Generalized Spatial Dirichlet Process Models

By JASON A. DUAN

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

jason@stat.duke.edu

MICHELE GUINDANI

Istituto di Metodi Quantitativi, Università Bocconi, 20136, Milano, Italy

michele.guindani@unibocconi.it

and ALAN E. GELFAND

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

alan@stat.duke.edu

Summary

In many applications, spatial data are assumed to be point referenced, e.g., observed at geo-coded locations. Modelling for this kind of data usually introduces spatial structure in the form of spatial random effects where a term capturing residual spatial association is explicitly introduced. This pure spatial effect is customarily modelled as a mean-zero stationary Gaussian process (GP). The SDP 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 SDP arises as a probability weighted collection of random surfaces. This can be unattractive for modelling, hence inferential purposes since it insists that a process

realization is one of these surfaces. 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. 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 (GSDP). We illustrate with a simulated dataset of independent replications and then demonstrate how to embed the GSDP within a dynamic model specification to remove the restrictive independence assumption, again providing an illustrative example. Finally, the GSDP is considerably more computationally demanding to work with than the SDP and so we also offer a collection of justifications for its use over the SDP.

Some key words: Dirichlet process mixing; dynamic models; latent processes; non-Gaussian; nonstationary; stick-breaking.

1 Introduction

In many applications, spatial data are assumed to be point referenced, e.g., observed at geo-coded locations. For example, this is the typical geostatistics setting, where many phenomena can be seen as realizations of (possibly) vector valued random fields at a set of known locations, referenced by their geographical coordinates. Modelling for this kind of data usually introduces spatial structure in the form of spatial random effects models, where a term capturing residual spatial association is explicitly introduced. This pure spatial effect is often modelled as a mean-zero stationary Gaussian process (GP). Within a Bayesian framework, the resulting model specification can be viewed as hierarchical and model fitting typically requires the use of Markov Chain Monte Carlo methods. See, e.g. Agarwal & Gelfand (2005) and Ecker & Gelfand (2003).

However, in many cases, either the stationarity or the Gaussian assumption will be

viewed as inappropriate. Flexible and computationally tractable modelling to remove the stationarity assumption includes the spatially varying kernel approach of Higdon et al. (1999) and the local stationarity approach of Fuentes & Smith (2001) but both are still within the setting of GP's. The fundamental paper of Sampson & Guttorp (1992) introduces a nonparametric specification for the covariance function, as does followup work by Damian et al. (2001) and Schmidt & O'Hagan (2003) but they still employ a GP in the likelihood.

The Gaussian assumption can be criticized when the spatial variability is caused by more than one latent processes so that, for example, a mixture of Gaussian processes would probably be more appropriate. See Brown et al. (2003) for a recent discussion of this issue related to the study of product quality in the paper-making industry or Palacios & Steel (2004) for the development of a class of models able to cope with heavy tail behavior.

Recently, Gelfand et al. (2005) have proposed a spatial Dirichlet process (SDP) mixture model to produce a random spatial process that is neither Gaussian nor stationary. The SDP explicitly adopts the distribution of a stochastic process as its base measure. This is assumed to be stationary and Gaussian; nevertheless the resulting process is nonstationary and the joint finite dimensional distributions are not normal. The use of the SDP specification to model the distribution of the spatial component in a spatial random effect model leads to a fully Bayesian semiparametric approach that, for fitting purposes, relies on well-known results and algorithms developed for Dirichlet process (DP) mixing. See, among others, Escobar & West, 1995 and MacEachern & Müller, 1998.

Since the SDP is essentially a Dirichlet process defined on a space of surfaces, its realizations are discrete probability measures with countable support with probability one (Ferguson, 1973 and Sethuraman, 1994). Thus, mixing against a Gaussian ker-

nel yields an error specification that can be characterized as a countable mixture of normals, and so in principle it is able to capture virtually any distribution for the observables. However, the way this is achieved can be unsatisfactory for inferential purposes. This is because the SDP 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. The development is offered constructively by providing a multivariate extension of the stick-breaking representation of the weights which is known to characterize the usual Dirichlet process (Sethuraman, 1994). Hence, we define a new class of random probability measures for random vectors and processes, which includes the customary Dirichlet process specification as a special case. We refer to this new class as generalized spatial Dirichlet process models (GSDP). In fact, this class can be seen as an extension of the generic class of priors described in Hjort (2000) and Ishwaran & James (2001), which, as well, take their aim explicitly from the stick-breaking representation. Also, we clarify modeling specifications under which the GSDP would be anticipated to have advantages over the SDP.

Fitting DP mixing models require that data come as a set of replications at the observed sites. This is not unexpected since replications are typically needed for a full nonparametric approach (see, e.g. Sampson & Guttorp, 1992 and Damian et al., 2001). Usefully, in Section 7, with replications that are discretized across time, we show that 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 under

such modelling 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.

Notice that, although we develop our model in the context of Bayesian nonparametric analysis for spatial data, the theory is general and can be used in other contexts. For example, rather than indexing our responses by spatial location, we could index them by a covariate as in usual regression settings. As a result, our model can be used as an alternate choice in most of the problems where mixtures of products of Dirichlet processes (Cifarelli & Regazzini, 1978) and/or the dependent Dirichlet processes (MacEachern, 2000) have been employed. See, for example, De Iorio et al., 2004.

In the context of Bayesian analysis of spatial data, 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. However, they consider problems where it is only the weights in the mixture that vary from one location to another. Our model differs from theirs since it applies to general point referenced data and both the weights and the parameters of the mixed distribution are allowed to vary spatially. Griffin & Steel (2004) present an implementation of the dependent Dirichlet process in the context of spatial processes using Sethuraman's constructive representation, providing a random marginal distribution at each site. The components of the marginal stick breaking 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 define a joint stick-breaking construction for any number and choice of locations, and also allow the marginal components to vary in space. Moreover, in our approach the closeness between the random distributions is ruled directly by the topology of the space, rather than the realizations of an underlying point process.

The format of the paper is as follows. In Section 2 we briefly review the spatial Dirichlet process model presented in Gelfand et al. (2005). Section 3 formalizes the generalized spatial Dirichlet process (GSDP) and develops its properties. It also treats mixing of Gaussian kernels using this process. Section 4 elaborates the spatially varying probabilities model that is a component of the GSDP. Section 5 presents the computational strategy for fitting such models while section 6 offers an example. Section 7 shows how to embed the GSDP within a dynamic model, again with an example. Section 8 concludes with a summary as well as a discussion of attractive modeling contexts for the GSDP.

2 Review of Dirichlet Process and Spatial Dirichlet Process Modelling

Here, we briefly review the SDP as developed in Gelfand et al. (2005). Denote the stochastic process by $\{Y(s) : s \in D\}$, $D \subseteq R^d$. Let $s^{(n)} = (s_1, ..., s_n)$ be the specific distinct locations in D where the observations are collected. Assume that we have available replicate observations at each location and therefore that the full data set consists of the collection of vectors $Y_t = \{Y_t(s_1), ..., Y_t(s_n)\}^T$, t = 1,...,T. In fact, imbalance or missingness can be accommodated in the $Y_t(s_i)$ through customary latent variable methods.

For a measurable space $(\Theta, \mathcal{B})$, the Dirichlet process (DP), (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 a specified base distribution defined on $(\Theta, \mathcal{B})$. A random distribution function on $(\Theta, \mathcal{B})$ arising from $DP(\nu G_0)$ is almost surely discrete and admits the representation $\sum_{l=1}^{\infty} \omega_l \delta_{\theta_l}$, where δ_z denotes a point mass at z, $\omega_1 = z_1$,

$\omega_l = z_l \prod_{r=1}^{l-1}(1 - z_r)$, l = 2,3,..., with z_r, r = 1,2,..., independently and identically distributed as Beta$(1, \nu)$. The θ_l's are independent and identically distributed as G_0 and also independent of the z_r's, $l = 1, 2, ...$ (Sethuraman, 1994). In this notation, θ_l is assumed to be scalar or perhaps vector valued, the latter case leading to a multivariate DP.

To model $Y_D \equiv \{Y(s) : s \in D\}$, following Gelfand et al. (2005), one can conceptually extend θ_l to a realization of a random field by replacing it with $\theta_{l,D} = \{\theta_l(s) : s \in D\}$. For instance, G_0 might be a stationary GP with each $\theta_{l,D}$ being a realization from G_0, i.e., a surface over D. The resulting random process or distribution, G, for Y_D is denoted by $\sum_{l=1}^{\infty} \omega_l \delta_{\theta_{l,D}}$ and the construction is referred to as a spatial Dirichlet process (SDP) model. The interpretation is that for $s^{(n)}$ as above, G induces a random probability measure $G^{(s^{(n)})}$ on the space of distribution functions for $\{Y(s_1), ..., Y(s_n)\}$. (To simplify notation, we will use $G^{(n)}$ instead of $G^{(s^{(n)})}$ in what follows.) Thus, we have that $G^{(n)} \sim DP(\nu G_0^{(n)})$, where $G_0^{(n)} \equiv G_0^{(s^{(n)})}$ is the n-variate distribution for $\{Y(s_1), ..., Y(s_n)\}$ induced by G_0. E.g., $G_0^{(n)}$ is an n-variate normal if G_0 is taken to be a GP).

Gelfand et al. note a connection between the spatial DP above and the notion of a dependent Dirichlet process (DDP) as developed by MacEachern (2000). The DDP provides a formal framework within which to describe a stochastic process of random distributions. These distributions are dependent but such that, at each index value, the distribution is a univariate DP. In the above setting, G induces a random distribution $G(Y(s))$ for each s, hence the set $\mathcal{G}_{\mathcal{D}} \equiv \{G(Y(s)) : s \in D\}$ which, under sufficient conditions (MacEachern, 2000, Theorem 3.1) will be a DDP.

For a stationary G_0 (i.e., cov$\{\theta_l(s_i), \theta_l(s_j)\}$ depends upon s_i and s_j through $s_i - s_j$), the choice of the covariance function determines how smooth process realizations are. Kent (1989), for instance, shows that, if the covariance function admits a second order

Taylor-series expansion with remainder that goes to 0 at a rate of $2+\delta$ for some $\delta > 0$ then $\theta(s_i) - \theta(s_j) \to 0$, almost surely, as $||s_i - s_j|| \to 0$. But then, in the representation of G as $\sum \omega_l \delta_{\theta_{l,D}}$, the continuity of $\theta_{l,D}$ implies that the random marginal distribution of $Y(s_i)$ given G, $G(Y(s_i))$, and the random marginal distribution of $Y(s_j)$ given G, $G(Y(s_j))$, are such that the difference between them tends to 0 almost surely, as $||s_i - s_j|| \to 0$. The implication is that we can learn about $G(Y(s))$ more from data at neighboring locations than from data at locations further away, as in usual spatial prediction.

For G arising from G_0 and ν, note that given G, $\mathrm{E}\{Y(s) \mid G\} = \sum \omega_l \theta_l(s)$ and $\mathrm{var}\{Y(s) \mid G\} = \sum \omega_l \theta_l^2(s) - \{\sum \omega_l \theta_l(s)\}^2$. Moreover for a pair of sites s_i and s_j,

$$\mathrm{cov}\{Y(s_i), Y(s_j) \mid G\} = \sum \omega_l \theta_l(s_i)\theta_l(s_j) - \left\{\sum \omega_l \theta_l(s_i)\right\}\left\{\sum \omega_l \theta_l(s_j)\right\}. \quad (1)$$

Hence, the random process G has heterogeneous variance and is nonstationary. If G_0 is a mean zero stationary GP with variance σ^2 and correlation function $\rho_\phi(s_i - s_j)$, where the (possibly vector valued) parameter ϕ specifies $\rho_\phi(\cdot)$, then, marginalizing over G, $\mathrm{E}\{Y(s)\} = 0$, $\mathrm{var}\{Y(s)\} = \sigma^2$ and $\mathrm{cov}\{Y(s_i), Y(s_j)\} = \sigma^2 \rho_\phi(s_i - s_j)$. That is, G is centered around a stationary process with constant variance but it has nonconstant variance and is nonstationary. Also, with almost surely continuous process realizations, (1) makes it clear that the SDP is mean square continuous. That is, given G, $\lim_{||s-s'||\to 0} E[\{Y(s) - Y(s')\}^2 | G] = 0$.

Since the almost sure discreteness of G will be undesirable in practice, mixing a pure error process with variance τ^2 with respect to G creates a random process F which has continuous support. If θ_D given G is a realization from G and $Y_D - \theta_D$ is a realization from the pure error process, then, operating formally, we find that, marginally, Y_D arises from the process F which can be defined as the convolution

$$F\left(Y_D \mid G, \tau^2\right) = \int \mathcal{K}\left(Y_D - \theta_D \mid \tau^2\right) G\left(d\theta_D\right).$$

Differentiating to densities,

$$f\left(Y_D \mid G, \tau^2\right) = \int k\left(Y_D - \theta_D \mid \tau^2\right) G\left(d\theta_D\right). \tag{2}$$

Here $\mathcal{K}$ and k denote the joint distribution function and density function, respectively, of the pure error process over D. k might denote a $N(0,1)$ or $t_r(0,1)$ density. Hence for any s, $f\left(Y(s) \mid G, \tau^2\right) = \int k\left(Y(s) - \theta(s) \mid \tau^2\right) G\left(d\theta(s)\right)$. In other words, $Y(s) = \theta(s) + \epsilon(s)$ where $\theta(s)$ arises from the above spatial DP prior model and $\epsilon(s)$ is $\mathrm{N}(0, \tau^2)$. The customary partitioning into a spatial component and a pure error or nugget component results. The process model is created by convolving distributions rather than convolving process variables as in Higdon et al. (1999) or Fuentes & Smith (2001).

For the finite set of locations $s^{(n)} = (s_1, ..., s_n)$, (2) implies that the joint density of $Y = \{Y(s_1), ..., Y(s_n)\}^T$, given $G^{(n)}$ (where $G^{(n)} \sim DP(\nu G_0^{(n)})$) and τ^2, is

$$f\left(Y \mid G^{(n)}, \tau^2\right) = \int N_n\left(Y \mid \theta, \tau^2 I_n\right) G^{(n)}\left(d\theta\right), \tag{3}$$

where, to simplify notation, $\theta \equiv \theta^{(s^{(n)})} = \{\theta(s_1), ..., \theta(s_n)\}^T$ and $N_p(\cdot \mid \lambda, \Sigma)$ denotes the p-variate normal density/distribution (depending on the context) with mean vector λ and covariance matrix Σ. Again, the almost sure representation of $G^{(n)}$ as $\sum \omega_l \delta_{\theta_l}$, where θ_l is the vector $\{\theta_l(s_1), ..., \theta_l(s_n)\}^T$, yields that $f(Y \mid G^{(n)}, \tau^2)$ is almost surely of the form $\sum_{l=1}^{\infty} \omega_l N_n(Y \mid \theta_l, \tau^2 I_n)$, i.e. a countable location mixture of normals. In fact, assuming the existence of expectations given $G^{(n)}$ and τ^2, one can obtain that $\mathrm{E}(Y \mid G^{(n)}, \tau^2) = \sum \omega_l \theta_l$ and the covariance matrix $\Sigma_Y \mid G^{(n)}, \tau^2 = \tau^2 I_n + \Sigma_\theta^{(s^{(n)})}$, where $(\Sigma_\theta^{(s^{(n)})})_{i,j} = \mathrm{cov}\{\theta(s_i), \theta(s_j) \mid G^{(n)}\}$ the covariance arising from (1).

A regression term, $X^T\beta$, would typically be added to the kernel of the mixture model in (3) leading to

$$f\left(Y \mid G^{(n)}, \beta, \tau^2\right) = \int N_n\left(Y \mid X^T\beta + \theta, \tau^2 I_n\right) G^{(n)}\left(d\theta\right). \tag{4}$$

That is, $\mathrm{E}(Y \mid G^{(n)}, \beta, \tau^2) = X^T\beta + \sum \omega_l \theta_l$ where X is a $p \times n$ matrix and β is a $p \times 1$ vector of regression coefficients.

Consider the data $Y_t = \{Y_t(s_1), ..., Y_t(s_n)\}^T$ with associated X_t, t = 1,...,T. Given X_t, the Y_t are assumed independent from $f(Y_t \mid G^{(n)}, \beta, \tau^2)$ as in (4). A DP prior is placed on $G^{(n)}$, i.e., $G^{(n)} \sim DP(\nu G_0^{(n)})$ (induced by the spatial DP prior for G in (2)), with $G_0^{(n)}$ being a multivariate normal with mean zero and covariance matrix $\sigma^2 H_n(\phi)$. The full Bayesian model is completed by placing (independent) priors on β, τ^2, ν, σ^2 and ϕ. Associating with each Y_t a $\theta_t = \{\theta_t(s_1), ..., \theta_t(s_n)\}^T$ where the θ_t, t = 1,...,T are independent realizations from $G^{(n)}$, the following semiparametric hierarchical model emerges

$$\begin{aligned}
Y_t \mid \theta_t, \beta, \tau^2 &\sim N_n(Y_t \mid X_t^T\beta + \theta_t, \tau^2 I_n), t = 1, ..., T \\
\theta_t \mid G^{(n)} &\sim G^{(n)}, t = 1, ..., T \\
G^{(n)} \mid \nu, \sigma^2, \phi &\sim DP(\nu G_0^{(n)});\ G_0^{(n)}(\cdot \mid \sigma^2, \phi) = N_n(\cdot \mid 0_n, \sigma^2 H_n(\phi)) \\
\beta, \tau^2 &\sim N_p(\beta \mid \beta_0, \Sigma_\beta) \times \mathrm{IGamma}(\tau^2 \mid a_\tau, b_\tau) \\
\nu, \sigma^2, \phi &\sim \mathrm{Gamma}(\nu \mid a_\nu, b_\nu) \times \mathrm{IGamma}(\sigma^2 \mid a_\sigma, b_\sigma) \times [\phi],
\end{aligned} \tag{5}$$

where $[\phi]$ indicates a convenient prior distribution for ϕ, according to the bracket notation of Gelfand & Smith (1990).

3 The Generalized Spatial Dirichlet Process Model

In subsection 3.1 we formally develop the GSDP model. In subsection 3.2 we employ this model as a mixing distribution, mixing against a Gaussian kernel.

3.1 Model details

In the spatial Dirichlet Process developed by Gelfand et al. (2005), the random distribution of the pure spatial effect is essentially a Dirichlet Process defined on the space

of the random surfaces over D generated by a mean 0 base spatial process. Then the almost sure characterization of the process implies that the random G for s is not the same as that for s' since $\{\theta_l^*(s)\}$ is not the same as $\{\theta_l^*(s')\}$. However, each distribution has the same set of random stick-breaking probabilities. Indeed, for any group of n locations, the joint distribution uses the same set of stick-breaking probabilities inducing common surface selection for all locations in the group. The spatial dependence is introduced only through the underlying base measure, and it is not possible to capture the situation in which spatial effects can be selected from different surfaces at different locations. This limitation of the SDP is common to other recent work relating to the so-called dependent Dirichlet process (MacEachern, 2000). See, for example, De Iorio et al. (2004).

We introduce a random distribution for the spatial effects that allows different finite dimensional distributions across locations in the sense that surface selection can vary with location and that the joint selection of surfaces for the n locations can vary with the choice of locations. Moreover, we still preserve the property that the marginal distribution at each location comes from a usual univariate Dirichlet Process. This is achieved constructively, defining a new multivariate stick-breaking prior in which spatial dependence structure is also introduced in the modeling of the weights. See Ishwaran & Zarepour, 2002b for a review of stick-breaking univariate priors.

Accordingly, we start by considering a base random field G_0, which, for convenience, we take to be stationary and Gaussian, and indicate with $\theta_l^* = \{\theta_l^*(s), s \in D\}$ a realization from G_0, i.e., a surface over D. 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

$(s_1, \ldots, s_n) \in D$, and any collection of sets $\{A_1, \ldots, A_n\}$ in $\mathcal{B}(\mathbb{R})$,

$$pr\{Y(s_1) \in A_1, \ldots, Y(s_n) \in A_n\} = \sum_{i_1=1}^{\infty} \ldots \sum_{i_n=1}^{\infty} p_{i_1,\ldots,i_n} \quad \delta_{\theta^*_{i_1}(s_1)}(A_1) \ldots \delta_{\theta^*_{i_n}(s_n)}(A_n), \tag{6}$$

where the θ^*_j's are independent and identically distributed as G_0, i_j is an abbreviation for $i(s_j)$, $j = 1, 2, \ldots, n$, and the weights $\{p_{i_1,\ldots,i_n}\}$, conditionally on the locations, have a distribution defined on the infinite dimensional simplex $\mathbb{P} = \{p_{i_1,\ldots,i_n} \geq 0 : \sum_{i_1=1}^{\infty} \ldots \sum_{i_n=1}^{\infty} p_{i_i,\ldots,i_n} = 1\}$. Following customary assumptions in Dirichlet process specification, the $\{p_{i_1,\ldots,i_n}\}$ arise from a spatial process described below, independent of that for the $\theta' s$.

The generalization of the usual Dirichlet process setting is apparent and it is evident that we allow the possibility to choose different surfaces at different locations. We will return to this point later in the section. For now, it will be enough to notice that the weights need to satisfy a consistency condition in order to properly define a random process for $Y(\cdot)$. Specifically, we need that for any set of locations $(s_1, \ldots, s_n)$, $n \in \mathbb{N}$ and for all $k \in \{1, \ldots, n\}$,

$$p_{i_1,\ldots,i_{k-1},i_{k+1},\ldots,i_n} = p_{i_1,\ldots,i_{k-1},\cdot,i_{k+1},\ldots,i_n} \equiv \sum_{j=1}^{\infty} p_{i_1,\ldots,i_{k-1},j,i_{k+1},\ldots,i_n}. \tag{7}$$

In addition, we insist that the weights satisfy a continuity property; we want the random laws associated with locations s_1 and s_2 near to each other to be similar. Equivalently, for locations s and s_0, as $s \to s_0$, $p_{i_1,i_2} = pr\{Y(s) = \theta^*_{i_1}(s), Y(s_0) = \theta^*_{i_2}(s_0)\}$, tends to the marginal probability $p_{i_2} = pr(\{Y(s_0) = \theta^*_{i_2}(s_0)\}$ when $i_1 = i_2$, and to 0 otherwise. Analogously, if we consider three locations (s_1, s_2, s_3), if s_3 is close to say, s_2, 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 n locations is clear; we avoid introducing further notation, and from now on 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). Recall that a univariate spatial process $\theta(s)$, $s \in D$ is said to be almost surely continuous at a point s_0 if $\theta(s) \to \theta(s_0)$ with probability one as $||s - s_0|| \to 0$. We take up an illustrative construction of almost surely continuous weights in Section 4.2 with associated formal arguments supplied in Appendix I. Now, if we also assume the random field G_0 to be almost surely continuous, we are able to establish the following proposition whose proof is also given in Appendix I.

Proposition 1 *Let $\{Y(s),\, s \in D\}$ be a random field, whose random finite dimensional distributions are given by (6) for all $n \in \mathbb{N}$. If the set of weights $\{p_{i_1,\ldots,i_n}\}$ 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 $||s - s_0|| \to 0$.*

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 stated in MacEachern (2000) and Gelfand et al. (2005).

Conditionally on the realized distribution G, the process has first and second moments given by

$$E\{Y(s)|G\} = \sum_{l=1}^{\infty} p_l(s)\, \theta_i^*(s) \tag{8}$$

$$var\{Y(s)|G\} = \sum_{l=1}^{\infty} p_l(s)\, \theta_l^{*^2}(s) - \Big\{ \sum_{l=1}^{\infty} p_l\, \theta_l^*(s) \Big\}^2, \tag{9}$$

and, for a pair of sites s_i, s_j,

$$\begin{aligned} cov\{Y(s_i), Y(s_j)|G\} = & \sum_{l=1}^{\infty}\sum_{m=1}^{\infty} p_{l,m}(s_i, s_j)\, \theta_l^*(s_i)\, \theta_m^*(s_j) + \\ & - \Big\{ \sum_{l=1}^{\infty} p_l(s_i)\, \theta_l^*(s_i) \Big\} \Big\{ \sum_{m=1}^{\infty} p_m(s_j)\, \theta_m^*(s_j) \Big\}. \end{aligned} \tag{10}$$

The latter result generalizes (1) above from Gelfand et al. Also, analogous to the case for the SDP, (10) shows that with almost surely continuous realizations from the base process and of the weights, the GSDP is mean square continuous.

Again, the process $Y(s)$ has heterogeneous variance and is nonstationary. However, when we marginalize over G, we can see more clearly the difference between the models. Again, suppose G_0 is a mean zero stationary Gaussian process with finite variance σ^2 and correlation function $\rho_\phi(s_i - s_j)$. Then, $E\{Y(s)\} = 0$ and $var\{Y(s)\} = \sigma^2$ as before, but now

$$cov\{Y(s_i), Y(s_j)\} = \sigma^2 \rho_\phi(s_i - s_j) \sum_{l=1}^{\infty} E\{p_{ll}(s_i, s_j)\}. \tag{11}$$

Notice that $\sum_{l=1}^{\infty} E\{p_{ll}(s_i, s_j)\} < 1$, unless $p_{ll'}(s_i, s_j) = 0$, $l \neq l'$, as it is in Gelfand et al. (2005) or, more generally, in the single-p dependent Dirichlet process discussed by MacEachern (2000). We can interpret this limiting situation as the one of maximum concordance among the surfaces chosen at the two locations. In all other cases, the association structure is diminished by the amount of mass that the process (6) is expected to place on the not equally indexed θ^*'s. Moreover, from (11) it follows that, although the base measure G_0 is stationary, the process Y(s) is centered around a stationary process only when $E\{p_{ll}(s_i, s_j)\}$ is a function of $s_i - s_j$ for all s_i and s_j.

We now turn to the specification of $p_{i_1,...,i_n}$ for any choice of n and $s_1, ..., s_n$. We propose a theoretically attractive and computationally feasible approach through a multivariate extension of the stick-breaking construction that usually characterizes the univariate Dirichlet process. For the sake of simplicity, we present our approach in a bivariate setting, considering the random measure (6) for a pair of sites s_i, s_j, providing details on extension to the general multivariate case when necessary. First, we define a convenient process which retains the same Dirichlet process structure marginally at each site and then we move to a more general setting.

We start by recalling that in the Sethuraman's univariate stick-breaking construction the random measure $\sum_{l=1}^{\infty} p_l \delta_{\theta_l^*}$ has weights p_l defined by $p_1 = q_1$, $p_l = q_l \prod_{m=1}^{l}(1 - q_m), l \geq 2$ where, for all $l \geq 1$, q_l are independent $Beta(1, \nu)$ random variables also in-

dependent of θ_l. Any realization of such a measure has evidently support on the set of realized θ_l^*'s, $l = 1, 2, \ldots$. Then, it is immediate to define the random events $\{Y = \theta_l^*\}$, denoted by Θ_l^1, as elements of the σ-algebra of the space on which the θ_l^*'s take values, together with their complements Θ_l^0, and interpret the sequence of weights $\{p_1, p_2, \ldots\}$ as arising from $q_1 = pr(\Theta_l^1)$, $q_l = pr(\Theta_l^1 | \Theta_m^0, m < l) = pr(Y = \theta_l^* | Y \neq \theta_m^*, m < l)$, $l = 1, 2, \ldots$. Turning back to our model, at each location s we can define events $\Theta_l^u(s)$, $u = 0, 1$, such that $\Theta_l^1(s) = \{Y(s) = \theta_l^*(s)\}$ and $\Theta_l^0(s) = \{Y(s) \neq \theta_l^*(s)\}$. Then, for any two locations s_i, s_j, we can consider the probabilities $q_{1,u,v}(s_i, s_j) = pr\{\Theta_1^u(s_i), \Theta_1^v(s_j)\}$, $q_{l,u,v}(s_i, s_j) = pr\{\Theta_l^u(s_i), \Theta_l^v(s_j) | \Theta_m^0(s_i), \Theta_m^0(s_j), m < l\}$, $l \geq 2$, $u, v \in \{0, 1\}$. For all $l = 1, 2, \ldots$, we can enter these probabilities in the form of Table 1. Note that, formally, e.g., $q_{l,1,1}(s_i, s_j) + q_{l,1,0}(s_i, s_j) = q_{l,1,+}(s_i, s_j)$ and we need to argue that $q_{l,1,+}(s_i, s_j) = q_l(s_i)$. Similarly, $q_{l,+,1}(s_i, s_j) = q_l(s_j)$. The argument is supplied in Appendix I as Lemma 1. Then, accordingly, we can define the weights in (6) as

$$
\begin{aligned}
p_{lm} &= pr\{Y(s_i) = \theta_l^*(s_i), Y(s_j) = \theta_m^*(s_j)\} \\
&= pr\{\Theta_l^1(s_i), \Theta_m^1(s_j), \Theta_k^0(s_i), k < l, \Theta_r^0(s_j), r < m\} \\
&= \begin{cases} \prod_{k=1}^{l-1} q_{k,0,0}\; q_{l,1,0} \prod_{r=l+1}^{m-1} (1 - q_r)\; q_m & \text{if } l < m \\ \prod_{r=1}^{m-1} q_{r,0,0}\; q_{m,0,1} \prod_{k=m+1}^{l-1} (1 - q_k)\; q_l & \text{if } m < l \\ \prod_{r=1}^{l-1} q_{r,00}\; q_{l,11} & \text{if } l = m \end{cases} \qquad (12)
\end{aligned}
$$

where we have suppressed s_i and s_j. Although not immediate, close inspection of expression (12) reveals that the weights are determined through a partition of the unit square similar to the one induced on the unit segment by the usual stick-breaking construction, so that indeed the former can be considered as a bivariate extension of the latter. We can see this clearly from the illustration in Figure 1. At the first stage, if both the events $\Theta_1^1(s_i)$ and $\Theta_1^1(s_j)$ are true, we break off a region of the unit square

of the same size as the realized value of $q_{1,1,1}(s_i, s_j)$. This is region A in Figure 1. If only $\Theta_1^1(s_i)$ (or $\Theta_1^1(s_j)$) is true, we remain only with a piece corresponding to region B (D). In fact, given $\Theta_1^1(s_i)$ ($\Theta_1^1(s_j)$), we go on with a univariate stick-breaking procedure so that we break off a part of region B (C) according to the values of $q_l(s_j)$ ($q_l(s_i)$), $l = 2, 3, \ldots$. If neither $\Theta_1^1(s_i)$ nor $\Theta_1^1(s_j)$ are true, then we discard all regions A, B, and D and remain only with region C, whose size is determined by $q_{1,0,0}(s_i, s_j)$. Then, at stage two, we repeat the same arguments as above for region C, and so on (see Figure 1).

Following the same steps, the preceding arguments can be easily extended for the n-locations problem to define an n-dimensional stick breaking construction on the unit n-dimensional hypercube.

The construction relies on the specification of probabilities $q_{l,u_1,\ldots,u_n}, u_j \in \{0, 1\}, j = 1, 2, \ldots, n$, where u_j is an abbreviation for $u(s_j)$, at any set of locations $(s_1, \ldots, s_n)$. This is generally difficult, since it entails defining a spatial process which, conditionally on the locations, has values on the simplex $\mathbb{Q} = \{q_{l,u_1,\ldots,u_n} \geq 0 : \sum_{u_1,\ldots,u_n=0}^{1} q_{l,u_1,\ldots,u_n} = 1\}$, and also satisfies consistency conditions of the type (7) for all $l = 1, 2, \ldots$ and any set of locations $(s_1, \ldots, s_n), n \in \mathbb{N}$ and for all $k = 1, \ldots, n$, that is

$$q_{l,u_1,\ldots,u_{k-1},u_{k+1},\ldots,u_n} = q_{l,u_1,\ldots,u_{k-1},\cdot,u_{k+1},\ldots,u_n} \equiv \sum_{u_k=0}^{1} q_{l,u_1,\ldots,u_{k-1},u_k,u_{k+1},\ldots,u_n}.$$

However, in the next section, we offer a flexible construction under which this can be done consistently. For the remainder of this section, as a special case, suppose the process retains the same marginal distribution at each location. Referring to Table 1, this can be achieved by imposing $q_l(s) = q_l$, together with the symmetry condition $q_{l,1,0}(s_i, s_j) = pr\{Y(s_i) = \theta_l^*(s_i), Y(s_j) \neq \theta_l^*(s_j)\} = pr\{Y(s_i) \neq \theta_l^*(s_i), Y(s_j) = \theta_l^*(s_j)\} = q_{l,0,1}(s_i, s_j)$, for all $l = 1, 2 \ldots$ and $s \in D$. But, given q_l, if we can compute say $q_{l,1,1}(s_i, s_j)$ as a function of q_l, the remainder of the table is determined. Then, ac-

cording to Sethuraman's construction, if we allow q_l to be Beta$(1, \nu)$, we get a process which marginally is a Dirichlet process with precision parameter ν and base measure G_0. Together with (12), this illuminates the role of the distribution of the q's in specifying the dependence structure in a multivariate Dirichlet process.

Notice that there are other ways to achieve this particular result. For example, we might consider a process such that each q_l given $q_{l,0,0}$ has a Beta-Stacy distribution with parameters $1, \nu - 1$, $1 - q_{l,0,0}$. If $q_{l,0,0}$ is assumed to be Beta$(1, \nu)$, then q_l is Beta$(1, \nu)$. The model we present in section 4 offers an alternative spatially-explicit way to specify q_l and $q_{l,1,1}$. For the n-dimensional case, symmetry conditions similar to the one stated above must be assumed in order to obtain the same marginal behaviour at each site.

Modelling the marginals to be Dirichlet processes allows direct comparison with the models described by Gelfand et al. (2005) and De Iorio et al. (2004). However, it is worth noting that, though we employ a generalized stick-breaking construction and achieve DP marginal distributions, our model doesn't generally describe a joint Dirichlet Process for a collection of locations. In particular, it follows that, given the dependence between the θ^*'s in the sum representation (6), we are not able to trace a joint urn scheme, but only a marginal one. The SDP model described in Gelfand et al. (2005) stands as a particular case of the model described here, where in Table 1 we set $q_{l,0,1} = q_{l,1,0} = 0$ and $q_{l,1,1} = q_l$ for all locations and for all l.

We can see the generalization from the SDP model also by looking at the random conditional distribution associated with $Y(s_i)|Y(s_j)$ for any pair of locations s_i, s_j. In fact, in the SDP this is just a random indicator function. In our model, it turns out to be another random measure. In fact, the random distribution $Y(s_i)|Y(s_j) = \theta^*_m(s_j)$ is

discrete with probability one and of the form $\sum_{l=1}^{\infty} p_{l|m}(s_i, s_j)\delta_{\theta_l^*(s_i)}$, where

$$\begin{aligned} p_{l|m}(s_i, s_j) &= pr(Y(s_i) = \theta_l^*(s_i)|Y(s_j) = \theta_m^*(s_j)) = \\ &= \frac{p_{lm}(s_i, s_j)}{\prod_{k=1}^{m-1}\{1 - q_k(s_j)\}\, q_m(s_j)}, \end{aligned} \tag{13}$$

since $\sum_l p_{l,m}(s_i, s_j) = p_m(s_j)$ due to marginal stick-breaking. But, substituting the expressions in (12),

$$p_{l|m} = \begin{cases} \prod_{k=1}^{l-1} \frac{q_{k,0,0}}{(1-q_k)} \frac{q_{l,1,0}}{1-q_l} & \text{if } l < m \\ \prod_{k=1}^{m-1} \frac{q_{k,0,0}}{(1-q_k)} \frac{q_{m,0,1}}{q_m} \prod_{k=m+1}^{l-1}(1 - q_k)\, q_l & \text{if } m < l \\ \prod_{k=1}^{l-1} \frac{q_{k,0,0}}{(1-q_k)} \frac{q_{l,1,1}}{q_l} & \text{if } l = m. \end{cases} \tag{14}$$

If we proceed along the lines that lead us to (12), we can show that for any given m, based on conditional reasoning, (14) defines a stick-breaking partition of the unit segment. However, this is not obtained through the usual Beta$(1, \nu)$ random variables, even if the process is marginally Dirichlet. In fact, the random measure arising from (14) can be seen as a generalized Dirichlet process, in the spirit of the more general definitions of Hjort (2000) and Ishwaran & James (2001).

As a final remark, notice that defining a stick-breaking construction does not necessarily ensure that the random weights sum to one with probability one. This depends on the distribution of the weights. In the context of univariate stick-breaking priors, however, it is possible to provide a necessary and sufficient condition for that to happen (see Lemma 1 in Ishwaran & James (2001)). We can expect that this condition holds for our model too, as long as we marginally get a DP prior (or, more in general, a stick-breaking prior). The precise argument is a direct extension of the result of Ishwaran and James and is developed for the bivariate case in Appendix I as Lemma 2. Extension to the general n-dimensional case is again straightforward.

3.2 Mixing using a Generalized Spatial Dirichlet Process.

Following Gelfand et al. (Gelfand et al. (2005)), as in Section 2, the GSDP will be used to model the distribution of the spatial component $\theta(s)$ in a random effect model of the type

$$Y(s) = \mu(s) + \theta(s) + \varepsilon(s),$$

where $\mu(s)$ 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 pure error (nugget) component with mean zero and variance τ^2. If $\theta(\cdot)$ follows the GSDP as above, we can provide analogues of expressions (3) and (4). Again, if we denote by $G^{(n)}$ the finite dimensional distributions defined by (6), for any finite set of locations $s^{(n)} = (s_1, \ldots, s_n)$, $n \in \mathbb{N}$, the joint distribution of $Y = \{Y(s_1), \ldots, Y(s_n)\}^T$, given $G^{(n)}$, μ and τ^2 is given by $F\left(y|\, G^{(n)}, \mu, \tau^2\right) = \int N_n\left(y|\, \theta + \mu, \tau^2 I_n\right)\, G^{(n)}(d\theta)$, where $\theta = \{\theta(s_1), \ldots, \theta(s_n)\}^T$, $\mu = \{\mu(s_1), \ldots, \mu(s_n)\}^T$. Again, differentiating to densities,

$$f\left(y|G^{(n)}, \mu, \tau^2\right) = \int N_n\left(y|\, \theta + \mu, \tau^2 I_n\right)\, G^{(n)}(d\theta). \tag{15}$$

As with the SDP, since $G^{(n)}$ is almost surely discrete, with probability one the conditional density (15) can be rewritten as a countable location mixture of normals,

$$f\left(Y|G^{(n)}, \mu, \tau^2\right) = \sum_{i_1=1}^{\infty} \ldots \sum_{i_n=1}^{\infty} p_{i_1,\ldots,i_n} N_n(Y|\, \theta_{i_1,\ldots,i_n} + \mu, \tau^2 I_n), \tag{16}$$

where, for simplicity, we have suppressed the locations in $p_{i_1,\ldots,i_n}$ and set the vector $\theta_{i_1,\ldots,i_n} = \{\theta_{i_1}(s_1), \ldots, \theta_{i_n}(s_n)\}^T$. Computation of the moments of this distribution is immediate. Y is a random vector which with probability one has density absolutely continuous with respect to the Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, expected value $E(Y|G^{(n)}, \mu, \tau^2) = \sum_{i_1=1}^{\infty} \ldots \sum_{i_n=1}^{\infty} p_{i_1,\ldots,i_n}\theta_{i_1,\ldots,i_n} + \mu$, and covariance matrix $\Sigma_Y|G^{(n)}, \mu, \tau^2 = \tau^2\, I_n + \Sigma_\theta^s$, where $\left(\Sigma_\theta^s\right)_{i,j} = cov\left\{\theta(s_i), \theta(s_j)|G^{(n)}\right\}$ is given by (10).

Under the assumptions of Proposition 1, if, in addition, the mean vector μ describes a continuous surface over D, it is easy to prove that an analogous statement 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 to (15), 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 $||s - s_0|| \to 0$.

4 The Spatially Varying Probabilities Model

In this section we discuss how to specify the stick-breaking components in a way that is appealing for modelling purposes and ensures the existence of the processes sampled from G. In particular, our constructive approach is through latent variables. In the implementation, using MCMC, we never sample or even estimate the q's or p's. We conclude this section with a discussion that distinguishes our approach from that of Griffin & Steel (2004).

Recalling the notation of Section 3.1, for any $n = 1, 2, \ldots$ and any $l = 1, 2, \ldots$ the stick-breaking components $q_{l,u_1,\ldots,u_n}(s_1, \ldots, s_n), u_j \in \{0, 1\}, j = 1, 2, \ldots, n$ arise through probabilities associated with the events $\Theta_l^{u_j}(s_j)$, $l = 1, 2, \ldots$. Therefore, it is possible to assign a distribution to the stick-breaking components directly by specifying a law for these events. In particular, we can consider the process $\{\delta_{\Theta_l^1(s)}, s \in D, l = 1, 2, \ldots, \}$, such that at any $l = 1, 2, \ldots$,

$$\delta_{\Theta_l^1(s)} = 1 \quad \text{if } \Theta_l^1(s) \text{ occurs}$$

$$\delta_{\Theta_l^1(s)} = 0 \quad \text{if } \Theta_l^1(s) \text{ does not occur.}$$

In particular, suppose $\Theta_l^1(s)$ occurs if and only if $Z_l(s) \in A_l(s)$. Then, we can work

with the equivalent stochastic process $\{\delta^*_{A_l(s)},\, s \in D,\, l = 1, 2, \ldots\}$ defined by

$$\delta^*_{A_l(s)} = 1 \quad \text{if } Z_l(s) \in A_l(s)$$

$$\delta^*_{A_l(s)} = 0 \quad \text{if } Z_l(s) \notin A_l(s),$$

where $\{Z_l(s),\, s \in D,\, l = 1, 2, \ldots\}$ is a latent random field. Furthermore, we can write

$$\begin{aligned} q_{l,u_1,\ldots,u_n}(s_1, \ldots, s_n) &= pr\{\delta_{\Theta^1_l(s_1)} = u_1, \ldots, \delta_{\Theta^1_l(s_n)} = u_n |\, \delta_{\Theta^1_i(s_j)} = 0,\, i < l,\, j = 1, \ldots, n\} \\ &= pr\{\delta^*_{A_l(s_1)} = u_1, \ldots, \delta^*_{A_l(s_n)} = u_n |\, \delta^*_{A_i(s_j)} = 0,\, i < l,\, j = 1, \ldots, n\}. \end{aligned}$$

It is easy to see that such a characterization guarantees that (7) is true, hence the existence of the processes sampled from the random distribution (6).

In the following we assume that $\{Z_l(s),\, s \in D,\, l = 1, 2, \ldots\}$ 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 lth process, say $\mu_l(s)$, is unknown and we put a convenient prior on it, so that the distribution of $Z_l(s)$ (and hence of the q_l's) can be viewed as random. We also choose $A_l(s) = \{Z_l(s) \geq 0\}$. With these assumptions, it follows that

$$q_{l,u_1,\ldots,u_n}(s_1, \ldots, s_n) = pr\{\delta^*_{\{Z_l(s_1)\geq 0\}} = u_1, \ldots, \delta^*_{\{Z_l(s_n)\geq 0\}} = u_n |\, \mu_l(s_1), \ldots, \mu_l(s_n)\},$$

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

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

$$q_{l,1}(s) = pr\{Z_l(s) \geq 0\} = 1 - \Phi\{-\mu_l(s)\} = \Phi\{\mu_l(s)\}, \tag{17}$$

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 $\text{Be}(1, \nu)$, $l = 1, 2, \ldots$, then for each s, the

marginal distribution of $\theta(s)$ is a DP with probabilities that vary with location. In the special case that $\mu_l(s) = \mu_l$, for all s, with $\Phi(\mu_l)$ independent Be$(1, \nu)$ then, again marginally, the $\theta(s)$ follow a DP where the marginal weights are same for each s but the marginal distributions are not the same since $\theta_l^*(s) \neq \theta_l^*(s')$.

Marginal reduction to a DP is not necessary for the definition of the GSDP (although it can be useful for purposes of comparison with the SDP or other competing approaches). For instance, if we retain the $\mu_l(s)$, then, since we would like to encourage $Z_l(s)$ to resemble $Z_l(s')$ when s is close to s', we could take $\mu_l(s)$ to be a realization of say a Gaussian spatial process rather than say independent as above.

We have described the construction of a flexible model for the spatial random effects $\theta(s)$ in order that they can come from different random spatial surfaces at different locations. Following the discussion above Proposition 1, we require two properties for this construction: (i) the random finite dimensional distribution $G^{(n)}$ satisfies the Kolmogorov consistency condition and (ii) the continuity property should be satisfied, that is, if location s is near s', we want the probability of picking up the same sample surface for s and s' to be high.

To recapitulate, we will never actually calculate the random weights $p_{i_1,\ldots,i_n}$. Rather, as is frequently done in hierarchical modeling, we have introduced latent variables, in this case a countable collection of independent Gaussian process realizations. And we let

$$\begin{aligned} p_{i_1,\ldots,i_n} = & pr\Big[Z_1(s_1) < 0, \ldots, Z_{i_1-1}(s_1) < 0, Z_{i_1}(s_1) \geq 0; \\ & Z_1(s_2) < 0, \ldots, Z_{i_2-1}(s_2) < 0, Z_{i_2}(s_2) \geq 0; \ldots; \\ & Z_1(s_n) < 0, \ldots, Z_{i_n-1}(s_n) < 0, Z_{i_n}(s_n) \geq 0 | \{\mu_l(s_i)\}\Big], \end{aligned} \tag{18}$$

In Propositions 2 and 3 of the Appendix we prove that the construction above satisfies the Kolmogorov consistency and the continuity conditions.

Finally, spatially varying weights have recently been considered by Griffin & Steel (2004), who work in the framework of dependent Dirichlet processes. They proceed from the assumption that the distribution of a $DP(\nu\, G_0)$ is unaffected by a permutation of the atoms $\{\theta_l^*(\cdot), q_l(\cdot), l = 1, 2, \ldots\}$ in Sethuraman's constructive representation. Then, if $\{\pi(s),\, s \in D\}$ is a process of permutations of the set of integers $\{1, 2, \ldots\}$, it is possible to define an order-based dependent stick-breaking prior over D, abbreviated πDDP as a process $\{F_\pi(s),\, s \in D\}$, such that at any $s \in D$, given a realization of the process $\pi(s)$, $F_\pi(s) = \sum_{l=1}^{\infty} p_l(s)\, \delta_{\theta_l(s)}$, where $p_l(s) = q_{\pi_l(s)} \prod_{j<l} \left\{1 - q_{\pi_j(s)}\right\}$.

With regard to surface selection, the difference between their approach and ours is as follows. We define a joint random distribution for any grouping of the locations $(s_1, \ldots, s_n)$, $n = 1, 2, \ldots$ and the probabilities of picking up the different surfaces are directly assigned. For instance, for n=2 and any integers l and m, we have seen that $pr\left\{Y(s_i) = \theta_l^*(s_i), Y(s_j) = \theta_m^*(s_j)\right\} = p_{l,m}(s_i, s_j)$. For Griffin and Steel's πDDP, this probability is given by

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

that is, as the expected value of the marginal probabilities with respect to the distribution of the permutation field at the two locations. By the definition of πDDP, it follows that the dependence among the marginal random distribution functions is directly deduced by the permutation at each s. In particular, this is given by means of an auxiliary latent point process Z. In fact, Griffin and Steel first associate each atom $\{\theta_i^*(s), q_i\}$ with a realization z_i from Z, for $i = 1, 2, \ldots$. Then, at any s, the πDDP is defined permuting the set of $q's$ according to the realizations of the latent point process Z. In fact, $\pi(s)$ is defined to satisfy $||s - z_{\pi_1(s)}|| < ||s - z_{\pi_2(s)}|| < \ldots$. It follows that a realization from this process will necessarily be the same for some regions of D, while allowing different stick-breaking constructions for points far apart

from each other. However, the representation at any s depends on how the process Z is associated with the atoms of the process, so that the representation does not seem to be invariant to a reordering of the z's. Moreover, for practical purposes it can be difficult to model the type of dependence induced by the point process mechanism, unless we choose simple processes, such as a stationary Poisson process. On the other hand, in our approach the spatial behavior of the stick-breaking components depends on the distribution of the latent Gaussian process Z and can vary across locations if this is true for the mean of Z.

5 Simulation Based Model fitting for the GSDP

Assembling Sections 3 and 4, we work with the following spatial model. Let the vectors $Y_t = \{Y_t(s_1), Y_t(s_2), \ldots, Y_t(s_n)\}^T, t = 1, \ldots, T$ indicate T groups of independent observations collected at the same set of locations $(s_1, \ldots s_n) \in D \subset \mathbb{R}^2$. The mean surface $\mu(s), s \in D$ is modelled by a linear regression $\mu(s) = x(s_i)^T\beta$. The spatial random effect $\theta(s), s \in D$ has the nonparametric rule as defined in Section 4. The

overall model has the following hierarchical structure

$$
\begin{aligned}
Y_t \mid \theta_t, \beta, \tau^2 &\sim N_n(Y_t \mid X_t^T\beta + \theta_t, \tau^2 I_n), \quad t = 1, ..., T \\
\theta_t \mid G^{(n)} &\sim G^{(n)}, \quad t = 1, ..., T \\
G^{(n)} \mid p_{i_1,...,i_n}, \theta_l^* &= \sum_{i_1,...,i_n=1}^{\infty} p_{i_1,...,i_n}\, \delta_{\theta^*_{i_1}(s_1)}(\cdot) \cdots \delta_{\theta^*_{i_n}(s_n)}(\cdot), \quad l = 1, 2, \ldots \\
p_{i_1,...,i_n} &= pr\left\{Z_1(s_l) < 0, \ldots, Z_{i_l}(s_l) \geq 0,\, l = 1, \ldots, n\right\}, i_j = 1, 2, \ldots, j = 1, 2, \ldots, n. \\
\left\{\theta_l^*(s_1), \ldots \theta_l^*(s_n)\right\}^T &\sim N_n\left(0, \sigma^2 R_n(\phi)\right), \quad l = 1, 2, \ldots \\
\left\{Z_{t,l}(s_1), \ldots, Z_{t,l}(s_n)\right\}^T &\sim N_n\left(\mu_l 1_n, H_n(\eta)\right), \quad l = 1, 2, \ldots, \, t = 1, 2, ...T \\
\mu_l \text{ s.t. } \Phi(\mu_l) &\sim Beta(1, \nu), \quad l = 1, 2, \ldots \\
\beta, \tau^2 &\sim N_p(\beta \mid \beta_0, \Sigma_\beta) \times \text{IGamma}(\tau^2 \mid a_\tau, b_\tau) \\
\sigma^2,\, \phi,\, \eta &\sim \text{IGamma}(\sigma^2 \mid a_\sigma, b_\sigma) \times [\phi] \times [\eta],
\end{aligned}
\tag{19}
$$

The priors for ϕ and η depend on the specific form of covariance structure of $R_n(\phi)$ and $H_n(\eta)$. For convenience, in our examples we have set $\nu = 1$. In the more general version, we have $\mu_l(s)$ replacing μ_l and, for each l, we obtain a realization from a Gaussian process with mean 0 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 the marginal random distribution at an individual location s follows a Dirichlet process, the joint random distribution $G^{(n)}$ does not. The traditional method of marginalizing over $G^{(n)}$ so that the θ_t, $t = 1, \ldots T$ follow a Polya urn scheme can not be used in this case. Instead, we approximate $G^{(n)}$ with a finite sum

$$
G_K^{(n)} = \sum_{(i_1,...,i_n) \in \{1,2,...,K\}^n} p_{i_1,...,i_n} \quad \delta_{\theta^*_{i_1}(s_1)}(\cdot)\, \delta_{\theta^*_{i_2}(s_2)}(\cdot) \ldots \delta_{\theta^*_{i_n}(s_n)}(\cdot), \tag{20}
$$

for K suitably large. In this finite mixture model, we only need $\theta_l^*, l = 1, \ldots, K$ and

$Z_l, l = 1, \ldots, K-1$. Note that $p_K(s) = pr\{Z_1(s) < 0, \ldots, Z_{K-1}(s) < 0\}$.

Recalling remarks in Section 4, actual computation of the weights $p_{i_1,\ldots,i_n}$ in (20) is very difficult because it involves evaluation of multivariate normal cdf's. It is avoided if we sample and use the latent variables Z_l's directly. We eliminate the sampling of the conditional distribution $[\theta_t|G_K^{(n)}]$ by referring to the following equivalent structure:

$$\theta_t(s) = \theta_1^*(s) I_{\{Z_t^1(s)\geq 0\}} + \theta_2^*(s) I_{\{Z_t^1(s)<0, Z_t^2(s)\geq 0\}} + \ldots + \theta_K^*(s) I_{\{Z_t^1(s)<0,\ldots,Z_t^{K-1}(s)<0\}}. \quad (21)$$

In equation (21), $\theta_t(s)$ is a deterministic function of $\theta_l^*(s); l = 1, \ldots, K$ and $Z_t^l(s); l = 1, \ldots, K-1$. We rewrite the first stage of the hierarchical model as $[Y_t|\mu, \theta_t] = [Y_t|\mu, \theta^*, Z_t]$. Then, the likelihood function for Y_t can be expressed as

$$\begin{aligned}
[Y_t|\mu, \theta^*, Z_t] &\propto \exp\left[-\frac{1}{2\tau^2}\sum_{i=1}^{n}\left\{Y_t(s_i) - X_t^T\beta(s_i) - \theta_t(s_i)\right\}^2\right] \\
&\propto \exp\left[-\frac{1}{2\tau^2}\sum_{l=1}^{K}\sum_{i=1}^{n}\left\{Y_t(s_i) - X_t^T\beta(s_i) - \theta_l^*(s_i)\right\}^2 I_{\{Z_t^1(s)<0,\ldots,Z_t^{l-1}(s)<0, Z_t^l(s)\geq 0\}}\right] \\
&\propto \prod_{i=1}^{n}\left(\sum_{l=1}^{K}\exp\left[-\frac{1}{2\tau^2}\left\{Y_t(s_i) - X_t^T\beta(s_i) - \theta_l^*(s_i)\right\}^2\right]\times\right. \\
&\left.\times\, I_{\{Z_t^1(s)<0,\ldots,Z_t^{l-1}(s)<0, Z_t^l(s)\geq 0\}}\right),
\end{aligned}$$

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

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

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

6 Data Illustration

We illustrate the fitting of our model (18) with a simulated data set where we simulate from a finite mixture model of Gaussian processes that allows different joint multimodal distributions for different pairs of locations.

We first simulate a specified number of locations in a given region. They are denoted as $(s_1, \ldots, s_n)$. Suppose there are T independent replicates $\{y_t(s_1), \ldots, y_t(s_n)\}$, $t = 1, \ldots, T$ sampled from a mixture distribution.

In particular, we proceed as follows. For $t = 1, \ldots, T$, let $\{\theta_t^1(s_1), \ldots, \theta_t^1(s_n)\}^T \sim N_n^{(1)}(-\mu 1_n, \sigma_1^2 R_n(\phi_1))$ and $\{\theta_t^2(s_1), \ldots, \theta_t^2(s_n)\}^T \sim N_n^{(2)}(\mu 1_n, \sigma_2^2 R_n(\phi_2))$. Also, let $\{Z_t(s_1), \ldots, Z_t(s_n)\}^T \sim N_n(0, H_n(\eta))$. Then, for $i = 1, \ldots, n$, if $Z_t(s_i) \geq 0$, we set $y_t(s_i) = \theta_t^1(s_i)$; if $Z_t(s_i) < 0$, let $y_t(s_i) = \theta_t^2(s_i)$.

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 $y_t(s_i)$ and $y_t(s_j)$ very likely to be 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, therefore the component choices for $y_t(s_i)$ and $y_t(s_j)$ are almost independent. The joint histogram plots in Figure 3 below demonstrate these special properties of our simulation model.

In our experiment we simulate at 50 design locations in a rectangular area shown in Figure 2. Notice that some of the locations are numbered for future reference. 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 mixture model above. We fit the model specified in section 5 to this data set. 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 $\{y_t(s_1), \ldots, y_t(s_n)\}$ to be zero.

The Bayesian goodness of fit is illustrated by the posterior predictive densities. In this example, we show not only the marginal posterior predictive density at each location, but also the joint posterior predictive densities for two locations. In Figure 3 we plot the posterior predictive density for four randomly selected locations. They are locations $s_5, s_{11}, s_{26}, s_{33}$. The thick density curves are the predictive densities estimated from our model. The thin density curves represent the true densities of the model from which we simulated the data. The +'s mark the values of the 40 observations at each of these 4 locations.

We select 2 pairs of sites to show the predictive joint densities. The first pair $\{y(s_{50}), y(s_{23})\}$ are close to each other. The second pair $\{y(s_{50}), y(s_{49})\}$ are distant. The left most column of plots in Figure 4 show the predictive joint histograms of the couples $\{y(s_{50}), y(s_{23})\}$ and $\{y(s_{50}), y(s_{49})\}$. The joint histogram of $\{y(s_{50}), y(s_{23})\}$ shows two highly correlated sample clouds. The joint histogram of $\{y(s_{50}), y(s_{49})\}$ shows four less correlated sample clouds. The middle column shows samples from the true joint densities, while the last column shows the histogram formed from the 40 observations at the two pairs of locations. We can see that, even with a relatively small sample size, our model reasonably well captures the joint distribution.

Based upon the posterior samples, Figure 5 is the plot of 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.

7 A Spatio-temporal Dynamic Model Version

In Section 5, we assumed the $Y_t = \{y_t(s_1), \ldots, y_t(s_n)\}^T, t = 1, \ldots, T$ to be independent replicates. In practice, these observations are usually made in T consecutive time

periods, so it is more realistic to model the evolution of the spatial process over time. In this section we present a version of the spatio-temporal model by embedding the GSDP in a dynamic linear model. We illustrate this spatio-temporal model also by fitting it to a simulated data set.

Preserving the notation in (18), the observations at time t can be modelled by the following dynamic linear model structure:

$$\begin{aligned} Y_t &= X_t^T \beta + \theta_t + \varepsilon_t;\ \varepsilon_t \sim N\left(0, \tau^2 I_n\right) \\ \theta_t &= \gamma \theta_{t-1} + \omega_t;\ \omega_t \sim GSDP\left(\nu G_0\right) \end{aligned} \tag{22}$$

These dynamics yield spatial random effects θ_t that evolve autoregressively over time with autocorrelation coefficient, $\gamma\left(|\gamma| \leq 1\right)$. Only the second hierarchical specification in (18) changes to reflect (22). Updating of the full conditional distributions and the associated MCMC algorithm for the dynamic version is straightforward but careful attention to bookkeeping is required. We detail it in Appendix II.

We illustrate the model above by with a simulated data set. We still use the region and the 50 locations given in Figure 2. However, we add 4 new locations (with no observations) labelled 51-54 where we seek to predict. Also, a simple linear regression of $\beta_0 + \beta_1 X\left(s_i\right)$ is added to the model. $X\left(s_i\right)$ denotes the distance from location s_i to a fixed point source represented by the diamond in Figure 2.

The simulated observations $\left\{y_t\left(s_1\right), \ldots, y_t\left(s_n\right)\right\}^T$, $t = 1, \ldots, T$ are sampled again from a mixture of two distributions as follows. Following the specifications from the previous section, now consider $\left\{\omega_t^1\left(s_1\right), \ldots, \omega_t^1\left(s_n\right)\right\} \sim N_n^{(1)}\left(-\mu 1_n, \sigma_1^2 R_n\left(\phi_1\right)\right)$ and $\left\{\omega_t^2\left(s_1\right), \ldots, \omega_t^2\left(s_n\right)\right\} \sim N_n^{(2)}\left(\mu 1_n, \sigma_2^2 R_n\left(\phi_2\right)\right)$. Also, let $\left\{Z_t\left(s_1\right), \ldots, Z_t\left(s_n\right)\right\} \sim N_n\left(0, H_n\left(\eta\right)\right)$. Then, for $i = 1, \ldots, n$, if $Z_t\left(s_i\right) \geq 0$, we set $\theta_t\left(s_i\right) = \gamma\, \theta_{t-1}\left(s_i\right) + \omega_t^1\left(s_i\right)$; if $Z_t\left(s_i\right) < 0$, $\theta_t\left(s_i\right) = \gamma\, \theta_{t-1}\left(s_i\right) + \omega_t^2\left(s_i\right)$ for $i = 1, \ldots, n$. Then we obtain $y_t\left(s_i\right) = \beta_0 + \beta_1 X\left(s_i\right) + \varepsilon_t\left(s_i\right)$, where $\varepsilon_t\left(s_i\right)$ is sampled independently from the normal distribution $N\left(0, \tau^2\right)$.

We choose the same values for the parameters of $\mu_1 = -\mu_2 = 3$. $\sigma_1 = 2\sigma_2 = 2$, $\phi_1 = \phi_2 = 0.3$ and $\eta = 0.3$ as in Section 6. Also, γ is chosen as 0.7, τ^2 is 9. β_0 and β_1 are 2 and -1 respectively, and T is equal to 40.

In fitting a model to the data, we use the same GSDP for ω_t as given in Section 6 with $K = 20$ components. The Bayesian posterior mean of γ is found to be 0.8 in our experiment. The Bayesian goodness of fit is again illustrated by the posterior predictive densities at $T+1$. We show not only the marginal posterior predictive density at each location, but also the joint posterior predictive densities for two locations.

In Figure 6 we plot the posterior predictive density at $T+1(=41)$ for two locations with observation (labelled 6 and 12 on Figure 2) and two new locations (labelled 51 and 54 on Figure 2). The thick density curves are the predictive densities estimated from our model. The thin density curves represent the true densities of the model from which we simulated the data. The results are interesting in that, despite the small sample size and the introduction of bimodality only through the innovations at the second stage, we find bimodal behavior at 6 and 12. Location 51 is not very close to any of the sampled locations and, in the absence of data, yields a unimodal predictive density. However, location 54 is very close to sampled locations and, reflecting the mean square continuity of the GSDP, an indication of two modes emerges.

Turning to bivariate predictive densities, we select 2 pairs of sites to show the predictive joint density at $T+1$. The first pair s_{50}, s_{23} are close to each other. The second pair s_{50}, s_{49} are much farther apart. In Figure 7, we provide perspective plots of the predictive and true joint densities. The first pair reveals a bimodal joint density while the second pair shows a density with four modes. If one were interested in developing simultaneous highest posterior density (HPD) confidence sets, one needs to identify the "footprint" associated with a level surface of the joint density. In particular, one must choose the level to provide a specified posterior probability. Of course, these bi-

variate densities are unavailable analytically so, using the posterior samples, we obtain a bivariate kernel density estimator. However, since level surfaces associated with this density estimate are still difficult to obtain, we evaluate the density estimate at the observed samples, providing an ordering for the samples. Then, according to the desired probability, we choose the density ordinate such that the proportion of the sample with ordinate above this value is the probability we seek. Figure 8 provides illustrative 80 % (inner curve) and 95 % (outer curve) HPD's for the site pairs in Figure 7.

8 Discussion and Summary

We have introduced the GSDP as a more flexible successor to the SDP proposed by Gelfand et al. (2005). However, any multivariate density can be approximated by a suitable countable mixture of multivariate normal densities. Since, for any finite set of locations, with probability one the SDP mixture model is such a countable mixture model, what practical advantages can the GSDP offer over the SDP? Why would we take the trouble to implement the much more computationally demanding GSDP? For example, with a bivariate distribution that is the product of two bimodal univariate distributions, while the GSDP might capture such a distribution using essentially two components, wouldn't the SDP be able to do it with four components?

In fact, while, in principle, the SDP can equally well find multiple modes in say a bivariate distribution, it will have a more difficult time distinguishing the joint distribution for points close to each other from the joint distribution for points far apart. In other words, in practice, the normal mixture model in (16) can more quickly adapt to the data than the normal mixture model below (3). Expressed in different terms, in requiring additional components, the SDP will run into the, a priori, geometrically decaying weights, so it may struggle to properly allocate mass to the modes. Furthermore,

consider the situation where we might have varying numbers of mixing components and these might vary with spatial location. The version of the GSDP which allows different marginal distributions at each s is better suited to handle this. In summary, one can not lose by adopting the more general GSDP and there are circumstances where the use of the SDP model might require much more data to capture desired features than the GSDP would require.

Other extensions of the SDP can be envisioned. For instance, in a future manuscript we will report on the use of the representation of Ishwaran & Zarepour (2002a), Theorem 3 rather than the Sethuraman representation, to create a different constructive formulation. Other future investigation will take us to the case of modeling discrete data, e.g., binary or count data at the first stage with a GSDP to model the spatial structure in the mean on a transformed scale. We are also interested in the case where we observe multivariate data at each location. GSDP's centered around multivariate spatial process models provide an obvious place to start.

9 Acknowledgement.

The authors acknowledge Sonia Petrone for insightful comments.

10 Appendices

We offer two appendices. In Appendix I we provide the arguments for the technical results concerning the GSDP, given in Sections 3 and 4. In Appendix II, we provide the full conditional distribution theory needed to fit models incorporating the GSDP.

Appendix I: Theoretical arguments

Proof of Proposition 1:

Proof. Consider two sites s, s_0 in D. According to (6) the joint distribution of the process is almost surely a realization of the random element

$$pr\{Y(s) \in A, Y(s_0) \in B\} = \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} p_{l,m}(s, s_0)\, \delta_{\theta_l^*(s)}(A)\, \delta_{\theta_m^*(s_0)}(B),$$

for all $A, B \in \mathcal{B}(\mathbb{R})$. Notice that

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

because of the almost sure continuity property of the weights. The interchange between limit and sum operations in the equation above follows from the dominated convergence theorem, since $p_{l,m}(s, s_0) \leq p_m(s_0)$ for all m. Since $0 \leq p_l(s)\, \delta_{\theta_l^*(s)}(A) \leq p_l(s)$ and $\sum_{l=1}^{\infty} p_l(s) = 1$ for all s, we can apply Fatou's Lemma for the series in order to justify

$$\begin{aligned}\lim_{||s-s_0||\to 0} pr\{Y(s) \in A\} &= \lim_{||s-s_0||\to 0} \sum_{l=1}^{\infty} p_l(s)\, \delta_{\theta_l^*(s)}(A) = \sum_{l=1}^{\infty} p_l(s_0)\, \delta_{\theta_l^*(s_0)}(A) \\ &= pr\{Y(s_0) \in A\},\end{aligned}$$

which shows the almost sure convergence of the marginal random distributions. ■

Lemma 1 *The probabilities*

$$\begin{aligned} q_{1,u,v}(s_i, s_j) &= pr\{\Theta_1^u(s_i), \Theta_1^v(s_j)\} \\ q_{l,u,v}(s_i, s_j) &= pr\{\Theta_l^u(s_i), \Theta_l^v(s_j) | \Theta_m^0(s_i), \Theta_m^0(s_j), m < l\},\ l \geq 2, \end{aligned}$$

$u, v \in \{0, 1\}$, are such that $q_{l,1,+}(s_i, s_j) = q_l(s_i)$ and $q_{l,+,1}(s_i, s_j) = q_l(s_j)$, for any $l = 1, 2, \ldots$.

Proof. Recall that

$$q_{1,u,v}(s_i, s_j) = pr\{\Theta_1^u(s_i), \Theta_1^v(s_j)\}$$

$$q_{l,u,v}(s_i, s_j) = pr\{\Theta_l^u(s_i), \Theta_l^v(s_j) | \Theta_m^0(s_i), \Theta_m^0(s_j), m < l\}, \quad l \geq 2, u, v \in \{0, 1\}.$$

Hence,

$$q_{l,1,+}(s_i, s_j) = pr\{\Theta_l^u(s_i) | \Theta_m^0(s_i), \Theta_m^0(s_j), m < l\}, \quad l \geq 2, u, v \in \{0, 1\}.$$

But $\Theta_l^u(s_i)$ is independent of $\{\Theta_m^0(s_j), m < l\}$ given $\{\Theta_m^0(s_i), m < l\}$ by the definition of stick-breaking. Since $q_l(s_i) = pr\{\Theta_l^u(s_i) | \Theta_m^0(s_i), m < l\}$, we are done. ■

Lemma 2 *For any given s_i, s_j in D,*

$$\sum_{l=1}^{\infty} \sum_{m=1}^{\infty} p_{l,m}(s_i, s_j) = 1 \quad \textit{if and only if} \quad \sum_{l=1}^{\infty} E\left[\log\{1 - q_l(s_i)\}\right] = -\infty. \tag{23}$$

Proof. Necessity follow after noticing that, if we marginalize with respect to s_i, condition (23) reduces to condition (5) in Ishwaran & James (2001). Now, consider for any $N, M = 1, 2, \ldots$, the remainder term

$$R_{N,M}(s_i, s_j) = 1 - \sum_{l=1}^{N} \sum_{m=1}^{M} p_{l,m}(s_i, s_j),$$

and assume condition (23) holds. We need to prove that $R_{N,M}(s_i, s_j) \to 0$ with probability one as $N, M \to \infty$. It's easy to see that

$$R_{N,M}(s_i, s_j) = \sum_{l=1}^{N} \sum_{m=M+1}^{\infty} p_{l,m}(s_i, s_j) + \sum_{m=1}^{M} \sum_{l=N+1}^{\infty} p_{l,m}(s_i, s_j) + \sum_{l=N+1}^{\infty} \sum_{m=M+1}^{\infty} p_{l,m}(s_i, s_j). \tag{24}$$

Since all the terms in the sums are positive, to show $R_{N,M}(s_i, s_j) \to 0$ it is necessary and sufficient that all the series tend to zero, as $N, M \to \infty$. Then we can work with each of them separately. Consider the first term in the sum and substitute (12) to all $p_{l,m}(s_i, s_j)$, so that

$$\sum_{l=1}^{N} \sum_{m=M+1}^{\infty} p_{l,m}(s_i, s_j) = \sum_{l=1}^{N} \prod_{k=1}^{l-1} q_{k,0,0}(s_i, s_j)\, q_{l,1,0}(s_i, s_j) \sum_{m=M+1}^{\infty} \prod_{r=l+1}^{m-1} \{1 - q_r(s_j)\}\, q_m(s_j). \tag{25}$$

Notice that, for any $l = 1, 2, \ldots, N$,

$$\sum_{m=M+1}^{\infty} \prod_{r=l+1}^{m-1} \{1 - q_r(s_j)\}\, q_m(s_j) = \frac{\sum_{m=M+1}^{\infty} \prod_{r=1}^{m-1} \{1 - q_r(s_j)\}\, q_m(s_j)}{\prod_{r=1}^{l} \{1 - q_r(s_j)\}}$$
$$= \frac{\sum_{m=M+1}^{\infty} p_m(s_j)}{1 - \sum_{m=1}^{l} p_m(s_j)}.$$

Therefore, if we let $M \to \infty$, the numerator tends to 0, because it is the remainder of the sum of the weights for the marginal model in s_j, and we can refer again to the result in Ishwaran & James (2001). Then, each term of the series in (25) tends to 0 as $M \to \infty$. So,

$$\lim_{N\to\infty} \lim_{M\to\infty} \sum_{l=1}^{N} \sum_{m=M+1}^{\infty} p_{l,m}(s_i, s_j) = 0.$$

We can follow a similar argument for the second remainder term in (24). Now consider

$$\sum_{l=N+1}^{\infty} \sum_{m=M+1}^{\infty} p_{l,m}(s_i, s_j).$$

Let $\tau = min(N, M)$. Then,

$$\sum_{l=N+1}^{\infty} \sum_{m=M+1}^{\infty} p_{l,m}(s_i, s_j) \leq \sum_{l=\tau+1}^{\infty} \sum_{m=\tau+1}^{\infty} p_{l,m}(s_i, s_j) = \prod_{k=1}^{\tau} q_{k,00}(s_i, s_j).$$
$$\leq \prod_{k=1}^{\tau} \{1 - q_k(s_i)\},$$

since $q_{k,0,0}(s_i, s_j) < 1 - q_k(s_i)$, $k = 1, \ldots,$. Then the desired result follows again from the Lemma 1 in Ishwaran & James (2001) for the marginal model in s_i. ■

We next turn to the argument regarding satisfaction of the Kolmogorov consistency and continuity conditions.

Proposition 2 *Let* $\{Y(s_1), Y(s_2), \ldots Y(s_n), s_i \in D, i = 1, \ldots n\}$ *have random finite dimensional distribution given by* (6)*, for* $n = 1, 2, \ldots$. *If the set of weights* $\{p_{i_1,\ldots,i_n}\}$ *is defined by means of a latent process as in* (18)*, then the collection of random finite dimensional distributions define a random field* $Y(s)$ *on* D.

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

$$p_{i_1,\ldots,i_{l-1},i_{l+1},\ldots i_n} = p_{i_1,\ldots,i_{l-1},\cdot,i_{l+1},\ldots i_n} = \sum_{k=1}^{\infty} p_{i_1,\ldots,i_{l-1},k,i_{l+1},\ldots i_n}. \tag{26}$$

In fact, let $Z(s_i) = \{Z_1(s_i), \ldots, Z_k(s_i), \ldots\}$, $i = 1, \ldots, n$. Note that if $\theta(s_i) = \theta_k^*(s_i)$, then $Z(s_i) \in S_{i,k}$, where $S_{i,k} = (-\infty, 0)_1 \times \cdots \times (-\infty, 0)_{k-1} \times [0, \infty)_k \times \mathbb{R} \times \cdots$. Therefore, we can rewrite

$$\begin{aligned} p_{i_1,\ldots,i_{l-1},i_{l+1},\ldots i_n} =& pr\Big\{ Z(s_1) \in S_{1,i_1}, \ldots, Z(s_{l-1}) \in S_{l-1,i_{l-1}}, \\ & Z(s_{l+1}) \in S_{l+1,i_{l+1}}, \ldots, Z(s_n) \in S_{n,i_n} \Big\} \\ \text{and} \quad p_{i_1,\ldots,i_{l-1},k,i_{l+1},\ldots i_n} =& pr\Big\{ Z(s_1) \in S_{1,i_1}, \ldots, Z(s_{l-1}) \in S_{l-1,i_{l-1}}, Z(s_l) \in S_{l,k}, \\ & Z(s_{l+1}) \in S_{l+1,i_{l+1}}, \ldots, Z(s_n) \in S_{n,i_n} \Big\}. \end{aligned}$$

By the continuity of probability measure,

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

Since $\bigcup_{k=1}^{\infty} S_{l,k} = \bigotimes_{k=1}^{\infty} \mathbb{R}$, (26) follows.

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

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

Proposition 3 *Let* $\{Y(s),\, s \in D\}$ *be as in Proposition 2. If the base random field* G_0 *is almost sure continuous, then for all* $s_0 \in D$, $Y(s)$ *converges weakly to* $Y(s_0)$ *with probability one as* $||s - s_0|| \to 0$.

Proof. The proof follows immediately from Proposition 1, once we notice that, under our assumptions, for any $n = 1, 2, \ldots$,

$$\lim_{||s_n - s_{n-1}|| \to 0} p_{i_1,\ldots,i_n} = p_{i_1,\ldots,i_{n-1}} \quad \text{if } i_n = i_{n-1}$$

$$= 0 \quad \text{otherwise,}$$

independently of the particular mean around which we center the process Z, i.e. the weights are almost surely continuous. ■

Appendix II: Full conditionals for the Gibbs sampler

1. Full conditionals for the Z's.

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

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

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

$$\tilde{\mu}^i_{t,l} = \mu_l - h_i(\eta)^T H^{-1}_{(-i)}(\eta)\, Z^{(-i)}_{t,l}, \tag{27}$$

$$\tilde{H}_i(\eta) = 1 - h_i(\eta)^T H^{-1}_{(-i)}(\eta) h_i(\eta), \tag{28}$$

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

Now consider the full conditionals. We start considering the full conditional of $Z_{t,1}(s_i)$, for some $i = 1, \ldots, n$. Let us 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 $Z_{t,1}(s_i)$ is given by

$$[Z_{t,1}(s_i)|Y_t, Z_{t,1}^{(-i)}, Z_{t,l}(s_j), l > 2, \psi] \propto [Z_{t,1}(s_i)|Z_{t,1}(s_j), j \neq i, \theta^*, \mu, \eta, \phi] \times$$
$$\times \sum_{k=1}^{K} \exp\left[-\frac{1}{2\tau^2}\left\{Y_t(s_i) - X_t(s_i)^T\beta - \theta_k^*(s_i)\right\}^2\right] I_{\{Z_{t,1}(s_i)<0,\ldots,Z_{t,k-1}(s_i)<0, Z_{t,k}(s_i)\geq 0\}}, \tag{29}$$

where $Z_{t,2}(s_i), \ldots, Z_{t,K-1}(s_i)$ are all known. For the purpose of exemplification, we suppose that $Z_{t,2}(s_i), \ldots, Z_{t,k-1}(s_i)$ are less than zero and $Z_{t,k}(s_i)$ is greater than zero. Then, if $Z_{t,1}(s_i)$ is sampled to be greater than zero, $\theta_1^*(s_i)$ will be observed, i.e. $\theta_t(s_i) = \theta_1^*(s_i)$. On the other hand, if $Z_t^1(s_i)$ is sampled to be less than zero, then it is evident that $\theta_t(s_i) = \theta_l^*(s_i)$. In fact, by the binary nature of the rule that we have set for the weights we can define the two quantities

$$\omega^- = \exp\left[-\frac{1}{2\tau^2}\{Y_t(s_i) - X_t(s_i)^T\beta - \theta_1^*(s_i)\}^2\right]$$
$$\omega^+ = \exp\left[-\frac{1}{2\tau^2}\{Y_t(s_i) - X_t(s_i)^T\beta - \theta_k^*(s_i)\}^2\right].$$

These are the kernels of two gaussian distributions. Therefore, if we consider the weights

$$\pi_1 = \frac{\omega^-\Phi\left\{\frac{\tilde{\mu}_{t,1}^i}{\sqrt{\tilde{H}_i(\eta)}}\right\}}{\omega^-\Phi\left\{\frac{\tilde{\mu}_{t,1}^i}{\sqrt{\tilde{H}_i(\eta)}}\right\} + \omega^+\Phi\left\{-\frac{\tilde{\mu}_{t,1}^i}{\sqrt{\tilde{H}_i(\eta)}}\right\}}, \text{and } \pi_k = \frac{\omega^+\Phi\left\{-\frac{\tilde{\mu}_{t,1}^i}{\sqrt{\tilde{H}_i(\eta)}}\right\}}{\omega^-\Phi\left\{\frac{\tilde{\mu}_{t,1}^i}{\sqrt{\tilde{H}_i(\eta)}}\right\} + \omega^+\Phi\left\{-\frac{\tilde{\mu}_{t,1}^i}{\sqrt{\tilde{H}_i(\eta)}}\right\}},$$

we can see that (29) is a mixture of two truncated gaussian. Therefore, with probability π_1, we sample $Z_{t,1}(s_i)$ from the truncated normal distribution $N(\tilde{\mu}_{t,1}^i, \tilde{H}_i(\eta))I_{\{Z_{t,1}(s_i)\geq 0\}}$; with π_l sample $Z_{t,1}(s_i)$ from the truncated normal distribution $N(\tilde{\mu}_{t,1}^i, \tilde{H}_i(\eta))I_{\{Z_{t,1}(s_i)<0\}}$. We next proceed repeating the same arguments for $Z_{t,2}(s_i)$.

Let us now consider the full conditional for the general term $Z_{t,l}(s_i)$. If $Z_{t,j}(s_i) \geq 0$, for some $j < l$, then $\theta_t(s_i) = \theta^*_j(s_i)$ and $Z_{t,l}(s_i)$ is sampled directly from the unrestricted distribution $N(\tilde{\mu}^i_{t,l}\,, \tilde{H}_i(\eta))$.

Otherwise if $Z_{t,j}(s_i) < 0,$ for $j < l$, the full conditional is again a binary mixture of truncated normals as we have seen for $Z^1_t(s_i)$ (see equation 29). Say $Z_{t,k}(s_i) \geq 0$ for some $k > l$, again let

$$\omega^- = \exp\left[-\frac{1}{2\tau^2}\{Y_t(s_i) - X_t(s_i)^T\beta - \theta^*_l(s_i)\}^2\right]$$
$$\omega^+ = \exp\left[-\frac{1}{2\tau^2}\{Y_t(s_i) - X_t(s_i)^T\beta - \theta^*_k(s_i)\}^2\right],$$

and

$$\pi_l = \frac{\omega^-\Phi\left\{\frac{\tilde{\mu}^i_{t,l}}{\sqrt{\tilde{H}_i(\eta)}}\right\}}{\omega^-\Phi\left\{\frac{\tilde{\mu}^i_{t,l}}{\sqrt{\tilde{H}_i(\eta)}}\right\} + \omega^+\Phi\left\{-\frac{\tilde{\mu}^i_{t,l}}{\sqrt{\tilde{H}_i(\eta)}}\right\}} \text{ and } \pi_k = \frac{\omega^+\Phi\left\{-\frac{\tilde{\mu}^i_{t,l}}{\sqrt{\tilde{H}_i(\eta)}}\right\}}{\omega^-\Phi\left\{\frac{\tilde{\mu}^i_{t,l}}{\sqrt{\tilde{H}_i(\eta)}}\right\} + \omega^+\Phi\left\{-\frac{\tilde{\mu}^i_{t,l}}{\sqrt{\tilde{H}_i(\eta)}}\right\}}.$$

Therefore, the full conditional for $Z_{t,l}(s_i)$ is again a mixture of two truncated normals. In particular, with probability π_l, we sample $Z_{t,l}(s_i)$ from the truncated normal distribution $N(\tilde{\mu}^i_{t,l}\,, \tilde{H}_i(\eta))I_{\{Z_{t,l}(s_i)\geq 0\}}$; with probability π_k, we sample $Z_{t,l}(s_i)$ from the truncated normal distribution $N(\tilde{\mu}^i_{t,l}\,, \tilde{H}_i(\eta))I_{\{Z_{t,l}(s_i)<0\}}$. Next, proceed repeating similar arguments for $Z_{t,l+1}(s_i)$.

The modification for the full conditionals for the Z's in the spatio-temporal dynamic model is as follows: for $t = 1$, follow the same steps as in the original sampler of the independent-sample case. Suppose $z_{m,l}$, $l = 1, ..K-1$; $m = 1, \ldots, t-1$ is already sampled. Calculate ω_m by $z_{m,l}$ and θ^*_l. For t, let $\tilde{y}_t = y_t - \sum_{m=1}^{t-1}\gamma^{t-m}\omega_m$. With $\tilde{y}_t$ replacing y_t, follow the same steps as in the independent sampler to get $z_{t,l}$, and calculate ω_t.

2. Full conditional for the θ^*'s.

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

It is immediate to see that the full conditional for $\theta_l^* = (\theta_l^*(s_1), \ldots, \theta_l^*(s_n))$ is given by

$$[\theta_l^* | Y_t, Z_t, t = 1, \ldots, T, \beta, \tau^2, \sigma^2, \phi] \propto \exp\left\{-\frac{1}{2\tau^2}\sum_{t=1}^{T}(Y_t - X_t^T\beta - \theta_k^*)^T I(\mathcal{Z}_t)(Y_t - X_t^T\beta - \theta_k^*)\right\} \times$$
$$\times \exp\left\{-\frac{1}{2\sigma^2}\,\theta_k^{*T} R_n^{-1}(\phi)\,\theta_k^*\right\}$$

Then,

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

where $\Lambda = \left(\frac{1}{\tau^2}\sum_{t=1}^{T} I(\mathcal{Z}_{t,l}) + \frac{1}{\sigma^2}R_n^{-1}(\phi)\right)^{-1}$.

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

The full conditionals for θ_l^*'s in the spatio-temporal dynamical model are far more complicated. We are still able to update θ_l^* at all locations, but it has to be conditioned on all the other value θ_j^*'s with $j \neq l$.

Rewrite model (22) with the expanded accumulated spatial random effect

$$Y_t(s) = X_t(s)^T\beta + \sum_{m=1}^{t}\gamma^{t-m}\omega_m(s) + \varepsilon_t(s)$$

Then, if we write ω_m as a function θ_l^*'s and Z_t's, we obtain

$$Y_t(s) = X_t(s)^T\beta + \sum_{m=1}^{t}\gamma^{t-m}\sum_{j=1}^{K-1} I(\mathcal{Z}_{m,j})\theta_j^*(s) + \varepsilon_t(s)$$

Hence, the likelihood × prior can be written as

$$\propto \exp\left[-\frac{1}{2\tau^2}\sum_{t=1}^{T}\left\{Y_t-\sum_{m=1}^{t}\gamma^{t-m}\sum_{j\neq l=1}^{K-1}I(\mathcal{Z}_{m,j})\theta_j^*-\sum_{m=1}^{t}\gamma^{t-m}I(\mathcal{Z}_{m,l})\theta_l^*-X_t^T\beta\right\}^T\right.$$
$$\left.\left\{Y_t-\sum_{m=1}^{t}\gamma^{t-m}\sum_{j\neq l=1}^{K-1}I(\mathcal{Z}_{m,j})\theta_j^*-\sum_{m=1}^{t}\gamma^{t-m}I(\mathcal{Z}_{m,l})\theta_l^*-X_t^T\beta\right\}\right]\times$$
$$\times\exp\left\{-\frac{1}{2\sigma^2}\theta_l^{*T}R^{-1}(\phi)\theta_l^*\right\}.$$

Let us define

$$\tilde{y}_t=y_t-\sum_{m=1}^{t}\left\{\gamma^{t-m}\sum_{j\neq l=1}^{K-1}I(\mathcal{Z}_{m,j})\theta_j^*\right\}-X_t^T\beta,$$

The expression above becomes

$$\exp\left[-\frac{1}{2\tau^2}\sum_{t=1}^{T}\left\{\tilde{y}_t-\sum_{m=1}^{t}\gamma^{t-m}I(\mathcal{Z}_{m,l})\theta_l^*\right\}^T\left\{\tilde{y}_t-\sum_{m=1}^{t}\gamma^{t-m}I(\mathcal{Z}_{m,l})\theta_l^*\right\}\right]\times$$
$$\times\exp\left\{-\frac{1}{2\sigma^2}\theta_l^{*T}R^{-1}(\phi)\theta_l^*\right\},$$

from which we can deduce

$$\left[\theta_l^*|\theta_j^*\left(j\neq l\right),z_t,y_t,\beta_t,\tau^2,\sigma^2,\phi\right]\sim N\left(\frac{1}{\tau^2}\Lambda\sum_{t=1}^{T}\left\{\sum_{m=1}^{t}\gamma^{t-m}I(\mathcal{Z}_{m,l})\right\}\tilde{y}_t,\Lambda\right),$$

with $\Lambda=\left[\frac{1}{\tau^2}\sum_{t=1}^{T}\left\{\sum_{m=1}^{t}\gamma^{t-m}I(\mathcal{Z}_{m,l})\right\}^2+\frac{1}{\sigma^2}R^{-1}(\phi)\right]^{-1}$.

3. Full conditional for β.

Assuming $\beta\sim N_p(\beta_0,\Sigma_0)$, we get

$$\left[\beta|X_t,Y_t,Z_t,\theta_t,\tau^2\right]\sim N(\hat{\beta},\hat{\Sigma}_\beta),$$

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

4. Full conditional for τ^2.

Assume $\tau^2 \sim IGamma(\alpha_\tau, \beta_\tau)$. Then

$$[\tau^2 | X_t, Y_t, \theta_t, \beta] \sim IG(\tilde{\alpha}_\tau, \tilde{\beta}_\tau),$$

where $\tilde{\alpha}_\tau = \alpha_\tau + \frac{nT}{2}$ and $\tilde{\beta}_\tau = \beta_\tau + \frac{1}{2}\sum_{t=1}^{T}(Y_t - \beta X_t - \theta_t)^T (Y_t - \beta X_t - \theta_t)$.

5. Full conditional for σ^2.

Assume $\sigma^2 \sim IG(\alpha_\sigma, \beta_\sigma)$. Then,

$$[\sigma^2 | \theta_l^*, \phi] \sim IGamma(\tilde{\alpha}_\sigma, \tilde{\beta}_\sigma),$$

where $\tilde{\alpha}_\sigma = \alpha_\sigma + \frac{nK}{2}$, and $\tilde{\beta}_\sigma = \beta_\sigma + \frac{1}{2}\sum_{l=1}^{K} \theta_l^{*T} R_n^{-1}(\phi)\theta_l^*$.

6. Full conditional for ϕ.

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

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

7. Full conditional for μ.

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

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

8. Full conditional for η.

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

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

References

Agarwal, D., & Gelfand, A. (2005). Slice Gibbs sampling for simulation based fitting of spatial data models. *To appear in Statistics and Computing.*

Banerjee, S., Gelfand, A. & Sirmans, C. (2003). Directional rates of change under spatial process models. *Journal of the American Statistical Association* **98**, 946–954.

Brown, P., Diggle, P. & Henderson, R. (2003). A non-gaussian spatial process model for opacity of flocculated paper. *Scandinavian Journal of Statistics.* **30**, 355–368.

Cifarelli, D. & Regazzini, E. (1978). Problemi statistici non parametrici in condizioni di scambiabilità parziale. Impiego di medie associative. *Quaderni dell'Istituto di Matematica Finanziaria dell'Universita' di Torino* **Serie III, n. 12**, 1–13.

Damian, D., Sampson, P. & Guttorp, P. (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics*, **12**, 161–178.

De Iorio, M., Müller, P., Rosner, G. & MacEachern, S. (2004). An anova model for dependent random measures. *Journal of the American Statistical Association* **99**, 205–215.

Ecker, M. & Gelfand, A. (2003). Spatial modeling and prediction under stationary non-geometric range anysotropy. *Environment and Ecological Statistics.* **10**, 165–178.

Escobar, M. D. & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**, 209–230.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics*, **2**, 615–629.

Fernandez, C. & Green, P. (2002). Modelling spatially correlated data via mixtures: a Bayesian approach. *J Royal Statistical Soc B* **64**, 805–805.

Gelfand, A. E. & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

Fuentes, M., & Smith, R.L. (2001). A New Class of Nonstationary Spatial Models. *Tech. rep.*, Department of Statistics, North Carolina State University.

Gelfand, A., Kottas, A. & MacEachern, S. (2005). Bayesian nonparametric spatial modeling with Dirichlet processes mixing. *To appear in the Journal of the American Statistical Association.*

Griffin, J. & Steel, M. (2004). Order-Based Dependent Dirichlet Processes. Warwick Statistics Research Report **430**.

Higdon, D., Swall, J., & Kern, J. (1999). Non-Stationary Spatial Modeling. *Bayesian Statistics 6*, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Oxford University Press.

Hjort, N. (2000). Bayesian analysis for a generalized Dirichlet process prior. Tech. rep., University of Oslo.

Ishwaran, H. & James, L. (2001). Gibbs sampling methods for stik breaking priors. *Journal of the American Statistical Association* **96**, 161–173.

Ishwaran, H. & Zarepour, M. (2002a). Dirichlet prior sieves in finite normal mixtures. *Statistica Sinica* **12**, 941–963.

Ishwaran, H. & Zarepour, M. (2002b). Exact and approximate sum-representations for the Dirichlet process. *Canad. J. Statist.* **30**, 269–283.

Kent, J. (1989). Continuity properties of random fields. *Ann. Probab.* **17**, 1432–1440.

MacEachern, S. (2000). Dependent Dirichlet processes. Tech. rep., Department of Statistics, The Ohio State University.

MacEachern, S. N. & Müller, P. (1998). Estimating mixture of Dirichlet process models,. *Journal of Computational and Graphical Statistics* **7**, 223–238.

Palacios, M. & Steel, M. (2004). Non-gaussian Bayesian geostatistical modeling. Warwick Statistics Research report 426, University of Warwick.

Sampson, P. & Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association* **87**, 108–119.

Schmidt, A. M. & O'Hagan, A. (2003). Bayesian inference for nonstationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society, Series B*, **65**, 745–758.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

	$\Theta_l^1(s_j)$	$\Theta_l^0(s_j)$	
$\Theta_l^1(s_i)$	$q_{l,1,1}(s_i, s_j)$	$q_{l,1,0}(s_i, s_j)$	$q_l(s_i)$
$\Theta_l^0(s_i)$	$q_{l,0,1}(s_i, s_j)$	$q_{l,0,0}(s_i, s_j)$	$1 - q_l(s_i)$
	$q_l(s_j)$	$1 - q_l(s_j)$	1

Table 1: Relevant probabilities in the multivariate stick-breaking construction in the special case of $n = 2$ locations, for $l = 1, 2, \ldots$.

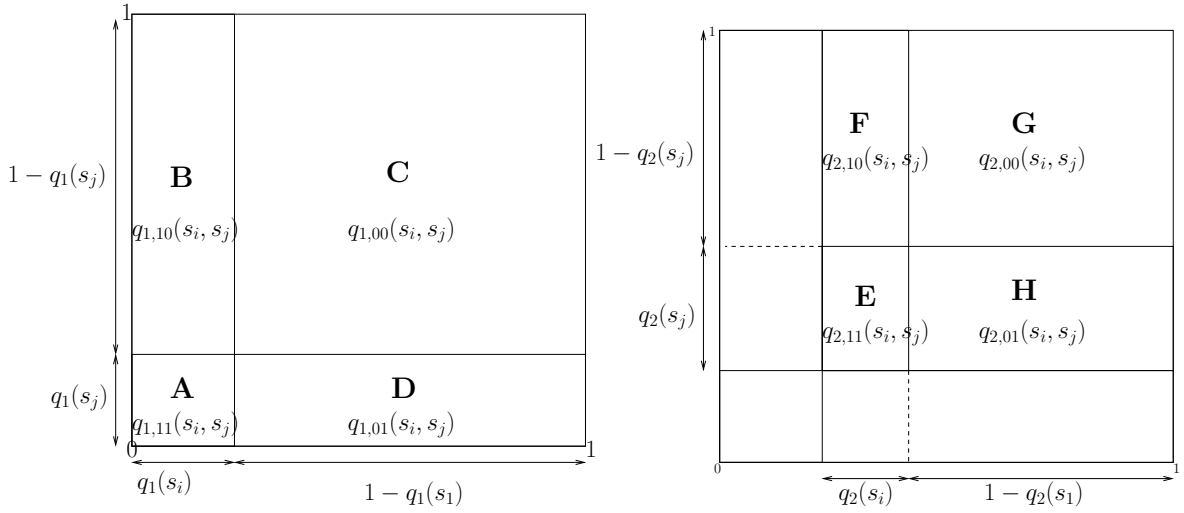

Figure 1: An exemplification of the multivariate stick-breaking procedure for the special case of $n = 2$ locations

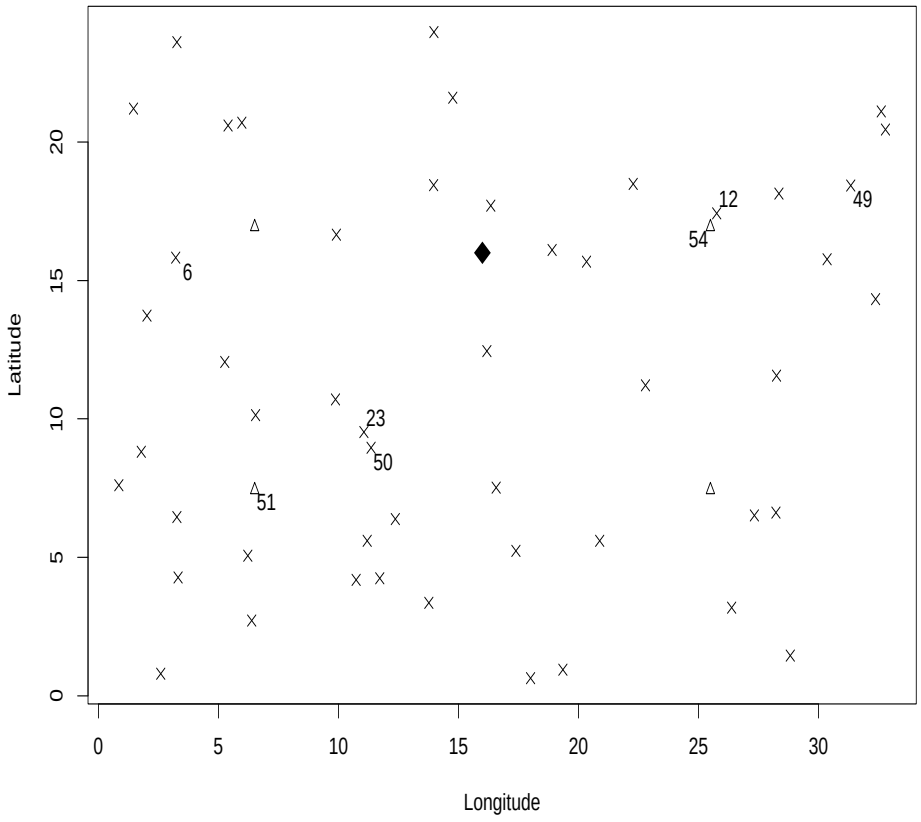

Figure 2: The Design Locations, New Locations and the Point Source in a Region for the Simulation Example in Section 6 and 7.

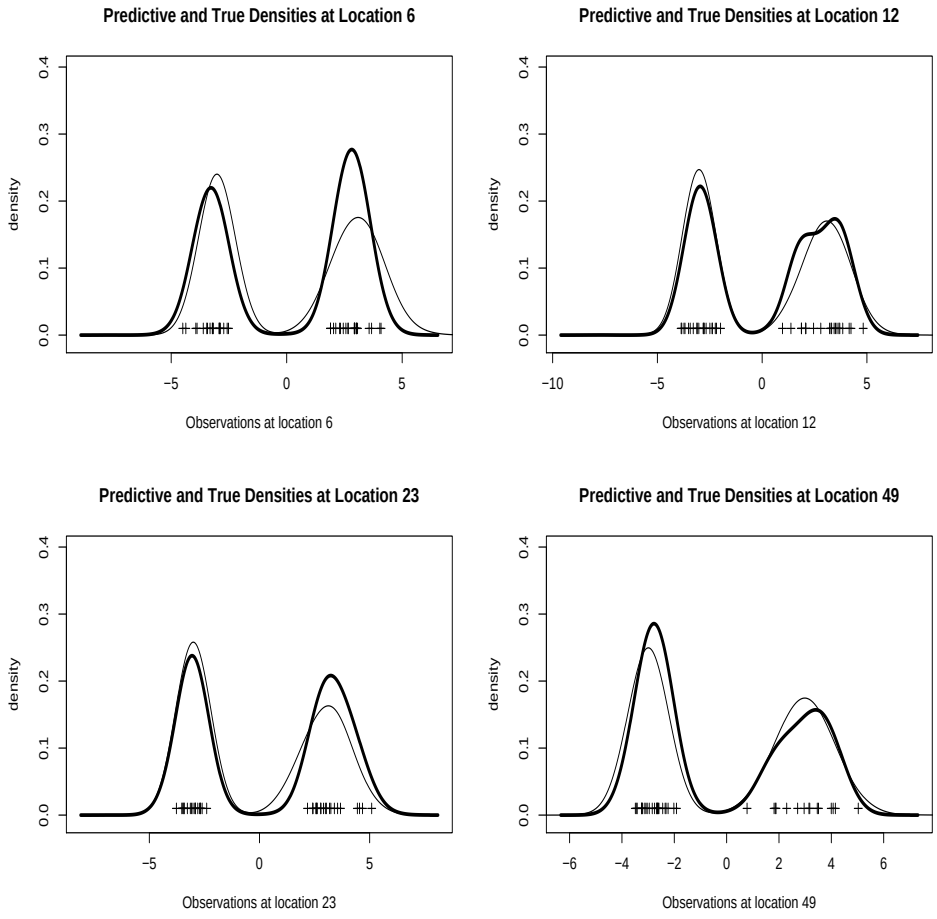

Figure 3: For the Simulation Example in Section 6, Posterior Predictive Densities, True Densities, and Observed Samples for Four Randomly Selected Locations(See text for details)

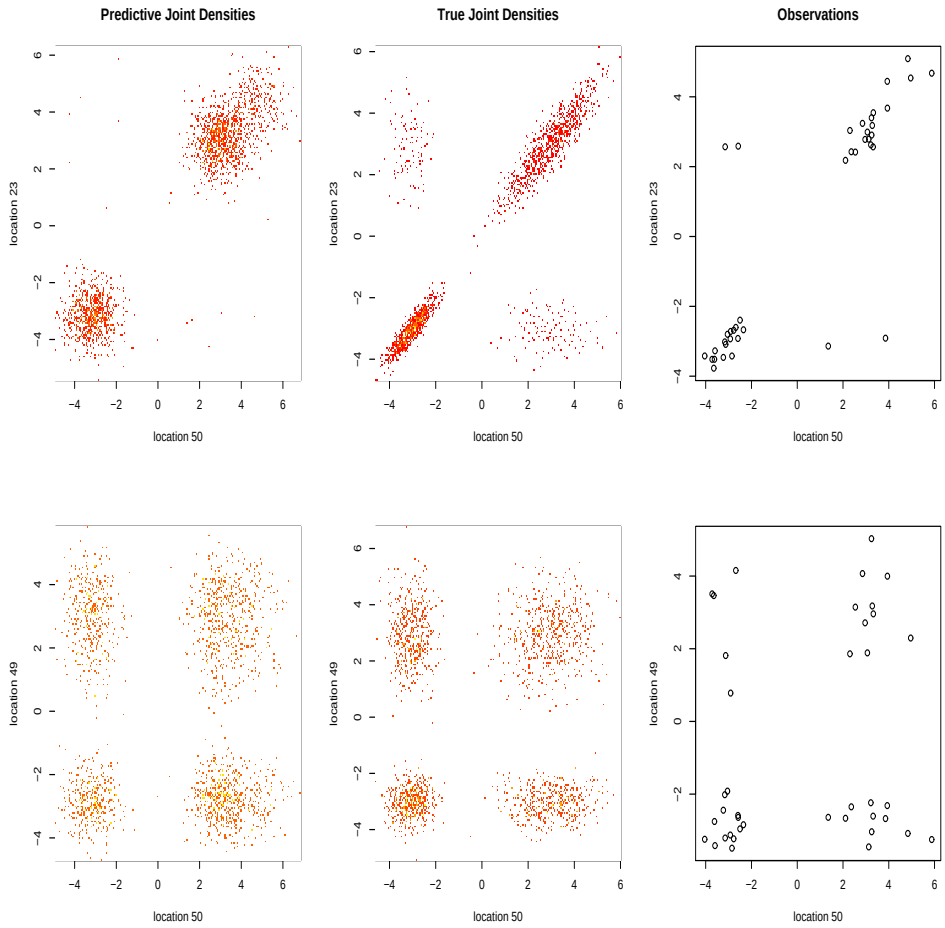

Figure 4: Predictive and True Bivariate Distribution for the Simulated Data Example of Section 6(See text for details)

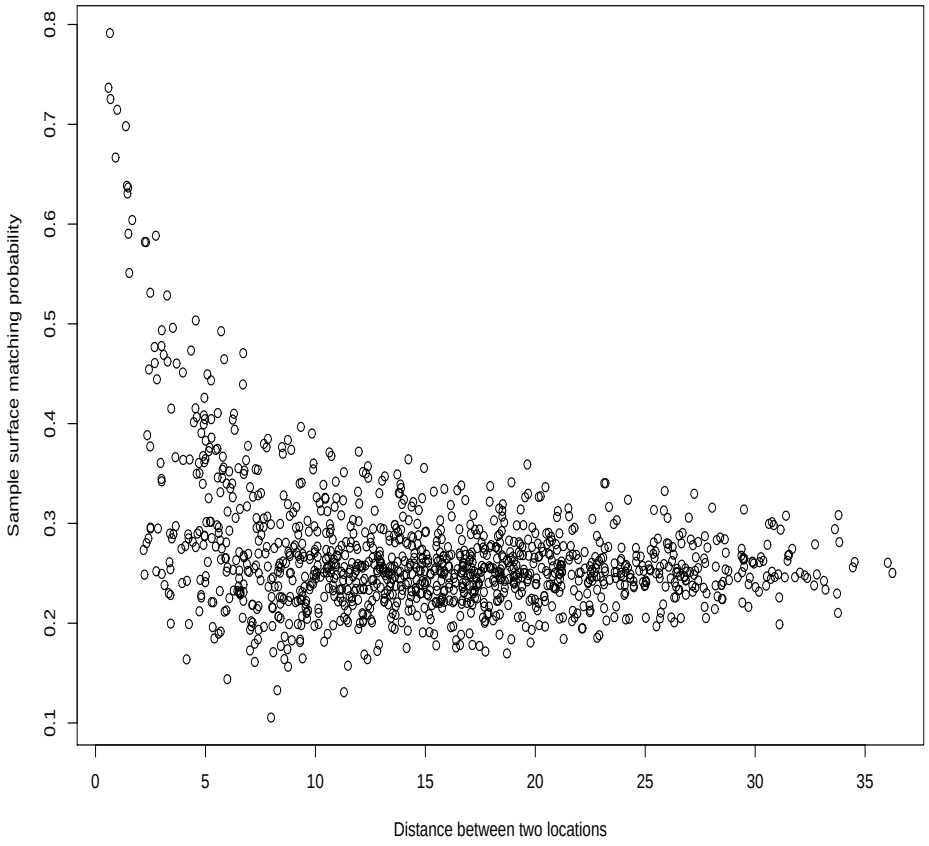

Figure 5: Decay in Probability of Common Surface Selection as a Function of Distance(See Section 6 for details)

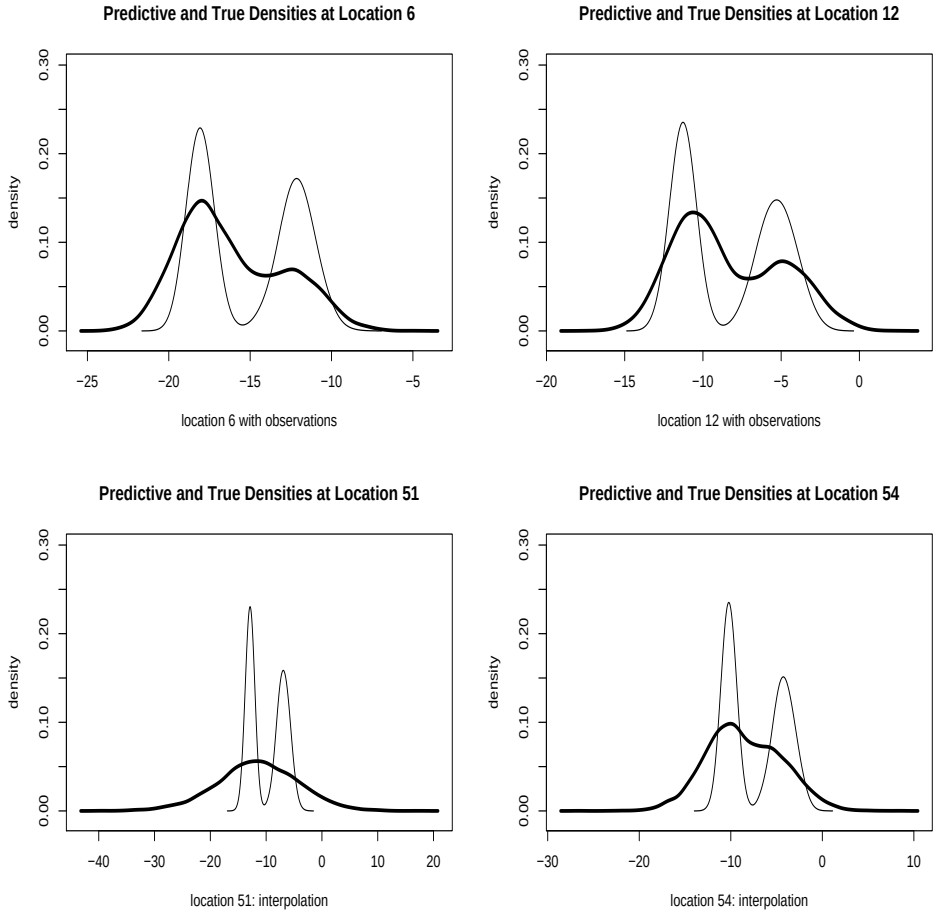

Figure 6: For the Simulation Example in Section 7, Posterior Predictive Densities, True Densities for Two Locations with Observations and Two New Locations at T+1(See text for details)

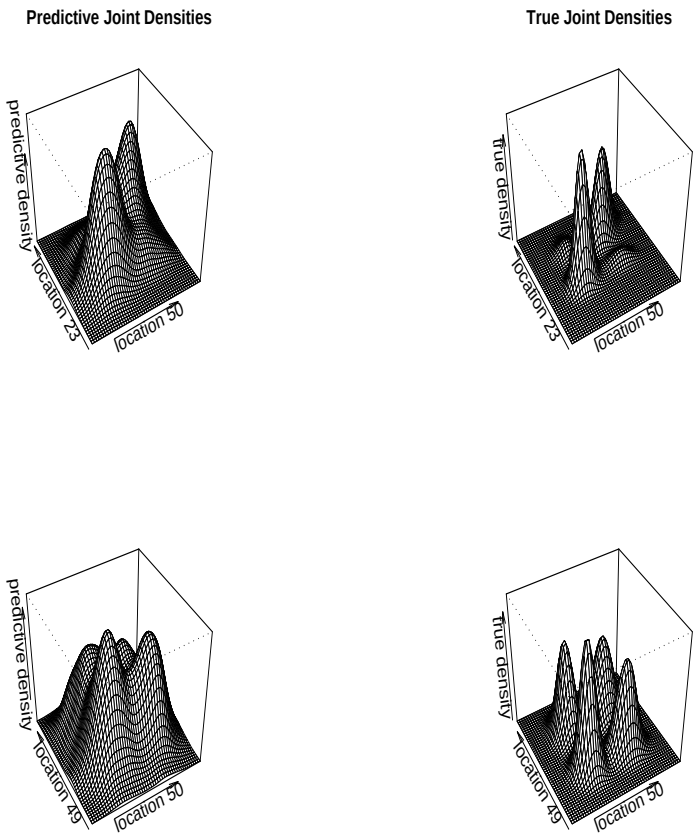

Figure 7: Predictive and True Bivariate Predictive Distribution at T+1 for the Simulated Data Example of Section 7(See text for details)

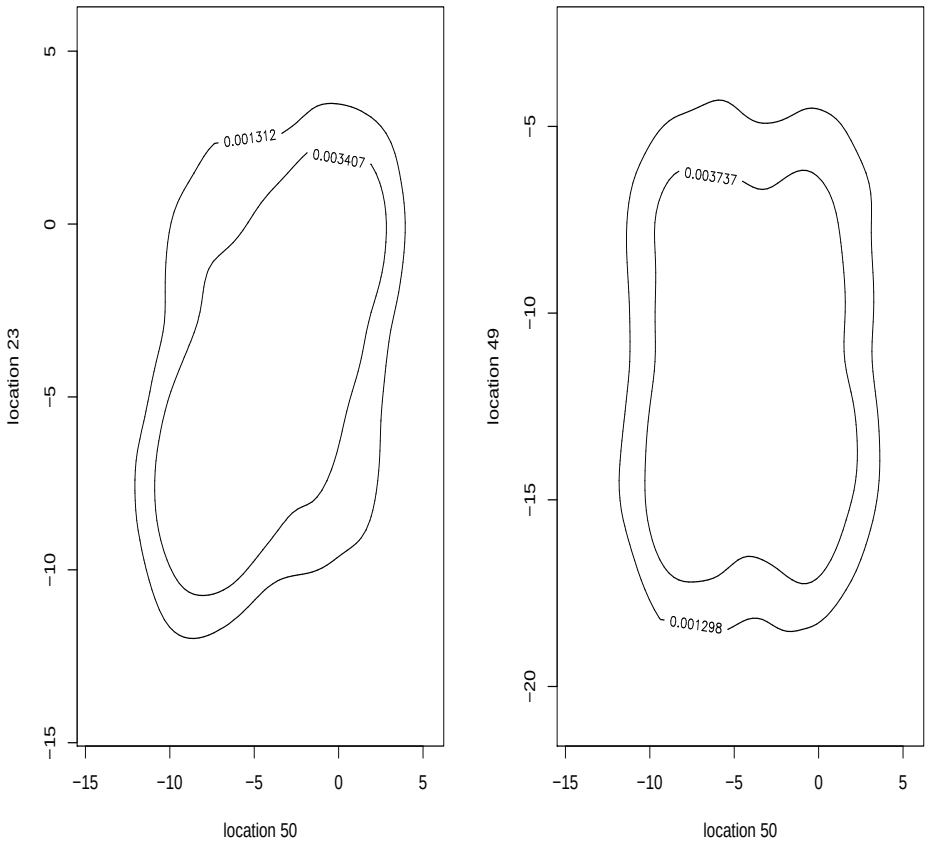

Figure 8: 95% and 80% Simultaneous Bivariate Posterior Density Confidence Sets(See section 7 for details)