R}^d\}$ denotes a spatial process where s is a spatial location over a fixed domain D, $\mathbb{R}^d$ is a d-dimensional Euclidean space. The spatial covariance function is defined as $C(s_i - s_j) \equiv \text{cov}\{Y(s_i), Y(s_j)\}$, where $s_i = (s_1^i, \cdots, s_d^i)'$ and C is positive definite. Under the assumption of stationarity, we have $C(s_i - s_j) \equiv C(h)$, where $h \equiv (h_1, \cdots, h_d)' = s_i - s_j$. However, in real applications, especially with air pollution data, it may not be reasonable to assume that the covariance function is stationary. Spatial models often assume the outcomes follow normal distributions, or that the observations are a realization of a Gaussian process. However, the Gaussian assumption is overly restrictive for pollution data, which often display heavy tails that change across space. We propose here a new nonparametric (Dirichlet-type) multivariate spatial model flexible enough to characterize extreme events in the presence of complex spatial structures, that allows for nonstationarity in both the covariance and cross-covariance between outcomes.

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's distribution. The SB prior is a special case of the priors neutral to the right (Doksum, 1974). 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, $p_i = V_i \prod_{j=1}^{i-1}(1 - V_j)$, $V_i \sim$

Beta(a_i, b_i) independent across i, $\delta(X_i)$ is the Dirac distribution with point mass at X_i, $X_i \overset{iid}{\sim} F_o$, and F_o is a known distribution. A special case of this prior is the Dirichlet process prior with infinite M and $V_i \overset{iid}{\sim}$ Beta$(1, \nu)$ (Ferguson, 1973). The stick-breaking prior has been extended to the univariate spatial setting by incorporating spatial information into the either the model for the locations 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 locations 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. 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 (iid) bandwidths, 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 in the spatial cross-dependency between the outcomes of interest. 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 we introduce here is the first to allow both the probabilities p_i and the locations X_i to depend on space, and that way there is no need for replications and we obtain a continuous process in the limit. We extend the model to the multivariate setting, by using a multivariate version of the space-dependent kernel SB prior, having for each process a different kernel function, but sharing the location of the kernel knots across the different processes. The resulting marginal covariance for the multivariate process is in general nonstationary and nonseparable.

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 determinants that we have in the likelihood function might not be feasible. The modelling framework proposed here is computationally efficient because it avoids inverting large matrices and calculating determinants.

The application presented in this paper is in air pollution, to model and characterize the complex spatial structure of speciated fine particulate matter. Fine particulate matter ($PM_{2.5}$) is an atmospheric pollutant that has been linked to serious health problems, including mortality. 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. Özkaynak and Thurston (1987) conducted an analysis of the association between several particle measures and mortality. Their results showed the importance of considering particle size, composition, and source information when modeling particle pollution health effects. $PM_{2.5}$ is a mixture of pollutants, and it has five main components: sulfate, nitrate, total carbonaceous mass, ammonium, and crustal material. These components have complex spatial dependency and cross dependency structures. The $PM_{2.5}$ chemistry changes with space and time so its association with mortality could change across space and time. Dominici et al. (2002) showed that different cities have different relative risk of mortality due to $PM_{2.5}$ exposure. It is important to gain insight and better understanding about the spatial distribution of each component of the total $PM_{2.5}$ mass, and also to estimate how the composition of $PM_{2.5}$ might change across space. This type of analysis is needed to conduct epidemiological studies of the association of these pollutants and adverse health effects. We introduce a innovative multivariate spatial model for speciated $PM_{2.5}$ across the entire United States. The proposed framework captures the cross dependency structure among the $PM_{2.5}$ components and explains their spatial dependency structures. Due to the complex cross-dependency between $PM_{2.5}$ components, it is important to introduce a statistical framework flexible enough to allow the cross-dependency between pollutants to be space-dependent, while allowing for nonstationarities in the spatial structure of each component. This is the first study with speciated particulate matter that allows the cross-dependency between components vary spatially. In the context of air pollution we are interested in extreme values (hot spots), so we need spatial models that go beyond the assumption of normality. The multivariate spatial SB model introduced in this paper seems to provide an ideal framework to characterize the complex structure of the speciated $PM_{2.5}$.

The paper is organized as follows. In Section 2, we introduce a new spatial univariate spatial model using an extension of the stick-breaking prior, that 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 introduce a particular case of the multivariate spatial model presented in Section 3, which is a simple model and very easy to implement, but general and flexible enough for many applications. In Section 5, we present the conditional properties of the spatial SB process prior. In Section 6, we present marginal properties. In Section 7, we study some asymptotic properties of the spatial SB prior. In Section 8, we present and prove an important feature of our model, which is that the limiting process has continuous realizations. In Section 9, we include some computational approaches and Markov Chain Monte Carlo(MCMC) details for its implementation. In Section 10, we illustrate the proposed methods with a simulation study. In Section 11, we present an application to air pollution. We conclude with Section 12 that has some remarks and final comments.

2 Univariate 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 $D \in \mathbb{R}^d$, in particular for $d = 3$ this could correspond to a continuous spatial temporal process. To simplify notation throughout this section we will assume $d = 2$.

Extending the stick-breaking prior to depend on location, the prior for $F_s(Y)$ is the potentially infinite mixture

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

where X is a Gaussian process 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. The weight function, K_i, is a spatial kernel centered at ϕ_i 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 ellipse. We represent $B_i = T(\phi_i)T(\phi_i)'$, where $T(\phi_i)'$ denotes the transpose of $T(\phi_i)$, and $T_{kk'}(\phi_i)$ is the (k, k') element of the matrix $T(\phi_i)$, $\log(T_{11}(\phi_i))$, $\log(T_{22}(\phi_i))$, and $T_{12}(\phi_i)$ are independent spatial Gaussian processes.

As we will see in Sections 5 and 6 the spatial correlation of the process $Y(s)$ is controlled by the bandwidth parameters associated with knots in the vicinity of s. To allow the correlation to vary from region to region, the bandwidth parameters are modeled as a spatially varying parameters. We assign to $\log(T(\phi_i))$ a spatial Gaussian prior with non-zero mean and a Matérn correlation function (Matérn, 1960). We assume the Matérn spatial correlation with the parameterization proposed by Handcock and Wallis (1994), that is,

$$\text{corr}(\log(T(\phi_i)), \log(T(\phi_j)) = \frac{(2\nu^{1/2}|\phi_i - \phi_j|/\rho)^\nu}{2^{\nu-1}\Gamma(\nu)}\mathcal{K}_\nu(2\nu^{1/2}|\phi_i - \phi_j|/\rho),$$

where $\mathcal{K}_\nu$ is the modified Bessel function, and $|\phi_i - \phi_j|$ denotes the Euclidean distance between ϕ_i and ϕ_j. The Matérn correlation has two parameters: ν controls the smoothness of the process (i.e., the degree of differentiability), and ρ controls the range of spatial correlation. If $\nu = 0.5$ we have the exponential correlation $\exp(|\phi_i - \phi_j|/\rho)$, and if $\nu = \infty$ we have the squared-exponential correlation $\exp(|\phi_i - \phi_j|^2/\rho^2)$.

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),

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

where $\sigma_i = \sigma(\phi_i)$ is the the variance of the process $X(\phi_i)$ and it is space-dependent, and ρ is a correlation function (e.g. a Matérn). We assign to $\log(\sigma(\phi_i))$ a spatial Gaussian prior, with non-zero mean and a Matérn covariance function.

In Appendix A.1 we prove that the representation in (1) introduces a properly defined process prior, using the Kolmogorov existence theorem. In Sections 6 and 7 we study the marginal and conditional properties of model (1). If the knots are uniformly distributed across space, the bandwidth parameters are not space-dependent functions, $\sigma_i = \sigma$ for all i, and the prior distributions for the knots, bandwidth parameters and V_i's are independent, then, the corresponding covariance (integrating out V_i, ϕ_i, T) assuming the kernel is not space-dependent, i.e. is stationary (as shown in Section 7). Allowing the bandwidth to be space-dependent will make it nonstationary (Section 7). The conditional covariance will be in general nonstationary, even when the kernels are not space-dependent (Section 6).

Model (1) is a Dirichlet Process (DP) mixture model with spatially varying weights. The almost sure discreteness of $F_s(Y)$ is not desirable in practice. Thus, we mix it with a pure error process (nugget process) with respect to $F_s(Y)$ to create a random distribution process $G_s(Z)$ with continuous support (Gelfand et al., 2005). Thus, 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 $e(s) \sim N(0, \sigma_0^2)$. We denote Σ_e the diagonal covariance of the nugget process e. Thus, our model requires the introduction of a nugget effect.

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, $Y_1(s), \ldots, Y_p(s)$, by introducing the following model for the distribution of each $Y_k(s)$,

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

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 knots of the kernel functions $K_{i,k}$ are shared across the p space-time processes, and the p-dimensional process $\mathbf{X} = (X_1, \ldots, X_p)$ has a multivariate normal prior, with cross covariance

$$\text{cov}((X_1(\phi_1), \ldots, X_p(\phi_1)), (X_1(\phi_2), \ldots, X_p(\phi_2))) = C(\phi_1, \phi_2) = \sum_{j=1}^{k} \rho_j(\phi_1 - \phi_2) g_j(\phi_1) g_j^t(\phi_2) \tag{4}$$

with g_j the jth column of a $p \times p$ full rank matrix G, and the ρ_j are correlation functions (e.g., Matérn functions). Equation (4) presents a nonstationary extension of a linear model of coregionalization (Gelfand et al, 2004). Gelfand et al. (2004) characterize the cross-covariance of a multivariate Gaussian process $\mathbf{X}$ using a spatial Wishart prior. Here, we propose an alternative framework, that could be more computationally efficient.

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 $\{1, \ldots, p\}$, $a_{kk'}$ 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 $(X_1(\phi_i), \ldots, X_p(\phi_i))$ 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 change with location, we obtain a cross covariance between the Y_k processes that varies with space (nonstationarity), 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, 2008).

4 Particular case

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 $Y_k(s)$,

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

where the V and K random weights are different for each process Y_k. This allows the spatial processes Y_k, with $k = 1, \ldots, p$, 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 nonstationarity in the sign.

5 Conditional properties

5.1 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 $B_i = \epsilon_i I$, 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,

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

$$\begin{aligned} = & \sum_i \sigma_i^2 \left[K_i(s) K_i(s') V_i^2 \prod_{j<i} (1 - ((K_j(s) + K_j(s'))V_j + K_j(s) K_j(s') V_j^2)) \right] \\ + & \sum_{i_1 \neq i_2} [K_{i_1}(s) K_{i_2}(s') V_{i_1} V_{i_2} C(\phi_{i_1} - \phi_{i_2}) \\ & \prod_{j_1 < i_1} \prod_{j_2 < i_2} (1 - (K_{j_1}(s) V_{j_1} + K_{j_2}(s') V_{j_2}) + K_{j_1}(s) K_{j_2}(s') V_{j_1} V_{j_2})], \end{aligned} \tag{5}$$

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 presented in (1), C is the covariance function of X, and $\sigma_i^2 = \text{cov}(X(\phi_i), X(\phi_i))$.

The conditional covariance of the data process Y in (5) between any two locations s and s' is stationary and approximates the covariance function C of the underlying process X, $\text{cov}(X(s), X(s')) = C(|s - s'|)$, as the bandwidths of the kernel functions become smaller. 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. $\epsilon_i < \epsilon$, for all i, such that $\epsilon \to 0$. Assuming that K_i is a kernel with compact support for all i, and that C has a first order derivative, C', that is a bounded nonnegative function.

Proof of Theorem 1: In Appendix A.2.

5.2 Conditional multivariate spatial covariance

Assuming the multivariate spatial stick-breaking process prior $(F_s(Y_1), \ldots, F_s(Y_p))$ in (3) for the data processes $Y_1(s), \ldots, Y_p(s)$, 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 $Y_2(s_2)$, is

$$\begin{aligned}
& \text{cov}(Y_1(s), Y_2(s')|p_1(s), p_2(s'), C) = \sum_i p_{i,1}(s) p_{i,2}(s') C_{1,2}(\phi_i, \phi_i) \\
+ & \sum_{i_1 \neq i_2} p_{i_1,1}(s) p_{i_2,2}(s') C_{1,2}(\phi_{i_1}, \phi_{i_2}) \\
= & \sum_i [K_{i,1}(s) K_{i,2}(s') V_{i,1} V_{i,2} C_{1,2}(\phi_i, \phi_i) \\
& \prod_{j<i} (1 - K_{j,1}(s) V_{j,1} - K_{j,2}(s') V_{j,2} + K_{j,1}(s) K_{j,2}(s') V_{j,1} V_{j,2})] \\
+ & \sum_{i_1 \neq i_2} [K_{i_1,1}(s) K_{i_2,2}(s') V_{i_1,1} V_{i_2,2} C_{1,2}(\phi_{i_1}, \phi_{i_2}) \\
& \left. \prod_{j_1<i_1} \prod_{j_2<i_2} (1 - (K_{j_1,1}(s) V_{j_1,1} + K_{j_2,2}(s') V_{j_2,2}) + K_{j_1,1}(s) K_{j_2,2}(s') V_{j_1,1} V_{j_2,2}) \right] . \qquad (6)
\end{aligned}$$

The conditional cross-covariance between any pair $Y_1(s)$ and $Y_2(s')$ presented in (6) 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 defined 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 $(F_s(Y_1), \ldots, F_s(Y_p))$ in (3) for the data processes $Y_1(s), \ldots, Y_p(s)$, 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 $Y_2(s_2)$, $\text{cov}(Y_1(s), Y_2(s'))$, approximates the cross-covariance function $C_{1,2}(s, s')$ of the underlying process $\mathbf{X}$, as the bandwidth parameters $\epsilon_{i,1}$ and $\epsilon_{i,2}$ of the kernel functions in $F_s(Y_1)$ and $F_s(Y_1)$ go uniformly to zero for all i, i.e. $\epsilon_{i,k} < \epsilon$, for all i and for $k = 1, 2$, with $\epsilon \rightarrow 0$.

Proof of Theorem 2: In Appendix A.3.

6 Marginal properties

6.1 Univariate

We study the marginal properties of the spatial stick prior representation in (1). We start 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. Throughout this section, without loss of generality, we assume $B_i = \epsilon_i I$, where I is the 2×2 identity matrix. We also assume the covariance, C, of the underlying process X is 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

$$
\begin{aligned}
& \text{cov}(Y(s), Y(s')) \\
= \; & \sigma^2 c_2 E(V^2) \sum_i \left[1 - 2c_1 E(V) + c_2 E(V^2)\right]^{i-1} \\
+ \; & \sum_{i_1 \neq i_2} c_{1,2} [1 - 2c_1 E(V) + c_2 E(V)^2]^{(i_1-1)(i_2-1)} \\
= \; & \sigma^2 \frac{\gamma(s, s')}{2(1 + b/(a+1)) - \gamma(s, s')} + E(V)^2 \sum_{i_1 \neq i_2} c_{1,2} [1 - 2c_1 E(V) + c_2 E(V)^2]^{(i_1-1)(i_2-1)}
\end{aligned} \tag{7}
$$

where

$$\gamma(s,s') = \frac{\int\int K_i(s)K_i(s')p(\phi_i,\epsilon_i)d\phi_i d\epsilon_i}{\int\int K_i(s)p(\phi_i,\epsilon_i)d\phi_i d\epsilon_i},$$

$$c_1 = \int\int K_i(s)p(\phi_i,\epsilon_i)d\phi_i d\epsilon_i,$$

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

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

In the expression for the marginal covariance between observations $Y(s)$ and $Y(s')$ in (7), 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 (7) is due to the spatial dependency of the process X. The first term in (7) is a function of s and s' through the function $\gamma(s,s')$. $\gamma(s,s')$ 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. In expression (7), $c_{1,2}(s,s')$ denoted as $c_{1,2}$, is the only component that is a function of the covariance of the underlying spatial process X. Even, when the covariance of X is stationary, $c_{1,2}(s,s')$ would be in general nonstationary, i.e. a function of the spatial locations of s and s'. This is due to the edge effect, a common problem in spatial statistics when working with fixed spatial domains. This is also the case for conditional autoregressive models (CAR), that have a marginal nonstationary covariance, even when generated by aggregation of a stationary point-reference spatial process (e.g. Song, Fuentes and Ghosh, 2008). In the expression for the marginal covariance of the process Y as presented in (7), 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 (7), 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)).

6.2 Univariate with space-dependent Kernels

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

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

where

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

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.

6.3 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, they are a function of the knots, i.e. for data process Y_k the kernel function $K_{i,k}$ has a bandwidth $\epsilon_{i,k}(\phi_{i,k})$ that is a function of the kernel knot $\phi_{i,k}$. Then, integrating over the probability masses, the covariance between two observations is

$$\begin{aligned} & \mathrm{cov}(Y_1(s), Y_2(s')) = \\ & \sum_{i_1,i_2} E(V_{i_1,1}V_{i_2,2}) c^*_{i_1,i_2}(s,s')[1 - c^*_{1,1}(s)E(V_{i_1,1}) - c^*_2(s')E(V_{i_2,2}) \\ + \quad & c^*_{i_1,i_2}(s,s')E(V_{i_1,1}V_{i_2,2})]^{(i_1-1)(i_2-1)}, \end{aligned} \tag{9}$$

where,

$$\begin{aligned} c^*_{1,k}(s) &= \int\int K_{i,k}(s) p(\phi_{i,k}) p(\epsilon_{i,k}/\phi_{i,k}) d\phi_{i,k} d\epsilon_{i,k} \\ c^*_2(s,s') &= \int\int\int\int K_{i,1}(s) K_{i,2}(s') p(\phi_{i,1}) p(\epsilon_{i,1}/\phi_{i,1}) d\phi_{i,1} d\epsilon_{i,1} p(\phi_{i,2}) p(\epsilon_{i,2}/\phi_{i,2}) d\phi_{i,1} d\epsilon_{i,2} \\ c^*_{i_1,i_2}(s,s') &= \int\int\int\int K_{i_1,1}(s) K_{i_2,2}(s') C_{1,2}(\phi_{i_1}, \phi_{i_2}) p(\phi_{i_1,1}) p(\epsilon_{i_1,1}/\phi_{i_1,1}) d\phi_{i_1,1} d\epsilon_{i_1,1} \\ & p(\phi_{i_2,2}) p(\epsilon_{i_2,2}/\phi_{i_2,2}) d\phi_{i_2} d\epsilon_{i_2,2} \end{aligned}$$

The marginal covariance in (9) allows for lack of stationarity in the sign of the spatial cross-dependency structure. Thus, the covariance, $\mathrm{cov}(Y_1(s), Y_2(s'))$, 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 dependency between Y_1 and Y_2 versus the spatial dependency between locations s and s'. This still would hold ((9) would be nonseparable), even if the covariance of the multivariate process X is separable, in the sense that is modelled as a product of a component that explains the dependency across the X components and another component that explains purely the spatial dependency.

7 Weak convergence

In this section we study the weakly 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 $|s_1 - s_2| \to 0$, then $p_i(s_1) - p_i(s_2) \to 0$ 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 $|s_1 - s_2| \to 0$.

Proof of Theorem 3: In Appendix A.4.

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 sure continuous in s, and $F_{s_1}(Y)$ will approximate $F_{s_2}(Y)$ as $|s_1 - s_2| \to 0$ (by Theorem 3). From Theorem 3, we do not need X (underlying process for locations) to be almost surely continuous to have weakly convergence for Y. However, to have almost surely continuous realizations for Y, we need X to be almost surely continuous (see Theorem 4).

8 Continuous realizations for the limiting process

In this section, we present one of the main features of our model, which is the fact that the limiting process has continuous realizations. 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 $s_0 \in \mathcal{R}^2$.

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 sure (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 $B_i = \epsilon_i I$, of each kernel function K_i in (1) goes uniformly to zero for all i, the process Y has a.s. continuous realizations.

Proof of Theorem 4: In Appendix A.5.

In Theorem 4, the continuity of the realizations for the process Y is established based on having kernels with small bandwidths 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.

9 Computing methods

9.1 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),

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

resulting in a distribution $F_{s,N}$ such that $F_{s,N} \to F_s$, with F_s the distribution of our stick-breaking process at location s. Letting $p_0(s)$ denote the probability mass on $X(\phi_0)$, 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 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.

9.2 MCMC details

For computational purposes we introduce auxiliary variables $g(s) \sim \text{Categorical}(p_1(s), ..., p_M(s))$ to indicate $Z(s)$'s mixture component, so that $Z(s)|g(s)$ is normal with mean $\text{X}(\phi_{g(s)})$ 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, ..., U_p are independent spatial process with mean zero, $\text{cov}(U_j(\phi_1), ..., U_j(\phi_M)) = \Omega_j$, 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

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

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

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

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.

Conjugacy does not hold in general for the stick-breaking parameters V_i, B_i or ψ_i or the spatial range parameters; these parameters are updated using Metropolis sampling with Gaussian candidate draws. Candidates with no posterior mass are rejected and the candidate standard deviations were tuned to give acceptance rate near 0.4. We draw 20,000 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.

10 Simulation Study

Here we generate and analyze four univariate data sets using the model of Section 2. In all cases we generate a sample size of $n=250$ by adding standard normal noise to a true surface $\mu^*(s)$. The true surfaces are

1. $\mu^*(s) = 5\sqrt{s_1^2 + s_2^2}\sin(10(s_1^2 + s_2^2))$

2. $\mu^*(s) = 2\sin(20s_1) + 2\sin(20s_2)$

3. $\mu^*(s) = 0$

4. $\mu^*(s) \sim N(0, \Sigma)$

In simulation #4, $\mu^*(s)$ is generated from a multivariate normal with mean zero and exponential covariance $C(d) = exp(-d)$, i.e. a covariance with range and sill parameters equal to 1. The data from simulations #1 and #2 are plotted in Figures 1a and 1e. The spatial domain for all four simulations is $[0, 1]^2$.

We fit four models to each simulated data set by varying the kernels and the spatial correlation structure. We fit both uniform kernels $K_i(s) = I(\delta_i(s) < 1)$ and squared-exponential kernels $K_i(s) = \exp(-0.5\delta_i^2(s))$, where $\delta_i(s) = (s - \phi_i)B_i^{-1}(s - \phi_i)'$, and I is an indicator function. We also compare the stationary model with constant kernel bandwidth and iid latent process $X(s)$ with the full non-stationary model described in Section 2.

We compare these models in terms mean squared error

$$MSE = \frac{1}{250}\sum_{j=1}^{250}[Z(s_j) - \widehat{E(Z(s_j))}]^2],$$

where $\widehat{E(Z(s_j))}$ is the posterior mean of $E(Z(s_j)) = \sum_{i=1}^{M} p_i(s_j)X_i$. For each data set, we also compute the deviance information criteria (DIC) of Speigelhalter et al. (2002) and the predictive criteria of Laud and Ibrahim (LI, 1995). DIC is defined as $DIC = \bar{D} + p_D$, where $\bar{D}$ is the posterior mean deviance and p_D is the effective number of parameters. To calculate LI, at each MCMC iteration we generate a replicate dataset from the current model, denoted

$Z^{(l)}$ for the l^{th} MCMC iteration. We compare these replicate data sets with the observed data $Z(s_j)$ in terms of the posterior mean of the squared-error loss

$$LI = \frac{1}{250}\sum_{i=1}^{250}(Z(s_j) - Z_j^{(l)})^2.$$

Models with small DIC and LI are preferred.

For the stick-breaking parameters we took $a = 1$ and $b\sim$Gamma(0.1,0.1) (parameterized to have mean 1, variance 10). All variances have Gamma(0.1,0.1) priors and the spatial range parameter for the latent mean process X has a Uniform(0,20) prior. As is common in spatial modeling, we found the spatial range of the spatially-varying bandwidth and log standard deviation to be poorly identified. To improve MCMC convergence, we fixed these spatial ranges at 2.5 to give correlation 0.78 between locations with $|s - s'| = 0.1$ (recall the spatial domain is $[0, 1]^2$) and correlation 0.08 for locations with $|s - s'| = 1$. In all model fits we truncated the infinite mixture model with a mixture of M=200 components, which gives satisfactory truncation probability.

Table 1 gives the results for the four simulated examples. Simulation #3 would correspond to a *null* model, and we would expect, as it is the case, the model comparison statistics to give a similar answer for the different models. The model comparison statistics DIC and LI both select the model with the smallest MSE for simulations #1,#2 and #4, supporting their use for these models. The first simulated data set is clearly anisotropic with stronger correlation in the northwest/southeast direction and nonstationary with increased variance in the northeast corner. The nonstationary model with squared-exponential kernel minimizes MSE. The posterior means of the bandwidth and standard deviation for this model in Figures 1c and 1d, respectively, show that our model is able to capture these complicated effects. Both stationary models and the nonstationary model with uniform kernels have MSE between 0.50 and 0.55 for the second simulated example. The nonstationary model with uniform kernels gives slightly smaller MSE, possibly due to the elliptical bandwidths (not shown here) which pool information across nearby undulations with the same heights. Simulation #4 corresponds to a stationary model, and for both types of kernels, all the comparison statistics prefer the stationary model.

11 Analysis of speciated $PM_{2.5}$

In this section we analyze monthly average speciated $PM_{2.5}$ for January, 2007 at 209 monitoring stations in the US, plotted in Figure 2. These data were obtained from the US EPA `http://www.epa.gov/airexplorer/index.htm`. For our analysis the spatial locations are transformed to $[0,1]^2$ and the responses are centered and scaled.

We fit several variations of the multivariate spatial stick-breaking model to these data by varying the kernel (squared-exponential versus uniform), allowing the latent $X(s)$ processes to have independent or spatial correlation, and allowing the covariance parameters A to be constant or varying across space with spatial priors. In each of these models we assume the knots are shared across outcomes and that the kernel bandwidth matrix B is constant across space and outcome. These nonparametric models are compared to the fully Gaussian model with stationary, separable covariance

$$\mathrm{cov}(Z(s), Z(s')) = \Sigma_e + \exp(\rho|s - s'|)\Sigma,$$

where Σ_e (nugget effect) is the $p \times p$ diagonal matrix with diagonal elements $\sigma_j^2 \sim$ InvGamma(0.01,0.01), ρ is the spatial range parameter, and $\Sigma \sim \mathrm{Wishart}(p + 0.1, 0.1 * I_p)$ is the $p \times p$ cross-covariance matrix.

We compare these models using both the DIC and LI, described in Section 10. In addition to these model comparisons, we conduct predictive checks of model adequacy using the Bayesian p-value of Rubin (1984). We compare the replicate data sets with the observed data Z in terms of a discrepancy measure $D(Z)$. The Bayesian p-value for the discrepancy measure is

$$\text{p-value} = \frac{1}{m}\sum_{i=1}^{m} I\left[D(Z) < D(Z^{(i)})\right],$$

where m is the number of MCMC samples. P-values near zero or one indicate that the predictive model does not fit the data adequately. We consider several discrepancy measures including the maximum (over space) $D(Z) = \max(Z_{1j}, ..., Z_{nj})$, where Z_{ij} denotes $Z_j(s_i)$, separately for each of the five pollutants, the skewness of each of the five pollutants $D(Y) = \text{skew}(Z_{1j}, ..., Z_{nj})$, and the difference between the cross-correlation of each pair of pollutants between two locations in the eastern and western halves of the US (see Fig. (12) showing the location of the pair points), $D(Y) = \text{cor}(Z_{ij}, Z_{ik}|i \text{ in the east}) - \text{cor}(Z_{ij}, Z_{ik}|i \text{ in the west})$. These measures are chosen to determine, respectively, whether the models adequately capture tail probabilities, the shape of the response distribution, and the potential non-stationarity in the cross-covariance between pollutants.

The fully parametric and stationary Gaussian model has the highest DIC and LI of the models in Table 2. The p-values in Table 3 indicate that the observed data generally have larger maximums and more skewness than the data sets generated from this model. These p-values remain unsatisfactory even after a log transformation (not shown), motivating a nonparametric analysis. Also, several of the p-values for the difference in correlation between east and west are near zero, suggesting non-stationary.

Both DIC and LI are smaller for the nonparametric models in Table 2 than the parametric Gaussian model. The most distinguishing factor between the eight nonparametric models is the correlation of the latent process $X(\phi)$; allowing the latent processes to have spatial correlation improves DIC and LI for all combinations of kernels and cross-covariance models. The smooth squared-exponential kernels are preferred for the model with independent $X(\phi)$. In contrast, uniform kernels are preferred when $X(\phi)$ is a spatial process. In this case, the Gaussian process $X(\phi)$ captures large-scale spatial trends and the nonparametric stick-breaking model with uniform kernels accounts for local spatial effects.

The best-fitting model with squared-exponential kernels allows for nonstationarity in the cross-covariance. Figure 3 plots the posterior mean and standard deviation of the correlation between carbon and sulfate and between ammonium and nitrate. Although the standard deviations of these correlation functions are fairly large, there are clear trends. For example, the correlation between carbon and sulfate increases from west to east. The Bayesian p-values for the nonparametric models with spatial correlation in the latent process and cross-covariance are all between 0.05 and 0.95 (Table 3). This provides evidence that these models adequately capture the shape of the response distribution and the cross-correlation.

The best-fitting model with uniform kernels has stationary cross-covariance. After modeling the spatial processes adequately with both a Gaussian spatial process $X(s)$ and local nonparametric model with uniform kernels, the cross-covariance can be taken to be constant over space.

12 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 main advantages of the proposed multivariate spatial stick breaking approach is that 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). One useful feature of this univariate version is that the limiting process is continuous, rather than discrete as in the other kernel-type stick breaking approaches.

An advantange 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 an extend the presented methods to spatial temporal processes.

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.

Choi J, Reich BJ, Fuentes M, and Davis JM. (2008). Multivariate spatial-temporal modeling and prediction of speciated fine particles. In press, *Journal of Statistical Theory and Practice.*

Dominici, F, Daniels, M, Zeger, S.L, and Samet, J.M., (2002) Air Pollution and Mortality: Estimating Regional and National Dose-Response Relationships. *Journal of the American Statistical Association,* **97**, 100-111.

Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior d istributions. *The Annals of Probability*, **2**, 183-201

Dunson, D.B. and J. H. Park. (2008). Kernel stick-breaking processes. *Biometrika,* **95**, 307-323.

Gelfand, A.E., A.M. Schmidt, S. Banerjee, and C.F. Sirmans, (2004). Nonstationary Multivariate Process Modelling through Spatially Varying Coregionalization (with discussion), *Test*, **13**, 2, 1-50.

Gelfand AE, Kottas A, and MacEachern SN (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*, **100**, 1021–1035.

Gelfand, A.E.,Guindani, M., and Petrone, S. (2007). Bayesian nonparametric modeling for spatial data analysis using Dirichlet processes. *Bayesian Statistics,* **8**. J. Bernardo et al., eds.

Griffin J.E., and Steel, M.F.J. (2006) Order-based dependent Dirichlet processes. *Journal of the American Statistical Association*, **101**, 179–194.

Laud P., and Ibrahim J. (1995). Predictive model selection. *Journal of the Royal Statistical Society B*, **57**, 247-262.

Matérn, B. (1960). *Spatial Variation.* Meddelanden från Statens Skogsforskningsinstitut, **49**, No. 5. Almaenna Foerlaget, Stockholm.

Özkaynak, H., and Thurston, G. D. (1987), Associations between 1980 U.S. mortality rates and alternative measures of airborne particle concentration *Risk Analysis*, **7**, 449-461.

Palacios, M.B., and Steel, M.F.J. (2006). Non-Gaussian Bayesian geostatistical modelling. *Journal of the American Statistical Association, Theory and Methods,* **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 at the Statistics Department at Duke University. Durham, NC.

Reich B.J., and Fuentes M. (2007). A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *Annals of Applied Statistics,* **1** 249-264.

Sethuraman J. (1994). A constructive definition of Dirichlet priors. *Statistica 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). *Journal of the Royal Statistical Society B*, **64**, 583–639.

Song, H.R., Fuentes, M., Ghosh, S. (2008). A comparative study of Gaussian geostatistical models and Gaussian Markov random field models. *Journal of Multivariate analysis,* **99**, 1681-1697.

Rubin D.B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics*, **12**, 1151–1172.

Table 1: Results for the simulated examples.

Design	Kernel	Correlated $X(\phi)$	Spatially-varying B and σ	$\bar{D}$	p_D	DIC	LI	MSE
1	Squared Expo	No	No	603	144	746	1.32	0.59
1	Uniform	No	No	669	125	794	1.72	0.80
1	Squared Expo	Yes	Yes	535	156	691	1.01	0.33
1	Uniform	Yes	Yes	599	137	735	1.30	0.49
2	Squared Expo	No	No	693	87	779	1.91	0.53
2	Uniform	No	No	756	65	821	2.44	0.77
2	Squared Expo	Yes	Yes	656	109	765	1.67	0.54
2	Uniform	Yes	Yes	604	126	730	1.35	0.51
3	Squared Expo	No	No	701	6.8	708	1.93	0.00
3	Uniform	No	No	688	16.8	705	1.83	0.01
3	Squared Expo	Yes	Yes	689	17.8	701	1.83	0.01
3	Uniform	Yes	Yes	697	11.0	708	1.90	0.01
4	Squared Expo	No	No	625	113	738	1.46	0.11
4	Uniform	No	No	683	74	756	1.80	0.12
4	Squared Expo	Yes	Yes	657	90	746	1.66	0.12
4	Uniform	Yes	Yes	689	72	762	1.86	0.12

Table 2: Model selection results for the speciated $PM_{2.5}$ data analysis.

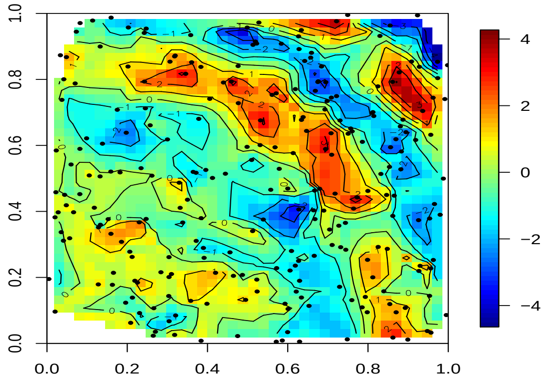
Stick-Breaking Kernel	Correlated $X(\phi)$	Nonstationary covariance, A	$\bar{D}$	p_D	DIC	LI
Squared Expo	No	No	-304	426	121	4.41
Squared Expo	No	Yes	-294	457	162	4.37
Squared Expo	Yes	No	-556	458	-97	3.65
Squared Expo	Yes	Yes	-666	522	-143	3.36
Uniform	No	No	-154	366	212	4.85
Uniform	No	Yes	-401	480	78	3.82
Uniform	Yes	No	-588	406	-182	3.41
Uniform	Yes	Yes	-512	477	-35	3.59
Parametric			-289	533	246	13.9

Table 3: Model adequacy results for the speciated $PM_{2.5}$ data analysis. The table gives the Bayesian p-value for three discrepancies: component-wise maximum and skewness, and the pairwise difference between cross-correlations in the Eastern and Western US. The labels for the species are "A" for Ammonium, "Cr" crustal materials, "TC" for total carbon, "N" for nitrate, and "S" for sulfate.

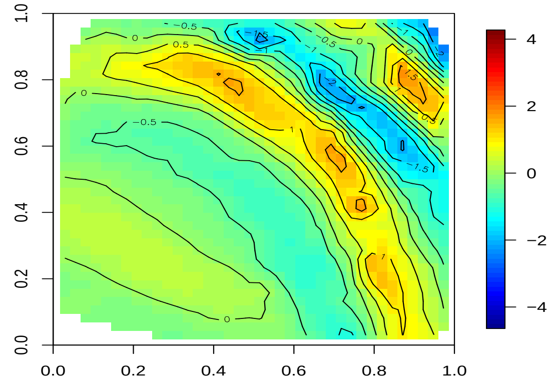
	Maximum					Skewness				
Model	A	Cr	TC	N	S	A	Cr	TC	N	S
SSB - sq exp	0.51	0.56	0.51	0.57	0.59	0.58	0.61	0.65	0.61	0.60
SSB - unif	0.50	0.57	0.53	0.60	0.60	0.59	0.63	0.69	0.63	0.68
Gaussian	0.47	1.00	0.98	0.75	0.76	0.46	1.00	1.00	0.82	0.95

	Cross-correlation, east - west									
Model	A/Cr	A/TC	A/N	A/S	Cr/TC	Cr/N	Cr/S	TC/N	TC/S	N/S
SSB - sq exp	0.28	0.27	0.10	0.48	0.23	0.39	0.32	0.46	0.30	0.80
SSB - unif	0.29	0.27	0.05	0.45	0.25	0.35	0.32	0.44	0.31	0.79
Gaussian	0.03	0.09	0.41	0.32	0.00	0.12	0.09	0.53	0.03	0.59

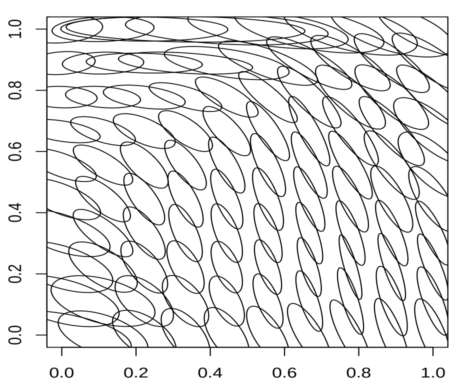
Figure 1: Data and results for simulated data set #1 using the nonstationary model with squared-exponential kernels, and for data set #2 using the nonstationary model with uniform kernels. The dots are observation locations, "Fitted values" is the posterior mean of the posterior mean of $\sum_{i=1}^{M} p_i(s_j)X(\phi_i)$, "bandwidth" is the posterior mean of B_i, and "standard deviation" is the posterior mean of σ.



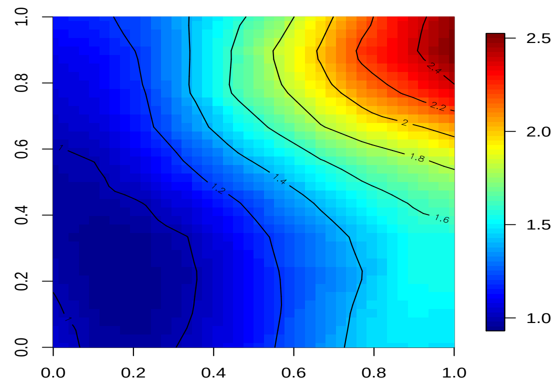
(a) Raw data for dataset #1.



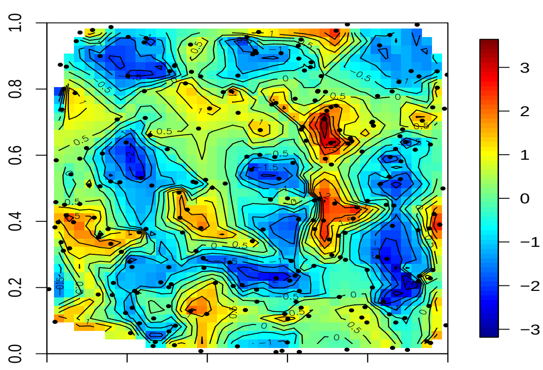
(b) Fitted values for dataset #1.



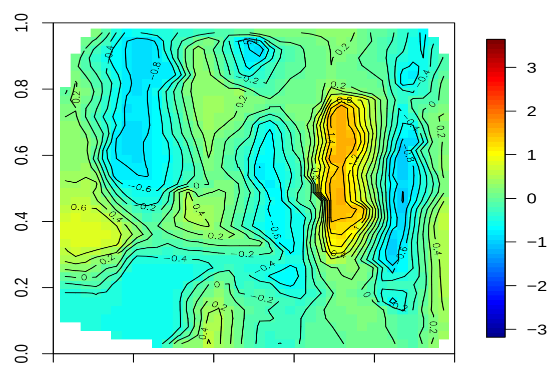
(c) Bandwidth for dataset #1.



(d) Standard deviation for dataset #1.



(e) Raw data for dataset #2.



(f) Fitted values for dataset #2.

Figure 2: Maps of monitoring locations and speciated $PM_{2.5}$ data. The two red triangles in the map with the monitoring locations, show the locations of the pair of points where the cross-correlation is evaluated.

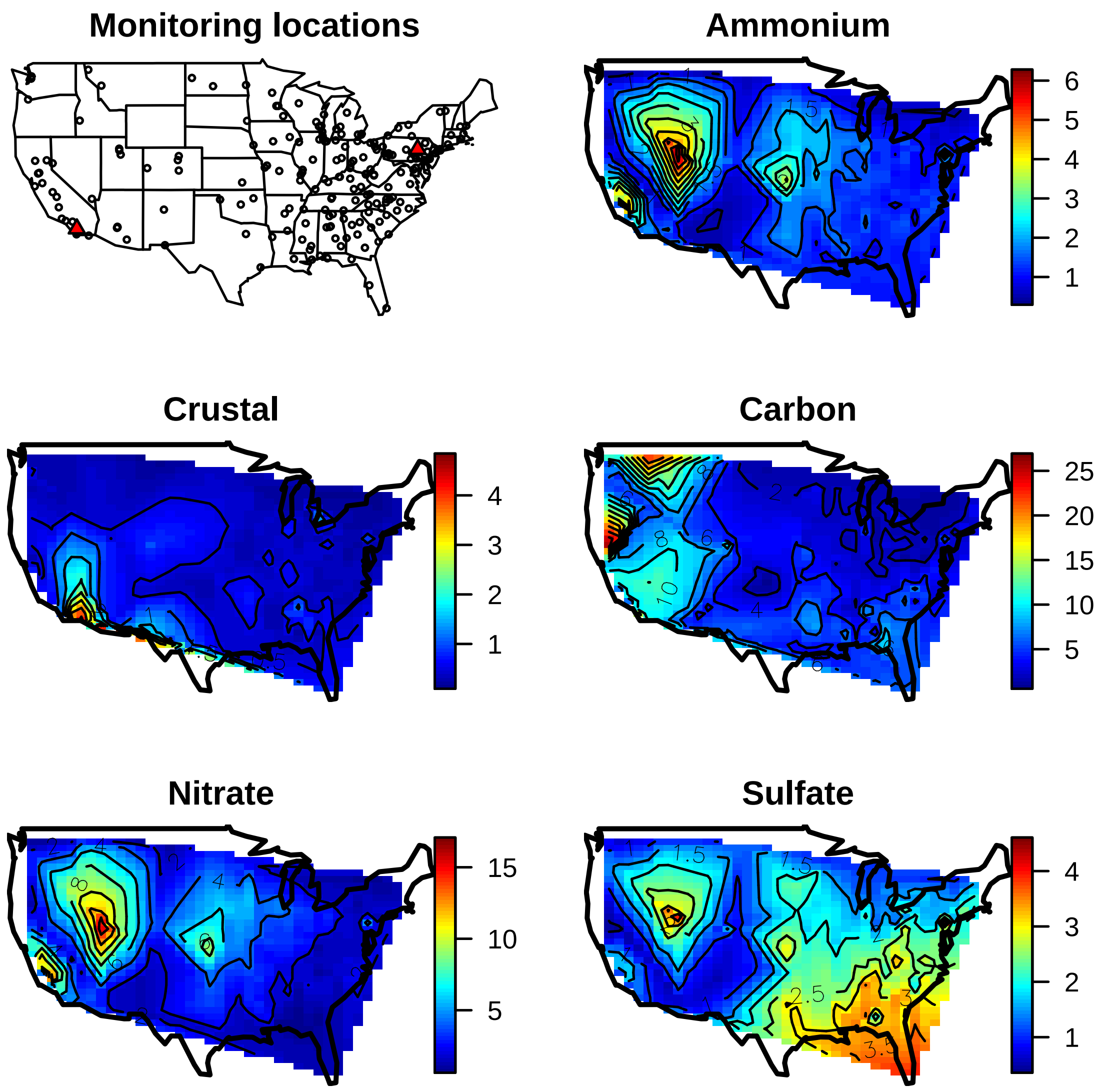
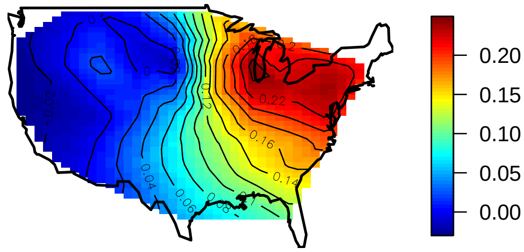
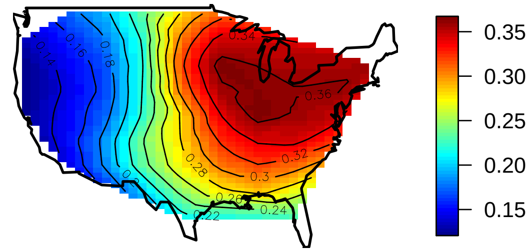


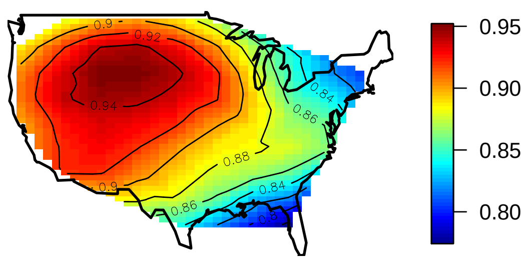
Figure 3: Posterior mean and standard deviation of the spatially-varying cross-correlation processes.



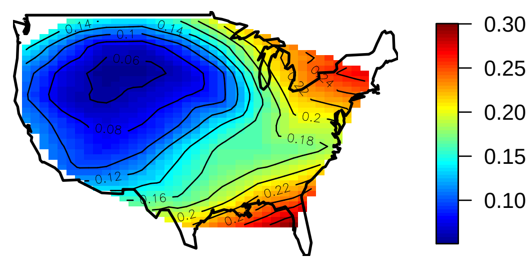
(a) Carbon and Sulfate, posterior mean



(b) Carbon and Sulfate, posterior sd



(c) Ammonium and Nitrate, posterior mean



(d) Ammonium and Nitrate, posterior sd

1 Appendix

A.1

Properly defined process prior

Kolmogorov existence theorem

We need to prove that the collection of finite-dimensional distributions introduced in (1) define a stochastic process $Y(s)$. We use the two Kolmogorov consistency conditions (symmetry under permutation, and dimensional consistency) to show that (1) defines a proper random process for Y.

Proposition 1. The collection of finite-dimensional distributions introduced in (1) properly define a stochastic process $Y(s)$. We use the two Kolmogorov consistency conditions: symmetry under permutation, and dimensional consistency, to define properly a random process for Y.

Proof of Proposition 1:

Symmetry under permutation.

Let $p_{i_j} = p(Y(s_j) = X(\phi_{i(s_j)}))$, where $\phi_{i(s_j)}$ is the centering knot of the kernel $i(s_j)$ in the representation of $F_{s_j}(Y)$ in (1), and let ϕ_{i_j} be an abbreviation for $\phi_{(i(s_j))}$. Then, $p_{i_1,\ldots,i_n}$ determine the site-specific joint selection probabilities. If $\pi(1),\ldots,\pi(n)$ is any permutation of $\{1,\ldots,n\}$, then we have

$$p_{i_{\pi(1)},\ldots,i_{\pi(n)}} = p(Y(s_{\pi(1)}) = X(\phi_{\pi(i_1)}),\ldots,Y(s_{\pi(n)}) = X(\phi_{\pi(i_n)}))$$

$$= p(Y(s_1) = X(\phi_{i_1}),\ldots,Y(s_n) = X(\phi_{i_n})) = p_{i_1,\ldots,i_n}, \tag{12}$$

since the observations are conditionally independent. Then,

$$
\begin{aligned}
& p(Y(s_1) \in A_1, \ldots, Y(s_n) \in A_n) \\
= & \sum_{i_1,\cdots,i_n} p(Y(s_1) = X(\phi_{i_1}), \ldots, Y(s_n) = X(\phi_{i_n}))\delta_{X(\phi(i_1))}(A_1) \ldots \delta_{X(\phi(i_n))}(A_n) \\
= & \sum_{i_1,\cdots,i_n} p_{i_1,\ldots,i_n} \delta_{X(\phi(i_1))}(A_1) \ldots \delta_{X(\phi(i_n))}(A_n) \\
= & \sum_{i_1,\cdots,i_n} p_{i_{\pi(1)},\ldots,i_{\pi(n)}} \delta_{X(\phi(i_{\pi(1)}))}(A_{\pi(1)}) \ldots \delta_{X(\phi(i_{\pi(n)}))}(A_{\pi(n)}) \\
= & p(Y(s_{\pi(1)}) \in A_{\pi(1)}, \ldots, Y(s_{\pi(n)}) \in A_{\pi(n)}).
\end{aligned}
$$

And, the *symmetry under permutation* condition holds.

Dimensional consistency.

$$
\begin{aligned}
& p(Y(s_1) \in (A_1), \ldots, Y(s_k) \in \mathcal{R}, \ldots, Y(s_n) \in (A_n)) \\
= & \sum_{(i_1,\ldots,i_n) \in \{1,2,\ldots\}^n} p_{i_1,\ldots,i_n} \delta_{X(\phi_{i_1})}(A_1) \cdots \delta_{X(\phi_{i_k})}(\mathcal{R}) \cdots \delta_{X(\phi_{i_n})}(A_n) \\
= & \sum_{(i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_n) \in \{1,2,\ldots\}^{n-1}} \delta_{X(\phi_{i_1})}(A_1) \cdots \delta_{X(\phi_{i_{k-1}})}(A_{k-1}) \delta_{X(\phi_{i_{k+1}})}(A_{k+1}) \\
& \cdots \delta_{X(\phi_{i_n})}(A_n) \sum_{j=1}^{\infty} p_{i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_n} \\
= & p(Y(s_1) \in (A_1), \ldots, Y(s_{k-1}) \in A_{k-1}, Y(s_{k+1}) \in A_{k+1}, \ldots, Y(s_n) \in (A_n)). \qquad (13)
\end{aligned}
$$

In (13), we need

$$
p_{i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_n} = \sum_{j=1}^{\infty} p_{i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_n}
$$

which holds by Fubini Theorem and the fact that X is a properly defined Gaussian process.

A.2

Proof of Theorem 1.

The covariance function C of the underlying process X has a first order derivative, C'. We introduce a Taylor expansion for C with a Lagrange remainder term,

$$C(|\phi_{i_1} - \phi_{i_2}|) = C(|s - s'|) + C'(\psi_{i_1,i_2})\varepsilon_{i_1,i_2}, \tag{14}$$

where $\varepsilon_{i_1,i_2} = (|\phi_{i_1} - \phi_{i_2}| - |s - s'|)$ and ψ_{i_1,i_2} is in between $|s - s'|$ and $|\phi_{i_1} - \phi_{i_2}|$.

Assuming that s and s' lie on the support of the kernels K_{i_1} and K_{i_2}, respectively, i.e. $|\phi_{i_1} - s| < \epsilon_{i_1}$ and $|\phi_{i_2} - s'| < \epsilon_{i_2}$. We have,

$$\varepsilon_{i_1,i_2} \leq ||\phi_{i_1} - \phi_{i_2}| - |s - s'|| \leq |(\phi_{i_1} - \phi_{i_2}) - (s - s')| \leq \epsilon_{i_1} + \epsilon_{i_2} \leq 2\epsilon,$$

and,

$$\varepsilon_{i_1,i_2} \geq -||\phi_{i_1} - \phi_{i_2}| - |s - s'|| \geq -|(\phi_{i_1} - \phi_{i_2}) - (s - s')| \geq -(\epsilon_{i_1} + \epsilon_{i_2}) \geq -2\epsilon,$$

Thus, $-2\epsilon \leq \varepsilon_{i_1,i_2} \leq 2\epsilon$.

Let $p(s)$ be the potentially infinite vector with all the probabilities masses $p_i(s)$ in $F_s(Y)$. The conditional covariance of the data process Y is written in terms of the covariance C of X,

$$\text{cov}(Y(s), Y(s')|p(s), p(s'), C) = \sum_{i_1 i_2} p_{i_1}(s)p_{i_2}(s')C(|\phi_{i_1} - \phi_{i_2}|),$$

since the kernels all have compact support, the expression above is the same as

$$\sum_{i_1,i_2;|\phi_{i_1}-s|<\epsilon_{i_1},|\phi_{i_2}-s'|<\epsilon_{i_2}} p_{i_1}(s)p_{i_2}(s')C(|\phi_{i_1} - \phi_{i_2}|).$$

Using the Taylor approximation in (14), the $\text{cov}(Y(s), Y(s')|p(s), p(s'), C)$ can be written

$$C(|s - s'|)\left[\sum_{i_1} p_{i_1}(s)\right]\left[\sum_{i_2} p_{i_2}(s')\right] + \sum_{i_1,i_2;|\phi_{i_1}-s|<\epsilon_{i_1},|\phi_{i_2}-s'|<\epsilon_{i_2}} p_{i_1}(s)p_{i_2}(s')C'(\psi_{i_1,i_2})\varepsilon_{i_1,i_2}.$$

Since C' is nonnegative and $\varepsilon_{i_1,i_2} \in (-2\epsilon, 2\epsilon)$, we have that for $J_{i_1,i_2} = \{(i_1,i_2); |\phi_{i_1} - s| < \epsilon_{i_1}, |\phi_{i_2} - s'| < \epsilon_{i_2}\}$,

$$\begin{aligned} -2\epsilon \sum_{i_1,i_2 \in J_{i_1,i_2}} p_{i_1}(s) p_{i_2}(s') C'(\psi_{i_1,i_2}) &\leq \sum_{(i_1,i_2)\in J_{i_1,i_2}} p_{i_1}(s) p_{i_2}(s') C'(\psi_{i_1,i_2}) \varepsilon_{i_1,i_2} \\ &\leq 2\epsilon \sum_{(i_1,i_2)\in J_{i_1,i_2}} p_{i_1}(s) p_{i_2}(s') C'(\psi_{i_1,i_2}), \end{aligned} \tag{15}$$

where,

$$2\epsilon \sum_{i_1,i_2 \in J_{i_1,i_2}} p_{i_1}(s) p_{i_2}(s') C'(\psi_{i_1,i_2}) \longrightarrow_{\epsilon \to 0} 0,$$

because C' is bounded and the sum of probability masses is always bounded by 1. The sum of probability masses would be always bounded by 1, because by proposition 1 the prior process is properly defined.

Thus, we obtain

$$\text{cov}(Y(s), Y(s') | p(s), p(s'), C) \longrightarrow_{\epsilon \to 0} C(|s - s'|) \left[\sum_{i_1} p_{i_1}(s) \right] \left[\sum_{i_2} p_{i_2}(s') \right] = C(|s - s'|).$$

Therefore, the conditional covariance of the data process Y approximates the covariance C of the underlying process X as the bandwiths of the kernel functions go to zero.

A.3

Proof of Theorem 2.

The cross-covariance function $C_{1,2}(s, s')$ of the underlying process $X = (X_1, X_2)$ has first order partial derivatives $\delta C_{1,2}(s, s')/\delta s$ and $\delta C_{1,2}(s, s')/\delta s'$. We introduce a Taylor expansion for $C_{1,2}$ with a Lagrange remainder term,

$$C_{1,2}(\phi_{i_1}, \phi_{i_2}) = C_{1,2}(s, s') + (\phi_{i_1} - s) \left[\delta C_{1,2}(s, s')/\delta s\right]_{(s,s')=(\psi_{i_1},\psi_{i_2})} + (\phi_{i_2} - s') \left[\delta C_{1,2}(s, s')/\delta s'\right]_{(s,s')=(\psi_{i_1},\psi_{i_2})}$$

where ψ_{i_1} is in between s and ϕ_{i_1}, and ψ_{i_2} is in between s' and ϕ_{i_2}. We follow the same steps as in Theorem 1, to bound the first order term of the Taylor expansion, and obtain that

$$\text{cov}(Y_1(s), Y_2(s')|p_1(s), p_2(s'), C_{1,2}) \longrightarrow_{\epsilon \to 0} C_{1,2}(s, s'),$$

where $p_1(s)$, and $p_2(s')$ are the potentially infinite dimensional vectors with the all probability masses in the spatial stick-breaking prior processes $F_s(Y_1)$, and $F_{s'}(Y_2)$ respectively.

A.4

Proof of Theorem 3.

Let $\psi(t, s)$ the characteristic function of $Y(s)$. Then,

$$\begin{aligned} & \psi(t, s_1) - \psi(t, s_2) = E_Y[\exp\{itY(s_1)\}] - E_Y[\exp\{itY(s_2)\}] \\ = \; & E_X\left\{\sum_j p_j(s_1)\exp\{itX(\phi_j)\}\right\} - E_X\left\{\sum_j p_j(s_2)\exp\{itX(\phi_j)\}\right\} \\ = \; & E_X\left\{\sum_j (p_j(s_1) - p_j(s_2))\exp\{itX(\phi_j)\}\right\} \longrightarrow_{|s_1 - s_2| \to 0} 0. \end{aligned} \tag{16}$$

Then, $F_{s_1}(Y)$ converges to $F_{s_2}(Y)$ for any locations s_1, s_2, as long as $|s_1 - s_2| \to 0$.

A.5

Proof of Theorem 4.

The probability masses $p_i(s)$ in (1) are $p_i(s) = V_i K_i(s) \prod_{j=1}^{i-1}(1 - V_j K_j(s))$. Since the bandwiths ϵ_i converge uniformly to zero, then, $p_i(s) \to 1$, as $|\phi_i - s| \to 0$, where ϕ_i is the knot of kernel K_i. This holds because $\sum_j p_j(s) = 1$ a.s. (since the process Y is properly defined).

Assume now $|s_1 - s_2| \to 0$, we need to prove that $Y(s_1)$ converges a.s. to $Y(s_2)$.

Let ϕ_1 and ϕ_2 satisfy,

$$|\phi_1 - s_1| \to 0, \quad \text{and} \quad |\phi_2 - s_2| \to 0. \tag{17}$$

Thus, we obtain that with probability 1, $Y(s_1)$ converges to $X(\phi_1)$, and $Y(s_2)$ to $X(\phi_2)$.

Since $|s_1 - s_2| \to 0$, and $|\phi_1 - \phi_2| \leq |\phi_1 - s_1| + |\phi_2 - s_2| + |s_1 - s_2|$. Then, by (17)

$$|\phi_1 - \phi_2| \to 0. \tag{18}$$

We have,

$$|Y(s_1) - Y(s_2)| \leq |Y(s_1) - X(\phi_1)| + |Y(s_2) - X(\phi_2)| + |X(\phi_1) - X(\phi_2)|,$$

where $|Y(s_1) - X(\phi_1)| \to 0$ a.s., as $|s_1 - \phi_1| \to 0$; $|Y(s_2) - X(\phi_2)| \to 0$ a.s., as $|s_2 - \phi_2| \to 0$; and since X is a.s. continuous, $|X(\phi_1) - X(\phi_2)| \to 0$ a.s., as $|\phi_1 - \phi_2| \to 0$ (which holds by 18).

Therefore, $|Y(s_1) - Y(s_2)| \to 0$ a.s.