Smoothness Properties and Gradient Analysis Under Spatial Dirichlet Process Models

Michele Guindani · Alan E. Gelfand

Abstract When analyzing point-referenced spatial data, interest will be in the first order or global behavior of associated surfaces. However, in order to better understand these surfaces, we may also be interested in second order or local behavior, e.g., in the rate of change of a spatial surface at a given location in a given direction. In a Bayesian parametric setting, such smoothness analysis has been pursued by Banerjee and Gelfand (2003) and Banerjee et al. (2003). We study continuity and differentiability of random surfaces in the Bayesian nonparametric setting proposed by Gelfand et al. (2005), which is based on the formulation of a spatial Dirichlet process (SDP). We provide conditions under which the random surfaces sampled from a SDP are smooth. We also obtain complete distributional theory for the directional finite difference and derivative processes associated with those random surfaces. We present inference under a Bayesian framework and illustrate our methodology with a simulated dataset.

1 Introduction

In many applications dealing with the analysis of point-referenced spatial data, it is often desired to study local features of the processes under investigation like smoothness and gradient behaviors at given locations and for given directions. For

M. Guindani
Department of Biostatistics and Applied Mathematics,
UT MD Anderson Cancer Care Center, Houston, TX 77030, USA
e-mail: Mguindani@mdanderson.org

A. E. Gelfand (✉)
Institute of Statistics and Decision Sciences, Duke University,
Durham, NC 27708-0251, USA
e-mail: alan@stat.duke.edu

example, in modelling for real estate data, land values are predicted to vary with the distance from a Central Business District (CBD)(see Majumdar et al., 2004). However, at any particular location, the gradient in land values may depend upon direction, e.g., perhaps steeper in the direction toward the CBD, perhaps more flat in the directional orthogonal to this. In digital-terrain models, estimation of slopes is crucial for creating digital representations of topographic surfaces that are widely used in earth and environmental sciences. Other domains of application include meteorology for recognizing temperature or rainfall gradients and environmental monitoring for understanding pollution gradients.

Such models presume for a region of study D, a collection of random variables $Y(s)$, where s indexes the locations in D. The set $\{Y(s) : s \in D\}$ can be viewed as a randomly realized surface over the region. In practice, this surface is only observed at a finite set of locations, say $s_1, s_2, ..., s_n$. It is evident that detecting smoothness of process realizations from finitely sampled data is challenging, perhaps hopeless. Rather, it is more suitable to capture such smoothness in the specification of the process, motivated by mechanistic considerations associated with the process yielding the data.

Customarily, the model incorporates both a mean trend along with local adjustments to this trend arising as realizations from a random surface (or surfaces). Of course, continuity and differentiability issues are handled differently in the two cases. While gradients for the nonspatial component of the mean are obtained through ordinary multivariable calculus, smoothness of random surfaces needs an entirely stochastic framework.

In this regard, Kent (1989) provides formal discussion of the notion of almost sure (a.s.) continuity for stationary random fields. Building upon the existing theory for stochastic processes on the real line (see Stein, 1999), Banerjee and Gelfand (2003) examine mean square continuity and mean square differentiability when D is a subset of the d-dimensional Euclidean space ($d \geq 2$). All notions can be characterized through the process covariance structure and can be formalized in the context of univariate and multivariate processes. Banerjee et al. (2003) consider the subject from a modelling perspective, introducing directional finite difference and directional derivative processes in the case of stationary Gaussian spatial processes (GP). They provide distributional details for individual and multiple gradients and show how to implement full inference with regard to rates of change in a Bayesian framework.

However, in many cases, both the Gaussian and stationarity assumptions will be viewed as inappropriate. Distribution of real data is often clearly non-Gaussian, e.g., it exhibits heavy tail behavior or multimodal distributions. The problem has been overcome in different ways in the literature, for example assuming that the random field of interest is the result of an unknown transformation of a GP (De Oliveira et al., 1997) or by means of generalized linear spatial models (Diggle et al., 1998). Palacios and Steel (2004) accommodate non-Gaussian behavior by means of scale mixing of the spatial components in a spatial random effects model. On the other hand, 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 and Smith (2001) but both are still within the setting of GP's. The fundamental paper of Sampson and Guttorp (1992) introduces a nonparametric specification for the covariance function, as does

followup effort by Damian et al. (2001) and Schmidt and O'Hagan (2003) but again, all of this work employs a GP in the likelihood.

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 semi-parametric approach that, for fitting purposes, relies on well-known results and algorithms developed for Dirichlet process (DP) mixing. See, among others, Escobar and West (1995) and MacEachern and Müller (1998).

Our contribution is to provide conditions under which the random surfaces sampled from a SDP are smooth. As might be expected, such conditions are related to the behavior of the base spatial process. We also consider the gradient processes associated with those random surfaces, obtaining complete distributional theory results. In particular, we show that the directional finite difference and derivative processes are themselves samples from a SDP, whose base measure is the distribution of the corresponding gradient for the original base stochastic process. The use of nonparametric specifications typically requires the availability of a set of replicates, as in the foregoing work of Sampson and Guttorp (1992), Damian et al. (2001) and Schmidt and O'Hagan (2003). More precisely, as noted in Gelfand et al. (2005), replication is needed in order to learn about the unknown distribution function of the spatial component in a nonparametric approach. Without replicates, we would fall back on a conventional parametric specification. Well-known results in the literature ensure consistency of DP mixtures (see Ghosal et al., 1999). Moreover, we can embed the replications within a dynamic model as in Duan et al. (2005) to remove the independence assumption. In the sequel we work with independent replications of the spatial surface, employing a simulated dataset in order that we can validate our performance.

Gradient analysis of the sort we undertake, but confined to Gaussian processes, has been presented in the context of temperature surfaces in Banerjee and Gelfand (2005) and in the context of land value surfaces in Majumdar et al. (2004). Using annual data as replications, in both cases our nonparametric SDP approach could be implemented.

The format of the paper is as follows. A summary of basic definitions and results on smoothness of spatial processes is presented in Section 2.1, while Section 2.2 reviews the spatial Dirichlet process model as developed in Gelfand et al. (2005). In Section 3, we discuss some theoretical results on continuity and differentiability of random surfaces sampled from a SDP. Distribution theory of the gradient processes is provided in Section 4. In particular, we show that directional finite differences and derivatives are still samples from a SDP. In Section 5, the SDP is used to model the spatial component in a random spatial effect model, as in Gelfand et al. (2005). There we also discuss model fitting and inferential issues related to gradients. Section 6 presents the computational strategy for fitting such models in a Bayesian setting. Finally, Section 7 offers an example, intended to reveal the relevant features of our model with respect to competing parametric ones. Section 8 concludes with a summary and discussion.

2 A Review

2.1 Smoothness Properties of Spatial Processes

In this section, we review smoothness properties for spatial processes. For simplicity, we assume that the topology on the space where the random field is defined is induced by the usual Euclidean norm. However, the results extend to an arbitrary Banach space. In our investigation of smoothness properties we look at two types of continuity for process realizations, continuity in the L_2 sense and almost sure (a.s.) continuity. Unless otherwise noted, we assume the processes to have 0 mean and finite second-order moments.

DEFINITION *A process* $\{Y(s), s \in \mathbb{R}^d\}$ *is mean square continuous (also said, L_2-continuous) at* s_0 *if* $\lim_{||s-s_0||\to 0} E[Y(s) - Y(s_0)]^2 = 0$. *We will denote mean square continuity at* s_0 *as* $Y(s)\xrightarrow{L_2}Y(s_0)$.

DEFINITION *A real valued process* $\{Y(s), s \in \mathbb{R}^d\}$ *is almost surely continuous at* s_0 *if* $Y(s) \to Y(s_0)$ *with probability one as* $||s - s_0|| \to 0$. *If the process is almost surely continuous for every* $s_0 \in \mathbb{R}^d$ *then the process is said to have continuous realizations.*

In general, one form of continuity does not imply the other since one form of convergence does not imply the other. However, if $Y(s)$ is a bounded process then a.s. continuity implies L_2 continuity. Of course, each implies that $Y(s)p\to Y(s_0)$.

Notice that such definitions apply to an arbitrary (possibly nonstationary) stochastic process. However, Kent (1989) has provided sufficient conditions to ensure a.s. continuity of stationary random fields. Specifically, let $\{Y(s), s \in \mathbb{R}^d\}$ be a real valued (weakly) stationary random field. Kent proves that if the covariance function $Cov(Y(s), Y(s+\mathbf{h})) = K(\mathbf{h})$ is d-times continuously differentiable with respect to $\mathbf{h}$ and $K_d(\mathbf{h}) = K(\mathbf{h}) - P_d(\mathbf{h})$, where $P_d(\mathbf{h})$, the Taylor polynomial of degree d for $K(\mathbf{h})$ about $\mathbf{h} = \mathbf{0}$, satisfies the following condition

$$|K_d(\mathbf{h})| = O\left(||\mathbf{h}||^{d+\beta}\right) \quad \text{as } ||\mathbf{h}|| \to 0 \tag{1}$$

for some $\beta > 0$, then there exists a version of the random field $\{Y(s), s \in \mathbb{R}^d\}$ with continuous realizations. In particular, the previous condition is true whenever $K(\mathbf{h})$ is $d+1$-times differentiable.

For $Y(s)$ a weakly stationary process with covariance function $K(\mathbf{h})$, where $\mathbf{h} = s - s_0$, mean square continuity reduces to whether $\lim_{||h||\to 0} 2[K(\mathbf{0}) - K(\mathbf{h})] = 0$, so that $Y(s)$ is mean square continuous if and only if $K(\mathbf{h})$ is continuous at the origin. Notice that if $K(\mathbf{h})$ is continuous at the origin, then it is continuous everywhere. As Banerjee and Gelfand (2003) note, if $Y(s)$ satisfies Kent's sufficient condition for a.s. continuity, $K(\mathbf{h})$ is d-times differentiable, therefore continuous at 0. It follows that processes satisfying Kent condition are also mean square continuous.

Following Banerjee and Gelfand (2003) and Banerjee et al. (2003), mean square differentiability is motivated by the analogous definition of total differentiability of a function in $\mathbb{R}^d$ in multivariate calculus.

DEFINITION *A process* $\{Y(s), s \in \mathbb{R}^d\}$ *is said to be mean square differentiable at* s_0 *if there exists a vector* $\nabla_Y(s_0)$ *such that, for any scalar* h *and any unit vector* $\mathbf{u} \in \mathbb{R}^d (||\mathbf{u}|| = 1)$,

$$Y(s_0 + h\mathbf{u}) = Y(s_0) + h\mathbf{u}^T \nabla_Y(s_0) + r(s_0, h\mathbf{u}), \qquad (2)$$

where $r(s_0, h\mathbf{u}) \to 0$ *in the* L^2 *sense as* $h \to 0$.

This definition ensures that if $Y(s)$ is a mean square differentiable process on $\mathbb{R}^d$, then it is mean square continuous as well. This result follows directly from (2), since $r(s_0, h\mathbf{u}) \to 0$ in the L^2 sense as $h \to 0$ and $\mathbf{u}^T \nabla_{\mathbf{Y}}(s_0)$ is a constant with respect to h.

For $d = 1$, Stein (1999) shows that a stationary process $Y(s)$ on the real line is mean square differentiable if and only if its covariance function is twice differentiable and $K''(0)$ exists and is finite.

Next, we define a gradient process. We start defining the *directional finite difference process* $Y_{\mathbf{u},h}(s)$ at scale h in direction $\mathbf{u}$ as

$$Y_{\mathbf{u},h}(s) = \frac{Y(s + h\mathbf{u}) - Y(s)}{h}.$$

Finite difference processes measure the rate of change of a process in a given direction $\mathbf{u}$ and at a certain scale h. They can be usefully employed whenever scale is of importance.

Note that if $E(Y(s)) = 0$ for all $s \in \mathbb{R}^d$, then also $E(Y_{\mathbf{u},h}(s)) = 0$. Let $C_{\mathbf{u}}^{(h)}(s, s')$ denote the covariance function associated with the process $Y_{\mathbf{u},h}(s)$ and let $\Delta = s - s'$ denote the separation vector. Then, if $Y(s)$ is stationary

$$C_{\mathbf{u}}^{(h)}(s, s') = \frac{2K(\Delta) - K(\Delta + h\mathbf{u}) - K(\Delta - h\mathbf{u})}{h^2}. \qquad (3)$$

In particular, for $\Delta = 0$, $Var(Y_{\mathbf{u},h}(s)) = 2(K(0) - K(h\mathbf{u}))/h^2$.

If $Y(s)$ is isotropic,

$$C_{\mathbf{u}}^{(h)}(s, s') = \frac{2K(||\Delta||) - K(||\Delta + h\mathbf{u}||) - K(||\Delta - h\mathbf{u}||)}{h^2}. \qquad (4)$$

Hence, although the original process is isotropic, the finite difference process is only stationary. Also, $Var(Y_{\mathbf{u},h}(s)) = 2(K(0) - K(h))/h^2$.

The *directional derivative process* $D_{\mathbf{u}}Y(s)$ in direction $\mathbf{u}$ is obtained as the limit in the L_2 sense of the finite difference process $Y_{\mathbf{u},h}(s)$, as $h \to 0$, that is

$$\lim_{h \to 0} E[Y_{\mathbf{u},h}(s) - D_{\mathbf{u}}Y(s)]^2 = 0,$$

if the limit exists. If $E(Y(s)) = 0$ for all $s \in \mathbb{R}^d$, then also $E(D_{\mathbf{u}}Y(s)) = 0$. Let $C_{\mathbf{u}}(s, s')$ define the covariance function associated with the process $D_{\mathbf{u}}Y(s)$. Then, it can be shown that for any pair of locations $s, s' \in D$

$$C_{\mathbf{u}}(s, s') = \lim_{h \to 0} C_{\mathbf{u}}^{(h)}(s, s'), \qquad (5)$$

that is the covariance function of the process is obtained as the limit of the covariance function of the associated finite difference process, as $h \to 0$. In particular, if $Y(s)$ is a stationary process and its covariance function has continuous second order partial and mixed derivatives, then

$$C_{\mathbf{u}}(\Delta) = -\mathbf{u}^T \Omega(\Delta)\mathbf{u}, \tag{6}$$

where $(\Omega(\Delta))_{i,j} = \frac{\partial^2 K}{\partial \Delta_i \partial \Delta_j}$. Thus, $Var(D_{\mathbf{u}}Y(s)) = -\mathbf{u}^T \Omega(0)\mathbf{u}$.

If $Y(s)$ is isotropic, the directional derivative process will be stationary but not isotropic, since we can show that (6) becomes

$$C_{\mathbf{u}}(\Delta) = -\left\{ \left(1 - \frac{(\mathbf{u}^T \Delta)^2}{||\Delta||^2}\right) \frac{K'(||\Delta||)}{||\Delta||} + \frac{(\mathbf{u}^T \Delta)^2}{||\Delta||^2} K''||\Delta||) \right\}. \tag{7}$$

If $K''(\cdot)$ is continuous at 0 and we let $\Delta \to 0$ in (7), then $\lim_{\Delta \to 0} C_{\mathbf{u}}(s, s') = -K''(0)$. It follows that $C_{\mathbf{u}}(\Delta)$ is a valid covariance function continuous at zero. Therefore, the underlying directional derivative process is mean square continuous.

Notice that the above results rely on the existence of the second derivatives of the covariance function $K(\cdot)$ implying that directional derivatives processes do not exist for all $K(\cdot)$. For instance, the only function differentiable at 0 in the so called *power exponential family*, $K(||\Delta||) = \sigma^2 \exp(-\phi||\Delta||^\nu)$, $0 < \nu \le 2$, is the Gaussian ($\nu = 2$) covariance function. However, the Gaussian covariance function produces process realizations that are too smooth, in fact analytic, as can be seen from the fact that $K(\cdot)$ is infinitely differentiable (see Stein, 1999).

Instead, consider the Matèrn class (see Matern (1986) and Handcock and Stein (1993)), $K(||\mathbf{h}||) = \frac{2^{-\nu+1}\sigma^2}{\Gamma(\nu)}(\phi||\mathbf{h}||)^\nu \mathcal{H}_\nu(\phi||\mathbf{h}||)$, for $\phi > 0$, $\nu > 0$, and $\sigma > 0$, where $\mathcal{H}_\nu(\cdot)$ is the modified Bessel function of order ν (see Abramowitz and Stegun (1965), ch. 9). For $\nu = \frac{1}{2}$, the Matèrn covariance function reduces to the exponential case; for $\nu = \frac{3}{2}$, we get $K(||\mathbf{h}||) = \sigma^2 \exp(-\phi||\mathbf{h}||)(1 + \phi||\mathbf{h}||)$. The parameter ϕ is a decay parameter, while ν controls the degree of smoothness of the process. More precisely, Stein (1999) shows that a real process $Y(s)$ with Matèrn covariance function is m-times mean square differentiable if and only if $\nu > m$.

The existence of the directional derivative process in all directions $\mathbf{u}$ does not necessarily imply that the process is mean square differentiable (see Banerjee and Gelfand (2003) for further discussion and a counterexample). However, if $Y(s)$ is a mean square differentiable process in $\mathbb{R}^d$, i.e., (2) holds for every s in $\mathbb{R}^d$, then the directional derivative process $D_{\mathbf{u}}Y(s)$ exists for every $\mathbf{u}$ and $D_{\mathbf{u}}Y(s) = \mathbf{u}^T \nabla_{\mathbf{Y}}(s_0)$ a.s..

Note that if the unit vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_d$ form an orthonormal basis set for $\mathbb{R}^d$, any unit vector $\mathbf{u}$ in $\mathbb{R}^d$ can be written as $\mathbf{u} = \sum_{i=1}^d w_i \mathbf{e}_i$, with $w_i = \mathbf{u}^T \mathbf{e}_i$ and $\sum_{i=1}^d w_i^2 = 1$. It follows that

$$D_{\mathbf{u}}Y(s) = \mathbf{u}^T \nabla_Y(s) = \sum_{i=1}^d w_i \mathbf{e}_i^T \nabla_Y(s) = \sum_{i=1}^d w_i D_{\mathbf{e}_i}Y(s). \tag{8}$$

Hence, to study directional derivative processes in an arbitrary direction $\mathbf{u}$, we need only to know a basis set of directional derivative processes. These can always be taken to be defined by the coordinate axes, so that $\mathbf{e}_i$ is a $d \times 1$ vector with all 0's except for a 1 in the i-th row. In fact, with this basis, $\nabla_Y(s) = (D_{\mathbf{e}_1}Y(s), \ldots, D_{\mathbf{e}_n}Y(s))^T$. Notice that reduction to a basis set is not possible for finite difference processes, as it is evident from the presence of h in (2). From (8), it is also clear that

$D_{-\mathbf{u}} Y(s) = -D_{\mathbf{u}}(s)$. Applying the Cauchy-Schwarz inequality to (8), for every unit vector $\mathbf{u}$, $D_{\mathbf{u}}^2 Y(s) \le \sum_{i=1}^d D_{\mathbf{e}_1}^2 Y(s)$. Hence, $\sum_{i=1}^d D_{\mathbf{e}_1}^2 Y(s)$ is the maximum over all directions of $D_{\mathbf{u}}^2 Y(s)$.

2.2 The Spatial Dirichlet Process

We begin by developing a model for point referenced spatial data assumed to arise as a sample from a realization of a random field $\mathbf{Y}_D \equiv \{Y(s) : s \in D\}$, $D \subseteq R^d$. Denote by $s^{(n)} = (s_1, ..., s_n)$ the specific distinct locations in D where the observations are collected. When a Gaussian random field is assumed, a multivariate normal specification for the observed data results. In order to allow for deviations from this, arguably restrictive, assumption, we propose a nonparametric model for the random field with an associated induced model for the distribution of $(Y(s_1), ..., Y(s_n))$. We assume that we have available replicate observations at each location and therefore that the full data set consists of the collection of vectors $Y_t = (Y_t(s_1), ..., Y_t(s_n))'$, $t = 1,...,T$. In fact, we can accommodate imbalance or missingness in the $Y_t(s_i)$ through customary latent variable methods.

A frequent approach for specifying random distributions is the Dirichlet process (DP) (Ferguson, 1973, 1974). In particular, given the space Θ (equipped with a σ-field $\mathcal{B}$), let $DP(\nu G_0)$ denote the DP, 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 $\sum_{j=1}^\infty p_j \delta_{\theta_j^*}$, where δ_z denotes a point mass at z, $p_1 = q_1$, $p_j = q_j \prod_{r=1}^{j-1}(1 - q_r)$, $j = 2,3,...$, with $\{q_r, r = 1, 2, ...\}$ i.i.d. from Beta$(1, \nu)$ and independently $\{\theta_j^*, j = 1, 2, ...\}$ i.i.d. from G_0 (Sethuraman, 1994). In this notation θ_j^* is assumed to be scalar or perhaps vector valued, the latter case leading to a multivariate DP.

To model $\mathbf{Y}_D$, following Gelfand et al. (2005), one can conceptually extend θ_j^* to a realization of a random field by replacing it with $\boldsymbol{\theta}_{j,D}^* = \{\theta_j^*(s) : s \in D\}$. For instance, G_0 might be a stationary GP with each $\boldsymbol{\theta}_{j,D}^*$ being a realization from G_0, i.e., a surface over D. The resulting random distribution, G, for $\mathbf{Y}_D$ is denoted by $\sum_{j=1}^\infty p_j \delta_{\theta_{j,D}^*}$ and the construction will be referred to as a spatial Dirichlet process model. From now on, $Y_D|G \equiv \{Y(s), s \in D|G\}$ will explicitly denote a field whose distribution is a given realization of G. 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., 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 $G_s(\cdot) \equiv G(Y(s))$ for each s, hence the set $\mathcal{G}_D \equiv \{G(Y(s)) : s \in D\}$. MacEachern (2000, Theorem 3.1) provides sufficient conditions such that $\mathcal{G}_D$ will be a DDP.

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

$$Cov(Y(s_i), Y(s_j) \mid G) = \sum p_l \theta_l^*(s_i) \theta_l^*(s_j) - \left\{\sum p_l \theta_l^*(s_i)\right\}\left\{\sum p_l \theta_l^*(s_j)\right\}. \quad (9)$$

Hence, the random process G has heterogeneous variance and is nonstationary. Marginalizing over G simplifies the above expressions. For example, assume a mean zero stationary GP for G_0 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 $E(Y(s)) = 0$, $Var(Y(s)) = \sigma^2$ and $Cov(Y(s_i), Y(s_j)) = \sigma^2 \rho_\phi(s_i - s_j)$. Though G is centered around a stationary process with constant variance, it has nonconstant variance and is nonstationary.

3 Smoothness Properties of the Spatial Dirichlet Process

In this section, we consider the random surfaces arising as samples from a SDP and investigate their smoothness properties. In particular, their continuity is examined in Section 3.1, while we discuss differentiability in Section 3.2.

3.1 Almost Sure and Mean Square Continuity of Samples from a SDP

In Section 2.1, the smoothness of the random surface Y_D is studied on the basis of a given distributional assumption, in fact a given covariance function. Instead, if we assume a SDP model (or any other nonparametric specification), the distribution G of Y_D is itself random. Let $\mathcal{P}$ denote the distribution of the random G. In other words, $\mathcal{P} \equiv \mathcal{DP}(\nu \mathcal{G}_0)$. In a SDP, the observed data are assumed to come from a given realization of G. Therefore, the object of our interest is the smoothness of the surfaces sampled from the realized G, that is the behavior of $Y_D|G = \{Y(s), s \in D|G\}$. This is analogous to the usual parametric context, where we are interested in properties of a field whose distribution is indexed by a finite number of parameters. Evidently, we could ask whether this smoothness holds a.s. over the parameter space. With a finite dimensional space, this is actually immediate.

Turning to the SDP, if G admits the Sethuraman representation and G_0 is a.s. continuous then G will be as well. From a different perspective, let $C_{s_0} = \{Y_D : \lim_{\|s-s_0\| \to 0} Y(s) = Y(s_0)\}$, i.e., the set of all surfaces continuous at s_0. Then we say that Y_D is a.s. continuous at a point s_0 if $G(C_{s_0}) = 1$ with $\mathcal{P}$-probability one. This is not a probability statement given G but rather a statement over an infinite dimensional space. In particular, if we denote with S the set of random distribution functions chosen according to the Sethuraman's representation, we know that $P(S) = 1$ and in the previous definition we can consider all G in S. Analogously, we say that Y_D is mean square continuous at a point s_0 if $E[Y(s)^2|G] < \infty$ and $E[(Y(s) - Y(s_0))^2|G] \to 0$ as $\|s - s_0\| \to 0$ with $\mathcal{P}$-probability one. Again, to investigate the mean square continuity of samples from a SDP, we can limit the study to random surfaces drawn according to a G in S. With regard to a.s. continuity, we first show the connection between the behavior of the base random field and the surfaces sampled from a SDP in the following theorem, whose proof is provided in the Appendix.

PROPOSITION 1 *Any random field Y_D sampled from a SDP is a.s. continuous* IFF *the random field G_0 is a.s. continuous.*

Proposition 1 is actually an extension of a previous result in Gelfand et al. (2005) and MacEachern (2000). There it is shown that if G_0 is a.s. continuous, then $Y(s)$ converges weakly to $Y(s_0)$ as $||s - s_0|| \to 0$ with $\mathcal{P}$-probability one. Theorem 1 proves that Y_D is indeed a.s. continuous. Nevertheless, in the non-parametric context, weak convergence of the samples translates into a.s. convergence of $G(Y(s))$, the marginal distribution of $Y(s)$, to $G(Y(s_0))$, the marginal distribution of $Y(s_0)$. Therefore, the result in Gelfand et al. (2005) and MacEachern (2000) can be more conveniently restated as in the following proposition, for which, again, we provide a proof in the Appendix.

PROPOSITION 2 *In a SDP, if the base measure G_0 is a.s. continuous in s_0, then the random probability measure $G(Y(s_0))$ converges a.s. to $G(Y(s_0))$ as $||s - s_0|| \to 0$.*

Now, we turn to mean square continuity. One point is critical here. If we marginalize with respect to the unknown G, mean square continuity of Y_D follows easily from mean square continuity of the base process, since $E[(Y(s) - Y(s_0))^2] = E_{G_0}[(\theta_1^*(s) - \theta_1^*(s_0))^2]$. In fact, we are not interested in Y_D, which after marginalization is indeed an expected sample from $\mathcal{P}$, but in $Y_D|G$, the realized surface. Since $E[(Y(s) - Y(s_0))^2] = E\left\{ E\left[(Y(s) - Y(s_0))^2 | G\right] \right\}$, we could expect that mean square continuity of the base measure implies mean square continuity of $Y_D|G$. But, of course, this is not true, since L_1 convergence does not imply a.s. convergence, as would be required. Therefore, when we fix a distribution G, mean square continuity of the base measure is not enough to claim mean square continuity of the samples Y_D from G. This is not totally unexpected, since any G is a discrete probability measure with probability one, and therefore we expect its smoothness properties to depend on the smoothness of the $\theta_{j,D}^*$, $j = 1, 2, \ldots$ which define the support of the realized G. In fact, we can prove the following sufficient condition to ensure mean square continuity with respect to G (see the Appendix for a proof):

PROPOSITION 3 *Let G_0 be a separable process a.s. continuous on a compact $K \subset D$. Then, any random field Y_D sampled from a SDP is mean square continuous on K.*

For constant mean (*centered*) Gaussian processes, a.s. continuity on a compact is equivalent to a.s. boundedness (see Theorem 2.6.4 in Adler and Taylor, 2003), so that Proposition 3 can be restated in terms of Gaussian stationary fields with a.s. bounded realizations, e.g., such that $E(\sup_{s \in K} \theta^*(s)) < \infty$. Under such condition, we can apply the bounded convergence theorem to obtain mean square continuity of the base measure, a fortiori of the marginal Y_D, since $\lim_{||s-s_0|| \to 0} E[(\theta^*(s) - \theta^*(s_0))^2] = E[\lim_{||s-s_0|| \to 0} (\theta^*(s) - \theta^*(s_0))^2]$.

Let $G_s \equiv G(Y(s))$ and $G_{0,s} \equiv G_0(\theta^*(s))$ indicate, respectively, the random marginal distribution of $Y(s)$ given G and the marginal base measure in s. Also, consider the marginal model described by $Y(s)|G_s \sim G_s$ where $G_s \sim DP(\nu, G_{0,s})$. It is possible to prove that if $G_{0,s}$ converges weakly to G_{0,s_0}, then $G(Y(s))$ converges in distribution to $G(Y(s_0))$, i.e., $Y(s_0)|G_{s_0} \sim G_{s_0}$ and $G_{s_0} \sim DP(\nu, G_{0,s_0})$. Notice that this result involves only the marginal DP. Therefore, it remains valid for a greater class of models, e.g., the DDP models (see Section 2.2). Intuitively, we can see this

by the following argument. Recall that $E[Y(s)|G] = \sum_{j=1}^{\infty} p_j \theta_j^*(s)$ a.s.-$\mathcal{P}$ by Sethuraman's representation, where the θ_j^*'s are i.i.d. draws from G_0, chosen independently from p_j, $j = 1, 2, \ldots$. Since each $\theta_j^*(s)$ converges in distribution to $\theta_j^*(s_0)$, we can conclude that $E[Y(s)|G_s]$ converges in distribution to $E[Y(s_0)|G_{s_0}] = \sum_{j=1}^{\infty} p_j \theta_j^*(s_0)$, which is the mean of the probability mass function $G_{s_0}(\cdot) = \sum_{j=1}^{\infty} p_j \delta_{\theta_j^*}(s_0)(\cdot)$. Therefore, we can expect the limit $Y(s_0)$ to be a sample from a DP with smooth parameter ν and base measure given by G_{0,s_0}. In fact, the above is a particular case of a more general result, which is stated in the next proposition (see the Appendix for the proof).

PROPOSITION 4   *Let $Y(s)|G_s \sim G_s$, $G_s \sim DP(\nu G_{0,s})$, $s \in D$. Let $g(\cdot)$ be a real valued measurable function, integrable with respect to $G_{0,s}$ and denote with $G_{0,s}^g$ the distribution of $g(\theta^*(s))$ induced from $G_{0,s}$. Further, suppose that $G_{0,s}^g$ converges weakly to a distribution H_{0,s_0}, as $||s - s_0|| \to 0$, where s_0 is a point in D. Then, $g(Y(s))$ converges in distribution to a random variable $Z(s_0)$, which is a sample from a DP with parameter ν and base measure H_{0,s_0}.*

3.2 Mean Square Differentiability of Samples from a SDP

Now we turn attention to mean square differentiability of a process arising from a SDP. For any given G, unit vector $\mathbf{u}$ and scalar $h > 0$, we consider the finite differences $Y_{\mathbf{u},h}(s)$ and define the directional derivative $D_{\mathbf{u}}Y(s)$ as the L_2 limit of the finite difference process with respect to G, i.e.,

$$\lim_{h \to 0} E\left[\left(Y_{\mathbf{u},h}(s) - D_{\mathbf{u}}Y(s)\right)^2 | G\right] = 0, \tag{10}$$

if the limit exists. If $D_{\mathbf{u}}Y(s)$ exists for all $s \in D$, then we will denote the directional derivative process by $D_{\mathbf{u}}Y_D$. In particular, if $D_{\mathbf{u}}Y(s)$ is a linear function of $\mathbf{u}$, we say that $Y_D|G$ is mean square differentiable.

Again, if we marginalize with respect to the unknown G, the differentiability of Y_D follows immediately from that of the base measure. In fact, for any scalars $h_n, h_m > 0$, and any $s \in D$, $E\left[\left(Y_{\mathbf{u},h_n}(s) - Y_{\mathbf{u},h_m}(s)\right)^2\right] = E_{G_0}\left[\left(\theta_{\mathbf{u},h_n}^*(s) - \theta_{\mathbf{u},h_m}^*(s)\right)^2\right]$, that is any Cauchy sequence in L^2 with respect to G_0 is a Cauchy sequence with respect to $\mathcal{P}$, and the limits are the same. However, when we condition to a given realization G of the SDP, mean square differentiability of the base measure is not enough to conclude about mean square differentiability of $Y_D|G$. In fact, the latter relies on the analytical properties of the surfaces specifying the realized support of G, similarly to what discussed in Proposition 3 above.

PROPOSITION 5   *Let G_0 be a.s. continuously differentiable on a compact $K \subset D$. Then, any random field Y_D sampled from a SDP is mean square differentiable on $\dot{K} = int(K)$.*

Banerjee and Gelfand (2003) discuss conditions on the covariance function $K(s)$ of a stationary process in order that the directional derivative process $D_{\mathbf{u}}\theta_D^*$ has a.s. continuous realizations. Suppose $K(s) \in C^{d+2}$, meaning it is $d + 2$-times continuously differentiable. Let $P_d(s)$ denote the Taylor polynomial in s of degree d and $K_d(s)$

denote the remainder term, i.e., $K_d(s) = K(s) - P_d(s)$. Then, applying Kent's (1989) result to the covariance of the directional derivative process, they conclude that a sufficient condition for a stationary process to have a.s. continuous derivatives is that $K_d(s) = O(||s||^{d+2+\beta})$, for some $\beta > 0$.

In conclusion, even if the base process G_0 is mean square differentiable, it does not follow that $Y_D|G$ is, with respect to the observed realization of the random probability measure. However, if we marginalize with respect to G, the marginal process is of course mean square differentiable, since its distribution coincides with that of G_0. In the next section, we will apply Proposition 4 to show that the limit $D_\mathbf{u} Y_D$ is indeed obtained as a sample from a SDP whose base measure is the distribution of $D_\mathbf{u} \theta_D^*$, i.e., the directional derivative process associated with G_0. Therefore, we can say that the smoothness properties of the base measure are reflected in the samples, in the sense of the induced convergence of the random probability measures.

4 Some Distribution Theory

Let $Y_D \equiv \{Y(s), s \in D\}$ be a random field sampled from a $SDP(\nu G_0)$ and $Y_{\mathbf{u},h}(s)$ be the associated directional finite difference process. Then it is easy to prove that also $Y_{\mathbf{u},h}(s)$ is a sample from a SDP with same precision parameter ν and with base measure the distribution of the finite difference process $\theta_{\mathbf{u},h}^*(s)$, say $G_0^{\mathbf{u},h}$. In fact, for any pair $(s, s + h\mathbf{u}) \in D$ and any real t, consider

$$P\Big(Y_{\mathbf{u},h}(s) \le t \Big| G\Big) = P\Big(Y(s + h\mathbf{u}) \le Y(s) + th \Big| G\Big)$$
$$= \sum_{j=1}^{\infty} p_j \, I_{(-\infty, \theta_j^*(s)+th]}(\theta_j^*(s + h\mathbf{u})), \quad a.s.\text{-}\mathcal{P}$$

by the Sethuraman representation of the SDP. We can rewrite the indicator function as

$$I_{(-\infty, \theta_j^*(s)+th]}(\theta_j^*(s + h\mathbf{u})) = I_{(-\infty, t]}\Big(\theta_{\mathbf{u},h}^{*j}(s)\Big),$$

so to conclude that $Y_{\mathbf{u},h}(s)| G$ is a sample from $\sum_{j=1}^{\infty} p_j \, \delta_{\theta_{\mathbf{u},h}^{*j}(s)}(\cdot)$, i.e., from a SDP with precision parameter ν and base distribution $G_0^{\mathbf{u},h}$. We denote the random probability measure so defined as $G^{\mathbf{u},h}$ and notice that it is directly induced from G for any given $\mathbf{u}$ and h. Therefore, the necessary distribution theory for the directional finite difference process is obtained from the general theory of the SDP. In particular, its first and second moments are given by

$$E(Y_{\mathbf{u},h}(s)|G) = \sum_{j=1}^{+\infty} p_j \, \theta_{\mathbf{u},h}^{*j}(s),$$
$$E(Y_{\mathbf{u},h}^2(s)|G) = \sum_{j=1}^{+\infty} p_j \left(\theta_{\mathbf{u},h}^{*j}(s)\right)^2,$$

and for any pair of locations (s, s') in D,

$$Cov\left(Y_{\mathbf{u},h}(s), Y_{\mathbf{u},h}(s')|G\right) = \sum_{j=1}^{+\infty} p_j \theta_{\mathbf{u},h}^{*j}(s) \theta_{\mathbf{u},h}^{*j}(s') - \left\{\sum_{j=1}^{+\infty} p_j \theta_{\mathbf{u},h}^{*j}(s)\right\}$$

$$\times \left\{\sum_{j=1}^{+\infty} p_j \theta_{\mathbf{u},h}^{*j}(s')\right\},$$

while the distribution of the marginal process $Y_{\mathbf{u},h}(s)$ is the same as the distribution of the base process $\theta_{\mathbf{u},h}^{*}(s)$ (see Section 2.1).

Now consider the directional derivative process $D_{\mathbf{u}}Y(s)$ and suppose that G_0 admits a directional derivatives process for each $\mathbf{u}$. Let $G_{0,\mathbf{u}}'$ denote the distribution of the process $D_{\mathbf{u}}\theta^*(s)$. Then, we can apply Proposition 4 in order to prove that $D_{\mathbf{u}}Y(s)$ is a sample from a SDP with smooth parameter ν and base measure $G_{0,\mathbf{u}}'$. In symbols, $D_{\mathbf{u}}Y(s)|G_{\mathbf{u}}' \sim G_{\mathbf{u}}'$ and $G_{\mathbf{u}}' \sim SDP(\nu G_{0,\mathbf{u}}')$. In fact, we have shown that for any direction $\mathbf{u}$ and any fixed h, the finite difference process $Y_{\mathbf{u},h}(s)$ is a sample from a SDP, i.e., $Y_{\mathbf{u},h}(s)|G^{\mathbf{u},h} \sim G^{\mathbf{u},h}$, $G^{\mathbf{u},h} \sim SDP(\nu G_0^{\mathbf{u},h})$. Therefore, marginally at any given site s, $Y_{\mathbf{u},h}(s)$ is just a sample from a DP($\nu G_{0,s}^{\mathbf{u},h}$), where $G_{0,s}^{\mathbf{u},h}$ is the marginal distribution of $G_0^{\mathbf{u},h}$ in s. Since G_0 admits directional derivatives in all directions, $G_{0,s}^{\mathbf{u},h}$ converges weakly to $G_{0,\mathbf{u}}'$. Hence, the conditions of Proposition 4 are satisfied and the previous assertion is proved.

Intuitively, this can also be seen in another way. In fact, since

$$E\left(\left(Y_{\mathbf{u},h}(s) - D_{\mathbf{u}}Y(s)\right)^2|G\right) \geq \left(E\left(Y_{\mathbf{u},h}(s) - D_{\mathbf{u}}Y(s)|G\right)\right)^2,$$

from the mean square convergence of $\theta_{\mathbf{u},h}^{*}(s)$ to $D_{\mathbf{u}}\theta^*(s)$, it follows that $E\left(Y_{\mathbf{u},h}(s)|G\right)$, that is the mean of the DP for $Y_{\mathbf{u},h}(s)$, converges in L_2 (with respect to the Dirichlet Measure $\mathcal{P}$) to the random variable $E(D_{\mathbf{u}}Y(s)|G)$. But $E\left(Y_{\mathbf{u},h}(s)|G\right) = \sum_{j=1}^{\infty} p_j \theta_{\mathbf{u},h}^{*}(s)$ a.s.-$\mathcal{P}$ by Sethuraman's representation and the $\theta_{\mathbf{u},h}^{*}$'s are i.i.d. draws from $G_0^{\mathbf{u},h}$, chosen independently from p_j, $j = 1, 2, \ldots$. Then, $E(Y_{\mathbf{u},h}(s)|G)$ converges in distribution to $E(D_{\mathbf{u}}Y(s)|G) = \sum_{j=1}^{\infty} p_j D_{\mathbf{u}}\theta_j^*(s)$, that is the mean of the probability mass function $\sum_{j=1}^{\infty} p_j \delta_{D_{\mathbf{u}}\theta_j^*(s)}(\cdot)$, which is the almost sure representation of $G_{\mathbf{u}}'$. Therefore, it is immediate to guess that $G(Y_{\mathbf{u},h}(s))$ converges in distribution to $G_{\mathbf{u}}'(D_{\mathbf{u}}Y(s))$, and by the uniqueness of the limit, that $D_{\mathbf{u}}Y(s)$ is a sample from a SDP with smooth parameter ν and base measure $G_{0,\mathbf{u}}'$.

From the discussion above, it is easy to extract the first and second moments of the directional derivative process. In fact,

$$E(D_{\mathbf{u}}Y(s)|G_{\mathbf{u}}') = \sum_{j=1}^{\infty} p_j D_{\mathbf{u}}\theta_j^*(s),$$

$$E\left(D_{\mathbf{u}}^2 Y(s)|G_{\mathbf{u}}'\right) = \sum_{j=1}^{\infty} p_j D_{\mathbf{u}}^2 \theta_j^*(s),$$

and for any pair of locations $(s, s') \in D$, we have

$$Cov(D_{\mathbf{u}}Y(s), D_{\mathbf{u}}Y(s')|G_{\mathbf{u}}') = \sum_{j=1}^{\infty} p_j D_{\mathbf{u}}\theta_j^*(s) D_{\mathbf{u}}\theta_j^*(s'),$$

$$-\left\{\sum_{j=1}^{\infty} p_j D_{\mathbf{u}}\theta_j^*(s)\right\}\left\{\sum_{j=1}^{\infty} p_j D_{\mathbf{u}}\theta_j^*(s')\right\}.$$

Again, the distribution for the marginal process coincides with that of $D_{\mathbf{u}}\theta^*(s)$, which has been described in Section 2.1.

In particular, if G_0 is mean square differentiable, then $D_{\mathbf{u}}Y(s) = \mathbf{u}^T\nabla_Y(s)$, where $\nabla_Y(s)$ is a vector valued process, whose distribution is a realization from a SDP, defined for all borel sets A as

$$P(\nabla_Y(s) \in A) = \sum_{j=1}^{\infty} p_j \, \delta_{\nabla_{\theta_j^*(s)}}(A),$$

according to Sethuraman's representation. Here $\nabla_{\theta_j^*(s)} = (D_{e_1}\theta^*(s), \ldots, D_{e_d}\theta^*(s))$ is the vector of directional derivatives of G_0 with respect to an orthonormal basis set of directions $(\mathbf{e}_1, \ldots, \mathbf{e}_d)$.

In accordance to the discussion in Section 2.1, if the base measure is mean square differentiable, it is possible to study the behavior of $D_{\mathbf{u}}Y(s)|G'_{\mathbf{u}}$ in arbitrary directions by means of an orthonormal basis $(\mathbf{e}_1, \ldots, \mathbf{e}_d)$, and $D_{\mathbf{u}}Y(s) = \mathbf{u}^T\nabla_Y(s) = \sum_{i=1}^{d} w_i D_{\mathbf{e}_i}Y(s)$. For example, the first and second moments of the process can be obtained as a linear combination of a basis set of moments, that is

$$E(D_{\mathbf{u}}Y(s)|G'_{\mathbf{u}}) = \sum_{j=1}^{\infty} p_j \sum_{i=1}^{d} w_i D_{\mathbf{e}_i}(\theta_j^*(s)) = \sum_{i=1}^{d} w_i E(D_{\mathbf{e}_i}Y(s)|G'_{\mathbf{u}})$$

and

$$Cov(D_{\mathbf{u}}Y(s), D_{\mathbf{u}}Y(s')|G'_{\mathbf{u}}) = \sum_{i=1}^{d} w_i^2 \, Cov(D_{\mathbf{e}_i}Y(s), D_{\mathbf{e}_i}Y(s')|G'_{\mathbf{u}}).$$

5 Model Fitting and Inference

We work in $d = 2$ dimensional space and assume we observe T replicates of a random field $\{Y(s), s \in D\}$ at n locations $(s_1, \ldots, s_n)$. In particular, let $Y(s)$ arise as

$$Y_t(s) = \mu_t(s) + \theta_t(s) + \varepsilon_t(s). \tag{11}$$

The mean structure component can be either constant or, more frequently, a regression form $x_t(s)^T\beta$. The elements of the p-dimensional vector $x_t(s)$ can be functions of geographical coordinates to capture a trend surface (therefore it is constant across replicates). In addition, $x_t(s)$ can include covariates varying with t. For example, in a study of land values gradients, Majumdar et al. (2004) consider $x(s) = e^{-\|s-s^*\|}$ or $x(s) = 1/(a + \|s - s^*\|)^b$, for some reals a, b, according to the economic theory that prescribes a decline in land values as we move away from a central business district located at s^*. As an example of the second choice, in studying selling prices of single family homes, Banerjee et al. (2003) consider a vector of home specific covariates including its age, square feet of living area, other area, and number of bathrooms.

For now, we don't consider any dynamic evolution of the model so that the same model is assumed across the t's. In other words, let us denote with $Y_t = (Y_t(s_1), \ldots, Y_t(s_n))^T$ the vector of observed values at the n locations for each $t = 1, \ldots, T$. Correspondingly, let X_t be the $n \times p$ matrix whose i-th column is the vector $x_t(s_i) = (x_{t,1}(s_i), \ldots, x_{t,p}(s_i))^T$, $i = 1, \ldots, n$, and $\theta_t = (\theta_t(s_1), \ldots, \theta_t(s_n))^T$ be the vector of the spatial components, $t = 1, \ldots, T$. Then, we assume that Y_t given β and θ_t are drawn independently from a density $f(Y_t|X_t^T\beta + \theta_t, \tau^2)$, usually assumed to be

Gaussian. The vector of spatial effects θ_t is a sample from a SDP, such that $\theta_t|G^{(n)} \sim G^{(n)}$. Here $G^{(n)}$ is the prior induced from the SDP G. Therefore, $G^{(n)} \sim SDP(\nu G_0^{(n)})$, with $G_0^{(n)}$ being a multivariate normal with mean zero and covariance matrix $\sigma^2 H_n(\phi)$, where $(H_n(\phi))_{i,j} = \rho_\phi(s_i, s_j)$ is the correlation function, indexed by some vector of parameters ϕ.

Hence, we can specify the following semiparametric hierarchical model,

$$Y_t|\beta,\theta_t,\tau^2 \stackrel{ind.}{\sim} N_n(Y_t|X_t^T\beta + \theta_t, \tau^2 I_n), \quad t = 1,\ldots,T \tag{12}$$

$$\theta_t|G^{(n)} \stackrel{i.i.d.}{\sim} G^{(n)}, \quad t = 1,\ldots,T$$

$$G^{(n)}|\nu,\sigma^2,\phi \sim DP(\nu G_0^{(n)}), \quad G_0^{(n)}(\cdot|0_n, \sigma^2 H_n(\phi))$$

$$\beta,\tau^2 \sim N_p(\beta|\beta_0,\Sigma_\beta) \times IGamma(\tau^2|a_\tau, b_\tau)$$

$$\nu,\sigma^2,\phi \sim Gamma(\nu|a_\nu, b_\nu) \times IGamma(\sigma^2|a_\sigma, b_\sigma) \times [\phi],$$

where we placed appropriate conventional priors on the hyperparameters $\beta, \tau^2, \nu, \sigma^2, \phi$ and the prior on ϕ is denoted as $[\phi]$ by means of the simple brackets notation in Gelfand and Smith (1990). Notice that such prior depends on the specific form of $\rho_\phi(\cdot)$. Hereafter we consider ρ belonging to the Matèrn covariance class, with decay parameter ϕ and smoothness parameter $\nu \in (1,2)$, i.e., our base process is exactly once differentiable. Accordingly, a Gamma prior is considered for ϕ and a uniform on $(1,2)$ for ν.

Since the pure error term is a white noise, hence nowhere differentiable, inferential interest is on directional finite differences and derivatives of the mean process defined at the first level of the hierarchy. However, gradients for the mean surface can be estimated at one of the observed sites, or at a set of entirely new locations (Bayesian "kriging"). Also, we could be interested in the behavior of one of the available replicates or in prediction for a future realization of the process. Therefore, in the following discussion, we will comprise all possible cases by suppressing t and writing simply $m(s) = E(Y(s)|X,\beta,\theta) = x(s)^T\beta + \theta(s)$, where s can denote an element either in the original set of locations $(s_1,\ldots,s_n)$ or in a new set $(\tilde{s}_1,\ldots,\tilde{s}_m)$. Notice that, in the second case, we need to know the value of the covariates at these new locations. In practice, s may be a specified set of surfaces, e.g., distance to a fixed point or elevation. However, if they are not fully known, we will need to provide a stochastic model for them, e.g., that the covariate surfaces be viewed as a realization from a multivariate random field, i.e., the p-dimensional vector process $\{x(s), s \in D\}$.

Then, the finite difference and the directional derivative processes of $m(s)$ are, respectively, given by

$$m_{\mathbf{u},h}(s) = \frac{E(Y(s+\mathbf{u}h)|X,\beta,\theta) - E(Y(s)|X,\beta,\theta)}{h} \tag{13}$$

$$= x_{\mathbf{u},h}(s)^T\beta + \theta_{\mathbf{u},h}(s),$$

and

$$D_{\mathbf{u}}m(s) = D_{\mathbf{u}}E(Y(s)|X,\beta,\theta) = D_{\mathbf{u}}x(s)^T\beta + D_{\mathbf{u}}\theta(s), \tag{14}$$

for some unit vector $\mathbf{u}$ and scalar h. The latter result can be obtained as an L^2-limit for $h \to 0$ of $m_{\mathbf{u},h}(s)$. However, it can also be seen as a special case of a more general

result provided by Majumdar et al. (2004) for directional derivatives of functional forms of a random field $X(s)$. Let $Z(s) = g(X(s))$ for some arbitrary functional g. Then, Majumdar et al. prove that $Z(s)$ has directional derivative process given by $D_{\mathbf{u}}Z(s) = g'(X(s))D_{\mathbf{u}}X(s)$. In other words, this defines a simple chain rule for directional derivative processes. Of course, if we specify a covariate process, we need to take into account the distributional assumptions on $x(s)$.

Simulation based model fitting proceeds by marginalizing over the random mixing distribution, resulting in a finite dimensional parameter vector. Gibbs sampling for the posterior distribution $[\theta, \beta, \tau^2, \nu, \sigma^2, \phi \mid \text{data}]$, where $\theta = (\theta_1, ..., \theta_T)$ and data $= \{Y_1, ..., Y_T\}$ is carried over according to one of the standard Pólya urn based algorithms developed by Escobar (1994), Escobar and West (1995) and Bush and MacEachern (1996). Implementation specific for the semiparameteric hierarchical model (12) has been described in Gelfand et al. (2005) and will not be detailed here. We do note that the θ are updated using the approach of Bush and MacEachern (1996). Updating β and τ^2 does not involve the DP part of the model resulting in normal and inverse gamma full conditional distributions, respectively. Lastly, σ^2 also has an inverse gamma full conditional, ν is handled through a Metropolis step and ϕ is handled by discretizing the parameter space. The a.s. discreteness of the realizations from the Dirichlet process implies that there is a positive probability of clustering in the samples. Let us denote with T^* the number of distinct elements in θ and with $\theta^* = (\theta_1^*, \ldots, \theta_T^*)$ the vector that collects only the distinct θ_t's. Notice that we can switch back and forth from θ to θ^*, once we define a vector of labels $w = (w_1, \ldots, w_T)$, such that $w_t = j$ if and only if $\theta_t = \theta_j^*$, $t = 1, \ldots, T$. Then, (θ^*, w) is an equivalent representation of θ and posterior draws from $[\theta, \beta, \tau^2, \sigma^2, \phi, \nu, Y_t, t = 1, \ldots, T]$ are the same as posterior draws from $[\theta^*, w, T^*, \beta, \tau^2, \sigma^2, \phi, \nu, Y_t, t = 1, \ldots, T]$.

A key point to be made is that our proposed gradient analysis is a post-model fitting activity. Once we have obtained posterior samples, we can study gradients at any location, in any direction, as many as we wish, using one-for-one sampling with the posterior output. Hence, the consequential computational demand is in the model fitting. But again, as noted above, fitting models as in (12) is now fairly standard. We make the priors fairly noninformative. As long as the number of sampled locations, n, and the number of replicates, T, are not too small, the MCMC tends to be well-behaved and there is little prior sensitivity. Run times are roughly linear in T and quadratic in n. We have used both R code and C++ code, the latter being typically an order of magnitude faster.

As anticipated, interest is in prediction of the directional finite differences and derivatives of the mean process $m(s)$ at locations where the random field $\{Y(s), s \in D\}$ is not observed. Hereafter, we use $V_{\mathbf{u}}(s)$ to denote either $m_{\mathbf{u},h}(s)$ or $D_{\mathbf{u}}m(s)$. In obvious notation, we use $V_{\mathbf{u}}^X(s)$ and $V_{\mathbf{u}}^\theta(s)$ to denote the components in (13) and (14), writing $V_{\mathbf{u}}(s) = V_{\mathbf{u}}^X(s) + V_{\mathbf{u}}^\theta(s)$. If we assume independence between the processes for the covariates and the spatial component, we get

$$F_{V_{\mathbf{u}}}(y \mid \text{data}) = \int F_{V_{\mathbf{u}}^X}(y - z \mid \text{data}) \, dF_{V_{\mathbf{u}}^\theta}(z \mid \text{data}),$$

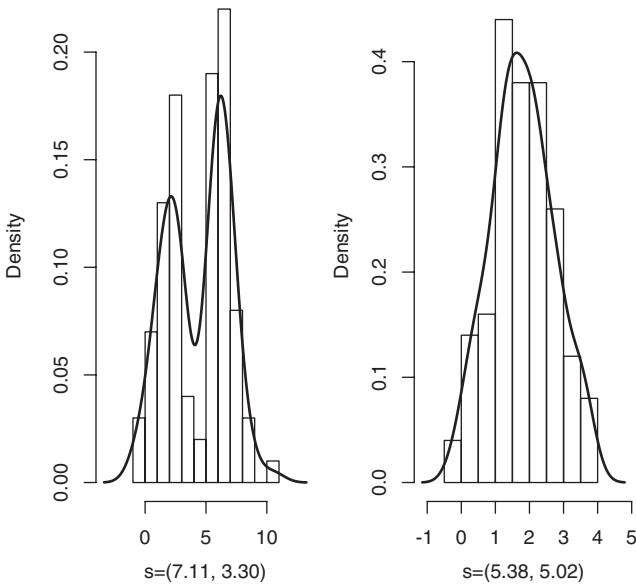where $F(\cdot)$ concisely denotes the distribution function of the process. Turning to densities,

$$f_{V_{\mathbf{u}}}(y \mid \text{data}) = \int f_{V_{\mathbf{u}}^X}(y - z \mid \text{data}) f_{V_{\mathbf{u}}^\theta}(z \mid \text{data}) \, dz,$$
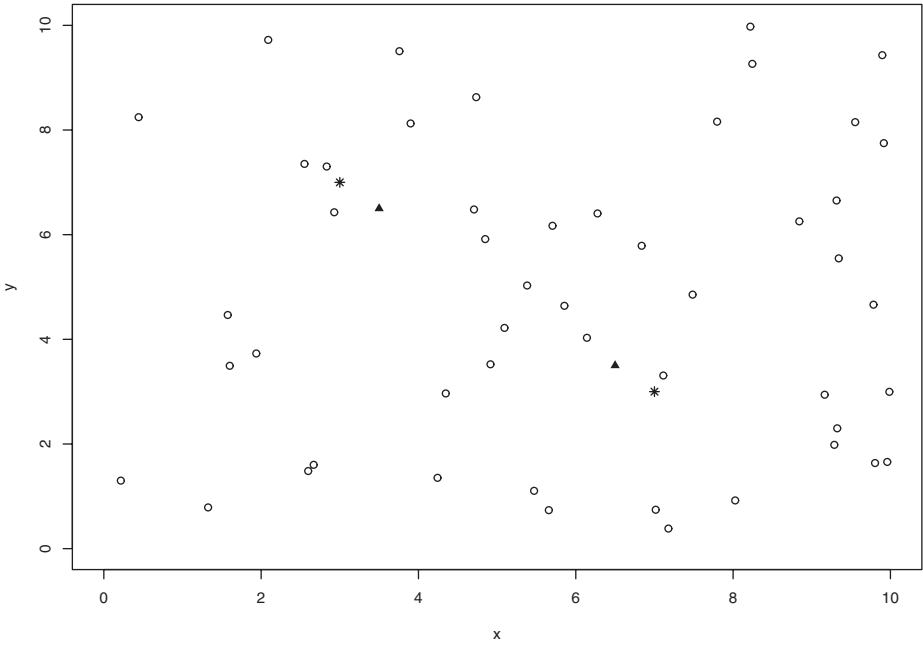
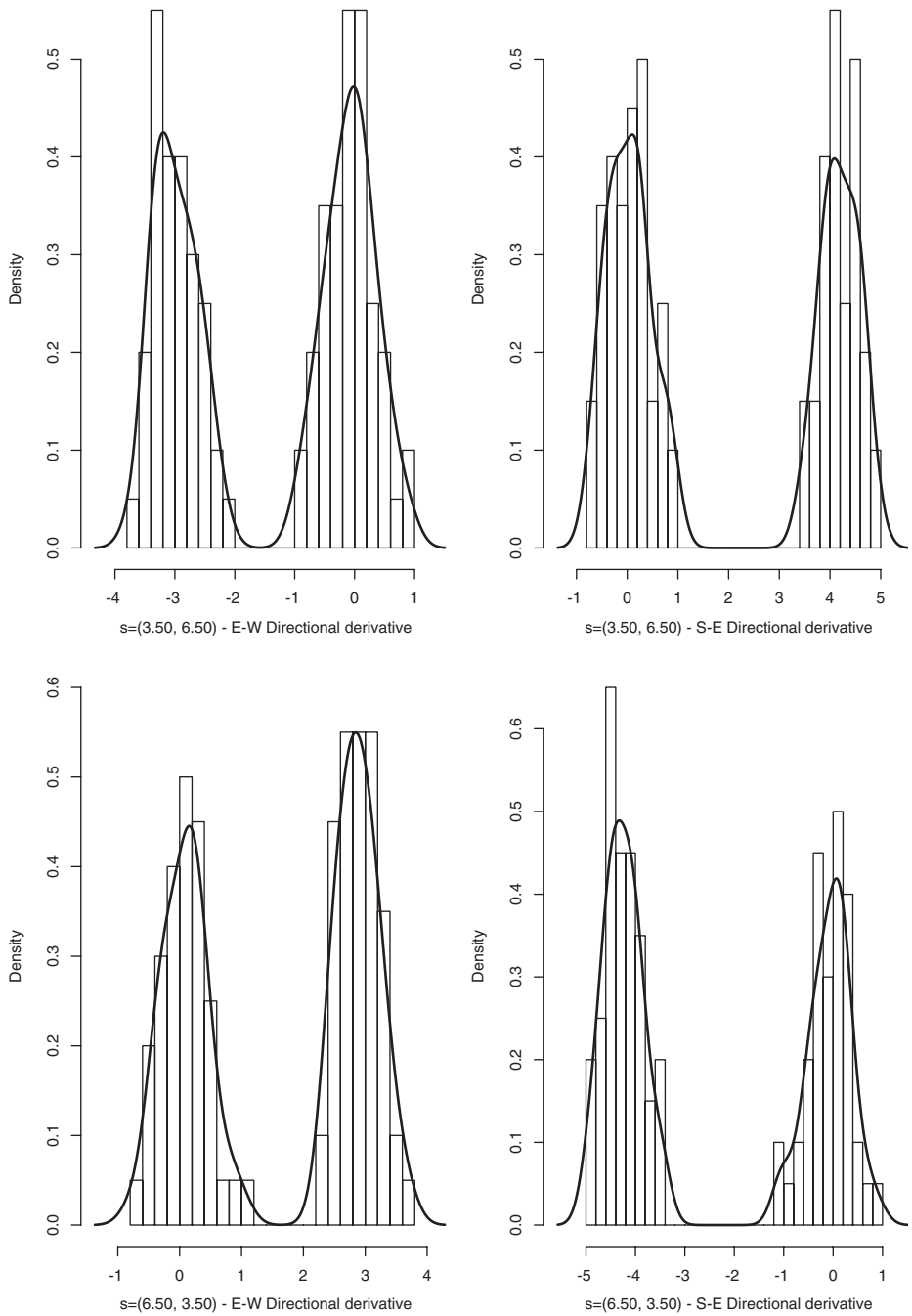**Fig. 1** Location of the 50 sites where the random field has been observed, together with the indication of the sources (✳) and prediction sites (▲)

Fig. 2 Histogram and estimated density in two sites, one close to a source (7.11,3.30), one distant (5.38, 5.02)

Fig. 3 Finite differences ($h = 0.01$) at angles 0 and 135, observed at the two locations $s*_1 = (3.5, 6.5)$ (above) and $s*_2 = (6.5, 3.5)$ (below)

from which it follows that we can get samples from the posterior predictive distribution of $V(s)$ via composition sampling. For instance, if $x(s)$ is fixed $V_{\mathbf{u}}^X(s)$ is a parametric function of β at each s, immediately sampled given posterior samples for the β's. Such samples may be added to associated samples from the posterior predictive of $V_{\mathbf{u}}^\theta(s)$ to obtain samples of $V_{\mathbf{u}}(s)$. Hereafter we concentrate only on inference over $V_{\mathbf{u}}^\theta(s)$, which is equivalent to assume a spatially constant mean structure in (12), so that we can simply consider $V_{\mathbf{u}}(s) = V_{\mathbf{u}}^\theta(s)$ for notational simplicity. In our simulation example, this facilitates isolating the contribution of the pure spatial effect to the pattern observed in the data.

6 Computational Issues

According to the prediction discussion in Gelfand et al. (2005), in principle we can study the behavior of the gradient process on each of the observed replicates or on a totally new predictive surface. However, in most applications it is natural to study the pattern of gradients for one of the available replicates. Such prediction is enhanced by borrowing strength across all replicates through the nonparametric specification and enables the possibility of clustering the gradients according to the spatial effect in place at the moment the observation was collected. Predicting gradients for a new unobserved surface would not appear to be of much interest. Apart from the mean structure, any patterns in such gradients would arise by chance, i.e., we would be seeing a random realization from the posterior gradient process.

Therefore, for $t = 1, \ldots, T$, let the vectors $V_{\mathbf{u},t} = (V_{\mathbf{u},t}(s_1), \ldots, V_{\mathbf{u},t}(s_n))$ and $\widetilde{V}_{\mathbf{u},t} = (V_{\mathbf{u},t}(\tilde{s}_1), \ldots, V_{\mathbf{u},t}(\tilde{s}_m))$ collect the values of the finite difference or directional derivative process respectively at old and new locations. We are interested in predicting $(V_{\mathbf{u},J}, \widetilde{V}_{\mathbf{u},J})$ given the data, for some $J = 1, \ldots, T$.

We recall the equivalence between θ and (θ^*, w) from Section 5. Hence, let $\tilde{\theta}_t = (\theta_t(\tilde{s}_1), \ldots, \theta_t(\tilde{s}_m))$ denote the vector of spatial component for new locations for replicate $t = 1, \ldots, T$. Then, $\tilde{\theta} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_T)$ and $\tilde{\theta}^* = (\tilde{\theta}_1, \ldots, \tilde{\theta}_{T^*})$ denote the vectors corresponding to θ and θ^* for the new locations. Notice that, by the nature of the SDP, each couple $(V_{\mathbf{u},t}, \widetilde{V}_{\mathbf{u},t})$ captures the gradient associated with the surface where $(\theta_t, \tilde{\theta}_t)$ belongs. Therefore, given the vector w of configuration indicators, it is possible to define vectors $(V_{\mathbf{u},t}^*, \widetilde{V}_{\mathbf{u},t}^*)$ corresponding to $(\theta_t^*, \tilde{\theta}_t^*)$, $t = 1, \ldots, T^*$, such that the joint predictive posterior distribution $[V_{\mathbf{u},t}, \widetilde{V}_{\mathbf{u},t}, t = 1, \ldots, T | \text{data}]$ can be rewritten as

$$[V_{\mathbf{u},t}, \widetilde{V}_{\mathbf{u},t}, t = 1, \ldots, T | \text{ data}] = \int [V_{\mathbf{u},t}, \widetilde{V}_{\mathbf{u},t}, t = 1, \ldots, T | \theta_t, \tilde{\theta}_t, t = 1, \ldots, T, \psi] \quad (15)$$

$$\times [\theta_t, \tilde{\theta}_t, t = 1, \ldots, T, \psi | \text{ data}]$$

$$= [V_{\mathbf{u},t}^*, \widetilde{V}_{\mathbf{u},t}^*, w, t = 1, \ldots, T^* | \text{ data}]$$

$$= \int \prod_{t=1}^{T^*} [V_{\mathbf{u},t}^*, \widetilde{V}_{\mathbf{u},t}^* | \theta_t^*, \tilde{\theta}_t^*, \sigma^2, \phi, \nu]$$

$$\times \prod_{t=1}^{T^*} [\tilde{\theta}_t^* | \theta_t^*, \sigma^2, \phi, \nu] \, [\theta^*, w, T^*, \psi | \text{ data}],$$

where $\psi = (\beta, \tau^2, \sigma^2, \phi, \nu)$ denotes the vector of parameters of the model. We can sample from (15) via composition sampling. In fact, $[\theta^*, w, T^*, \psi | \text{ data}]$ is the output
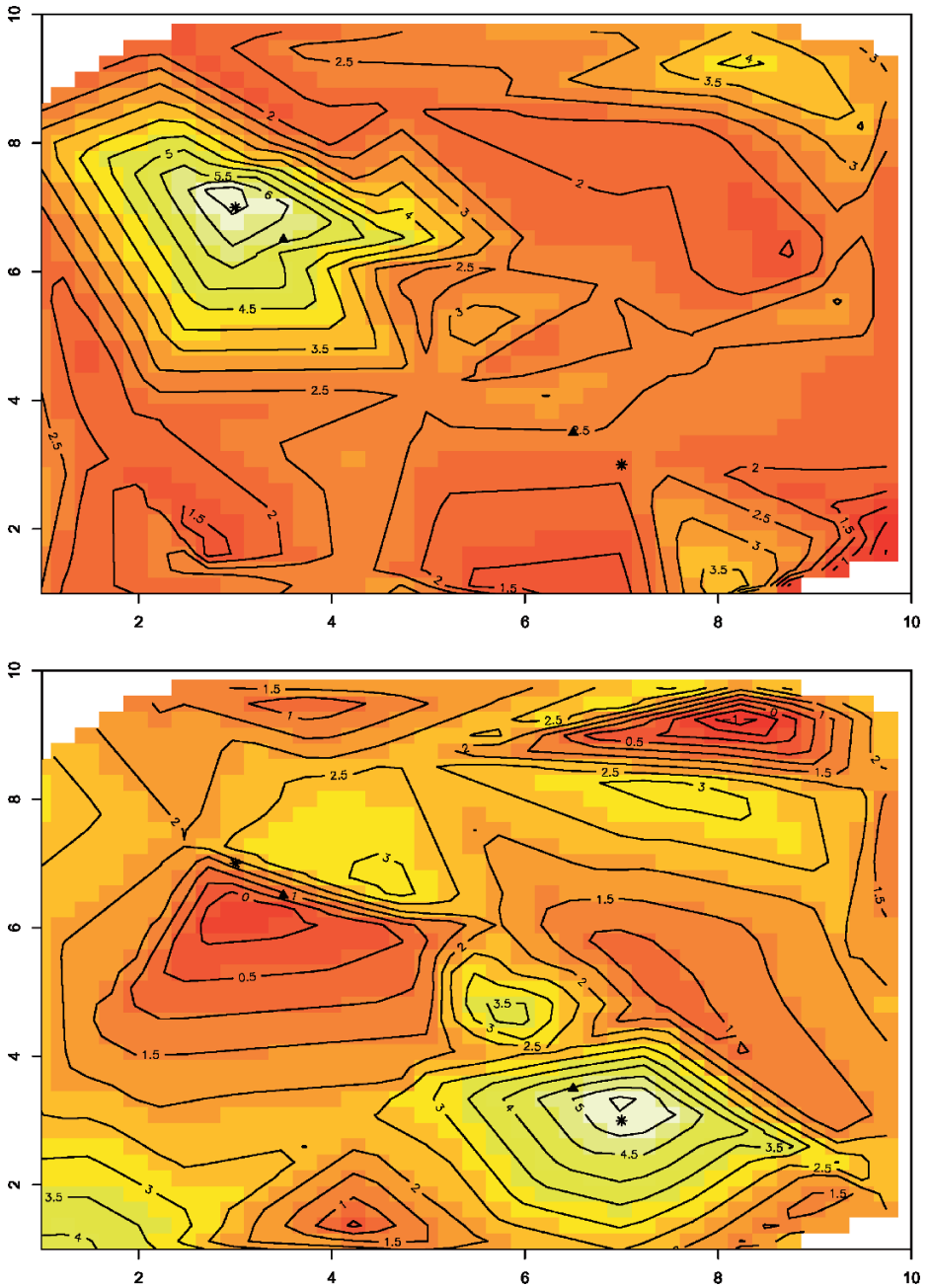
Fig. 4 Image plots with contour lines of data generated at the 99th (above) and 100th (below) replicates
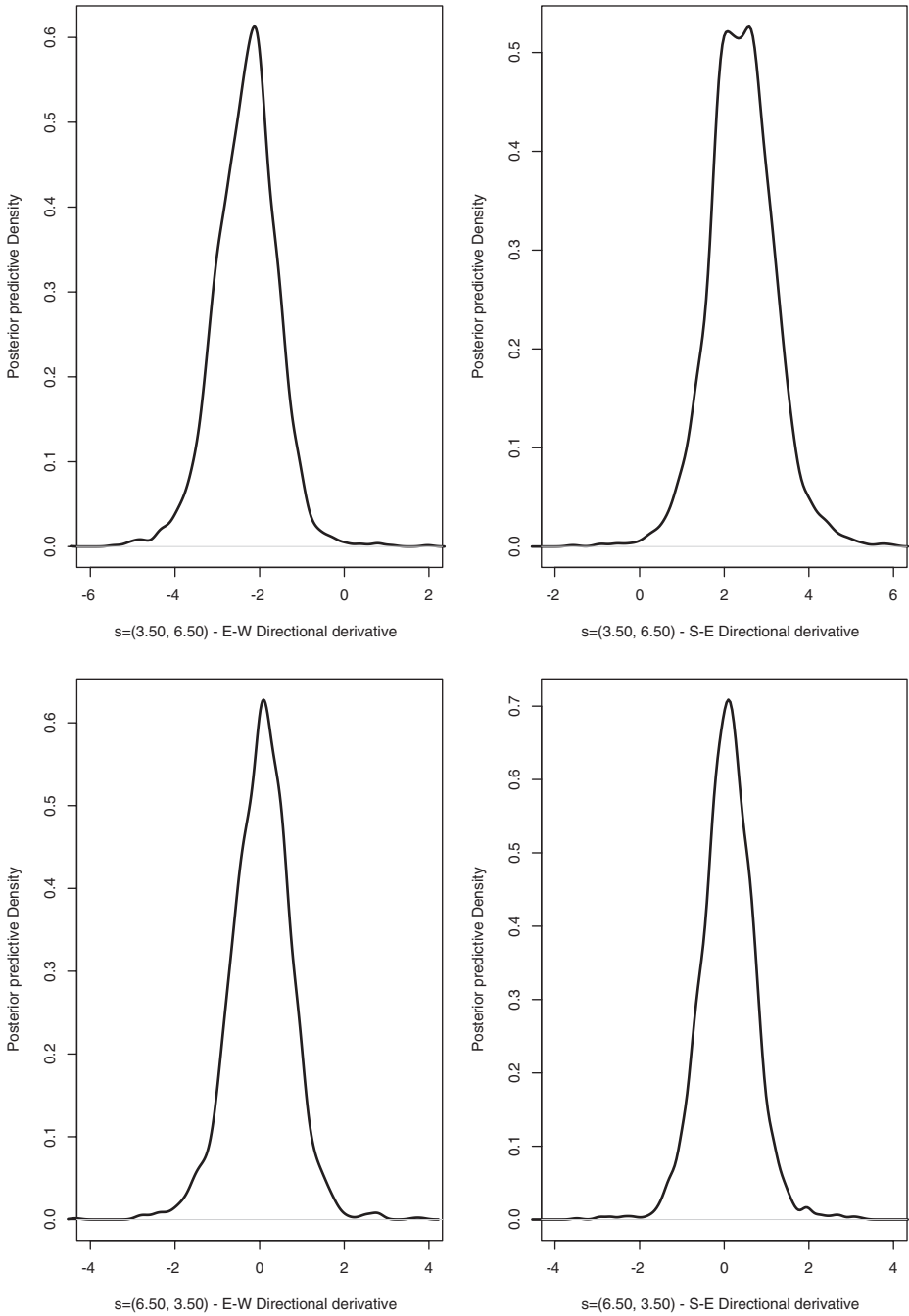
Fig. 5 Predictive posterior distribution of the directional derivatives at angles 0 and 135 and at sites $s_1^* = (3.5, 6.5)$ and $s_2^* = (6.5, 3.5)$ for the 99th replicate
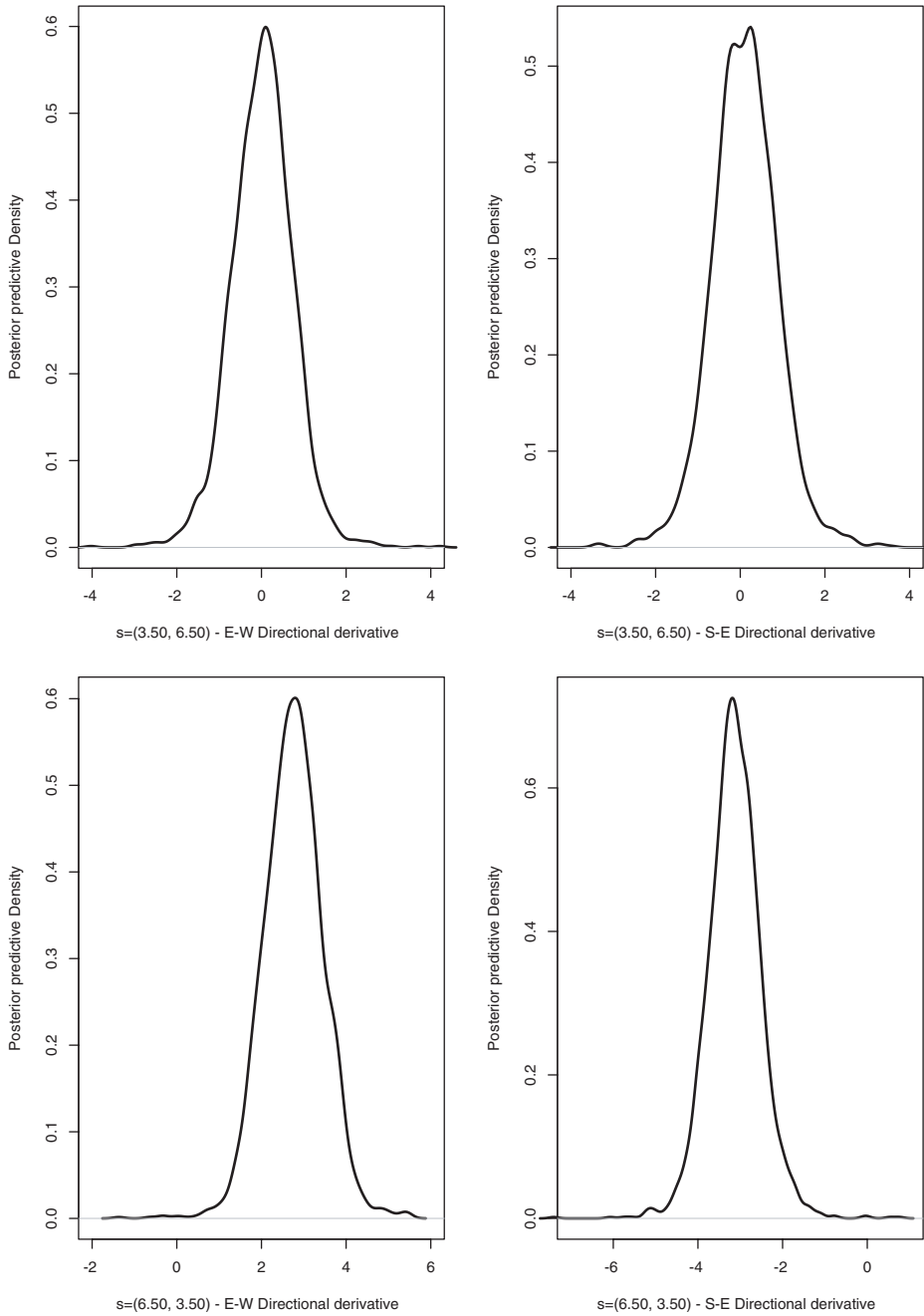
Fig. 6 Predictive posterior distribution of the directional derivatives at angles 0 and 135 and at sites $s_1^* = (3.5, 6.5)$ and $s_2^* = (6.5, 3.5)$ for the 100th replicate

Table 1 99th replicate - gradients at $s_1^* = (3.5, 6.5)$: the first column shows the values of the finite differences ($h = 0.01$) computed on the generated data; the second column shows the posterior median and predictive intervals for the density in Fig 3-above; all the other columns provide the posterior medians and (2.5%, 97.5%) predictive intervals for directional derivatives and finite differences ($h = 0.01, 0.1, 1$)

Angle	$m_{\mathbf{u},0.01}(\mathbf{s_1}^*)$ obs. at 99th	Quant. of $m_{\mathbf{u},0.01}(\mathbf{s_1}^*)$ obs. over all replic.	Posterior predictive (2.5%, 97.5%)			
			$D_{\mathbf{u}}m(\mathbf{s_1}^*)$	$m_{\mathbf{u},0.01}(\mathbf{s_1}^*)$	$m_{\mathbf{u},0.1}(\mathbf{s_1}^*)$	$m_{\mathbf{u},1}(\mathbf{s_1}^*)$
0	−2.80	−0.70(−3.45, 0.60)	−2.31(−3.78, −0.81)	−2.32(−3.79, −0.88)	−2.40(−3.76, −1.07)	−2.39(−3.01, −1.73)
45	0.60	−0.1(0.68, 0.70)	−0.87(−2.36, 0.62)	−0.88(−2.37, 0.56)	−1.00(2.43, 0.43)	−1.71(−2.53, 0.82)
90	3.71	0.62(−0.61, 3.66)	1.11(−0.5, 2.65)	1.09(−0.49, 2.64)	1.00(−0.5, 2.52)	−0.24(−1.01, 0.58)
135	4.64	0.76(−0.59, 4.7)	2.43(0.84, 3.9)	2.39(0.93, 3.97)	2.36(0.94, 3.85)	1.03(0.35, 1.83)
180	2.82	0.66(−0.55, 3.48)	2.31(0.81, 3.78)	2.30(0.87, 3.71)	2.21(0.91, 3.56)	0.43(−0.58, 1.51)
225	−0.60	0.07(−0.73, 0.64)	0.87(−0.62, 2.36)	0.83(−0.57, 2.37)	0.72(−0.71, 2.13)	−0.58(−1.68, 0.60)
270	−3.64	−0.56(−3.61, 0.62)	−1.11(−2.65, 0.50)	−1.11(−2.70, 0.49)	−1.18(−2.74, 0.36)	−1.50(−2.57, −0.39)
315	−4.56	−0.73(−4.70, 0.60)	−2.43(−3.90, −0.84)	−2.40(−3.94, −0.90)	−2.46(3.91, −1.02)	−2.21(−3.02, −1.42)

Table 2 100th replicate - gradients at $s_2^* = (6.5, 3.5)$: the first column shows the values of the finite differences ($h = 0.01$) computed on the generated data; the second column shows the posterior median and (2.5%, 97.5%) predictive intervals for the density in Fig 3-below; all the other columns provide the posterior medians and (2.5%, 97.5%) predictive intervals for directional derivatives and finite differences ($h = 0.01, 0.1, 1$)

Angle	$m_{\mathbf{u},0.01}(\mathbf{s}_2^*)$ obs. at 100 th	Quant. of $m_{\mathbf{u},0.01}(\mathbf{s}_1^*)$ obs. over all replic.	Posterior predictive (2.5%, 97.5%) intervals			
			$D_{\mathbf{u}}m(\mathbf{s}_2^*)$	$m_{\mathbf{u},0.01}(\mathbf{s}_2^*)$	$m_{\mathbf{u},0.1}(\mathbf{s}_2^*)$	$m_{\mathbf{u},1}(\mathbf{s}_2^*)$
0	2.87	2.45 (−0.43, 3.41)	2.76 (1.44, 4.08)	2.73 (1.37, 4.17)	2.71 (1.51, 4.09)	0.84 (−0.08, 1.84)
45	−0.59	−0.06 (−0.73, 0.57)	0.76 (−0.80, 2.40)	0.73 (−0.84, 2.49)	0.61 (−0.91, 2.20)	−0.90 (−1.82, 0.04)
90	−3.82	−2.58 (−3.7, 0.50)	−1.68 (−3.18, −0.10)	−1.73 (−3.24, −0.14)	−1.85 (−3.30, −0.34)	−2.09 (−2.94, −1.18)
135	−4.80	−3.61 (−4.84, 0.56)	−3.14 (−4.31, −1.89)	−3.15 (−4.37, −1.87)	−3.2 (−4.32, −2.03)	−2.49 (−3.20, −1.76)
180	−2.96	−2.43 (−3.51, 0.44)	−2.76 (−4.31, −1.44)	−2.75 (−4.19, −1.35)	−2.74 (−4.11, −1.41)	−2.09 (−2.85, −1.38)
225	0.60	0.02 (−0.57, 0.68)	−0.76 (−2.40, 0.80)	−0.74 (−2.51, 0.85)	−0.87 (−2.47, 0.69)	−1.32 (−2.37, −0.44)
270	3.70	2.51 (−0.57, 3.66)	1.68 (0.10, 3.18)	1.69 (0.08, 3.20)	1.52 (−0.07, 3.00)	−0.29 (−1.41, 0.75)
315	4.68	3.61 (−0.62, 4.81)	3.14 (1.88, 4.31)	3.14 (1.83, 4.35)	3.06 (1.93, 4.25)	0.93 (−0.05, 1.9)

of the Gibbs sampling procedure described in Gelfand et al. (2005) for model (12).If the base measure G_0 is Gaussian, then also $[V_{\mathbf{u},t}^*, \tilde{V}_{\mathbf{u},t}^* | \theta_t^*, \tilde{\theta}_t^*, \sigma^2, \phi, \nu]$ and $[\tilde{\theta}_t^* | \theta_t^*, \sigma^2, \phi, \nu]$ are gaussian. In fact, for any fixed $t = 1, \ldots, T^*$, the joint distribution $[\theta_t^*, \tilde{\theta}_t^*, V_{\mathbf{u},t}^*, \tilde{V}_{\mathbf{u},t}^*]$ is multivariate gaussian and has been thoroughly described by Banerjee et al. (2003). Hence, if we are interested in the predictive distribution only for the new set of locations, we can consider just

$$
\begin{aligned}
[\tilde{V}_{\mathbf{u},t}, t = 1, \ldots, T \,|\, \text{data}] &= [\tilde{V}_{\mathbf{u},t}^*, t = 1, \ldots, T^*, w \,|\, \text{data}] \\
&= \int \prod_{t=1}^{T^*} [\tilde{V}_{\mathbf{u},t}^* | \theta_t^*, \sigma^2, \phi, \nu] \, [\theta^*, w, T^*, \psi \,|\, \text{data}],
\end{aligned}
\tag{16}
$$

where $[\tilde{V}_{\mathbf{u},t}^* | \theta_t^*, \sigma^2, \phi, \nu]$ is again conditionally gaussian.When the base measure of the SDP is Gaussian an alternative method to obtain samples from $[V_{\mathbf{u},t}, \tilde{V}_{\mathbf{u},t}, t = 1, \ldots, T | \text{data}]$ is the following. Model (12) can be restated so that the conditional distribution of the observables at the first level of the hierarchy is parametrized by the vector of spatial gradients instead of the vector of spatial effects. It is sufficient to replace $Y_t | \beta, \theta_t, \tau^2$ with $Y_t | V_{\mathbf{u},t}, \psi$, where

$$
[Y_t | V_{\mathbf{u},t}, \psi] = \int [Y_t | \theta_t, \beta, \tau^2] \, [\theta_t | V_{\mathbf{u},t}, \sigma, \phi, \nu] \quad t = 1, \ldots, T,
$$

Since $[\theta_t, V_{\mathbf{u},t} | \sigma, \phi, \nu]$ is gaussian, then $[\theta_t | V_{\mathbf{u},t}, \sigma, \phi, \nu]$ is obtained by usual conditioning and it can be shown that

$$
Y_t | V_{\mathbf{u},t}, \psi \sim N_n \left(X^T \beta - \left[I + \frac{1}{\tau^2} \Lambda \right]^{-1} \Lambda K_2^{-1} K_1 V_{\mathbf{u},t}, \ \tau^2 \left[I + \frac{1}{\tau^2} \Lambda \right]^{-1} \right)
$$

where K_1 is the matrix of regression coefficients of θ_t on $V_{\mathbf{u},t}$, K_2 is the conditional variance, and $\Lambda = \left(\frac{1}{\tau^2} I_n + K_2^{-1} \right)^{-1}$ (see Harville (1997) and Anderson (2003)). Once we have reparametrized the model in terms of gradients, we can exploit the fact that $V_{\mathbf{u},t}(s)$ is a sample from a SDP (see Section 4) and apply the algorithm outlined by Gelfand et al. (2005) to obtain samples from the posterior distribution $[V_{\mathbf{u},t}, t = 1, \ldots, T, \psi \,|\, data]$ for old locations. Then, for new locations it is sufficient to consider the distinct $V_{\mathbf{u},t}$'s and the distribution $[\tilde{V}_{\mathbf{u},t}^* | V_{\mathbf{u},t}^*, \sigma^2, \phi, \nu]$, for $t = 1, \ldots, T^*$.

7 A Simulation Example

We present a simulation example with the intended purpose of stressing some features of our model. In particular, we consider a situation where usual Gaussian modelling is inappropriate. We take advantage of the clustering capability of the SDP to show the possibility of finding several types of spatial effects using the independent replicates.

Thus, we consider $Y(s) = Z(s) + \varepsilon(s)$, where $\varepsilon(s)$ is a pure error process with variance τ^2 and $Z(s)$ denotes a random field with distribution $F_Z(\cdot)$ given by

$$
F_Z(s) = \alpha F_{Z_1}(s) + (1 - \alpha) F_{Z_2}(s),
$$

for some $\alpha \in (0, 1)$, that is, a probability mixture of two independent Gaussian fields $Z_1(s)$ and $Z_2(s)$, with mean $E(Z_i(s)) = \eta_i(s)$ with respective covariance structure

specified by $\sigma^2 \rho_{\phi_i}(s, s')$, $i = 1, 2$. Data are generated so that each replicate could come from one of the two spatial processes, $Z_1(s)$ or $Z_2(s)$.

For example, usually temperature decreases with altitude. However, during the passage of a cold front or under overnight radiative cooling, temperature inversions occur and the temperature of the atmosphere increases with altitude. Temperature inversions are relevant, for example, in studying pollution phenomena, sound propagation, or thunderstorms. Therefore, it could be appropriate to use a non-unimodal model to describe the nature of the spatial effects acting at the different times when we collect data and we might be interested in the directions of the sharpest temperature gradients. As another example, suppose a signal comes from different sources, e.g., emission sites, at different times. We may not know either the locations of the sources or whether the sources are emitting at the time of collection. Then, we might seek both local intensity of the signal as well as local gradients to learn about both exposure and the direction of sources relative to the location. These are all situations where a simple Gaussian assumption for replicates is not appropriate. Our dataset can be seen as a simplified version of one of these settings.

Let

$$\eta_i(s) = \beta_{0i} + \beta_{1i} \, e^{-\psi_i \|s - s_{0i}\|^2}, \quad i = 1, 2 \tag{17}$$

where s_{01} and s_{02} are two distinct sources emitting a signal decaying to a long range mean level β_{0i} (possibly, a mean structure term depending on some covariates) as a function of the squared distance from the source. Thus, β_{1i} can be interpreted as coefficient of amplification of the signal and ψ_i is a decay parameter. Then,

$$E(Z(s)) = \alpha\beta_{01} + (1 - \alpha)\beta_{02} + \alpha\beta_{11} \, e^{-\psi_1(\|s - s_{01}\|)} + (1 - \alpha)\beta_{12} \, e^{-\psi_2(\|s - s_{02}\|)}, \tag{18}$$

and

$$\begin{aligned} Cov(Z(s), Z(s')) &= \alpha(1 - \alpha)[\eta_1(s') - \eta_2(s')][\eta_1(s) - \eta_2(s)] \\ &\quad + \sigma^2 \{\alpha \, \rho_{\phi_1}(s, s') + (1 - \alpha) \, \rho_{\phi_2}(s, s')\}. \end{aligned} \tag{19}$$

The field is observed on a randomly sampled set of points within a 10×10 square. In the subsequent illustration, we consider $n = 50$ sites, which are shown in Fig. 1, together with the location of the sources $s_{01} = (3, 7)$ and $s_{02} = (7, 3)$ (indicated with ✳). The maximum observed distance in our generated field is approximately 11.65 units. We consider 18 more points, which are excluded from the monitoring sites and will be used for validation of spatial prediction. These are the two points $s_1^* = (3.5, 6.5)$ and $s_2^* = (6.5, 3.5)$ (denoted by ▲ in Fig. 1), together with the 8 points around them, which are positioned along the eight main directions at angles of $0, 45, 90, 135, 180, 225, 270$ and 315 degrees and distant $h = 0.01$ units from the center.

We assume $T = 100$ independent observations of the random field $Y(s)$ specified above. In particular, we have chosen the following values for the parameters. The variance of the pure error component is $\tau^2 = 1$, while the parameters for $Z(s)$ are $\alpha = 0.5$, $\beta_{0i} = 2$, $\beta_{1i} = 5$, $i = 1, 2$. The value of $\psi = 1$ is such that the signal over the long range mean is expected to become negligible (i.e., $\beta_{01} \exp\{-\psi\|s - s_{0i}\|\} \leq 0.05$) at points further than 2.15 units from the sources. Here $\rho_\phi(\cdot)$ is the Matèrn correlation function with smoothness parameter $\nu = 3/2$, so that the covariance is

given by $\sigma^2 (1 + \phi d) \exp(-\phi d)$, with $\sigma^2 = 0.25$ and $\phi = 1.5$. That value ensures an effective isotropic range of 3.16 units.

In Fig. 2, we plot the histogram of the replications as well as the estimated density for two sites, one close to a source, one more distant. It is apparent that in a neighborhood of the sources, the distribution of the observations is bimodal. In fact, at each replicate they come from one of two distinct fields according to whether we sample from Z_1 or Z_2. At points far from the sources, the signal is negligible relative to its mean and thus, the density is unimodal and centered around this mean.

Based on observations collected at s_1^* and s_2^* and points around, we can plot the finite differences $m_{\mathbf{u},h}(s_i^*) = \theta_{\mathbf{u},h}(s_i^*)$, $i = 1, 2$ for $h = 0.01$. In Fig. 3, we show the distribution of the observed gradients (histograms and density estimates) at angles 0 and 135. It is immediate to see that it is bimodal, again reflecting the data generating mechanism. Careful analysis of the modes of the distribution also seems to suggest that, by chance, more of the data come from Z_2 rather than Z_1.

We suppose that the existence of the sources is unknown and fit the observations by means of the hierarchical model (12). That is, we assume $Y(s) = \mu + \theta(s) + \varepsilon(s)$, where μ is a constant and the distribution of $\theta(s)$ is centered around a mean zero stationary gaussian process, with Matèrn covariance function. The aim is to see if our model is able to capture the long range mean together with the spatial component specific to each replicate. In particular, we are interested in the ability to capture the gradient behavior typical of the spatial component operating at every single replicate.

We adopt a normal prior for μ, $\mu \sim N(0, 1)$, an $IG(2, 1)$ (mean $= 1$, infinite variance) for τ^2 and σ^2, a $G(2, .1)$ prior (mean $= 20$, variance $= 200$) for ϕ, and a uniform prior on $(1, 2)$ for ν. After fitting the model, we obtain samples from the posterior predictive distributions of the gradients for the 99th and 100th replicates. In particular, we are interested in prediction at points s_1^* and s_2^*, where we can compare the results either with Fig. 3 or the observed gradients in order to validate our inference.

In Fig. 4 we consider two exploratory image plots of data generated, respectively, at the 99th and 100th replicate. It is evident that the upper is a sample from Z_1 and the lower is from Z_2. This conjecture is confirmed by looking at Figs. 5 and 6, where for both replicates we report the posterior predictive distribution for the directional derivative in s_1^* and s_2^* at angles 0 and 135. For example, consider the 100th replicate (see Fig. 6). We can see that at site s_2^* there are a significant uphill gradient moving along the E-W direction and a significant downhill gradient toward S-W. On the other hand, the posterior predictive distributions are centered around zero in s_1^*, thus suggesting lack of significant gradients along those directions. This conclusion is actually confirmed when we look at all the other directions, so that we can conclude that s_1^* is located in an essentially flat portion of the region. Similar arguments can of course be developed to outline the characteristics of the directional derivative process at s_1^* and s_2^* for the 99th replicate (Fig. 5).

Tables 1 and 2 provide a summary of the posterior predictive distributions of directional finite differences ($h = 0.01, 0.1$ and 1.0) and derivatives at the angles of $0, 45, 90, 135, 180, 225$ and 270 degrees. Those are compared with the gradients computed on the basis of the data generated at sites s_1^* and s_2^*, as specified above. Even more important, if we refer to Fig. 1, we see that source 1 is northwest of site s_1^* while source 2 is southeast of site s_2^*. Hence gradients in those directions would

be expected to be much larger than in say the northeast or southwest directions. Tables 1 and 2 show the former to be very significant, the latter to be insignificant.

In conclusion, by means of the preceding analysis we are able to detect the presence of distinct gradient behaviors across the replicates. The usual assumption of a single Gaussian process for $\theta(s)$ in (11) would lead to smoothing of the estimated gradient processes through all the replicates. Such smoothing is also present in our estimates, but to a lesser degree and only for certain directions, depending on the number of sites sampled and whether the magnitude of the ε's hides the gradient behavior of the mean process $m(s)$.

8 Summary and Extensions

Historically, in working with data from a spatial process, interest has focused on global behavior of the realized surface, e.g., on the estimated mean surface and associated uncertainty. Recently, there has been increasing interest in the local behavior of such surfaces, captured through finite differences and infinitesimal gradients. All of this work has proceeded in the context of Gaussian processes. Here, we have shed such restriction, working with spatial Dirichlet processes, processes that are nonGaussian, nonstationary with nonhomogeneous variances. Understanding smoothness properties and developing gradient theory for surfaces resulting from sampling such processes requires careful analysis as we have provided in Section 3. Computation and data analysis to infer about gradient behavior is demanding as we have illuminated. Future work in this area will lead to the extension of our results to very recently proposed generalized SDP's (see Duan et al., 2005). Other natural extensions include gradient analysis for nonparametric space-time models and for multivariate models.

Appendix

Proof of Proposition 1: The proof follows after noticing that for all $s, s_0 \in D$ and for all $A \in \mathcal{B}$

$$G(Y(s) - Y(s_0) \in A) = \sum_{j=1}^{\infty} p_j \, \delta_{\theta_j^*(s) - \theta_j^*(s_0)}(A), \quad a.s.\text{-}\mathcal{P}, \tag{20}$$

that is $Y(s) - Y(s_0)$ is a sample from a SDP, with smoothness parameter ν and base measure the distribution of $\theta_1^*(s) - \theta_1^*(s_0)$. Recalling the properties of the DP (see Proposition 4 (Ferguson, 1973)), we get $\mathcal{P}(A) = E(G(A)) = P_{G_0}(A)$, where the expectation is taken marginalizing with respect to G. Now consider the set C_{s_0} as defined in the text.

If $G(C_{s_0}) = 1$ for all G in a set of $\mathcal{P}$-measure one, of course $E(G(C_{s_0})) = 1$ and from the previous identities we get that G_0 is a.s. continuous.

Conversely, if G_0 is a.s. continuous, $E(G(C_{s_0})) = 1$. Now, $G(C_{s_0})$ is a random probability measure when considered as a function of G; therefore, it is nonnegative. It follows that $G(C_{s_0}) = 1$ a.s. with respect to $\mathcal{P}$. ∎

Proof of Proposition 2:   For an arbitrary set $A \in \mathcal{B}$, let $P(Y(s) \in A|G) \equiv G(Y(s) \in A)$ be the probability under G that a sample takes values in A at a site s. Then, consider the set

$$W_{s_0} = \{G : \lim_{||s-s_0|| \to 0} P(Y(s) \in A|G) = P(Y(s_0) \in A|G)\},$$

i.e., the set of the probability measures which are convergent in s_0. For any fixed G in S, we have

$$\lim_{||s-s_0|| \to 0} G(Y(s) \in A) = \sum_{j=1}^{\infty} p_j \lim_{||s-s_0|| \to 0} \delta_{\theta_j^*(s)}(A)$$

$$= \sum_{j=1}^{\infty} p_j \, \delta_{\theta_j^*(s_0)}(A) = G(Y(s_0) \in A), \quad a.s.\text{-}G_0$$

where the first equality follows easily from the dominated convergence theorem. Notice that the previous result is obtained regardless of the particular realization of the vector of weights $p = (p_1, p_2, \ldots)$. Therefore, we can conclude that $S \cap W_{s_0}$ is indeed of the form $S \cap W_{s_0} = P \times E_{s_0}^{\infty}$, where P is the support of the GEM distribution defined on the weights and $E_{s_0} = \{\theta_D^* : \lim_{||s-s_0|| \to 0} \theta^*(s) = \theta^*(s_0)\}$. If we denote with $\pi(p)$ the distribution of p and exploit the independence structure of the SDP measure, we get

$$P(S \cap W_{s_0}) = \int_{P \times E_{s_0}^{\infty}} \pi(p) \, G_0^{\infty}(d\theta)$$

$$= \int_P \pi(p) \int_{E_{s_0}} G_0(d\theta) = 1,$$

since E_{s_0} has measure 1 w.r.t. G_0 by hypothesis.   ∎

Proof of Proposition 3:   Let s_0 be an arbitrary point in K and consider a realization of the random probability measure G. We know that $G \in S$ with probability 1. Therefore, for all $s \in D$, $E[(Y(s) - Y(s_0))^2|G] = \sum_{j=1}^{\infty} p_j (\theta_j^*(s) - \theta_j^*(s_0))^2$, and we need to prove that

$$\lim_{||s-s_0|| \to 0} E[(Y(s) - Y(s_0))^2|G] = \lim_{||s-s_0|| \to 0} \sum_{j=1}^{\infty} p_j (\theta_j^*(s) - \theta_j^*(s_0))^2 = 0, \quad a.s.\text{-}\mathcal{P}. \quad (21)$$

The base measure is almost surely continuous by hypothesis. Therefore, all we have to prove is the admissibility of the interchange between sum and limit operations in (21) by Lebesgue dominated convergence theorem.   Consider the process $Z = \max_{s \in K} |\theta^*(s)|$. Since G_0 is a.s. continuous on the compact K, then Z is a.s. bounded on K, therefore integrable. Analogously, $Z^2 = (\max_{s \in K} \theta^*(s))^2$ is a.s. bounded and integrable on K. It follows that

$$E(Z^2|G) = \sum_{j=1}^{\infty} p_j (\max_{s \in K} |\theta_j^*(s)|)^2 < \infty,$$

by Theorem 3 in Ferguson (1973). Now consider an arbitrary term in the sum representation (21). It's immediate to show that

$$(\theta_j^*(s) - \theta_j^*(s_0))^2 \leq \max_{s \in K}(\theta_j^*(s) - \theta_j^*(s_0))^2 \leq 4\,Z^2,$$

for all $j = 1, 2, \ldots$, by Minkowski inequality. Therefore, any term of the sum (21) is bounded by the integrable function Z^2 and the conditions for the dominated convergence theorem are satisfied. ∎

Proof of Proposition 4: First, notice that $g(Y(s))$ is a sample from a DP, with parameter ν and base measure $G_{0,s}^g$, i.e., we can define $G_s(g) = \sum_{j=1}^{\infty} p_j \delta_{g(\theta_j^*(s))}(\cdot)$ and $g(Y(s))|G_s(g) \sim G_s(g)$ and $G_s(g) \sim G_{0,s}^g$. By the usual properties of the Dirichlet process, it follows that if $G_{0,s}^g$ converges weakly, then also $g(Y(s))$ converges in distribution (with respect to the collection of Dirichlet measures defined at each $s \in D$). Therefore, all we need to establish is that the limit is indeed a sample from a DP with parameter ν and base measure H_{0,s_0}. In other words, we need to prove that $G_s(g)$ converges in distribution to the random probability measure $\widetilde{G} = \sum_{j=1}^{\infty} p_j \delta_{\xi_i(s_0)}(\cdot)$, where $\xi_i(s_0) \overset{i.i.d.}{\sim} H_{0,s_0}$.

In order to prove this result, we use the characteristic function method illustrated in Ishwaran and Zarepour (2002).

We start recalling that Sethuraman (1994) proves that $G \sim DP(\nu G_0)$ is the unique solution satisfying the following distributional equation

$$G \overset{\mathcal{D}}{=} q_1 \delta_{\theta_1^*(s)} + (1 - q_1)\,G, \tag{22}$$

where, on the right hand side, q_1 has distribution $Beta(1, \nu)$, $\theta_1^*(s)$ is independent of q_1 and G is independent of $(q_1, \theta_1^*(s))$.

Now, let $\psi(t, s) = E\{\exp itG(g, s)\}$ denote the characteristic function of $G_s(g)$. By (22), it follows that

$$\psi(t, s) = E\{\exp\{it[q_1\,g(\theta_1^*(s)) + (1 - q_1)\,G_s(g)]\}\}. \tag{23}$$

Analogously, let $\phi(t, s)$ denote the characteristic function of $g(\theta_1^*(s))$. Then, we can exploit the independence relations in order to obtain

$$\psi(t, s) = E_{q_1}\{\phi(t\,q_1, s)\,\psi(t\,(1 - q_1), s)\}, \tag{24}$$

where the expected value is taken with respect to the distribution of q_1. Any characteristic function satisfying (24) must be the characteristic function for $G_s(g)$.

Now consider $\psi^*(t, s_0) = \lim_{\|s - s_0\| \to 0} \psi(t, s)$, and $\phi^*(t, s_0) = \lim_{\|s - s_0\| \to 0} \phi(t, s)$. Since $g(\theta_1^*(s))$ converges in distribution to $\xi(s_0)$, $\phi^*(t, s_0)$ is the characteristic function of $\xi(s_0)$.

Therefore, since $|\phi(t, s)| \leq 1$ and $|\psi(t, s)| \leq 1$, for all $s \in D$, we can apply the bounded convergence theorem in order to get

$$\psi^*(t, s_0) = E_{q_1}\{\phi^*(t\,q_1, s_0)\,\psi^*(t\,(1 - q_1), s_0)\},$$

which is the characteristic function of $\widetilde{G}$, by the uniqueness of the solution of (24).

Then, we have proved that $g(Y(s))$ converges to a random variable, say $Z(s_0)$, such that $Z(s_0)|\widetilde{G} \sim \widetilde{G}$ and $\widetilde{G} \sim DP(\nu H_{0,s_0})$. ∎

Proof of Proposition 5:   The proof mimics that of Proposition 3. In fact, let $h_n, h_m > 0$ be arbitrary scalars, and $s_0 \in \dot{K}$. Then,

$$E\left[(Y_{\mathbf{u},h_n}(s_0) - Y_{\mathbf{u},h_m}(s_0))^2 | G\right] = \sum_{j=1}^{\infty} p_j \left(\theta^*_{\mathbf{u},h_n}(s_0) - \theta^*_{\mathbf{u},h_m}(s_0)\right)^2 \quad a.s.\text{-}\mathcal{P}. \qquad (25)$$

Since $\theta^*_j(s) \in C^1$, $\lim_{h_n,h_m \to 0}(\theta^*_{\mathbf{u},h_n}(s_0) - \theta^*_{\mathbf{u},h_m}(s_0))^2 = 0$ and $\theta^*_j(s_0)$ has directional derivative $D_{\mathbf{u}}\theta^*_j(s_0) = \mathbf{u}^T \nabla_{\theta^*}(s_0)$. Therefore, $Y_{\mathbf{u},h_n}(s_0)$ converges in L^2 with respect to G to a random variable $D_{\mathbf{u}}Y(s_0) = \mathbf{u}^T \nabla_Y(s_0)$, whose distribution is $\sum_{j=1}^{\infty} p_j \delta_{D_{\mathbf{u}}\theta^*_j(s_0)}(\cdot)$, since

$$\lim_{h_n \to 0} \sum_{j=1}^{\infty} p_j \left(\theta^*_{\mathbf{u},h_n}(s_0) - \mathbf{u}^T \nabla_{\theta^*}(s_0)\right)^2 = \sum_{j=1}^{\infty} p_j \lim_{h_n \to 0} \left(\theta^*_{\mathbf{u},h_n}(s_0) - \mathbf{u}^T \nabla_{\theta^*}(s_0)\right)^2,$$

the interchange between limits and sum being justified by the existence of the directional derivatives and the continuity hypotheses.   ∎

References

M. Abramowitz, and I. A. Stegun, *Handbook of Mathematical Functions*, Dover: New York, 1965.

R. J. Adler, and J. E. Taylor, "Random fields and their geometry," Unpublished, 2003.

T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, third edition, Wiley: Hoboken, NJ, 2003.

S. Banerjee, and A. E. Gelfand, "On smoothness properties of spatial processes," *Journal of Multivariate Analysis* vol. 84 pp. 85–100, 2003.

S. Banerjee, and A.E. Gelfand, "Curvilinear boundary analysis under spatial process models," *Journal of the American Statistical Association*, 2005. (in press).

S. Banerjee, A. E. Gelfand, and C. F. Sirmans, "Directional rates of change under spatial process models," *Journal of the American Statistical Association* vol. 98 pp. 946–954, 2003.

C. A. Bush, and S. N. MacEachern, "A semiparametric Bayesian model for randomised block designs," *Biometrika* vol. 83 pp. 275–285, 1996.

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

V. De Oliveira, B. Kedem, and D. A. Short, "Bayesian prediction of transformed Gaussian random fields" *Journal of the American Statistical Association* vol. 96 pp. 1361–1374, 1997.

P. J. Diggle, J. A. Tawn, and R. A. Moyeed, "Model-based geostatistics (with discussion)," *Applied Statistics* vol. 47 pp. 299–350, 1998.

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

M. D. Escobar, "Estimating normal means with a Dirichlet process prior," *Journal of the American Statistical Association* vol. 89 pp. 268–277, 1994.

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

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

M. Fuentes, and R. L. Smith, "A new class of nonstationary spatial models," Technical report, Department of Statistics, North Carolina State University, 2001.

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

A. E. Gelfand, A. Kottas, and S. N. MacEachern, "Bayesian nonparametric spatial modeling with Dirichlet processes mixing," *Journal of the American Statistical Association* vol. 100 pp. 1021–1035, 2005.

S. Ghosal, J. K. Ghosh, and R. V. Ramamoorthi, "Posterior consistency of Dirichlet mixture in density estimation," *The Annals of Statistics* vol. 27 pp. 143–158, 1999.

M. S. Handcock, and M. Stein, "A Baesian analysis of kriging," *Technometrics* vol. 35(4) pp. 403–410, 1993.

D. A. Harville, Matrix Algebra from a Statistician's Viewpoint, Springer Series in Statistics, Springer-Verlag: New York, 1997.

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

H. Ishwaran, and M. Zarepour, "Exact and approximate sum-representations for the Dirichlet process," *Canadian Journal of Statistics* vol. 30 pp. 269–283, 2002.

J. T. Kent, "Continuity properties of random fields," *Annals of Probability* vol. 17 pp. 1432–1440, 1989.

S. N. MacEachern, Dependent Dirichlet processes. Technical report, Department of Statistics, The Ohio State University, 2000.

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

A. Majumdar, H. Munneke, S. Banerjee, A.E. Gelfand, and F. Sirmans, "Gradients in spatial response surfaces, land value gradients, and endogenous central business districts," *Submitted to The Journal of Business and Economic Studies*, 2004.

B. Matern, *Spatial Variation*, 2nd edition, Springer Verlag: Berlin, 1986.

M. B. Palacios, and M. F. J. Steel, "Non-gaussian Bayesian geostatistical modeling," Warwick Statistics Research report 426, University of Warwick, 2004.

P. D. Sampson, and P. Guttorp, "Nonparametric estimation of nonstationary spatial covariance structure," *Journal of the American Statistical Association* vol. 87(417) pp. 108–119, 1992.

A. M. Schmidt, and A. O'Hagan, "Bayesian inference for non-stationary spatial covariance structure via spatial deformations," *Journal of the Royal Statistical Society. Series B, Statistical methodology* vol. 65(3) pp. 743–758, 2003.

J. Sethuraman, "A constructive definition of Dirichlet priors," *Statistica Sinica* vol. 4 pp. 639–650, 1994.

M. L. Stein, *Interpolation of Spatial Data—Some theory of Kriging*, Springer: New York, 1999.