

Generalized spatial Dirichlet process models

BY JASON A. DUAN

School of Management, Yale University, New Haven, Connecticut 06520-8200, U.S.A.

jd522@som.yale.edu

MICHELE GUINDANI

*Department of Mathematics and Statistics, University of New Mexico, Albuquerque,
NM 87131, U.S.A.*

michele@stat.unm.edu

AND ALAN E. GELFAND

*Institute of Statistics and Decision Sciences, Duke University, Durham,
North Carolina 27708-0251, U.S.A.*

alan@stat.duke.edu

SUMMARY

Many models for the study of point-referenced data explicitly introduce spatial random effects to capture residual spatial association. These spatial effects are customarily modelled as a zero-mean stationary Gaussian process. The spatial Dirichlet process introduced by [Gelfand et al. \(2005\)](#) produces a random spatial process which is neither Gaussian nor stationary. Rather, it varies about a process that is assumed to be stationary and Gaussian. The spatial Dirichlet process arises as a probability-weighted collection of random surfaces. This can be limiting for modelling and inferential purposes since it insists that a process realization must be one of these surfaces. We introduce a random distribution for the spatial effects that allows different surface selection at different sites. Moreover, we can specify the model so that the marginal distribution of the effect at each site still comes from a Dirichlet process. The development is offered constructively, providing a multivariate extension of the stick-breaking representation of the weights. We then introduce mixing using this generalized spatial Dirichlet process. We illustrate with a simulated dataset of independent replications and note that we can embed the generalized process within a dynamic model specification to eliminate the independence assumption.

Some key words: Dirichlet process mixing; Dynamic model; Latent process; Non-Gaussian; Nonstationary; Stick breaking.

1. INTRODUCTION

Distributional modelling for point-referenced spatial data usually introduces spatial random effects to capture residual spatial association. These spatial effects are often modelled as a zero-mean stationary Gaussian process. The stationarity or the Gaussian assumption may be inappropriate. Flexible and computationally tractable modelling to eliminate the stationarity assumption includes the spatially varying kernel approach of [Higdon et al. \(1999\)](#) and the local stationarity approach described in an unpublished North Carolina State University technical report by M. Fuentes and R.L. Smith; however, both are still within the setting of Gaussian

processes. The fundamental paper of [Sampson & Guttorp \(1992\)](#) introduces a nonparametric specification for the covariance function, as does follow-up work by [Damian et al. \(2001\)](#) and [Schmidt & O'Hagan \(2003\)](#) but all still employ a Gaussian Process in the likelihood.

The Gaussian assumption can be criticized when the spatial variability is attributable to more than one latent process, so that, for example, a mixture of Gaussian processes might be more appropriate; see [Brown et al. \(2003\)](#) for a recent example or [Palacios & Steel \(2006\)](#) for the development of a class of models able to cope with heavy tail behaviour. Recently, [Gelfand et al. \(2005\)](#) proposed a spatial Dirichlet process mixture model which adopts a stationary and Gaussian base measure. However, the resulting random stochastic process is nonstationary and its joint finite-dimensional distributions are not normal. The use of this specification to model the distribution of the spatial component in a spatial random effects model leads to a fully Bayesian semiparametric approach that, for fitting purposes, relies on well-known results and algorithms developed for Dirichlet process mixing; see, among others, [Escobar & West \(1995\)](#) and [MacEachern & Müller \(1998\)](#).

Since the spatial Dirichlet process is essentially a Dirichlet process defined on a space of surfaces, with probability one its realizations are discrete probability measures with countable support ([Ferguson, 1973](#); [Sethuraman, 1994](#)). Thus, mixing against a Gaussian kernel yields an error specification that can be characterized as a countable location mixture of normals. However, the spatial Dirichlet process insists that, given the countable collection of surfaces, we actually sample only one of them and then the process realization is that surface. In this paper, we introduce a random distribution for the spatial effects that allows different surface selection at different sites. Moreover, we can specify the model to preserve the property that the marginal distribution of the effect at each site still comes from a Dirichlet process. We refer to these new random probability measures as generalized spatial Dirichlet process models; the customary Dirichlet process specification is included as a special case.

Our formulation envisages a setting in which few random surfaces are needed to achieve an adequate random spatial effects model; the novel modelling ingredient is a latent covariate that determines surface selection. Hence, resulting realizations are complex functions of these surfaces. As an illustrative example, in brain imaging, with regard to the neurological activity level, researchers imagine healthy brain images, i.e. surfaces, as well as diseased or impaired brain images. For an actual image only a portion of the brain will be impaired; we can envisage surface selection according to a latent indicator of where the brain is diseased.

An advantage of working with generalized spatial Dirichlet processes in such settings is that we need not define the latent covariate: we need not specify the number of components or surfaces. The posterior inference will provide information about this and may illuminate the interpretation of the covariate. Lastly, it is evident that the spatial Dirichlet process is not well equipped to handle such situations. For instance, in the brain imaging illustration, the spatial Dirichlet process is likely to require a different random surface for each image to be analysed.

Replicated observations are typically needed for a fully nonparametric approach ([Sampson & Guttorp, 1992](#); [Damian et al., 2001](#)) and this is true in our case as well. However, with replications that arise across discretized time periods, we can shed the independence assumption by embedding our methodology within a dynamic model, retaining the temporal dependence. These methods allow the possibility to infer about the random distribution function that is operating at any given location, at any time, in the region. Nonparametric spatial prediction can be pursued not only at new locations for each replicate, but more generally through the generation of an entire new predictive surface at a future time. Also, though we develop our model in the context of spatial data, the theory is general and can be used when our responses are indexed by covariates in usual regression settings. Hence, we offer an alternative for most of the

problems where mixtures of products of Dirichlet processes (Cifarelli & Regazzini, 1978) and/or the dependent Dirichlet processes of an unpublished Ohio State University technical report of S. MacEachern have been employed; see for example De Iorio et al. (2004).

We are aware of only two other recent approaches that also consider mixture models for spatial data where the weights are allowed to vary across locations. Fernandez & Green (2002) confine their attention to Markov random fields over lattices and Poisson-distributed data, and only the weights in the mixture vary from one location to another. We work with general point-referenced data allowing both the weights and the parameters of the mixed distribution to vary spatially. Griffin & Steel (2006) present an implementation of the dependent Dirichlet process using Sethuraman's constructive representation, providing a random marginal distribution at each site. The marginal stick-breaking weights are the same at each location, but they are randomly permuted according to the realizations of a latent point process, so that at each site the resulting weights are assigned to different surfaces, inducing spatial dependence. Instead, we present a process model for joint surface selection for any number and choice of locations, and also allow the marginal components to vary in space. Our construction may also be viewed as a multivariate stick-breaking specification. Moreover, the closeness between the random distributions is ruled directly by the topology of the space, rather than by realizations of an underlying point process.

2. A BRIEF REVIEW OF THE SPATIAL DIRICHLET PROCESS

Denote the stochastic process by $\{Y(s) : s \in D\}$, $D \subseteq R^d$, and let $s^{(n)} = (s_1, \dots, s_n)$ be the specific distinct locations in D where observations are collected. Assume that we have replicate observations at each location so that the full dataset consists of the collection of vectors $Y_t = \{Y_t(s_1), \dots, Y_t(s_n)\}^T$, $t = 1, \dots, T$.

For a measurable space (Θ, \mathcal{B}) , the Dirichlet process (Ferguson, 1973, 1974) specifies random distributions on Θ denoted by $DP(\nu G_0)$ where $\nu > 0$ is a scalar precision parameter and G_0 is a specified base distribution defined on (Θ, \mathcal{B}) . A random distribution function on (Θ, \mathcal{B}) arising from $DP(\nu G_0)$ is almost surely discrete and admits the representation $\sum_{l=1}^{\infty} p_l \delta_{\theta_l^*}$, where δ_z denotes a point mass at z , $p_1 = q_1$, $p_l = q_l \prod_{r=1}^{l-1} (1 - q_r)$, $l = 2, 3, \dots$, with q_r , $r = 1, 2, \dots$, independently and identically distributed as $Be(1, \nu)$. The θ_l^* 's are independent and identically distributed as G_0 and are also independent of the q_r 's, $r = 1, 2, \dots$ (Sethuraman, 1994).

To model $Y_D \equiv \{Y(s) : s \in D\}$, one can conceptually replace θ_l^* with a realization of a random field $\theta_{l,D}^* = \{\theta_l^*(s) : s \in D\}$. The process G_0 might be a stationary Gaussian process with each $\theta_{l,D}^*$ being a realized surface over D from G_0 . The resulting random process or distribution, G , for Y_D is denoted by $\sum_{l=1}^{\infty} p_l \delta_{\theta_{l,D}^*}$. The interpretation is that, for $s^{(n)}$ as above, G induces a random probability measure $G^{(n)}$ on the space of distribution functions for $\{Y(s_1), \dots, Y(s_n)\}$ with $G^{(n)} \sim DP(\nu G_0^{(n)})$, where $G_0^{(n)}$ is the n -variate distribution for $\{Y(s_1), \dots, Y(s_n)\}$ induced by G_0 . The representation of G clarifies that the spatial Dirichlet process is a nonstationary, non-Gaussian process that is centred around a stationary process. In addition, G induces a random distribution $G\{Y(s)\}$ for each s , and hence the set $\mathcal{G}_D \equiv \{G\{Y(s)\} : s \in D\}$. For a stationary G_0 , the choice of the covariance function determines whether process realizations are almost surely continuous (Kent, 1989). Again, from the representation of G , the continuity of $\theta_{l,D}^*$ implies that $G\{Y(s_i)\}$ and $G\{Y(s_j)\}$ are such that the difference between them tends to 0 almost surely, as $\|s_i - s_j\| \rightarrow 0$.

Mixing a pure error process with the variance τ^2 with respect to G creates a random process F that has continuous support; that is, if k is a density function on R^1 , for any s , $f\{Y(s) | G, \tau^2\} = \int k\{Y(s) - \theta(s) | \tau^2\} G\{d\theta(s)\}$, i.e. $Y(s) = \theta(s) + \epsilon(s)$, where $\theta(s)$ arises from the above spatial Dirichlet process prior model and $\epsilon(s)$ is a pure error term. When k is the $N(0, \tau^2)$ density, the joint

density of $Y = \{Y(s_1), \dots, Y(s_n)\}^T$, given $G^{(n)}$ and τ^2 , using the almost sure representation of $G^{(n)}$, is almost surely of the form $\sum_{l=1}^\infty p_l N_n(Y \mid \theta_l^*, \tau^2 I_n)$, i.e. a countable location mixture of normals.

3. THE GENERALIZED SPATIAL DIRICHLET PROCESS MODEL

3.1. Model details

In [Gelfand et al. \(2005\)](#), the spatial Dirichlet process is defined using a countable collection of random surfaces over D generated by a zero-mean base spatial process. For any collection of n locations, the joint distribution uses the same set of stick-breaking probabilities, inducing common surface selection for all locations in the collection; it is not possible to capture the situation in which spatial effects can be selected from different surfaces at different locations. This limitation is common to other recent work relating to dependent Dirichlet processes; see, for example, [De Iorio et al. \(2004\)](#). We introduce a random distribution for the spatial effects such that the surface selection can vary with the location and the joint selection of surfaces for the n locations can vary with the choice of locations. Moreover, we can still preserve the property that the marginal distribution at each location comes from a usual univariate Dirichlet process.

Again, we start with a base random field G_0 , which for convenience we take to be stationary and Gaussian. Again, we draw a countable collection of independent realizations from G_0 . Then, we define a random probability measure G on the space of surfaces over D as that measure whose finite-dimensional distributions almost surely have the following representation: for any set of locations $(s_1, \dots, s_n) \in D$, and any collection of sets $\{A_1, \dots, A_n\}$ in $\mathcal{B}(\mathbb{R})$,

$$\text{pr}\{Y(s_1) \in A_1, \dots, Y(s_n) \in A_n\} = \sum_{i_1=1}^\infty \cdots \sum_{i_n=1}^\infty p_{i_1, \dots, i_n} \delta_{\theta_{i_1}^*(s_1)}(A_1) \cdots \delta_{\theta_{i_n}^*(s_n)}(A_n), \quad (1)$$

where the θ_j^* 's are independent and identically distributed as G_0 , i_j is an abbreviation for $i(s_j)$, $j = 1, \dots, n$, and the weights $\{p_{i_1, \dots, i_n}\}$ determine the site-specific joint selection probabilities. In fact, conditionally on the locations, the $\{p_{i_1, \dots, i_n}\}$ have a distribution defined on the infinite-dimensional simplex $\mathbb{P} = \{p_{i_1, \dots, i_n} \geq 0 : \sum_{i_1=1}^\infty \cdots \sum_{i_n=1}^\infty p_{i_1, \dots, i_n} = 1\}$ independent of that for the θ^* 's.

The weights need to satisfy a consistency condition in order to define properly a random process for $Y(\cdot)$. To be specific, it is necessary that, for any set of locations (s_1, \dots, s_n) , $n \in \mathbb{N}$, and for all $k \in \{1, \dots, n\}$,

$$p_{i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_n} = p_{i_1, \dots, i_{k-1}, j, i_{k+1}, \dots, i_n} \equiv \sum_{j=1}^\infty p_{i_1, \dots, i_{k-1}, j, i_{k+1}, \dots, i_n}. \quad (2)$$

In addition, we insist that the weights satisfy a continuity property; we want the random laws associated with the locations s_1 and s_2 near to each other to be similar. Equivalently, for the locations s and s_0 , as $s \rightarrow s_0$, $p_{i_1, i_2} = \text{pr}\{Y(s) = \theta_{i_1}^*(s), Y(s_0) = \theta_{i_2}^*(s_0)\}$ tends to the marginal probability $p_{i_2} = \text{pr}\{Y(s_0) = \theta_{i_2}^*(s_0)\}$ when $i_1 = i_2$, and to 0 otherwise. Analogously, if we consider three locations (s_1, s_2, s_3) , when s_3 is close to s_2 say, we require p_{i_1, i_2, i_3} to be close to p_{i_1, i_2} if $i_2 = i_3$ and to 0 otherwise. Extension to the n locations is clear and we refer to this property simply as almost sure continuity of the weights. The name is suggested by the almost sure continuity of the paths of a univariate spatial process, as defined in [Kent \(1989\)](#) or [Banerjee et al. \(2003\)](#). If we also assume the random field G_0 to be almost surely continuous, a univariate spatial process $\theta(s)$, $s \in D$, being said to be almost surely continuous at a point s_0 if $\theta(s) \rightarrow \theta(s_0)$ with probability one as $\|s - s_0\| \rightarrow 0$, we are able to establish the following proposition whose proof is given in Appendix 1.

PROPOSITION 1. Let $\{Y(s), s \in D\}$ be a random field, whose random finite-dimensional distributions are given by (1) for all $n \in \mathbb{N}$. If the set of weights $\{p_{i_1, \dots, i_n}\}$ and the base random field G_0 are almost surely continuous, then, for all $s_0 \in D$, $Y(s)$ converges weakly to $Y(s_0)$ with probability one as $\|s - s_0\| \rightarrow 0$.

In fact, the proof demonstrates almost sure convergence of the random probability measures. Note that Proposition 1 is an extension to our case of analogous results given in S. MacEachern’s technical report and in Gelfand et al. (2005).

Conditional on the realized distribution G , the process has first and second moments

$$E\{Y(s)|G\} = \sum_{l=1}^{\infty} p_l(s) \theta_l^*(s)$$

$$\text{var}\{Y(s)|G\} = \sum_{l=1}^{\infty} p_l(s) \theta_l^{*2}(s) - \left\{ \sum_{l=1}^{\infty} p_l(s) \theta_l^*(s) \right\}^2,$$

and, for a pair of sites s_i, s_j ,

$$\text{cov}\{Y(s_i), Y(s_j)|G\} = \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} p_{l,m}(s_i, s_j) \theta_l^*(s_i) \theta_m^*(s_j) - \left\{ \sum_{l=1}^{\infty} p_l(s_i) \theta_l^*(s_i) \right\} \left\{ \sum_{m=1}^{\infty} p_m(s_j) \theta_m^*(s_j) \right\}. \tag{3}$$

With almost surely continuous realizations from the base process and of the weights, (3) shows that the generalized spatial Dirichlet process is mean-square continuous.

Suppose G_0 is a zero-mean stationary Gaussian process with finite variance σ^2 and correlation function $\rho_\phi(s_i - s_j)$. Then $E\{Y(s)\} = 0$ and $\text{var}\{Y(s)\} = \sigma^2$ but

$$\text{cov}\{Y(s_i), Y(s_j)\} = \sigma^2 \rho_\phi(s_i - s_j) \sum_{l=1}^{\infty} E\{p_{ll}(s_i, s_j)\}.$$

Note that $\sum_{l=1}^{\infty} E\{p_{ll}(s_i, s_j)\} < 1$, unless $p_{ll}(s_i, s_j) = 0$, for $l \neq l'$, as it is in Gelfand et al. (2005) or, more generally, in the single- p -dependent Dirichlet process discussed in S. MacEachern’s technical report. The implication is that, in allowing local surface selection, for a given pair of locations, we expect weaker spatial association under our generalization than under the spatial Dirichlet process.

3.2. Mixing using a generalized spatial Dirichlet process

The generalized spatial Dirichlet process will be used to model the distribution of the spatial component $\theta(s)$ in a random effect model of the type $Y(s) = u(s) + \theta(s) + \varepsilon(s)$, where $u(s)$ is a constant mean term, typically assumed to be a regression term $X(s)^T \beta$ for some vector of covariates $X(s)$ and some vector of parameters β , and $\varepsilon(s)$ is a Gaussian pure random error component with mean zero and variance τ^2 . Again, denote by $G^{(n)}$ the finite-dimensional distributions defined by (1), for any finite set of locations $s^{(n)} = (s_1, \dots, s_n)$, $n \in \mathbb{N}$. Then, since $G^{(n)}$ is almost surely discrete, with probability one the joint distribution of $Y = \{Y(s_1), \dots, Y(s_n)\}^T$, given $G^{(n)}$, u and τ^2 , is

$$f(Y|G^{(n)}, u, \tau^2) = \sum_{i_1=1}^{\infty} \cdots \sum_{i_n=1}^{\infty} p_{i_1, \dots, i_n} N_n(Y | \theta_{i_1, \dots, i_n} + u, \tau^2 I_n),$$

where $u = \{u(s_1), \dots, u(s_n)\}^T$ and, for simplicity, we have suppressed the locations in p_{i_1, \dots, i_n} and set the vector $\theta_{i_1, \dots, i_n} = \{\theta_{i_1}(s_1), \dots, \theta_{i_n}(s_n)\}^T$. Immediately, we obtain $E(Y|G^{(n)}, u, \tau^2) = \sum_{i_1=1}^\infty \dots \sum_{i_n=1}^\infty p_{i_1, \dots, i_n} \theta_{i_1, \dots, i_n} + u$ and $\Sigma_{Y|G^{(n)}, \mu, \tau^2} = \tau^2 I_n + \Sigma_\theta^s$, where $(\Sigma_\theta^s)_{i,j} = \text{cov}\{\theta(s_i), \theta(s_j)|G^{(n)}\}$ is given by (3).

If the mean vector u describes a continuous surface over D , it is easy to prove that a statement analogous to Proposition 1 holds for the convolved process Y . In fact, the normal density is a bounded continuous function of the mean. Then the bounded convergence theorem applies and, together with almost sure convergence of the random probability measures $G^{(n)}$ proved in Proposition 1, this implies that, with probability 1, $Y(s)$ converges weakly to $Y(s_0)$ for any $s, s_0 \in D$, as $\|s - s_0\| \rightarrow 0$.

4. THE SPATIALLY VARYING PROBABILITIES MODEL

We now turn to an illustrative specification of p_{i_1, \dots, i_n} for any choice of n and s_1, \dots, s_n . We provide a constructive approach which can be viewed as multivariate stick-breaking, employing computationally convenient Gaussian thresholding. Attractively, with such thresholding, we never need to work with the p 's. In fact, we assume that $\{Z_l(s), s \in D, l = 1, 2, \dots\}$ is a countable collection of independent stationary Gaussian random fields on D having variance 1 and correlation function $\rho_Z(\cdot, \eta)$. We further assume that the mean of the l th process, $\mu_l(s)$, is unknown and we put a convenient prior on it, so that the distribution of $Z_l(s)$ can be viewed as random. With these assumptions, given s_1, \dots, s_n , we let

$$\begin{aligned}
 p_{i_1, \dots, i_n} = \text{pr}\{ & Z_1(s_1) < 0, \dots, Z_{i_1-1}(s_1) < 0, Z_{i_1}(s_1) \geq 0; \\
 & Z_1(s_2) < 0, \dots, Z_{i_2-1}(s_2) < 0, Z_{i_2}(s_2) \geq 0; \dots; \\
 & Z_1(s_n) < 0, \dots, Z_{i_n-1}(s_n) < 0, Z_{i_n}(s_n) \geq 0 | \{\mu_l(s_i)\}\}.
 \end{aligned}
 \tag{4}$$

Since the distribution of the Z 's is random, so is the induced distribution of the p 's.

To clarify the multivariate stick-breaking implicit in (6) consider the stochastic process $\{\delta_{A_l(s)}^*, s \in D, l = 1, 2, \dots\}$ defined by $\delta_{A_l(s)}^* = 1$ if $Z_l(s) \in A_l(s)$, and $\delta_{A_l(s)}^* = 0$ if $Z_l(s) \notin A_l(s)$, where $\{Z_l(s), s \in D, l = 1, 2, \dots\}$ is a latent random field. Let

$$q_{l, u_1, \dots, u_n}(s_1, \dots, s_n) = \text{pr}\{\delta_{A_l(s_1)}^* = u_1, \dots, \delta_{A_l(s_n)}^* = u_n | \delta_{A_l(s_j)}^* = 0, i < l, j = 1, \dots, n\}.$$

If we choose $A_l(s) = \{Z_l(s) \geq 0\}$, it follows that

$$q_{l, u_1, \dots, u_n}(s_1, \dots, s_n) = \text{pr}\{\delta_{\{Z_l(s_1) \geq 0\}}^* = u_1, \dots, \delta_{\{Z_l(s_n) \geq 0\}}^* = u_n | \mu_l(s_1), \dots, \mu_l(s_n)\},$$

because of the independence of the processes $\{Z_l(s)\}$ over the index l . For example, for $n = 2$, we obtain $q_{l, 0, 1} = \text{pr}\{Z_l(s_1) < 0, Z_l(s_2) \geq 0 | \mu_l(s_1), \mu_l(s_2)\}$. If the $\mu_l(s)$ surfaces are independent, $l = 1, 2, \dots$, then so are the $q_{l, u_1, \dots, u_n}(s_1, \dots, s_n)$'s.

Since $Z_l(s)$ is assumed to be Gaussian, at any location s we obtain

$$q_{l, 1}(s) = \text{pr}\{Z_l(s) \geq 0\} = 1 - \Phi\{-\mu_l(s)\} = \Phi\{\mu_l(s)\},$$

where $\Phi(\cdot)$ denotes the univariate standard normal distribution function. If the $\mu_l(s)$ are such that the $\Phi\{\mu_l(s)\}$ are independent $\text{Be}(1, \nu), l = 1, 2, \dots$, then, for each s , the marginal distribution of $\theta(s)$ is a Dirichlet process with probabilities that vary with location. If the $\mu_l(s)$ are a continuous realization, $Z_l(s)$ will resemble $Z_l(s')$ when s is close to s' . For the special case in which $\mu_l(s) = \mu_l$, for all s , with $\Phi(\mu_l)$ independent $\text{Be}(1, \nu)$ then, marginally, the $\theta(s)$ follow a Dirichlet process where the weights are the same for each s but the marginal distributions are not the same since $\theta_l^*(s) \neq \theta_l^*(s')$.

The impact of using a thresholding function on prediction is discussed in §6. In this regard, Gaussian modelling has been routinely used, for example in the binary regression setting (Albert & Chib, 1993). Moreover, as a model for second-stage random effects, there will be little posterior sensitivity to this choice.

Following the discussion preceding Proposition 1, we require two properties for this construction: the random finite-dimensional distribution $G^{(n)}$ should satisfy the Kolmogorov consistency condition and the continuity property should be satisfied, in that, if location s is near s' , the probability of choosing the same sample surface for s and s' is high. In Propositions A1 and A2 in Appendix 1, we prove that these conditions are satisfied. Note, however, that (6) is not invariant to simultaneous permutation of the indices of the i_j and s_j , as would be the case in the usual Dirichlet process. However, the assumed exchangeability of the surfaces ensures that the distributions for the random p 's are exchangeable.

Finally, spatially varying weights have recently been considered by Griffin & Steel (2006), who work in the framework of dependent Dirichlet processes. They proceed from the assumption that the distribution of a DP(νG_0) is unaffected by a location-specific permutation, $\{\pi(s), s \in D\}$, of the atoms $\{\theta_l^*(\cdot), q_l(\cdot), l = 1, 2, \dots\}$ in Sethuraman's constructive representation. In particular, when $n = 2$, for any integers l and m , under their prior

$$\text{pr}\{Y(s_i) = \theta_l^*(s_i), Y(s_j) = \theta_m^*(s_j)\} = \int p_l(s_i) p_m(s_j) dH\{\pi(s_i), \pi(s_j)\},$$

where $H\{\cdot\}$ is the distribution of the permutation field at the two locations. We define a joint random distribution for any grouping of the locations $(s_1, \dots, s_n), n = 1, 2, \dots$, directly assigning the probabilities of selecting the different surfaces. For us, $\text{pr}\{Y(s_i) = \theta_l^*(s_i), Y(s_j) = \theta_m^*(s_j)\} = p_{l,m}(s_i, s_j)$.

5. SIMULATION-BASED MODEL FITTING

Sections and 3 establish that we work with the following spatial model. Let the vectors $Y_t = \{Y_t(s_1), \dots, Y_t(s_n)\}^T, t = 1, \dots, T$, denote T groups of independent observations collected at the same set of locations $(s_1, \dots, s_n) \in D \subset \mathbb{R}^2$. The mean surface $u(s), s \in D$, is modelled by a linear regression $u(s) = x(s)^T \beta$. The spatial random effect $\theta(s), s \in D$, has the nonparametric rule as defined in §3. The overall model has the following hierarchical structure:

$$\begin{aligned} Y_t \mid \theta_t, \beta, \tau^2 &\sim N_n(X_t^T \beta + \theta_t, \tau^2 I_n), & t = 1, \dots, T, \\ \theta_t \mid G^{(n)} &\sim G^{(n)}, & t = 1, \dots, T, \\ G^{(n)} \mid p_{i_1, \dots, i_n}, \theta_l^* &= \sum_{i_1, \dots, i_n=1}^{\infty} p_{i_1, \dots, i_n} \delta_{\theta_{i_1}^*(s_1)}(\cdot) \cdots \delta_{\theta_{i_n}^*(s_n)}(\cdot), & l = 1, 2, \dots, \\ p_{i_1, \dots, i_n} &= \text{pr}\{Z_1(s_l) < 0, \dots, Z_{i_j}(s_l) \geq 0, & l = 1, \dots, n\}, \\ & i_j = 1, 2, \dots, & j = 1, 2, \dots, n, \\ \{\theta_l^*(s_1), \dots, \theta_l^*(s_n)\}^T &\sim N_n\{0, \sigma^2 R_n(\phi)\}, & l = 1, 2, \dots, \\ \{Z_{t,l}(s_1), \dots, Z_{t,l}(s_n)\}^T &\sim N_n\{\mu_l 1_n, H_n(\eta)\}, & l = 1, 2, \dots, \quad t = 1, 2, \dots, T, \\ \mu_l \text{ s.t. } \Phi(\mu_l) &\sim \text{Be}(1, \nu), & l = 1, 2, \dots, \\ \beta, \tau^2 &\sim N_p(\beta_0, \Sigma_\beta) \times \text{IG}(a_\tau, b_\tau), \\ \sigma^2, \phi, \eta &\sim \text{IG}(a_\sigma, b_\sigma) \times [\phi] \times [\eta], \end{aligned} \tag{5}$$

where we use the brackets notation in Gelfand & Smith (1990) to denote densities. The priors for ϕ and η depend on the specific form of covariance function in $R_n(\phi)$ and $H_n(\eta)$. For convenience, in our examples we have set $\nu = 1$. In the version with $\mu_l(s)$ replacing μ_l , for each l , we obtain a realization from a Gaussian process with mean zero and stationary covariance function $C(\cdot, \psi)$. In either case, the replications across t enable us to learn about the μ_l or the process driving the $\mu_l(s)$.

Although in (5) the marginal random distribution at an individual location s follows a Dirichlet process, the joint random distribution $G^{(n)}$ does not; we cannot marginalize over $G^{(n)}$. For model fitting, therefore, we approximate $G^{(n)}$ with a finite sum

$$G_K^{(n)} = \sum_{(i_1, \dots, i_n) \in \{1, 2, \dots, K\}^n} p_{i_1, \dots, i_n} \delta_{\theta_{i_1}^*(s_1)}(\cdot) \delta_{\theta_{i_2}^*(s_2)}(\cdot) \dots \delta_{\theta_{i_n}^*(s_n)}(\cdot), \tag{6}$$

for K suitably large. In this finite mixture model, we only need θ_l^* , $l = 1, \dots, K$, and Z_l , $l = 1, \dots, K - 1$. Note that $p_K(s) = \text{pr}\{Z_1(s) < 0, \dots, Z_{K-1}(s) < 0\}$.

Again, we sample the latent variables Z_l 's directly to avoid computation of the weights p_{i_1, \dots, i_n} in (6). For $t = 1, \dots, T$ and $l = 1, \dots, K - 1$, consider the sets $\mathcal{Z}_{t,l}(s) = \{s \in D \mid Z_{t,1}(s) < 0, \dots, Z_{t,i-1}(s) < 0, Z_{t,l}(s) \geq 0\}$, a $\mathcal{Z}_{t,K}(s) = \{s \in D \mid Z_{t,1}(s) < 0, \dots, Z_{t,K-1}(s) < 0\}$. We eliminate the sampling of the conditional distribution of $\theta_t \mid G_K^{(n)}$ by referring to the following equivalent structure:

$$\theta_t(s) = \theta_1^*(s) I_{\mathcal{Z}_{t,1}(s)} + \theta_2^*(s) I_{\mathcal{Z}_{t,2}(s)} + \dots + \theta_K^*(s) I_{\mathcal{Z}_{t,K}(s)}. \tag{7}$$

In equation (7), $\theta_t(s)$ is a deterministic function of $\theta_l^*(s)$, $l = 1, \dots, K$, and $Z_l^t(s)$, $l = 1, \dots, K - 1$. We rewrite the first stage of the hierarchical model as $[Y_t \mid \mu, \theta_t] = [Y_t \mid \mu, \theta^*, Z_t]$. The likelihood function for Y_t can then be expressed as

$$\begin{aligned} [Y_t \mid \mu, \theta^*, Z_t] &\propto \exp \left[-\frac{1}{2\tau^2} \sum_{i=1}^n \{Y_t(s_i) - X_t(s_i)^\top \beta - \theta_t(s_i)\}^2 \right] \\ &\propto \exp \left[-\frac{1}{2\tau^2} \sum_{l=1}^K \sum_{i=1}^n \{Y_t(s_i) - X_t(s_i)^\top \beta - \theta_l^*(s_i)\}^2 I_{\mathcal{Z}_{t,l}(s_i)} \right] \\ &\propto \prod_{i=1}^n \left(\sum_{l=1}^K \exp \left[-\frac{1}{2\tau^2} \{Y_t(s_i) - X_t(s_i)^\top \beta - \theta_l^*(s_i)\}^2 \right] \times I_{\mathcal{Z}_{t,l}(s_i)} \right). \end{aligned}$$

The posterior distributions for the latent variables and parameters are proportional to this likelihood function multiplied by the priors,

$$\prod_{t=1}^T [Y_t \mid \theta^*, Z_t, \tau^2] \times \prod_{l=1}^K [\theta_l^* \mid \sigma^2, \phi] \times \prod_{t=1}^T \prod_{l=1}^{K-1} [Z_{t,l} \mid \mu_{t,l}, \eta][\mu_{t,l}] \times [\sigma^2][\phi][\tau^2]$$

This model can be fitted by a Gibbs sampler; the details of all the full-conditional distributions are given in Appendix 2.

We can easily interpolate $\{Y_t(\tilde{s}_1), \dots, Y_t(\tilde{s}_m)\}$ at m new locations on any of the T realized replicates or predict the posterior predictive distribution for $\{Y(\tilde{s}_1), \dots, Y(\tilde{s}_m)\}$ on a new replicate. We can expand $[G_K^{(n)} \mid \text{data}]$ to $[G_K^{(n+m)} \mid \text{data}]$ with m new locations by the method as follows. The $\{\theta_l^*(\tilde{s}_1), \dots, \theta_l^*(\tilde{s}_m)\}$, $l = 1, \dots, K$, in $[G_K^{(n+m)} \mid \text{data}]$ are normal conditionally on $\{\theta_l^*(s_1), \dots, \theta_l^*(s_n)\}$, $l = 1, \dots, K$, and the model parameters. Hence they are sampled accordingly, conditioning on the posterior samples of $\{\theta_l^*(s_1), \dots, \theta_l^*(s_n)\}$, $l = 1, \dots, K$. If we

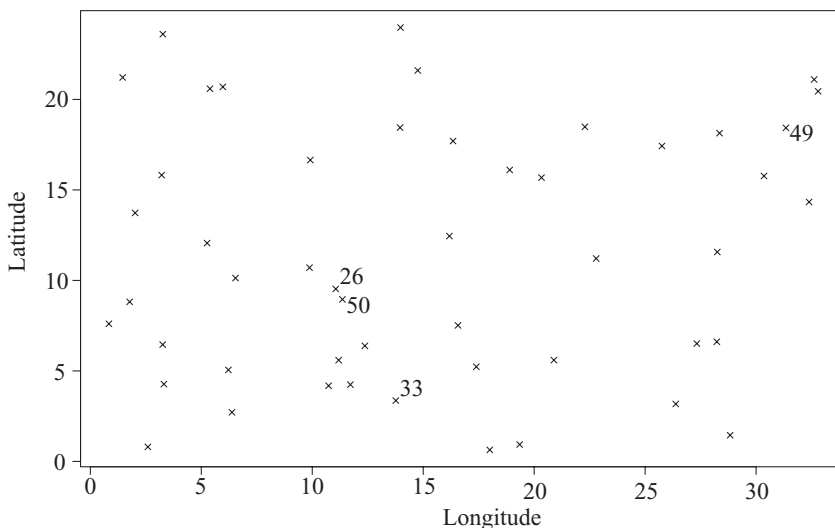


Fig. 1. The design locations for the simulation example in §6.

want to interpolate $\{Y_t(\tilde{s}_1), \dots, Y_t(\tilde{s}_m)\}$ on any realized replicate t , the $\{Z_{t,l}(\tilde{s}_1), \dots, Z_{t,l}(\tilde{s}_m)\}$ will be sampled conditioning on the posterior samples of $\{Z_{t,l}(s_1), \dots, Z_{t,l}(s_n)\}$ and the corresponding model parameters. If we want the posterior predictive distribution on a new replicate for $\{Y(\tilde{s}_1), \dots, Y(\tilde{s}_m)\}$, we will sample $\{Z_l(\tilde{s}_1), \dots, Z_l(\tilde{s}_m)\}$ independently of the posterior samples of $\{Z_{t,l}(s_1), \dots, Z_{t,l}(s_n)\}$, $t = 1, \dots, T$.

6. AN ILLUSTRATIVE EXAMPLE

We illustrate the fitting of (5) with a simulated dataset from a finite mixture model of Gaussian processes that allows different joint multimodal distributions for different pairs of locations. We first draw a specified number of locations in a given region. They are denoted by (s_1, \dots, s_n) . Suppose there are T independent replicates $\{y_t(s_1), \dots, y_t(s_n)\}$, $t = 1, \dots, T$. We proceed as follows. For $t = 1, \dots, T$, let $\{\theta_t^1(s_1), \dots, \theta_t^1(s_n)\}^T \sim N_n^{(1)}\{-\mu_1 1_n, \sigma_1^2 R_n(\phi_1)\}$ and $\{\theta_t^2(s_1), \dots, \theta_t^2(s_n)\}^T \sim N_n^{(2)}\{\mu_2 1_n, \sigma_2^2 R_n(\phi_2)\}$. Also, let $\{Z_t(s_1), \dots, Z_t(s_n)\}^T \sim N_n\{0, H_n(\eta)\}$. Then, for $i = 1, \dots, n$, if $Z_t(s_i) \geq 0$, we set $y_t(s_i) = \theta_t^1(s_i)$; if $Z_t(s_i) < 0$, we set $y_t(s_i) = \theta_t^2(s_i)$.

Each $y_t(s_i)$ has a bimodal distribution of the form $\frac{1}{2}N^{(1)}(-\mu, \sigma_1^2) + \frac{1}{2}N^{(2)}(\mu, \sigma_2^2)$. For two locations s_i and s_j near each other, the strong association between $Z_t(s_i)$ and $Z_t(s_j)$ makes it very likely that $y_t(s_i)$ and $y_t(s_j)$ come from the same component $N^{(k)}(\mu_k, \sigma_k^2)$, $k = 1, 2$. If s_i and s_j are distant, the linkage between $Z_t(s_i)$ and $Z_t(s_j)$ is weak, and therefore the component choices for $y_t(s_i)$ and $y_t(s_j)$ are almost independent.

We simulate at 50 design locations in a rectangular region shown in Fig. 1. Then, 40 independent replicates are sampled for these 50 locations. We choose the values of the parameters as $\mu_1 = -\mu_2 = 3$, $\sigma_1 = 2\sigma_2 = 2$, $\phi_1 = \phi_2 = 0.3$ and $\eta = 0.3$ in the above mixture model. We fit the model in (5) to this dataset. We approximate $G^{(n)}$ with a finite sum of $K = 20$ components. To focus on the modelling of spatial dependence, we fixed the mean structure of $\{y_t(s_1), \dots, y_t(s_n)\}$ to be zero. For comparison, we considered the analogous spatial Dirichlet process models, using the same base measure and the same prior for all the parameters, as well as the standard Gaussian process model for spatial data. With regard to the choice of the Gaussian model for comparison,

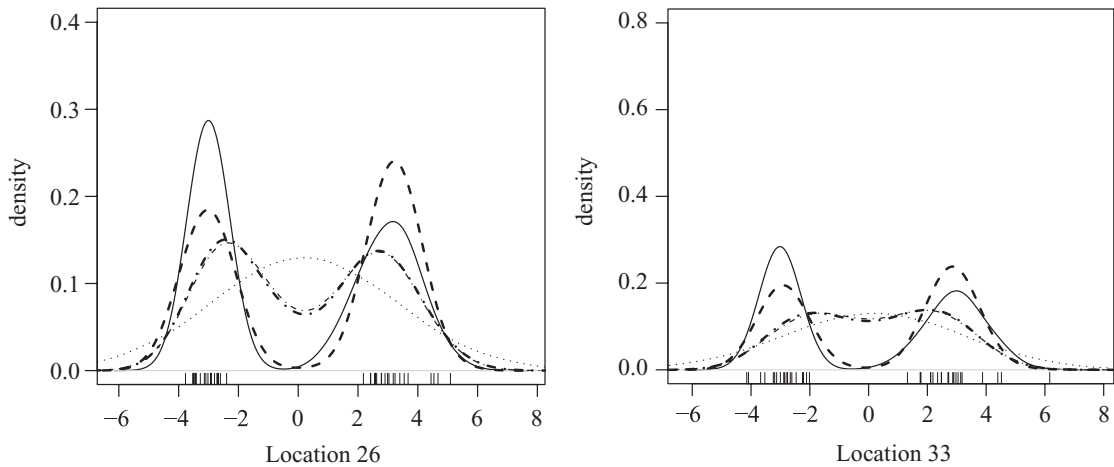


Fig. 2. Posterior predictive densities for the generalized spatial Dirichlet process (thick —) and the spatial Dirichlet process (thick ·-·), and the Gaussian process (both for the independent case (thin —), and the common spatial effect (thin ··)) for locations (a) 26 and (b) 33 in Fig. 1. The continuous line (—) is the estimated true density and the observed sample is reported on the x -axis. Note that the ‘independent’ Gaussian process and the spatial Dirichlet process have similar predictive densities, which indicates that the spatial Dirichlet process shows essentially no clustering.

we considered both the case where the $\theta_t(s)$ are independent across replicates and the case where $\theta_t(s) = \theta(s)$, for all t . The latter can be seen as the model which arises when we fix the precision parameter of the Dirichlet process at zero, so that all replicates are clustered together, whereas the former corresponds to setting ν at a value large enough so that no clustering is induced among the replicates.

We obtain 30 000 samples from our Gibbs sampler for the generalized spatial Dirichlet process and discard the first 10 000 as burn-in. For the posterior inference, we use 4000 subsamples from the remaining 20 000 samples, with a thinning equal to 5. We test each of the models on the same computational platform. All model fitting can be conveniently implemented in R. It takes about 12 hours for the generalized Dirichlet sampler to complete 30 000 iterations. By contrast, the spatial Dirichlet takes 6 hours, or 2 hours in C++ and the Gaussian model takes 2 hours, or half an hour in C++. Evidently, the code for the generalized spatial Dirichlet process requires more bookkeeping than the code for the spatial Dirichlet process, which, in turn, requires more than the code for the Gaussian model. However, since we update the base surfaces $\theta_j^*(\cdot)$ in a single sweep from a multivariate normal distribution in Appendix 2, we avoid problems of slow mixing associated with the basic Dirichlet process (Bush & MacEachern, 1996). In fact, the sampling of the Z 's amounts to sampling which surface is selected at each replicate and each location, as when sampling the configuration indicators in the Pólya urn. Therefore, the number of elements involved affects the computational speed, not the mixing of the chain.

Performance is examined through posterior predictive densities, both marginal and joint. In Fig. 2, for two selected locations (s_{26}, s_{33}), we plot the true density together with the posterior predictive densities estimated under the three types of model. The values of the 40 observations at each location are shown along the x -axis. It is evident that the estimates for our generalization agree most closely with the true densities of the model. Next, we select three pairs of sites and for each pair we show the predictive joint densities. In Fig. 3, the first pair $\{y(s_{50}), y(s_{26})\}$ is very close to each other, the second pair $\{y(s_{50}), y(s_{33})\}$ is slightly distant and the third pair $\{y(s_{50}), y(s_{49})\}$ is very distant; see Fig. 1. We show the results for the generalized and basic Dirichlet process models only. As expected, the corresponding plots for the Gaussian model, corresponding to

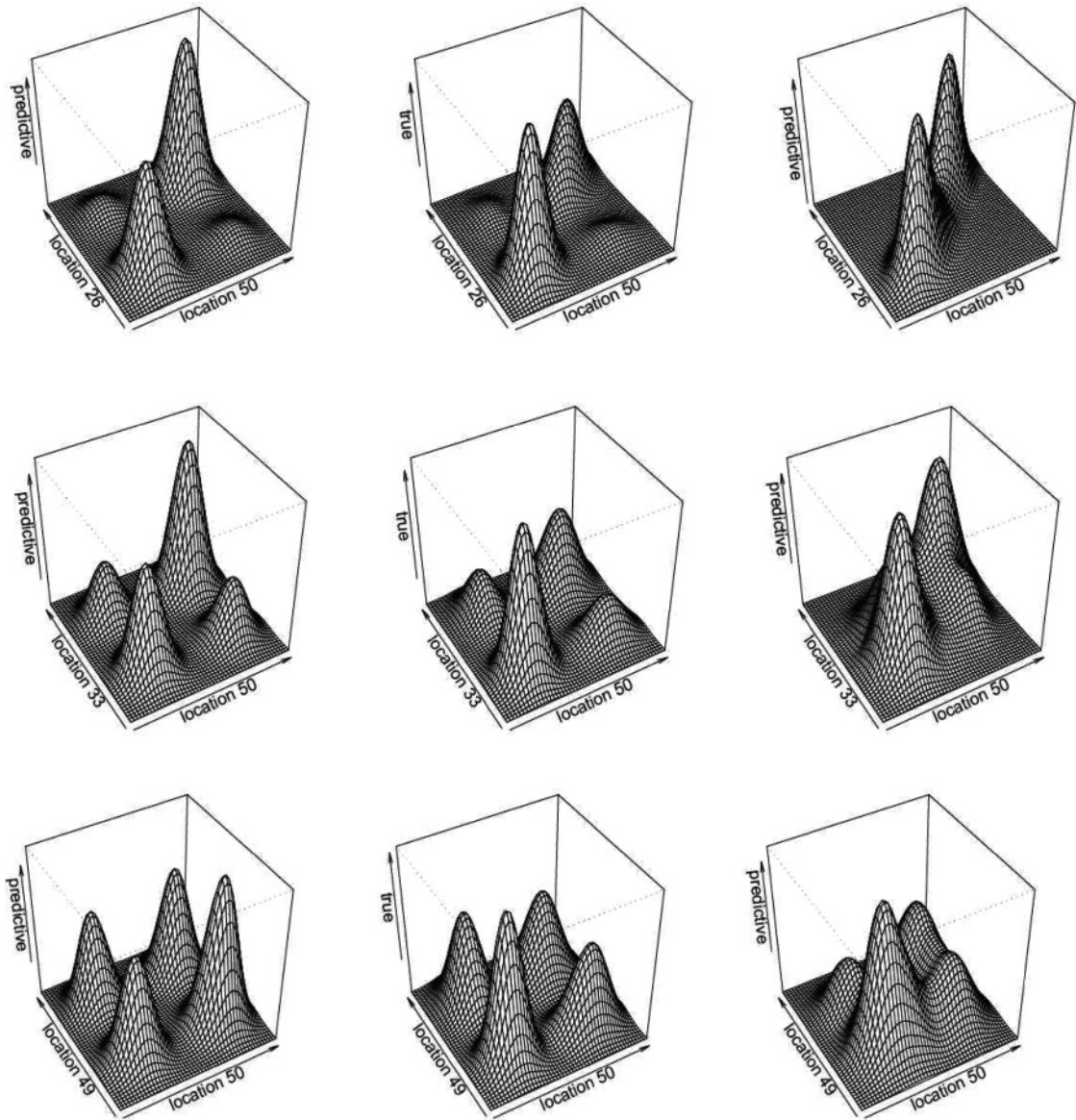


Fig. 3. Predictive bivariate posterior densities for the generalized spatial Dirichlet process, on the left, and the spatial Dirichlet process, on the right, models and the data in §5. The middle column shows the true bivariate posterior density from a sample from the true model; see the text for details.

$\nu \rightarrow 0$, are elliptical in shape, regardless of the locations. Again, the generalized spatial Dirichlet process is better than the spatial Dirichlet process in capturing the local details and in particular the heights of the local modes.

Figure 4 plots the probability that a common sample surface is selected for a pair of locations against the distance between the two locations. We can see the decay in this probability as locations become further apart. Finally, for the spatial Dirichlet process, we see no clustering; essentially, a separate surface is needed for each replication. For the generalized spatial Dirichlet process, the modal number of surfaces is 3 and the maximum number of surfaces is 5. Clearly,

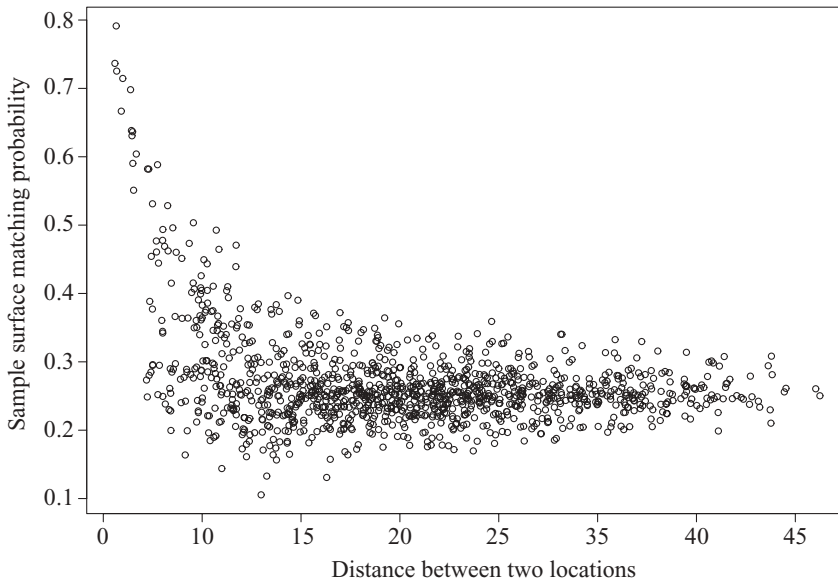


Fig. 4. Simulated example. Decay in probability of common surface selection as a function of distance.

the generalized spatial Dirichlet process is recognizing the local surface selection in the true model.

In many applications replicate observations are made in time, in T consecutive time periods for example, so it may be appropriate to model the evolution of the spatial process over time. A convenient spatio-temporal specification arises if we embed the generalized Dirichlet process in a dynamic linear model. If we preserve the notation in (5), the observations at time t can be modelled by the following dynamic linear model structure:

$$\begin{aligned} Y_t &= X_t^T \beta + \theta_t + \varepsilon_t; & \varepsilon_t &\sim N_n\{\mathbf{0}, \tau^2 I_n\}, \\ \theta_t &= \gamma \theta_{t-1} + \omega_t; & \omega_t &\sim \text{GSDP}(vG_0). \end{aligned} \quad (8)$$

These dynamics yield spatial random effects θ_t that evolve autoregressively over time with autocorrelation coefficient, γ , with $|\gamma| \leq 1$. Only the second hierarchical specification in (5) changes to reflect (8). Updating of the full-conditional distributions and the associated Markov chain Monte Carlo algorithm for the dynamic version is straightforward but, again, requires attention to bookkeeping.

Finally, with regard to the spatial prediction, consider the case of independent replications. Returning to the brain images example, we could imagine interest in prediction at new locations for one of the observed images but we would not seek to predict a random future image. However, when the generalized spatial Dirichlet process is built into a dynamic model, temporal prediction will then be of interest; the methodology follows that described in Gelfand et al. (2005) and is not presented in detail here. An important related point regarding interpolation is the difference between usual kriging associated with a single surface and prediction here with local choice over a countable collection of surfaces. For locations far from those where we have observations, we will have essentially independent surface selection, rendering the interpolation of little use. However, for locations close to observed ones, the underlying latent process will be able to illuminate proper surface selection at each of the replicates.

7. DISCUSSION

Other extensions of the spatial Dirichlet process can be envisaged. For instance, in a future paper we will report on the use of the representation of Theorem 3 of [Ishwaran & Zarepour \(2002a\)](#), rather than the Sethuraman representation, to create a different constructive formulation. In another future investigation we shall model discrete data, such as binary or count data, at the first stage with a generalized spatial Dirichlet process to capture the spatial structure in the mean on a transformed scale. Application to the setting of functional data analysis, in which covariate space replaces geographic space, is also under study. Lastly, we are interested in the case where we observe multivariate data at each location. The generalized spatial Dirichlet process centred around a multivariate spatial process model provides an obvious place to start.

ACKNOWLEDGEMENT

The authors acknowledge Sonia Petrone for insightful comments. The work of the third author was facilitated by funding from the U.S. National Science Foundation.

APPENDIX 1

Theoretical arguments

Proof of Proposition 1. Let $\psi(t, s)$ be the characteristic function of $Y(s)$. From definition (1), it follows that

$$\psi(t, s) = E[\exp\{itY(s)\}] = E\left\{\sum_{l=1}^{\infty} p_l(s) \exp\{it\theta_l^*(s)\}\right\} \rightarrow E[\exp\{itY(s_0)\}]$$

as $\|s - s_0\| \rightarrow 0$, since $\psi(t, s) \leq 1$ and the continuity of the weights implies that

$$\lim_{\|s-s_0\| \rightarrow 0} p_l(s) = \lim_{\|s-s_0\| \rightarrow 0} \sum_{m=1}^{\infty} p_{l,m}(s, s_0) = \sum_{m=1}^{\infty} \lim_{\|s-s_0\| \rightarrow 0} p_{l,m}(s, s_0) = p_l(s_0). \quad \square$$

PROPOSITION A1. *Let $\{Y(s_1), \dots, Y(s_n), s_i \in D, i = 1, \dots, n\}$ have random finite-dimensional distribution given by (1), for $n = 1, 2, \dots$. If the set of weights $\{p_{i_1, \dots, i_n}\}$ is defined by means of a latent process as in (4), then the collection of random finite-dimensional distributions define a random field $Y(s)$ on D .*

Proof. First we show that, for any $l = 1, \dots, n$,

$$p_{i_1, \dots, i_{l-1}, i_{l+1}, \dots, i_n} = p_{i_1, \dots, i_{l-1}, i_{l+1}, \dots, i_n} = \sum_{k=1}^{\infty} p_{i_1, \dots, i_{l-1}, k, i_{l+1}, \dots, i_n}.$$

In fact, let $Z(s_i) = \{Z_1(s_i), \dots, Z_k(s_i), \dots\}, i = 1, \dots, n$. Note that if $\theta(s_i) = \theta_k^*(s_i)$ then $Z(s_i) \in S_{i,k}$, where $S_{i,k} = (-\infty, 0)_1 \times \dots \times (-\infty, 0)_{k-1} \times [0, \infty)_k \times \mathbb{R} \times \dots$. By the continuity of the probability measure,

$$\begin{aligned} \sum_{k=1}^{\infty} p_{i_1, \dots, i_{l-1}, k, i_{l+1}, \dots, i_n} &= \text{pr}\left\{Z(s_1) \in S_{1,i_1}, \dots, Z(s_{l-1}) \in S_{l-1,i_{l-1}}, Z(s_l) \right. \\ &\quad \left. \in \bigcup_{k=1}^{\infty} S_{l,k}, Z(s_{l+1}) \in S_{l+1,i_{l+1}}, \dots, Z(s_n) \in S_{n,i_n}\right\}, \end{aligned}$$

with straightforward calculation, since $\bigcup_{k=1}^{\infty} S_{l,k} = \bigotimes_{k=1}^{\infty} \mathbb{R}$.

The theorem is proven after showing that, for any $A_i \in \mathcal{B}(\mathbb{R})$, $i = 1, \dots, k$, we have

$$\begin{aligned}
& \text{pr}\{\theta(s_1) \in A_1, \dots, \theta(s_{l-1}) \in A_{l-1}, \theta(s_l) \in \mathbb{R}, \theta(s_{l+1}) \in A_{l+1}, \dots, \theta(s_n) \in A_n\} \\
&= \sum_{(i_1, \dots, i_n) \in \{1, 2, \dots\}^n} p_{i_1, \dots, i_n} \delta_{\theta_{i_1}^*(s_1)}(A_1) \cdots \delta_{\theta_{i_l}^*(s_l)}(\mathbb{R}) \cdots \delta_{\theta_{i_n}^*(s_n)}(A_n) \\
&= \sum_{(i_1, \dots, i_{l-1}, i_{l+1}, \dots, i_n) \in \{1, 2, \dots\}^{n-1}} \delta_{\theta_{i_1}^*(s_1)}(A_1) \cdots \delta_{\theta_{i_n}^*(s_n)}(A_n) \left(\sum_{k=1}^{\infty} p_{i_1, \dots, i_{l-1}, k, i_{l+1}, \dots, i_n} \right) \\
&= \sum_{(i_1, \dots, i_{l-1}, i_{l+1}, \dots, i_n) \in \{1, 2, \dots\}^{n-1}} p_{i_1, \dots, i_{l-1}, i_{l+1}, \dots, i_n} \delta_{\theta_{i_1}^*(s_1)}(A_1) \cdots \delta_{\theta_{i_n}^*(s_n)}(A_n) \\
&= \text{pr}\{\theta(s_1) \in A_1, \dots, \theta(s_{l-1}) \in A_{l-1}, \theta(s_{l+1}) \in A_{l+1}, \dots, \theta(s_n) \in A_n\}. \quad \square
\end{aligned}$$

PROPOSITION A2. *Let $\{Y(s), s \in D\}$ be as in Proposition A1. If the base random field G_0 is almost surely continuous, then, for all $s_0 \in D$, $Y(s)$ converges weakly to $Y(s_0)$ with probability one as $\|s - s_0\| \rightarrow 0$.*

Proof. The proof follows immediately from Proposition A1, once we notice that, under our assumptions, for any $n = 1, 2, \dots$, $\lim_{\|s_n - s_{n-1}\| \rightarrow 0} p_{i_1, \dots, i_n} = p_{i_1, \dots, i_{n-1}}$ if $i_n = i_{n-1}$, and is zero otherwise, independently of the particular mean around which we centre the process Z , i.e., the weights are almost surely continuous. \square

APPENDIX 2

Full-conditional distributions for the Gibbs sampler

Full conditionals for the Z 's. To write down the full conditionals for the Z 's, we first write the conditional distributions

$$[Z_{t,l}(s_i) | Z_{t,l}(s_j), j \neq i, \mu_l, \eta] = N\{\tilde{\mu}_{t,l}^i, \tilde{H}_i(\eta)\},$$

for all $i = 1, \dots, n$, $l = 1, \dots, K - 1$, $t = 1, \dots, T$, where

$$\begin{aligned}
\tilde{\mu}_{t,l}^i &= \mu_l - h_i(\eta)^\top H_{(-i)}^{-1}(\eta) Z_{t,l}^{(-i)}, \\
\tilde{H}_i(\eta) &= 1 - h_i(\eta)^\top H_{(-i)}^{-1}(\eta) h_i(\eta),
\end{aligned}$$

in which $h_i(\eta)$ is the i th column vector of $H_n(\eta)$, $H_{(-i)}(\eta)$ is the $(n-1) \times (n-1)$ matrix obtained from $H_n(\eta)$ by deleting the i th row and column and $Z_{t,l}^{(-i)}$ is the $(n-1)$ -dimensional vector obtained from $Z_{t,l}$ by deleting the i th element. Note that both $\tilde{\mu}_{t,l}^i$ and $\tilde{H}_i(\eta)$ are scalars.

We indicate with $\psi = (X_t, \beta, \theta^*, \tau^2, \sigma^2, \phi, \mu_l, l > 1, \eta)$ the vector of parameters of the model other than the Z 's. Then the full conditional of $Z_{t,l}(s_i)$ is given by

$$\begin{aligned}
& [Z_{t,l}(s_i) | Y_t, Z_{t,l}(s_j), Z_{t,m}(s_i), m \neq l, j \neq i, \psi] \propto [Z_{t,l}(s_i) | Z_{t,l}(s_j), j \neq i, \psi] \\
& \times \sum_{m=1}^K \exp \left[-\frac{1}{2\tau^2} \{Y_t(s_i) - X_t(s_i)^\top \beta - \theta_m^*(s_i)\}^2 \right] I_{Z_{t,m}(s_i)},
\end{aligned}$$

where $Z_{t,m}(s_i)$, $m \neq l$, are all known. If $Z_{t,m}(s_i) \geq 0$, for some $m < l$, then $\theta_t(s_i) = \theta_m^*(s_i)$ and $Z_{t,l}(s_i)$ is sampled directly from the unrestricted distribution $N\{\tilde{\mu}_{t,l}^i, \tilde{H}_i(\eta)\}$.

Otherwise, if $Z_{t,m}(s_i) < 0$, for $m < l$, the full conditional is a binary mixture of truncated normals. If $Z_{t,k}(s_i) \geq 0$ for the first $k > l$, let

$$\begin{aligned} \omega^- &= \exp \left[-\frac{1}{2\tau^2} \{Y_t(s_i) - X_t(s_i)^T \beta - \theta_l^*(s_i)\}^2 \right], \\ \omega^+ &= \exp \left[-\frac{1}{2\tau^2} \{Y_t(s_i) - X_t(s_i)^T \beta - \theta_k^*(s_i)\}^2 \right], \\ \pi_l &= \frac{\omega^- \Phi \{ \zeta_{t,l}^i \}}{\omega^- \Phi \{ \zeta_{t,l}^i \} + \omega^+ \Phi \{ -\zeta_{t,l}^i \}}, \quad \pi_k = \frac{\omega^+ \Phi \{ -\zeta_{t,l}^i \}}{\omega^- \Phi \{ \zeta_{t,l}^i \} + \omega^+ \Phi \{ -\zeta_{t,l}^i \}}, \end{aligned}$$

where $\zeta_{t,l}^i = \tilde{\mu}_{t,l}^i / \sqrt{\tilde{H}_i(\eta)}$. Therefore, the full conditional for $Z_{t,l}(s_i)$ is a mixture of two truncated normals. In particular, with probability π_l , we sample $Z_{t,l}(s_i)$ from the truncated normal distribution $N\{\tilde{\mu}_{t,l}^i, \tilde{H}_i(\eta)\}I_{\{Z_{t,l}(s_i) \geq 0\}}$; with probability π_k , we sample $Z_{t,l}(s_i)$ from the truncated normal distribution $N\{\tilde{\mu}_{t,l}^i, \tilde{H}_i(\eta)\}I_{\{Z_{t,l}(s_i) < 0\}}$.

Full conditional for the θ^ 's.* We can update all the θ^* 's at once for all locations. Let us consider at each point $s \in D$ the partition induced on the space of the Z 's by the allocation process; that is, for $t = 1, \dots, T$ and $l = 1, \dots, K - 1$, consider the sets $\mathcal{Z}_{t,l}(s) = \{s \in D : Z_{t,1}(s) < 0, \dots, Z_{t,l-1}(s) < 0, Z_{t,l}(s) \geq 0\}$, and $\mathcal{Z}_{t,K}(s) = \{s \in D : Z_{t,1}(s) < 0, \dots, Z_{t,K-1}(s) < 0\}$. Then $I(\mathcal{Z}_{t,l}) = \text{diag}\{I_{\mathcal{Z}_{t,l}(s_1)}, \dots, I_{\mathcal{Z}_{t,l}(s_n)}\}$ is the diagonal matrix whose i th entry is equal to one when the component l is chosen at location s_i .

Immediately, the full conditional for $\theta_l^* = \{\theta_l^*(s_1), \dots, \theta_l^*(s_n)\}$ is given by

$$\begin{aligned} &[\theta_l^* | Y_t, Z_t, t = 1, \dots, T, \beta, \tau^2, \sigma^2, \phi] \\ &\propto \exp \left\{ -\frac{1}{2\tau^2} \sum_{t=1}^T (Y_t - X_t^T \beta - \theta_l^*)^T I(\mathcal{Z}_{t,l}) (Y_t - X_t^T \beta - \theta_l^*) \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2\sigma^2} \theta_l^{*T} R_n^{-1}(\phi) \theta_l^* \right\}. \end{aligned}$$

Then, with $\Lambda = \{ \sum_{t=1}^T I(\mathcal{Z}_{t,l}) / \tau^2 + R_n^{-1}(\phi) / \sigma^2 \}^{-1}$,

$$[\theta_l^* | Y_t, Z_t, t = 1, \dots, T, \beta, \tau^2, \sigma^2, \phi] = N \left\{ \frac{1}{\tau^2} \Lambda \sum_{t=1}^T I(\mathcal{Z}_{t,l}) (Y_t - X_t^T \beta), \Lambda \right\}.$$

Once we know θ_l^* and Z_t for all $l = 1, \dots, K$ and $t = 1, \dots, T$, we can compute each θ_t as a function of θ_l^* and Z_t .

Full conditionals for $\beta, \tau^2, \sigma^2, \phi, \mu$ and η . Assume that $\beta \sim N_p\{\beta_0, \Sigma_0\}$. Then

$$[\beta | X_t, Y_t, Z_t, \theta_t, \tau^2] = N\{\hat{\beta}, \hat{\Sigma}_\beta\}, \text{ where } \hat{\Sigma}_\beta = \left(\frac{1}{2} \sum_{t=1}^T X_t^T X_t + \Sigma_0^{-1} \right)^{-1} \text{ and}$$

$$\hat{\beta} = \hat{\Sigma}_\beta \left[\frac{1}{2} \sum_{t=1}^T X_t^T (Y_t - \theta_t) + \Sigma_0^{-1} \beta_0 \right].$$

Assume that $\tau^2 \sim \text{IG}(\alpha_\tau, \beta_\tau)$. Then $[\tau^2 | X_t, Y_t, \theta_t, \beta] = \text{IG}(\tilde{\alpha}_\tau, \tilde{\beta}_\tau)$, where $\tilde{\alpha}_\tau = \alpha_\tau + \frac{1}{2}nT$ and $\tilde{\beta}_\tau = \beta_\tau + \frac{1}{2} \sum_{t=1}^T (Y_t - \beta X_t - \theta_t)^T (Y_t - \beta X_t - \theta_t)$.

Assume that $\sigma^2 \sim \text{ig}(\alpha_\sigma, \beta_\sigma)$. Then $[\sigma^2 | \theta_l^*, \phi] = \text{ig}(\tilde{\alpha}_\sigma, \tilde{\beta}_\sigma)$, where $\tilde{\alpha}_\sigma = \alpha_\sigma + \frac{1}{2}nK$, and $\tilde{\beta}_\sigma = \beta_\sigma + \frac{1}{2} \sum_{l=1}^K \theta_l^{*T} R_n^{-1}(\phi) \theta_l^*$.

Depending on the prior $[\phi]$, the full conditional of ϕ can be sampled with a Metropolis-within-Gibbs step,

$$[\phi|\theta_l^*, \sigma^2] \sim [\phi] \times \exp \left\{ -\frac{1}{2\sigma^2} \sum_{l=1}^K \theta_l^{*\top} R_n^{-1}(\phi) \theta_l^* \right\}.$$

Generally, we must use a Metropolis step for $\mu_l, l = 1, \dots, (K - 1)$, unless the α in the $\text{Be}(1, \alpha)$ is equal to 1. Note that $\text{pr}\{Z_l(s) \geq 0\} = \Phi(\mu_l)$ and $\text{pr}\{Z_l(s) \geq 0\} \sim \text{Be}(1, \alpha)$ induce a prior for μ_l that is proportional to $\{1 - \Phi(\mu_l)\}^{\alpha-1} \times \exp(-\frac{1}{2}\mu_l^2)$. If $\alpha = 1$, the prior for μ_l is a normal distribution and thus conjugate. The full conditional for μ_l is

$$\begin{aligned} [\mu_l|Z_l^l, \eta] &\propto \{1 - \Phi(\mu_l)\}^{\alpha-1} \times \exp \left(-\frac{1}{2}\mu_l^2 \right) \\ &\times \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (Z_t^l - \mu_l \mathbf{1}_n)^\top H_n^{-1}(\eta) (Z_t^l - \mu_l \mathbf{1}_n) \right\}. \end{aligned}$$

Depending on the prior $[\eta]$, the full conditional of ψ can be sampled with a Metropolis-within-Gibbs step,

$$[\eta|Z_t, \mu_l] \sim [\eta] \times \exp \left\{ -\frac{1}{2} \sum_{t=1}^T \sum_{l=1}^{K-1} (Z_t^l - \mu_l \mathbf{1}_n)^\top H_n^{-1}(\eta) (Z_t^l - \mu_l \mathbf{1}_n) \right\}.$$

REFERENCES

- AGARWAL, D. & GELFAND, A. (2005). Slice Gibbs sampling for simulation based fitting of spatial data models. *Statist. Comp.* **15**, 61–9.
- ALBERT, J. & CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Am. Statist. Assoc.* **88**, 669–79.
- BANERJEE, S., GELFAND, A. & SIRMANS, C. (2003). Directional rates of change under spatial process models. *J. Am. Statist. Assoc.* **98**, 946–54.
- BROWN, P., DIGGLE, P. & HENDERSON, R. (2003). A non-Gaussian spatial process model for opacity of flocculated paper. *Scand. J. Statist.* **30**, 355–68.
- BUSH, C. A. & MACEachern, S. N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika* **83**, 275–85.
- CIFARELLI, D. & REGAZZINI, E. (1978). Problemi statistici non parametrici in condizioni di scambiabilità parziale. Impiego di medie associative. *Quad. Istit. Mat. Finanz. Univ. Torino* **III**, 1–13.
- DAMIAN, D., SAMPSON, P. & GUTTORP, P. (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics* **12**, 161–78.
- DE IORIO, M., MÜLLER, P., ROSNER, G. & MACEachern, S. (2004). An ANOVA model for dependent random measures. *J. Am. Statist. Assoc.* **99**, 205–15.
- ECKER, M. & GELFAND, A. (2003). Spatial modeling and prediction under stationary non-geometric range anisotropy. *Envir. Ecol. Statist.* **10**, 165–78.
- ESCOBAR, M. D. & WEST, M. (1995). Bayesian density estimation and inference using mixtures. *J. Am. Statist. Assoc.* **90**, 577–88.
- FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–30.
- FERGUSON, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615–29.
- FERNANDEZ, C. & GREEN, P. (2002). Modelling spatially correlated data via mixtures: a Bayesian approach. *J. R. Statist. Soc. B* **64**, 805–26.
- GELFAND, A. E. & SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Am. Statist. Assoc.* **85**, 398–409.
- GELFAND, A., KOTTAS, A. & MACEachern, S. (2005). Bayesian nonparametric spatial modeling with Dirichlet processes mixing. *J. Am. Statist. Assoc.* **100**, 1021–35.
- GRIFFIN, J. & STEEL, M. (2006). Order-based dependent dirichlet processes. *J. Am. Statist. Assoc.* **101**, 179–94.
- HIGDON, D., SWALL, J. & KERN, J. (1999). Non-stationary spatial modeling. In *Bayesian Statistics 6*, Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp. 761, Oxford: Oxford University Press.
- ISHWARAN, H. & JAMES, L. (2001). Gibbs sampling methods for stick breaking priors. *J. Am. Statist. Assoc.* **96**, 161–73.
- ISHWARAN, H. & ZAREPOUR, M. (2002a). Dirichlet prior sieves in finite normal mixtures. *Statist. Sinica* **12**, 941–63.
- ISHWARAN, H. & ZAREPOUR, M. (2002b). Exact and approximate sum-representations for the Dirichlet process. *Can. J. Statist.* **30**, 269–83.
- KENT, J. (1989). Continuity properties of random fields. *Ann. Prob.* **17**, 1432–40.

- MACÉACHERN, S. (1994). Estimating normal means with a conjugate style Dirichlet prior. *Commun. Statist. B* **23**, 727–41.
- MACÉACHERN, S. N. & MÜLLER, P. (1998). Estimating mixture of Dirichlet process models. *J. Comp. Graph. Statist.* **7**, 223–38.
- PALACIOS, M. & STEEL, M. (2006). Non-Gaussian Bayesian geostatistical modeling. *J. Am. Statist. Assoc.* **101**, 604–18.
- SAMPSON, P. & GUTTORP, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *J. Am. Statist. Assoc.* **87**, 108–19.
- SCHMIDT, A. M. & O'HAGAN, A. (2003). Bayesian inference for nonstationary spatial covariance structure via spatial deformations. *J. R. Statist. Soc. B* **65**, 745–58.
- SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4**, 639–50.

[Received December 2005. Revised April 2007]