Bayesian Nonparametric Modelling for Spatial Data Using Dirichlet Processes

ALAN E. GELFAND
Duke University, USA
`alan@stat.duke.edu, USA`

MICHELE GUINDANI
MD Anderson Cancer Center, USA
`mguindani@mdanderson.org`

SONIA PETRONE
Università Bocconi, Milano, Italy
`sonia.petrone@unibocconi.it`

SUMMARY

Modelling for spatially referenced data is receiving increased attention in the statistics and the more general scientific literature with applications in, e.g., environmental, ecological and health sciences. Bayesian nonparametric modelling for unknown population distributions, i.e., placing distributions on a space of distributions is also enjoying a resurgence of interest thanks to their amenability to MCMC model fitting. Indeed, both areas benefit from the wide availability of high speed computation. Until very recently, there was no literature attempting to merge them. The contribution of this paper is to provide an overview of this recent effort including some new advances. The nonparametric specifications that underlie this work are generalizations of Dirichlet process mixture models. We attempt to interrelate these various choices either as generalizations or suitable limits. We also offer data analytic comparison among these specifications as well as with customary Gaussian process alternatives.

Keywords and Phrases: GAUSSIAN PROCESS, GENERALIZED STICK-BREAKING PROCESS, LOCAL SURFACE SELECTION, NONSTATIONARY PROCESS, SPATIAL RANDOM EFFECTS

1. INTRODUCTION

Point-referenced spatial data is collected in a wide range of contexts, with applications, among others, in environmental, ecological, and health sciences. Modelling for such data introduces a spatial process specification either for the data directly or for a set of spatial random effects associated with the mean structure for the data,

perhaps on a transformed scale. In virtually all of this work the spatial process specification is parametric. In fact, it is almost always a Gaussian process (GP), which is often assumed to be stationary.

Within a Bayesian framework the resulting model specification can be viewed as hierarchical (see, e.g., Banerjee, Carlin and Gelfand, 2004). The parameters of the spatial process are unknown, and so they are assigned prior distributions resulting in a random GP. Fitting, using Markov chain Monte Carlo (MCMC) methods, is by now fairly straightforward. See, e.g., Agarwal and Gelfand (2005) and further references therein. Bayesian inference is attractive in analyzing point-referenced spatial data particularly with regard to assessing uncertainty. The alternative likelihood analysis will impose arguably inappropriate assumptions to achieve asymptotic inference, and hence to obtain asymptotic variability (Stein, 1999).

Spatially varying kernel convolution ideas as in Higdon, Swall and Kern (1999) as well as a local stationarity approach as in Fuentes and Smith (2001) remove the stationarity assumption but are still within the setting of GPs. Zidek and colleagues (see, e.g., Le and Zidek (1992)) introduce nonstationarity through a Wishart model for a random covariance matrix with mean based upon a stationary covariance function. However, this construction sacrifices the notion of a spatial process and, given the covariance matrix, the spatial effects are still Gaussian.

There is a rich literature on nonparametric modeling for an expected spatial surface, much of it drawing from the nonparametric regression literature. See, e.g., Stein (1999) and further references therein. Our interest is in nonparametric modeling for the stochastic mechanism producing the spatial dependence structure. In this regard, the literature is very limited. The *nonparametric* variogram fitting approaches, e.g., Shapiro and Botha (1991) and Barry and Ver Hoef (1996) do not fully specify the process; they are nonparametric only in the second moment structure. Arguably, the most significant nonparametric specification of the covariance function is the "deformation" approach of Sampson and Guttorp (1992). The observed locations in the actual (geographic) space are viewed as a nonlinear transformation of locations in a conceptual (deformed) space where the process is assumed stationary and, in fact, isotropic. This approach has been pursued in a Bayesian context by both Damian et al. (2001) and Schmidt and O'Hagan (2003) but again confined to a GP for the likelihood.

Regarding spatiotemporal data, while space is usually taken to be continuous, a primary distinction is whether time is taken to be continuous or discrete. In particular, with continuous time, the space-time dependence structure is provided by a covariance function which may be separable (Mardia and Goodall, 1993) or nonseparable (Stein, 2005 and references therein). With discrete time, the dependence structure is usually conceived as temporal evolution of a spatial process described in the form of a dynamic model. (See, e.g., Gelfand, Banerjee and Gamerman, 2005a.) Such evolution can yield a time series at each spatial location, as with weather or pollution measurements at monitoring stations. It can also occur with cross-sectional data, as with real estate transactions over a region across time. All of this literature is fully parametric.

The goal of this paper is to review and extend recent nonparametric modeling approaches for spatial and spatiotemporal data through the use of the Dirichlet process (DP) (Ferguson, 1973, 1974). DPs have been used to provide random univariate and multivariate distributions, requiring only a baseline or centering distribution and a precision parameter. Using the DP, we describe a probability law for the stochastic process, $\{Y(s) : s \in D\}$. We refer to such a model as a spatial Dirichlet process

(SDP) (Gelfand et al., 2005b). As with GPs, we provide this specification through finite dimensional distributions, that is, a random distribution for $(Y(s_1), ..., Y(s_n))$ where n and the set of s_i are arbitrary. The resulting process is nonstationary and the resulting joint distributions are not normal. With regard to the set of random univariate distributions, $\{F(Y(s)) : s \in D\}$ we can achieve a dependent Dirichlet process (DDP)(MacEachern (2000)). More precisely, the $F(Y(s))$ are dependent and, as $s \to s_0$, the realized $F(Y(s))$ converge to the realized $F(Y(s_0))$. This enables us to pool information from nearby spatial locations to better estimate say $F(Y(s_0))$ yielding fully model-based nonparametric spatial prediction analogous to parametric kriging using predictive distributions.

For a fully nonparametric modeling approach, replication of some form will be required; with only a single observation from a multivariate distribution, a nonparametric model is not viable.[1] Replicates from a spatial process typically arise over time. Depending upon the data collection, an assumption of independence for temporal replicates may be inappropriate; fortunately, we can directly embed our methodology within a dynamic model, retaining temporal dependence.

Our contribution is best described as random effects modelling. Just as DP priors are often used to extend random effects specifications beyond usual Gaussian random effects, here we broaden Gaussian spatial process specifications. In this regard, we recall the finite-dimensional Dirichlet process priors (Ishwaran and Zarepour, 2002b references therein), which we denote as DP_K priors, and which converge, under suitable conditions, to Dirichlet process priors. Spatial analogues are easily formulated (we call them SDP_K priors) with similar limiting behavior and thus provide another extension of GP's. Recent work by Duan, Guindani and Gelfand (2005) extends the SDP to a generalized SDP (GSDP) by introducing "local surface selection". In essence, for a usual multivariate DP, vectors are drawn from a probability weighted countable collection of vectors. A generalization would allow for individual components of the drawn vector to come from different choices of the countable collection of vectors. Here, we can also formulate a finite-dimensional version which we refer to as a $GSDP_K$. The cartoon presented as Figure 1 attempts to capture the modelling world that we are focusing on. G denotes "Gaussian" and GP a Gaussian process. The arrows suggest either that the model being pointed to is an extension or generalization of the model it emanates from.
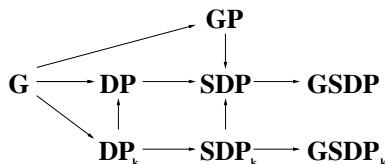

Figure 1: *The modelling world for this paper.*

Possible advantages offered by our approach are the following. We can draw upon the well-developed theory for DP mixing to facilitate interpretation of our analysis. We can implement the required simulation-based model fitting for posterior

---

[1]The deformation setting (Sampson and Guttorp (1992)) requires replications to obtain the sample covariance estimate, a *nonparametric* estimate of the process covariance matrix at the observed sites.

inference and spatial prediction easily, availing ourselves of established strategies for DP mixture models. (See, e.g., Neal (2000) and, more recently, Jain and Neal (2004).) We can infer about the random distribution that is operating at any given location or set of locations in the region. An alternative line of nonparametric development using Lévy processes is summarized in Clyde and Wolpert (2006).

The models presented here are in terms of random process realizations over a geographic space. However, we can also envisage random curves or surfaces over a covariate space, i.e., a space of covariate levels. Covariate space might be dose levels with interest in response to treatment, ocean depths with interest in temperature or salinity, or time yielding, say, hormone levels across menstrual cycles. Hence, our models are directly applicable to so-called functional data analysis where we consider replicates of random curves or surfaces as process realizations. Lastly, we can imagine spatial functional data analysis where there is a conceptual random curve at each spatial location. The curves are only observed (with error) at a finite set of locations and a finite number of points at each location. For instance there is an unknown temperature vs depth curve at every location in an ocean and we may gather data at a set of locations and a set of depths at these locations. We would expect that curves would be more similar for locations close to each other, less so when the locations are distant from each other.

The format for the paper is as follows. In Section 2 we review the DP and DP_K specifications. In Section 3 we bring in the spatial aspect with Section 4 providing two examples to compare the GP, SDP, and SDP_K. Section 5 develops the generalized local surface selection versions, the GSDP and $GSDP_K$ leaving Section 6 to provide a comparison among them and the SDP. Section 7 concludes with a summary and some issues that require further examination.

2. A REVIEW

Hierarchical models are frequently built through random effects which typically are specified to be i.i.d. from a mean 0 Gaussian "population" model. The random effects need not be univariate, e.g., random slopes and intercepts led to random lines and more generally to random curves for instance in growth curve modelling. See, e.g., Diggle, Liang and Zeger (1994) for a parametric treatment with the recent work of Scaccia and Green (2003) offering a nonparametric view. With the computational revolution of the '90's (Gibbs sampling and MCMC), handling hierarchical models with Gaussian random effects became routine (arguably the most widely used specification in the BUGS software. Of course, concern regarding the Gaussian assumption led to more flexible specifications through mixture modelling including t distributions, Dirichlet processes, and finite mixtures within which we distinguish the finite dimensional Dirichlet process models (what we refer to as DP_K models). We now review this modelling. In the process, we clarify the arrows in Figure 1 involving G, DP and DP_K.

There has been a growing literature on the use of DP priors primarily because they are easy to specify, attractive to interpret, and ideally suited for model fitting within an MCMC framework. In particular, we recall the stickbreaking representation of the DP (Sethuraman (1994)). Let $\theta_1^*, \theta_2^*, \ldots$ be i.i.d. random elements independently and identically distributed according to the law G_0. G_0 can be a distribution over a general probability space, allowing the θ^*'s to be random objects such as vectors, a stochastic process of random variables (Section 3), or a distribution (Rodriguez et al, 2006). Let $q_1, q_2, \ldots$ be random variables independent of the θ^*'s and i.i.d. among themselves with common distribution Beta$(1, \alpha)$. If we

set $p_1 = q_1$, $p_2 = q_2 (1 - q_1), \ldots$, $p_k = q_k \prod_{j=1}^{k-1}(1 - q_j), \ldots$ the random probability measure defined by

$$G(\cdot) = \sum_{k=1}^{\infty} p_k \, \delta_{\theta_k^*}(\cdot), \tag{1}$$

is distributed according to a DP.

A wide range of random probability measures can be defined by means of a stick-breaking construction, where the q's are drawn independently from an arbitrary distribution Q on $[0, 1]$ (see Hjort (2000)). For example, the beta-two parameter process (Ishwaran and Zarepour, 2000) is defined by choosing $q_k \sim Beta(a, b)$. If $q_k \sim Beta(1 - a, b + ka)$, $k = 1, 2, \ldots$, for some $a \in [0, 1)$ and $b \in (-a, \infty)$ we obtain the two-parameter Poisson-Dirichlet Process (Pitman and Yor, 97). Ishwaran and James (2001) discuss the general case, $q_k \sim Beta(a_k, b_k)$. More generally, we can consider the discrete random probability measure

$$G_K(\cdot) = \sum_{k=1}^{K} p_k \, \delta_{\theta_k^*}(\cdot), \tag{2}$$

where K is an integer random variable (allowed to be infinite), θ_k^* are i.i.d. from some base distribution G_0 (not necessarily nonatomic) and, independently, the weights p_k can have any distribution on the simplex $\{\mathbf{p} : \sum_{i=1}^{K} p_i = 1, p_k \geq 0\}$ (see Ongaro and Cattaneo (2004)). In these constructions the stick-breaking is *one-dimensional*; the probability p_k is for the selection of the entire random quantity, θ_k^*.

If K is finite, and $(p_1, \ldots, p_K) \sim Dir(\alpha_{1,K}, \ldots, \alpha_{K,K})$, we obtain the class of *finite dimensional Dirichlet priors*, discussed by Ishwaran and Zarepour (2002a). We refer to this class of priors as DP_K priors and denote them by $G_K \sim DP_K(\alpha, G_0)$. In fact, such priors have often been considered as a general approximation of models based on 1, e.g., it has been a common choice to set $\alpha_{i,K} = \lambda_K$ (see i,KKRichardson and Green (1997)). However, an important result from Ishwaran and Zarepour (2002a) clarifies when finite sums such as 2 converge to a DP. In particular, let $G_K \sim DP_K(\alpha, G_0)$ and $E_{G_k}(h(x)) = \int h(x) G_K(dx)$ denote a random functional of G_K, where h is a non-negative continuous function with compact support. Then, if $\alpha_{k,K} = \lambda_K$, where $K \lambda_K \to \infty$, then $E_{G_k}(h(x)) \xrightarrow{p} E_{G_0}(h(x))$, that is the choice of a uniform Dirichlet prior leads to a limiting parametric model. Instead, if $\sum_{k=1}^{K} \alpha_{k,K} \to \alpha > 0$ and $\max \alpha_{1,K}, \ldots, \alpha_{K,K} \to 0$ as $K \to \infty$, the limit distribution of G_K is really nonparametric, since $E_{G_k}(h(x)) \xrightarrow{\mathcal{D}} (h(x))$, where G is the usual Dirichlet process with finite measure αG_0. The result follows directly from 0 Kingman (1975) and the properties of the Poisson-Dirichlet distribution. In particular, the former includes the common case of $\alpha_{k,K} = \alpha/K$ for some $\alpha > 0$, whose weak convergence to the DP had already been proved by Muliere and Secchi (1995).

In some situations, there could be an interest in indexing the random probability distribution of the observables according to the values of underlying covariates. Following MacEachern (see MacEachern (2000)), we refer to these models as dependent random probability measures (see also De Lorio et al. (2004), Griffin and Steel (2004)) noting that there is a previous literature in this vein summarized in Petrone and Raftery (1997). In this setting, the weights and point masses of the random probability measure 2 are indexed by a vector of covariates $\mathbf{z}$, that is $G_K(\mathbf{z}, \cdot) = \sum_{k=1}^{K} p_k(\mathbf{z}) \, \delta_{\theta_k^*(\mathbf{z})}(\cdot)$.

3. RANDOM DISTRIBUTIONS FOR SPATIAL LOCATIONS: THE SDP AND SDP_K MODELS

We assume data from a random field $\{Y(\mathbf{s}), \mathbf{s} \in D\}$, $D \in \mathbb{R}^d$, such that

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \theta(\mathbf{s}) + \varepsilon(\mathbf{s}), \tag{3}$$

where the mean structure $\mu(\mathbf{s})$ is, say $\mathbf{x}(\mathbf{s})^T\boldsymbol{\beta}$ with $\mathbf{x}(\mathbf{s})$ a p-dimensional vector comprised of covariates at location $\mathbf{s}$. $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients. The residual term is partitioned into two pieces. The first, $\theta(\mathbf{s})$, accounts for spatial variability that is not captured by the mean. The second term, $\varepsilon(\mathbf{s})$, is intended to capture variability of a non spatial nature; it can be interpreted either as a pure error term or as a term incorporating measurement errors or microscale variability. When $\theta(\mathbf{s})$ is a mean 0 Gaussian process and $\varepsilon(\mathbf{s})$ is Gaussian white noise, the process $Y(\mathbf{s})$ is Gaussian and stationary. A mixture of Gaussian processes would permit greater flexibility (see, e.g., Brown et al. (2003)) as would the Gaussian/logGaussian process described in Palacios and Steel (2006).

Gelfand, Kottas and MacEachern (2005b) generalize in a different way, replacing the Gaussian specification with a DP. They define a spatial Dirichlet process (SDP) by considering the base measure G_0 itself to be a mean zero stationary Gaussian process defined over D. Hence, recalling 1, the θ_k^*'s are realizations of a random field, i.e. surfaces over D, $\theta_k^* = \{\theta_k^*(\mathbf{s}), \mathbf{s} \in D\}$, $k = 1, \ldots$. However, the random weights do not depend on the locations and the spatial dependence is introduced only through the underlying base measure.

More generally, we could extend the random probability measures in 2 so that K is finite, $(p_1, \ldots, p_K) \sim Dir(\alpha_1, \ldots, \alpha_K)$ and θ_k^* are surfaces, which are realizations of a specified random field G_0, thus defining a finite dimensional spatial Dirichlet prior (SDP_K). The above results from Ishwaran and Zarepour (2002a) show that, for certain choices of the α parameters, the SDP_K process provides a finite-approximation to the SDP process. Let G denote a random distribution obtained according to either 1 or 2. In both cases, we need only consider finite dimensional distributions of G at locations, say $(\mathbf{s}_1, \ldots, \mathbf{s}_n)$, denoted by $G^{(n)}$. Then, $G^{(n)} \sim DP(\alpha G_0^{(n)})$ or $G^{(n)} \sim DP_K(\alpha G_0^{(n)})$ where $G_0^{(n)}$ is the associated n-variate finite dimensional (e.g., multivariate normal) distribution of the process G_0.

Then, if $\{\theta(\mathbf{s}), \mathbf{s} \in D\}$ is a random field, such that $\theta(\cdot)|G \sim G$, where $G \sim SDP(\alpha, G_0)$ or $G \sim SDP_K(\alpha, G_0)$, we obtain

$$
\begin{aligned}
E(\theta(\mathbf{s})|G) &= \sum_{k=1}^{K} p_k\, \theta_k^*(\mathbf{s}), \\
Var(\theta(\mathbf{s})|G) &= \sum_{k=1}^{K} p_k\, (\theta_k^*(\mathbf{s}))^2 - \left\{\sum_{k=1}^{K} p_k\, \theta_k^*(\mathbf{s})\right\}^2,
\end{aligned}
$$

and for any two locations $\mathbf{s}_i, \mathbf{s}_j \in D$,

$$Cov(\theta(\mathbf{s}_i), \theta(\mathbf{s}_j)|G) = \sum_{k=1}^{K} p_k\, \theta_k^*(\mathbf{s}_i)\, \theta_k^*(\mathbf{s}_j) - \left\{\sum_{k=1}^{K} p_k\, \theta_k^*(\mathbf{s}_i)\right\} \left\{\sum_{k=1}^{K} p_k \theta_k^*(\mathbf{s}_j)\right\}, \tag{4}$$

where $K = \infty$ if G is a SDP. Hence, for any given G, the process $\theta(\cdot)$ has heterogenous variance and is nonstationary. However, G is "centered" around G_0 which is typically stationary.
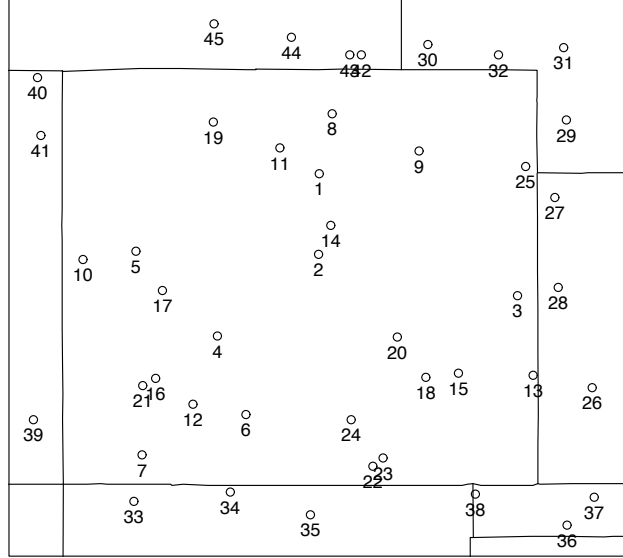
Figure 2: *A region comprising the state of Colorado where we have collected data used in Section 4.*

Finally, consider data $\mathbf{Y}_t = (Y_t(\mathbf{s}_1), \ldots, Y_t(\mathbf{s}_n))^T$, for replications $t = 1, \ldots, T$, and associate with $\mathbf{Y}_t$ a vector $\boldsymbol{\theta}_t = (\theta_t(\mathbf{s}_1), \ldots, \theta_t(\mathbf{s}_n))^T$, such that $\boldsymbol{\theta}_t | G^{(n)} \overset{i.i.d.}{\sim} G^{(n)}$, $t = 1, \ldots, T$. Then, the following semiparametric hierarchical model arises:

$$
\begin{array}{ll}
\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_T | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_T, \boldsymbol{\beta}, \tau^2 & \sim \prod_{t=1}^{T} N(\mathbf{X}_t^T \boldsymbol{\beta} + \boldsymbol{\theta}_t, \tau^2 I_n) \\
\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_T | G^{(n)} & \overset{i.i.d.}{\sim} G^{(n)}
\end{array}
\tag{5}
$$

where X_t is a suitable design matrix for replicate t. The model is completed by choosing an appropriate specification for the unknown G as above together with a suitable choice of prior for the other parameters of the model.

4. A DATA EXAMPLE.

Here, we compare the behavior of the GP, the SDP and the SDP_K using temperature and precipitation data collected at 45 weather stations monthly over 40 years (1958-1997) in a region encompassing the state of Colorado (in the US) (see Figure 2). We assume independence across years by restricting ourselves to average monthly temperature for the single month of July (though embedding (5) within a dynamic model could also be straightforwardly done)

We assume a spatial random effect model as in 3, where $Y_t(s)$ is the average July temperature in year t and $\mu_t(\mathbf{s}) = \beta_0 + \beta_1^T X_t(s)$, with $X_t(s)$ indicating associated
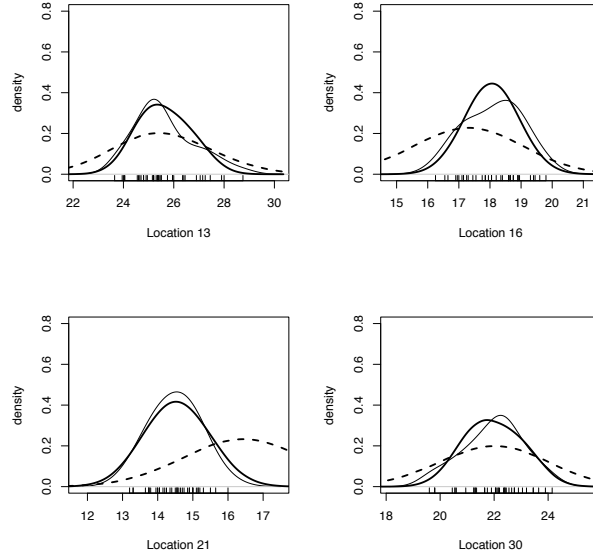
Figure 3:   *Posterior predictive densities $Y_{new}(s)|data$ for the SDP(thick line −) and GP (thick dotted line = −). The lighter dotted line (−) is the estimated density from the 40 replicates in the Colorado dataset (real data).*
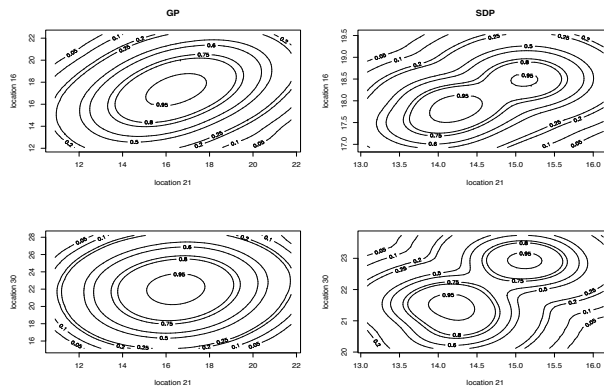

Figure 4:   *Bivariate contour plots of the posterior distributions of the mean in the SDP and GP models for locations 16 and 21 as well as 21 and 30 and the 40 replicates in the Colorado dataset (real data). See Section 4.*

precipitation. The correlation function for the spatial effect is assumed exponential, $\rho(\mathbf{s}-\mathbf{s}') = \sigma^2 \exp\{-\phi \, ||\mathbf{s}-\mathbf{s}'||\}$. In all other respects the prior distributions and the hyperparameters in 5 are chosen as in Gelfand *et al.* (2005b), except that here ϕ is given a Gamma prior and is sampled with a Metropolis-Hastings step. Also, in order to facilitate comparison with the SDP_K we fix $\alpha = 10$ (trials with random α in the SDP yielded similar results. For the SDP, this choice corresponds to a number of clusters *apriori* equal to 16 (see Antoniak (1974)), a weak restriction on the number of clusters. Also, for $K = 10$ and $\alpha_{i,K} = \alpha/K$, in the SDP_K the distribution of the p_k is uniform Dirichlet.

With regard to the choice of the GP model for comparison, following (3), we are concerned with the residual $\theta_t(s)+\epsilon_t(s), t = 1, 2, ..., T$. The two extreme cases correspond to $\alpha \to \infty$ (where the $\theta_t(s)$ are all distinct) and $\alpha \to 0$ (where $\theta_t(s) = \theta(s)$). From this perspective, the SDP (and the SDP_K) falls in between. Structurally, the differences are clear. $\theta_t(s) + \epsilon_t(s)$ has dependence within a replication but independence across replications while $\theta(s) + \epsilon_t(s)$ has dependence both within and across replications (see Sethuraman and Tiwari (1982)). With regard to comparison, according to context, a case can be made for each of the extremes. For instance, suppose the spatial locations are associated with apartment buildings, and the replications are the selling prices of condominium units within the building. Each unit should receive a common building level spatial effect (see Gelfand *et al.* (2006)). However, in, e.g., Duan and Gelfand , 2006 the $\alpha = \infty$ case was chosen as the "fair" comparison with the SDP, arguing that, in both cases, "dependence within a replication with independence across replications" is retained. (Of course, for the former, the mixing distribution is "known" while for the latter it is not.)

The simple GP ($\alpha \to 0$) is unable to capture the variability of multimodal data. In particular, using $d^2 = \sum_{s,t} [Y_t(s) - E(Y_{new}(s)|\text{data})]^2$, a predictive sum of squared deviations, we find d^2 is roughly 950 for the SDP, SDP_K and 860 for GP ($\alpha \to \infty$), while it is 1,600 for the GP ($\alpha \to 0$). Also, as long as the number of mixing components is small relative to K, little difference is seen between the SDP_K and SDP models, in accordance with the theoretical results in Ishwaran and Zarepour (2002a). so, we restrict comparison to the SDP and the GP corresponding to $\alpha \to \infty$.

In Figure 3, we show the marginal predictive distributions at sites $13, 16, 21$, and 30, noting that the SDP captures the estimated density better than the GP at these locations. Next, sites 16 and 21 are close to each other while sites 21 and 30 are far apart. In both cases, the bivariate predictive distribution under the SDP suggests the presence of two clusters, while the GP suggests unimodality (see Figure 4).

5. GENERALIZING THE SDP AND SDP_K.

Here we motivate and formulate the generalized spatial Dirichlet process models.

5.1. *Motivating the generalized DP*

The DP in (2) selects a θ_k^* with probability p_k. Though clustering can be encouraged say with small α, there is always the chance to draw a new, distinct θ^*. For example, in the problem of species sampling it is attractive to have a mechanism that enables new species types (Pitman, 1996) and a similar perspective would apply in developing say, image classifications.

Though this is possible with usual finite mixture models by allowing the number of mixture components to be random, it is more elegantly handled through the DP.

Still, the DP can be viewed as inefficient in this regard. For instance, assume a species is *defined* through a vector of morphological and genetic traits. In the DP setting, a new species would be characterized by a clearly distinct vector. However, suppose a new species is a hybrid of two existing species. The present DP can not distinguish this case from any other distinct vector. However, a DP that allows for different components of the θ_t's to be drawn from different θ_k^*'s would. Evidently, if some new species arise as such hybrids, the latter process could introduce far fewer clusters yielding a much simpler story for speciation.

If the θ_t's are say $r \times 1$, i.e., $\theta_t^T = (\theta_{t1}, \theta_{t2}, ..., \theta_{tr})$, process specification requires joint probabilities of the form $p_{i_1, i_2, ... i_r} = Pr(\theta_{t1} = \theta_{i_1, 1}^*, \theta_{t2} = \theta_{i_2, 2}^*, ..., \theta_{tr} = \theta_{i_r, r}^*)$ where the $i_l \in \{1, 2, ..., K\}$. Such specification will be referred to as r-dimensional stick-breaking and is discussed in some detail in Section 5.3.

Our goal is a bit more ambitious. In the spatial setting, we want local surface selection among the process realizations that define the SDP or SDP_k. Thus, we need to provide such selection for any number of and choice of locations. Moreover, we seek to do this in a spatially structured way. The closer two locations are the more likely they are to select the same surface; when sufficiently far apart, surface selection will be essentially independent. Rather than attempting to sample the p's above, it will prove easier to simulate indicator variables that produce these p's.

5.2. *The Basics of the GSDP*

Though the SDP realizes a countable collection of surfaces it can not capture the situation in which different surfaces can be selected at different locations. For example, in brain imaging, with regard to neurological activity level, researchers imagine healthy brain images (surfaces) as well as diseased or impaired brain images; however, for an actual image, only a portion of the brain is diseased and it is appropriate to envision surface selection according to where the brain is diseased.

To formalize the GSDP, 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 $\mathcal{B}(\mathbb{R})$,

$$pr\{\theta(s_1) \in A_1, \ldots, \theta(s_n) \in A_n\} = \sum_{i_1=1}^{K} ... \sum_{i_n=1}^{K} p_{i_1, ..., i_n} \, \delta_{\theta_{i_1}^*(s_1)}(A_1) \, ... \, \delta_{\theta_{i_n}^*(s_n)}(A_n),$$

(6)

where the θ_j^*'s are independent and identically distributed as $G_0^{(n)}$, i_j is an abbreviation for $i(s_j)$, $j = 1, 2, \ldots, n$, and the weights $\{p_{i_1, ..., i_n}\}$, conditionally on the locations, have a distribution defined on the simplex $\mathbb{P} = \{p_{i_1, ..., i_n} \geq 0 : \sum_{i_1=1}^{K} ... \sum_{i_n=1}^{K} p_{i_i, ..., i_n} = 1\}$ and independent of that for the $\theta's$.

Evidently, (6) allows the possibility to choose different surfaces at different locations. However, the weights need to satisfy a consistency condition in order to properly define a random process for $\theta(\cdot)$. Specifically, for any set of locations $(s_1, \ldots, s_n)$, $n \in \mathbb{N}$ and for all $\ell \in \{1, \ldots, n\}$, we need

$$p_{i_1, ..., i_{\ell-1}, i_{\ell+1}, ..., i_n} = p_{i_1, ..., i_{\ell-1}, \cdot, i_{\ell+1}, ..., i_n} \equiv \sum_{j=1}^{K} p_{i_1, ..., i_{\ell-1}, j, i_{\ell+1}, ..., i_n}. \qquad (7)$$

So, although the suppression of the $s_1, s_2, ..., s_n$ in (7) may disguise it, the collection of probabilities is really a process. Hence, the weights must satisfy a continuity property (essentially, Kolmogorov consistency of the finite dimensional laws); for locations s and s_0, as $s \to s_0$, $p_{i_1,i_2} = pr\{\theta(s) = \theta_{i_1}^*(s), \theta(s_0) = \theta_{i_2}^*(s_0)\}$, tends to the marginal probability $p_{i_2} = pr\{\theta(s_0) = \theta_{i_2}^*(s_0)\}$ when $i_1 = i_2$, and to 0 otherwise.

Extension to three or more locations is clear and we refer this property as almost sure continuity of the weights (suggested by the almost sure continuity of the paths of a univariate spatial process, as defined in Kent (1989) or Banerjee *et al.* (2003). Suppose we also assume the random field G_0 to be almost surely continuous (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 $||s - s_0|| \to 0$). Then, for the random field given by (6) if the set of weights $\{p_{i_1,...,i_n}\}$ and the base random field G_0 are almost surely continuous, $\theta(s)$ converges weakly to $\theta(s_0)$ with probability one as $||s - s_0|| \to 0$ for all $s_0 \in D$.

Conditionally on the realized distribution G, the process has first and second moments given by (and comparable with (4))

$$
\begin{aligned}
E\{\theta(s)|G\} &= \sum_{l=1}^{K} p_l(s)\, \theta_l^*(s) \\
var\{\theta(s)|G\} &= \sum_{l=1}^{K} p_l(s)\, \theta_l^{*^2}(s) - \left\{ \sum_{l=1}^{K} p_l\, \theta_l^*(s) \right\}^2,
\end{aligned}
$$

and, for a pair of sites s_i, s_j,

$$
\begin{aligned}
cov\{\theta(s_i), \theta(s_j)|G\} &= \sum_{l=1}^{K} \sum_{m=1}^{\infty} p_{l,m}(s_i, s_j)\, \theta_l^*(s_i)\, \theta_m^*(s_j) + \\
&\quad - \left\{ \sum_{l=1}^{K} p_l(s_i)\, \theta_l^*(s_i) \right\} \left\{ \sum_{m=1}^{K} p_m(s_j)\, \theta_m^*(s_j) \right\}.
\end{aligned}
\tag{8}
$$

(8) shows that with almost surely continuous realizations from the base process and of the weights, the GSDP is mean square continuous.

As with the SDP, the process $\theta(\mathbf{s})$ has heterogenous variance and is nonstationary. However, marginalizing over G clarifies the difference. 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, $E\{\theta(s)\} = 0$ and $var\{\theta(s)\} = \sigma^2$ as before, but now

$$
cov\{Y(s_i), Y(s_j)\} = \sigma^2 \rho_\phi(s_i - s_j) \sum_{l=1}^{K} E\{p_{ll}(s_i, s_j)\}.
\tag{9}
$$

Notice that $\sum_{l=1}^{K} E\{p_{ll}(s_i, s_j)\} < 1$ so, marginally, 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 (9), the process $\theta(s)$ 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.

5.3. *Multi-dimensional Stick-breaking*

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 multivariate stick-breaking construction which we detail in the bivariate setting. Sethuraman's univariate stick-breaking construction has weights p_l defined above (1). Denote the random events $\{\theta = \theta_l^*\}$ by Θ_l^1 (with their complements Θ_l^0) and interpret the sequence of weights $\{p_1, p_2, \ldots\}$ as arising from $q_1 = pr\{\Theta_l^1\}$, $q_l = pr\{\Theta_l^1|\Theta_m^0, m < l\} = pr\{Y = \theta_l^*|Y \neq \theta_m^*, m < l\}$, $l = 1, 2, \ldots$. At each location let

$\Theta_l^u(s)$, $u = 0, 1$, be such that $\Theta_l^1(s) = \{\theta(s) = \theta_l^*(s)\}$ and $\Theta_l^0(s) = \{\theta(s) \neq \theta_l^*(s)\}$. Then, for any two locations s_i, s_j, we can consider the probabilities $q_{1,u,v}(s_i, s_j) = pr\{\Theta_1^u(s_i), \Theta_1^v(s_j)\}$, $q_{l,u,v}(s_i, s_j) = pr\{\Theta_l^u(s_i), \Theta_l^v(s_j) | \Theta_m^0(s_i), \Theta_m^0(s_j), m < l\}$, $l \geq 2$, $u, v \in \{0, 1\}$. Note that $q_{l,1,1}(s_i, s_j) + q_{l,1,0}(s_i, s_j) = q_{l,1,+}(s_i, s_j) = q_l(s_i)$. Similarly, $q_{l,+,1}(s_i, s_j) = q_l(s_j)$. Then, we can define

$$
\begin{aligned}
p_{lm} &= pr\{\theta(s_i) = \theta_l^*(s_i), \theta(s_j) = \theta_m^*(s_j)\} \\
&= pr\{\Theta_l^1(s_i), \Theta_m^1(s_j), \Theta_k^0(s_i), k < l, \Theta_r^0(s_j), r < m\} \\
&= \begin{cases}
\prod_{k=1}^{l-1} q_{k,0,0}\, q_{l,1,0} \prod_{r=l+1}^{m-1}(1 - q_r)\, q_m & \text{if } l < m \\
\prod_{r=1}^{m-1} q_{r,0,0}\, q_{m,0,1} \prod_{k=m+1}^{l-1}(1 - q_k)\, q_l & \text{if } m < l \\
\prod_{r=1}^{l-1} q_{r,00}\, q_{l,11} & \text{if } l = m.
\end{cases}
\end{aligned}
\tag{10}
$$

Inspection of expression (10) reveals that the weights are determined through a partition of the unit square. For n locations the construction is on the unit n-dimensional hypercube and requires 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 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), that is

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

We specify the weight process through latent spatial processes.

5.4. *A fully-specified GSDP*

To generalize an SDP we can specify the foregoing components consistently if we allow a latent process to determine surface selection, that is, the stick-breaking components $q_{l,u_1,\ldots,u_n}(s_1, \ldots, s_n)$ arise through probabilities associated with the events $\Theta_l^{u_j}(s_j)$. In particular, consider the process $\{\delta_{\Theta_l^1(s)}, s \in D, l = 1, 2, \ldots, \}$, such that at any $l = 1, 2, \ldots$, $\delta_{\Theta_l^1(s)} = 1$ if $\Theta_l^1(s)$ occurs, $\delta_{\Theta_l^1(s)} = 0$ if $\Theta_l^1(s)$ does not occur. In turn, 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 defined by $\delta_{A_l(s)}^* = 1$ if $Z_l(s) \in A_l(s)$, $\delta_{A_l(s)}^* = 0$ 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

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

It is easy to see that such a characterization guarantees that 7 is true.

We employ Gausssian thresholding to provide binary outcomes, i.e., $A_l(s) = \{Z_l(s) \geq 0\}$ as in Albert and Chib (1993). This is computationally convenient and, as a model for second stage random effects, there will be little posterior sensitivity to this choice. Suppose $\{Z_l(s), s \in D, l = 1, 2, \ldots\}$ is a countable collection of

independent stationary Gaussian random fields with unknown (hence random) mean $\mu_l(s)$, variance 1, and correlation function $\rho_Z(\cdot, \eta)$. 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, $l = 1, 2, \ldots$, then so are the $q_{l,u_1,\ldots,u_n}(s_1,\ldots,s_n)$'s.

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

$$q_{l,1}(s) = pr\{Z_l(s) \geq 0\} = 1 - \Phi\{-\mu_l(s)\} = \Phi\{\mu_l(s)\}. \tag{11}$$

If $\mu_l(s) = \mu_l$, for all s, with $\Phi(\mu_l)$ independent $\mathrm{Beta}(1,\nu)$ then, 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 $\theta_l^*(s) \neq \theta_l^*(s')$. More generally, if the $\mu_l(s)$ are such that the $\Phi\{\mu_l(s)\}$ are independent $\mathrm{Beta}(1,\nu)$, $l = 1, 2, \ldots$, now the marginal distribution of $\theta(s)$ is a DP with probabilities that vary with location. Spatially varying weights have recently been considered by Griffin and Steel (2004), who work in the framework of dependent Dirichlet processes.

5.5. *Generalized finite-dimensional spatial Dirichlet Process (GSDP$_K$).*

The above GSDP is, in fact, fitted with a finite sum approximation. But now we consider a much different construction when K is finite in 6. If the random weights are Dirichlet distributed for any choice of n and $s_1, \ldots, s_n$, i.e., $p_{i_1 \ldots i_n} \sim Dir(\{\alpha_{i_1,\ldots,i_n}(s_1,\ldots,s_n)\})$, we say that G is a generalized finite dimensional spatial Dirichlet process, in symbols, $G \sim GSDP_K(\boldsymbol{\alpha}, G_0)$, where $\boldsymbol{\alpha}$ denotes the measure that specifies the $\alpha_{i_1,\ldots,i_n}(s_1,\ldots,s_n)$. In particular, a $GSDP_K$ is a SDP_K process if, for any choice of n and $(s_1,\ldots,s_n)$, $\alpha_{i_1,\ldots,j_n}(s_1,\ldots,s_n) = 0$ unless $i_1 = \cdots = i_n$. We want to allow the possibility of choosing different surfaces at different locations, but, again, as $s \to s'$ we want $\theta(s) \to \theta(s')$ in distribution.

Suppose we want the measure $\alpha(\cdot)$ to have uniform marginals, i.e., $\alpha_i(s) = \alpha/K$, for some real constant α. Again, we illustrate in the case of two locations (s_1, s_2). Let $\alpha_{i+}(s_1) = \sum_{j=1}^{K} \alpha_{i,j}(s_1, s_2)$ and $\alpha_{+j}(s_2) = \sum_{i=1}^{K} \alpha_{i,j}(s_1, s_2)$. To obtain uniform marginals, we have to set $\alpha_{i+}(s_1) = \alpha_{+j}(s_2) = \alpha/K$ for any $i, j = 1, \ldots, K$. We can achieve this with $\alpha_{i,i} = \alpha/K$ and $\alpha_{i,j} = 0$ for $i \neq j$ (the SDP_K). The choice $\alpha_{i,j} = \alpha/K^2, i, j = 1, \ldots, K$ achieves independent surface selection across locations. To take into account spatial dependence we take up ideas in Petrone, Guindani and Gelfand (2006). Let $\alpha_{i,j} = \alpha E(p_{i,j}) = \alpha a_{i,j}$, $i, j = 1, \ldots, K$, where the $a_{i,j}$'s are defined so that $a_i = a_{i+} = \sum_{j=1}^{K} a_{i,j} = 1/K$. Now, let $H(\cdot, \cdot; \tau)$ be a distribution function on $[0,1]^2$, with uniform marginals. In other words, let $\boldsymbol{\mathcal{U}} = (\mathcal{U}_1, \mathcal{U}_2)^T$ be a random vector such that $(\mathcal{U}_1, \mathcal{U}_2) \sim H(\cdot, \cdot; \tau)$ and $\mathcal{U}_r \sim U(0,1)$, $r = 1, 2$. Then, for given K, we can partition the unit interval in the K intervals $(\frac{i-1}{K}, \frac{i}{K}]$, $i = 1, \ldots, K$, and correspondingly consider the induced partition of the unit square made of sets $Q_{i,j} = (\frac{i-1}{K}, \frac{i}{K}] \times (\frac{j-1}{K}, \frac{j}{K}]$, $i, j = 1, \ldots, K$.

Therefore, we can set $E(p_{i,j})$ as the probability that $(\mathcal{U}_1, \mathcal{U}_2)$ belong to $Q_{i,j}$, that is

$$a_{i,j} = P_H\left(\mathcal{U}_1 \in \left(\frac{i-1}{K}, \frac{i}{K}\right], \mathcal{U}_2 \in \left(\frac{j-1}{K}, \frac{j}{K}\right]\right), \quad i, j = 1, \ldots, K. \tag{12}$$

Of course, marginally $a_i = a_{i+} = a_{+i} = P_H\left(\mathcal{U}_1 \in (\frac{i-1}{K}, \frac{i}{K}]\right) = \frac{1}{K}$. Then, $\alpha_{i,j} = \alpha P_H(Q_{i,j}; \tau)$ and $\alpha_i = \alpha/K$ as desired.

Next, we introduce spatial dependence through a "copula argument". Let V be say, a mean 0 GP on D with covariance function $\gamma^2 \rho_\psi(s, s')$. For locations s and s' with $\mathbf{V} = (V(s), V(s'))$ let $\mathcal{U}_1 = \Phi_1\left(\frac{V(s)}{\gamma}\right), \mathcal{U}_2 = \Phi_2\left(\frac{V(s')}{\gamma}\right)$ and

$$
\begin{aligned}
H_V(u_1, u_2; \tau) &= P(\mathcal{U}_1 \leq u_1, \mathcal{U}_2 \leq u_2) \\
&= P(V(s) \leq \gamma\,\Phi_1^{-1}(u_1), V(s') \leq \gamma\,\Phi_2^{-1}(u_2)) \\
&= \Phi^{(2)}\left(\gamma\,\Phi^{-1}(u_1), \gamma\,\Phi^{-1}(u_2)\right).
\end{aligned}
\tag{13}
$$

The distribution function $H_V(u_1, u_2; \tau)$ is the copula of the distribution of V. If $\hat{Q}_{i,j} = \left(\gamma\Phi^{-1}\left(\frac{i-1}{K}\right), \gamma\Phi^{-1}\left(\frac{i}{K}\right)\right] \times \left(\gamma\Phi^{-1}\left(\frac{j-1}{K}\right), \gamma\Phi^{-1}\left(\frac{j}{K}\right)\right]$, for any $i, j = 1, \ldots, K$ then, $a_{i,j} = P_{H_V}\left(\mathcal{U} \in Q_{i,j}\right) = P_V\left((V(s), V(s')) \in \hat{Q}_{i,j}\right)$. Extension to the case of n locations is straightforward. It is enough to consider a distribution $H(\cdot)$ on the n-dimensional unit hypercube having uniform marginals coupled with the associated n-dimensional multivariate normal, arising from the GP.

A remaining question concerns the behavior of the $GSDP_K$ as $K \to \infty$. Clearly, the limiting behavior of the $GDP_K(\boldsymbol{\alpha}^{(K)} G_0(s, s'))$ for $K \to \infty$ depends on the specification of the Dirichlet parameters $\boldsymbol{\alpha}^{(K)} = (\alpha_{i,j}^{(K)}, i, j = 1, \ldots, K)$. The limiting behavior of the vector $\boldsymbol{p}^{(K)}$ of the K^n random probabilities $p_{i,j}^{(K)}$ can be obtained from results in Kingman (1975). That is, if $\max(\alpha_{i,j}^{(K)}, i, j = 1, \ldots, K)$ converges to zero and $\sum_{i=1}^{K} \sum_{j=1}^{K} \alpha_{i,j}^{(K)} \to \alpha$, $0 < \alpha < \infty$, then the *ordered* vector of the $p_{i,j}^{(K)}$'s converges in law to a Poisson-Dirichlet distribution. However, the limiting distribution depends more specifically on the behavior of partial sums such as $\sum_{j=1}^{K} \alpha_{i,j}^{(K)}$. Furthermore, since the support points of $G \sim GSPD_K(\boldsymbol{\alpha}^{(K)}, G_0)$ are dependent, the limit of the random process G does not follow directly from that of the vector $\boldsymbol{p}^{(K)}$. For instance, when the $GSDP_K$ is an SDP_K it converges weakly to a SDP with scale parameter α and base measure G_0. At the other extreme, with independent surface selection, the $GDP_K(\boldsymbol{\alpha}^{(K)} G_0(s, s'))$ converges weakly to a DP with scale parameter α and base measure $G_0(s)G_0(s')$. Finally, an intermediate choice might be $\alpha_{i,i} = a/k + b/k^2, \alpha_{i,j} = a/k^2, i \neq j$. In this case, we can show that the GDP_K converges weakly to a DP with base measure given by the mixture $aG_0(s, s') + bG_0(s)G_0(s')$ (general discussion is in Petrone et al.(2006)).

6. SIMULATED EXAMPLE COMPARING SDP, SDP_K AND $GSDP$, $GSDP_K$

We compare the SDP and our generalizations by means of a simulated data set (see Duan *et al.* (2005)). Data are generated from a finite mixture model of GPs, whose weights are assumed spatially varying. Let $\mathbf{Y}_t = (Y_t(s_1), \ldots, Y_t(s_n))^T, t = 1, 2..T$ be a set of independent observations at a set of locations $(s_1, \ldots, s_n)$. Then, each $Y_t(s)$ arises from a mixture of two GPs, $G_{0,s}^1$ and $G_{0,s}^2$, respectively, with mean ξ_i and covariance function $\sigma_i^2 \rho_{\psi_i}(s, s')$, $i = 1, 2$, $s, s' \in D$, such that $Y_t(s) \sim \alpha(s)G_{0,s}^1 + (1 - \alpha(s))G_{0,s}^2$. The marginal weight is $\alpha(s) = P(Z(s) \geq 0)$, where $Z(s)$ is a mean zero stationary GP with covariance function $\rho_\eta(s - s')$. Therefore, we choose $Y_t(s)$ from $G_{0,s}^1$ if $Z_t(s) \geq 0$ or from $G_{0,s}^2$ if $Z_t(s) < 0$. Since $Z(s)$ is centered at zero, marginally we have $Y_t(s) \sim \frac{1}{2}N_n\left(\xi_1, \sigma_1^2\right) + \frac{1}{2}N_n\left(\xi_2, \sigma_2^2\right)$. However, the joint
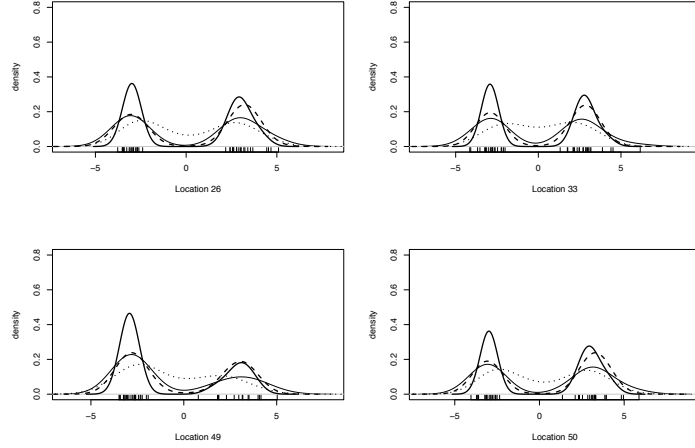
Figure 5: *True density (lighter line −) and predictive posterior density, respectively, under the SDP (light dotted line), the GSDP (thick dotted line) and the $GSDP_K$ models for the simulation example in Section 6.*

distribution for a pair s, s' in D is

$$
\begin{aligned}
(Y_t(s), Y_t(s')) \sim \quad & \alpha_{1,1}(s, s') \, G_{0,s,s'}^1 + \alpha_{2,2}(s, s') \, G_{0,s,s'}^2 + \\
& + \alpha_{1,2}(s, s') \, G_{0,s}^1 \, G_{0,s'}^2 + \alpha_{2,1}(s, s') \, G_{0,s}^2 \, G_{0,s'}^1,
\end{aligned}
\tag{14}
$$

where $\alpha_{i,j} = P\left((-1)^{i+1} Z(s) > 0, (-1)^{j+1} Z(s') > 0\right)$, $i, j = 1, 2$. Therefore, when s and s' are near to each other, it is very likely that $Y_t(s)$ and $Y_t(s')$ are from the same component. On the other hand, if s and s' are distant, the linkage between $Z(s)$ and $Z(s')$ is weak, so that $Y_t(s)$ and $Y_t(s')$ are chosen almost independently.

In our experiment, we have $n = 50$ random sites in a rectangle and $T = 40$. We set the parameters as follows: $\xi_1 = -\xi_2 = 3$, $\sigma_1 = 2\,\sigma_2 = 2$, $\phi_1 = \phi_2 = 0.3$, and $\eta = 0.3$. We fit the data using the SDP, the GSDP and the $GSDP_K$ with $K = 20$. To focus on the modeling of the spatial association, we assume $\mu(s) = 0$ in 3.

In Figure 5, we plot the true density and the posterior predictive densities under the three models for four selected locations ($s_{26}, s_{33}, s_{49}, s_{50}$). The values of the 40 observations at each of these 4 locations are shown along the x-axis. The nonparametric models with spatially varying weights provide estimates closer to the true densities of the model and the data. This is confirmed by the bivariate plots in Figure 6 (locations 26 and 50 are close, 49 and 50 are far apart) where we plot the probability contours of the true density along with the posterior distribution of the mean for the other models. It's interesting to compare these contours with 14. The posterior distributions of the $GSDP$ and $GSDP_K$ capture the expected behavior of the true density; explicit spatial modeling of the weights enables us to capture local details, as is also revealed by the heights of the local modes.
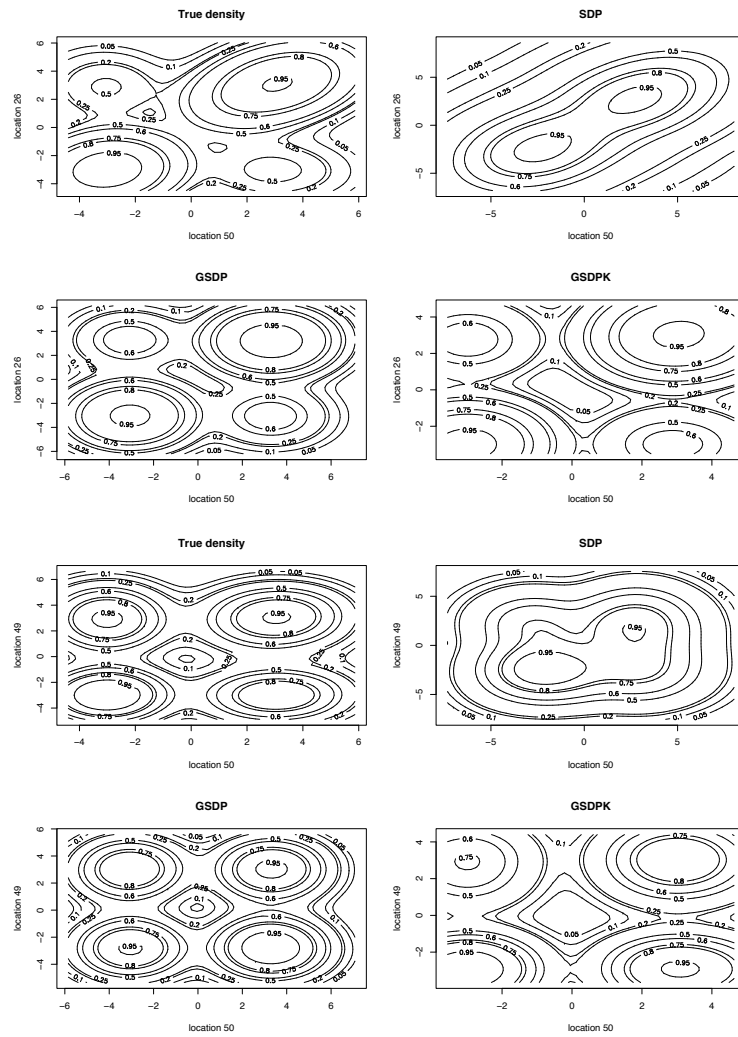
Figure 6:   *Bivariate contour plots of the true densities and the poste-*
rior distributions of the mean of the SDP, GSDP and $GSDP_K$ models for
locations 26 *and* 50 *(a) and* 49 *and* 50 *(b). See section 6.*

7. SUMMARY AND FURTHER ISSUES

We have developed spatial Dirichlet process specifications that can be used as an alternative to customary Gaussian process choices to model spatial random effects. Such models provide nonstationary, nonGaussian processes that can have locally varying DP marginals and can both marginally and jointly capture multimodalities. We have demonstrated this with both real and simulated datasets.

In the case where the replications might be time dependent, we can embed any of the foregoing specifications within a dynamic model. See Gelfang *et al.* (2005) for the SDP case and Duan *et al.* (2005) for the GSDP.

Multivariate spatial process modelling using GP's is reviewed in Gelfand et al. (2004) but the nonparametric setting has not been addressed. Gelfand et al. advocate coregionalization (random linear transformation of independent univariate process models). Here, coregionalization could be applied to the base measure or, possibly, by random linear transformation of independent spatial DP's.

Finally, as noted in the Introduction, an alternative role for this modelling is for functional data analysis (FDA). Here, we replace geographic space, $s \in D$, with covariate space, $z \in Z$, seeking to model a random function of z to explain responses Y. For a collection of individuals, modelling these functions as individual-level process realizations leads to DP's where the atoms are random functions. In further extension to spatial FDA, model development would use DP specifications for both the spatial and the functional aspects of the modelling.

REFERENCES

Agarwal, D. and Gelfand, A. (2005). Slice gibbs sampling for simulation based fitting of spatial data models. *Statist. Computing* **15**, 61–69.

Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88**, 669–679.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2**, 1152–1174.

Banerjee, S., Carlin, B. and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. London: Chapman and Hall

Banerjee, S., Gelfand, A. and Sirmans, C. (2003). Directional rates of change under spatial process models. *J. Amer. Statist. Assoc.* **98**, 946–954.

Barry, R. P. and Ver Hoef, J. M. (1996). Blackbox kriging: Spatial prediction without specifying variogram models. *J. Agricultural, Biological and Envir. Statistics*, **1**, 297–322.

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

Cifarelli, D. and 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, 12**, 1–13.

Clyde, M. and Wolpert, R. L. (2006). Bayesian modelling with overcomplete representations. *In this volume.*

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

De Iorio, M., Müller, P., Rosner, G. and MacEachern, S. (2004). An anova model for dependent random measures. *J. Amer. Statist. Assoc.* **99**, 205–215.

Diggle, P. J., Liang, K. and Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford: Clarendon Press.

Duan, J. A., Guindani, M. and Gelfand, A. E. (2005). Generalized Spatial Dirichlet Process Models. Submitted.

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

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

Fuentes, M. and Smith, R. (2001). A new class of nonstationary spatial models. *Tech. Rep.*, North Carolina State University, USA.

Gelfand, A. E., Schmidt, A. M, Banerjee, S. and Sirmans, C .F. (2004) Nonstationary multivariate process modelling through spatially varying coregionalization. *Test* **13**, 1–50, (with discussion).

Gelfand, A., Banerjee, S. and Gamerman, D. (2005a). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, **16**, 1–15.

Gelfand, A., Kottas, A. and MacEachern, S. (2005b). Bayesian nonparametric spatial modeling with Dirichlet processes mixing. *J. Amer. Statist. Assoc.* **100**, 1021–1035.

Griffin, J. and Steel, M. (2004). A class of dependent Dirichlet processes. *Tech. Rep.*, University of Kent at Canterbury, UK.

Higdon, D., Swall, J. and Kern, J. (1999). Non-stationary spatial modeling. *Bayesian Statistics 6* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) Oxford: University Press, 287–302.

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

Ishwaran, H. and James, L. (2001). Gibbs sampling methods for stik breaking priors. *J. Amer. Statist. Assoc.* **96**, 161–173.

Ishwaran, H. and Zarepour, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika* **87**, 371–390.

Ishwaran, H. and Zarepour, M. (2000a). Dirichlet prior sieves in finite normal mixtures. *Statistica Sinica* **12**, 941–963.

Ishwaran, H. and Zarepour, M. (2000b). Exact and approximate sum-representations for the Dirichlet process. *Can. J. Statist.* **30**, 269–283.

Jain, S. and Neal, R. M. (2004). A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. *J. Comp. Graphical Statist.* **13**, 158–182.

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

Kingman, J. F. C. (1975). Random discrete distributions. *J. Roy. Statist. Soc. B* **37**, 1–22.

Kottas, A. , Duan, J. and Gelfand, A. E. (2006). Modeling Disease Incidence Data with Spatial and Spatio-temporal Dirichlet Process Mixtures *Tech. Rep.*, ISDS, Duke University, USA.

Le, N. and Zidek, J. (1992). Interpolation with uncertain spatial covariances: a Bayesian alternative to kriging. *J. Multivariate Analysis* **43**, 351–374.

MacEachern, S. (2000). Dependent Dirichlet processes. *Tech. Rep.*, The Ohio State University, USA.

Mardia, K. V. and Goodall, C. R. (1993). Spatial-temporal analysis of multivariate environmental monitoring data. *Multivariate Environmental Statistics*, (G. P. Patil and C. R. Rao, eds.) Amsterdam: North-Holland, 347–386.

Muliere, P. and Secchi, P. (1995). A note on a proper Bayesian bootstrap. *Tech. Rep.*, Università di Pavia, Italy.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *J. Comp. Graphical Statist.* **9**, 249–265.

Ongaro, A. and Cattaneo, C. (2004). Discrete random probability measures: a general framework for nonparametric Bayesian inference. *Statistics and Probability Letters* **67**, 33–45.

Palacios, M. and Steel, M. (2006). Non-gaussian Bayesian geostatistical modeling. *J. Amer. Statist. Assoc.* (to appear).

Petrone, S., Guindani, M. and Gelfand, A. (2006). Finite mixtures of spatial processes. *Tech. Rep.*, ISDS, Duke University, USA.

Petrone, S. and Raftery, A. (1997). A note on the Dirichlet process prior in Bayesian non-parametric inference with partial exchangeability. *Statistics and Probability Letters* **36**, 69–83.

Pitman, J. and Yor, M. (1997). The two-parameter poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Prob.* **25**, 855–900.

Richardson, S. and Green, P. (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. B* **59**, 731–792.

Sampson, P. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *J. Amer. Statist. Assoc.* **87**, 108–119.

Scaccia, L. and Green, P. J. (2003). Bayesian growth curves using normal mixtures with nonparametric weights. *J. Comp. Graphical Statist.* **12**, 308–331.

Schmidt, A. and O'Hagan, A. (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *J. Roy. Statist. Soc. B* **65**, 743–758.

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

Sethuraman, J. and Tiwari, R. (1982). Convergence of Dirichlet measures and the interpretation of their parameter. *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) New York: Academic Press, 305–315.

Shapiro, A. and Botha, J. D. (1991) Variogram fitting with a general class of conditionally nonnegative definite functions. *Comput. Statist. Data Anal.* **11**, 87–96.

Stein, M. L. (1999). *Interpolation of Spatial Data – Some theory of Kriging.* Berlin: Springer.

Stein, M. L. (2005) Space-time covariance functions. *J. Amer. Statist. Assoc.* **100**, 310–321.

DISCUSSION

NILS LID HJORT (*University of Oslo, Norway*)

I have followed the work of Alan Gelfand and various co-authors on Bayesian spatial statistics over the past few years with interest. There are many aspects and sub-themes, from detailed technicalities of modelling and simulation strategies to the challenges of actual applications, and I very much welcome the present survey paper.

It's about time

We have had spatial statistics since around 1950 (starting with Bertil Matérn and Danie Krige) and nonparametric Bayesian statistics since around 1970 (with the early work of Thomas Ferguson, Kjell Doksum, David Blackwell and Charles Antoniak). These brooks have grown into strong rivers, and are according to current estimates of spatial-temporal derivatives in the process of extending themselves into veritable floods. So it was a question of time until the twain met, and so they did: I view the spatial Dirichlet processes of Gelfand and co-authors as among the first serious attempts at intersecting 'nonparametric Bayes' with 'spatial statistics'.

Both of these areas are inherently large, and need to be so: Nonparametrics lives in infinite-dimensional spaces, and the ratio of the number of ways in which variables can be dependent divided by the number of ways they can be independent is also infinite. It would follow that the new, fledgling intersection area of 'spatial Bayesian nonparametrics' also will need to grow big. In this perspective I think we must realise that spatial Dirichlet processes, although a broadly versatile machinery, must not rule alone, or for too long. I expect Bayesian spatial statistics to grow healthily in many directions over the coming ten years (and beyond), and view the spatial Dirichlet process direction as one of several.

It is perhaps unexpected that Dirichlet processes should turn out to be such a broadly flexible tool also in spatial contexts. Perhaps parallel developments may be

expected in the model world of Gaußian and nearly Gaußian processes. Consider the basic application example of Gelfand, Guindani and Petrone, which takes the form

$$Y_t(s) = \mu_t(s) + \theta_t(s) + \epsilon_t(s) \tag{15}$$

for modelling say monitoring stations over time, at time points $t_1, \ldots, t_{40}$ and spatial positions $s_1, \ldots, s_{45}$, where $\mu_t(s)$ is trend, $\epsilon_t(s)$ is white noise, and $\theta_t(s)$ is the main spatial object. I wish to express a modicum of polite surprise that statisticians after staring at (15) for a while would say, "Of course! Dirichlet processes!". But, apparently, they do. I would expect there to be alternative fruitful models, e.g. built around Gaußian processes or relatives.

For a given spatial correlation function, giving rise to an $n \times n$ variance matrix for $\theta(s_1), \ldots, \theta(s_n)$ of the form $\sigma^2 \Sigma_n(\phi)$, say, the pure Gaußian model amounts to

$$\begin{pmatrix} \theta(s_1) \\ \vdots \\ \theta(s_n) \end{pmatrix} = \sigma \Sigma(\phi)^{1/2} \begin{pmatrix} N_1 \\ \vdots \\ N_n \end{pmatrix} \tag{16}$$

in terms of independent standard normals. There are several ways in which nonparametric envelopes, or sausages, can be built around the normal, with a parameter dictating the degree of concentration around the given normal distribution. Thus different forms of spatial nonparametric models would emerge by taking

$$N_1, \ldots, N_n \sim \text{envelope}(\text{N}(0,1); \delta),$$

in concert with (16). Here δ indicates some parameter that governs the tightness around the standard normal. It would be interesting to have such modelling attempts contrasted with those of Gelfand *et al.*, see also Gamerman *et al.* (2007).

Different versions of spatial Dirichlet processes

How can one make two Dirichlet processes G and G' dependent? Suppose both of them are required to have parameters aG_0, say. Again, there must be several different fruitful ways of achieving this. Gelfand et al. work with the Tiwari–Sethuraman representation, where

$$G = \sum_{j=1}^{\infty} p_j \delta(\theta_j) \quad \text{and} \quad G' = \sum_{j=1}^{\infty} p'_j \delta(\theta'_j),$$

say, where the $\{\theta_j\}$ sequence needs to be i.i.d. from G_0, as does the $\{\theta'_j\}$ sequence. One class of models emerges by keeping $p_j = p'_j$ but allowing dependence between θ_j and θ'_j; cf. MacEachern's (2000) early work that set off some of the later developments. One may also simultaneously work with dependence between the $\{p_j\}$ and $\{p'_j\}$ sequences. In the usual set-up,

$$p_j = \bar{q}_1 \cdots \bar{q}_{j-1} q_j \quad \text{and} \quad p'_j = \bar{q}'_1 \cdots \bar{q}'_{j-1} q'_j,$$

with $q_j \sim \text{Beta}(1, a)$ and $q'_j \sim \text{Beta}(1, a)$, writing $\bar{q}_j = 1 - q_j$ and $\bar{q}'_j = 1 - q'_j$. A natural recipe is now to take

$$q_j = \text{Beta}^{-1}(\Phi(N_j), 1, a) \quad \text{and} \quad q'_j = \text{Beta}^{-1}(\Phi(N'_j), 1, a),$$

where the (N_j, N_j')s are independent pairs, but with internal correlation ρ. This creates two dependent Dirichlet processes, with dependence both from θ_j to θ_j' and from p_j to p_j', and the construction is easily generalised to full random fields, via normal processes. Here $\mathrm{Beta}^{-1}(q, 1, a)$ is the inverse of the cumulative distribution function $\mathrm{Beta}(x, 1, a)$ for the $\mathrm{Beta}(1, a)$ distribution.

A different construction that also looks fruitful on paper is as follows. Let

$$G = \sum_{j=1}^{K} p_j \delta(\theta_j) \quad \text{and} \quad G' = \sum_{j=1}^{K} p_j' \delta(\theta_j'),$$

for a moderate or large K. As long as

$$p \sim \mathrm{Dir}(a/K, \ldots, a/K) \quad \text{and} \quad p' \sim \mathrm{Dir}(a/K, \ldots, a/K), \tag{17}$$

it is known that G and G' converge in distribution to the $\mathrm{Dir}(aG_0)$ process; see discussion about this in Hjort (2003). The problem is therefore to make p and p' dependent without losing (17). Here one method is to define $p_j = A_j/A$ and $p_j' = A_j'/A'$, where A and A' are the sums of A_js and A_j's, respectively, and where

$$A_j = \mathrm{Gam}^{-1}(\Phi(N_j), a/K, 1) \quad \text{and} \quad A_j' = \mathrm{Gam}^{-1}(\Phi(N_j'), a/K, 1),$$

and again $\mathrm{corr}(N_j, N_j') = \rho$. Here $\mathrm{Gam}^{-1}(q, a/K, 1)$ is the quantile inverse of the distribution function for a $\mathrm{Gam}(a/K, 1)$ variable. Other copulae (also called 'the emperor's new clothes', see Mikosch, 2007) can also be employed here.

One might also contemplate versions of this where only the biggest Dirichlet process jumps are allowed to enter the approximations, say

$$G = \sum_{j=1}^{K} b_j \delta(\theta_j) \Big/ \sum_{j=1}^{K} b_j,$$

where $b_1 > \cdots > b_K$ are the K biggest jumps in a $\mathrm{Dir}(aG_0)$ process. This is moderately awkward, but doable, via results in Hjort and Ongaro (2006).

Is the spatial component really required?

This section title is not meant too provocatively, since I of course acknowledge that the spatial part of the model often needs to be there. But I wish to have in the toolbox various checks and tests for helping me decide whether I need the spatial component $\theta_t(s)$ in (15) or not. It is important to realise that in (15), both zero-mean terms $\theta_t(s)$ and $\epsilon_t(s)$ change values and interpretation with what we put into the trend part $\mu_t(s)$. Specifically, adding one more clever covariate with strong explanatory power will sometimes carry enough spatial information to make $\theta_t(s)$ become too small in size to really matter. I do suppose Gelfand et al.'s machinery can give me $\pi(\sigma \,|\, \text{data})$ relatively easily, about the standard deviation of the $\theta_t(s)$ part, or perhaps even more informatively the joint posterior of (σ, τ). But it would be nice to have a formalised procedure specifically for answering the question 'can I just as well set $\theta_t(s)$ to zero?' for each given application. This could perhaps be in the form of a post-processed posterior predictive p-value, as in Hjort, Dahl and Steinbakk (2006), or via a suitable Bayesian test, as in Rousseau (2007).

There might be Bernshteǐn–von Mises theorems to prove in these models, to the effect that with enough data things will go very well and different Bayesians will be in decent agreement a posteriori. Very often there would not be overwhelmingly much data, however, particularly in view of the somewhat evasive character of parameters relating to spatial correlation. This is an argument for being very careful with the prior construction. Gelfand et al. use reasonably generic prior descriptions, but they seem to consistently take σ (noise level for $\theta_t(s)$) and τ (noise level for the white part $\epsilon_t(s)$) as independent. This worries me, since it would clash both with Jeffreys' type priors and with other work about setting priors in variance component models?

Side effects

Bayesian and empirical Bayesian constructions are often successful in terms of achieving better overall accuracy of predictions, better average precision for estimating a flock of parameters, etc. This is achieved by intelligent amounts of smoothing and shrinking. There are sometimes less fortunate side effects, however; one might more easily miss 'alarm situations' and anomalies, which in many practical applications would be of great concern, e.g. for pollution monitoring. Or are there good fixes by setting up suitable loss functions, to be combined and assessed from the posterior calculations?

The last point I wish to make is that there is a potential problem with the Dirichlet process as the basic Lego building block in these contexts, related to the two roles of the 'a' parameter in the $\mathrm{Dir}(aG_0)$. As we know, the size of a seriously influences two different aspects of the distribution, (i) the degree of tightness around its prior mean G_0, and (ii) the level of clustering in repeated samples from G. These are not quite reconcilable. This might point to the need for extending the infinite-dimensional prior processes with one more parameter.

STEVEN N. MACEACHERN (*The Ohio State University, USA*)

The authors have written a stimulating paper that introduces a new element to models based on dependent nonparametric processes. A selection surface allows some portions of a realized surface to come from one component of a mixture and other portions of the realized surface to come from other components of the mixture. This innovation generates tremendous flexibility for modelling and will prove useful in many applications.

The description of a formal hierarchical model often lends additional insight to a modelling effort. The subsequent description (at the risk of modestly warping the authors' model) casts the model in a more traditional framework. In doing so, connections to existing processes are seen, and the importance of qualitative features of the selection surface become apparent.

The authors write the model (3) at the observed sites as

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \theta(\mathbf{s}) + \epsilon(\mathbf{s}).$$

This can be divided into three portions–one portion that is far-removed from the data, a second portion that provides the guts of the model and a third portion that connects the model to the data. The model includes a potential regression structure for $\mu(\cdot)$, including dependence on covariates, and it allows for distributions on hyperparameters that govern this regression structure. This is essential for modelling, although not the focus of this work.

The middle of the hierarchical model consists of a dependent nonparametric process. To generate the SDP or the GSDP, this mid-level process is a dependent Dirichlet process (DDP) as defined, for example, in MacEachern (2000). The DDP is defined by two countable sets of independent and identically distributed stochastic processes. The first set of processes is for the locations: Here, $\theta^*(\cdot)$, with $\theta_1^*(\cdot), \theta_2^*(\cdot), \ldots$ a random sample of these processes. These $\theta_l^*(\cdot)$ lead directly to $\theta_l^*(\mathbf{s})$ for vectors of sites in D through the usual finite-dimensional specifications (although there is often little loss in thinking of the entire surface, $\theta^*(\cdot)$ as having been realized).

The second set of processes determines the probabilities associated with the locations. Following MacEachern (2000), $V(\cdot)$ follows a stochastic process where $V(s)$ is distributed as Beta$(1, M)$ with $M > 0$ a mass parameter. From the surfaces $V_1(\cdot), V_2(\cdot), \ldots$, and following Sethuraman (1994), arithmetic takes one to the $p_1(\cdot), p_2(\cdot), \ldots$. Again, $V_l(\mathbf{s})$ is a finite dimensional realization from $V_l(\cdot)$. The choices for the distribution on $V(\cdot)$ are endless. A convenient choice for $V(\cdot)$ maps a Gaussian process with standard normal marginals into the Beta$(1, M)$ with a pair of transformations. Provided the covariance function of the Gaussian process allows $V(s) \neq V(s')$, this produces the "multivariate stick-breaking" that is needed to ensure independence of the random distributions at s and s' (or independence in the limit) s and s' become distant.

At this point, the middle stage of the model is a DDP. The DDP includes multivariate as well as univariate $\theta^*(\cdot)$ (DeIorio et al., 2004). It can be specialized (e.g., $V_l(s) = V_l$ for all s and all l results in the single-p DDP which falls under Sethuraman's definition of the Dirichlet process), or it can be generalized to DDPs with spatially varying mass parameters (see MacEachern, 1999) or, more generally, to a wide variety of dependent nonparametric processes (for example, Hjort, 2000, or MacEachern, Kottas and Gelfand, 2001). In essence, the DDP and its variants describe distribution-valued stochastic processes indexed by covariate values, in this case the spatial location, s.

The final stage of the model presents the novelty. It consists of a likelihood to smooth out the discreteness inherent in mixture models, and it includes a "selector surface". The selector surface is driven by a latent process. Under one view, the surface, say $U(\cdot)$, has the property that $U(s)$ is uniformly distributed on the interval $(0, 1)$ for each s. The value of $U(s)$ determines which component of the mixture is active at site s through the rule that assigns the site to component k if both $\sum_{l=1}^{k} p_l \geq U(s)$ and $\sum_{l=1}^{k-1} p_l < U(s)$. Alternatively, one can pursue an approach whereby spatial regions are assigned to mixture components successively. A latent surface, $Z_1(\cdot)$, with appropriate marginal distribution at site s, is compared to the cutoff determined by $V_1(\cdot)$ at the same site, s. If $Z_1(s) \leq V_1(s)$, the site is assigned to the first component of the mixture; if $Z_1(s) > V_1(s)$, the site is not assigned to the first component of the mixture. A similar process unfolds for further components of the mixture, with comparisons of $Z_l(\cdot)$ to $V_l(\cdot)$. It is evident that the latter approach has several advantages over typical implementations of the former approach.

The selector surfaces have many interesting implications. If there is no variation in either $V(\cdot)$ or $Z(\cdot)$ across sites, the model reduces to the SDP model, a version of the single-p DDP model, and also a version of the Dirichlet process. If the realizations of $V(\cdot)$ and $Z(\cdot)$ show variation and are relatively smooth, one would expect to find edges, where there would be a nice curve separating one mixture component from another. If the realizations are very jagged, one would expect "transition zones" without a clean jump from one component to another. In these zones, one

would expect many switches back and forth between components. With the structure of the mixture model, where $\theta_l^*(s)$ and $\theta_{l'}^*(s)$ will typically be very different, one would expect discontinuities at the transition between mixture components. Properties such as clean breaks in the surface or transition zones are important in applications such as image analysis, where one typically searches for the edges (clean breaks) in the surface and expects discontinuities at these edges. Nicely, the authors' pursuit of the approach with the $Z(\cdot)$ rather than the $U(\cdot)$ approach with continuous spatial paths avoids the problem of always moving to adjacent components when switching from one component to another.

I would like to make a plea to the authors and to others working in the area to refrain from overloading notation, as is sometimes done for the Dirichlet process. For the Dirichlet process, the time-honored tradition (see Ferguson's early work, for example) uses the symbol α for the base measure. With respect for this tradition, and adapting notation to the spatial setting, α would represent a common (across sites) base measure while α_s would represent the base measure at the site s. Making use of an already taken symbol to represent a different quantity can only lead to confusion and make reading the literature more difficult for those entering the area.

Acknowledgement

REPLY TO THE DISCUSSION

We appreciate the positive comments from Hjort and MacEachern and agree that, because there are so many ways to bridge spatial process modelling and Bayesian nonparametric modelling, we can expect to see growth in various directions in the near future.

A key point that we feel needs to be made in this rejoinder is the distinction in perspective and perhaps in specification, between MacEachern's DDP's and our SDP's. The issue centers on concern with joint or marginal models.

MacEachern's DDP in the spatial setting specifies the distribution of $\theta(s)$ as F_s, i.e., a collection of random distributions indexed by spatial location. Focusing on the random distributions F_{s_1} and F_{s_2}, we have $v_l = \begin{pmatrix} v_l(s_1) \\ v_l(s_2) \end{pmatrix}, l = 1, 2, ..$ $\Leftrightarrow p_l = \begin{pmatrix} p_l(s_1) \\ p_l(s_2) \end{pmatrix}, l = 1, 2, ...$ and $\{\theta_l^*\}$ where $\theta_l^* = \begin{pmatrix} \theta_l^*(s_1) \\ \theta_l^*(s_2) \end{pmatrix}$ and $F_{s_j} \Leftrightarrow \{p_l(s_j)\}, \{\theta_l^*(s_j)\}, j = 1, 2$. Immediately, we can see that F_{s_1} and F_{s2} come from a DP. If the components of v_l and θ_l^* are dependent then F_{s_1} and F_{s_2} are dependent. But what about the joint distribution of $\theta(s_1)$ and $\theta(s_2)$? If we assume conditional independence given F_{s_1} and F_{s_2}, as in, e.g., Cifarelli and Regazzini (1978), we have $P(\theta(s_1) = \theta_l^*(s_1), \theta(s_2) = \theta_{l'}^*(s_2)) = p_l(s_1)p_{l'}(s_2)$ with marginal dependence through mixing. Introducing dependence for the v's and θ^*'s across $s, s \in D$ is straightforward, e.g., suppose we have $v_l(s) = G^{-1}(\Phi(Z_l(s)))$ where G^{-1} is the inverse Beta$(1, \alpha)$ c.d.f., $Z_{l,D}$ are iid realizations from a mean 0 GP and, independently, the $\theta_{l,D}^*$'s are iid realizations from a GP. Indeed, this strategy is also mentioned by Hjort along with a different construction using transformation of a Gaussian process to gamma variables.

Amplifying MacEachern's subsequent suggestion, suppose we introduce a usual stickbreaking, $\{v_l\} \Leftrightarrow \{p_l\}$. Additionally, Z_D a mean 0, variance 1 GP realization yields $U(s) = \Phi(Z(s))$.

$$
\begin{aligned}
P(\theta(s_1) &= \theta_l^*(s_1), \theta(s_2) = \theta_{l'}^*(s_2)) \\
&= P(U(s_1) \in (\textstyle\sum_{j=1}^{l-1} p_j, \sum_{j=1}^{l} p_j), U(s_2) \in (\sum_{j=1}^{l'-1} p_j, \sum_{j=1}^{l'} p_j)) \\
&= P(Z(s_1) \in (\Phi^{-1}(\textstyle\sum_{j=1}^{l-1} p_j), \Phi^{-1}(\sum_{j=1}^{l} p_j)), Z(s_2) \\
&\qquad\qquad\qquad \in (\Phi^{-1}(\textstyle\sum_{j=1}^{l'-1} p_j), \Phi^{-1}(\sum_{j=1}^{l'} p_j)) \\
&\equiv p_{l,l'}(s_1, s_2).
\end{aligned}
$$

Evidently, we are constructing joint distributions as opposed to adopting the conditional independence assumption above. However, it is still the case that $F(\theta(s))$ comes from a DP.

Then, we come to the GSDP. Here, we introduce iid copies $\{Z_{l,D}\}$ of a GP with mean μ_D to create

$$
p_{l,l'}(s_1, s_2) = P(Z_j(s_1) < 0, j < l, Z_l(s_1) > 0, Z_j(s_2) < 0, j < l', Z_{l'}(s_2) > 0).
$$

We need μ_D to be random in order that the p's are and if $\mu(s)$ varies with s, $p_l(s)$ varies with s. Again, we are constructing joint distributions.

Finally, the GSDP-k introduces $Z_D \Leftrightarrow U_D = \Phi(Z_D)$. Now, $l, l' = 1, 2, ...k$ and

$$
\{p_{l,l'}(s_1, s_2\} \sim Dir(\{a_{l,l'}(s_1, s_2)\})
$$

with

$$
\begin{aligned}
a_{l,l'}(s_1, s_2) &= P(U(s_1) \in (\tfrac{l-1}{k}, \tfrac{l}{k}), U(s_2) \in (\tfrac{l'-1}{k}, \tfrac{l'}{k})) \\
&= P(Z(s_1) \in (\Phi^{-1}(\tfrac{l-1}{k}), \Phi^{-1}(\tfrac{l}{k})), Z(s_2) \in (\Phi^{-1}(\tfrac{l'-1}{k}), \Phi^{-1}(\tfrac{l'}{k}))).
\end{aligned}
$$

We see that automatically $\{p_l(s)\}$ varies with s by the random selection from the Dirichlet. Again, by definition, we are constructing joint probabilities for surface selection, hence joint distributions.

Next, we take up Hjort's discussion around his expression (2). Caution is needed here; (2) provides the joint distribution for a set of n locations under a GP but this is, in fact, a property of the GP. In general, for a given collection of locations, starting with i.i.d. random variables from some distribution, say f, using (2) produces a joint distribution. However, outside of the normal (or mixtures of normals), such transformation need not uniquely determine finite dimensional distributions. Hence, in general, we can not use (2) to define a process.

One aspect MacEachern helps to illuminate is the nature of the expected surface selection. However, we confess to being unclear as to his concern about "moving to adjacent components" with the $U(\cdot)$ approach. The $Z(\cdot)$ approach will tend to yield similar adjacent selection with regard to the θ_l^*'s. Moreover, with the independence of the θ^*'s across l, why is this a worry?

Hjort raises some general issues regarding spatial process modelling (that are apart from the nonparametric aspect). First, he questions the need for the spatial component in the model. Here, he joins the ongoing debate regarding the mean-covariance trade-off - the size of the mean specification will, for a given dataset, determine the need for spatial random effects. Next, he expresses concern about prior assumptions. Here, we face the usual issues of parameter independence, choice

of parametrization, identifiability of parameters, etc. Then, he raises the matter of smoothing away potentially interesting extremes in the spatial surface. Here, there are, in fact, two issues. One is the inherent smoothness associated with spatial process realizations under customary covariance functions and its impact with regard to spatial interpolation. The second is how, in particular, a Bayesian implementation of such interpolation imposes smoothness. Lastly, the dual role for the precision parameter in the DP is important to note. Here, we would see the issue as being distinguished by whether we mix or not, whether we introduce random effects or not. Without mixing, as in say customary Bayesian bioassay examples (see, e.g., Gelfand and Kuo, 1991), we would be looking at the DP through Ferguson's original perspective whence the precision parameter really is intended to reflect proximity of the random DP realization to some baseline CDF. Under mixing, interest turns to the random effects which, seen through the stickbreaking perspective, encourage clustering of individual observations into common populations. Furthermore, we can enrich the stickbreaking specification beyond a single precision parameter as we alluded to in generalizations above our expression (2) and also in Section 5.5.

Finally, we thank the discussants for their thoughtful input. Practical use of spatial and spatio-temporal models requires care in implementation and interpretation. Adding a nonparametric aspect certainly exacerbates this requirement.

ADDITIONAL REFERENCES IN THE DISCUSSION

Gamerman, D., Salazar, E. and Reis, E. (2007). Dynamic Gaussian process priors, with applications to the analysis of space-time data. *In this volume.*

Gelfand, A. E. and Kuo, L. (1991) Nonparametric Bayesian bioassay including ordered polytomous response. *Biometrika* **78**, 657–666.

Hjort, N. L. (2003). Topics in nonparametric Bayesian statistics . *Highly Structured Stochastic Systems* (P. Green, N.L. Hjort, S. Richardson, eds.) Oxford: University Press455–478, (with discussion).

Hjort, N. L. and Ongaro, A. (2006). On the distribution of random Dirichlet jumps. *Metron* **64**, 61–92.

Hjort, N. L, Dahl, F. A. and Steinbakk, G. H. (2006). Post-processing posterior predictive p-values. *J. Amer. Statist. Assoc.* (to appear).

MacEachern, S. N. (1999). Dependent nonparametric processes. *ASA Proceedings of the Section on Bayesian Statistical Science.* Alexandria, VA: American Statistical Association, 50–55.

MacEachern, S. N., Kottas, A, and Gelfand, A. E. (2001). Spatial nonparametric Bayesian models *ASA Proceedings of the Joint Statistical Meetings.* Alexandria, VA: American Statistical Association.

Mikosch, T. (2007). Copulas: tales and facts. *Extremes* (to appear).

Rousseau, J. (2007). Approximating interval hypothesis: *p*-values and Bayes factors. *In this volume.*